Tests non paramétriques minimax pour de grandes matrices de covariance

Rania Zgheib

UNIVERSITÉ PARIS-EST

École Doctorale Mathématiques et Sciences et Technologie de l'Information et de la Communication (MSTIC)

Thèse de doctorat

Discipline : Mathématiques

Présentée par

Rania ZGHEIB

Tests non paramétriques minimax pour de grandes matrices de covariance

---

Tests non paramétriques minimax pour de grandes matrices de covariance: vitesses minimax de séparation, équivalents asymptotiques exacts et adaptation

Ces travaux contribuent à la théorie des tests non paramétriques minimax dans le modèle de grandes matrices de covariance. Plus précisément, nous observons n vecteurs indépendants, de dimension p, $X_1, \dots, X_n$, ayant la même loi gaussienne $\mathcal{N}_p(0, \Sigma)$, où Σ est la matrice de covariance inconnue. Nous testons l'hypothèse nulle $H_0 : \Sigma = I$, où I est la matrice identité. L'hypothèse alternative est constituée d'un ellipsoïde avec une boule de rayon φ autour de I enlevée. Asymptotiquement, n et p tendent vers l'infini. La théorie minimax des tests, les autres approches considérées pour le modèle de matrice de covariance, ainsi que le résumé de nos résultats font l'objet de l'introduction.

Le deuxième chapitre est consacré aux matrices de covariance Σ de Toeplitz. Le lien avec le modèle de densité spectrale est discuté. Nous considérons deux types d'ellipsoïdes, décrits par des pondérations polynomiales (dits de type Sobolev) et exponentielles, respectivement. Dans les deux cas, nous trouvons les vitesses de séparation minimax. Nous établissons également des équivalents asymptotiques exacts de l'erreur minimax de deuxième espèce et de l'erreur minimax totale. La procédure de test asymptotiquement minimax exacte est basée sur une U-statistique d'ordre 2 pondérée de façon optimale.

Le troisième chapitre considère une hypothèse alternative de matrices de covariance pas nécessairement de Toeplitz, appartenant à un ellipsoïde de type Sobolev de paramètre α. Nous donnons des équivalents asymptotiques exacts des erreurs minimax de 2ème espèce et totale. Nous proposons une procédure de test adaptative, c-à-d libre de α, quand α appartient à un compact de $(1/2, +\infty)$.

L'implémentation numérique des procédures introduites dans les deux premiers chapitres montrent qu'elles se comportent très bien pour de grandes valeurs de p, en particulier elles gagnent beaucoup sur les méthodes existantes quand p est grand et n petit.

Le quatrième chapitre se consacre aux tests adaptatifs dans un modèle de covariance où les observations sont incomplètes. En effet, chaque coordonnée du vecteur est manquante de manière indépendante avec probabilité $1 - a$, $a \in (0, 1)$, où a peut tendre vers 0. Nous traitons ce problème comme un problème inverse. Nous établissons ici les vitesses minimax de séparation et introduisons de nouvelles procédures adaptatives de test. Les statistiques de test définies ici ont des poids constants. Nous considérons les deux cas: matrices de Toeplitz ou pas, appartenant aux ellipsoïdes de type Sobolev.

Non parametric minimax tests for high dimensional covariance matrices : minimax separation rate, sharp asymptotics and adaptation.

Our work contributes to the theory of non-parametric minimax tests for high dimensional covariance matrices. More precisely, we observe n independent, identically distributed vectors of dimension p, $X_1, \ldots, X_n$ having Gaussian distribution $\mathcal{N}_p(0, \Sigma)$, where Σ is the unknown covariance matrix. We test the null hypothesis $H_0 : \Sigma = I$, where I is the identity matrix. The alternative hypothesis is given by an ellipsoid from which a ball of radius φ centered in I is removed. Asymptotically, n and p tend to infinity. The minimax test theory, other approaches considered for testing covariance matrices and a summary of our results are given in the introduction.

The second chapter is devoted to the case of Toeplitz covariance matrices Σ. The connection with the spectral density model is discussed. We consider two types of ellipsoids, describe by polynomial weights and exponential weights, respectively. We find the minimax separation rate in both cases. We establish the sharp asymptotic equivalents of the minimax type II error probability and the minimax total error probability. The asymptotically minimax test procedure is a U-statistic of order 2 weighted by an optimal way.

The third chapter considers alternative hypothesis containing covariance matrices not necessarily Toeplitz, that belong to an ellipsoid of parameter α. We obtain the minimax separation rate and give sharp asymptotic equivalents of the minimax type II error probability and the minimax total error probability. We propose an adaptive test procedure free of α, for α belonging to a compact of $(1/2, +\infty)$.

We implement the tests procedures given in the previous two chapters. The results show their good behavior for large values of p and that, in particular, they gain significantly over existing methods for large p and small n.

The fourth chapter is dedicated to adaptive tests in the model of covariance matrices where the observations are incomplete. That is, each value of the observed vector is missing with probability $1-a$, $a \in (0,1)$ and a may tend to 0. We treat this problem as an inverse problem. We establish the minimax separation rates and introduce new adaptive test procedures. Here, the tests statistics are weighted by constant weights. We consider ellipsoids of Sobolev type, for both cases : Toeplitz and non Toeplitz matrices.

Contents

Chapter 1

Introduction

Dans ce chapitre, nous rappelons la théorie des tests et, plus en détail, la théorie des tests minimax et adaptatifs non paramétriques. Ensuite, nous donnons un aperçu des travaux établis dans la littérature, sur les tests minimax pour différents modèles statistiques et sur les tests pour les matrices de covariance, qui font l'objet de cette thèse. Enfin, nous détaillons les résultats obtenus dans la thèse, sur des tests minimax et adaptatifs, pour des grandes matrices de covariance.

1.1 Formalisme des tests minimax non paramétriques

Soit une expérience statistique $(\Omega, \mathcal{A}, \{P_\Sigma, \Sigma \in \mathcal{F}\})$, où $(\Omega, \mathcal{A})$ est un espace mesurable et P_Σ une mesure de probabilité sur $\mathcal{A}$, avec Σ inconnu et appartenant à l'ensemble $\mathcal{F}$. Soient $\mathcal{F}_0$ et $\mathcal{F}_1$ deux sous-ensembles de $\mathcal{F}$, tels que $\mathcal{F}_0 \cap \mathcal{F}_1 = \varnothing$ et $\mathcal{F}_0 \cup \mathcal{F}_1 \subseteq \mathcal{F}$. Le problème de test se présente de la manière suivante: dans un premier temps, nous énonçons deux hypothèses sur le paramètre inconnu

$$H_0 : \Sigma \in \mathcal{F}_0 \quad \text{contre} \quad H_1 : \Sigma \in \mathcal{F}_1 \,, \tag{1.1}$$

où H_0 est l'hypothèse la plus plausible, appelée hypothèse nulle, et H_1 l'hypothèse alternative. Ensuite, nous devons construire une procédure de test, c.à.d une fonction mesurable des observations à valeurs dans $[0, 1]$, qui permet de choisir entre ces deux hypothèses.

On distingue deux types de test, les *tests randomisés* et les *tests non randomisés*. Un test randomisé Δ est un test à valeurs dans $(0, 1)$, qui accepte l'hypothèse H_0 avec une probabilité $1 - \Delta$ et rejette H_0 avec une probabilité Δ. Un test non randomisé Δ est un test qui ne prend que les valeurs 0 ou 1, il accepte l'hypothèse H_0 quand il prend la valeur 0, et rejette H_0 quand il prend la valeur 1. Pour la suite nous ne considérons que des tests non randomisés.

Quand $\mathcal{F}_0 = \{\Sigma_0\}$, donc réduit à un élément, on parle d'*hypothèse simple*, dans le cas contraire on parle d'*hypothèse composite*. L'ensemble $\mathcal{F}$ peut être un intervalle de $\mathbb{R}$ ($\mathcal{F} \subseteq \mathbb{R}$), ou un ensemble de vecteurs ($\mathcal{F} \subseteq \mathbb{R}^d$) ou un ensemble de matrices ($\mathcal{F} \subseteq \mathbb{R}^{d\times d}$). Quand $\mathcal{F}$ est inclus dans un espace vectoriel de dimension finie, le problème de test est dit *paramétrique*. Par contre, si $\mathcal{F}$ est inclus dans un espace vectoriel de Hilbert ou de Banach de dimension infinie, (comme par exemple un espace de fonctions régulières, un espace de suites de carrés sommables, etc), le test est dit *non paramétrique*.

Pour un test Δ donné, on associe deux types d'erreurs, qui permettent d'évaluer sa qualité:

- *l'erreur de première espèce* est la fonction définie sur $\mathcal{F}_0$ par:

$$\eta(\Delta, \Sigma_0) = P_{\Sigma_0}(\Delta = 1) = \mathbb{E}_{\Sigma_0}(\Delta), \quad \text{pour tout } \Sigma_0 \in \mathcal{F}_0$$

qui représente la probabilité que le test choisisse l'hypothèse H_1 alors que la vraie valeur de Σ appartient à $\mathcal{F}_0$.

- *l'erreur de deuxième espèce* est la fonction définie sur $\mathcal{F}_1$ par:

$$\beta(\Delta, \Sigma_1) = P_{\Sigma_1}(\Delta = 0) = \mathbb{E}_{\Sigma_1}(1 - \Delta), \quad \text{pour tout } \Sigma_1 \in \mathcal{F}_1,$$

qui représente la probabilité que le test choisisse l'hypothèse H_0 alors que la vraie valeur de Σ appartient à $\mathcal{F}_1$. Souvent, au lieu de l'erreur de deuxième espèce, on utilise la fonction *puissance* définie sur $\mathcal{F}_1$ par:

$$\delta(\Delta, \Sigma_1) = 1 - \beta(\Delta, \Sigma_1), \quad \text{pour tout } \Sigma_1 \in \mathcal{F}_1.$$

D'après ces définitions, nous voyons que le but du problème est de trouver un test pour lequel les erreurs de première et deuxième espèce soient minimales ou, de manière équivalente, trouver un test dont son erreur de première espèce soit minimale et sa puissance soit maximale. Si les deux hypothèses sont simples, les erreurs sont des quantités réelles et dans ce cas la solution du problème est donnée par le lemme de Neyman-Pearson, voir [68]. Par contre, si au moins l'une des deux hypothèses est composite, alors pour comparer les tests entre eux, il faut comparer les erreurs de chaque test entres elles.

Comme, dans toute la suite, on traite des problèmes de test avec une hypothèse nulle simple et des hypothèses alternatives composites, nous allons dès à présent restreindre l'étude à ce cas. Donc pour la suite nous considérons $\mathcal{F}_0 = \{\Sigma_0\}$.

Une méthode possible pour comparer les tests, est de définir une relation d'ordre partiel sur l'ensemble de tous les tests possibles $\boldsymbol{\Delta}$. Soient $\Delta_1, \Delta_2 \in \boldsymbol{\Delta}$, la relation d'ordre $\geq$ définie par:

$$\Delta_1 \geq \Delta_2 \Leftrightarrow \begin{cases} \eta(\Delta_1, \Sigma_0) \leq \eta(\Delta_2, \Sigma_0) \quad \text{et} \\ \beta(\Delta_1, \Sigma_1) \leq \beta(\Delta_2, \Sigma_1) \quad \forall \Sigma_1 \in \mathcal{F}_1 \end{cases}$$

Alors, le problème se formalise de la façon suivante, trouver un test $\Delta^\star$ tel que $\Delta^\star \geq \Delta$ pour tout test $\Delta \in \boldsymbol{\Delta}$. Cependant de tels tests n'existent que dans des conditions très particulières. D'où le besoin de développer de nouvelles méthodes qui permettent de comparer plusieurs tests.

Nous distinguons deux types d'approches. Les approches locales se limitent à l'étude de l'erreur de deuxième espèce en une suite d'éléments fixés de l'alternative. Les approches globales cherchent à contrôler l'erreur de deuxième espèce uniformément, pour tous les éléments de l'alternative. Parmi ces dernières, nous trouvons l'approche minimax, qui est choisie pour les problèmes de tests considérés dans ce travail de thèse.

1.1.1 Théorie minimax

La théorie minimax consiste à contrôler la pire erreur possible sous l'alternative. Donc la qualité d'un test Δ est mesurée par les quantités suivantes:
- *l'erreur de première espèce*

$$\eta(\Delta) = P_{\Sigma_0}(\Delta = 1) = \mathbb{E}_{\Sigma_0}(\Delta),$$

- *l'erreur maximale de deuxième espèce*

$$\beta(\Delta, \mathcal{F}_1) = \sup_{\Sigma \in \mathcal{F}_1} P_\Sigma(\Delta = 0) = \sup_{\Sigma \in \mathcal{F}_1} \mathbb{E}_\Sigma(1 - \Delta).$$

En général, le rôle des deux erreurs n'est pas symétrique, le problème est traité en privilégiant l'hypothèse nulle par rapport à l'hypothèse alternative. Il existe dans la littérature deux approches qui traduisent ce privilège.

L'approche de Neyman-Pearson consiste à fixer un seuil pour l'erreur de première espèce, et trouver un test qui minimise l'erreur de deuxième espèce sous cette contrainte. Autrement dit, pour un η fixé dans $(0,1)$, il faut trouver un test Δ_η qui, vérifie:

$$\beta(\Delta_\eta, \mathcal{F}_1) = \inf_{\chi; \eta(\chi) \leq \eta} \beta(\chi, \mathcal{F}_1). \tag{1.2}$$

Dans ce cas on dit que le test Δ_η est le plus puissant parmi tous les tests dont l'erreur maximale de première espèce est inférieure à η.

La deuxième approche est basée sur la somme des erreurs, qui est définie par :

$$\gamma(\Delta, \mathcal{F}_1) = \eta(\Delta) + \beta(\Delta, \mathcal{F}_1), \tag{1.3}$$

pour Δ un test donné. On appelle $\gamma(\Delta, \mathcal{F}_1)$ *erreur totale* du test Δ. Cette approche consiste à trouver un test qui minimise l'erreur totale sur l'ensemble de tous les tests possibles. Donc, nous cherchons un test Δ qui vérifie

$$\gamma_(\Delta, \mathcal{F}_1) = \inf_\chi \ \gamma(\chi, \mathcal{F}_1) = \inf_\chi \Big(\eta(\chi) + \beta(\chi, \mathcal{F}_1)\Big)$$

où l'infimum est pris sur toutes les procédures de test χ possibles. Notons qu'il est montré dans [58] que ces deux approches sont sont liées par la relation

$$\inf_{\chi} \gamma(\chi, \mathcal{F}_1) = \inf_{\eta \in (0,1)} \Big(\eta + \inf_{\chi; \eta(\chi) \le \eta} \beta(\chi, \mathcal{F}_1) \Big).$$

Pour pouvoir construire un test qui minimise l'erreur maximale de deuxième espèce (1.2) ou l'erreur totale (1.3), il faut que les deux ensembles d'hypothèses soient séparés. Soit d une distance sur $\mathcal{F}$; on définit

$$\bar{\mathcal{B}}_{\mathcal{F}}(\varphi) := \bar{\mathcal{B}}_{\mathcal{F}}(\Sigma_0, \varphi) = \{\Sigma \in \mathcal{F};\ d(\Sigma, \Sigma_0) \ge \varphi\},$$

le complémentaire dans $\mathcal{F}$ muni de la métrique d, de la boule centrée en Σ_0 et de rayon φ. Par conséquent, nous remplaçons l'alternative $\mathcal{F}_1$ par $\mathcal{F}_1 \cap \bar{\mathcal{B}}_{\mathcal{F}}(\varphi)$. Nous notons $\beta(\chi, \mathcal{F}_1, \varphi)$ l'erreur maximale de deuxième espèce et $\gamma(\chi, \mathcal{F}_1, \varphi)$ l'erreur totale d'un test χ, sous cette nouvelle alternative . Suite à la définition de cette nouvelle alternative, un nouveau problème concernant le paramètre φ, qui va dépendre du nombre d'observations n et qu'on notera par φ_n, se pose: trouver la suite $\tilde{\varphi}_n$ qui définit la vitesse de séparation minimax asymptotique.

Approche minimax asymptotique

Définition 1.1. *Pour η fixé dans $(0,1)$, l'erreur minimax de deuxième espèce est définie par*

$$\beta_\eta(\varphi_n) := \inf_{\chi; \eta(\chi) \le \eta} \beta(\chi, \mathcal{F}_1, \varphi_n),$$

où l'infimum est pris sur toutes les procédures de test χ tel que $\eta(\chi) \le \eta$.

L'erreur totale minimax est définie par

$$\gamma(\varphi_n) := \inf_{\chi} \gamma(\chi, \mathcal{F}_1, \varphi_n) = \inf_{\chi} \Big(\eta(\chi) + \beta(\chi, \mathcal{F}_1, \varphi_n) \Big)$$

où l'infimum est pris sur toutes les procédures de test χ possibles.

Définition 1.2. *Une suite $\tilde{\varphi}_n$ est appelée vitesse minimax de séparation si elle vérifie que*

-(Borne inférieure) pour toute suite φ_n telle que $\varphi_n / \tilde{\varphi}_n \to 0$, on a:

$$\gamma(\varphi_n) \underset{n \to +\infty}{\longrightarrow} 1,$$

(ou, pour $\eta \in (0,1)$ fixé, vérifie: $\beta_\eta(\varphi_n) \underset{n \to +\infty}{\longrightarrow} 1 - \eta$).

-(Borne supérieure) pour tout suite φ_n telle que $\varphi_n / \tilde{\varphi}_n \to +\infty$, il existe un test Δ qui vérifie

$$\gamma(\Delta, \mathcal{F}_1, \varphi_n) = \eta(\Delta) + \beta(\Delta, \mathcal{F}_1, \varphi_n) \underset{n \to +\infty}{\longrightarrow} 0,$$

(respectivement, pour un η fixé dans $(0,1)$, trouver un test Δ_η tel que $\eta(\Delta_\eta) \le \eta$ et $\beta(\Delta_\eta, \mathcal{F}_1, \varphi_n) \underset{n \to +\infty}{\longrightarrow} 0$). On dit alors que le test Δ (respectivement Δ_η) est asymptotiquement minimax.

En d'autres termes $\tilde{\varphi}_n$ est vue comme la frontière qui sépare la région où il existe une procédure de test dont l'erreur totale tend vers 0, de la région où toute procédure aura une erreur totale qui tend vers 1.

Définition 1.3. *On dit que l'erreur minimax de deuxième espèce (l'erreur totale minimax) possède un équivalent asymptotique exact de type Gaussien dans un voisinage de $\tilde{\varphi}_n$, si il existe une fonction f (respectivement une fonction g) telle que pour toute suite φ_n vérifiant $\varphi_n \asymp \tilde{\varphi}_n$, on a $f(\varphi_n) \in]-\infty, \Phi^{-1}(1-\eta)]$ avec $\eta \in (0,1)$ (respectivement $g(\varphi_n) \in]-\infty, \Phi^{-1}(1)]$) et*

$$\beta(\varphi_n) = \Phi(g(\varphi_{n,p})) + o_n(1) \quad (\textit{respectivement } \gamma(\varphi_{n,p}) = \Phi(g(\varphi_{n,p})) + o_n(1))$$

où Φ est la fonction de répartition de la loi normale centrée et réduite.

Définition 1.4. *Un test Δ est asymptotiquement minimax exact si pour toute suite $\varphi_n > 0$, on a:*

$$\gamma(\Delta, \mathcal{F}_1, \varphi_n) = \inf_{\psi} \gamma(\psi, \mathcal{F}_1, \varphi_n) + o(1)$$

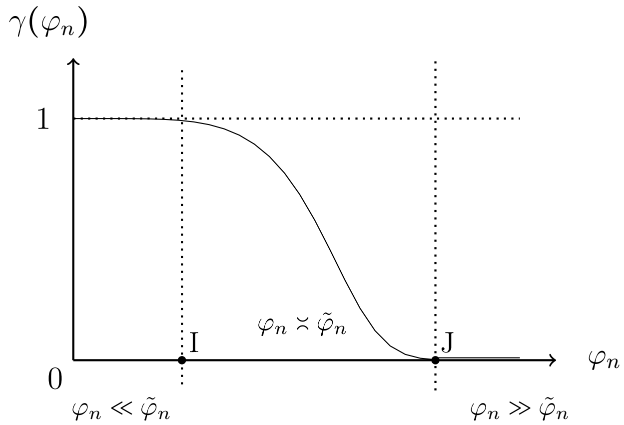

Figure 1.1 – Variation de l'erreur totale γ en fonction du rayon φ_n, qui caractérise l'écart entre l'hypothèse nulle et l'hypothèse alternative.

La figure 1.1, représente l'intervalle $[0, I]$ dans lequel, il est impossible de distinguer entre les deux hypothèses, l'intervalle $[I, J]$ où on a l'équivalent asymptotique exact de l'erreur totale qui est de type gaussien, et l'intervalle $[J, +\infty[$ où il est possible de distinguer entre les deux hypothèses avec une erreur totale qui tend vers zero.

1.1.2 Adaptation

En général les procédures de tests optimales nécessitent de connaître la régularité du paramètre inconnu Σ. Donc l'alternative est constituée de classes qui dépendent de α, que nous notons par $\mathcal{F}_1(\alpha)$. Cette restriction aboutit à des vitesses minimax de séparation qui dépendent non seulement de n mais aussi de α. Or, supposer que la régularité est connue, est une hypothèse qui semble irréaliste. D'où le besoin de développer des procédures de test qui ne dépendent pas du paramètre inconnu, mais en même temps qui s'adapte au mieux à α inconnu. En général ces tests sont construits en agrégeant plusieurs tests. Spokoiny dans [85] introduit un concept d'adaptation pour les problèmes de tests, alors que ce concept avait été introduit pour les problèmes d'estimation par Lepski dans [71] et [72]. Plusieurs travaux ont succédé sur le sujet des tests minimax adaptatifs pour différents modèles, voir Autin et Pouet [3], Butucea *et al.* [15], Gayraud et Pouet [43].

Le problème de test adaptatif est défini pour une hypothèse alternative élargie, contenant une collection de classes $\mathcal{F}_1(\alpha)$, pour des valeurs de α dans un ensemble A. Soit le problème de test des hypothèses suivantes:

$$H_0 : \Sigma = \Sigma_0 \qquad \text{contre} \qquad H_1 : \Sigma \in \bigcup_{\alpha \in A} \Big(\mathcal{F}_1(\alpha) \cap \bar{\mathcal{B}}_{\mathcal{F}}(\mathcal{C}\Phi_{n,\alpha}) \Big)$$

où A est un ensemble de paramètres, $\Phi_{n,\alpha} = \rho_{n,\alpha} \cdot \tilde{\varphi}_{n,\alpha}$, $\rho_{n,\alpha}$ est la perte dans la vitesse de séparation due à l'adaptation et $\widetilde{\varphi}_{n,\alpha}$ est la vitesse de séparation minimax pour le problème de test où l'alternative est donnée par $\mathcal{F}_1(\alpha) \cap \bar{\mathcal{B}}_{\mathcal{F}}(\varphi)$.

Définition 1.5. *La suite $\Phi_{n,\alpha}$ est appelée vitesse (minimax) adaptative de séparation s'il existe une constante $\mathcal{C}_0 > 0$ telle que:*

-d'une part, pour toute constante $\mathcal{C} < \mathcal{C}_0$

$$\inf_{\chi} \Big(\eta(\chi) + \sup_{\alpha \in A} \beta(\chi, \mathcal{F}_1(\alpha), \mathcal{C}\Phi_{n,\alpha}) \Big) \underset{n \to +\infty}{\longrightarrow} 1.$$

-d'autre part, pour toute constante $\mathcal{C} > \mathcal{C}_0$, il existe une procédure de test Δ libre de α qui vérifie

$$\eta(\Delta) + \sup_{\alpha \in A} \beta(\Delta, \mathcal{F}_1(\alpha), \mathcal{C}\Phi_{n,\alpha}) \underset{n \to +\infty}{\longrightarrow} 0.$$

Dans ce cas nous disons que Δ est une procédure de test (minimax) adaptative.

1.2 Aperçu de la littérature

Dans cette section, nous introduisons plusieurs exemples de classes non paramétriques $\mathcal{F}$. Deuxièmement, nous décrivons plusieurs modèles statistiques, pour lesquelles des problèmes des tests minimax ont été étudiés. Finalement, nous revenons sur les résultats connus pour les tests concernant, plus particulièrement, les matrices de covariance.

1.2.1 Classes non paramétriques

On considère la base trigonométrique $\{f_j\}_{j \in \mathbb{Z}}$ sur $\mathbb{L}_2([-\pi, \pi])$,

$$f_0(x) = 1, f_j(x) = \cos(jx) \text{ et } f_{-j}(x) = \sin(-jx)\,, \quad |j| \geq 1.$$

Pour toute fonction $f \in \mathbb{L}_2([-\pi,\pi])$, on note

$$\sigma_0 = \int_{-\pi}^{\pi} f(x)dx\,, \quad \sigma_j = \int_{-\pi}^{\pi} f(x)f_j(x)dx\,, \quad |j| \geq 1$$

les coefficients de Fourier de f dans la base $\{f_j\}_{j\in\mathbb{Z}}$.

Les classes non paramétriques suivantes apparaissent dans la littérature sur les tests minimax et adaptatifs:

i) Ellipsoïde de Sobolev: soient $\alpha, L > 0$,

$$\mathcal{S}(\alpha,L) = \Big\{ f \in \mathbb{L}_2([-\pi,\pi]) : f = \sum_{j\in\mathbb{Z}} \sigma_j f_j \ ; \ \sum_{j\in\mathbb{Z}} \sigma_j^2 j^{2\alpha} \leq L \text{ et } \sigma_0 = 1 \Big\}.$$

ii) Ellipsoïde de fonctions analytiques: soient $\alpha, L > 0$,

$$\mathcal{A}(\alpha,L) = \Big\{ f \in \mathbb{L}_2([-\pi,\pi]) : f = \sum_{j\in\mathbb{Z}} \sigma_j f_j \ ; \ \sum_{j\in\mathbb{Z}^*} \sigma_j^2 e^{2\alpha j} \leq L \text{ et } \sigma_0 = 1 \Big\}$$

iii) Classe de Hölder: soient J un intervalle de $\mathbb{R}$, $\alpha, L > 0$

$$\begin{aligned} H(\alpha,L) \ = \ & \Big\{ f : J \to \mathbb{R} : f \text{ est } l \text{ fois différentiable avec } l = \lfloor \alpha \rfloor \text{ et} \\ & |f^{(l)}(x) - f^{(l)}(y)| \leq L|x-y|^{\alpha - l}, \text{ pour tout } x, y \in J \Big\} \end{aligned}$$

iv) Classe de matrices de Toeplitz: soient $\alpha, L > 0$

$$\mathcal{T}(\alpha,L) = \{\Sigma > 0 : [\Sigma]_{i,i+j} = \sigma_{|j|} \ ; \sum_{j\geq 1} \sigma_j^2 j^{2\alpha} \leq L \text{ et } \sigma_0 = 1\}. \tag{1.4}$$

$$\mathcal{E}(\alpha,L) = \{\Sigma > 0, [\Sigma]_{i,i+j} = \sigma_{|j|} \ ; \sum_{j\geq 1} e^{2\alpha j} \sigma_j^2 \leq L \text{ et } \sigma_0 = 1\}. \tag{1.5}$$

On note $\Sigma > 0$ pour une matrice Σ définie positive. On rappelle que les matrices de Toeplitz ont les éléments diagonaux constants.

v) Classe de matrices non-Toeplitz: soient $\alpha, L > 0$

$$\begin{aligned} \mathcal{F}(\alpha,L) \ = \ & \Big\{ \Sigma > 0 : [\Sigma]_{ij} = \sigma_{ij}, \text{ symétrique } ; \frac{1}{p} \sum_{1\leq i<j\leq p} \sigma_{ij}^2 |i-j|^{2\alpha} \leq L \\ & \text{pour tout } p \text{ et } \sigma_{ii} = 1 \text{ pour tout } \ i = 1,\ldots,p \Big\}. \end{aligned} \tag{1.6}$$

Par analogie aux ensembles de fonctions, nous appelons α paramètre de régularité de $\mathcal{T}(\alpha,L)$, $\mathcal{E}(\alpha,L)$ et $\mathcal{F}(\alpha,L)$.

1.2.2 Modèles statistiques

a) Modèle de densité de probabilité: nous observons n réalisations indépendantes, $X_1, \dots, X_n$, de X de loi P_f admettant une densité de probabilité f inconnue et appartenant à $\mathcal{F}$.

b) Modèle de bruit blanc gaussien ou modèle de signal: nous observons le processus $\mathbf{X} = (X_t)_{t\in[0,1]}$, qui vérifie

$$dX_t = f(t)dt + \epsilon dW(t), \;\; t \in [0,1], \epsilon > 0$$

où la fonction f est inconnue et appartient à $\mathcal{F}$ et $W(t)$ est un mouvement brownien.

Des tests minimax asymptotiques ont été développés pour le modèle de densité de probabilité et le modèle de signal pour des classes de type ellipsoïde de Sobolev $\mathcal{S}(\alpha, L)$ et de fonctions analytiques $\mathcal{A}(\alpha, L)$. La distance induite par la norme $\mathbb{L}_2$ est utilisée pour séparer les deux hypothèses dans les papiers de Ermakov [34], [36] et [35]. Ingster dans [53], [54] et [55], traite des problèmes de tests pour les deux modèles précédents en considérant la distance induite par la norme $\mathbb{L}_p$ pour séparer entre les hypothèses, pour $1 \le p \le +\infty$, pour des classes de type Hölder et des classes de type ellipsoïde de Sobolev. Lepski et Sobolev [69] ont considéré des ellipsoïdes de type Besov. Des équivalents asymptotiques exacts pour la semi-distance en un point fixé et norme $\mathbb{L}_\infty$ sont donnés par Lepski et Tsybakov [70]. Ces résultats se trouvent résumés également dans Giné et Nickel [45]. Notons que des problèmes d'estimation et de test pour différents modèles de signal multidimensionnels, c.à.d quand $t \in [0,1]^d$ avec $1 \le d \le +\infty$, ont été traités dans Ingster et Suslina [52], [59] et Ingster et Stepanova [57], [60].

c) Modèle de régression: soient $Y_1, \dots, Y_n$, n observations définies par

$$Y_i = f(X_i) + \xi_i \qquad \text{pour } i = 1, \dots, n$$

où f est la fonction de régression inconnue, $\{\xi_i\}_{1\le i\le n}$ est une suite de variables aléatoires i.i.d, $\{X_i\}_{1\le i\le n}$ une suite de variables qui sont, soit déterministes, soit aléatoires indépendantes entre-elles, équidistribuées et indépendantes de $\{\xi_i\}_{1\le i\le n}$.

d) Modèle de densité spectrale: soit $\mathbf{X}$ un processus de second ordre, de fonction d'auto-covariance Γ définie par $\Gamma(r,s) := \mathrm{Cov}(X_r, X_s)$, pour tout $r, s \in \mathbb{N}$. Soit, de plus, $\mathbf{X}$ stationnaire, ce qui implique que sa fonction d'auto-covariance ne dépend que de l'écart $r - s$, $\Gamma(r,s) = \Gamma(r+h, s+h)$ pour tout $h \in \mathbb{Z}$. Notons $\sigma_{|r-s|} = \Gamma(r,s)$. Si

$$\sum_{j=-\infty}^{+\infty} |\sigma_j| < +\infty,$$

on appelle densité spectrale du processus $\mathbf{X}$ la fonction f

$$f(x) = \frac{1}{2\pi} \sum_{j=-\infty}^{+\infty} \sigma_j e^{-ijx} \text{ pour tout } x \in \mathbb{R}.$$

Cette densité spectrale est continue, positive, paire et 2π-périodique, ce qui permet de l'étudier uniquement sur $[-\pi, \pi]$:

$$f(x) = \frac{1}{2\pi}\Big(\sigma_0 + 2\sum_{j\geq 1} \sigma_j \cos(jx)\Big) \text{ pour } x \in [-\pi, \pi].$$

On appelle $\Sigma = \Sigma_p(f)$ la matrice de covariance du vecteur $(X_r, \dots, X_{r+p-1})$, pour tout $r, p \in \mathbb{N}$, de taille $p \times p$ de la forme

$$\Sigma = \begin{pmatrix} \sigma_0 & \sigma_1 & \cdots & \sigma_{p-2} & \sigma_{p-1} \\ \sigma_1 & \sigma_0 & \sigma_1 & \cdots & \sigma_{p-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \sigma_{p-1} & \sigma_{p-2} & \cdots & \sigma_1 & \sigma_0 \end{pmatrix}$$

Remarquons que la matrice Σ est de Toeplitz. Si, en plus, $\sigma_0 = 1$, alors la matrice Σ est la matrice d'auto-corrélation de $(X_r, \dots, X_{r+p-1})$.

Le modèle de densité spectrale consiste à observer $X_1, \dots, X_p$, p réalisations du processus stationnaire du second ordre, $\mathbf{X} = (X_r)_{r\in\mathbb{N}}$, admettant une densité spectrale f inconnue, appartenant à $\mathcal{F}$. Ces observations ont la matrice covariance $\Sigma = \Sigma_p(f)$ inconnue. On note que $f(x) = (1/2\pi) \cdot \mathbb{1}([-\pi, \pi])$ si et seulement si $\Sigma = I$ est la matrice identité. Le problème de test de l'hypothèse nulle $f(x) = (1/2\pi) \cdot \mathbb{1}([-\pi, \pi])$ est étudié par Ermakov [37].

e) Grande matrice de covariance de vecteurs Gaussiens stationnaires: soient $X_1, \dots, X_n$, n observations d'un p-vecteur Gaussian stationnaire le loi $\mathcal{N}_p(0, \Sigma)$, centré et de matrice de covariance Σ inconnue, $\Sigma \in \mathcal{F}$. Nous notons $X_k = (X_{k,1}, \dots, X_{k,p})^\top$, pour $k = 1, \dots, n$. Nous remarquons que Σ est de Toeplitz et que ce modèle peut être vu comme le modèle précédent, à observations indépendantes et répétées.

f) Grande matrice de covariance de vecteurs Gaussiens: soient $X_1, \dots, X_n$, n observations d'un vecteur Gaussian de dimension p, centré et de matrice de covariance Σ inconnue, $\Sigma \in \mathcal{F}$. Dans ce modèle Σ n'est pas une matrice de Toeplitz.

g) Grande matrice de covariance et données manquantes: soient $X_1, \dots, X_n$, i.i.d de dimension p qui suivent la $\mathcal{N}_p(0, \Sigma)$, avec Σ inconnu. Nous observons $Y_1, \dots, Y_n$, n vecteurs aléatoires i.i.d, tels que

$$Y_k = (\varepsilon_{k,1} \cdot X_{k,1}, \dots, \varepsilon_{k,p} \cdot X_{k,p}) \text{ pour tout } k = 1, \dots, n\,,$$

où $\{\varepsilon_{k,j}\}_{1\leq k\leq n, 1\leq j\leq p}$ est une suite i.i.d de variables de Bernoulli de paramètre $a \in (0,1)$ et indépendante de $\{X_k\}_{k=1,\dots,n}$. Par conséquent, chaque composante de chaque vecteur Y_k est observée avec une probabilité a.

Les modèles e), f) et g) sont considérés dans les travaux de Butucea et Zgheib [18], [19] et [20], et font l'objet d'une description plus détaillée dans la section 1.3. Ces travaux constituent les contributions principales de cette thèse.

1.2.3 Tests pour des matrices de covariance

Dans nos travaux, nous nous intéressons à des problèmes de tests de grandes matrices de covariance ayant une structure de bande, en utilisant l'approche minimax asymptotique. Nous traitons également la question de l'adaptation au paramètre de régularité de la matrice sous-jacente. Faisons un aperçu général sur ce qui a été fait dans la littérature sur ce sujet.

Soient $X_1, \ldots, X_n$, n réalisations de $X \sim \mathcal{N}_p(0, \Sigma)$, où Σ est de taille $p \times p$. Le problème de tester à partir de ces observations l'hypothèse nulle

$$H_0 : \Sigma = I \tag{1.7}$$

contre l'hypothèse alternative $H_1 : \Sigma \neq I$, a été étudié au début dans un cadre paramétrique, pour p fixe. Le premier test pour les matrices de covariance a été proposé par Mauchly [75], ce test est basé sur le rapport du maximum de vraisemblance. Le rapport du maximum de vraisemblance est défini par:

$$\mathrm{RV} := \frac{1}{(\det S_n)^{\frac{n}{2}}} \cdot \exp\Big(-\frac{n}{2}(tr(S_n - I))\Big)$$

où $S_n = (1/n) \sum_{i=1}^n X_n X_n^\top$ l'estimateur de maximum de vraisemblance Σ. Une étude développée du test basé sur le rapport du maximum de vraisemblance RV se trouve dans Anderson [1] ainsi que dans Muirhead [76]. Il est montré que $-2\log(RV)$ converge sous l'hypothèse nulle vers une $\chi^2_{p(p+1)/2}$ quand p est fixe.

Une autre approche basée sur la forme quadratique suivante:

$$\mathrm{FQ} = \frac{n}{2} tr(S_n - I)^2$$

est suggérée par Nagao [77] pour tester l'hypothèse nulle (1.7). Il montre la convergence en loi de FQ vers $\chi^2_{p(p+1)/2}$ sous H_0, quand n tend vers $+\infty$ et p est fixe.

Or, de nos jours, suite au développement dans plusieurs domaines, tels que la biologie, la télécommunication, la finance, l'économie, etc, il est possible de collecter un nombre immense de données par individu. Il se peut que le nombre de données soit beaucoup plus grand que le nombre d'individus. Par exemple, en génomique une molécule d'ADN comporte des milliers de gènes, en cosmologie une image d'un objet céleste est constituée par des millions de pixels. D'où l'intérêt de traiter les problèmes statistiques dans le cadre des grandes dimensions. En particulier, plusieurs travaux ont effectué le test de l'hypothèse (1.7) dans le cas d'une grande dimension p, où les statistiques de test définies précédemment sont dégénérées asymptotiquement.

Pour faire face à ce problème, de nouvelles statistiques de test ont été proposées. Certaines dérivent du rapport de vraisemblance ou bien de la forme quadratique. Srivastava dans son papier [87], propose pour couvrir le cas $p > n$, de construire une statistique de test à partir des n valeurs propres non nulles de nS_n et en permutant les rôles de n et p. Soient $\hat{\lambda}_1, \dots, \hat{\lambda}_n$ ces valeurs propres, la statistique de test est définie par:

$$RVC_1 = \Big(1 - \frac{2n^2 + 3n + 1}{6p(n+1)}\Big)\Big(\sum_{i=1}^{n} \hat{\lambda}_i - p \log\Big(\frac{\prod_{i=1}^{n} \hat{\lambda}_i}{p^n}\Big) - np\Big)$$

Bai *et al.* [4] proposent une correction de RV dans le cas où $p/n \to y\,;\, y \in (0,1)$ en utilisant des résultats de la théorie des matrice aléatoires. La version corrigée qu'ils proposent est :

$$RVC_2 = \frac{-2\log(RV)/n - p(1 - (1 - y^{-1})\log(1-y)) - 1/2\log(1-y)}{\sqrt{-2\log(1-y) - 2y}}$$

et montre que RVC_2 converge en loi vers $\mathcal{N}(0,1)$ sous H_0. Jiang *et al.* [62] étendent ce résultat pour le cas $c = 1$. Une version généralisée du rapport de vraisemblance est proposée dans [8], pour établir le test le l'hypothèse nulle $\Sigma = \sigma^2 I$, pour σ inconnu.

Nous citons aussi le travail de Ledoit et Wolf [67] qui proposent une modification de FQ pour obtenir la statistique suivante:

$$FQM_1 = \frac{2FQ}{np} - \frac{p}{n}\Big(\frac{1}{p}tr(S_n)\Big)^2 + \frac{p}{n}.$$

Ils montrent que $n \cdot FQM_1 - p$ converge en loi vers $\mathcal{N}(1,4)$, sous H_0, pour $n, p \to +\infty$.

Notons que $(2/np)FQ$ vaut $(1/p)tr(S_n - I)^2$. Remarquons que, pour tester l'hypothèse $\Sigma = I$ contre l'alternative $\Sigma \neq I$, il suffit d'étudier le comportement de la fonction

$$\frac{1}{p}tr(\Sigma - I)^2 = \frac{1}{p}\sum_{i=1}^{p}(\lambda_i - 1)^2 = \frac{1}{p}\sum_{i=1}^{p}\lambda_i^2 - \frac{2}{p}\sum_{i=1}^{p}\lambda_i + 1 \tag{1.8}$$

Srivastava [86] propose des estimateurs sans biais et consistents des moyennes arithmétiques $(1/p)\sum_{i=1}^{p}\lambda_i^2$ et $(1/p)\sum_{i=1}^{p}\lambda_i$, pour construire la statistique de test donnée par :

$$FQM_2 = \frac{n^2}{p(n-1)(n+2)}\Big(tr(S_n)^2 - \frac{1}{n}(tr(S_n))^2\Big) - \frac{2}{p}tr(S_n) + 1$$

De plus, il est montré dans [88] que cette statistique de test peut être utilisée pour des observations non gaussiennes mais sous des conditions sur les moments de X. De manière plus générale, Fisher [39] propose de construire des statistiques de tests basées sur des estimateurs de $(1/p)\sum_{i=1}^{p}(\lambda_i^r - 1)^{2s}$, pour $r, s > 1$. Il traite le cas $r = 1$ et $s = 2$ ainsi que le cas $r = 2$ et $s = 1$.

Des procédures de test basées sur la déviation maximale des entrées non diagonales [92] et d'autres basées sur la plus grande valeur propre de la matrice de covariance empirique [64], ont été considérées aussi dans la littérature. Pour traiter le cas des observations non gaussiennes, Chen *et al.* [29] suggèrent une U-statistique d'ordre 2 comme un estimateur de

(1.8), qui n'est rien d'autre qu'une nouvelle modification de $(2/np)FQ$, et qui est donnée par l'expression suivante:

$$U_n = \frac{1}{p}\Big(\frac{1}{n(n-1)}\sum_{1\le l\ne k\le n}(X_l^\top X_k)^2 - \frac{2}{n}\sum_{k=1}^{n} X_k^\top X_k + p\Big). \tag{1.9}$$

Ils montrent que nU_n converge en loi vers $\mathcal{N}(0,4)$, sous H_0, quand n et p tendent vers l'infini. En plus, ils minorent la puissance du test basée sur U_n pour une matrice Σ arbitrairement fixée, sous l'alternative. Dans la plupart des travaux que nous avons cités, les résultats portent sur l'étude de l'erreur de première espèce. [86] et [29] ont étudié l'erreur de deuxième espèce pour une matrice arbitrairement fixée sous l'alternative.

Récemment, Cai et Ma dans leur papier [23], proposent de traiter le problème d'un point de vue minimax. Ils proposent de tester l'hypothèse (1.7) contre l'hypothèse alternative donnée par:

$$H_1 : \Sigma \ne I \quad \text{tel que } \|\Sigma - I\|_F^2 \ge \varphi_{n,p}^2.$$

Ils construisent une procédure de test basée sur la U-statistique U_n définie dans (1.9). Ils montrent que la vitesse de séparation minimax $\tilde{\varphi}_{n,p}$ est de l'ordre de $\sqrt{p/n}$.

1.3 Résultats de la thèse.

Le modèle étudié est celui de n vecteurs gaussiens indépendants, de dimension p, $X_1,\dots,X_n$, de loi $\mathcal{N}_p(0,\Sigma)$.

Le premier chapitre est consacré au problème de test minimax de l'hypothèse (1.7), $H_0 : \Sigma = I$, contre une alternative de matrices de Toeplitz. Nos procédures sont asymptotiquement minimax exactes. Le second chapitre se consacre au même test pour des classes de matrices, pas nécessairement Toeplitz. De plus, une procédure qui agrège les tests asymptotiquement exacts pour créer un test adaptatif est étudiée. Le troisième chapitre se concentre sur les tests adaptatifs dans le problème inverse où nous disposons de données incomplètes. En effet, des coordonnées des vecteurs $X_1,\dots,X_n$ sont manquantes de manière indépendante, avec même probabilité.

Dans les trois chapitres, la séparation entre les deux hypothèses est donnée par des normes de type $\mathbb{L}_2$. En effet, ceci permet de construire des procédures de test qui atteignent des vitesses minimax de séparation plus rapides que les vitesses d'estimation.

1.3.1 Grande matrice de covariance de Toeplitz.

On considère le modèle de grande matrice de covariance de vecteurs gaussiens stationnaires (voir modèle e) section 1.2.1) avec n et p qui tendent vers $+\infty$. En premier lieu, nous étudions le problème de test de l'hypothèse nulle (1.7) contre l'alternative

$$H_1 : \Sigma \in \mathcal{T}(\alpha, L) \quad \text{tel que } \sum_{j\ge1}\sigma_j^2 \ge \psi^2 \tag{1.10}$$

où $\mathcal{T}(\alpha, L)$ est la classe de matrices définie dans (1.4). Remarquons que pour toute matrice $\Sigma \in \mathcal{T}(\alpha, L)$ avec $\alpha > 1/2$, on a $\frac{1}{2p}\|\Sigma - I\|_F^2 \underset{p\to+\infty}{\sim} \sum_{j\geq 1}\sigma_j^2$. En effet,

$$\sum_{j=1}^{p-1}\sigma_j^2 - \frac{1}{p}\sum_{j=1}^{p-1} j^{2\alpha}\sigma_j^2 \leq \frac{1}{2p}\|\Sigma - I\|_F^2 = \frac{1}{p}\sum_{j=1}^{p-1}(p-j)\sigma_j^2 \leq \sum_{j=1}^{p-1}\sigma_j^2$$

or $\frac{1}{p}\sum_{j=1}^{p-1} j^{2\alpha}\sigma_j^2 \leq \frac{L}{p} \underset{p\to+\infty}{\to} 0$. Ce problème est donc analogue au test dans le modèle de la densité spectrale f associée à $\mathbf{X}$:

$$H_0 : f = \frac{1}{2\pi}\cdot \mathbb{1}([-\pi,\pi]) \ \text{ contre } \ H_1 : f \in \mathcal{S}(\alpha, L) \text{ telle que } \left\| f - \frac{1}{2\pi}\right\|^2_{\mathbb{L}_2([-\pi,\pi])} \geq \psi^2.$$

Notons que le problème sur la densité spectrale énoncé précédemment est étudié dans [37] pour le cas $n = 1$. Nos résultats généralisent ceux d'Ermakov [37] au cas où nous avons des observations répétées du processus.

En second lieu, nous étudions le problème de test de la même hypothèse nulle contre l'alternative

$$H_1 : \Sigma \in \mathcal{E}(\alpha, L) \quad \text{ tel que } \sum_{j\geq 1}\sigma_j^2 \geq \psi^2 \tag{1.11}$$

où $\mathcal{E}(\alpha, L)$ est la classe de matrices définie dans (1.5). Ce problème est similaire au problème de test de l'hypothèse $f = 1/2\pi$ contre l'alternative f appartient à l'ensemble $\mathcal{A}(\alpha, L)$ et l'écart entre les deux hypothèses est mesuré par la norme $\mathbb{L}_2$.

Ellipsoïde à coefficients polynomiaux

Pour tester l'hypothèse nulle contre l'alternative (1.10), nous introduisons une statistique de test similaire à celle proposée par Ermakov [37] et suivant la théorie des tests minimax de Ingster et Suslina [58]. À la différence de Ermakov [37], notre statistique de test ne contient pas de termes croisés grâce aux observations répétées. Elle est pondérée par des poids optimaux, qui sont solution du problème d'optimisation:

$$\sum_{j\geq 1} w_j^*\sigma_j^{*2} = \sup_{\left\{\substack{(w_j)_j \ : \ w_j\geq 0; \\ \sum_{j\geq 1} w_j^2 = \frac{1}{2}}\right\}} \inf_{\left\{\substack{\Sigma \,:\, \Sigma\in\mathcal{T}(\alpha,L); \\ \sum_{j\geq 1}\sigma_j^2\geq\psi^2}\right\}} \sum_{j\geq 1} w_j\sigma_j^2. \tag{1.12}$$

Soit,

$$\widehat{\mathcal{A}}^{\mathcal{T}}_{n,p} = \frac{1}{n(n-1)(p-T)^2}\sum_{1\leq k\neq l\leq n}\sum_{j=1}^{T} w_j^* \sum_{T+1\leq i_1,i_2\leq p} X_{k,i_1}X_{k,i_1-j}X_{l,i_2}X_{l,i_2-j}$$

où

$$w_j^* \ = \ \frac{\lambda}{2b(\psi)}\Big(1-(\frac{j}{T})^{2\alpha}\Big) \quad \text{avec } T = \lfloor (L(4\alpha+1))^{\frac{1}{2\alpha}}\cdot\psi^{-\frac{1}{\alpha}}\rfloor, \tag{1.13}$$

$$\lambda = \frac{2\alpha+1}{2\alpha(L(4\alpha+1))^{\frac{1}{2\alpha}}}\cdot\psi^{\frac{2\alpha+1}{\alpha}} \qquad \text{et} \qquad b^2(\psi) = \frac{2\alpha+1}{L^{\frac{1}{2\alpha}}(4\alpha+1)^{1+\frac{1}{2\alpha}}}\cdot\psi^{\frac{4\alpha+1}{\alpha}}.$$

Nous calculons les moments d'ordre un et deux de $\widehat{\mathcal{A}}^{\mathcal{T}}_{n,p}$ sous l'hypothèse nulle ainsi que le moment d'ordre un et un majorant de la variance sous l'alternative. En plus nous montrons que $\widehat{\mathcal{A}}^{\mathcal{T}}_{n,p}$ converge en loi vers la loi normale standard sous H_0. Sous H_1 nous appliquons Hall [50] à $\widehat{\mathcal{A}}^{\mathcal{T}}_{n,p} - \mathbb{E}_\Sigma(\widehat{\mathcal{A}}^{\mathcal{T}}_{n,p})$ pour montrer sa normalité asymptotique, pour des matrices Σ qui sont proches de la matrice Σ^*, $[\Sigma^*]_{i,i+j} = \sigma^*_{|j|}$, solution du problème d'optimisation (1.12). En effet, une U-statistique dégénérée dont le noyaux dépend de n peut converger vers une loi normale au lieu d'une loi de χ^2, sous certaines hypothèses qui seront vérifiées ici.

Nous proposons un test qui rejette l'hypothèse nulle pour une valeur de $\widehat{\mathcal{A}}^{\mathcal{T}}_{n,p}$ plus grande qu'un seuil $t > 0$:

$$\chi^* = \chi^*(t) = \mathbb{1}(\widehat{\mathcal{A}}^{\mathcal{T}}_{n,p} > t)$$

où t est choisi strictement plus petit que $cb(\psi)$ avec $c \in (0,1)$ et tel que $npt \to +\infty$. Nos résultats sont établis sous les conditions suivantes $n, p \to +\infty$, $\psi \to 0$ et $p\psi^{\frac{1}{\alpha}} \to +\infty$.

D'abord, nous montrons que pour t vérifiant les conditions énoncées précédemment, l'erreur de première espèce $\eta(\chi^*(t))$ tend vers 0, et que si $\alpha > 1/4$ et $n^2p^2\psi^{\frac{4\alpha+1}{\alpha}} \to +\infty$, l'erreur maximale de deuxième espèce tend aussi vers 0. Nous montrons également les bornes inférieures. En d'autres termes, nous démontrons que pour $\alpha > 1$, si $n^2p^2\psi^{\frac{4\alpha+1}{\alpha}} \to 0$, alors

$$\gamma(\psi) = \inf_{\chi} \gamma(\chi, \mathcal{T}(\alpha, L), \psi) \to 1.$$

où l'inf est pris sur tous les tests possibles. Par conséquent, nous prouvons que la procédure de test χ^* est asymptotiquement minimax. Nous déduisons de ce théorème la vitesse minimax de séparation

$$\widetilde{\psi} = \Big(C(\alpha, L)\cdot n^2p^2\Big)^{-\frac{\alpha}{4\alpha+1}}, \text{ où } C(\alpha, L) = \frac{2\alpha+1}{L^{\frac{1}{2\alpha}}(4\alpha+1)^{1+\frac{1}{2\alpha}}}.$$

Dans un second théorème, nous évaluons le comportement des erreurs dans le voisinage de $\widetilde{\psi}$. Nous montrons que si $n^2p^2b^2(\psi) \asymp 1$ alors d'une part pour $\alpha > 1$ nous avons,

$$\gamma(\psi) = \inf_{\chi} \gamma(\chi, \mathcal{T}(\alpha, L), \psi) \geq 2\Phi\Big(-\frac{1}{2}\cdot npb(\psi)\Big).$$

D'autre part, la procédure de test χ^* vérifie les propriétés suivantes:

$$\eta(\chi^*(t)) = 1 - \Phi(npt) + o(1)$$

et pour $\alpha > 1$

$$\beta(\chi^*(t), \mathcal{T}(\alpha, L), \psi) \leq \Phi(np(t - b(\psi))) + o(1).$$

Par conséquent si nous prenons $t = b(\psi/2)$, nous obtenons

$$\gamma(\chi^*(t), \mathcal{T}(\alpha, L), \psi) = 2\Phi\Big(-\frac{1}{2}npb(\psi)\Big).$$

Pour obtenir les bornes inférieures de l'erreur totale minimax, nous réduisons l'alternative à une classe paramétrique de matrices de Toeplitz. Nous considérons le sous ensemble formé par les matrices Σ_U^* de diagonale principale 1 et dont les entrées non diagonales sont données par $[\Sigma_U^*]_{i,i+j} = u_{|j|} \cdot \sigma_{|j|}^* \cdot \mathbb{1}_{(1<|j|<T)}$, avec $u_{|j|} = \pm 1$ et les $\sigma_{|j|}^*$ sont solution du problème d'optimisation (1.12). Nous considérons ensuite, P_π la moyenne des lois de probabilité $P_{\Sigma_U^*}$, où $P_{\Sigma_U^*}$ est une loi normale centrée de matrice de covariance Σ_U^*. La preuve est basée sur le log du rapport de vraisemblance des observations $X_1, \dots, X_n$, $L_{n,p} = \log \dfrac{dP_\pi}{dP_I}(X_1 \dots, X_n)$.

Nous illustrons la performance de notre procédure par des simulations numériques.

Ellipsoïde à coefficients exponentiels

Dans cette partie nous traitons le problème de test avec l'alternative (1.11). Nous construisons un test à partir de la statistique $\widehat{\mathcal{A}}_{n,p}^{\mathcal{E}}$, qui a la même forme que $\widehat{\mathcal{A}}_{n,p}^{\mathcal{T}}$ mais pondérée différemment. Les poids w_j^* que nous utilisons dans ce cas sont donnés par:

$$w_j^* = \frac{\lambda}{2b(\psi)}\Big(1 - (\frac{e^j}{e^T})^{2\alpha}\Big)_+, \quad \text{où}$$

$$T = \Big\lfloor \frac{1}{\alpha} \ln\Big(\frac{1}{\psi}\Big)\Big\rfloor, \qquad \lambda = \frac{\alpha\psi^2}{\ln\Big(\dfrac{1}{\psi}\Big)}, \qquad b^2(\psi) = \frac{\alpha\psi^4}{2\ln\Big(\dfrac{1}{\psi}\Big)}.$$

Nous montrons que sous l'alternative (1.11) la vitesse minimax de séparation est

$$\widetilde{\psi} = \Big(\frac{2\ln(n^2p^2)}{\alpha n^2 p^2}\Big)^{1/4}.$$

qui est libre de L et nous l'obtenons pour toute valeur de $\alpha > 0$, sous la condition que $p \cdot \ln(\psi) \to \infty$.

1.3.2 Grande matrice de covariance et adaptation.

À partir du modèle f) et en supposant que $n, p \to +\infty$, nous abordons le problème de tester l'hypothèse (1.7) contre

$$H_1 : \Sigma \in \mathcal{F}(\alpha, L), \quad \text{tel que} \quad \frac{1}{2p}\|\Sigma - I\|_F^2 \geq \varphi^2. \tag{1.14}$$

où $\mathcal{F}(\alpha, L)$ est définie dans (1.6). Nous construisons un test basé sur une forme régularisée de la U-statistique d'ordre 2 définie dans (1.9). En effet, les techniques de régularisation ont été employées avec succès pour les problèmes d'estimation de grandes matrices de covariances appartenant à des classes possédant des propriétés similaires à celles de la classe $\mathcal{F}(\alpha, L)$. Parmi ces techniques, nous notons les différentes méthodes de seuillages proposées dans [9] et [10]. Notons que récemment Cai et *al* dans [24], proposent une étude minimax du problème d'estimation de plusieurs gammes de matrices de covariances structurées, parmi lesquelles un ensemble de matrices ayant des propriétés semblables à celle de $\mathcal{F}(\alpha, L)$.

Nous proposons une pondération par des poids optimaux $\{w_{ij}^*\}_{1\le i\neq j\le p}$ tel que $w_{ij}^* = w_{|i-j|}^*$ avec $w_{|i-j|}^*$ défini dans (1.13) comme solution du problème d'optimisation (1.12). Soit

$$\widehat{\mathcal{D}}_{n,p} = \frac{1}{n(n-1)p} \sum_{1\le k\neq l\le n} \sum_{1\le i<j\le p} w_{ij}^* X_{k,i} X_{k,j} X_{l,i} X_{l,j}$$

la nouvelle statistique de test. Le test construit à partir de $\widehat{\mathcal{D}}_{n,p}$ est défini par:

$$\Delta^* = \Delta^*(t) = \mathbb{1}(\widehat{\mathcal{D}}_{n,p} > t), \quad t > 0.$$

Nous énonçons deux propositions qui décrivent quelques propriétés de $\widehat{\mathcal{D}}_{n,p}$: ses moments du premier ordre sous chacune des hypothèses, son moment du second ordre sous H_0, un majorant de sa variance sous H_1, sa normalité asymptotique sous H_0, ainsi que sa normalité asymptotique sous H_1, pour des matrices Σ au voisinage de la solution du problème d'optimisation (1.12).

Grâce aux propriétés de $\widehat{\mathcal{D}}_{n,p}$, nous montrons premièrement que, sous les conditions $\alpha > 1/2$, $\varphi \to 0$ et $p\varphi^{\frac{1}{\alpha}} \to +\infty$, le test $\Delta^*(t)$ avec $0 < t < cb(\varphi), c \in (0,1)$, est minimax asymptotiquement consistant, si $n\sqrt{p}t \to +\infty$ et $n^2pb^2(\varphi) \to +\infty$. Deuxièmement, nous prouvons que, sous les mêmes conditions, l'erreur de première espèce vérifie $\eta(\Delta^*(t)) = 1 - \Phi(n\sqrt{p}t) + o(1)$ et l'erreur maximale de deuxième espèce est majorée comme suit

$$\beta(\Delta^*(t), \mathcal{F}(\alpha, L), \varphi) \le \Phi(n\sqrt{p}(t - b(\varphi))) \quad \text{si } n^2pb^2(\varphi) \asymp 1.$$

Nous implémentons notre procédure de test. Les résultats montrent que notre test a une puissance meilleure que le test basé sur la U-statistique non pondérée de Cai et Ma [23].

Nous montrons ensuite l'optimalité de nos résultats, ce qui prouve que

$$\widetilde{\varphi} = (C(\alpha, L)n^2p)^{-\alpha/(4\alpha+1)},$$

est la vitesse de séparation minimax. Plus précisément, nous montrons que sous les conditions $\alpha > 3/2$, $\varphi \to 0$ et $p\varphi^{\frac{1}{\alpha}} \to +\infty$,

$$\gamma(\varphi) = \inf_{\chi} \gamma(\chi, \mathcal{F}(\alpha, L), \varphi) \to 1 \quad \text{si } n^2pb^2(\varphi) \to 0.$$

En effet, pour obtenir cette borne inférieure, nous construisons un sous ensemble paramétrique plus large que celui considéré dans le cas de matrices de Toeplitz. Nous considérons l'ensemble des matrices Σ_U^* de diagonale principale 1, dont les entrées non diagonales vérifient $[\Sigma_U^*]_{ij} = u_{ij}\sigma_{|i-j|}^* \cdot \mathbb{1}_{(1<|i-j|<T)}$, avec $u_{ij} = \pm 1$ et les $\sigma_{|i-j|}^*$ sont les mêmes que ceux utilisées dans le cas de matrices de Toeplitz. Ici, P_π est la moyenne des lois $P_{\Sigma_U^*}$, avec $P_{\Sigma_U^*} \sim \mathcal{N}_p(0, \Sigma_U^*)$ et Σ_U^* appartenant à l'ensemble paramétrique défini précédemment. Enfin, comme

$$\gamma(\varphi) \ge 1 - \frac{1}{2}\|P_I - P_\pi\|_1 \ge 1 - \frac{1}{2}\Big(\mathbb{E}_I \left(\frac{dP_\pi}{dP_I}\right)^2 - 1\Big),$$

nous montrons que $\mathbb{E}_I \left(\frac{dP_\pi}{dP_I} \right)^2 \leq 1 + o(1)$, pour obtenir le résultat désiré.

De plus, sous les conditions $\alpha > 1$, $\varphi \to 0$, $p\varphi^{\frac{1}{\alpha}} \to +\infty$, $np\varphi^4 \to 0$ et $n^2pb^2(\varphi) \asymp 1$, nous montrons que le log du rapport de vraisemblance de $X_1 \ldots, X_n$ vérifie:

$$L_{n,p} := \log \frac{dP_\pi}{dP_I}(X_1 \ldots, X_n) = u_n Z_n - \frac{u_n^2}{2} + \xi,$$

où $u_n = n\sqrt{p}b(\varphi)$, Z_n converge en loi sous P_I, vers une loi normale standard et ξ une variable aléatoire qui converge vers 0 en P_I-probabilité. Par conséquent, nous obtenons une minoration de l'erreur maximale minimax de deuxième espèce et de l'erreur totale minimax:

$$\beta_\eta(\varphi) = \inf_{\chi:\eta(\chi)\leq\eta} \beta(\chi, \mathcal{F}(\alpha, L), \varphi) \geq \Phi(z_{1-\eta} - n\sqrt{p}b(\varphi)) + o(1),$$

et

$$\gamma(\varphi) = \inf_\chi \gamma(\chi, \mathcal{F}(\alpha, L), \varphi) \geq 2\Phi(-n\sqrt{p}\,\frac{b(\varphi)}{2}) + o(1).$$

Nous remarquons une perte d'un facteur p dans la vitesse minimax par rapport au cas des matrices de Toeplitz. Cette perte est due au nombre de paramètres à estimer en plus dans ce cas. En effet, pour le cas des matrices non-Toeplitz, on a $p(p-1)/2$ paramètres inconnus, or pour le cas des matrices de Toeplitz on a $p-1$ paramètres inconnus. En rassemblant ces résultats, nous déduisons les équivalents exacts des erreurs, pour des valeurs particulières de t.

En plus, nous montrons que $\widetilde{\varphi}$ est la vitesse de séparation minimax, pour ce problème de test établi pour l'inverse de la matrice de covariance, si ses valeurs propres sont dans un compact de $(0, +\infty)$.

En dernier, nous présentons un test adaptatif au paramètre α, pour α dans un compact de $(1/2, +\infty)$. Nous montrons que ce test atteint la vitesse

$$\widetilde{\psi} = \left(n\sqrt{p}/\sqrt{\ln\ln(n\sqrt{p})} \right)^{-2\alpha/(4\alpha+1)}.$$

1.3.3 Adaptation en présence de données manquantes.

Le problème d'estimation des matrices de covariance à partir d'un échantillon incomplet a été étudié dans la littérature, suivant différentes approches. La plus simple est d'éliminer de l'étude toutes variables dont les observations ne sont pas disponibles. Cette méthode n'est pas fiable dans le cas où le nombre des valeurs manquantes est assez important. Une autre méthode consiste à remplir les valeurs manquantes via de nouveaux modèles pour estimer ces valeurs manquantes. Récemment, le modèle de données incomplètes a été vu comme un problème inverse, voir [73] pour l'estimation non paramétrique de grandes matrices de covariance de rang petit. Le problème de test n'a pas été considéré auparavant dans ce contexte à notre connaissance.

Matrice non-Toeplitz

D'abord, nous considérons le modèle de grande matrice de covariance non-Toeplitz dans le cadre des données manquantes (voir modèle g) section 1.2.1) avec $n, p \to +\infty$ et a peut tendre vers 0. À partir des ces observations, qui contiennent des valeurs manquantes, nous voulons tester l'hypothèse nulle (1.7) contre (1.14). Nous prenons ici $L = 1$. Nous notons par $\mathcal{F}(\alpha)$ l'ensemble $\mathcal{F}(\alpha, 1)$.

Procédures de tests et vitesses.

Nous cherchons à construire une procédure de test adaptative de forme simple. Nous introduisons la statistique $\widehat{\mathcal{D}}_{n,p,m}$ qui est du même type que $\widehat{\mathcal{D}}_{n,p}$ mais de forme plus simple, pondérée par des poids constants. Soit m un entier naturel assez grand qui vérifie:

$$D \le m^{\alpha} \cdot \varphi \le K^{-2\alpha} \text{ pour des constantes } D > 1 \quad \text{et } K > 0. \tag{1.15}$$

Ici, m joue le rôle de T, le paramètre de troncature. Remarquons aussi que $m \asymp \varphi^{-\frac{1}{\alpha}}$. La statistique de test $\widehat{\mathcal{D}}_{n,p,m}$ est obtenue à partir de $\widehat{\mathcal{D}}_{n,p}$, en remplaçant les poids w_{ij}^* par les poids constants $1/\sqrt{2m}$ pour tout $1 \le |i-j| \le m-1$, et les observations X_k par les Y_k. Par conséquent,

$$\widehat{\mathcal{D}}_{n,p,m} = \frac{1}{n(n-1)p} \cdot \frac{1}{\sqrt{2m}} \sum_{1 \le k \ne l \le n} \sum_{\substack{1 \le i < j \le p \\ |i-j| < m}} Y_{k,i} Y_{k,j} Y_{l,i} Y_{l,j}.$$

Pour tester l'hypothèse (1.7) contre l'hypothèse (1.14), nous proposons le test suivant

$$\Delta_m(t) = \mathbb{1}(\widehat{\mathcal{D}}_{n,p,m} > t) \qquad \text{avec } t > 0.$$

Nous dérivons quelques propriétés de $\widehat{\mathcal{D}}_{n,p,m}$ et nous montrons que la vitesse de séparation minimax obtenue dans le cas des données incomplètes est

$$\widetilde{\varphi}_{M,\alpha} = \left(a^2 n \sqrt{p}\right)^{-\frac{2\alpha}{4\alpha+1}}.$$

Pour obtenir la borne inférieure de l'erreur totale minimax dans le cas des données manquantes, nous passons par des lois conditionnelles. D'abord, nous restreignons l'alternative à l'ensemble des matrices Σ_U; $[\Sigma_U]_{ij} = \mathbb{1}_{(i=j)} + u_{ij}\sigma \cdot \mathbb{1}_{(|1<|i-j|<T)}$, avec $\sigma \asymp \varphi^{1+\frac{1}{2\alpha}}$ et $T = \lfloor \varphi^{-\frac{1}{\alpha}} \rfloor$. Notons par P_U et $P_U^{(\varepsilon)}$ les lois de $Y = (Y_1, \dots, Y_n)$ et $Y|\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$, quand les X_k suivent la loi $\mathcal{N}(0, \Sigma_U)$, et par P_I et $P_I^{(\varepsilon)}$ les lois de Y et $Y|\varepsilon$, quand les X_k suivent la loi $\mathcal{N}(0, I)$. Soit P_π la moyenne des lois P_U. Nous contrôlons $K(P_I, P_\pi)$ en utilisant les vraisemblances conditionnelles:

$$K(P_I, P_\pi) = \mathbb{E}_I \log\left(\frac{dP_I}{dP_\pi}\right) = \mathbb{E}_I \log\left(\frac{d(P_\varepsilon \otimes P_I^{(\varepsilon)})}{d(P_\varepsilon \otimes P_\pi^{(\varepsilon)})}\right) = \mathbb{E}_\varepsilon \mathbb{E}_I^{(\varepsilon)} \log\left(\frac{dP_I^{(\varepsilon)}}{dP_\pi^{(\varepsilon)}}\right),$$

où P_ε est la loi de $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$.

Tests adaptatifs

Dans cette partie, nous considérons que le paramètre α est inconnu et appartient à l'intervalle A, tel que $A := [\alpha_*, \alpha^*_{n,p}] \subset]1/2, +\infty[$, où $\alpha^*_{n,p} \to +\infty$ mais en respectant la condition $\alpha^*_{n,p} = o(1)\ln(a^2 n\sqrt{p})$. Nous étudions le problème de test de l'hypothèse (1.7) contre l'alternative

$$H_1 : \Sigma \in \bigcup_{\alpha \in A} \Big\{ \mathcal{F}(\alpha) \; ; \; \frac{1}{2}\sum_{i<j} \sigma_{ij}^2 \geq (\mathcal{C}\phi_{M,\alpha})^2 \Big\}, \tag{1.16}$$

où $\mathcal{C}$ est une constante strictement positive et

$$\widetilde{\phi}_{M,\alpha} = \left(\frac{\sqrt{\ln\ln(a^2 n\sqrt{p})}}{a^2 n\sqrt{p}} \right)^{\frac{2\alpha}{4\alpha+1}} .$$

Soient $L_*, L^* \in \mathbb{N}^*$, qui vérifient

$$L_* = \Big(\frac{2}{(4\alpha^*_{n,p}+1)\ln 2} \Big) \ln(a^2 n\sqrt{p}) \quad \text{et} \quad L^* = \Big(\frac{2}{(4\alpha_*+1)\ln 2} \Big) \ln(a^2 n\sqrt{p}).$$

Alors, pour tout $\alpha \in A$, il existe $l \in \{L_*, \ldots, L^*\}$ tel que $2^{l-1} \leq (\phi_{M,\alpha})^{-\frac{1}{\alpha}} < 2^l$. Pour tester (1.7) contre (1.16), le test adaptatif Δ_{ad} est construit en agrégeant les tests $\Delta_{2^l}(t_l)$ pour $l \in \{L_*, \ldots, L^*\}$. Soit alors

$$\Delta_{ad} = \max_{L_* \leq l \leq L^*} \Delta_{2^l}(t_l) = \max_{L_* \leq l \leq L^*} \mathbb{1}(\widehat{\mathcal{D}}_{n,p,2^l} > t_l).$$

ce test. Remarquons que le test Δ_{ad} rejette l'hypothèse nulle dès que l'un des tests singuliers $\Delta_{2^l}(t_l)$ la rejette.

Pour contrôler l'erreur de première espèce, nous dérivons une inégalité de type Berry-Esseen pour $\widehat{\mathcal{D}}_{n,p,2^l}$. Cette inégalité est obtenue par le théorème de représentation de Skorokhod et le Lemme 3.3 de [51]. Ensuite, nous montrons que, si les seuils t_l sont choisis tels que $t_l = a^2\sqrt{\mathcal{C}^* \ln l}/n\sqrt{p}$ avec $\mathcal{C}^* > 4$ alors $\eta(\Delta_{ad}) \to 0$.

Nous montrons en utilisant les majorations des variances des statistiques de test $\widehat{\mathcal{D}}_{n,p,2^l}$, que si

$$a^2 n\sqrt{p} \to +\infty, \;\; 2^{L^*}/p \to 0, \;\; \ln(a^2 n\sqrt{p})/n \to 0 \text{ et } \mathcal{C}^2 > 1 + 4\sqrt{\mathcal{C}^*}$$

alors $\sup_{\alpha \in A} \beta\Big(\Delta_{ad}, \mathcal{F}(\alpha), \mathcal{C}\phi_{M,\alpha}\Big) \to 0$.

Matrice de Toeplitz

Pour le cas des matrices de Toeplitz, nous voulons tester (1.7) contre (1.10). Soit, m un entier vérifiant les conditions (1.15) pour ψ au lieu de φ. Nous proposons un test basé sur la statistique $\widehat{\mathcal{A}}_{n,p,m}$, qui est construite à partir de $\widehat{\mathcal{A}}_{n,p}$, de façon similaire à la constructions de $\widehat{\mathcal{D}}_{n,p,m}$ à partir de $\widehat{\mathcal{D}}_{n,p}$.

Nous montrons que dans ce cas, la vitesse de séparation minimax est détériorée de la même façon par le paramètre a. En effet, la vitesse de séparation minimax obtenue dans ce cas est

$$\widetilde{\Psi}_{M,\alpha} = \left(a^2 np\right)^{-\frac{2\alpha}{4\alpha+1}}.$$

Dans ce cas aussi, nous abordons le problème d'adaptation par rapport à $\alpha \in A$, où $A = [\alpha_*, \alpha^*_{n,p}] \subset]1/4, +\infty[$, où $\alpha^*_{n,p} \to +\infty$ mais en respectant la condition $\alpha^*_{n,p} = o(1)\ln(a^2 np)$. Nous considérons l'hypothèse alternative suivante

$$H_1 : \Sigma \in \bigcup_{\alpha \in A} \Big\{ \mathcal{T}(\alpha)\ ;\ \sum_{j=1}^{p-1} \sigma_j^2 \geq (\mathcal{C}\Psi_{M,\alpha})^2 \Big\}, \tag{1.17}$$

où $\mathcal{C} > 0$ et $\Psi_{M,\alpha} = \left(\dfrac{\sqrt{\ln\ln(a^2 np)}}{a^2 np} \right)^{\frac{2\alpha}{4\alpha+1}}$. Pour tester (1.7) contre (1.17), nous proposons le test adaptatif

$$\chi_{ad} = \max_{L_* \leq l \leq L^*} \mathbb{1}(\widehat{\mathcal{A}}_{n,p,2^l} > t_l),$$

où $L_*, L^* \in \mathbb{N}^*$ et vérifient les mêmes conditions que dans le cas précédent, mais pour p au lieu de $\sqrt{p}$.

Pour ce cas, nous montrons que, si $t_l = a^2 \sqrt{\mathcal{C}^* \ln l}/np$ avec $\mathcal{C}^* > 4$, alors $\eta(\chi_{ad}) \to 0$. Si, en plus,

$$a^2 np \to +\infty,\ 2^{L^*}/p \to 0,\ \ln(a^2 np)/n \to 0 \text{ et } \mathcal{C}^2 > 1 + 4\sqrt{\mathcal{C}^*},$$

alors $\sup_{\alpha \in A} \beta\Big(\chi_{ad}, \mathcal{T}(\alpha), \mathcal{C}\Psi_{M,\alpha}\Big) \to 0$.

1.3.4 Bilan des vitesses

La table 1.1 résume les résultats obtenus pour le modèle de données complètes: vitesses minimax de séparation et équivalents asymptotiques exacts. Ici n et p tendent vers l'infini, tels que $\psi \to 0$ et $p\psi^{\frac{1}{\alpha}} \to \infty$ (cas des matrices de Toeplitz) et $\varphi \to 0$ et $p\varphi^{\frac{1}{\alpha}} \to \infty$ (cas des matrices pas forcément de Toeplitz). La procédure adaptative du troisième chapitre n'est pas incluse ici.

La table 1.2 résume le modèle aux données incomplètes: vitesses minimax de séparation et vitesses atteintes par nos procédures adaptatives par agrégation des tests. Asymptotiquement, n et p tendent vers l'infini, a peut tendre vers 0, tels que, de nouveau, $\psi \to 0$ et $p\psi^{\frac{1}{\alpha}} \to \infty$ (cas des matrices de Toeplitz) et $\varphi \to 0$ et $p\varphi^{\frac{1}{\alpha}} \to \infty$ (cas des matrices pas forcément de Toeplitz).

Les bornes supérieures des vitesses adaptatives contiennent le modèle de données complètes pour $a = 1$. En particulier, même si la procédure adaptative de test est différente de celle du châpitre 3, la même perte de vitesse est observée.

Σ	Toeplitz	non-Toeplitz
Vitesse minimax de séparation	$\widetilde{\psi} = \left(C^{\frac{1}{2}}(\alpha, L) \cdot np\right)^{-\frac{2\alpha}{4\alpha+1}}$	$\widetilde{\varphi} = \left(C^{\frac{1}{2}}(\alpha, L) \cdot n\sqrt{p}\right)^{-\frac{2\alpha}{4\alpha+1}}$
Borne supérieure	Si $\alpha > 1/4$ et $n^2p^2b^2(\psi) \to +\infty$ alors $\exists \chi^*; \gamma(\chi^*, \mathcal{T}(\alpha, L), \psi) \to 0$	Si $\alpha > 1/2$ et $n^2pb^2(\varphi) \to +\infty$ alors $\exists \Delta^*;\ \gamma(\Delta^*, \mathcal{F}(\alpha, L), \varphi) \to 0$
Borne inférieure	si $\alpha > 1$ et $n^2p^2b^2(\psi) \to 0$ alors $\gamma(\psi) \to 1$	si $\alpha > 3/2$ et $n^2pb^2(\varphi) \to 0$ alors $\gamma(\varphi) \to 1$
Borne supérieure exacte	si $\alpha > 1/4$ et $n^2p^2b^2(\psi) \asymp 1$ alors $\gamma(\chi^*, \mathcal{T}(\alpha, L), \psi)$ $\leq 1 - \Phi(npt) + \Phi(np(t - b(\psi)))$	si $\alpha > 1/2$ et $n^2pb^2(\varphi) \asymp 1$ alors $\gamma(\Delta^*, \mathcal{F}(\alpha, L), \varphi)$ $\leq 1 - \Phi(n\sqrt{p}t) + \Phi(n\sqrt{p}(t - b(\varphi)))$
Borne inférieure exacte	si $\alpha > 1$ et $n^2p^2b^2(\psi) \asymp 1$ alors $\gamma(\psi) \geq 2\Phi(npb(\psi)/2)$	si $\alpha > 1$, $np\varphi^4 \to 0$ et $n^2pb^2(\varphi) \asymp 1$ alors $\gamma(\varphi) \geq 2\Phi(n\sqrt{p}b(\varphi)/2)$

Table 1.1 – Bilan des résultats obtenus pour les problèmes de test de l'hypothèse nulle $H_0 : \Sigma = I$ contre les alternatives (1.10) et (1.14) dans le cas des données complètes.

1.4 Perspectives

Les résultats décrits dans ce manuscrit peuvent être prolongés et étendus de plusieurs façons.

Vitesses de séparation non asymptotiques. Une théorie minimax non asymptotique des tests s'est développée depuis Baraud [5], Laurent, Loubes et Marteau [66] dans le modèle de signal. Des vitesses minimax non asymptotiques sont envisageables.

Cadre non gaussien. Si $X_1, \ldots, X_n$ n sont des vecteurs indépendants, tels que

$$X_i = \Gamma Z_i + \mu \qquad \text{pour tout } i = 1, \ldots, n,$$

où la moyenne $\mu \in \mathbb{R}^p$, la matrice de covariance est $\Sigma = \Gamma\Gamma^\top$ et les $Z_1, \ldots, Z_n$ sont des vecteurs i.i.d de dimension $m \geq p$ centrés et réduits. Suivant [29] qui regardent les conditions de moments sur Z_i pour que les erreurs de test convergent vers 0, nous pouvons chercher les vitesses minimax de séparation et les équivalents asymptotiques des erreurs minimax de deuxième espèce et totale.

Σ	Toeplitz	non-Toeplitz
Vitesse minimax de séparation	$\widetilde{\psi}_{M,\alpha} = \left(a^2 np\right)^{-\frac{2\alpha}{4\alpha+1}}$	$\widetilde{\varphi}_{M,\alpha} = \left(a^2 n\sqrt{p}\right)^{-\frac{2\alpha}{4\alpha+1}}$
Borne supérieure	Si $\alpha > 1/4$ et $a^4n^2p^2\psi^{4+\frac{1}{\alpha}} \to +\infty$ alors $\exists\chi$: $\gamma(\chi, \mathcal{T}(\alpha), \psi) \to 0$	Si $\alpha > 1/2$ et $a^4n^2p\varphi^{4+\frac{1}{\alpha}} \to +\infty$ alors $\exists\Delta$: $\gamma(\Delta, \mathcal{F}(\alpha), \varphi) \to 0$
Borne inférieure	Si $\alpha > 1/2$, $a^2np \to +\infty$ et $a^2np\psi^{2+\frac{1}{2\alpha}} \to 0$ alors $\gamma(\psi) \to 1$	Si $\alpha > 1/2$, $a^2n \to +\infty$, $p = o(a^2n)^{4\alpha-1}$ et $a^2n\sqrt{p}\varphi^{2+\frac{1}{2\alpha}} \to 0$ alors $\gamma(\varphi) \to 1$
Vitesse adaptative atteinte	$\Psi_{M,\alpha} = \left(\frac{\sqrt{\ln\ln(a^2np)}}{a^2np}\right)^{\frac{2\alpha}{4\alpha+1}}$	$\phi_{M,\alpha} = \left(\frac{\sqrt{\ln\ln(a^2n\sqrt{p})}}{a^2n\sqrt{p}}\right)^{\frac{2\alpha}{4\alpha+1}}$
Borne supérieure adaptative	Si $a^2np \to +\infty$, $2^{L^*}/p \to 0$, $\ln(a^2np)/n \to 0$, $\mathcal{C}^* > 4$, et $\mathcal{C}^2 \geq 1 + 4\sqrt{\mathcal{C}^*}$ alors $\sup_{\alpha\in A} \gamma(\chi_{ad}, \mathcal{T}(\alpha), \mathcal{C}\Psi_{M,\alpha}) \to 0$	Si $a^2n\sqrt{p} \to +\infty$, $2^{L^*}/p \to 0$, $\ln(a^2n\sqrt{p})/n \to 0$, $\mathcal{C}^* > 4$, et $\mathcal{C}^2 \geq 1 + 4\sqrt{\mathcal{C}^*}$ alors $\sup_{\alpha\in A} \gamma(\Delta_{ad}, \mathcal{F}(\alpha), \mathcal{C}\phi_{M,\alpha}) \to 0$

Table 1.2 – Bilan des résultats obtenus pour les problèmes de test de l'hypothèse nulle $H_0 : \Sigma = I$ contre les alternatives (1.10), (1.14), (1.16) et (1.17) dans le cas des données incomplètes.

Tests d'adéquation: tester $H_0 : \Sigma = \Sigma_0$, pour Σ_0 une matrice définie positive donnée. Une façon de procéder est de changer d'échelle en posant $Z_i = \Sigma_0^{-1/2} X_i$ pour i de 1 à n et tester l'hypothèse que la covariance des Z_i est l'identité. Dans ce cas, l'hypothèse alternative porte sur les Z_i. Une procédure utilisant les X_i pourrait s'envisager pour cette raison. Le cas de Σ_0 sparse reste à considérer (p. ex., Σ_0 de petit rang).

Séries temporelles multivariées. L'estimation des matrices de densité spectrale pour les séries temporelles multivariées a été considérée, entre autres, par [11], [12] et [38]. Dans un premier temps, nous pourrons considérer des vecteurs gaussiens de même loi, mais pas forcément indépendants et tester la matrice de covariance de $(X_1^\top, \ldots, X_n^\top)$. Il serait également intéressant de proposer des tests d'adéquation ou d'homogénéité pour les matrices de densités spectrales.

Hypothèse nulle composite. Le problème de test avec une hypothèse nulle composite a été considéré pour le modèle de régression dans [30], [44], [61], pour le modèle de densité dans [80], [42] et le modèle de signal [79]. Des tests d'hypothèses qualitatives ont été proposés par [33], [6]. On observe des fois des pertes de vitesses dans ce cas et ce n'est pas toujours évident quand ces pertes sont inévitables où simplement dues à des techniques de preuves. Dans le modèle de matrice de covariance de vecteurs gaussiens ou pas nécessairement gaussiens, si la moyenne des observations est inconnue nous sommes en présence d'une hypothèse nulle composite. D'autres hypothèses nulles composites sur la matrice de covariance sont: la sphéricité $\Sigma = \sigma^2 \cdot I$ avec $\sigma > 0$ inconnu, l'homogénéité $\Sigma_1 = \Sigma_2$, etc.

Chapter 2

Sharp minimax tests for large Toeplitz covariance matrices with repeated observations

Abstract.

We observe a sample of n independent p-dimensional Gaussian vectors with Toeplitz covariance matrix $\Sigma = [\sigma_{|i-j|}]_{1 \le i,j \le p}$ and $\sigma_0 = 1$. We consider the problem of testing the hypothesis that Σ is the identity matrix asymptotically when $n \to \infty$ and $p \to \infty$. We suppose that the covariances σ_k decrease either polynomially ($\sum_{k \ge 1} k^{2\alpha} \sigma_k^2 \le L$ for $\alpha > 1/4$ and $L > 0$) or exponentially ($\sum_{k \ge 1} e^{2Ak} \sigma_k^2 \le L$ for $A, L > 0$).

We consider a test procedure based on a weighted U-statistic of order 2, with optimal weights chosen as solution of an extremal problem. We give the asymptotic normality of the test statistic under the null hypothesis for fixed n and $p \to +\infty$ and the asymptotic behavior of the type I error probability of our test procedure. We also show that the maximal type II error probability, either tend to 0, or is bounded from above. In the latter case, the upper bound is given using the asymptotic normality of our test statistic under alternatives close to the separation boundary. Our assumptions imply mild conditions: $n = o(p^{2\alpha - 1/2})$ (in the polynomial case), $n = o(e^p)$ (in the exponential case).

We prove both rate optimality and sharp optimality of our results, for $\alpha > 1$ in the polynomial case and for any $A > 0$ in the exponential case.

A simulation study illustrates the good behavior of our procedure, in particular for small n, large p.

2.1 Introduction

In the last decade, both functional data analysis (FDA) and high-dimensional (HD) problems have known an unprecedented expansion both from a theoretical point of view (as they offer many mathematical challenges) and for the applications (where data have complex structure and grow larger every day). Therefore, both areas share a large number of trends, see [13] and the review by [32], like regression models with functional or large-dimensional covariates, supervised or unsupervised classification, testing procedures, covariance operators.

Functional data analysis proceeds very often by discretizing curve datasets in time domain or by projecting on suitable orthonormal systems and produces large dimensional vectors with size possibly larger than the sample size. Hence methods and techniques from HD problems can be successfully implemented (see e.g. [2]).However, in some cases, HD vectors can be transformed into stochastic processes, see [28], and then techniques from FDA bring new insights into HD problems. Our work is of the former type.

We observe independent, identically distributed Gaussian vectors $X_1, ..., X_n$, $n \geq 2$, which are p-dimensional, centered and with a positive definite Toeplitz covariance matrix Σ. We denote by $X_k = (X_{k,1}, ..., X_{k,p})^\top$ the coordinates of the vector X_k in $\mathbb{R}^p$ for all k.

Our model is that of a stationary Gaussian time series, repeatedly and independently observed n times, for $n \geq 2$. We assume that n and p are large. In functional data analysis, it is quite often that curves are observed in an independent way: electrocardiograms of different patients, power supply for different households and so on, see other data sets in [13]. After modelisation of the discretized curves, the statistician will study the normality and the whiteness of the residuals in order to validate the model. Our problem is to test from independent samples of high-dimensional residual vectors that the standardized Gaussian coordinates are uncorrelated.

Let us denote by $\sigma_{|j|} = Cov(X_{k,h}, X_{k,h+j})$, for all integer numbers h and j, for all $k \in \mathbb{N}^*$, where $\mathbb{N}^*$ is the set of positive integers. We assume that $\sigma_0 = 1$, therefore σ_j are correlation coefficients. We recall that $\{\sigma_j\}_{j\in\mathbb{N}}$ is a sequence of non-negative type, or, equivalently, the associated Toeplitz matrix Σ is non-negative definite. We assume that the sequence $\{\sigma_j\}_{j\in\mathbb{N}}$ belongs to to $\ell_1(\mathbb{N}) \cap \ell_2(\mathbb{N})$, where $\ell_1(\mathbb{N})$ (resp. $\ell_2(\mathbb{N})$) is the set all absolutely (resp. square) summable sequences. It is therefore possible to construct a positive, periodic function

$$f(x) = \frac{1}{2\pi}\left(1 + 2\sum_{j=1}^{\infty} \sigma_j \cos(jx)\right), \qquad \text{for } x \in (-\pi, \pi),$$

belonging to $\mathbb{L}_2(-\pi,\pi)$ the set of all square-integrable functions f over $(-\pi,\pi)$. This function is known as the spectral density of the stationary series $\{X_{k,i}, i \in \mathbb{Z}\}$.

We solve the following test problem,

$$H_0 : \Sigma = I \tag{2.1}$$

versus the alternative

$$H_1 : \Sigma \in \mathcal{T}(\alpha, L) \text{ such that } \sum_{j \geq 1} \sigma_j^2 \geq \psi^2, \tag{2.2}$$

for $\psi = (\psi_{n,p})_{n,p}$ a positive sequence converging to 0. From now on, $C_{>0}$ denotes the set of squared symmetric and positive definite matrices. The set $\mathcal{T}(\alpha, L)$ is an ellipsoid of Sobolev type

$$\mathcal{T}(\alpha, L) = \{\Sigma \in C_{>0}, \Sigma \text{ is Toeplitz };\sum_{j \geq 1} \sigma_j^2 j^{2\alpha} \leq L \text{ and } \sigma_0 = 1\},\ \alpha > 1/4,\ L > 0.$$

We shall also test (2.1) against

$$H_1 : \Sigma \in \mathcal{E}(A, L) \text{ such that } \sum_{j \geq 1} \sigma_j^2 \geq \psi^2, \text{ for } \psi > 0, \tag{2.3}$$

where the ellipsoid of covariance matrices is given by

$$\mathcal{E}(A, L) = \{\Sigma \in C_{>0}, \Sigma \text{ is Toeplitz };\sum_{j \geq 1} \sigma_j^2 e^{2Aj} \leq L \text{ and } \sigma_0 = 1\}, A, L > 0.$$

This class contains the covariance matrices whose elements decrease exponentially, when moving away from the diagonal. We denote by $G(\psi)$ either $G(\mathcal{T}(\alpha, L), \psi)$ the set of matrices under the alternative (2.2) or $G(\mathcal{E}(A, L), \psi)$ under the alternative (2.3).

We stress the fact that a matrix Σ in $G(\psi)$ is such that $1/(2p)\|\Sigma - I\|_F^2 \geq \sum_{j \geq 1} \sigma_j^2 \geq \psi^2$, i.e. Σ is outside a neighborhood of I with radius ψ in Frobenius norm.

Our test can be applied in the context of model fitting for testing the whiteness of the standard Gaussian residuals. In this context, it is natural to assume that the covariance matrix under the alternative hypothesis has small entries like in our classes of covariance matrices. Such tests have been proposed by [40], where it is noted that weighted test statistics can be more powerful.

Note that, most of the literature on testing the null hypothesis (2.1), either focus on finding the asymptotic behavior of the test statistic under the null hypothesis, or control in addition the type II error probability for one fixed unknown matrix under the alternative, whereas our main interest is to quantify the worst type II error probabilities, i.e. uniformly over a large set of possible covariance matrices.

Various test statistics in high dimensional settings have been considered for testing (2.1), as it was known for some time that likelihood ratio tests do not converge when dimension grows. Therefore, a corrected Likelihood Ratio Test is proposed in [4] when $p/n \to c \in (0, 1)$, and its asymptotic behavior is given under the null hypothesis, based on the random matrix theory. In [62] the result is extended to $c = 1$. An exact test based on one column of the covariance matrix is constructed by [49]. A series of papers propose test statistics based on the Frobenius norm of $\Sigma - I$, see [67], [86], [89] and [29]. Different test statistics are introduced and their asymptotic distribution is studied. In particular in [29] the test statistic is a U-statistic with constant weights. An unbiased estimator of $tr(\Sigma - B_k(\Sigma))^2$ is constructed in [81], where $B_k(\Sigma) = (\sigma_{ij} \cdot I\{|i-j| \leq k\})$, in order to develop a test statistic for the problem of testing the bandedness of a given matrix. Another extension of our test problem is to test the sphericity hypothesis $\Sigma = \sigma^2 I$, where $\sigma^2 > 0$ is unknown. [41] introduced a test statistic based on functionals of order 4 of the covariance matrix. Motivated by these results, the test $H_0 : \Sigma = I$ is revisited by [39]. The maximum value of non-diagonal elements of the empirical covariance matrix was also investigated as a test statistic. Its asymptotic extreme-value distribution was given under the identity covariance matrix by [22] and for other covariance matrices by [93]. We propose here a new test statistic to test (2.1) which is a weighted U-statistic of order 2 and study its probability errors uniformly over the set of matrices given by the alternative hypothesis.

The test problem with alternative (2.2) and with one sample ($n = 1$) was solved in the sharp asymptotic framework, as $p \to \infty$, by [37]. Indeed, [37] studies sharp minimax testing of the spectral density f of the Gaussian process. Note that under the null hypothesis we have a constant spectral density $f_0(x) = 1/(2\pi)$ for all x and the alternative can be described in $\mathbb{L}_2$ norm as we have the following isometry $\|f - f_0\|_2^2 = (2\pi)^{-1}\|\Sigma - I\|_F^2$. Moreover, the ellipsoid of covariance matrices $\mathcal{T}(\alpha, L)$ are in bijection with Sobolev ellipsoids of spectral densities f. Let us also recall that the adaptive rates for minimax testing are obtained for the spectral density problem by [47] by a non constructive method using the asymptotic equivalence with a Gaussian white noise model. Finding explicit test procedures which adapt automatically to parameters α and/or L of our class of matrices will be the object of future work. Our efforts go here into finding sharp minimax rates for testing.

Our results generalize the results in [37] to the case of repeatedly observed stationary Gaussian process. We stress the fact that repeated sampling of the stationary process $(X_{1,1}, ..., X_{1,p})$ to $(X_{n,1}, ..., X_{n,p})$ can be viewed as one sample of size $n \times p$ under the null hypothesis. However, this sample will not fit the assumptions of our alternative. Indeed, under the alternative, its covariance matrix is not Toeplitz, but block diagonal. Moreover, we can summarize the n independent vectors into one p-dimensional vector $X = n^{-1/2} \sum_{k=1}^{n} X_k$ having Gaussian distribution $\mathcal{N}_p(0, \Sigma)$. The results by [37] will produce a test procedure with rate that we expect optimal as a function of p, but more biased and suboptimal as a function of n. The test statistic that we suggest removes cross-terms and has smaller bias. Therefore, results in [37] do not apply in a straightforward way to our setup.

A conjecture in the sense of asymptotic equivalence of the model of repeatedly observed Gaussian vectors and a Gaussian white noise model was given by [21]. Our rates go in the sense of the conjecture.

The test of $H_0 : \Sigma = I$ against (2.2), with Σ not necessary Toeplitz, is given in [18]. Their rates show a loss of a factor p when compared to the rates for Toeplitz matrices obtained here. This can be interpreted heuristically by the size of the set of unknown parameters which is $p(p-1)/2$ for [18] whereas here it is p. We can see that the family of Toeplitz matrices is a subfamily of general covariance matrices in [18]. Therefore, the lower bounds are different, they are attained through a particular family of Toeplitz large covariance matrices. The upper bounds take into account as well the fact that we have repeated information on the same diagonal elements. The test statistic is different from the one used in [18].

The test problem with alternative hypothesis (2.3) has not been studied in this model. The class $\mathcal{E}(A, L)$ contains matrices with exponentially decaying elements when further from the main diagonal. The spectral density function associated to this process belongs to the class of functions which are in $\mathbb{L}_2$ and admit an analytic continuation on the strip of complex numbers z with $|Im(z)| \le A$. Such classes of analytic functions are very popular in the literature of minimax estimation, see [48] .

In times series analysis such covariance matrices describe among others the linear ARMA processes. The problem of adaptive estimation of the spectral density of an ARMA process has been studied by [46] (for known α) and adaptively to α via wavelet based methods by [78] and by model selection by [31]. In the case of an ARFIMA process, obtained by fractional differentiation of order $d \in (-1/2, 1/2)$ of a casual invertible ARMA process, [84] gave adaptive estimators of the spectral density based on the log-periodogram regression model when the covariance matrix belongs to $\mathcal{E}(A, L)$.

Before describing our results let us define more precisely the quantities we are interested in evaluating.

2.1.1 Formalism of the minimax theory of testing

Let χ be a test, that is a measurable function of the observations $X_1, \dots, X_n$ taking values in $\{0, 1\}$ and recall that $G(\psi)$ corresponds to the set of covariance matrice under the alternative hypothesis. Let

$$\begin{aligned} \eta(\chi) &= \mathbb{E}_I(\chi) \quad \text{be its type I error probability, and} \\ \beta(\chi, G(\psi)) &= \sup_{\Sigma \in G(\psi)} \mathbb{E}_\Sigma(1-\chi) \quad \text{be its maximal type II error probability.} \end{aligned}$$

Σ	$\mathcal{T}(\alpha, L)$	$\mathcal{E}(A, L)$	not Toeplitz and $\mathcal{T}(\alpha, L)$ [18]
$\widetilde{\psi}$	$\left(C(\alpha, L)\cdot n^2p^2\right)^{-\frac{\alpha}{4\alpha+1}}$	$\left(\dfrac{2\ln(n^2p^2)}{An^2p^2}\right)^{1/4}$	$\left(C(\alpha, L)\cdot n^2p\right)^{-\frac{\alpha}{4\alpha+1}}$
$b(\psi)^2$	$C(\alpha, L)\cdot \psi^{\frac{4\alpha+1}{\alpha}}$	$\dfrac{A\psi^4}{2\ln\left(\frac{1}{\psi}\right)}$	$C(\alpha, L)\cdot \psi^{\frac{4\alpha+1}{\alpha}}$

Table 2.1 – Separation rates $\widetilde{\psi}$ and $b(\psi)$ in the sharp asymptotic bounds where $C(\alpha, L) = (2\alpha+1)(4\alpha+1)^{-(1+\frac{1}{2\alpha})}L^{-\frac{1}{2\alpha}}$.

We consider two criteria to measure the performance of the test procedure. The first one corresponds to the classical Neyman-Pearson criterion. For $w \in (0,1)$, we define,

$$\beta_w(G(\psi)) = \inf_{\chi\,;\,\eta(\chi)\le w} \beta(\chi, G(\psi)).$$

The test χ_w is asymptotically minimax according to the Neyman-Pearson criterion if

$$\eta(\chi_w) \le w + o(1) \quad \text{and} \quad \beta(\chi_w\,, G(\psi)) = \beta_w(G(\psi)) + o(1).$$

The second criterion is the total error probability, which is defined as follows:

$$\gamma(\,\chi\,, G(\psi)) = \eta(\chi) + \beta(\,\chi\,, G(\psi)).$$

Define also the minimax total error probability γ as $\gamma(G(\psi)) = \inf_{\chi} \gamma(\,\chi\,, G(\psi))$, where the infimum is taken over all possible tests.

Note that the two criteria are related since $\gamma(G(\psi)) = \inf_{w\in(0,1)}(w + \beta_w(G(\psi)))$ (see Ingster and Suslina [58]).

A test χ is asymptotically minimax if: $\gamma(G(\psi)) = \gamma(\,\chi\,, G(\psi)) + o(1)$. We say that $\widetilde{\psi}$ is a (asymptotic) separation rate, if the following lower bounds hold

$$\gamma(G(\psi)) \longrightarrow 1 \quad \text{as } \frac{\psi}{\widetilde{\psi}} \longrightarrow 0$$

together with the following upper bounds: there exists a test χ such that,

$$\gamma(\,\chi\,, G(\psi)) \longrightarrow 0 \quad \text{as } \frac{\psi}{\widetilde{\psi}} \longrightarrow +\infty.$$

The sharp optimality corresponds to the study of the asymptotic behavior of the maximal type II error probability $\beta_w(G(\psi))$ and the total error probability $\gamma(G(\psi))$. In our study we obtain asymptotic behavior of Gaussian type, i.e. we show that, under some assumptions,

$$\beta_w(G(\psi)) = \Phi(z_{1-w} - npb(\psi)) + o(1) \quad \text{and} \quad \gamma(G(\psi)) = 2\Phi(-npb(\psi)) + o(1), \tag{2.4}$$

where Φ is the cumulative distribution function of a standard Gaussian random variable, z_{1-w} is the $1-w$ quantile of the standard Gaussian distribution for any $w \in (0,1)$, and $b(\psi)$ has an explicit form for each ellipsoid of Toeplitz covariance matrices.

Separation rates and sharp asymptotic results for different testing problem were studied under this formalism by [56]. We refer for precise definitions of sharp asymptotic and non asymptotic rates to [74]. Note that throughout this paper, asymptotics and symbols o, O, $\sim$ and $\asymp$ are considered as p tends to infinity, unless we specify that n tends to infinity. Recall that, given sequences of real numbers u and real positive numbers v, we say that they are asymptotically equivalent, $u \sim v$, if $\lim u/v = 1$. Moreover, we say that the sequences are asymptotically of the same order, $u \asymp v$, if there exist two constants $0 < c \leq C < \infty$ such that $c \leq \liminf u/v$ and $\limsup u/v \leq C$.

2.1.2 Overview of the results

In this paper, we describe the separation rates $\widetilde{\psi}$ and sharp asymptotics for the error probabilities for testing the identity matrix against $G(\mathcal{T}(\alpha, L), \psi)$ and $G(\mathcal{E}(A, L), \psi)$ respectively.

We propose here a test procedure whose type II error probability tends to 0 uniformly over the set of $G(\psi)$, that is even for a covariance matrix that gets closer to the identity matrix at distance $\widetilde{\psi} \to 0$ as n and p increase. The radius $\widetilde{\psi}$ in Table 2.1 is the smallest vicinity around the identity matrix which still allows testing error probabilities to tend to 0. Our test statistic is a weighted quadratic form and we show how to choose these weights in an optimal way over each class of alternative hypotheses.

Under mild assumptions we obtain the sharp optimality in (2.4), where $b(\psi)$ is described in Table 2.1 and compared to the case of non Toeplitz matrices in [18].

This paper is structured as follows. In Section 2.2, we study the test problem with alternative hypothesis defined by the class $G(\mathcal{T}(\alpha, L), \psi)$, $\alpha > 1/4$, L, $\psi > 0$. We define explicitly the test statistic and give its first and second moments under the null and the alternative hypotheses. We derive its Gaussian asymptotic behavior under the null hypothesis and under the alternative submitted to the constraints that ψ is close to the separation rate $\widetilde{\psi}$ and that Σ is closed to the solution of an extremal problem Σ^*. We deduce the asymptotic separation rates. Their optimality is shown only for $\alpha > 1$. Our lower bounds are original in the literature of minimax lower bounds, as in this case we cannot reduce the proof to the vector case, or diagonal matrices. We give the sharp rates for $\psi \asymp \widetilde{\psi}$. Our assumptions imply that necessarily $n = o(p^{2\alpha - 1/2})$ as $p \to \infty$. That does not prevent n to be larger than p for sufficiently large α.

In Section 2.3, we derive analogous results over the class $G(\mathcal{E}(A, L), \psi)$, with A, $L, \psi > 0$. We show how to choose the parameters in this case and study the test procedure similarly. We give asymptotic separation rates. The sharp bounds are attained as $\psi \asymp \widetilde{\psi}$.

Our assumptions involve that $n = o(\exp(p))$ which allows n to grow exponentially fast with p. That can be explained by the fact that the elements of Σ decay much faster over exponential ellipsoids than over the polynomial ones. In Section 2.4 we implement our procedure and show the power of testing over two families of covariance matrices.

The proofs of our results are postponed to the Section 2.5 and to the Supplementary material.

2.2 Testing procedure and results for polynomially decreasing covariances

We introduce a weighted U-statistic of order 2, which is an estimator of the functional $\sum_{j\geq 1} \sigma_j^2$ that defines the separation between a Toeplitz covariance matrix under the alternative hypothesis from the identity matrix under the null. Indeed, in nonparametric estimation of quadratic functionals such as $\sum_{j\geq 1} \sigma_j^2$ weighted estimators are often considered (see e.g. [17]). These weights have finite support of length T, where T is optimal in some sense. Intuitively, as the coefficients $\{\sigma_j\}_j$ belong to an ellipsoid, they become smaller when j increases and thus the bias due to the truncation and the weights becomes as small as the variance for estimating the weighted finite sum.

2.2.1 Test Statistic

Let us denote by $T_p(\{\sigma_j\}_{j\geq 1})$ the symmetric $p \times p$ Toeplitz matrix $\Sigma = [\sigma_{lk}]_{1\leq l,k\leq p}$ such that the diagonal elements of Σ are equal to 1, and $\sigma_{lk} = \sigma_{kl} = \sigma_{|l-k|}$, for all $l \neq k$. Now we define the weighted test statistic in this setup

$$\widehat{\mathcal{A}}_n := \widehat{\mathcal{A}}_n^{\mathcal{T}} = \frac{1}{n(n-1)(p-T)^2} \sum_{1\leq k\neq l\leq n} \sum_{j=1}^{T} w_j^* \sum_{T+1\leq i_1,i_2\leq p} X_{k,i_1} X_{k,i_1-j} X_{l,i_2} X_{l,i_2-j} \tag{2.5}$$

where the weights $\{w_j^*\}_j$ and the parameters $T, \lambda, b^2(\psi)$ are obtained by solving the following extremal problem:

$$b(\psi) := \sum_{j\geq 1} w_j^* \sigma_j^{*2} = \sup_{\left\{\substack{(w_j)_j \;:\; w_j\geq 0; \\ \sum_{j\geq 1} w_j^2 = \frac{1}{2}}\right\}} \inf_{\left\{\substack{\Sigma \;:\; \Sigma = T_p(\{\sigma_j\}_{j\geq 1}); \\ \Sigma\in\mathcal{T}(\alpha,L), \;\; \sum_{j\geq 1}\sigma_j^2\geq\psi^2}\right\}} \sum_{j\geq 1} w_j \sigma_j^2. \tag{2.6}$$

This extremal problem appears heuristically as we want that the expected value of our test statistic for the worst parameter Σ under the alternative hypothesis (closest to the null) to be as large as possible for the weights we use. This problem will provide the optimal weights $\{w_j^*\}_{j\geq 1}$ in order to control the worst type II error probability, but also the critical matrix $\Sigma^* = T_p(\{\sigma_j^*\})$ that will be used in the lower bounds. Indeed, Σ^* is positive definite for small enough ψ (see [18]).

The solution of the extremal problem (2.6) can be found in [58]:

$$
\begin{aligned}
w_j^* &= \frac{\lambda}{2b(\psi)}\Big(1-(\frac{j}{T})^{2\alpha}\Big), \quad \sigma_j^{*2} = \lambda\Big(1-(\frac{j}{T})^{2\alpha}\Big), \quad T = \lfloor (L(4\alpha+1))^{\frac{1}{2\alpha}} \cdot \psi^{-\frac{1}{\alpha}} \rfloor \\
\lambda &= \frac{2\alpha+1}{2\alpha(L(4\alpha+1))^{\frac{1}{2\alpha}}} \cdot \psi^{\frac{2\alpha+1}{\alpha}}, \quad b^2(\psi) = \frac{1}{2}\sum_j \sigma_j^{*4} = \frac{2\alpha+1}{L^{\frac{1}{2\alpha}}(4\alpha+1)^{1+\frac{1}{2\alpha}}} \cdot \psi^{\frac{4\alpha+1}{\alpha}}
\end{aligned}
\tag{2.7}
$$

Remark that T is a finite number but grows to infinity as $\psi \to 0$. Moreover, the test statistic will have optimality properties under the additional condition that $T/p \to 0$ which is equivalent to $p\psi^{1/\alpha} \to \infty$. It is obvious that in practice it might happen that $T \geq p$ and then we have no solution but to use $T = p-1$, with the inconvenient that the procedure does not behave as well as the theory predicts.

Proposition 2.1. *Under the null hypothesis, the test statistic $\widehat{\mathcal{A}}_n$ is centered, $\mathbb{E}_I(\widehat{\mathcal{A}}_n) = 0$, with variance :*

$$
Var_I(\widehat{\mathcal{A}}_n) = \frac{1}{n(n-1)(p-T)^2}.
$$

Moreover, under the alternative hypothesis with $\alpha > 1/4$, if we assume that $\psi \to 0$ we have:

$$
\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) = \sum_{j=1}^{T} w_j^* \sigma_j^2 \geq b(\psi) \quad \text{and} \quad Var_\Sigma(\widehat{\mathcal{A}}_n) = \frac{R_1}{n(n-1)(p-T)^4} + \frac{R_2}{n(p-T)^2},
$$

uniformly over Σ in $G(\mathcal{T}(\alpha, L), \psi)$, where

$$
R_1 \leq (p-T)^2 \cdot \{1 + o(1) + \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) \cdot (O(\sqrt{T}) + O(T^{3/2-2\alpha})) + \mathbb{E}_\Sigma^2(\widehat{\mathcal{A}}_n) \cdot O(T^2)\}
\tag{2.8}
$$

$$
R_2 \leq (p-T) \cdot \{\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) \cdot o(1) + \mathbb{E}_\Sigma^{3/2}(\widehat{\mathcal{A}}_n) \cdot (O(T^{1/4}) + O(T^{3/4-\alpha})) + \mathbb{E}_\Sigma^2(\widehat{\mathcal{A}}_n) \cdot O(T)\}.
\tag{2.9}
$$

In the next Proposition we prove asymptotic normality of the test statistic under the null and under the alternative hypothesis with additional assumptions. More precisely, we need that ψ is of the same order as the separation rate and that the matrix Σ is close to the optimal Σ^*. This is not a drawback, since the asymptotic constant for probability errors are attained under the same assumptions or tend to 0 otherwise.

Proposition 2.2. *Suppose that n, $p \to +\infty$, $\alpha > 1/4$, $\psi \to 0$, $p\psi^{1/\alpha} \to +\infty$ and moreover assume that $n(p-T)b(\psi) \asymp 1$, the test statistic $\widehat{\mathcal{A}}_n$ defined by (2.5) with parameters given in (2.7), verifies :*

$$
n(p-T)\Big(\widehat{\mathcal{A}}_n - \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n)\Big) \longrightarrow \mathcal{N}(0,1)
$$

for all $\Sigma \in G(\mathcal{T}(\alpha, L), \psi)$, such that $\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) = O(b(\psi))$.

Moreover, $n(p-T)\widehat{\mathcal{A}}_n$ has asymptotical $\mathcal{N}(0,1)$ distribution under H_0, as $p \to \infty$ for any fixed $n \geq 2$.

2.2.2 Separation rate and sharp asymptotic optimality

Based on the test statistic $\widehat{\mathcal{A}}_n$, we define the test procedure

$$\chi^* = \chi^*(t) = \mathbb{1}(\widehat{\mathcal{A}}_n > t), \tag{2.10}$$

for conveniently chosen $t > 0$, where $\widehat{\mathcal{A}}_n$ is the estimator defined in (2.5) with parameters in (2.7).

The next theorem gives the separation rate under the assumption that $T = o(p)$, or equivalently, that $p\psi^{1/\alpha} \to \infty$. The upper bounds are attained for arbitrary $\alpha > 1/4$, but the lower bounds require $\alpha > 1$.

Theorem 2.3. *Suppose that asymptotically*

$$\psi \to 0 \quad \text{and} \quad p\psi^{1/\alpha} \to +\infty \tag{2.11}$$

Lower bound. *If* $\alpha > 1$ *and* $n^2p^2\, b^2(\psi) = C(\alpha, L)n^2p^2\,\psi^{\frac{4\alpha+1}{\alpha}} \to 0$ *then*

$$\gamma = \inf_{\chi} \gamma(\,\chi\,, G(\mathcal{T}(\alpha, L), \psi)) \longrightarrow 1,$$

where the infimum is taken over all test statistics χ.

Upper bound. *The test procedure* χ^* *defined in (2.10) with* $t > 0$ *has the following properties:*

Type I error probability : if $np \cdot t \to +\infty$ *then* $\eta(\chi^*) \to 0$.

Type II error probability : if

$$\alpha > 1/4 \quad \text{and} \quad n^2p^2\, b^2(\psi) = C(\alpha, L)n^2p^2\,\psi^{\frac{4\alpha+1}{\alpha}} \to +\infty \tag{2.12}$$

then, uniformly over t *such that* $t \le c \cdot C^{1/2}(\alpha, L) \cdot \psi^{\frac{4\alpha+1}{2\alpha}}$ *, for some constant* $0 < c < 1$, *we have*

$$\beta(\,\chi^*, G(\mathcal{T}(\alpha, L), \psi)) \longrightarrow 0.$$

Under the assumptions given in (2.11) and (2.12), with t verifying the assumptions of Theorem 2.3, we get :

$$\gamma(\,\chi^*\,, G(\mathcal{T}(\alpha, L), \psi)) \longrightarrow 0$$

As a consequence of the previous theorem, we get that χ^* is an asymptotically minimax test procedure if $\psi/\widetilde{\psi} \longrightarrow +\infty$. From the lower bounds we deduce that, if $\psi/\widetilde{\psi} \longrightarrow 0$, there is no test procedure to distinguish between the null and the alternative hypotheses, with errors tending to 0. The minimax separation rate $\widetilde{\psi}$ is therefore :

$$\widetilde{\psi} = \left(\frac{2\alpha + 1}{L^{\frac{1}{2\alpha}}(4\alpha+1)^{1+\frac{1}{2\alpha}}} \cdot n^2p^2 \right)^{-\frac{\alpha}{4\alpha+1}} \tag{2.13}$$

It is obtained from the relation $n^2p^2b^2(\psi) = 1$. Naturally the constant does not play any role here. Remark that the condition $T/p \to 0 \asymp p\widetilde{\psi}^{1/\alpha} \to +\infty$ implies that $n = o(p^{2\alpha - \frac{1}{2}})$.

The maximal type II error probability either tends to 0, see Theorem 2.3, or is less than $\Phi(np(t - b(\psi))) + o(1)$ when $npt < npb(\psi) \asymp 1$. The latter case is the object of the next theorem giving sharps bounds for the asymptotic errors. The upper bounds are attained for arbitrary $n \geq 2$ and for $\alpha > 1/4$, while our proof of the sharp lower bounds requires additionally that $n \to \infty$ and $\alpha > 1$.

Theorem 2.4. *Suppose that $\psi \to 0$ such that $p/T \asymp p\psi^{1/\alpha} \to +\infty$ and, moreover, that*

$$n^2 p^2 b^2(\psi) \asymp 1. \tag{2.14}$$

Lower bound. *If $\alpha > 1$, then*

$$\inf_{\chi:\eta(\chi)\leq w} \beta(\chi, G(\mathcal{T}(\alpha, L), \psi)) \geq \Phi(z_{1-w} - npb(\psi)) + o(1),$$

where the infimum is taken over all test statistics χ with type I error probability less than or equal to w. Moreover,

$$\gamma = \inf_{\chi} \gamma(\chi, G(\mathcal{T}(\alpha, L), \psi)) \geq 2\Phi(-np\frac{b(\psi)}{2}) + o(1).$$

Upper bound. *The test procedure χ^* defined in (2.10) with $t > 0$ has the following properties.*
Type I error probability : $\eta(\chi^*) = 1 - \Phi(np \cdot t) + o(1)$.
Type II error probability : under the assumption (2.14), and for all $\alpha > 1/4$, we have that, uniformly over t :

$$\beta(\chi^*, G(\mathcal{T}(\alpha, L), \psi)) \leq \Phi(np \cdot (t - b(\psi))) + o(1).$$

In particular, for $t = t^w$, such that $np \cdot t^w = z_{1-w}$, we have $\eta(\chi^*(t^w)) \leq w + o(1)$ and also,

$$\beta(\chi^*(t^w), G(\mathcal{T}(\alpha, L), \psi)) = \Phi(z_{1-w} - np \cdot b(\psi)) + o(1).$$

Another important consequence of the previous theorem, is that the test procedure χ^*, with $t^* = b(\psi)/2$ is such that

$$\gamma(\chi^*(t^*), G(\mathcal{T}(\alpha, L), \psi)) = 2\,\Phi\left(-np\frac{b(\psi)}{2}\right) + o(1).$$

Then we can deduce that the minimax separation rate $\widetilde{\psi}$ defined in (2.13) is sharp.

2.3 Exponentially decreasing covariances

In this section we want to test (2.1) against (2.3), where the alternative set is $G(\mathcal{E}(A, L), \psi)$, for some $A, L, \psi > 0$. It is well known in the nonparametric minimax theory that $\mathcal{E}(A, L)$ is in bijection with ellipsoids of analytic spectral densities admiting analytic continuation on the strip $\{z \in \mathbb{C} : |Im(z)| \le A\}$ of the complex plane. On this class nearly parametric rates are attained for testing in the Gaussian noise model, see Ingster [53].

Let us define $\widehat{\mathcal{A}}_n^{\mathcal{E}}$ in (2.5)

$$\widehat{\mathcal{A}}_n^{\mathcal{E}} = \frac{1}{n(n-1)(p-T)^2} \sum_{1\le k\ne l\le n} \sum_{j=1}^{T} w_j^* \sum_{T+1\le i_1,i_2\le p} X_{k,i_1} X_{k,i_1-j} X_{l,i_2} X_{l,i_2-j}, \tag{2.15}$$

where the weights $\{w_j^*\}_{j\ge1}$, are obtained by solving the optimization problem (2.6), with the class $\mathcal{T}(\alpha, L)$ replaced by $\mathcal{E}(A, L)$. The solution given in [53] is as follows :

$$\begin{aligned} w_j^* &= \frac{\lambda}{2b(\psi)}\Big(1-(\frac{e^j}{e^T})^{2A}\Big)_+, \quad \sigma_j^* = \sqrt{\lambda}\Big(1-(\frac{e^j}{e^T})^{2A}\Big)_+^{1/2}, \quad T = \Big\lfloor \frac{1}{A}\ln\Big(\frac{1}{\psi}\Big)\Big\rfloor, \\ \lambda &= \frac{A\psi^2}{\ln\Big(\dfrac{1}{\psi}\Big)}, \quad b^2(\psi) = \frac{A\psi^4}{2\ln\Big(\dfrac{1}{\psi}\Big)}. \end{aligned} \tag{2.16}$$

Note that all parameters above are free of the radius $L > 0$. Moreover, we have :

$$\sup_j w_j^* \le \frac{\lambda}{2b(\psi)} \asymp \frac{1}{2(\ln(1/\psi))^{1/2}} \longrightarrow 0$$

Under the null hypothesis, we still have $\mathbb{E}_I(\widehat{\mathcal{A}}_n^{\mathcal{E}}) = 0\,, \mathrm{Var}_I(\widehat{\mathcal{A}}_n^{\mathcal{E}}) = 1/(n(n-1)(p-T)^2)$ and

$$n(p-T)\widehat{\mathcal{A}}_n^{\mathcal{E}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \quad \text{for fixed } n \ge 2 \text{ and } p \to +\infty.$$

In the following proposition, we see how the upper bounds of the variance have changed under Σ in $G(\mathcal{E}(A, L), \psi)$.

Proposition 2.5. *Under the alternative, for all $\Sigma \in G(\mathcal{E}(A, L), \psi)$, we have :*

$$\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n^{\mathcal{E}}) = \sum_{j=1}^{T} w_j^* \sigma_j^2 \ge b(\psi) \quad \text{and} \quad Var_\Sigma(\widehat{\mathcal{A}}_n^{\mathcal{E}}) = \frac{R_1}{n(n-1)(p-T)^4} + \frac{R_2}{n(p-T)^2}$$

where, for all $A > 0$, and as $\psi \longrightarrow 0$:

$$R_1 \le (p-T)^2 \cdot \{1 + o(1) + \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n^{\mathcal{E}}) \cdot O(\sqrt{T}) + \mathbb{E}_\Sigma^2(\widehat{\mathcal{A}}_n^{\mathcal{E}}) \cdot O(T^2)\} \tag{2.17}$$

$$R_2 \le (p-T) \cdot \{\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n^{\mathcal{E}}) \cdot o(1) + \mathbb{E}_\Sigma^{3/2}(\widehat{\mathcal{A}}_n^{\mathcal{E}}) \cdot O(T^{1/4}) + \mathbb{E}_\Sigma^2(\widehat{\mathcal{A}}_n^{\mathcal{E}}) \cdot O(T)\} \tag{2.18}$$

Moreover, if $n(p-T)b(\psi) \asymp 1$, we show that $n(p-T)(\widehat{\mathcal{A}}_n^{\mathcal{E}} - \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n^{\mathcal{E}})) \to \mathcal{N}(0,1)$, for all $\Sigma \in \mathcal{E}(A, L)$, such that $\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n^{\mathcal{E}}) = O(b(\psi))$.

Now we define the test procedure as follows,

$$\Delta^* = \Delta^*(t) = \mathbb{1}(\widehat{\mathcal{A}}_n^{\mathcal{E}} > t).$$

We describe next the separation rate. We stress the fact that Lemma 2.8 shows that the optimal sequence $\{\sigma_j^*\}_j$ in (2.16) provides a Toeplitz positive definite covariance matrix. The sharp results are obtained under the additional assumption that $\psi \asymp \widetilde{\psi}$ and the lower bounds require that n tends also to infinity.

Theorem 2.6. *Suppose that asymptotically $\psi \to 0$ and $p/T \asymp p/\ln(1/\psi) \to \infty$.*

1. Separation rate.

Lower bound: *if* $n^2p^2b^2(\psi) = n^2p^2 \cdot A\psi^4/(2\ln(1/\psi)) \longrightarrow 0$ *then*

$$\gamma = \inf_{\Delta} \gamma(\Delta, G(\psi)) \longrightarrow 1,$$

where the infimum is taken over all test statistics Δ.

Upper bound: *the test procedure Δ^* defined previously with $t > 0$ has the following properties:*

Type I error probability: if $np \cdot t \to +\infty$ then $\eta(\Delta^*) \to 0$.

Type II error probability: if $n^2p^2\, b^2(\psi) = n^2p^2 \cdot A\psi^4/(2\ln(1/\psi)) \longrightarrow +\infty$ then, uniformly over t such that $t \le c \cdot A^{\frac{1}{2}}\psi^2/(2\ln(1/\psi))^{\frac{1}{2}}$, for some constant c; $0 < c < 1$,

$$\beta(\Delta^*, G(\psi)) \longrightarrow 0.$$

2. Sharp asymptotic bounds.

Lower bound: *suppose that $n \to +\infty$ and that*

$$n^2p^2\, b^2(\psi) \asymp 1, \tag{2.19}$$

then we get $\inf\limits_{\Delta:\eta(\Delta)\le w} \beta(\Delta, G(\psi)) \ge \Phi(z_{1-w} - npb(\psi)) + o(1)$, where the infimum is taken over all test statistics Δ with type I error probability less than or equal to w for $w \in (0,1)$. Moreover,

$$\gamma = \inf_{\Delta} \gamma(\Delta, \psi)) \ge 2\Phi(-np\,\frac{b(\psi)}{2}) + o(1).$$

Upper bound: *we have*

Type I error probability : $\eta(\Delta^*) = 1 - \Phi(npt) + o(1)$.

Type II error probability : under the condition (2.19), we get that, uniformly over t,

$$\beta(\Delta^*, G(\psi)) \le \Phi(np \cdot (t - b(\psi))) + o(1).$$

In particular, the test procedure $\Delta^*(b(\psi)/2)$, is such that $\gamma(\Delta^*(b(\psi)/2), G(\psi)) = 2\Phi(-np\,\frac{b(\psi)}{2}) + o(1)$. We get the sharp minimax separation rate : $\widetilde{\psi} = \left(\frac{2\ln(n^2p^2)}{An^2p^2}\right)^{1/4}$. Remark that, in this case the condition $T/p \to 0$ implies that $n = o(e^p)$, which is considerably less restrictive than the condition $n = o(p^{2\alpha - \frac{1}{2}})$ of the previous case and allows for exponentially large n, e.g. $n = e^{p/2}$.

2.4 Numerical implementation and extensions

In this section we implement the test procedure χ in (2.10) with empirically chosen threshold $t > 0$ and study its numerical performance over two families of covariance matrices. We estimate the type I and type II errors by Monte Carlo sampling with 1000 repetitions.

First, we choose $\Sigma = \Sigma(M) = [\sigma_j]_j$; $\sigma_j = j^{-2}/M$ under the alternative hypothesis, for various values of $M \in \{2, 2.5, 3, 4, 6, 8, 16, 30, 60, 80\}$. We implement the test statistic $\widehat{\mathcal{A}}_n^T$ defined in (2.5) and (2.7), for parameters $\alpha = 1, L = 1$ and $\psi = \psi(M) = \left(\sum_{j=1}^{p-1} j^{-4}\right)^{\frac{1}{2}}/M$. Our choice of the values for M provides positive definite matrices. We denote by $A(M)$ the random variable $n(p-T)\widehat{\mathcal{A}}_n^T$ when $\Sigma = \Sigma(M)$, and by $A(0)$ when $\Sigma = I$. Note that large values of M give $\Sigma(M)$ with small off-diagonal entries, which is very close to the identity matrix.

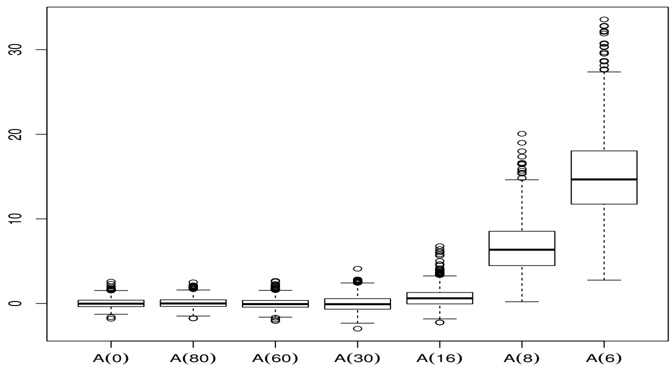

Figure 2.1 – Distributions of $A(M) = n(p-T)\widehat{\mathcal{A}}_n^T$ for $I = \Sigma(0)$ and $\Sigma = \Sigma(M)$, when $p = 60$ and $n = 40$.

Figure 2.1, shows that $n(p-T)\widehat{\mathcal{A}}_n^T$ is distributed as a standard normal random variable, when $\Sigma = I$ and $\Sigma(M)$ close enough to the identity. And as a non-centered normal distribution when $\Sigma(M)$ is far from the identity matrix.

To evaluate the performance of our test procedure we compute it's power. For each value of n and p, we estimate the 95th percentile t of the distribution of $n(p-T)\widehat{\mathcal{A}}_n^T$ under the null hypothesis $\Sigma = I$. We use t previously defined to estimate the type II error probability, and then plot the associated power. In Figure 2.2, we plot the power function of our test procedure χ-test as function of $\psi(M)$, for a fixed value of n and different values of p.

The vertical lines in Figure 2.2 represent the different $\tilde{\psi}(n,p)$ associated to different values of p and $n = 10$. We remark that, on the one hand the power grows with $\psi(M)$ for all $p \in \{10, 30, 50, 70\}$. On the other hand the power is an increasing function of p for a fixed covariance matrix $\Sigma(M)$.

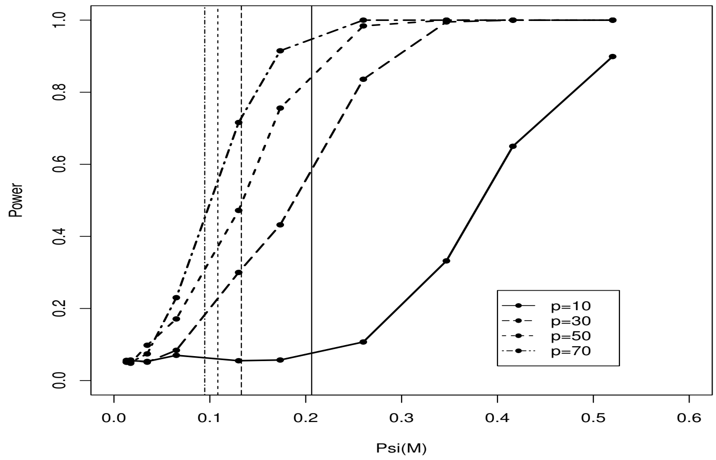

Figure 2.2 – Power curves of the χ-test as function of $\psi(M)$ for $n = 10$ and $p \in \{10, 30, 50, 70\}$

We also compare our test procedure with the one defined in [23]. Recall that the test statistic defined by [23] is given by:

$$\widehat{T}_n^{CM} = \frac{2}{n(n-1)} \sum\sum_{1 \le k < l \le n} \Big((X_k^\top X_l)^2 - X_k^\top X_k - X_l^\top X_l + p \Big).$$

Note that for matrices $\Sigma \in \mathcal{T}(1,1)$, we have $(1/p)\|\Sigma - I\|_F^2 \sim \sum_{j=1}^{p-1} \sigma_j^2$, thus we implement $\widehat{T}_n^{CM}/p$ as CM-test statistic. To have fair comparison, we estimate the 95th percentile under the null hypothesis for both tests. Figures 2.3, shows that when n is bigger than or equal to p the powers of the χ-test and the CM-test take close values. While when n is smaller then p, the gap between the power values of the two tests is large, and the χ-test is more powerful than the CM-test.

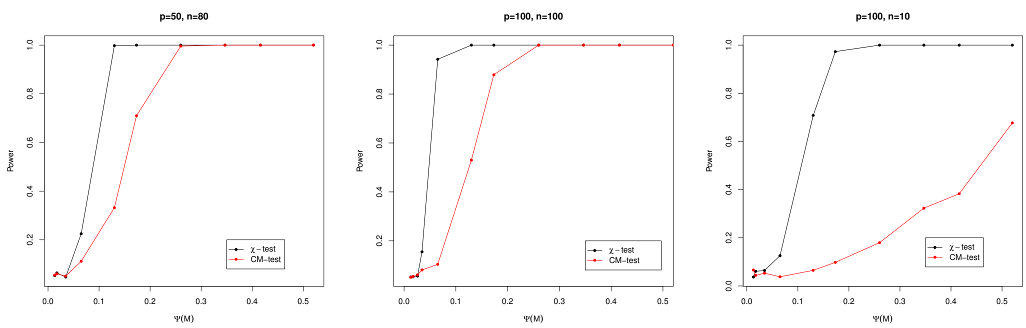

Figure 2.3 – Power curves of the χ-test and the CM-test as functions of $\psi(M)$, when the alternative consists of matrices whose elements decrease polynomially when moving away from the main diagonal

Second, we consider tridiagonal matrices under the alternative. We define $\Sigma = \Sigma(\rho) = [\sigma_j]_j$; $\sigma_j = \rho \cdot \mathbb{1}\{j = 1\}$, for $\rho \in (0,1)$. In this case the parameter ψ is $\psi(\rho) = \rho$, for a grid of 10 points ρ belonging to the interval $(0, 0.35]$ and as previously we take $\alpha = 1$ and $L = 1$.

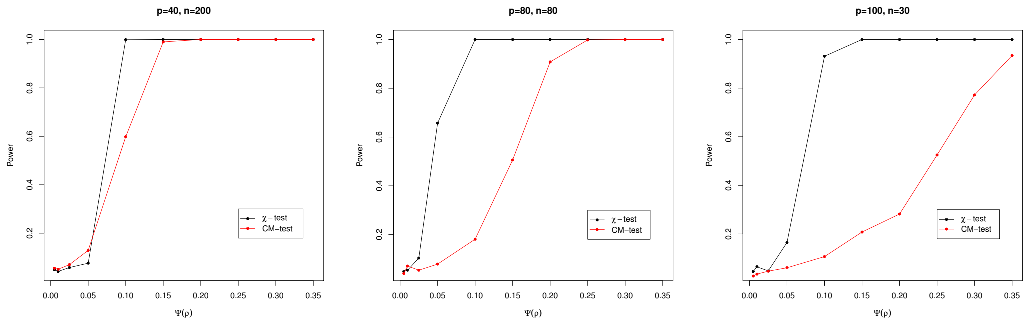

Figure 2.4 – Power curves of the χ-test and the CM-test as functions of $\psi(\rho)$, when the alternative consists of tridiagonal matrices

Figure 2.4 shows that, the χ-test performs better than the U-test, in the three cases : p smaller than n, p equal n and p larger than n. Moreover, we see that the power curves of the χ-test and the CM-test are closer, when the ratio p/n is smaller. We expect even better results in this particular example if we use a larger value of α, or the procedure defined by (2.15) and (2.16). The question arises of a test statistic free of parameters α, respectively A, which is beyond the scope of this paper.

2.5 Proofs

Proof of Theorems 2.3 and 2.4. Recall the assumptions $n, p \to +\infty$, $\psi \to 0$ and $T/p \asymp 1/(p\psi^{1/\alpha}) \to 0$.

Lower bounds : In order to show the lower bound, we first reduce the set of parameters to a convenient parametric family. Let $\Sigma^* = T_p(\{\sigma_k^*\}_{k\geq 1})$ be the Toeplitz matrix such that,

$$\sigma_k^* = \sqrt{\lambda}\left(1 - (\frac{k}{T})^{2\alpha}\right)_+^{\frac{1}{2}} \qquad \text{for } 1 \leq k \leq p-1, \tag{2.20}$$

with λ and T are given by (2.7).

Let us define G^* a subset of $G(\mathcal{T}(\alpha, L), \psi)$ as follows

$$G^* = \{\Sigma_U^* : \Sigma_U^* = T_p(\{u_k \sigma_k\}_{k\geq 1}) \ , \ U \in \mathcal{U}\},$$

where

$$\mathcal{U} = \{U = T_p(\{u_k\}_{k\geq 1}) - I_p \ \text{ and } \ u_k = \pm 1 \cdot I(k \leq T-1), \ \text{for } 1 \leq k \leq T-1\}.$$

The cardinality of $\mathcal{U}$ is 2^{T-1}.

From Proposition 3 in [18], we can see that if $\alpha > 1/2$, for all $U \in \mathcal{U}$, the matrix Σ_U^* is positive definite, for $\psi > 0$ small enough. In contrast with [18], we change the signs randomly on each diagonal of the upper triangle of Σ^* and not of all its elements. That allows us to stay into the model of Toeplitz covariance matrices and will actually change the rates of these lower bounds.

Assume that $X_1, \dots, X_n \sim N(0, I)$ under the null hypothesis and denote by P_I the likelihood of these random variables. Moreover assume that $X_1, \dots, X_n \sim N(0, \Sigma_U^*)$ under the alternative, and we denote P_U the associated likelihood. In addition let

$$P_\pi = \frac{1}{2^{T-1}} \sum_{U \in \mathcal{U}} P_U$$

be the average likelihood over G^*.

The problem can be reduced to the test $H_0 : X_1, ..., X_n \sim P_I$ against the averaged distribution $H_1 : X_1, ..., X_n \sim P_\pi$, in the sense that

$$\begin{aligned} &\inf_{\chi : \eta(\chi) \le w} \beta(\chi, G(\mathcal{T}(\alpha, L), \psi)) \\ = &\inf_{\chi : \eta(\chi) \le w} \sup_{\Sigma \in G(\mathcal{T}(\alpha, L), \psi)} \mathbb{E}_\Sigma(1 - \chi) \ge \inf_{\chi : \eta(\chi) \le w} \sup_{\Sigma \in G^*} \mathbb{E}_\Sigma(1 - \chi) \\ \ge &\inf_{\chi : \eta(\chi) \le w} \frac{1}{2^{T-1}} \mathbb{E}_\Sigma(1 - \chi) = \inf_{\chi : \eta(\chi) \le w} \mathbb{E}_\pi(1 - \chi) := \inf_{\chi : \eta(\chi) \le w} \beta(\chi, \{P_\pi\}) \end{aligned}$$

and that

$$\inf_\chi \gamma(\chi, G(\mathcal{T}(\alpha, L), \psi)) \quad \ge \quad \inf_\chi \gamma(\chi, \{P_\pi\}) + o(1)$$

where, with an abuse of notation,

$$\beta(\chi, \{P_\pi\}) = \mathbb{E}_\pi(1 - \chi) \quad \text{and} \quad \gamma(\chi, \{P_\pi\}) = \mathbb{E}_I(\chi) + \mathbb{E}_\pi(1 - \chi).$$

It is therefore sufficient to show that, when $u_n \asymp 1$,

$$\inf_{\chi : \eta(\chi) \le w} \beta(\chi, \{P_\pi\}) \ge \Phi(z_{1-w} - npb(\psi))) + o(1) \tag{2.21}$$

and that

$$\inf_\chi \gamma(\chi, \{P_\pi\}) \ge 2\Phi(-np \frac{b(\psi)}{2}) + o(1), \tag{2.22}$$

while, for $u_n = o(1)$, we need that

$$\gamma(\chi, \{P_\pi\}) \to 1. \tag{2.23}$$

Lemme 2.7. *Assume that $\psi \to 0$ such that $p\psi^{1/\alpha} \to \infty$ and let f_π be the probability density associated to the likelihood P_π previously defined. Then*

$$L_{n,p} := \log \frac{f_\pi}{f_I}(X_1, ..., X_n) = u_n Z_n - \frac{u_n^2}{2} + o_P(1), \text{ in } P_I \text{ probability,} \tag{2.24}$$

where Z_n is asymptotically distributed as a standard Gaussian distribution and $u_n = npb(\psi)$ is such that either $u_n \to 0$ or $u_n \asymp 1$.Moreover, $L_{n,p}$ is uniformly integrable.

In order to obtain (2.21) and (2.22), we apply results in Section 4.3.1 of [58] giving the sufficient condition is (2.24).

It is known that $\gamma(\chi, \{P_\pi\}) = 1 - \frac{1}{2}\|P_I - P_\pi\|_1$ and we bound the L_1 norm by the Kullback-Leibler divergence

$$\frac{1}{2}\|P_I - P_\pi\|_1^2 \le K(P_I, P_\pi).$$

Therefore to show (2.23), we apply Lemma 2.7 to see that the log likelihood $\log f_\pi/f_I(X_1, ..., X_n)$ is an uniformly integrable sequence. This implies that

$$K(P_I, P_\pi) = -\mathbb{E}_I(\log f_\pi/f_I(X_1, ..., X_n)) \to 0.$$

□

Upper bounds : By the Proposition 2.1, we have that under the null hypothesis $n(p-T)\widehat{\mathcal{A}}_n \to \mathcal{N}(0,1)$. Then we can deduce that the Type I error probability of χ^* has the following form :

$$\eta(\chi^*) = \mathbb{P}(\widehat{\mathcal{A}}_n > t) = 1 - \Phi(npt) + o(1).$$

For the Type II error probability of χ^*, we shall distinguish two cases, when $n^2p^2b^2(\psi)$ tends to infinity or is bounded by some finite constant. First, assume that $\psi/\widetilde{\psi} \to +\infty$ or, equivalently, that $n^2p^2b^2(\psi) \to +\infty$. Then by the Markov inequality,

$$\mathbb{P}_\Sigma(\widehat{\mathcal{A}}_n \le t) \quad \le \quad \mathbb{P}_\Sigma(|\widehat{\mathcal{A}}_n - \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n)| \ge \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) - t) \le \frac{\mathrm{Var}_\Sigma(\widehat{\mathcal{A}}_n)}{(\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) - t)^2}$$

for all $\Sigma \in G(\mathcal{T}(\alpha, L), \psi)$ and $t \le c \cdot b(\psi)$ such that $0 < c < 1$. Recall that under the alternative, we have $\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) \ge b(\psi)$ which gives:

$$\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) - t \ge (1-c)\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) \ge (1-c)b(\psi). \tag{2.25}$$

Therefore from the first part of the inequality (2.25) and the variance expression of $\widehat{\mathcal{A}}_n$ under H_1, given in Proposition 1, we have:

$$\begin{aligned}\mathbb{P}_\Sigma(\widehat{\mathcal{A}}_n \le t) \quad &\le \quad \frac{R_1}{n(n-1)(p-T)^4(1-c)^2\mathbb{E}^2_\Sigma(\widehat{\mathcal{A}}_n)} + \frac{R_2}{n(p-T)^2(1-c)^2\mathbb{E}^2_\Sigma(\widehat{\mathcal{A}}_n)} \\ &:= \quad U_1 + U_2.\end{aligned}$$

Let us bound from above U_1, using (2.8) and the second part of the inequality (2.25):

$$U_1 \quad \le \quad \frac{1 + o(1)}{n(n-1)(p-T)^2(1-c)^2b^2(\psi)} + \frac{O(\sqrt{T}) + O(T^{3/2-2\alpha})}{n(n-1)(p-T)^2b(\psi)} + \frac{O(T^2)}{n(n-1)(p-T)^2}.$$

We have $T^{(3/2-2\alpha)}b(\psi) \asymp T^2b^2(\psi) \asymp \psi^{4-\frac{1}{\alpha}} = o(1),$ for all $\alpha > 1/4$, which proves that :

$$U_1 \le \frac{1 + o(1)}{n(n-1)(p-T)(1-c)^2b^2(\psi)} = o(1).$$

Indeed, $n^2(p-T)^2b^2(\psi) \to +\infty$, since $n^2p^2b^2(\psi) \to +\infty$ and $T/p \to 0$.
We can check using (2.9) that the term U_2 tends to zero as well :

$$\begin{aligned} U_2 &\leq \frac{o(1)}{n(p-T)b(\psi)} + \frac{O(T^{1/4}) + O(T^{3/4-\alpha})}{n(p-T)b^{1/2}(\psi)} + \frac{O(T)}{n(p-T)} \\ &= o(1) \text{ for all } \alpha > 1/4, \text{ as soon as } n^2p^2b^2(\psi) \longrightarrow +\infty. \end{aligned}$$

Finally, when ψ is of the same order of the separation rate, i.e. $n^2p^2b^2(\psi) \asymp 1$, we may have either $\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n)/b(\psi)$ tends to infinity, or $\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) = O(b(\psi))$. In the first case it is easy to see that $U_1 + U_2 \longrightarrow 0$. In the latter the Proposition 2.2 gives the asymptotic normality of $n(p-T)(\widehat{\mathcal{A}}_n - \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n))$. Thereby,

$$\begin{aligned} \sup_{\Sigma \in G(\mathcal{T}(\alpha,L),\psi)} \mathbb{P}_\Sigma(\widehat{\mathcal{A}}_n \leq t) &\leq \sup_{\Sigma \in G(\mathcal{T}(\alpha,L),\psi)} \Phi(np \cdot (t - \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n))) + o(1) \\ &\leq \Phi(np \cdot (t - \inf_{\Sigma \in G(\mathcal{T}(\alpha,L),\psi)} \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n))) + o(1) \\ &= \Phi(np \cdot (t - b(\psi))) + o(1). \end{aligned}$$

2.6 Supplementary material

2.6.1 Additional proofs for the results in Section 2.2

Proof of Lemma 2.7. We need to study the log-likelihood ratio:

$$L_{n,p} := \log \frac{f_\pi}{f_I}(X_1, ..., X_n) = \log \mathbb{E}_U \exp\left(-\frac{1}{2} \sum_{k=1}^n X_k^\top ((\Sigma_U^*)^{-1} - I) X_k - \frac{n}{2} \log\det(\Sigma_U^*) \right),$$

where U is seen as a randomly chosen matrix with uniform distribution over the set $\mathcal{U}$.

Moreover, let us denote $\Delta_U = \Sigma_U^* - I$ which is a symmetric matrix with null diagonal. Recall that for all $U \in \mathcal{U}$, $tr(\Delta_U) = 0$ and that $\|\Delta_U\| = O(\psi^{1-1/(2\alpha)})$. Remember also that $\sigma_k^* = 0$ for all $|k| \geq T$.

The matrix Taylor expansion gives

$$\begin{aligned} (\Sigma_U^*)^{-1} - I &= -\Delta_U + \Delta_U^2 + O(1) \cdot \Delta_U^3, \\ \log\det(\Sigma_U^*) &= -\frac{1}{2} tr(\Delta_U^2) + O(1) \cdot tr(\Delta_U^3). \end{aligned}$$

On the one hand, $tr(\Delta_U^2) = \sum_{1 \leq i \neq j \leq p} (\sigma_{|i-j|}^*)^2$, does not depend on U. Moreover,

$$tr(\Delta_U^3) \leq \|\Delta_U\| \cdot \|\Delta_U\|_F^2 = O(p\psi^{3-\frac{1}{2\alpha}}) = O(np\psi^{2+\frac{1}{2\alpha}} \cdot \frac{\psi^{1-\frac{1}{\alpha}}}{n}) = o(1) \quad \text{for } \alpha > 1. \tag{2.26}$$

Thus we get

$$\frac{n}{2} \log\det(\Sigma_U^*) = \frac{n}{2} \sum_{1 \leq i \neq j \leq p} (\sigma_{|i-j|}^*)^2 + o(1). \tag{2.27}$$

On the other hand, we see that

$$X_k^\top \Delta_U X_k = \sum_{1\le i,j\le p} X_{k,i} u_{|i-j|}\sigma_{|i-j|} X_{k,j} = 2\sum_{1\le r<T} u_r\sigma_r^* \sum_{i=1+r}^{p} X_{k,i}X_{k,i-r} \tag{2.28}$$

and that

$$\begin{aligned} X_k^\top \Delta_U^2 X_k &= \sum_{1\le i,j\le p} X_{k,i}X_{k,j} \sum_{\substack{h=1\\ h\notin\{i,j\}}}^{p} u_{|i-h|}u_{|j-h|}\sigma^*_{|i-h|}\sigma^*_{|j-h|} \\ &= \sum_{i=1}^{p} X_{k,i}^2 \sum_{\substack{h=1\\ h\neq i}}^{p} (\sigma^*_{|i-h|})^2 + \sum_{1\le i\neq j\le p} X_{k,i}X_{k,j} \sum_{\substack{h=1\\ h\notin\{i,j\}}}^{p} u_{|i-h|}u_{|j-h|}\sigma^*_{|i-h|}\sigma^*_{|j-h|} \\ &:= S_1 + S_2. \end{aligned}$$

In the term S_2, we change the variables i and j into $l = i-h$ and $m = j-h$ and due to the constraints we have $|l|,|m| \in \{1,\dots,T-1\}$ and $l \neq m$, while h varies in the set $\{1\vee(1-l)\vee(1-m)\,,\,p\wedge(p-l)\wedge(p-m)\}$ for each fixed pair (l,m). Therefore,

$$S_2 = \sum_{\substack{l\neq m\\ 1\le|l|,|m|<T}} \sum_{h=1\vee(1-l)\vee(1-m)}^{p\wedge(p-l)\wedge(p-m)} u_{|l|}u_{|m|}\sigma^*_{|l|}\sigma^*_{|m|}X_{k,l+h}X_{k,m+h}.$$

We split the previous sums over $l\neq m$ such that $\operatorname{sign}(l\cdot m)>0$ and get

$$S_{2,1} := \sum_{1\le l\neq m<T} u_l u_m \sigma_l^*\sigma_m^* \Big(\sum_{h=1}^{(p-l)\wedge(p-m)} X_{k,h+l}X_{k,h+m} + \sum_{h=(1+l)\vee(1+m)}^{p} X_{k,h-l}X_{k,h-m} \Big)$$

respectively, over l,m of opposite signs: $\operatorname{sign}(l\cdot m)<0$ and get

$$\begin{aligned} S_{2,2} &= 2\sum_{1\le l,m<T} \sum_{h=1+m}^{p-l} u_l u_m \sigma_l^*\sigma_m^* X_{k,h+l}X_{k,h-m} \\ &= 2\sum_{l=1}^{T-1}\sum_{h=1+l}^{p-l} \sigma_l^{*2} X_{k,h+l}X_{k,h-l} + 2\sum_{1\le l\neq m<T}\sum_{h=1+m}^{p-l} u_l u_m \sigma_l^*\sigma_m^* X_{k,h+l}X_{k,h-m}. \end{aligned}$$

In conclusion, we can group terms differently and write

$$\begin{aligned} X_k^\top \Delta_U^2 X_k &= \sum_{1\le l\neq m<T} u_l u_m \sigma_l^*\sigma_m^* \left(\sum_{h=1}^{(p-l)\wedge(p-m)} X_{k,h+l}X_{k,h+m} + \sum_{h=(1+l)\vee(1+m)}^{p} X_{k,h-l}X_{k,h-m} \right. \\ &\quad \left. + 2\sum_{h=1+m}^{p-l} X_{k,h+l}X_{k,h-m} \right) + \sum_{i=1}^{p} X_{k,i}^2 \sum_{\substack{h=1\\ h\neq i}}^{p} (\sigma^*_{|i-h|})^2 + 2\sum_{l=1}^{T-1}\sum_{h=1+l}^{p-l} \sigma_l^{*2} X_{k,h+l}X_{k,h-l} \\ &= \sum_{1\le l\neq m<T} u_l u_m \sigma_l^*\sigma_m^* V_p(l,m,k) + \sum_{i=1}^{p} X_{k,i}^2 \sum_{\substack{h=1\\ h\neq i}}^{p} (\sigma^*_{|i-h|})^2 + 2\sum_{l=1}^{T-1}\sum_{h=1+l}^{p-l} \sigma_l^{*2} X_{k,h+l}X_{k,h-l}, \end{aligned} \tag{2.29}$$

where

$$V_p(l,m,k) := \sum_{h=1}^{(p-l)\wedge(p-m)} X_{k,h+l}X_{k,h+m} + \sum_{h=(1+l)\vee(1+m)}^{p} X_{k,h-l}X_{k,h-m} + 2\sum_{h=1+m}^{p-l} X_{k,h+l}X_{k,h-m}.$$

Now, let us see that:

$$\mathbb{E}_I(X_k^\top \Delta_U^3 X_k) = \mathbb{E}_I(tr(X_k^\top \Delta_U^3 X_k)) = \mathbb{E}_I(tr(X_k X_k^\top \Delta_U^3)) = tr(\Delta_U^3 \mathbb{E}_I(X_k X_k^\top)) = tr(\Delta_U^3)$$

and recall (2.26) to get

$$\mathbb{E}_I(\sum_{k=1}^n X_k^\top \Delta_U^3 X_k) = O(np\psi^{3-\frac{1}{2\alpha}}) = o(1).$$

Moreover, we have $\mathbb{E}_I(X_k^\top \Delta_U^3 X_k)^2 = tr^2(\Delta_U^3) + 2tr(\Delta_U^6)$ by Proposition A.1 in [29], which implies that

$$\mathrm{Var}_I(\sum_{k=1}^n X_k^\top \Delta_U^3 X_k) = 2ntr(\Delta_U^6) \le 2n\|\Delta_U\|^4 \|\Delta_U\|_F^2 = O(np\psi^{6-4/(2\alpha)}) = o(1).$$

Then, using Chebyshev's inequality we obtain,

$$\sum_{k=1}^n X_k^\top \Delta_U^3 X_k = o_P(1). \tag{2.30}$$

Thus we replace (2.27) to (2.30) in $L_{n,p}$ and get

$$\begin{aligned} L_{n,p} &= \log \mathbb{E}_U \exp\Bigg(\sum_{1\le r<T} u_r \sigma_r^* \sum_{i=1+r}^p \sum_{k=1}^n X_{k,i}X_{k,i-r} - \frac{1}{2} \sum_{1\le l\ne m<T} u_l u_m \sigma_l^* \sigma_m^* \sum_{k=1}^n V_p(l,m,k) \Bigg) \\ &- \frac{1}{2} \sum_{i=1}^p \sum_{k=1}^n X_{k,i}^2 \sum_{\substack{h=1\\h\ne i}}^p (\sigma_{|i-h|}^*)^2 - \sum_{1\le l\le T-1} \sigma_l^{*2} \sum_{h=1+l}^{p-l} \sum_{k=1}^n X_{k,h+l}X_{k,h-l} + \frac{n}{4} \sum_{1\le i\ne j\le p} (\sigma_{|i-j|}^*)^2 + o_P(1). \end{aligned}$$

Denote by $W_{l,m} := \sum_{k=1}^n X_{k,l}X_{k,m}$. Now, we evaluate the expected value with respect to the i.i.d. Rademacher variables u_r, $u_l u_m$ for all $1 \le r < T$ and $1 \le l \ne m < T$ to get

$$\begin{aligned} L_{n,p} &= \log\Big(\prod_{1\le r\le T-1} \cosh(\sigma_r^* \sum_{i=r+1}^p W_{i,i-r})\Big) + \log\Big(\prod_{1\le l\ne m<T} \cosh\Big(\frac{1}{2}\sigma_l^*\sigma_m^* \sum_{k=1}^n V_p(l,m,k)\Big)\Big) \\ &- \frac{1}{2}\sum_{i=1}^p W_{i,i} \sum_{j:j\ne i} (\sigma_{|i-j|}^*)^2 - \sum_{1\le l\le T-1} \sigma_l^{*2} \sum_{h=1+l}^{p-l} W_{h+l,h-l} + \frac{n}{4} \sum_{1\le i\ne j\le p} (\sigma_{|i-j|}^*)^2 + o_P(1). \end{aligned}$$

We get that

$$\begin{aligned} L_{n,p} &= \sum_{1\le r\le T-1} \log\cosh(\sigma_r^* \sum_{i=r+1}^p W_{i,i-r}) + \sum_{1\le l\ne m<T} \log\cosh\Big(\frac{1}{2}\sigma_l^*\sigma_m^* \sum_{k=1}^n V_p(l,m,k)\Big) \\ &- \frac{1}{2}\sum_{i=1}^p W_{i,i} \sum_{j:j\ne i} (\sigma_{|i-j|}^*)^2 - \sum_{1\le l\le T-1} \sigma_l^{*2} \sum_{h=1+l}^{p-l} W_{h+l,h-l} + \frac{n}{4} \sum_{1\le i\ne j\le p} (\sigma_{|i-j|}^*)^2 + o_P(1). \end{aligned}$$

Note that

$$\sum_{k=1}^n V_p(l,m,k) = \sum_{h=1}^{(p-l)\wedge(p-m)} W_{h+l,h+m} + \sum_{h=(1+l)\vee(1+m)}^p W_{h-l,h-m} + 2\sum_{h=1+m}^{p-l} W_{h+l,h-m}.$$

We use several times the Taylor expansion $\log\cosh(u) = \frac{u^2}{2} - \frac{u^4}{12}(1+o(1))$ for $|u| \to 0$. On the one hand, by Chebyshev's inequality, $|\sigma_r^* \sum_{i=r+1}^{p} W_{i,i-r}| = O_P(\sqrt{\lambda np}) = O_P(\psi^{1/4\alpha}\sqrt{npb(\psi)}) = o_P(1)$, as soon as $\psi \to 0$. Then,

$$\log\cosh(\sigma_r^* \sum_{i=r+1}^{p} W_{i,i-r}) = \frac{1}{2}\sigma_r^{*2}(\sum_{i=r+1}^{p} W_{i,i-r})^2 - \frac{(1+o_P(1))}{12}\cdot\sigma_r^{*4}(\sum_{i=r+1}^{p} W_{i,i-r})^4.$$

On the other hand,

$$\begin{aligned}
& \Big|\frac{\sigma_l^*\sigma_m^*}{2}\Big(\sum_{h=1}^{(p-l)\wedge(p-m)} W_{h+l,h+m} + \sum_{h=(1+l)\vee(1+m)}^{p} W_{h-l,h-m} + 2\sum_{h=1+m}^{p-l} W_{h+l,h-m}\Big)\Big| \\
\leq\ & \frac{\lambda}{2}\cdot\Big|\sum_{h=1}^{(p-l)\wedge(p-m)} W_{h+l,h+m}\Big| + \frac{\lambda}{2}\cdot\Big|\sum_{h=(1+l)\vee(1+m)}^{p} W_{h-l,h-m}\Big| + \lambda\cdot\Big|\sum_{l=1+m}^{p-l} W_{h+l,h-m}\Big| \\
\leq\ & O_p(\lambda\sqrt{np}) = O_p(\psi^{1/2\alpha}\sqrt{npb(\psi)}) = o_P(1).
\end{aligned}$$

Thus we have to study now

$$\begin{aligned}
L_{n,p} &= \log\frac{f_\pi}{f_I}(X_1,\dots,X_n) \\
&= \sum_{1\leq r<T}\Big\{\frac{1}{2}\cdot\sigma_r^{*2}(\sum_{i=r+1}^{p} W_{i,i-r})^2 - \frac{(1+o_P(1))}{12}\cdot\sigma_r^{*4}(\sum_{i=r+1}^{p} W_{i,i-r})^4\Big\} \\
&+ \frac{1}{4}\sum_{1\leq l\neq m<T}\sigma_l^{*2}\sigma_m^{*2}\Big(\sum_{k=1}^{n} V_p(l,m,k)\Big)^2(1+o_P(1)) \qquad (2.31)\\
&- \frac{1}{2}\sum_{i=1}^{p} W_{i,i}\sum_{j:j\neq i}(\sigma_{|i-j|}^*)^2 - \sum_{1\leq l\leq T-1}\sigma_l^{*2}\sum_{h=1+l}^{p-l} W_{h+l,h-l} + \frac{n}{4}\sum_{1\leq i\neq j\leq p}(\sigma_{|i-j|}^*)^2 + o_P(1).
\end{aligned}$$

Let us treat each term of (2.31) separately. We first decompose $(\sum_{i=r+1}^{p} W_{i,i-r})^2$ as follows,

$$\begin{aligned}
A &:= (\sum_{i=r+1}^{p} W_{i,i-r})^2 \\
&= \sum_{1+r\leq i_1,i_2\leq p}\Big(\sum_{k=1}^{n}\sum_{\substack{l=1\\l\neq k}}^{n} X_{k,i_1}X_{k,i_1-r}X_{l,i_2}X_{l,i_2-r} + \sum_{k=1}^{n} X_{k,i_1}X_{k,i_1-r}X_{k,i_2}X_{k,i_2-r}\Big) \\
&= \sum_{1+r\leq i_1,i_2\leq p}\ \sum_{k=1}^{n}\sum_{\substack{l=1\\l\neq k}}^{n} X_{k,i_1}X_{k,i_1-r}X_{l,i_2}X_{l,i_2-r} \\
&+ \sum_{1+r\leq i_1\neq i_2\leq p}\ \sum_{k=1}^{n} X_{k,i_1}X_{k,i_1-r}X_{k,i_2}X_{k,i_2-r} + \sum_{1+r\leq i\leq p}\ \sum_{k=1}^{n} X_{k,i}^2X_{k,i-r}^2 \\
&:= A_1 + A_2 + A_3
\end{aligned}$$

The term A_3 will be taken into account as it is later on.

The dominant term giving the asymptotic distribution is :

$$
\begin{aligned}
\frac{1}{2}\sum_{1\le r<T}\sigma_r^{*2}\cdot A_1 &= \frac{1}{2}\sum_{1\le r<T}\sigma_r^{*2}\sum_{1+r\le i_1,i_2\le p}\sum_{k=1}^{n}\sum_{\substack{l=1\\ l\neq k}}^{n}X_{k,i_1}X_{k,i_1-r}X_{l,i_2}X_{l,i_2-r}\\
&= \frac{1}{2}\sum_{1\le r<T}\sigma_r^{*2}\sum_{1+T\le i_1,i_2\le p}\sum_{k=1}^{n}\sum_{\substack{l=1\\ l\neq k}}^{n}X_{k,i_1}X_{k,i_1-r}X_{l,i_2}X_{l,i_2-r}\\
&+ \sum_{1\le r<T}\sigma_r^{*2}\sum_{1+r\le i_1\le T}\ \sum_{1+T\le i_2\le p}\sum_{k=1}^{n}\sum_{\substack{l=1\\ l\neq k}}^{n}X_{k,i_1}X_{k,i_1-r}X_{l,i_2}X_{l,i_2-r}\\
&+ \frac{1}{2}\sum_{1\le r<T}\sigma_r^{*2}\sum_{1+r\le i_1,i_2\le T}\sum_{k=1}^{n}\sum_{\substack{l=1\\ l\neq k}}^{n}X_{k,i_1}X_{k,i_1-r}X_{l,i_2}X_{l,i_2-r}\\
&:= A_{1,1}+A_{1,2}+A_{1,3}\,, \quad \text{say.}
\end{aligned}
$$

Recall that $\sigma_r^{*2} = 2w_r^* b(\psi)$ and then $A_{1,1} = n(p-T)\widehat{\mathcal{A}}_n \cdot n(p-T)b(\psi)$. By Proposition 2.1, $n(p-T)\widehat{\mathcal{A}}_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ and thus $A_{1,1}$ can be written $u_n Z_n$ with $Z_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$.

Next, under $\mathbb{P}_I$ all variables in the multiple sums of $A_{1,2}$ are uncorrelated (as well as for $A_{1,3}$). Thus,

$$
\begin{aligned}
\mathrm{Var}_I(A_{1,2}) &= 2\sum_{1\le r<T}\sigma_r^{*4}\sum_{1+r\le i_1\le T}\ \sum_{1+T\le i_2\le p}\sum_{k=1}^{n}\sum_{\substack{l=1\\ l\neq k}}^{n}\mathrm{Var}_I(X_{k,i_1}X_{k,i_1-r}X_{l,i_2}X_{l,i_2-r})\\
&= 2\sum_{1\le r<T}\sigma_r^{*4}(p-T)(T-r)n(n-1) \le n^2pT\sum_{1\le r<T}\sigma_r^{*4} = 2n^2pTb^2(\psi)\\
&= 2\cdot\frac{T}{p}\cdot u_n^2 = o(u_n^2), \qquad \text{as } T/p\to 0.
\end{aligned}
$$

And, similarly,

$$
\begin{aligned}
\mathrm{Var}_I(A_{1,3}) &= \frac{1}{4}\sum_{1\le r<T}\sigma_r^{*4}\sum_{1+r\le i_1,i_2\le T}2\sum_{k=1}^{n}\sum_{\substack{l=1\\ l\neq k}}^{n}\mathrm{Var}_I(X_{k,i_1}X_{k,i_1-r}X_{l,i_2}X_{l,i_2-r})\\
&\le \frac{1}{2}\cdot T^2n(n-1)b^2(\psi) = O\Big(\Big(\frac{T}{p}\Big)^2\cdot u_n^2\Big) = o(u_n^2).
\end{aligned}
$$

Therefore, $A_{1,1}+A_{1,2}+A_{1,3} = u_nZ_n + o_p(u_n)$, where $Z_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$. For the same reason, we have,

$$
\begin{aligned}
\mathrm{Var}_I\Big(\frac{1}{2}\sum_{1\le r<T}\sigma_r^{*2}\cdot A_2\Big) &= \frac{1}{4}\sum_{1\le r<T}\sigma_r^{*4}\sum_{1+r\le i_1\neq i_2\le p}\sum_{k=1}^{n}\mathrm{Var}_I(X_{k,i_1}X_{k,i_1-r}X_{k,i_2}X_{k,i_2-r})\\
&\le \frac{1}{4}\cdot np^2b^2(\psi) = O\Big(\frac{1}{n}\cdot u_n^2\Big) = o(1),
\end{aligned}
$$

as soon as $n\to\infty$ or $u_n\to 0$. We want to show that

$$
B = \frac{1}{12}\sum_{1\le r<T}\sigma_r^{*4}\Big(\sum_{i=r+1}^{p}W_{i,i-r}\Big)^4 = \frac{u_n^2}{2} + o_p(1).
$$

Indeed,

$$\begin{aligned}
\mathbb{E}_I(B) &= \frac{1}{12}\sum_{1\le r<T}\sigma_r^{*4}\cdot\mathbb{E}_I\Big(\sum_{i=r+1}^{p}W_{i,i-r}\Big)^4 = \frac{1}{12}\sum_{1\le r<T}\sigma_r^{*4}\cdot\mathbb{E}_I\Big(\sum_{i=r+1}^{p}\sum_{k=1}^{n}X_{k,i}X_{k,i-r}\Big)^4\\
&= \frac{1}{12}\sum_{1\le r<T}\sigma_r^{*4}\sum_{k=1}^{n}\Big(\sum_{i=r+1}^{p}\mathbb{E}_I(X_{k,i}^4X_{k,i-r}^4)+3\sum_{1+r\le i_1\ne i_2\le p}\mathbb{E}_I(X_{k,i_1}^2X_{k,i_1-r}^2)\mathbb{E}_I(X_{k,i_2}^2X_{k,i_2-r}^2)\Big)\\
&+ \frac{3}{12}\sum_{1\le r<T}\sigma_r^{*4}\sum_{1\le k_1\ne k_2\le n}\ \sum_{1+r\le i_1\ne i_2\le p}\mathbb{E}_I(X_{k_1,i_1}^2X_{k_1,i_1-r}^2)\mathbb{E}_I(X_{k_2,i_2}^2X_{k_2,i_2-r}^2)\\
&= \frac{3}{4}\sum_{1\le r<T}\sigma_r^{*4}\cdot n(p-r)+\frac{1}{4}\sum_{1\le r<T}\sigma_r^{*4}\cdot n(p-r)^2+\frac{1}{4}\sum_{1\le r<T}\sigma_r^{*4}\cdot n^2(p-r)^2 \qquad (2.32)
\end{aligned}$$

Recall that $2b^2(\psi)=\sum_j\sigma_j^{*4}$, thus

$$\mathbb{E}_I(B)=\frac{3}{2}\cdot npb^2(\psi)(1+o(1))+\frac{1}{2}\cdot np^2b^2(\psi)(1+o(1))+\frac{1}{2}\cdot n^2p^2b^2(\psi)(1+o(1))=\frac{u_n^2}{2}(1+o(1)).$$

Moreover,

$$\begin{aligned}
\mathrm{Var}_I(B) &= \frac{1}{12^2}\sum_{1\le r<T}\sigma_r^{*8}\cdot\mathrm{Var}_I\Big(\Big(\sum_{i=r+1}^{p}W_{i,i-r}\Big)^4\Big)\\
&+ \frac{1}{12^2}\sum_{1\le r\ne r'<T}\sigma_r^{*4}\sigma_{r'}^{*4}\mathrm{Cov}_I\Big(\Big(\sum_{i=r+1}^{p}W_{i,i-r}\Big)^4,\Big(\sum_{i'=r'+1}^{p}W_{i',i'-r'}\Big)^4\Big).
\end{aligned}$$

As in the calculation of the expected value of B, we can see that the term of higher order is obtained when we gather the indices into distinct pairs. Thus following the same reasoning we get

$$\mathrm{Var}_I\Big(\sum_{i=r+1}^{p}\sum_{k=1}^{n}X_{k,i}X_{k,i-r}\Big)^4=O(n^4p^4).$$

Through a very technical calculation, and using similar arguments as previously, we can prove that, for $r\ne r'$,

$$\mathrm{Cov}_I\Big(\Big(\sum_{i=r+1}^{p}\sum_{k=1}^{n}X_{k,i}X_{k,i-r}\Big)^4,\Big(\sum_{i'=r'+1}^{p}\sum_{k'=1}^{n}X_{k',i'}X_{k',i'-r'}\Big)^4\Big)=O(n^3p^4).$$

Thus,

$$\mathrm{Var}_I(B)=O(\lambda^4Tn^4p^4)+O(b^4(\psi)n^3p^4)=O(\psi^{\frac{3}{\alpha}}n^4p^4b^4(\psi))+O\Big(\frac{1}{n}\cdot n^4p^4b^4(\psi)\Big)=o(1).$$

By Chebyshev's inequality we deduce that

$$\begin{aligned}
\frac{1}{12}\sum_{r=1}^{T-1}\sigma_r^{*4}\Big(\sum_{i=r+1}^{p}W_{i,i-r}\Big)^4 &= \mathbb{E}_I\Big(\frac{1}{12}\sum_{r=1}^{T-1}\sigma_r^{*4}\Big(\sum_{i=r+1}^{p}W_{i,i-r}\Big)^4\Big)+o_P(1)\\
&= \frac{3(1+o(1))}{12}\cdot 2n^2(p-r)^2b^2(\psi)+o_P(1)=\frac{u_n^2}{2}(1+o_P(1)).
\end{aligned}$$

Also using that $\mathbb{E}_I(\sum_{i=r+1}^{p}W_{i,i-r})^4=O(n^2p^2)$, we get

$$\begin{aligned}
C &:= \sum_{1\le l\ne m<T}\frac{\sigma_l^{*2}\sigma_m^{*2}}{4}\Big(\sum_{h=1}^{(p-l)\wedge(p-m)}W_{h+l,h+m}+\sum_{h=(1+l)\vee(1+m)}^{p}W_{h-l,h-m}+2\sum_{h=1+m}^{p-l}W_{h+l,h-m}\Big)^2\\
&= O_P(\lambda^2T^2np)=o_P(\psi^{(2-\frac{1}{2\alpha})}\cdot u_n)=o_p(1) \quad \text{ for } \alpha>1/4 \text{ and since } \psi\to 0.
\end{aligned}$$

Moreover,

$$F := -\sum_{1\le l\le T-1}\sigma_l^{*2}\sum_{h=1+l}^{p-l}W_{h+l,h-l} = O_P(\sqrt{np}b(\psi)) = o_P(u_n) = o_P(1).$$

Finally, we group the remaining terms of (2.31) as follows,

$$\begin{aligned}
G &:= \frac{1}{2}\sum_{r=1}^{T-1}\sigma_r^{*2}A_3 - \frac{1}{2}\sum_{i=1}^{p}W_{i,i}\sum_{j:j\neq i}(\sigma^*_{|i-j|})^2 + \frac{n}{4}\sum_{1\le i\neq j\le p}(\sigma^*_{|i-j|})^2\\
&= \frac{1}{4}\sum_{1\le i\neq j\le p}(\sigma^*_{|i-j|})^2\sum_{k=1}^{n}X_{k,i}^2X_{k,j}^2 - \frac{1}{2}\sum_{1\le i\neq j\le p}(\sigma^*_{|i-j|})^2\sum_{k=1}^{n}X_{k-i}^2 + \frac{n}{4}\sum_{1\le i\neq j\le p}(\sigma^*_{|i-j|})^2\\
&= \frac{1}{4}\sum_{1\le i\neq j\le p}(\sigma^*_{|i-j|})^2\sum_{k=1}^{n}(X_{k,i}^2-1)(X_{k,j}^2-1) = O_P(\sqrt{np}\cdot b(\psi)) = o_P(u_n) = o_P(1).
\end{aligned}$$

Let us note that throughout the previous proof we also showed that the likelihood ratio $L_{n,p}$ has a variance which tends to 0, for all $n \ge 2$, when $u_n \to 0$. □

Proof of Proposition 2.1 . Under the null hypothesis, $\widehat{\mathcal{A}}_n$ is centered, and

$$\begin{aligned}
\mathrm{Var}_I(\widehat{\mathcal{A}}_n) &= \frac{2}{n(n-1)(p-T)^4}\mathrm{Var}_I\left(\sum_{j=1}^{T}w_j^*\sum_{1+T\le i_1,i_2\le p}X_{1,i_1}X_{1,i_1-j}X_{2,i_2}X_{2,i_2-j}\right)\\
&= \frac{2}{n(n-1)(p-T)^4}\sum_{j=1}^{T}w_j^{*2}\sum_{1+T\le i_1,i_2\le p}\mathbb{E}_I(X_{1,i_1}^2X_{1,i_1-j}^2X_{2,i_2}^2X_{2,i_2-j}^2)\\
&= \frac{2}{n(n-1)(p-T)^2}\sum_{j=1}^{T}w_j^{*2}
\end{aligned}$$

Recall that $\sum_{j=1}^{T}w_j^{*2} = 1/2$ to get the desired result. Under the alternative, for all $\Sigma \in G(\alpha, L, \psi)$, we decompose $\widehat{\mathcal{A}}_n - \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n)$ into a sum of two uncorrelated terms.

$$\begin{aligned}
\widehat{\mathcal{A}}_n - \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) &= \frac{1}{n(n-1)(p-T)^2}\sum_{1\le k\neq l\le n}\sum_{j=1}^{T}w_j^*\sum_{1+T\le i_1,i_2\le p}(X_{k,i_1}X_{k,i_1-j}-\sigma_j)(X_{l,i_2}X_{l,i_2-j}-\sigma_j)\\
&\quad + \frac{2}{n(p-T)}\sum_{k=1}^{n}\sum_{j=1}^{T}w_j^*\sum_{i_1=T+1}^{p}(X_{k,i_1}X_{k,i_1-j}-\sigma_j)\sigma_j. \qquad (2.33)
\end{aligned}$$

Then the variance of $\widehat{\mathcal{A}}_n$ will be given as a sum of two terms,

$$\mathrm{Var}_\Sigma(\widehat{\mathcal{A}}_n) = \frac{R_1}{n(n-1)(p-T)^4} + \frac{R_2}{n(p-T)^2},$$

where

$$\begin{aligned}
R_1 &= 2\mathbb{E}_\Sigma\Big(\sum_{j=1}^{T}w_j^*\sum_{T+1\le i_1,i_2\le p}(X_{1,i_1}X_{1,i_1-j}-\sigma_j)(X_{2,i_2}X_{2,i_2-j}-\sigma_j)\Big)^2,\\
R_2 &= 4\mathbb{E}_\Sigma\Big(\sum_{j=1}^{T}w_j^*\sum_{i_1=T+1}^{p}(X_{1,i_1}X_{1,i_1-j}-\sigma_j)\sigma_j\Big)^2.
\end{aligned}$$

Let us deal first with R_1:

$$\begin{aligned} R_1 &= 2\sum_{1\le j,j'<T} w_j^* w_{j'}^* \sum_{T+1\le i_1,i_3\le p} \mathbb{E}_\Sigma[(X_{1,i_1}X_{1,i_1-j}-\sigma_j)(X_{1,i_3}X_{1,i_3-j'}-\sigma_j')] \\ &\quad\cdot \sum_{T+1\le i_2,i_4\le p} \mathbb{E}_\Sigma[(X_{2,i_2}X_{2,i_2-j}-\sigma_j)(X_{2,i_4}X_{2,i_4-j'}-\sigma_j')] \\ &= 2\sum_{1\le j,j'<T} w_j^* w_{j'}^* \Big(\sum_{T+1\le i_1,i_3\le p} (\sigma_{|i_1-i_3|}\sigma_{|i_1-i_3-j+j'|}+\sigma_{|i_1-i_3-j|}\sigma_{|i_1-i_3+j'|})\Big)^2 \\ &= 2\sum_{1\le j,j'<T} w_j^* w_{j'}^* \Big(\sum_{r=-p+T+1}^{p-(T+1)} (p-T-|r|)(\sigma_{|r|}\sigma_{|r-j+j'|}+\sigma_{|r-j|}\sigma_{|r+j'|})\Big)^2 \end{aligned}$$

Our aim here is to find an upper bound of R_1. In R_1 we distinguish two cases: the first one when for $j=j'$ and the second one when $j\neq j'$. Let us begin with the case when $j=j'$:

$$\begin{aligned} R_{1,1} &:= 2\sum_{j=1}^T w_j^{*2}\Big(\sum_{r=-p+T+1}^{p-(T+1)}(p-T-|r|)(\sigma_{|r|}^2+\sigma_{|r-j|}\sigma_{|r+j|})\Big)^2 \\ &= 2\sum_{j=1}^T w_j^{*2}\Big((p-T)(\sigma_0^2+\sigma_j^2)+2\sum_{r=1}^{p-(T+1)}(p-T-r)(\sigma_r^2+\sigma_{|r-j|}\sigma_{|r+j|})\Big)^2 \\ &= 2\sum_{j=1}^T w_j^{*2}\Big[(p-T)^2(\sigma_0^2+\sigma_j^2)^2+4\Big(\sum_{r=1}^{p-(T+1)}(p-T-r)(\sigma_r^2+\sigma_{|r-j|}\sigma_{|r+j|})\Big)^2 \\ &\quad +4(p-T)(\sigma_0^2+\sigma_j^2)\sum_{r=1}^{p-(T+1)}(p-T-r)(\sigma_r^2+\sigma_{|r-j|}\sigma_{|r+j|})\Big]. \end{aligned}$$

Let us bound from above each term on the right-hand side of the previous equality:

$$\begin{aligned} R_{1,1,1} &:= 2\sum_{j=1}^T w_j^{*2}(p-T)^2(\sigma_0^2+\sigma_j^2)^2 = 2(p-T)^2\Big(\sum_{j=1}^T w_j^{*2}+2\sum_{j=1}^T w_j^{*2}\sigma_j^2+\sum_{j=1}^T w_j^{*2}\sigma_j^4\Big) \\ &\le 2(p-T)^2\Big(\frac{1}{2}+3L\cdot(\sup_j w_j^*)^2\Big) = (p-T)^2(1+o(1)). \end{aligned} \tag{2.34}$$

Now we give an upper bound for the second term of (2.38). Using Cauchy-Schwarz inequality we get,

$$\begin{aligned} R_{1,1,2} &:= 8\sum_{j=1}^T w_j^{*2}\Big[\sum_{r=1}^{p-(T+1)}(p-T-r)(\sigma_r^2+\sigma_{|r-j|}\sigma_{|r+j|})\Big]^2 \\ &\le 8(p-T)^2\sum_{j=1}^T w_j^{*2}\Big[\sum_{r=1}^{p-(T+1)}\sigma_r^2+(\sum_{r=1}^{p-(T+1)}\sigma_{|r-j|}^2)^{1/2}(\sum_{r=1}^{p-(T+1)}\sigma_{|r+j|}^2)^{1/2}\Big]^2 \\ &\le 16(p-T)^2\sum_{j=1}^T w_j^{*2}\Big[\Big(\sum_{r=1}^{p-(T+1)}\sigma_r^2\Big)^2+(\sum_{r=1}^{p-(T+1)}\sigma_{|r-j|}^2)(\sum_{r=1}^{p-(T+1)}\sigma_{|r+j|}^2)\Big]. \end{aligned}$$

Again we will treat each term of the previous inequality apart. Let us see first, that if

$(r \le j \Longrightarrow w_j^* \le w_r^*)$. In addition to the previous remark we use the class property to get:

$$
\begin{aligned}
R_{1,1,2,1} &:= \sum_{j=1}^{T} w_j^{*2}\Big(\sum_{r=1}^{p-(T+1)} \sigma_r^2\Big)^2 \le \sum_{j=1}^{T} w_j^{*2}\Big(\sum_{r=1}^{j}\sigma_r^2 + \sum_{r=j+1}^{p-(T+1)} \frac{r^{2\alpha}}{j^{2\alpha}}\sigma_r^2\Big)^2 \\
&\le 2\sum_{j=1}^{T}\Big(\sum_{r=1}^{j} w_r^*\sigma_r^2\Big)^2 + 2(\sup_j w_j^*)^2 \sum_{j=1}^{T}\frac{1}{j^{4\alpha}}\Big(\sum_{r=j+1}^{p-(T+1)} r^{2\alpha}\sigma_r^2\Big)^2 \\
&\le 2\cdot T\cdot \mathbb{E}_\Sigma^2(\widehat{\mathcal{A}}_n) + (\sup_j w_j^*)^2\cdot k_0(\alpha,L). \qquad (2.35)
\end{aligned}
$$

Indeed, for $\alpha > 1/4$, we have, $\sum_{j=1}^{T} j^{-4\alpha} \le (4\alpha-1)^{-1}$ and we can take $k_0(\alpha,L) = 2L^2(4\alpha-1)^{-1}$. Using similar arguments we prove that,

$$
\begin{aligned}
R_{1,1,2,2} &:= \sum_{j=1}^{T} w_j^{*2}\Big(\sum_{\substack{r=1\\|r-j|<j}}^{p-(T+1)} \sigma_{|r-j|}^2 + \sum_{\substack{r=1\\|r-j|\ge j}}^{p-(T+1)} \sigma_{|r-j|}^2\Big)\Big(\sum_{r=1}^{p-(T+1)}\sigma_{r+j}^2\Big) \\
&\le \sum_{j=1}^{T} w_j^*\Big(\sum_{\substack{r=1\\|r-j|<j}}^{p-(T+1)} w_{|r-j|}^*\sigma_{|r-j|}^2\Big)\Big(\sum_{r=1}^{p-(T+1)}\sigma_{r+j}^2\Big) \\
&+ \sum_{j=1}^{T} w_j^{*2}\Big(\sum_{\substack{r=1\\|r-j|\ge j}}^{p-1}\frac{|r-j|^{2\alpha}}{j^{2\alpha}}\sigma_{|r-j|}^2\Big)\Big(\sum_{r=1}^{p-(T+1)}\frac{(r+j)^{2\alpha}}{j^{2\alpha}}\sigma_{r+j}^2\Big) \\
&\le (\sup_j w_j^*)\cdot T\cdot \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n)\cdot L + (\sup_j w_j^*)^2\cdot k_0(\alpha,L). \qquad (2.36)
\end{aligned}
$$

The third term in $R_{1,1}$ is treated by similar arguments:

$$
\begin{aligned}
R_{1,1,3} &= (p-T)\sum_{j=1}^{T} w_j^{*2}(\sigma_0^2+\sigma_j^2)\sum_{r=1}^{p-(T+1)}(p-T-r)(\sigma_r^2 + \sigma_{|r-j|}\sigma_{|r+j|}) \\
&\le (p-T)^2\cdot \sup_j(\sigma_0^2+\sigma_j^2)\cdot\Big\{\sum_{j=1}^{T} w_j^*\sum_{r=1}^{j} w_r^*\sigma_r^2 + (\sup_j w_j^{*2})\sum_{j=1}^{T}\frac{1}{j^{2\alpha}}\sum_{r=j+1}^{p-(T+1)} r^{2\alpha}\sigma_r^2 \\
&+ (\sup_j w_j^{*2})\sum_{j=1}^{T}\Big(\sum_{r=1}^{p-(T+1)}\sigma_{|r-j|}^2\Big)^{1/2}\Big(\sum_{r=1}^{p-(T+1)}\frac{(r+j)^{2\alpha}}{j^{2\alpha}}\sigma_{r+j}^2\Big)^{1/2}\Big\} \\
&\le 2(p-T)^2\Big\{O(\sqrt{T})\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) + (\sup_j w_j^{*2})\cdot\Big(O(\max\{1, T^{-2\alpha+1}, T^{-\alpha+1}\})\Big)\Big\} \\
&\le 2(p-T)^2\cdot\Big\{O(\sqrt{T})\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) + o(1)\Big\}. \qquad (2.37)
\end{aligned}
$$

Put together bounds in (2.34) to (2.37), we can deduce that,

$$
R_{1,1} \le (p-T)^2(1+o(1)) + (p-T)^2\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n)\cdot O(\sqrt{T}) + (p-T)^2\cdot\mathbb{E}_\Sigma^2(\widehat{\mathcal{A}}_n)\cdot O(T). \quad (2.38)
$$

Now, we will treat the case when, $j \ne j'$.

$$
\begin{aligned}
R_{1,2} &:= 2\sum\sum_{1\le j\ne j'\le T} w_j^* w_{j'}^*\Big(\sum_{r=-p+T+1}^{p-(T+1)}(p-|r|)(\sigma_{|r|}\sigma_{|r-j+j'|} + \sigma_{|r-j|}\sigma_{|r+j'|})\Big)^2 \\
&\le 4(p-T)^2\sum\sum_{1\le j\ne j'\le T} w_j^* w_{j'}^*\Big[\Big(\sum_{r=-p+T+1}^{p-(T+1)}|\sigma_{|r|}\sigma_{|r-j+j'|}|\Big)^2 + \Big(\sum_{r=-p+T+1}^{p-(T+1)}|\sigma_{|r-j|}\sigma_{|r+j'|}|\Big)^2\Big].
\end{aligned}
$$

These last two terms are treated similarly, so let us deal with the first one. By using the same arguments as previously, we have

$$\begin{aligned} R_{1,2,2} &:= \sum_{1\le j\ne j'\le T} w_j^* w_{j'}^* \Big(|\sigma_{|j'-j|}| + \sum_{\substack{r=-p+T+1\\ r\ne 0}}^{p-(T+1)} |\sigma_{|r|}\sigma_{|r-j+j'|}| \Big)^2 \\ &\le 2 \sum_{1\le j\ne j'\le T} w_j^* w_{j'}^* \sigma_{|j'-j|}^2 + 4 \sum_{1\le j\ne j'\le T} w_j^* w_{j'}^* (\sum_{r=1}^{p-(T+1)} \sigma_r^2)(\sum_{\substack{r=-p+T+1\\ r\ne 0}}^{p-(T+1)} \sigma_{|r-j+j'|}^2). \end{aligned}$$

We decompose the sum over $j \ne j'$ over sets where $\{|j'-j| \le j\}$ and $\{|j'-j| > j\}$ and use $1 \le |j'-j|^{2\alpha}/j^{2\alpha}$ over the later, then similarly for sums over r:

$$\begin{aligned} R_{1,2,2} &\le 2 \sum_{\substack{1\le j\ne j'\le T\\ |j'-j|<j}} w_{j'}^* w_{|j'-j|}^* \sigma_{|j'-j|}^2 + 2 \sum_{\substack{1\le j\ne j'\le T\\ |j'-j|>j}} w_j^* w_{j'}^* \frac{|j'-j|^{2\alpha}}{j^{2\alpha}} \sigma_{|j'-j|}^2 + 4 \sum_{1\le j\ne j'\le T} \Big(\sum_{r=1}^{j} w_r^* \sigma_r^2 \\ &+ w_j^* \sum_{r=j+1}^{p-(T+1)} \frac{r^{2\alpha}}{j^{2\alpha}} \sigma_r^2 \Big) \Big(\sum_{\substack{r=-p+T+1\\ |r-j+j'|<j'}}^{p-(T+1)} w_{|r-j+j'|}^* \sigma_{|r-j+j'|}^2 + w_{j'}^* \sum_{\substack{r=-p+T+1\\ |r-j+j'|\ge j'}}^{p-(T+1)} \frac{|r-j+j'|^{2\alpha}}{(j')^{2\alpha}} \sigma_{|r-j+j'|}^2 \Big) \\ &\le 4 \cdot (\sup_j w_j^*) \cdot T \cdot \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) + 4L \cdot (\sup_j w_j^*)^2 \cdot O(\max\{1, T^{-2\alpha+1}\}) \\ &+ O(T^2) \cdot \mathbb{E}_\Sigma^2(\widehat{\mathcal{A}}_n) 16L \cdot (\sup_j w_j^*) \cdot T \cdot O(\max\{1, T^{-2\alpha+1}\}) \cdot \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) \\ &+ 16L^2 \cdot (\sup_j w_j^*)^2 \cdot O(\max\{1, T^{-4\alpha+2}\}). \end{aligned}$$

As consequence, for all $\alpha > 1/4$,

$$R_{1,2} \le (p-T)^2 \{ \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) \cdot O(\sqrt{T}) + \mathbb{E}_\Sigma^2(\widehat{\mathcal{A}}_n) \cdot O(T^2) + \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) \cdot O(T^{3/2-2\alpha}) + o(1) \}. \quad (2.39)$$

Finally put together (2.38) and (2.39) to get (2.8). In order to find an upper bound for the variance of $\widehat{\mathcal{A}}_n$ we still have to bound from above R_2.

$$\begin{aligned} R_2 &= 4 \sum_{1\le j,j'<T} w_j^* w_{j'}^* \sigma_j \sigma_{j'} \sum_{T+1\le i_1,i_2\le p} \mathbb{E}_\Sigma[(X_{i,i_1} X_{i,i_1-j} - \sigma_j)(X_{i,i_2} X_{i,i_2-j'} - \sigma_{j'})] \\ &= 4 \sum_{1\le j,j'<T} w_j^* w_{j'}^* \sigma_j \sigma_{j'} \sum_{T+1\le i_1,i_2\le p} (\sigma_{|i_1-i_2|}\sigma_{|i_1-i_2-j+j'|} + \sigma_{|i_1-i_2-j|}\sigma_{|i_1-i_2+j'|}) \\ &= 4 \sum_{1\le j,j'<T} w_j^* w_{j'}^* \sigma_j \sigma_{j'} \sum_{r=-p+T+1}^{p-(T+1)} (p-T-|r|)(\sigma_{|r|}\sigma_{|r-j+j'|} + \sigma_{|r-j|}\sigma_{|r+j'|}). \end{aligned}$$

Let us begin by the first case when $j = j'$. It is easily seen that,

$$\begin{aligned} R_{2,1} &:= 4 \sum_{j=1}^{T} w_j^{*2} \sigma_j^2 \sum_{r=-p+T+1}^{p-(T+1)} (p-T-|r|)(\sigma_{|r|}^2 + \sigma_{|r-j|}\sigma_{|r+j|}) \\ &\le 8L \cdot p \cdot (\sup_j w_j^*) \cdot \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) \end{aligned} \quad (2.40)$$

While, when $j \neq j'$, we can prove that,

$$
\begin{aligned}
R_{2,2} &:= 4 \sum\sum_{1\le j\neq j'\le T} w_j^* w_{j'}^* \sigma_j \sigma_{j'} \sum_{r=-p+T+1}^{p-(T+1)} (p-T-|r|)(\sigma_{|r|}\sigma_{|r-j+j'|} + \sigma_{|r-j|}\sigma_{|r+j'|}) \\
&\le 4 \Big(\sum\sum_{1\le j\neq j'\le T} w_j^* w_{j'}^* \sigma_j^2 \sigma_{j'}^2 \Big)^{\frac{1}{2}} \Big(\sum\sum_{1\le j\neq j'\le T} w_j^* w_{j'}^* \Big(\sum_{r=-p+T+1}^{p-(T+1)} (p-T-|r|)(\sigma_{|r|}\sigma_{|r-j+j'|} + \sigma_{|r-j|}\sigma_{|r+j'|}) \Big)^2 \Big)^{\frac{1}{2}} \\
&\le 4\, \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) \cdot (R_{1,2})^{1/2}.
\end{aligned}
$$

We use the bound obtained in (2.39) to deduce that:

$$
R_{2,2} \le (p-T)\Big(\mathbb{E}_\Sigma^2(\widehat{\mathcal{A}}_n) \cdot O(T) + \mathbb{E}_\Sigma^{3/2}(\widehat{\mathcal{A}}_n) \cdot (O(T^{1/4}) + O(T^{3/4-\alpha})) + \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) \cdot o(1) \Big). \tag{2.41}
$$

Put together (2.40) and (2.41) to get (2.9). □

Proof of Proposition 2.2. Assume that $n(p-T) \cdot \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) \asymp 1$, to prove the asymptotic normality of $n(p-T) \cdot \widehat{\mathcal{A}}_n$, we use the decomposition (2.33) of the test statistic. first let us show that,

$$
\widehat{\mathcal{A}}_{n,1} := 2 \sum_{k=1}^{n} \sum_{j=1}^{T} w_j^* \sum_{i_1=T+1}^{p} (X_{k,i_1} X_{k,i_1-j} - \sigma_j)\sigma_j \xrightarrow{P} 0
$$

By Markov inequality we have, $\forall \varepsilon > 0$,

$$
\mathbb{P}_\Sigma\Big(|2 \sum_{k=1}^{n} \sum_{j=1}^{T} w_j^* \sum_{i_1=T+1}^{p} (X_{k,i_1} X_{k,i_1-j} - \sigma_j)\sigma_j| > \varepsilon \Big) \quad \le \quad \frac{n \cdot R_2}{\varepsilon^2}
$$

According to (2.9), and under the assumption that $np \cdot \mathbb{E}(\widehat{\mathcal{A}}_n) \asymp 1$, we can see that,

$$
\begin{aligned}
n \cdot R_2 &\le n \cdot (p-T)\{ \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) \cdot o(1) + \mathbb{E}_\Sigma^{3/2}(\widehat{\mathcal{A}}_n) \cdot (O(T^{1/4}) + O(T^{3/4-\alpha})) + \mathbb{E}_\Sigma^2(\widehat{\mathcal{A}}_n) \cdot O(T) \} \\
&\le o(1) + O\Big(\frac{T^{1/4} + T^{3/4-\alpha}}{\sqrt{n(p-T)}} \Big) + O\Big(\frac{T}{n(p-T)} \Big) = o(1)
\end{aligned}
$$

since $T/p \longrightarrow 0$ and for all $\alpha > 1/4$. Which involves by Slutsky theorem that for proving the asymptotic normality it is sufficient to show that,

$$
\widehat{\mathcal{A}}_{n,2} := \frac{1}{n(p-T)} \sum_{1\le k\neq l\le n} \sum_{j=1}^{T} w_j^* \sum_{T+1\le i_1, i_2\le p} (X_{k,i_1} X_{k,i_1-j} - \sigma_j)(X_{l,i_2} X_{l,i_2-j} - \sigma_j) \xrightarrow{L} N(0,1) \tag{2.42}
$$

In order to prove this previous convergence, we are led to apply theorem 1 of [50]. This result is an application of the more general theorem of asymptotic normality for martingale differences, see e.g. [83]. $\widehat{\mathcal{A}}_{n,2}$ is a centered, 1-degenerate, U-Statistic of second order, with kernel $H_n(X_1, X_2)$ defined by,

$$
H_n(X_1, X_2) := \frac{1}{n(p-T)} \sum_{j=1}^{T} w_j^* \sum_{T+1\le i_1, i_2\le p} (X_{1,i_1} X_{1,i_1-j} - \sigma_j)(X_{2,i_2} X_{2,i_2-j} - \sigma_j)
$$

Therefore we should check that $\mathbb{E}_\Sigma(H_n^2(X_1,X_2)) < +\infty$ and

$$\frac{\mathbb{E}_\Sigma(G_n^2(X_1,X_2)) + n^{-1}\mathbb{E}_\Sigma(H_n^4(X_1,X_2))}{\mathbb{E}_\Sigma^2(H_n^2(X_1,X_2))} \longrightarrow 0 \tag{2.43}$$

where $G_n(x,y) := \mathbb{E}(H_n(X_1,x)H_n(X_1,y))$, for $x,y \in \mathbb{R}^p$. The proof of (2.43) is given separately hereafter.

The asymptotic normality under $\Sigma = I$ (the null hypothesis) is only simpler as $\sigma_j = 0$ for all $j \geq 1$, for n, $p \to \infty$. However, under the null hypothesis we prove separately (hereafter) that

$$n(p-T)\widehat{\mathcal{A}}_n \to \mathcal{N}(0,1), \text{ for } p \to \infty \text{ and for any fixed } n \geq 2. \tag{2.44}$$

□

Proof of (2.43). To show (2.43), we first calculate $G_n(x,y)$ and $\mathbb{E}_\Sigma(H_n^2(X_1,X_2))$. That is,

$$\begin{aligned} G_n(x,y) \;=\; & \frac{1}{n^2(p-T)^2} \sum_{1\leq j_1,j_2<T} w_{j_1}^* w_{j_2}^* \sum_{r=-p+T+1}^{p-(T+1)} (p-T-|r|)(\sigma_{|r|}\sigma_{|r-j_1+j_2|} + \sigma_{|r-j_1|}\sigma_{|r+j_2|}) \\ & \sum_{1\leq i_1,i_2\leq p} (x_{i_1}x_{i_1-j_1} - \sigma_{j_1})(y_{i_2}y_{i_2-j_2} - \sigma_{j_2}) \end{aligned} \tag{2.45}$$

Note that, under the assumption $np\cdot\mathbb{E}(\widehat{\mathcal{A}}_n) \asymp 1$, $\alpha > 1/4$, $p\psi^{1/\alpha} \to +\infty$ and using (2.8), we have,

$$\mathbb{E}_\Sigma(H_n^2(X_1,X_2)) = \frac{1+o(1)}{2n^2}$$

Now, let us verify that, uniformly over Σ,

$$\mathbb{E}_\Sigma(G_n^2(X_1,X_2))/\mathbb{E}_\Sigma^2(H_n^2(X_1,X_2)) = o(1). \tag{2.46}$$

We write

$$\begin{aligned} & \frac{\mathbb{E}_\Sigma(G_n^2(X_1,X_2))}{\mathbb{E}_\Sigma^2(H_n^2(X_1,X_2))} = 4n^4\cdot\mathbb{E}_\Sigma(G_n^2(X_1,X_2)) \\ = \; & \frac{4}{(p-T)^4} \sum_{1\leq j_1,j_2,j_3,j_4<T} w_{j_1}^* w_{j_2}^* w_{j_3}^* w_{j_4}^* \sum_{-p+T+1\leq r_1,r_2\leq p-(T+1)} (p-T-|r_1|)(p-T-|r_2|) \\ & \cdot(\sigma_{|r_1|}\sigma_{|r_1-j_1+j_2|} + \sigma_{|r_1-j_1|}\sigma_{|r_1+j_2|})(\sigma_{|r_2|}\sigma_{|r_2-j_3+j_4|} + \sigma_{|r_2-j_3|}\sigma_{|r_2+j_4|}) \\ & \cdot \sum_{T+1\leq i_1,i_3\leq p} \mathbb{E}_\Sigma[(X_{1,i_1}X_{1,i_1-j_1} - \sigma_{j_1})(X_{1,i_3}X_{1,i_3-j_3} - \sigma_{j_3})] \\ & \cdot \sum_{T+1\leq i_2,i_4\leq p} \mathbb{E}_\Sigma[(X_{2,i_2}X_{2,i_2-j_2} - \sigma_{j_2})(X_{2,i_4}X_{2,i_4-j_4} - \sigma_{j_4})] \end{aligned} \tag{2.47}$$

We calculate each expected value, and bound from above by the absolute value, we obtain:

$$\begin{aligned} & 4n^4\cdot\mathbb{E}_\Sigma(G_n^2(X_1,X_2)) \\ \leq \; & 4 \sum_{1\leq j_1,j_2,j_3,j_4<T} w_{j_1}^* w_{j_2}^* w_{j_3}^* w_{j_4}^* \sum_{-p+T+1\leq r_1,r_2,r_3,r_4\leq p-(T+1)} \\ & \cdot(|\sigma_{|r_1|}\sigma_{|r_1-j_1+j_2|}| + |\sigma_{|r_1-j_1|}\sigma_{|r_1+j_2|}|)(|\sigma_{|r_2|}\sigma_{|r_2-j_3+j_4|}| + |\sigma_{|r_2-j_3|}\sigma_{|r_2+j_4|}|) \\ & \cdot(|\sigma_{|r_3|}\sigma_{|r_3-j_1+j_3|}| + |\sigma_{|r_3-j_1|}\sigma_{|r_3+j_3|}|)(|\sigma_{|r_4|}\sigma_{|r_4-j_2+j_4|}| + |\sigma_{|r_4-j_2|}\sigma_{|r_4+j_4|}|) \end{aligned} \tag{2.48}$$

In (2.48) there are sixteen terms, that are all treated the same way, then we deal with,

$$\begin{aligned}\mathcal{G} &:= 4\sum_{1\le j_1,j_2,j_3,j_4<T} w^*_{j_1}w^*_{j_2}w^*_{j_3}w^*_{j_4}\sum_{-p+T+1\le r_1,r_2,r_3,r_4\le p-(T+1)}\\ &\quad\cdot\left|\sigma_{|r_1|}\sigma_{|r_1-j_1+j_2|}\sigma_{|r_2|}\sigma_{|r_2-j_3+j_4|}\sigma_{|r_3|}\sigma_{|r_3-j_1+j_3|}\sigma_{|r_4|}\sigma_{|r_4-j_2+j_4|}\right|\end{aligned}$$

To bound from above this previous quantity, we distinguish four cases, based on the indices j_1, j_2, j_3 and j_4. Let us begin by the the first case, when $j_1 = j_2 = j_3 = j_4$:

$$\begin{aligned}\mathcal{G}_1 &:= 4\sum_{j_1=1}^{T} w^{*4}_{j_1}\sum_{-p+T+1\le r_1,r_2,r_3,r_4\le p-(T+1)}\sigma^2_{|r_1|}\sigma^2_{|r_2}\sigma^2_{|r_3|}\sigma^2_{|r_4|}\\ &\le 4\cdot(\sup_j w^*_j)^4\cdot T\cdot(2L)^4 = O(\frac{1}{T}) = o(1)\end{aligned}$$

We consider the second case, where there are two different values of indices, either two groups of two, or one group of three and one separate index. For the first one, let us assume that ($j_1 = j_4$, $j_2 = j_3$ and $j_1 \neq j_2$),

$$\begin{aligned}\mathcal{G}_2 &:= 4\sum_{1\le j_1\neq j_2<T} w^{*2}_{j_1}w^{*2}_{j_2}\sum_{-p+T+1\le r_1,r_2,r_3,r_4\le p-(T+1)}\left|\sigma_{|r_1|}\sigma_{|r_1-j_1+j_2|}\sigma_{|r_2|}\sigma_{|r_2-j_2+j_1|}\right|\\ &\qquad\cdot\left|\sigma_{|r_3|}\sigma_{|r_3-j_1+j_2|}\sigma_{|r_4|}\sigma_{|r_4-j_2+j_1|}\right|\\ &= 4\sum_{1\le j_1\neq j_2<T} w^{*2}_{j_1}w^{*2}_{j_2}\cdot\Big(2\left|\sigma_0\,\sigma_{|j_1-j_2|}\right| + \sum_{\substack{r_1=-p+T+1\\ r_1\neq 0,\, r_1\neq j_1-j_2}}^{p-(T+1)}\left|\sigma_{|r_1|}\sigma_{|r_1-j_1+j_2|}\right|\Big)^2\\ &\quad\cdot\Big(2\left|\sigma_0\,\sigma_{|j_1-j_2|}\right| + \sum_{\substack{r_2=-p+T+1\\ r_2\neq 0,\, r_2\neq j_2-j_1}}^{p-(T+1)}\left|\sigma_{|r_2|}\sigma_{|r_2-j_2+j_1|}\right|\Big)^2\end{aligned}$$

We apply the Cauchy-Schwarz inequality with respect to r_1 and r_2 separately to get:

$$\begin{aligned}\mathcal{G}_2 &\le 4\cdot(2+2L)^2\sum_{1\le j_1\neq j_2<T} w^{*2}_{j_1}w^{*2}_{j_2}\cdot\Big\{4\,\sigma^2_{|j_1-j_2|} + 2\Big(\sum_{\substack{r_1=-p+T+1\\ r_1\neq 0,\, r_1\neq j_1-j_2}}^{p-(T+1)}\sigma^2_{|r_1|}\Big)\Big(\sum_{\substack{r_1=-p+T+1\\ r_1\neq 0,\, r_1\neq j_1-j_2}}^{p-(T+1)}\sigma^2_{|r_1-j_1+j_2|}\Big)\Big\}\\ &\le 16\cdot(2+2L)^2\cdot\Big(\sum_{1\le j_1\neq j_2<T} w^{*2}_{j_1}w^{*2}_{j_2}\sigma^2_{|j_1-j_2|} + 2L\cdot\sum_{1\le j_1\neq j_2<T} w^{*2}_{j_1}w^{*2}_{j_2}\sum_{r_1\neq 0}\sigma^2_{|r_1|}\Big)\\ &\le 16\cdot(2+2L)^2\cdot\Big\{(\sup_{j_2} w^{*2}_{j_2})\cdot\sum_{j_1=1}^{T} w^{*2}_{j_1}\sum_{j_2=1}^{T}\sigma^2_{|j_1-j_2|}\\ &+ 2L\cdot\sum_{j_2=1}^{T} w^{*2}_{j_2}\cdot\Big(\sum_{j_1=1}^{T} w^*_{j_1}\sum_{\substack{r_1;r_1\neq 0\\ |r_1|\le j_1}} w_{|r_1|}\sigma^2_{|r_1|} + \sum_{j_1=1}^{T} w^{*2}_{j_1}\sum_{\substack{r_1;r_1\neq 0\\ |r_1|>j_1}}\frac{|r_1|^{2\alpha}}{j_1^{2\alpha}}\sigma^2_{|r_1|}\Big\}\\ &\le O\Big(\frac{1}{T}\Big) + O(\sqrt{T})\cdot\mathbb{E}(\widehat{\mathcal{A}}_n) + O\Big(\frac{1}{T}\Big)\cdot\max\{1, T^{-2\alpha+1}\} = o(1)\end{aligned}$$

since $\mathbb{E}(\widehat{\mathcal{A}}_n) \asymp 1/np$ and $T/p \to 0$. Similar argument to prove that for $j_1 = j_3 = j_4$ and $j_1 \neq j_2$, we have,

$$4\sum_{1\le j_1\neq j_2<T} w^{*3}_{j_1}w^*_{j_2}\sum_{-p+T+1\le r_1,r_2,r_3,r_4\le p-(T+1)}\sigma_{|r_1|}\sigma_{|r_1-j_1+j_2|}\sigma^2_{|r_2|}\sigma^2_{|r_3|}\sigma_{|r_4|}\sigma_{|r_4-j_2+j_1|} = o(1)$$

which finishes the second case. Now let us assume that we have three different values, ($j_1 = j_4$ and $j_1 \neq j_2 \neq j_3$), we obtain,

$$\begin{aligned}
\mathcal{G}_3 &:= 4 \sum_{1 \le j_1 \neq j_2 \neq j_3 < T} w_{j_1}^{*2} w_{j_2}^{*} w_{j_3}^{*} \sum_{-p+T+1 \le r_1, r_2, r_3, r_4 \le p-(T+1)} \left| \sigma_{|r_1|} \sigma_{|r_1 - j_1 + j_2|} \sigma_{|r_2|} \sigma_{|r_2 - j_3 + j_1|} \right| \\
&\qquad \cdot \left| \sigma_{|r_3|} \sigma_{|r_3 - j_1 + j_3|} \sigma_{|r_4|} \sigma_{|r_4 - j_2 + j_1|} \right| \\
&= 4 \sum_{1 \le j_1 \neq j_2 \neq j_3 < T} w_{j_1}^{*2} w_{j_2}^{*} w_{j_3}^{*} \Big(2 \left| \sigma_{|j_2 - j_1|} \right| + \sum_{\substack{r_1 = -p+T+1 \\ r_1 \neq 0,\, r_1 \neq j_2 - j_1}}^{p-(T+1)} \left| \sigma_{|r_1|} \sigma_{|r_1 - j_1 + j_2|} \right| \Big) \\
&\quad \Big(2 \left| \sigma_{|j_1 - j_3|} \right| + \sum_{\substack{r_2 = -p+T+1 \\ r_2 \neq 0,\, r_2 \neq j_1 - j_3}}^{p-(T+1)} \left| \sigma_{|r_2|} \sigma_{|r_2 - j_1 + j_3|} \right| \Big) \cdot \Big(2 \left| \sigma_{|j_3 - j_1|} \right| + \sum_{\substack{r_3 = -p+T+1 \\ r_3 \neq 0,\, r_3 \neq j_3 - j_1}}^{p-(T+1)} \left| \sigma_{|r_3|} \sigma_{|r_3 - j_3 + j_1|} \right| \Big) \\
&\quad \Big(2 \left| \sigma_{|j_1 - j_2|} \right| + \sum_{\substack{r_4 = -p+T+1 \\ r_4 \neq 0,\, r_4 \neq j_1 - j_2}}^{p-(T+1)} \left| \sigma_{|r_4|} \sigma_{|r_4 - j_2 + j_1|} \right| \Big)
\end{aligned}$$

and hence

$$\begin{aligned}
\mathcal{G}_{3,1} &:= \sum_{1 \le j_1 \neq j_2 \neq j_3 < T} w_{j_1}^{*2} w_{j_2}^{*} w_{j_3}^{*} \sigma_{|j_1 - j_2|}^2 \sigma_{|j_1 - j_3|}^2 \le (\sup_j w_j^*)^2 \sum_{j_1=1}^{T} w_{j_1}^{*2} \sum_{j_2=1}^{T} \sigma_{|j_1 - j_2|}^2 \sum_{j_3=1}^{T} \sigma_{|j_1 - j_3|}^2 \\
&\le (\sup_j w_j^*)^2 \cdot \frac{1}{2} \cdot 4L^2 = o(1).
\end{aligned}$$

Note that $\sup_r \sigma_r \le 1$ and by Cauchy-Schwarz we have $\sum_{\substack{r_4 = -p+T+1 \\ r_4 \neq 0,\, r_4 \neq j_1 - j_2}}^{p-(T+1)} \left| \sigma_{|r_4|} \sigma_{|r_4 - j_2 + j_1|} \right| \le \sum_{\substack{r_4 \\ r_4 \neq 0}} \sigma_{r_4}^2$.

Thus we get,

$$\begin{aligned}
\mathcal{G}_{3,2} &:= \sum_{1 \le j_1 \neq j_2 \neq j_3 < T} w_{j_1}^{*2} w_{j_2}^{*} w_{j_3}^{*} \sigma_{|j_1 - j_2|} \sigma_{|j_1 - j_3|}^2 \sum_{\substack{r_4 = -p+T+1 \\ r_4 \neq 0,\, r_4 \neq j_1 - j_2}}^{p-(T+1)} \left| \sigma_{|r_4|} \sigma_{|r_4 - j_2 + j_1|} \right| \\
&\le (\sup_j w_j^*) \cdot \sum_{j_1} w_{j_1}^{*2} \sum_{j_2} w_{j_2}^{*} \sum_{j_3} \sigma_{|j_1 - j_3|}^2 \Big(\sum_{\substack{|r_4| \le j_2 \\ r_4 \neq 0}} \sigma_{r_4}^2 + \sum_{\substack{|r_4| > j_2 \\ r_4 \neq 0}} \frac{|r_4|^{2\alpha}}{j_2^{2\alpha}} \sigma_{|r_4|}^2 \Big) \\
&\le (\sup_j w_j^*) \cdot \frac{1}{2} \cdot 2L \cdot \Big(T \cdot \mathbb{E}(\widehat{\mathcal{A}}_n) + (\sup_j w_j^*) \cdot \max\{1, T^{-2\alpha+1}\} \cdot 2L \Big) = o(1).
\end{aligned}$$

Moreover,

$$\begin{aligned}
\mathcal{G}_{3,3} &:= \sum_{1 \le j_1 \neq j_2 \neq j_3 < T} w_{j_1}^{*2} w_{j_2}^{*} w_{j_3}^{*} \sigma_{|j_1 - j_2|} \sigma_{|j_1 - j_3|} \sum_{\substack{r_2 = -p+T+1 \\ r_2 \neq 0,\, r_2 \neq j_1 - j_3}}^{p-(T+1)} \left| \sigma_{|r_2|} \sigma_{|r_2 - j_1 + j_3|} \right| \sum_{\substack{r_4 = -p+T+1 \\ r_4 \neq 0,\, r_4 \neq j_1 - j_2}}^{p-(T+1)} \left| \sigma_{|r_4|} \sigma_{|r_4 - j_2 + j_1|} \right| \\
&\le \sum_{j_1} w_{j_1}^{*2} \cdot \sum_{j_2} w_{j_2}^{*} \Big(\sum_{\substack{|r_2| \le j_2 \\ r_2 \neq 0}} \sigma_{r_2}^2 + \sum_{\substack{|r_2| > j_2 \\ r_2 \neq 0}} \frac{|r_2|^{2\alpha}}{j_2^{2\alpha}} \sigma_{|r_2|}^2 \Big) \sum_{j_3} w_{j_3}^{*} \Big(\sum_{\substack{|r_4| \le j_3 \\ r_4 \neq 0}} \sigma_{r_4}^2 + \sum_{\substack{|r_4| > j_3 \\ r_4 \neq 0}} \frac{|r_4|^{2\alpha}}{j_3^{2\alpha}} \sigma_{|r_4|}^2 \Big) \\
&\le \frac{1}{2} \cdot \Big(T \cdot \mathbb{E}(\widehat{\mathcal{A}}_n) + (\sup_j w_j^*) \cdot \max\{1, T^{-2\alpha+1}\} \cdot 2L \Big)^2 = o(1)
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{G}_{3,4} &:= \sum_{1\le j_1\ne j_2\ne j_3<T} w_{j_1}^{*2} w_{j_2}^* w_{j_3}^* \sigma_{|j_2-j_1|} \sum_{\substack{r_2=-p+T+1\\ r_2\ne 0,\, r_2\ne j_1-j_3}}^{p-(T+1)} \left|\sigma_{|r_2|}\sigma_{|r_2-j_1+j_3|}\right| \sum_{\substack{r_3=-p+T+1\\ r_3\ne 0,\, r_3\ne j_3-j_1}}^{p-(T+1)} \left|\sigma_{|r_3|}\sigma_{|r_3-j_3+j_1|}\right| \\
&\quad \sum_{\substack{r_4=-p+T+1\\ r_4\ne 0,\, r_4\ne j_1-j_2}}^{p-(T+1)} \left|\sigma_{|r_4|}\sigma_{|r_4-j_2+j_1|}\right| \\
&\le \sum_{j_1} w_{j_1}^{*2} \cdot \sum_{j_2} w_{j_2}^* \Big(\sum_{\substack{|r_2|\le j_2\\ r_2\ne 0}} \sigma_{r_2}^2 + \sum_{\substack{|r_2|>j_2\\ r_2\ne 0}} \frac{|r_2|^{2\alpha}}{j_2^{2\alpha}} \sigma_{|r_2|}^2 \Big) \sum_{j_3} w_{j_3}^* \Big(\sum_{\substack{|r_4|\le j_3\\ r_4\ne 0}} \sigma_{r_4}^2 + \sum_{\substack{|r_4|>j_3\\ r_4\ne 0}} \frac{|r_4|^{2\alpha}}{j_3^{2\alpha}} \sigma_{|r_4|}^2 \Big) \cdot 2L \\
&= o(1).
\end{aligned}$$

Similarly we show that

$$\begin{aligned}
\mathcal{G}_{3,5} &:= \sum_{1\le j_1\ne j_2\ne j_3<T} w_{j_1}^{*2} w_{j_2}^* w_{j_3}^* \sum_{\substack{r_1=-p+T+1\\ r_1\ne 0,\, r_1\ne j_2-j_1}}^{p-(T+1)} \left|\sigma_{|r_1|}\sigma_{|r_1-j_1+j_2|}\right| \sum_{\substack{r_2=-p+T+1\\ r_2\ne 0,\, r_2\ne j_1-j_3}}^{p-(T+1)} \left|\sigma_{|r_2|}\sigma_{|r_2-j_1+j_3|}\right| \\
&\quad \cdot \sum_{\substack{r_3=-p+T+1\\ r_3\ne 0,\, r_3\ne j_3-j_1}}^{p-(T+1)} \left|\sigma_{|r_3|}\sigma_{|r_3-j_3+j_1|}\right| \sum_{\substack{r_4=-p+T+1\\ r_4\ne 0,\, r_4\ne j_1-j_2}}^{p-(T+1)} \left|\sigma_{|r_4|}\sigma_{|r_4-j_2+j_1|}\right| = o(1).
\end{aligned}$$

Finally, when all indices are pairwise distinct. We use the same arguments as previously, and we get,

$$\begin{aligned}
\mathcal{G}_4 &:= 4 \sum_{1\le j_1\ne j_2\ne j_3\ne j_4<T} w_{j_1}^* w_{j_2}^* w_{j_3}^* w_{j_4}^* \sum_{-p+T+1\le r_1,r_2,r_3,r_4\le p-(T+1)} \left|\sigma_{|r_1|}\sigma_{|r_1-j_1+j_2|}\sigma_{|r_2|}\sigma_{|r_2-j_3+j_4|}\right| \\
&\qquad \cdot\left|\sigma_{|r_3|}\sigma_{|r_3-j_1+j_3|}\sigma_{|r_4|}\sigma_{|r_4-j_2+j_4|}\right| \\
&= 4 \sum_{1\le j_1\ne j_2\ne j_3\ne j_4<T} w_{j_1}^* w_{j_2}^* w_{j_3}^* w_{j_4}^* \Big(2\,\sigma_0 \left|\sigma_{|j_1-j_2|}\right| + \sum_{r_1\ne 0,\, r_1\ne j_1-j_2} \left|\sigma_{|r_1|}\sigma_{|r_1-j_1+j_2|}\right| \Big) \\
&\quad \Big(2\,\sigma_0 \left|\sigma_{|j_3-j_4|}\right| + \sum_{r_2\ne 0,\, r_2\ne j_3-j_4} \left|\sigma_{|r_2|}\sigma_{|r_2-j_3+j_4|}\right| \Big)\Big(2\,\sigma_0 \left|\sigma_{|j_1-j_3|}\right| + \sum_{r_3\ne 0,\, r_3\ne j_1-j_3} \left|\sigma_{|r_3|}\sigma_{|r_3-j_1+j_3|}\right| \Big) \\
&\quad \Big(2\,\sigma_0 \left|\sigma_{|j_2-j_4|}\right| + \sum_{r_4\ne 0,\, r_4\ne j_2-j_4} \left|\sigma_{|r_4|}\sigma_{|r_4-j_2+j_4|}\right| \Big)
\end{aligned}$$

Now, we treat each term of $\mathcal{G}_4$ separately:

$$\begin{aligned}
\mathcal{G}_{4,1} &:= \sum_{1\le j_1\ne j_2\ne j_3\ne j_4<T} w_{j_1}^* w_{j_2}^* w_{j_3}^* w_{j_4}^* \left|\sigma_{|j_1-j_2|}\sigma_{|j_3-j_4|}\sigma_{|j_1-j_3|}\sigma_{|j_2-j_4|}\right| \\
&\le \cdot\Big(\sum_{j_1,j_2} w_{j_1}^* w_{j_2}^* \sigma_{|j_1-j_2|}^2\Big)^{\frac{1}{2}} \Big(\sum_{j_1,j_3} w_{j_1}^* w_{j_3}^* \sigma_{|j_1-j_3|}^2\Big)^{\frac{1}{2}} \Big(\sum_{j_2,j_4} w_{j_2}^* w_{j_4}^* \sigma_{|j_2-j_4|}^2\Big)^{\frac{1}{2}} \Big(\sum_{j_3,j_4} w_{j_3}^* w_{j_4}^* \sigma_{|j_3-j_4|}^2\Big)^{\frac{1}{2}} \\
&\le \cdot\Big(\sum_{j_1} w_{j_1}^* \sum_{\substack{j_2\\ j_2\le|j_1-j_2|}} w_{|j_1-j_2|}^* \sigma_{|j_1-j_2|}^2 + (\sup_j w_j^*) \cdot \sum_{j_1} \sum_{\substack{j_2\\ j_2>|j_1-j_2|}} \frac{|j_1-j_2|^{2\alpha}}{j_2^{2\alpha}} \sigma_{|j_1-j_2|}^2 \Big)^2 \\
&\le \Big(O(\sqrt{T}) \cdot \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) + O\Big(\frac{1}{\sqrt{T}}\Big) \cdot \max\{1, T^{-2\alpha+1}\} \cdot L\Big)^2 = o(1)
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{G}_{4,2} &:= \sum_{1\le j_1\ne j_2\ne j_3\ne j_4<T} w^*_{j_1}w^*_{j_2}w^*_{j_3}w^*_{j_4}|\sigma_{|j_1-j_2|}\sigma_{|j_3-j_4|}\sigma_{|j_1-j_3|}| \sum_{r_4\ne 0,\, r_4\ne j_2-j_4} |\sigma_{|r_4|}\sigma_{|r_4-j_2+j_4|}| \\
&\le \sum_{j_2,j_4}\sum w^*_{j_2}w^*_{j_4}\Big(\sum_{j_1}\sum_{j_3} w^*_{j_1}w^*_{j_3}\sigma^2_{|j_1-j_3|}\Big)^{\frac12}\Big(\sum_{j_1} w^*_{j_1}\sigma^2_{|j_1-j_2|}\Big)^{\frac12}\Big(\sum_{j_3} w^*_{j_3}\sigma^2_{|j_3-j_4|}\Big)^{\frac12} \\
&\quad \cdot\Big(\sum_{\substack{r_4\ne 0,\, r_4\ne j_2-j_4\\ |r_4|\le j_2}} \sigma^2_{|r_4|} + \sum_{\substack{r_4\ne 0,\, r_4\ne j_2-j_4\\ |r_4|\le j_2}} \sigma^2_{|r_4|}\Big) \\
&\le \Big(\sum_{j_1}\sum_{j_3} w^*_{j_1}w^*_{j_3}\sigma^2_{|j_1-j_3|}\Big)^{\frac12}\Big((\sup_j w^*_j)\cdot L\Big) \\
&\quad \cdot\Big(\sum_{j_2,j_4}\sum w^*_{j_4}\sum_{\substack{r_4\ne 0\\ |r_4|\le j_2}} w_{j_2}\sigma^2_{|r_4|} + \sum_{j_2,j_4}\sum w^*_{j_2}w^*_{j_4}\sum_{\substack{r_4\ne 0\\ |r_4|\le j_2}} \frac{|r_4|^{2\alpha}}{j_2^{2\alpha}}\sigma^2_{|r_4|}\Big) \\
&\le \Big(O(\sqrt{T})\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) + O\Big(\frac{1}{\sqrt{T}}\Big)\cdot\max\{1,T^{-2\alpha+1}\}\Big)^{\frac12}\cdot O\Big(\frac{1}{\sqrt{T}}\Big) \\
&\quad \cdot\Big(T^2\cdot(\sup_j w^*_j)\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) + T\cdot(\sup_j w^*_j)^2\cdot\max\{1,T^{-2\alpha+1}\cdot L\}\Big) \\
&\le \Big(O(\sqrt{T})\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) + O\Big(\frac{1}{\sqrt{T}}\Big)\cdot\max\{1,T^{-2\alpha+1}\}\cdot 2L\Big)^{\frac12} \\
&\quad \cdot\Big(T\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) + O\Big(\frac{1}{\sqrt{T}}\Big)\cdot\max\{1,T^{-2\alpha+1}\}\cdot L\Big) \\
&= o(1) \quad \text{since } \mathbb{E}(\widehat{\mathcal{A}}_n)\asymp 1/np \text{ and for all } \alpha>1/4.
\end{aligned}$$

We use similar argument as previously to show that the remaining terms in $\mathcal{G}_4$ tend to zero. To complete the proof, we need to verify that,

$$\mathbb{E}_\Sigma(H_n^4(X_1,X_2))/\mathbb{E}^2_\Sigma(H_n^2(X_1,X_2)) = o(n). \tag{2.49}$$

We write

$$\begin{aligned}
&\frac{\mathbb{E}_\Sigma(H_n^4(X_1,X_2))}{\mathbb{E}^2_\Sigma(H_n^2(X_1,X_2))} = \frac{1}{(p-T)^4}\sum_{j_1,j_2,j_3,j_4} w^*_{j_1}w^*_{j_2}w^*_{j_3}w^*_{j_4}\sum_{T+1\le i_1,i_3,i_5,i_7\le p}\ \sum_{T+1\le i_2,i_4,i_6,i_8\le p} \\
&\mathbb{E}_\Sigma[(X_{1,i_1}X_{1,i_1-j_1}-\sigma_{j_1})(X_{1,i_3}X_{1,i_3-j_2}-\sigma_{j_2})(X_{1,i_5}X_{1,i_5-j_3}-\sigma_{j_3})(X_{1,i_7}X_{1,i_7-j_4}-\sigma_{j_4})] \\
&\cdot\mathbb{E}_\Sigma[(X_{2,i_2}X_{2,i_2-j_1}-\sigma_{j_1})(X_{2,i_4}X_{2,i_4-j_2}-\sigma_{j_2})(X_{2,i_6}X_{2,i_6-j_3}-\sigma_{j_3})(X_{2,i_8}X_{2,i_8-j_4}-\sigma_{j_4})]
\end{aligned}$$

To bound from above the previous sum, we replace the expected value by it's value, which is a sum of many terms, that are all treated similarly. So let us give an upper bound for the following one :

$$\begin{aligned}
\mathcal{H} &:= \frac{1}{(p-T)^4}\sum_{j_1,j_2,j_3,j_4} w^*_{j_1}w^*_{j_2}w^*_{j_3}w^*_{j_4}\sum_{T+1\le i_1,i_3,i_5,i_7\le p}\ \sum_{T+1\le i_2,i_4,i_6,i_8\le p}\sigma_{|i_1-i_3|}\sigma_{|i_1-i_3-j_1+j_2|} \\
&\qquad \cdot\sigma_{|i_5-i_7|}\sigma_{|i_5-i_7-j_3+j_4|}\sigma_{|i_2-i_4|}\sigma_{|i_2-i_4-j_1+j_2|}\sigma_{|i_6-i_8|}\sigma_{|i_6-i_8-j_3+j_4|} \\
&\le \sum_{j_1,j_2,j_3,j_4} w^*_{j_1}w^*_{j_2}w^*_{j_3}w^*_{j_4}\sum_{-p+1\le r_1,r_2,r_3,r_4\le p-1}\sigma_{|r_1|}\sigma_{|r_1-j_1+j_2|}\sigma_{|r_2|}\sigma_{|r_2-j_1+j_2|} \\
&\qquad \cdot\sigma_{|r_3|}\sigma_{|r_3-j_3+j_4|}\sigma_{|r_4|}\sigma_{|r_4-j_3+j_4|}
\end{aligned}$$

We see that $\mathcal{H}$ can be treated in the same way as $\mathcal{G}$. However, we show that $\mathcal{H} = O(1) = o(n)$. Let us deal with one of the terms of $\mathcal{H}$, consider the term for which we have $j_1 = j_2$, $j_3 = j_4$, and $j_1 \neq j_3$ thus we get

$$\sum_{1\le j_1\neq j_3<T} w_{j_1}^{*2} w_{j_3}^{*2} \sum_{-p+1\le r_1,r_2,r_3,r_4\le p-1} \sigma_{|r_1|}^2\sigma_{|r_2|}^2\sigma_{|r_3|}^2\sigma_{|r_4|}^2 = \sum_{1\le j_1\neq j_3<T} w_{j_1}^{*2} w_{j_3}^{*2}\cdot\Big(\sigma_0^2 + \sum_{r_1\neq 0}\sigma_{|r_1|}^2\Big)^2$$
$$\le\ 2\sum\sum_{1\le j_1\neq j_3<T} w_{j_1}^{*2} w_{j_3}^{*2} + 2\sum\sum_{1\le j_1\neq j_3<T} w_{j_1}^{*2} w_{j_3}^{*2}\Big(\sum_{r_1\neq 0}\sigma_{|r_1|}^2\Big)^2.$$

It is easily seen that $\sum\sum_{1\le j_1\neq j_2<T} w_{j_1}^{*2} w_{j_2}^{*2} = O(1)$. And so on, we show that all terms in $\mathcal{H}$ are $O(1)$ and thus we get the desired result. Together with (2.46), this proves (2.43). In consequence, we apply theorem 1 of [50], to get (2.42). □

Proof of (2.44). We define $\widehat{\mathcal{B}}_{n,p}$ as follows,

$$\widehat{\mathcal{B}}_{n,p} = \frac{2}{\sqrt{n(n-1)(p-T)(p-T-1)}}\sum_{i=T+1}^{p}\sum_{h=i+1}^{p}\sum_{1\le k\neq l\le n}\sum_{j=1}^{T-1} w_j^*\ X_{k,i}X_{k,i-j}X_{l,h}X_{l,h-j}$$

We set

$$\begin{aligned} D_{n,p,i} &= \frac{2}{\sqrt{n(n-1)(p-T)(p-T-1)}}\sum_{h=i+1}^{p}\sum_{1\le k\neq l\le n}\sum_{j=1}^{T-1} w_j^*\ X_{k,i}X_{k,i-j}X_{l,h}X_{l,h-j} \\ &:= c(n,p,T)\sum_{h=i+1}^{p}\sum_{1\le k\neq l\le n}\sum_{j=1}^{T-1} w_j^*\ X_{k,i}X_{k,i-j}X_{l,h}X_{l,h-j} \end{aligned}$$

Note that the $\{D_{n,p,i}\}_{T+1\le i\le p}$ is a sequence of martingale differences with respect to the sequence of σ fields $\{\mathcal{F}_i, i\ge T+1\}$ such that $\mathcal{F}_i = \sigma\{X_{.,r}\ , r\le i\}$, we denote by $\mathbb{E}_i(\cdot) = \mathbb{E}(\cdot/\mathcal{F}_i)$, where $\mathbb{E}$ is the expected value under the null hypothesis. Indeed, for all $T+1\le i\le p$, we have, $\mathbb{E}_{i-1}(D_{n,p,i}) = 0$. We use sufficient conditions to show the asymptotic normality of a sum of martingale differences $\widehat{\mathcal{B}}_{n,p}$ for all $n\ge 2$, as $(p-T)\to\infty$, see e.g. [83]. Thus it suffices to show that,

$$\mathbb{E}\Big(\sum_{i=T+1}^{p}\mathbb{E}_{i-1}(D_{n,p,i}^2)-1\Big)^2\to 0 \quad\text{and}\quad \sum_{i=T+1}^{p}\mathbb{E}(D_{n,p,i}^4)\to 0. \tag{2.50}$$

We first show the first part of (2.50).

$$\begin{aligned} \mathbb{E}_{i-1}(D_{n,p,i}^2) &= (c(n,p,T))^2\sum_{h=i+1}^{p}\sum_{1\le k\neq l\le n}\sum_{1\le j,j_1\le<T-1} w_j w_{j_1}^* X_{k,i-j}X_{k,i-j_1}\mathbb{E}_{i-1}(X_{l,h-j}X_{l,h-j_1}) \\ &= (c(n,p,T))^2\cdot\Big(\sum_{1\le j,j_1\le<T-1}\sum_{1\le k\neq l\le n} w_j w_{j_1}^* X_{k,i-j}X_{k,i-j_1}\sum_{h=i+1}^{(i+j_1-1)\wedge(i+j-1)} X_{l,h-j}X_{l,h-j_1} \\ &\quad + (n-1)\sum_{k=1}^{n}\sum_{j=1}^{T} w_j^{*2}X_{k,i-j}^2(p-i-j+1)\Big) \end{aligned}$$

giving

$$
\begin{aligned}
& \mathbb{E}(\sum_{i=T+1}^{p} \mathbb{E}_{i-1}(D_{n,p,i}^2)) \\
= \; & c^2(n,p,T) \cdot \Big(n(n-1) \sum_{i=T+1}^{p} \sum_{j=1}^{T-1} w_j^{*2}(j-1) + n(n-1) \sum_{i=T+1}^{p} \sum_{j=1}^{T-1} w_j^{*2}(p-i-j+1) \Big) \\
= \; & \frac{4}{(p-T)(p-T-1)} \sum_{j=1}^{T-1} w_j^{*2} \sum_{i=T+1}^{p} (p-i) = 1.
\end{aligned}
$$

Thus, to show that $\mathbb{E}\Big(\sum_{i=T+1}^{p} \mathbb{E}_{i-1}(D_{n,p,i}^2) - 1 \Big)^2 \to 0$, it is sufficient to show that $\mathbb{E}\Big(\sum_{i=T+1}^{p} \mathbb{E}_{i-1}(D_{n,p,i}^2) \Big)^2 = 1 + o(1)$. Indeed,

$$
\mathbb{E}\Big(\sum_{i=T+1}^{p} \mathbb{E}_{i-1}(D_{n,p,i}^2) \Big)^2 = (c(n,p,T))^4 \cdot \Big(E_1 + E_2 + E_3 + E_4 \Big). \tag{2.51}
$$

where E_1, E_2 , E_3 and E_4 are given by the following.

$$
\begin{aligned}
E_1 \; = \; & \sum_{T+1 \le i,i' \le p} \sum_{1 \le k \ne l \le n} \sum_{1 \le k' \ne l' \le n} \sum_{1 \le j,j_1 \le T} \sum_{1 \le j',j_1' \le T-1} \sum_{h=i+1}^{(i+j-1)\wedge(i+j_1-1)} \sum_{h'=i'+1}^{(i'+j'-1)\wedge(i'+j_1'-1)} \\
& w_j^* w_{j_1}^* w_{j'}^* w_{j_1'}^* \, \mathbb{E}(X_{k,i-j} X_{k,i-j_1} X_{l,h-j} X_{l,h-j_1} X_{k',i'-j'} X_{k',i'-j_1'} X_{l',h'-j'} X_{l',h'-j_1'})
\end{aligned}
$$

Now we decompose E_1 into five sums that depends on the indices k, k', l and l'. We begin by the first case when $k = k'$ and $l = l'$,

$$
\begin{aligned}
E_{1,1} \; := \; & \sum_{1 \le k \ne l \le n} \Bigg(\sum_{i=T+1}^{p} \Big(\sum_{j=1}^{T-1} w_j^{*4} \, 3 \cdot (3(j-1) + (j-1)(j-2)) + \sum_{1 \le j \ne j' \le T} w_j^{*2} w_{j'}^{*2} \, (j-1)(j'-1) \\
& \qquad + 2 \sum_{1 \le j \ne j_1 \le T-1} w_j^{*2} w_{j_1}^{*2} \Big((j-1) \wedge (j_1 - 1) \Big) \\
+ \; & \sum_{T+1 \le i \ne i' \le p} \Big(\sum_{j=1}^{T-1} w_j^{*4} \, (3(j-1) + (j-1)(j-2)) + \sum_{1 \le j \ne j' \le T} w_j^{*2} w_{j'}^{*2} \, (j-1)(j'-1) \Bigg) \\
= \; & n(n-1) \Bigg(2 \sum_{i=T+1}^{p} \sum_{j=1}^{T-1} w_j^{*4} (j-1)(j+1) + 2 \sum_{T+1 \le i,i' \le p} \sum_{j=1}^{T-1} w_j^{*4} \, (j-1) \\
+ \; & \sum_{T+1 \le i,i' \le p} \sum_{1 \le j,j' \le T} w_j^{*2} w_{j'}^{*2} \, (j-1)(j'-1) + 2 \sum_{i=T+1}^{p} \sum_{1 \le j \ne j' \le T-1} w_j^{*2} w_{j'}^{*2} \Big((j-1) \wedge (j'-1) \Big) \Bigg)
\end{aligned}
$$

When $k = l'$ and $l = k'$, we have using similar arguments as previously that,

$$
E_{1,2} \; = \; n(n-1) \sum_{T+1 \le i,i' \le p} \sum_{1 \le j,j' \le T} w_j^{*2} w_{j'}^{*2} \, (j-1)(j'-1)
$$

We move to the term, when $k = k'$ and $l \neq l'$,

$$
\begin{aligned}
E_{1,3} &:= \sum_{\substack{1 \le k,l,l' \le n \\ k \neq l,l', l \neq l'}} \Big\{ \sum_{j=1}^{T-1} w_j^{*4} \Big(\sum_{i=T+1}^{p} 3\,(j-1)^2 + \sum_{T+1 \le i \neq i' \le p} (j-1)^2 \Big) \\
&+ \sum_{1 \le j \neq j' \le T} w_j^{*2} w_{j'}^{*2} \sum_{T+1 \le i,i' \le p} (j-1)(j'-1) \Big\} \\
&= n(n-1)(n-2) \Big(2 \sum_{i=T+1}^{p} \sum_{j=1}^{T-1} w_j^{*4} (j-1)^2 + \sum_{T+1 \le i,i' \le p} \sum_{1 \le j,j' \le T-1} w_j^{*2} w_{j'}^{*2} (j-1)(j'-1) \Big)
\end{aligned}
$$

Now we treat the case when , $k \neq k'$ and $l = l'$,

$$
\begin{aligned}
E_{1,4} &:= \sum_{\substack{1 \le k,k',l \le n \\ l \neq k,k', k \neq k'}} \Big\{ \sum_{j=1}^{T-1} w_j^{*4} \sum_{T+1 \le i,i' \le p} (3(j-1) + (j-1)(j-2)) \\
&+ \sum_{1 \le j \neq j' \le T} \sum_{T+1 \le i,i' \le p} w_j^{*2} w_{j'}^{*2} \cdot (j-1)(j'-1) \Big\} \\
&= n(n-1)(n-2) \sum_{T+1 \le i,i' \le p} \Big\{ \sum_{j=1}^{T-1} w_j^{*4} \,(j-1)(j+1) + \sum_{1 \le j \neq j' \le T} w_j^{*2} w_{j'}^{*2} \cdot (j-1)(j'-1) \Big\} \\
&= n(n-1)(n-2) \sum_{T+1 \le i,i' \le p} \Big(\sum_{1 \le j,j' \le T-1} w_j^{*2} w_{j'}^{*2} \,(j-1)(j'-1) + 2 \sum_{j=1}^{T-1} w_j^{*4} \,(j-1) \Big)
\end{aligned}
$$

Finally, we treat the term for $k \neq k'$ and $l \neq l'$,

$$
\begin{aligned}
E_{1,5} &:= \sum_{\substack{1 \le k \neq l \le n \\ k \neq k', l \neq l'}} \sum_{1 \le k' \neq l' \le n} \sum_{T+1 \le i,i' \le p} \sum_{1 \le j,j' \le T} w_j^{*2} w_{j'}^{*2} \,(j-1)(j'-1) \\
&= n(n-1)^2(n-2) \sum_{T+1 \le i,i' \le p} \sum_{1 \le j,j' \le T} w_j^{*2} w_{j'}^{*2} \,(j-1)(j'-1)
\end{aligned}
$$

We group the previous result to get,

$$
\begin{aligned}
E_1 &= \Big(2n(n-1) + 2n(n-1)(n-2) + n(n-1)^2(n-2) \Big) \sum_{T+1 \le i,i' \le p} \sum_{1 \le j,j' \le T} w_j^{*2} w_{j'}^{*2} \,(j-1)(j'-1) \\
&+ R_1(n,p,T)
\end{aligned}
$$

where,

$$
\begin{aligned}
R_1(n,p,T) &= 2\Big(n(n-1) + n(n-1)(n-2) \Big) \sum_{T+1 \le i,i' \le p} \sum_{j=1}^{T-1} w_j^{*4} (j-1) \\
&+ 2n(n-1) \sum_{i=T+1}^{p} \sum_{j=1}^{T-1} w_j^{*4} (j-1)(j+1) + 2n(n-1)(n-2) \sum_{i=T+1}^{p} \sum_{j=1}^{T-1} w_j^{*4} (j-1)^2 \\
&+ 2n(n-1) \sum_{i=T+1}^{p} \sum_{1 \le j \neq j' \le T-1} w_j^{*2} w_{j'}^{*2} \Big((j-1) \wedge (j'-1) \Big) \\
&= o((c(n,p,T)^{-4})
\end{aligned}
$$

Now, let us bound from above the term E_2 in (2.51):

$$
\begin{aligned}
E_2 &:= (n-1)\sum_{T+1\le i,i'\le p}\sum_{1\le k\ne l\le n}\sum_{1\le j,j_1\le T}\sum_{k'=1}^{n}\sum_{j'=1}^{T} w_j^* w_{j_1}^* w_{j'}^{*2}\mathbb{E}(X_{k,i-j}X_{k,i-j_1}X_{k',i'-j'}^2)\\
&\cdot(p-i'-j'+1)\sum_{h=i+1}^{(i+j-1)\wedge(i+j_1-1)}\mathbb{E}(X_{l,h-j}X_{l,h-j_1})
\end{aligned}
$$

We treat the two cases $k=k'$ and $k\ne k'$ each one apart. We begin by the case when $k\ne k'$,

$$
\begin{aligned}
E_{2,1} &:= n(n-1)^2\Big\{\sum_{j=1}^{T-1} w_j^{*4}\Big(\sum_{i=T+1}^{p}\cdot 3(p-i-j+1)(j-1)+\sum_{T+1\le i\ne i'\le p}(p-i'-j+1)(j-1)\Big)\\
&+\sum_{1\le j\ne j'\le T} w_j^{*2}w_{j'}^{*2}\Big(\sum_{i=T+1}^{p}(p-i-j'+1)(j-1)+\sum_{T+1\le i\ne i'\le p}(p-i'-j'+1)(j-1)\Big)\\
&= n(n-1)^2\Big\{2\sum_{i=T+1}^{p}\sum_{j=1}^{T-1} w_j^{*4}\,(p-i-j+1)(j-1)\\
&+\sum_{T+1\le i,i'\le p}\sum_{1\le j,j'\le T-1} w_j^{*2}w_{j'}^{*2}(p-i'-j'+1)(j-1)\Big\}
\end{aligned}
$$

When $k\ne k'$,

$$
E_{2,2} := n(n-1)^3\sum_{T+1\le i,i'\le p}\sum_{1\le j,j'\le T} w_j^{*2}w_{j'}^{*2}(p-i'-j'+1)(j-1).
$$

As consequence

$$
\begin{aligned}
E_2 &= \Big(n(n-1)^2+n(n-1)^3\Big)\sum_{T+1\le i,i'\le p}\sum_{1\le j,j'\le T-1} w_j^{*2}w_{j'}^{*2}(p-i'-j'+1)(j-1)\\
&+ o((c(n,p,T)^{-4}).
\end{aligned}
$$

Similarly we get,

$$
\begin{aligned}
E_3 &= (n-1)\sum_{T+1\le i,i'\le p}\sum_{1\le k'\ne l'\le n}\sum_{1\le j',j_1'\le T-1}\sum_{k=1}^{n}\sum_{j=1}^{T-1} w_{j'}^* w_{j_1'}^* w_j^{*2}\mathbb{E}(X_{k',i'-j'}X_{k',i'-j_1'}X_{k,i-j}^2)\\
&\cdot(p-i-j+1)\sum_{h'=i'+1}^{(i'+j'-1)\wedge(i'+j_1'-1)}\mathbb{E}(X_{l',h'-j'}X_{l',h'-j_1'})\\
&= n^2(n-1)^2\sum_{T+1\le i,i'\le p}\sum_{1\le j,j'\le T-1} w_j^{*2}w_{j'}^{*2}(p-i-j+1)(j'-1)\\
&+ 2n(n-1)^2\sum_{i=T+1}^{p}\sum_{j=1}^{T-1} w_j^{*4}\,(p-i-j+1)(j-1)
\end{aligned}
$$

The term E_4 of (2.51) is treated as follows,

$$
\begin{aligned}
E_4 &:= (n-1)^2 \sum_{T+1\le i,i'\le p}\sum_{1\le k,k'\le n}\sum_{1\le j,j'\le T} w_j^{*2} w_{j'}^{*2}\mathbb{E}(X_{k,i-j}^2 X_{k',i'-j'}^2)\\
&\qquad\qquad \cdot(p-i-j+1)(p-i'-j'+1)\\
&= n(n-1)^2\Big\{\sum_{i=T+1}^{p}\Big(\sum_{1\le j\le T} w_j^{*4} 3(p-i-j+1)^2\\
&+ \sum_{1\le j\ne j'\le T} w_j^{*2} w_{j'}^{*2}(p-i-j+1)(p-i-j'+1)\Big)\\
&+ \sum_{T+1\le i\ne i'\le p}\sum_{1\le j,j'\le T-1} w_j^{*2} w_{j'}^{*2}(p-i-j+1)(p-i'-j'+1)\Big\}\\
&+ n(n-1)^3 \sum_{T+1\le i,i'\le p}\sum_{1\le j,j'\le T-1} w_j^{*2} w_{j'}^{*2}(p-i-j+1)(p-i'-j'+1)\\
&= n^2(n-1)^2 \sum_{T+1\le i,i'\le p}\sum_{1\le j,j'\le T-1} w_j^{*2} w_{j'}^{*2}(p-i-j+1)(p-i'-j'+1)\\
&\quad +2n(n-1)^2\sum_{i=T+1}^{p}\sum_{1\le j\le T} w_j^{*4}(p-i-j+1)^2
\end{aligned}
$$

Finally we group all the previous terms and obtain,

$$
\begin{aligned}
&\mathbb{E}\Big(\sum_{i=T+1}^{p}\mathbb{E}_{i-1}(D_{n,p,i}^2)\Big)^2\\
&= c^4(n,p,T)\Big\{\cdot n^2(n-1)^2\sum_{T+1\le i,i'\le p}\sum_{1\le j,j'\le T-1} w_j^{*2} w_{j'}^{*2}\Big((j-1)(j'-1)\\
&+ (p-i'-j'+1)(j-1)+(p-i-j+1)(j'-1)\\
&+ (p-i-j+1)(p-i'-j'+1)\Big)+o((c(n,p,T))^{-4})\Big\}\\
&= \frac{16}{(p-T)^2(p-T-1)^2}\sum_{1\le j,j'\le T-1} w_j^{*2} w_{j'}^{*2}\sum_{T+1\le i,i'\le p}(p-i)(p-i'))+o(1).\\
&= \frac{16}{(p-T)^2(p-T-1)^2}\cdot\frac{1}{4}\cdot\Big(\frac{(p-T-1)(p-T)}{2}\Big)^2+o(1)=1+o(1)
\end{aligned}
$$

To achieve the proof, we show that the second condition given in (2.50) is also verified. Indeed,

$$
\begin{aligned}
&\sum_{i=T+1}^{p}\mathbb{E}(D_{n,p,i}^4)\\
&= (c(n,p,T))^4\sum_{i=T+1}^{p}\sum_{i+1\le h_1,h_2,h_3,h_4\le p}\sum_{1\le k_1\ne l_1\le n}\sum_{1\le k_2\ne l_2\le n}\sum_{1\le k_3\ne l_3\le n}\sum_{1\le k_4\ne l_4\le n}\\
&\sum_{1\le j_1,j_2,j_3,j_4\le T-1} w_{j_1}^* w_{j_2}^* w_{j_3}^* w_{j_4}^*\mathbb{E}(X_{k_1,i}X_{k_2,i}X_{k_3,i}X_{k_4,i}X_{l_1,h_1-j_1}X_{l_2,h_2-j_2}X_{l_3,h_3-j_3}X_{l_4,h_4-j_4})\\
&\qquad\cdot\mathbb{E}(X_{l_1,h_1}X_{l_2,h_2}X_{l_3,h_3}X_{l_4,h_4}X_{k_1,i-j_1}X_{k_2,i-j_2}X_{k_3,i-j_3}X_{k_4,i-j_4})\\
&= O(1)\cdot(c(n,p,T))^4\cdot\sum_{i=T+2}^{p}\sum_{T+1\le h_1,h_2\le p}\sum_{1\le k_1\ne l_1\le n}\sum_{1\le k_2\ne l_2\le n}\sum_{1\le j_1,j_2\le T} w_{j_1}^{*2} w_{j_2}^{*2}\\
&= \frac{O(1)}{(p-T)^2(p-T-1)^2}\cdot(p-T)^3=o(1).
\end{aligned}
$$

$\square$

2.6.2 Proofs of results in Section 2.3

Proof of Proposition 2.5. To show the upper bound for the variance of $\widehat{\mathcal{A}}_n^{\mathcal{E}}$, we follow the line of proof of Proposition 2.1. We use that $\sum_{j\geq 1} 1/(e^{nAj}) = 1/(e^{nA}-1)$ for all $A>0$ and n finite integer. As an example, let us bound from above one term of the variance of $\widehat{\mathcal{A}}_n^{\mathcal{E}}$:

$$
\begin{aligned}
R_{1,2,2} &:= \sum_{1\leq j\neq j'\leq T} w_j^* w_{j'}^* \Big(|\sigma_{|j'-j|}| + \sum_{\substack{r=-p+T+1\\ r\neq 0}}^{p-(T+1)} |\sigma_{|r|}\sigma_{|r-j+j'|}| \Big)^2 \\
&\leq 2 \sum_{1\leq j\neq j'\leq T} w_j^* w_{j'}^* \sigma_{|j'-j|}^2 + 4 \sum_{1\leq j\neq j'\leq T} w_j^* w_{j'}^* (\sum_{r=1}^{p-(T+1)} \sigma_r^2)(\sum_{\substack{r=-p+T+1\\ r\neq 0}}^{p-(T+1)} \sigma_{|r-j+j'|}^2) \\
&\leq 2 \sum_{\substack{1\leq j\neq j'\leq T\\ |j'-j|<j}} w_{j'}^* w_{|j'-j|}^* \sigma_{|j'-j|}^2 + 2 \sum_{\substack{1\leq j\neq j'\leq T\\ |j'-j|>j}} w_j^* w_{j'}^* \frac{e^{2A|j'-j|}}{e^{2Aj}} \sigma_{|j'-j|}^2 + 4 \sum_{1\leq j\neq j'\leq T} \Big(\sum_{r=1}^{j} w^* r \sigma_r^2 \\
&+ w_j^* \sum_{r=j+1}^{p-(T+1)} \frac{e^{2Ar}}{e^{2Aj}} \sigma_r^2 \Big) \Big(\sum_{\substack{r=-p+T+1\\ |r-j+j'|<j'}}^{p-(T+1)} w_{|r-j+j'|}^* \sigma_{|r-j+j'|}^2 + w_{j'}^* \sum_{\substack{r=-p+T+1\\ |r-j+j'|\geq j'}}^{p-(T+1)} \frac{e^{2A|r-j+j'|}}{e^{2Aj'}} \sigma_{|r-j+j'|}^2 \Big) \\
&\leq 4 \cdot (\sup_j w_j^*) \cdot T \cdot \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n^{\mathcal{E}}) + 4L \cdot (\sup_j w_j^*)^2 \cdot (1/(e^{2A}-1)) + O(T^2) \cdot \mathbb{E}_\Sigma^2(\widehat{\mathcal{A}}_n^{\mathcal{E}}) \\
&+ 16L \cdot (\sup_j w_j^*) \cdot T \cdot (1/(e^{2A}-1)) \cdot \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n^{\mathcal{E}}) + 16L^2 \cdot (\sup_j w_j^*)^2 \cdot (1/(e^{2A}-1))^2.
\end{aligned}
$$

The proof of the asymptotic normality of $n(p-T)(\widehat{\mathcal{A}}_n^{\mathcal{E}} - \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n^{\mathcal{E}}))$, when $n(p-T)b(\psi) \asymp 1$ and for $\Sigma \in G(\mathcal{E}(A,L),\psi)$ such that $\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n^{\mathcal{E}}) = O(b(\psi))$, is also due to Theorem 1 of [50]. That is, we have to check (2.43) as in Proposition 2.2. As an example, let us bound from above the term $\mathcal{G}_2$ in (2.49) with the parameters given in (2.16):

$$
\begin{aligned}
\mathcal{G}_2 &:= 4 \sum_{1\leq j_1\neq j_2<T} w_{j_1}^{*2} w_{j_2}^{*2} \sum_{-p+T+1\leq r_1,r_2,r_3,r_4\leq p-(T+1)} |\sigma_{|r_1|}\sigma_{|r_1-j_1+j_2|}\sigma_{|r_2|}\sigma_{|r_2-j_2+j_1|}| \\
&\qquad \cdot |\sigma_{|r_3|}\sigma_{|r_3-j_1+j_2|}\sigma_{|r_4|}\sigma_{|r_4-j_2+j_1|}| \\
&\leq 4 \sum_{1\leq j_1\neq j_2<T} w_{j_1}^{*2} w_{j_2}^{*2} (\sum_{r_1} \sigma_{|r_1|}^2)^2 (\sum_{r_1} \sigma_{|r_1-j_1+j_2|}^2)(\sum_{r_2} \sigma_{|r_2-j_2+j_1|}^2) \\
&\leq 16L^2 \sum_{1\leq j_1\neq j_2<T} w_{j_1}^{*2} w_{j_2}^{*2} (\sum_{\substack{r_1\\ |r_1|\leq j_1}} \sigma_{|r_1|}^2 + \sum_{\substack{r_1\\ |r_1|>j_1}} \sigma_{|r_1|}^2)^2 \\
&\leq 16L^2 \Big\{ \sum_{1\leq j_1\neq j_2<T} w_{j_2}^{*2} (\sum_{\substack{r_1\\ |r_1|\leq j_1}} w_{|r_1|}^* \sigma_{|r_1|}^2)^2 + \sum_{1\leq j_1\neq j_2<T} w_{j_1}^{*2} w_{j_2}^{*2} (\sum_{\substack{r_1\\ |r_1|>j_1}} \frac{e^{2Ar_1}}{e^{2Aj_1}} \sigma_{|r_1|}^2)^2 \Big\} \\
&\leq 16L^2 \Big\{ \sum_{j_1} (\sum_{j_2} w_{j_2}^{*2}) \cdot \mathbb{E}_\Sigma^2(\widehat{\mathcal{A}}_n) + 4L^2 (\sum_{j_2} w_{j_2}^*) \cdot (\sum_{j_1} w_{j_1}^{*2} \frac{1}{e^{2Aj}}) \Big\} \\
&\leq 16L^2 \Big\{ \frac{T}{2} \cdot \mathbb{E}_\Sigma^2(\widehat{\mathcal{A}}_n^{\mathcal{E}}) + 4L^2 \cdot \frac{1}{2} \cdot (\sup_j w_j^{*2}) \cdot \frac{1}{e^{2A}-1} \Big\} \\
&\leq \mathbb{E}_\Sigma^2(\widehat{\mathcal{A}}_n^{\mathcal{E}}) \cdot O(T) + o(1) = O(\frac{T}{n^2(p-T)^2}) + o(1) = o(1). \qquad (2.52)
\end{aligned}
$$

□

Proof of Theorem 2.6. To show the upper bound, we use first the asymptotic normality of the $n(p-T)\widehat{\mathcal{A}}_n^{\mathcal{E}}$ under H_0 to prove that the type I error probability of Δ^* : $\eta(\Delta^*) = 1 - \Phi(npb(\psi)) + o(1)$.

To bound from above the type II error probability, we shall distinguish 2 cases. First, when $n^2p^2b^2(\psi) \to +\infty$, we use the Markov inequality, (2.17) and (2.18), to show that $\beta(\Delta^*, G(\psi)) \to 0$. Then, when $n^2p^2b^2(\psi) \asymp 1$, we have two possibilities: either $\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n^{\mathcal{E}})/b(\psi)$ tends to infinity, or $\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n^{\mathcal{E}}) = O(b(\psi))$. We show respectively that either type II error probability tends to zero, or we use the asymptotic normality of $n(p-T)(\widehat{\mathcal{A}}_n^{\mathcal{E}} - \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n^{\mathcal{E}}))$ to get that $\beta(\Delta^*, G(\psi)) \le \Phi(np(t - b(\psi)) + o(1)$.

To show the lower bound, we follow the same sketch of proof of lower bounds of Theorems 2.3 and 2.4. The key point for ellipsoids $\mathcal{E}(A, L)$ is to check the positivity of the matrix

$$\Sigma^* = T_P(\{\sigma_j^*\}_{j\ge1}) \quad \text{where} \quad \sigma_j^* = \sqrt{\lambda}\Big(1 - (\frac{e^j}{e^T})^{2A}\Big)_+^{1/2} \quad \text{for all } j \ge 1.$$

Then we create a parametric family of matrices by changing the sign randomly on each diagonal of Σ^*, with parameters given in (2.16).

Lemme 2.8. *For $A > 0$, the symmetric Toeplitz matrix $\Sigma_U^* = T_p(\{u_j\sigma_j^*\}_{j\ge1})$, where $U = \{u_j\}_{j\ge0}$ with $u_0 = 1$, $u_j = \pm1$ for all $j \ge 1$, and σ_j^* defined as previously, is positive definite, for $\psi > 0$ small enough. Moreover, denote by $\lambda_{1,U}^*, \dots, \lambda_{p,U}^*$ the eigenvalues of Σ_U^*, then $|\lambda_{i,U}^* - 1| \le O(\psi \cdot \sqrt{\ln(1/\psi)})$, for all i from 1 to p.*

Proof of Lemma 2.8 . Using Gershgorin's Theorem we get that each eigenvalue of $\Sigma_U^* = T_p(\{u_j\sigma_j^*\}_{j\ge1})$ verifies, $|\lambda_{i,U}^* - u_0\sigma_0^*| \le 2\sum_{j\ge1}^{p} |u_j\sigma_j^*| = 2\sum_{j\ge1}\sigma_j^*$. We have,

$$\begin{aligned} \sum_{j\ge1}\sigma_j^* &= \sqrt{\lambda}\sum_{j\ge1}\Big(1 - (\frac{e^j}{e^T})^{2A}\Big)_+^{1/2} \le \sqrt{\lambda}\sum_{j=1}^{T}\Big(1 - (\frac{e^j}{e^T})^{2A}\Big)^{\frac{1}{2}} \\ &= O(1)\sqrt{\lambda}\cdot T \asymp \psi \cdot \sqrt{\ln(1/\psi)}. \end{aligned}$$

We deduce that the smallest eigenvalue is bounded from below by

$$\min_{i=1,\dots,p} \lambda_{i,U}^* \ge \sigma_0^* - 2\sum_{j\ge1}\sigma_j^* \ge 1 - O(1)\psi \cdot \sqrt{\ln(1/\psi)}.$$

which is strictly positive for $\psi > 0$ small enough. □

To complete the proof, we follow the steps of the proof of the lower bound in Section 2.2.2. □

Chapter 3

Sharp minimax tests for large covariance matrices and adaptation

Abstract.

We consider the detection problem of correlations in a p-dimensional Gaussian vector, when we observe n independent, identically distributed random vectors, for n and p large. We assume that the covariance matrix varies in some ellipsoid with parameter $\alpha > 1/2$ and total energy bounded by $L > 0$.

We propose a test procedure based on a U-statistic of order 2 which is weighted in an optimal way. The weights are the solution of an optimization problem, they are constant on each diagonal and non-null only for the T first diagonals, where $T = o(p)$. We show that this test statistic is asymptotically Gaussian distributed under the null hypothesis and also under the alternative hypothesis for matrices close to the detection boundary. We prove upper bounds for the total error probability of our test procedure, for $\alpha > 1/2$ and under the assumption $T = o(p)$ which implies that $n = o(p^{2\alpha})$. We illustrate via a numerical study the behavior of our test procedure.

Moreover, we prove lower bounds for the maximal type II error and the total error probabilities. Thus we obtain the asymptotic and the sharp asymptotically minimax separation rate $\widetilde{\varphi} = (C(\alpha, L)n^2 p)^{-\alpha/(4\alpha+1)}$, for $\alpha > 3/2$ and for $\alpha > 1$ together with the additional assumption $p = o(n^{4\alpha-1})$, respectively. We deduce rate asymptotic minimax results for testing the inverse of the covariance matrix.

We construct an adaptive test procedure with respect to the parameter α and show that it attains the rate $\widetilde{\psi} = (n^2 p/ \ln\ln(n\sqrt{p}))^{-\alpha/(4\alpha+1)}$.

3.1 Introduction

A large variety of applied fields collect and need to recover information from high-dimensional data. Among these we can cite communications and signal theory (functional magnetic resonance imaging, spectroscopic imaging), econometrics, climate studies, biology (gene expression micro-array) and finance (portfolio allocation). Testing large covariance matrix is an important problem and has recently been approached via several techniques: corrected likelihood ratio test using the theory of large random matrices, methods based on the sample covariance matrix and so on.

Let $X_1, \ldots, X_n$, be n independent and identically distributed p-vectors following a multivariate normal distribution $\mathcal{N}_p(0, \Sigma)$, where $\Sigma = [\sigma_{ij}]_{1 \le i,j \le p}$ is the normalized covariance matrix, with $\sigma_{ii} = 1$, for all $i = 1$ to p. Let us denote by $X_k = (X_{k,1}, \ldots, X_{k,p})^T$ for all $k = 1, \ldots, n$. In this paper we also assume that the size p of the vectors grows to infinity as well as the sample size n, $p \to \infty$ and $n \to \infty$.

We consider the following goodness-of-fit test, where we test the null hypothesis

$$H_0 : \Sigma = I, \quad \text{where } I \text{ is the } p \times p \text{ identity matrix} \tag{3.1}$$

against the composite alternative hypothesis

$$H_1 : \Sigma \in \mathcal{F}(\alpha, L), \quad \text{such that } \frac{1}{2p} \|\Sigma - I\|_F^2 \ge \varphi^2.$$

For any covariance matrix Σ, we recall that the Frobenius norm is computed as

$$\|\Sigma - I\|_F^2 = tr\big[(\Sigma - I)^2\big] = 2 \sum_{1 \le i < j \le p} \sigma_{ij}^2.$$

The class of matrices $\mathcal{F}(\alpha, L)$ is defined as follows, for $\alpha > 0$,

$$\mathcal{F}(\alpha, L) = \left\{ \Sigma \ge 0 \, ; \frac{1}{p} \sum_{1 \le i < j \le p} \sigma_{ij}^2 |i - j|^{2\alpha} \le L \text{ for all } p \text{ and } \sigma_{ii} = 1 \text{ for all } i = 1, \ldots, p \right\}$$

In order to test $H_0 : \Sigma = \Sigma_0$, for some given non negative definite covariance matrix Σ_0, we suggest rescaling the data $Z_i = \Sigma_0^{-1/2} X_i$ and then apply the same test procedure provided that $\Sigma_0^{-1/2} \Sigma \Sigma_0^{-1/2}$ belongs to $\mathcal{F}(\alpha, L)$. Let us denote by

$$Q(\alpha, L, \varphi) = \left\{ \Sigma \in \mathcal{F}(\alpha, L) \; ; \; \frac{1}{p} \sum_{1 \le i < j \le p} \sigma_{ij}^2 \ge \varphi^2 \right\}, \tag{3.2}$$

where $\varphi = \varphi_{n,p}(\alpha, L)$ is related to n and p, but also to α and L assumed fixed. The set of covariance matrices under the alternative hypothesis consists of matrices of size $p \times p$, whose elements decrease polynomially when moving away from the diagonal. This assumption is natural for covariances matrices and has been considered for estimation problems, see e.g [9], [26]. Regularization techniques, originally used for nonparametric estimation of functions, were successfully employed to the estimation of large covariance matrices. Among these works, let us mention minimax and adaptive minimax results: via banding the covariance matrix [9], thresholding the entries of the empirical covariance matrix [10], block-thresholding [25], tapering [26], ℓ_1-estimation [27] and so on. Unlike the estimation of the covariance matrix, there are very few works for testing in a minimax setup in the existing literature.

Several types of test statistics have been proposed in the literature in order to test the null hypothesis (3.1). The likelihood ratio (LR) statistic, was first designed for fixed p and $n \to +\infty$. To treat the high dimensional case when $n, p \to +\infty$, [4] proposed a correction to the LR statistic and showed its convergence in law under the null hypothesis, as soon as $p/n \to c$, for some fixed $c \in (0,1)$. Indeed, this correction is based on the asymptotic behavior of the spectrum of the covariance matrix. Another approach is based on the largest magnitude of the off-diagonal entries of the sample correlation matrix and was introduced by [63]. Later, [22] and [93] show an original limit behavior of Gumbel type for a self-normalized version of the maximum deviation of the sample covariance matrix. We also note that a non-asymptotic sphericity test for Gaussian vectors was studied by [7]. The alternative is given by a model with rank-one and sparse additive perturbation in the variance.

Furthermore, an approach based on the quadratic form $U_n = (1/p)tr(S_n - I)^2$, where $S_n = (1/n)\sum_{i=1}^n X_i X_i^\top$ is the sample covariance matrix, was proposed by [77], to test (3.1). Later, [67] shows that the test of H_0 based on U_n is not consistent for large p. They introduce a corrected version of U_n and study its asymptotic behavior when $n, p \to \infty$ and $p/n \to c \in (0 + \infty)$. In order to deal with non Gaussian random vectors, and without specifying any relation between n and p, [29] proposed a U-statistic of order 2, as a new correction of the previous quadratic form. They do moment assumptions in order to show the asymptotic behavior of their U-statistic, under the null and under a fixed alternative hypothesis. Motivated by their work, [23] used the U-statistic given in [29] to test (3.1) from a sample of Gaussian vectors, and studied the testing problem from a minimax point of view. They consider the alternative hypothesis $H_1 : \Sigma$ such that $\|\Sigma - I\|_F \geq \varphi$ and they establish the minimax rates of order $\sqrt{p/n}$ in this case. In our setup the restriction to the ellipsoid $\mathcal{F}(\alpha, L)$ leads to different rates for testing.

In this paper, we introduce a U-statistic, which is weighted in an optimal way for our problem. This can also be seen as a regularization technique for estimating a quadratic functional, as it is often the case in minimax nonparametric test theory (see [58]). We use this test statistic to construct an asymptotically minimax test procedure. Let us stress the fact that we study the type II error probability uniformly over the set of all matrices Σ under the alternative and that induces a separation rate saying how close Σ can be to the identity matrix I and still be distinguishable from I. We describe the sharp separation rates for fixed unknown α and give an adaptive procedure free of α that allows to test at the price of a logarithmic loss in the rate.

We describe here the rate asymptotics of the error probabilities from the minimax point of view. We recall that a test procedure Δ is a measurable function with respect to the observations, taking values in $[0,1]$. Set $\eta(\Delta) = \mathbb{E}_I(\Delta) = \mathbb{P}_I(\Delta = 1)$ its type I error probability, $\beta(\Delta, Q(\alpha, L, \varphi)) = \sup_{\Sigma \in Q(\alpha,L,\varphi)} \mathbb{E}_\Sigma(1-\Delta) = \sup_{\Sigma \in Q(\alpha,L,\varphi)} \mathbb{P}_\Sigma(\Delta = 0)$ its maximal type II error probability over the set $Q(\alpha, L, \varphi)$, and by

$$\gamma(\Delta, Q(\alpha, L, \varphi)) = \eta(\Delta) + \beta(\Delta, Q(\alpha, L, \varphi))$$

the total error probability of Δ. Let us denote by γ the minimax total error probability over $Q(\alpha, L, \varphi)$ which is defined by

$$\gamma = \gamma(\varphi) := \gamma(Q(\alpha, L, \varphi)) = \inf_{\Delta} \gamma(\Delta, Q(\alpha, L, \varphi))$$

where the infimum is taken over all test procedures. We want to describe the separation rate $\widetilde{\varphi} = \widetilde{\varphi}(n,p)$ such that, on the one hand,

$$\gamma \to 1 \quad \text{if } \frac{\varphi}{\widetilde{\varphi}} \to 0.$$

In this case we say that we can not distinguish between the two hypotheses. On the other hand, we exhibit an explicit test procedure Δ^* such that its total error probability tends to 0

$$\gamma(\Delta^*, Q(\alpha, L, \varphi)) \to 0 \quad \text{if } \frac{\varphi}{\widetilde{\varphi}} \to +\infty.$$

We say that Δ^* is asymptotically minimax consistent test and $\widetilde{\varphi}$ is the asymptotically minimax separation rate.

In this paper, we find asymptotically minimax rates for testing over the class $\mathcal{F}(\alpha, L)$. The minimax consistent test procedure is based on a U-statistic of second order, weighted in an optimal way. In this, our procedure is very different from known corrected procedures based on quadratic forms of the sample covariance matrix, see e.g. [67]. This is the first time a weighted test-statistic is used for testing covariance matrices.

Moreover, our rates are sharp minimax. We show a Gaussian asymptotic behaviour of the test statistic in the neighbourhood of the separation rate. We get the following sharp asymptotic expression for the maximal type II probability error, under some assumptions relating φ, n and p,

$$\inf_{\Delta:\eta(\Delta)\le w} \beta(\Delta, Q(\alpha, L, \varphi)) = \Phi(z_{1-w} - n\sqrt{p}b(\varphi)) + o(1),$$

where Φ denotes the cumulative distribution function (cdf) of the standard Gaussian distribution and z_{1-w} is the $1-w$ quantile of the standard Gaussian distribution for any $w \in (0,1)$. We deduce that the sharp minimax total error probability is of the type

$$\gamma(\varphi) = 2\Phi(-n\sqrt{p}\, b(\varphi)/2) + o(1),$$

where $b^2(\varphi) = C(\alpha, L)\varphi^{4+1/\alpha}$ as $\varphi \to 0$, $C(\alpha, L)$ is explicitly given. It is usual to call the asymptotically sharp minimax rate

$$\widetilde{\varphi} = (C(\alpha, L)n^2 p)^{-\alpha/(4\alpha+1)},$$

corresponding to $n^2 p b^2(\widetilde{\varphi}) = 1$ and to the asymptotic testing constant $C(\alpha, L)$.

Analogous results were obtained by [19] in the particular case where the covariance matrix is Toeplitz, that is $\sigma_{i,j} = \sigma_{|i-j|}$ for all different i and j from 1 to p. We note a gain of a factor p in the minimax rate. The asymptotically sharp minimax rate for Toeplitz covariance matrices is

$$\widetilde{\varphi}_T = (C(\alpha, L)n^2 p^2)^{-\alpha/(4\alpha+1)}.$$

This additional factor p can be heuristically explained by the number of parameters $p-1$ for a Toeplitz matrix, instead of $p(p-1)/2$ for an arbitrary covariance matrix. For $n = 1$ the test problem for Toeplitz covariance matrices was solved in the sharp asymptotic framework, as $p \to \infty$, by [37]. Let us also recall that the adaptive rates (to α) for minimax testing are obtained for the spectral density problem by [47] by a non constructive method using the asymptotic equivalence with a Gaussian white noise model. We also give an adaptive procedure for testing without prior knowledge on α, for α belonging to a closed subset of $(1/2, +\infty)$.

Important generalizations of this problem include testing in a minimax setup of composite null hypotheses like sphericity, $H_0 : \Sigma = v^2 \cdot I$, for unknown v^2 in some compact set separated from 0, or bandedness, $H_0 : \Sigma = \Sigma_0$ such that $[\Sigma_0]_{ij} = 0$ for all $i \neq j$ with $|i-j| > K$. Our proofs rely on the Gaussian distribution of Gaussian vectors. Generalizations to non Gaussian distributions with finite moments of some order can be proposed under additional assumptions on the behaviour of higher order moments, like e.g. [29].

Section 3.2 introduces the test statistic and studies its asymptotic properties. Next we give upper bounds for the maximal type II error probability and for the total error probability and refine these results to sharp asymptotics under the condition that $n = o(1)p^{2\alpha}$. In Section 3.2.3 we implement our test procedure and estimate its power. In Section 3.3 we prove sharp asymptotic optimality and deduce the optimality of the minimax separation rates for all $\alpha > 1$ and as soon as $p = o(n^{4\alpha-1})$. In Section 3.4 we present the rate minimax ressults for testing the inverse of the covariance matrix. In Section 3.5

we define an adaptive test procedure and show that the price of adaptation is a loss of $(\ln\ln(n\sqrt{p}))^{\alpha/(4\alpha+1)}$ in the separation rate.

Proofs are given in Sections 3.6 and 3.7.

3.2 Test procedure and sharp asymptotics

In the minimax theory of tests developed since [53] it is well understood that optimal test statistics are estimators (suitably normalized and tuned) of the functional which defines the separation of an element in the alternative from the element of the null hypothesis. In our case this is the Frobenius norm $\|\Sigma - I\|_F^2 = tr[(\Sigma - I)^2]$.

Weighting the elements of the sample covariance matrix appeared first as hard thresholding in minimax estimation of large covariance matrices. Let us mention [9] for banding i.e. truncation of the matrix to its k first diagonals (closest to the main diagonal), [10] for hard thresholding, then [26] where tapering was studied. It is a natural idea when coming from minimax nonparametric estimation.

However, that was never used for tests concerning large covariance matrices. In this section, we introduce a weighted U-statistic of order 2 for testing large covariance matrices, study its asymptotic properties and give asymptotic upper bounds for the minimax rates of testing.

From now on asymptotics and symbols o, O, $\sim$ and $\asymp$ are considered n and p tend to infinity. Recall that, given sequences of real numbers u and real positive numbers v, we say that they are asymptotically equivalent, $u \sim v$, if $\lim u/v = 1$. Moreover, we say that the sequences are asymptotically of the same order, $u \asymp v$, if there exist two constants $0 < c \le C < \infty$ such that $c \le \liminf u/v$ and $\limsup u/v \le C$.

3.2.1 Test statistic and its asymptotic behaviour

Our test statistic is a weighted U-statistic of order 2. It can be also seen as a weighted functional of the sample covariance matrix. The weights w_{ij}^* are constant on each diagonal (they depend on i and j only through $i - j$), non-zero only for $|i - j| \le T$ for some large integer T and decreasing polynomially for elements further from the main diagonal (as $|i - j|$ is increasing). More precisely, we consider the following test statistic:

$$\widehat{\mathcal{D}}_n = \frac{1}{n(n-1)p} \sum_{1\le k\ne l\le n} \sum_{1\le i<j\le p} w_{ij}^* X_{k,i}X_{k,j}X_{l,i}X_{l,j} \tag{3.3}$$

where

$$\begin{aligned} w_{ij}^* &= \frac{\lambda}{2b(\varphi)}\Big(1 - \Big(\frac{|i-j|}{T}\Big)^{2\alpha}\Big)_+, \quad T = \lfloor C_T(\alpha, L)\cdot \varphi^{-\frac{1}{\alpha}} \rfloor \\ \lambda &= C_\lambda(\alpha, L)\cdot \varphi^{\frac{2\alpha+1}{\alpha}}, \quad b(\varphi) = C^{1/2}(\alpha, L)\cdot \varphi^{2+\frac{1}{2\alpha}} \end{aligned} \tag{3.4}$$

with

$$
\begin{aligned}
C_T(\alpha, L) &= ((4\alpha+1)L)^{\frac{1}{2\alpha}}, \quad C_\lambda(\alpha, L) = \frac{2\alpha+1}{2\alpha}((4\alpha+1)L)^{-\frac{1}{2\alpha}}, \\
C(\alpha, L) &= \frac{2\alpha+1}{(4\alpha+1)^{1+1/(2\alpha)}} L^{-\frac{1}{2\alpha}}.
\end{aligned} \tag{3.5}
$$

The weights $\{w_{ij}^*\}_{i,j}$ and the parameters $T, \lambda, b^2(\varphi)$ are obtained by solving the following optimization problem :

$$
\frac{1}{p}\sum_{1\le i<j\le p} w_{ij}^* \sigma_{ij}^{*2} = \sup_{\left\{\begin{array}{c}(w_{ij})_{ij} \;:\; w_{ij}\ge 0; \\ \frac{1}{p}\sum_{1\le i<j\le p} w_{ij}^2 = \frac{1}{2}\end{array}\right\}} \inf_{\left\{\begin{array}{c}\Sigma \;:\; \Sigma=(\sigma_{ij})_{i,j}; \\ \Sigma\in Q(\alpha,L,\varphi)\end{array}\right\}} \frac{1}{p}\sum_{1\le i<j\le p} w_{ij}\sigma_{ij}^2 \tag{3.6}
$$

Indeed our test statistic $\widehat{\mathcal{D}}_n$ will concentrate asymptotically around the value $\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n) = (1/p)\sum_{1\le i<j\le p} w_{ij}\sigma_{ij}^2$ which is 0 for $\Sigma = I$. The minimax paradigm considers the worst parameter Σ^* in the class $Q(\alpha, L, \varphi)$, that will give the smallest value $\mathbb{E}_{\Sigma^*}(\widehat{\mathcal{D}}_n(w_{ij}))$ and then finds the parameters $\{w_{ij}^*\}_{i<j}$ of the test statistic to provide the largest value $\mathbb{E}_{\Sigma^*}(\widehat{\mathcal{D}}_n(w_{ij}^*))$. Such procedure performs uniformly well over all parameters $\Sigma \in Q(\alpha, L, \varphi)$. That explains why we solve the optimization problem (3.6).

Note that the weights in (3.4) have further properties:

$$
w_{ij}^* \ge 0\ , \quad \frac{1}{p}\sum_{1\le i<j\le p} w_{ij}^{*2} = \frac{1}{2}\ , \quad \sup_{i,j} w_{ij}^* \asymp \frac{1}{\sqrt{T}}, \text{ as } \varphi\to 0 \text{ and } p\varphi^{1/\alpha}\to\infty.
$$

The following Proposition gives the moments of $\widehat{\mathcal{D}}_n$ under the null and their bounds under the alternative hypothesis, respectively, as well as the asymptotic normality under the null hypothesis.

Proposition 3.1. *The test statistic $\widehat{\mathcal{D}}_n$ defined by (3.3) with parameters given by (3.4) and (3.5) has the following moments, under the null hypothesis:*

$$
\mathbb{E}_I(\widehat{\mathcal{D}}_n) = 0, \qquad Var_I(\widehat{\mathcal{D}}_n) = \frac{2}{n(n-1)p^2}\sum_{1\le i<j\le p} w_{ij}^{*2} = \frac{1}{n(n-1)p}
$$

Also we have that

$$
n\sqrt{p}\,\widehat{\mathcal{D}}_n \xrightarrow{d} \mathcal{N}(0,1).
$$

Moreover, under the alternative, if we assume that $\varphi\to 0$, $p\varphi^{1/\alpha}\to\infty$ and $\alpha > 1/2$, we have, for all Σ in $Q(\alpha, L, \varphi)$:

$$
\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n) = \frac{1}{p}\sum_{1\le i<j\le p} w_{ij}^*\sigma_{ij}^2 \ge b(\varphi) \quad \textit{and} \quad Var_\Sigma(\widehat{\mathcal{D}}_n) = \frac{T_1}{n(n-1)p^2} + \frac{T_2}{np^2},
$$

where

$$
T_1 \le p\cdot(1+o(1)) + p\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n)\cdot O(T\sqrt{T}), \tag{3.7}
$$

$$
T_2 \le p\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n)\cdot O(\sqrt{T}) + p^{3/2}\cdot\Big(\mathbb{E}_\Sigma^{3/2}(\widehat{\mathcal{D}}_n)\cdot O(T^{3/4}) + \mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n)\cdot o(1)\Big). \tag{3.8}
$$

Note that, under the alternative, we have the additional assumption that $p\varphi^{1/\alpha} \asymp p/T \to +\infty$, when p grows to infinity. This is natural in order to a have a meaningful weighted statistic.

Let us look closer at the optimization problem (3.6): for given $\varphi > 0$, $b(\varphi)$ is the least value that $\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n)$ can take over Σ in the alternative set of hypotheses.

Under the alternative, we shall establish the asymptotic normality under additional conditions that the underlying covariance matrix is close to the border of $\mathcal{F}(\alpha, L)$. This will be sufficient to give upper bounds of the total error probability of Gaussian type in next Section.

Proposition 3.2. *The test statistic $\widehat{\mathcal{D}}_n$ defined by (3.3) with parameters given by (3.4) and (3.5), such that $\varphi \to 0$, $p\varphi^{1/\alpha} \to \infty$ and under the aditionnal assumption that $n^2pb^2(\varphi) \asymp 1$, is asymptotically normal:*

$$n\sqrt{p}(\widehat{\mathcal{D}}_n - \mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n)) \xrightarrow{d} \mathcal{N}(0,1),$$

for any Σ in $Q(\alpha, L, \varphi)$ such that $\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n) = O(b(\varphi))$.

3.2.2 Upper bounds for the error probabilities

In order to distinguish between the two hypothesis H_0 and H_1 defined previously, we propose the following test procedure

$$\Delta^* = \Delta^*(t) = \mathbb{1}(\widehat{\mathcal{D}}_n > t), \quad t > 0 \tag{3.9}$$

where $\widehat{\mathcal{D}}_n$ is the estimator defined in (3.3).

The following theorem proves that the previously defined test procedure is minimax consistent if t is conveniently chosen.

Theorem 3.3. *The test procedure Δ^* defined in (3.9) with $t > 0$ has the following properties:*

Type I error probability : if $n\sqrt{p} \cdot t \to +\infty$ then $\eta(\Delta^*) \to 0$.

Type II error probability : if $\alpha > 1/2$ and if

$$\varphi \to 0, p\varphi^{1/\alpha} \to \infty \text{ and } n^2pb^2(\varphi) \to +\infty$$

then, uniformly over t such that $t \leq c \cdot C^{1/2}(\alpha, L) \cdot \varphi^{2+\frac{1}{2\alpha}}$, for some constant c in $(0,1)$, we have

$$\beta(\Delta^*(t), Q(\alpha, L, \varphi)) \to 0.$$

If t verifies all previous assumptions, then $\Delta^*(t)$ is asymptotically minimax consistent:

$$\gamma(\Delta^*(t), Q(\alpha, L, \varphi)) \to 0.$$

In the next Theorem we give sharp upper bounds of error probabilities of Gaussian type. The proof of this result explains the choice of the weights as solution of the optimization problem (3.6). Moreover, we will see that the Gaussian behavior is obtained near the separation rates.

Recall that Φ is the cumulative distribution function (cdf) of standard Gaussian random variable and, for any $w \in (0,1)$, z_{1-w} is defined by $\Phi(z_{1-w}) = 1 - w$.

Theorem 3.4. *The test procedure Δ^* defined in (3.9) with $t > 0$ has the following properties:*

Type I error probability : we have $\eta(\Delta^*(t)) = 1 - \Phi(n\sqrt{p}\cdot t) + o(1)$.

Type II error probability : if $\alpha > 1/2$ and if

$$\varphi \to 0,\, p\varphi^{1/\alpha} \to \infty \text{ and } n^2 p\, b^2(\varphi) \asymp 1\,, \tag{3.10}$$

then, uniformly over t, we have

$$\beta(\Delta^*(t), Q(\alpha, L, \varphi)) \le \Phi(n\sqrt{p}\cdot(t - b(\varphi))) + o(1).$$

In particular, for $t = t^w$ such that $n\sqrt{p}\cdot t^w = z_{1-w}$ we have $\eta(\Delta^*(t^w)) \le w + o(1)$ and

$$\beta(\Delta^*(t^w), Q(\alpha, L, \varphi)) \le \Phi(z_{1-w} - n\sqrt{p}\cdot b(\varphi)) + o(1).$$

Another important consequence of the previous theorem, is that the test procedure $\Delta^*(t^*)$, with $t^* = b(\varphi)/2$ has total error probability

$$\gamma(\Delta^*(t^*), Q(\alpha, L, \varphi)) \le 2\,\Phi\left(-n\sqrt{p}\,\frac{b(\varphi)}{2}\right) + o(1).$$

3.2.3 Simulation study

We include two examples, to illustrate the numerical behavior of our test procedure. First, we test the null hypothesis $\Sigma = I$ against the alternative hypothesis defined by the symmetric positive matrices $\Sigma(M) = \left(\mathbb{1}_{\{i=j\}} + \mathbb{1}_{\{i\neq j\}}\cdot(|i-j|^{-\frac{3}{2}}\cdot|i+j|^{\frac{1}{100}})/M\right)_{1\le i,j\le p}$. We implement the test statistic $\widehat{\mathcal{D}}_n$ defined in (3.3) and (3.4) for $\alpha = L = 1$, and $\varphi = \psi(M) = (1/p)\left(\sum_{i<j}|i-j|^{-3}\cdot|i+j|^{\frac{2}{100}}\right)^{1/2}/M$. We choose the threshold t of the test empirically, under the null hypothesis $\Sigma = Id$, from 1000 repetead samples of size n, such that the type I error probability is fixed at 0.05. We use t to estimate the type II error probability, also from 1000 repetitions and then plot the power as function of $\psi(M)$.

Figure 3.1 shows that the power is an increasing function of $\psi(M)$. Also, we can see that for a fixed value of $\psi(M)$ the power increases with p. Indeed, our procedure benefits from large values of p, which is not a nuisance parameter here.

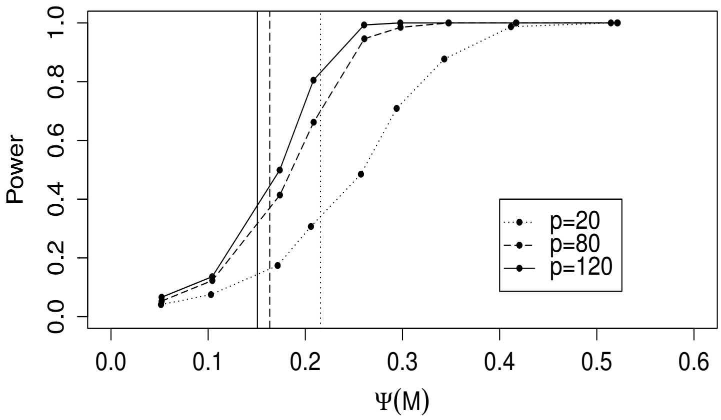

Figure 3.1 – Power curves of the Δ-test as function of $\psi(M)$ for $n = 20$ and $p \in \{20, 80, 120\}$

Second we consider the tridiagonal matrices $\Sigma(\rho) = (\mathbb{1}_{\{i=j\}} + \rho \mathbb{1}_{\{|i-j|=1\}})_{1 \le i,j \le p}$, for $\rho \in (0, 0.35]$ under the alternative hypothesis. We compare our test procedure to the one given in [23], which is based on a U-statistic of order 2, we denote it CM-test. Moreover, the matrices $\Sigma(\rho)$ are Toeplitz, we also compare our the procedure to the one proposed in [19] for Toeplitz covariance matrices, that we denote by BZ-test. The thresholds are evaluated empirically for each procedure at type I error probability smaller than 0.05. We finally plot the powers curves of the three test procedures.

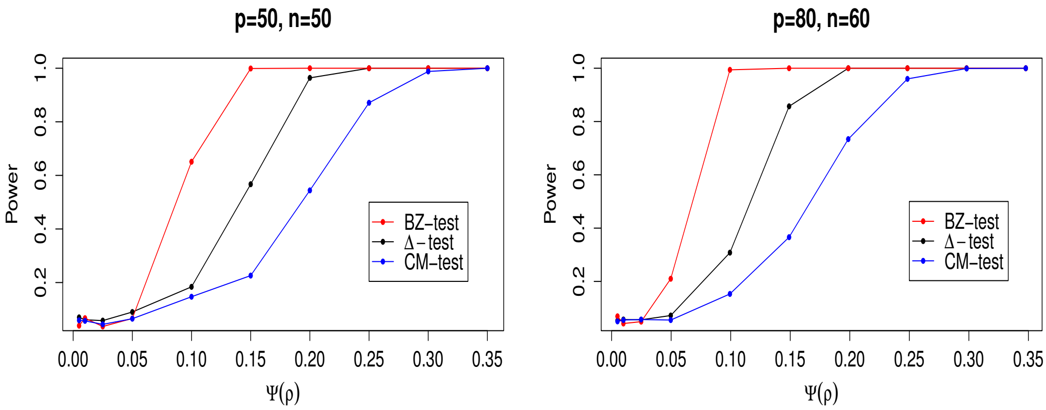

Figure 3.2 – Power curves of the BZ-test, Δ-test and CM-test as function of $\psi(\rho)$

Figure 3.2 shows that, when the alternative hypothesis consists of Toeplitz matrices the BZ-test has the better performance. However if we miss the information that the matrix is

Toeplitz, we see that the Δ-test is not bad and its power dominates the power of the CM-test.

3.3 Asymptotic optimality

In this section, we first state the lower bound for testing, which, in addition to the test procedure exhibited in the previous section, shows that the asymptotically minimax separation rate is

$$\widetilde{\varphi} = \left(n\sqrt{p}\,C^{1/2}(\alpha, L)\right)^{-\frac{2\alpha}{4\alpha+1}}, \tag{3.11}$$

where the constant $C(\alpha, L)$ is given by (3.5).

Theorem 3.5. *Assume that, either $\alpha > 3/2$, or $\alpha > 5/8$ and $np\varphi^{6-\frac{2}{\alpha}} \to 0$. If*

$$\varphi \to 0,\ p\varphi^{1/\alpha} \to \infty, \quad \text{and } n^2 p\, b^2(\varphi) \to 0,$$

then

$$\gamma = \inf_{\Delta} \gamma(\Delta, Q(\alpha, L, \varphi)) \to 1,$$

where the infimum is taken over all test statistics Δ.

Together with Theorem 3.3, the proof that $\widetilde{\varphi}$ is asymptotically minimax, under our assumptions, is complete. Note that the condition $np\varphi^{6-\frac{2}{\alpha}} \to 0$ is verified when $\alpha > 3/2$ for all n and $p \to +\infty$ giving a general result in this case. When $5/8 < \alpha < 3/2$, the same condition holds for $p = o(n^{\frac{8\alpha-5}{3-2\alpha}})$. This result is proven by showing that the χ^2 distance between the null hypothesis and an averaged likelihood under the alternative (that we explicitly construct) tends to 0.

Moreover, we give a sharp lower bound for the type II error probability which is of Gaussian type.

Theorem 3.6. *Assume that $\alpha > 1$ and if*

$$\varphi \to 0,\ p\varphi^{1/\alpha} \to \infty,\ \sqrt{p}\varphi^{2-\frac{1}{2\alpha}} \to 0 \text{ and } n^2 p\, b^2(\varphi) \asymp 1, \tag{3.12}$$

$$\inf_{\Delta:\eta(\Delta)\le w} \beta(\Delta, Q(\alpha, L, \varphi)) \ge \Phi(z_{1-w} - n\sqrt{p}b(\varphi)) + o(1),$$

where the infimum is taken over all test statistics Δ with type I error probability less than or equal to w. Moreover,

$$\gamma = \inf_{\Delta} \gamma(\Delta, Q(\alpha, L, \varphi)) \ge 2\Phi(-n\sqrt{p}\,\frac{b(\varphi)}{2}) + o(1).$$

Theorems 3.4 and 3.6 imply that for $\alpha > 1$, the sharp separation rate for minimax testing is $\widetilde{\varphi}$, under the additionnal assumptions (3.10) and (3.12). Note that a sufficient condition is that the separation rate verifies these assumptions, in particular $p\widetilde{\varphi}^{1/\alpha} \to \infty$ holds if $n = o(1)p^{2\alpha}$, and $\sqrt{p}\widetilde{\varphi}^{2-\frac{1}{2\alpha}} \to 0$ holds if $p = o(n^{4\alpha-1})$.

Note that, there is a more general test procedure independent of φ, for which it is possible to derive the upper bounds as in Theorems 3.3 and 3.4. It suffices to use the test statistic $\widehat{\mathcal{D}}_n$ with the weights w_{ij}^* replaced by the weights $w_{ij}^*(\widetilde{\varphi})$ defined as in (3.4) and (3.5) for φ replaced by $\widetilde{\varphi}$. For more details see section 4.2 in [14].

The proof of the lower bounds is given in Section 3.6. We construct a family of n large centered Gaussian vectors with covariance matrices based on $\{\sigma_{ij}^*\}_{1\le i,j\le p}$ given by the optimization problem (3.6) and a prior measure P_π on these covariance matrices. The logarithm of the likelihood ratio associated to an arbitrary Σ with respect to I under the null hypothesis is known to drift away to infinity (see [4], who corrected this ratio to get a proper limit). However, we show that the logarithm of the Bayesian likelihood ratio with our prior measure P_π verifies

$$\log \frac{f_\pi}{f_I}(X_1, ..., X_n) = u_n Z_n - \frac{u_n^2}{2} + \xi, \qquad \text{in } P_I \text{ probability}$$

where $u_n = n\sqrt{p} b(\varphi)$, Z_n is asymptotically distributed as a standard Gaussian distribution and ξ is a random variable which converges to zero under P_I probability.

3.4 Testing the inverse of the covariance matrix

Let us consider the same model, but the following test problem

$$H_0 : \Sigma^{-1} = I$$

against the alternative

$$H_1 : \Sigma \in \mathcal{G}(\alpha, L, \lambda) \text{ such that } \frac{1}{2p}\|\Sigma^{-1} - I\|_F^2 \ge \psi^2,$$

where $\mathcal{G}(\alpha, L, \lambda)$ is the class of covariance matrices Σ in $\mathcal{F}(\alpha, L)$ with the additional constraint that the eigenvalues $\lambda_i(\Sigma)$ are bounded from below by some $\lambda \in (0,1)$ for all i from 1 to p and all Σ in the set.

We prove here that previous results apply to this setup and we get the same rates, but not the sharp asymptotics. Note that, the additional hypothesis is mild enough so that it does not change the rates for testing. Indeed, we see this case as a well-posed inverse problem. The cases of ill-posed inverse problem where the smallest eigenvalue can be allowed to tend to 0 will most certainly imply a loss in the rate and is beyond the scope of this paper.

Theorem 3.7. *Suppose $\alpha > 3/2$, $L > 0$ and $\lambda \in (0,1)$. If n and p tend to infinity and if $\psi \to 0$ such that $p\psi^{\frac{1}{\alpha}} \to +\infty$, then $\widetilde{\varphi}$ defined in (3.11) is the asymptotically minimax rate for the previous test.*

Proof. Note that $\Sigma^{-1} = I$ if and only if $\Sigma = I$. Moreover, if Σ belongs to $\mathcal{G}(\alpha, L, \lambda)$ such that $\frac{1}{2p}\|\Sigma^{-1} - I\|_F^2 \geq \psi^2$, then Σ obviously belongs to $\mathcal{F}(\alpha, L)$ and is such that

$$\frac{1}{2p}\|\Sigma - I\|_F^2 \geq \frac{\lambda^2}{2p}\|\Sigma^{-1} - I\|_F^2 \geq \lambda^2\psi^2.$$

Thus we can proceed with our former test procedure, with φ replaced by $\lambda\psi$ and we obtain the upper bounds in the definition of the separation rates.

The lower bounds in the previous Section will also remain valid. Indeed, this proof is based on the construction of a subfamily $\{\Sigma_U^* : u \in \mathcal{U}\}$ on the set of alternatives. We have proven in Proposition 3.9, that

$$\min_i \lambda_i(\Sigma_U^*) \geq 1 - O(\varphi^{1-1/(2\alpha)}),$$

and we have $\alpha > 1$ and $\varphi = \lambda\psi \to 0$ as $\psi \to 0$ and therefore, $1 - O(\varphi^{1-1/(2\alpha)}) \geq \lambda$ for $\psi > 0$ small enough. Thus, this family belongs to the set of alternatives we consider here, as well. Moreover, Proposition 3.9 proves also that

$$\|\Sigma_U^*\|_2 := \max_i \lambda_i(\Sigma_U^*) \leq 1 + O(\varphi^{1-1/(2\alpha)}) \leq \lambda_{max},$$

for some fixed λ_{max} free of α and L. Thus,

$$\frac{1}{2p}\|(\Sigma_U^*)^{-1} - I\|_F^2 \geq \frac{1}{2p \cdot \|\Sigma_U^*\|_2^2} \cdot \|\Sigma_U^* - I\|_F^2 \geq \frac{1}{\lambda_{max}^2}\frac{1}{2p} \cdot \|\Sigma_U^* - I\|_F^2.$$

Thus we proceed the same way with φ replaced by $\lambda_{max}\psi$. □

3.5 Adaptive testing procedure

We want to built a test procedure of H_0 in (3.1) which is free of the parameter α belonging to some closed interval $\mathcal{A} = [\underline{\alpha}, \bar{\alpha}] \subset (1/2, +\infty)$. The radius L plays a minor role in the procedure and we suppose that it is known (w.l.o.g we assume that $L = 1$). Such a procedure is called adaptive and it solves the test problem H_0 in (3.1) against a much larger set of alternative hypotheses:

$$H_1 : \Sigma \in \bigcup_{\alpha \in \mathcal{A}} Q(\alpha, L, \mathcal{C}\psi_\alpha)\,, \tag{3.13}$$

where $\mathcal{C}$ is a large enough positive constant and

$$\psi_\alpha = \left(\frac{\rho_{n,p}}{n\sqrt{p}}\right)^{\frac{2\alpha}{4\alpha+1}}, \quad \rho_{n,p} = \sqrt{\ln\ln(n\sqrt{p})}\,, \tag{3.14}$$

depend on n and p, but also on α. In order to construct the adaptive test procedure, we define a finite regular grid over the set $\mathcal{A} = [\underline{\alpha}, \bar{\alpha}]$:

$$\mathcal{A}_N = \{\alpha_r = \underline{\alpha} + \frac{\bar{\alpha} - \underline{\alpha}}{N} \cdot r\,; r = 1, \ldots, N\}, \text{ where } N = \lceil \ln(n\sqrt{p}) \rceil.$$

To each $r \in \{1, \cdots, N\}$, we associate the weights :

$$w_{ij,r}^* = \frac{\lambda_r}{b_r}\Big(1 - \Big(\frac{|i-j|}{T_r}\Big)^{2\alpha_r}\Big)_+,$$

where the parameters λ_r, b_r and T_r are given in (3.4) and (3.5) with α replaced by α_r and φ by ψ_α. Define the adaptive test procedure, for some constant $\mathcal{C}^* > 0$ large enough

$$\Delta_{ad}^* = \max_{r=1,\dots,N} \mathbb{1}(\widehat{\mathcal{D}}_{n,r} > \mathcal{C}^* t_r), \quad \text{where } t_r = C_{\lambda_r} \cdot \rho_{n,p}/(n\sqrt{p}), \tag{3.15}$$

and where $\widehat{\mathcal{D}}_{n,r}$ is the test statistic in (3.3) with weights $\{w_{ij,r}\}_{i<j}$. Note that the test Δ_{ad}^* rejects the null hypothesis as soon as there exists at least on $r \in \{1, \dots, N\}$ for which $\widehat{\mathcal{D}}_{n,r} > \mathcal{C}^* t_r$.

Theorem 3.8. *Assume that*

$$p \cdot \Big(\frac{\rho_{n,p}}{n\sqrt{p}}\Big)^{\frac{2}{4\underline{\alpha}+1}} \to +\infty \quad \text{and} \quad \frac{\ln p}{n} \to 0$$

The test statistic defined in (3.15) with $\mathcal{C}^*$ large enough verifies :

$$\gamma(\Delta_{ad}^*, \underset{\alpha \in \mathcal{A}}{\cup} Q(\alpha, L, \mathcal{C}\psi_\alpha)) \to 0,$$

for all $\mathcal{C} > \Big(\mathcal{C}^* + \dfrac{1}{C(\underline{\alpha}, \bar{\alpha})}\Big)$, where ψ_α is given in (3.14) and $C(\underline{\alpha}, \bar{\alpha}) = \exp(-8(\bar{\alpha} - \underline{\alpha})/(4\underline{\alpha} + 1))$.

The proof that the adaptive procedure we propose attains the above rate is given in Section 3.6. By analogy to nonparametric testing of functions, we expect the loss $\rho_{n,p}$ to be optimal uniformly over the class in the alternative hypothesis (3.13) .

3.6 Proofs

Proof of Theorems 3.3 and 3.4. The proof is based on the Proposition 3.1 and the asymptotic normality of the weighted test statistic $n\sqrt{p}\widehat{\mathcal{D}}_n$ in Proposition 3.2. We get for the type I error probability of Δ^*

$$\eta(\Delta) = \mathbb{P}(\widehat{\mathcal{D}}_n > t) = 1 - \Phi(n\sqrt{p} \cdot t) + o(1).$$

For the type II error probability of Δ^*, uniformly in Σ over $Q(\alpha, L, \varphi)$, we have

$$\mathbb{P}_\Sigma(\widehat{\mathcal{D}}_n \le t) \quad \le \quad \mathbb{P}_\Sigma(|\widehat{\mathcal{D}}_n - \mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n)| \ge \mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n) - t) \le \frac{\mathrm{Var}_\Sigma(\widehat{\mathcal{D}}_n)}{(\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n) - t)^2},$$

for $t \le c \cdot b(\varphi)$ and $0 < c < 1$. It implies that $n\sqrt{p} \cdot t \le cn\sqrt{p}b(\varphi)$. Therefore, we distinguish the cases where $n^2pb^2(\varphi)$ tends to infinity or is bounded.

We use the fact that, under the alternative, $\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n) \geq b(\varphi)$. We bound from below as follows:

$$\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n) - t \geq (1-c)\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n).$$

Then, it gives

$$\mathbb{P}_\Sigma(\widehat{\mathcal{D}}_n \leq t) \quad \leq \quad \frac{T_1}{n(n-1)p^2(1-c)^2\mathbb{E}^2_\Sigma(\widehat{\mathcal{D}}_n)} + \frac{T_2}{np^2(1-c)^2\mathbb{E}^2_\Sigma(\widehat{\mathcal{D}}_n)} =: S_1 + S_2.$$

Let us bound from above S_1 using (3.7):

$$S_1 \leq \frac{1+o(1)}{n(n-1)p(1-c)^2b^2(\varphi)} + \frac{O(T^{3/2})}{n(n-1)p\,b(\varphi)}.$$

We have $T^{3/2}b(\varphi) \asymp \varphi^{2-\frac{1}{\alpha}} = o(1)$, for all $\alpha > 1/2$, which proves that :

$$S_1 \leq \frac{1+o(1)}{n(n-1)p(1-c)^2b^2(\varphi)}$$

which tends to 0 provided that $n^2pb^2(\varphi) \to +\infty$. We will see using (3.8) that the term S_2 tends to 0 as well:

$$\begin{aligned} S_2 \quad &\leq \quad \frac{O(\sqrt{T})}{np\,b(\varphi)} + \frac{O(T^{3/4}b^{1/2}(\varphi))}{n\sqrt{p}\,b(\varphi)} + \frac{o(1)}{n\sqrt{p}\,b(\varphi)} \\ &= \quad o(1) \text{ for all } \alpha > 1/2, \text{ as soon as } n\sqrt{p}b(\varphi) \to +\infty. \end{aligned}$$

Now, if φ is close to the separation rate: $n^2pb^2(\varphi) \asymp 1$, we see that whenever $\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n)/b(\varphi)$ tends to infinity, the bound is trivial ($S_1 + S_2 \to 0$).

The nontrivial bound is obtained when Σ under the alternative is close to the optimal matrix $\Sigma^* = (\sigma^*_{ij})_{1\leq i,j\leq p}$, in the sense that $\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n) = O(b(\varphi))$ together with the fact that φ is close to the separation rate: $n^2pb^2(\varphi) \asymp 1$. We apply Proposition 3.2 to get the asymptotic normality

$$n\sqrt{p}(\widehat{\mathcal{D}}_n - \mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n)) \to \mathcal{N}(0,1).$$

Thus,

$$\begin{aligned} \sup_{\Sigma\in Q(\alpha,L,\varphi))} \mathbb{P}_\Sigma(\widehat{\mathcal{D}}_n \leq t) \quad &\leq \quad \sup_{\Sigma\in Q(\alpha,L,\varphi))} \Phi(n\sqrt{p}\cdot(t - \mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n))) + o(1) \\ &\leq \quad \Phi(n\sqrt{p}\cdot(t - \inf_{\Sigma\in Q(\alpha,L,\varphi))} \mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n))) + o(1). \end{aligned}$$

At this point, choosing optimal weights translates into

$$\begin{aligned} &\inf_{w_{ij}>0:\sum_{i\neq j} w_{ij}^2 = 1/2} \quad \sup_{\Sigma\in Q(\alpha,L,\varphi))} \mathbb{P}_\Sigma(\widehat{\mathcal{D}}_n \leq t) \\ &\quad\leq \quad \Phi(n\sqrt{p}\cdot(t - \sup_{w_{ij}>0:\sum_{i\neq j} w_{ij}^2=1/2} \inf_{\Sigma\in Q(\alpha,L,\varphi))} \mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n))) + o(1) \\ &\quad\leq \quad \Phi(n\sqrt{p}\cdot(t - b(\varphi))) + o(1), \end{aligned}$$

after solving the optimization problem, which ends the proof of the Theorem. □

Proof of Theorems 3.5 and 3.6. The first step of the proof is to reduce the set of parameters to a convenient parametric family. Let $\Sigma^* = [\sigma^*_{ij}]_{1\le i,j\le p}$ be the matrix which has 1 on the diagonal and off-diagonal entries σ^*_{ij} where

$$\sigma^*_{ij} = \sqrt{\lambda}\left(1 - (\frac{|i-j|}{T})^{2\alpha}\right)_+^{\frac{1}{2}} \qquad \text{for } i \neq j, \tag{3.16}$$

with λ and T are given by (3.4) and (3.5).

Let us define Q^* a subset of $Q(\alpha, L, \varphi)$ as follows

$$Q^* = \{\Sigma^*_U : [\Sigma^*_U]_{ij} = I(i=j) + u_{ij}\sigma^*_{ij} I(i \neq j) \text{ for all } 1 \le i,j \le p \ , \ U = [u_{ij}]_{1\le i,j\le p} \in \mathcal{U}\},$$

where

$$\mathcal{U} = \{U = [u_{ij}]_{1\le i,j\le p} : u_{ii} = 0, \forall i \text{ and } \ u_{ij} = u_{j\,i} = \pm 1 \cdot I(|i-j| \le T), \text{ for } i \neq j\}.$$

The cardinality of $\mathcal{U}$ is $p(T-1)/2$.

Proposition 3.9. *For $\alpha > 1/2$, the symmetric matrix $\Sigma^*_U = [u_{ij}\sigma^*_{ij}]_{1\le i,j\le p}$, with $\sigma^*_{ii} = 1$, for all i from 1 to p, and σ^*_{ij} defined in (3.16) is non-negative definite, for $\varphi > 0$ small enough, and for all $U \in \mathcal{U}$.*

Moreover, denote by $\lambda_{1,U}, ..., \lambda_{p,U}$ the eigenvalues of Σ^*_U, then $|\lambda_{i,U} - 1| \le O(1)\varphi^{1-1/(2\alpha)}$, for all i from 1 to p.

We deduce that

$$\|\Sigma^*_U\| \le 1 + O(\varphi^{1-\frac{1}{2\alpha}}) \text{ and } \|\Sigma^*_U - I\| \le O(\varphi^{1-\frac{1}{2\alpha}}). \tag{3.17}$$

Indeed, $\|\Sigma^*_U\| = \max_{i=1,\dots,p} \lambda_{i,U} \le 1 + O(\varphi^{1+\frac{1}{2\alpha}})$ and $\Sigma^*_U - I$ has eigenvalues $\lambda_{i,U} - 1$.

Proposition 3.9 shows that for all $\Sigma^*_U \in Q^*$, Σ^*_U is non-negative definite, for $\varphi > 0$ small enough.

Assume that $X_1, \dots, X_n \sim N(0, I)$ under the null hypothesis and denote by P_I the likelihood of these random variables. We assume that $X_1, \dots, X_n \sim N(0, \Sigma^*_U)$, under the alternative, and we denote P_U the associated likelihood. In addition let

$$P_\pi = \frac{1}{2^{p(T-1)/2}} \sum_{U \in \mathcal{U}} P_U$$

be the average likelihood over Q^*.

The problem can be reduced to the test $H_0 : X_1, ..., X_n \sim P_I$ against the averaged distribution $H_1 : X_1, ..., X_n \sim P_\pi$, in the sense that

$$\inf_{\Delta:\eta(\Delta)\le w} \beta(\Delta(t), Q(\alpha, L, \varphi)) \ge \inf_{\Delta:\eta(\Delta)\le w} \beta(\Delta(t), P_\pi) + o(1)$$

and that

$$\inf_{\Delta} \gamma(\Delta, Q(\alpha, L, \varphi)) \ge \inf_{\Delta} \gamma(\Delta, P_\pi) + o(1).$$

It is, therefore, sufficient to show that, when $u_n \asymp 1$,

$$\inf_{\Delta:\eta(\Delta)\le w} \beta(\Delta(t), P_\pi) \ge \Phi(n\sqrt{p}\cdot(t - b(\varphi))) + o(1) \tag{3.18}$$

and that

$$\inf_{\Delta} \gamma(\Delta, P_\pi) \ge 2\Phi(-n\sqrt{p}\,\frac{b(\varphi)}{2}) + o(1). \tag{3.19}$$

While, for $u_n = o(1)$, we need to show that

$$\inf_{\Delta} \gamma(\Delta, P_\pi) \to 1. \tag{3.20}$$

In order to obtain (3.18) and (3.19), we apply results in Section 4.3.1 of [58] giving the sufficient condition that, in P_I probability:

$$L_{n,p} := \log \frac{f_\pi}{f_I}(X_1, ..., X_n) = u_n Z_n - \frac{u_n^2}{2} + \xi, \tag{3.21}$$

where $u_n = n\sqrt{p}b(\varphi) \asymp 1$, $b(\varphi) = C^{\frac{1}{2}}(\alpha, L)\cdot\varphi^{2+\frac{1}{2\alpha}}$, Z_n is asymptotically distributed as a standard Gaussian distribution and ξ is a random variable which converges to zero under P_I probability. Moreover, to show (3.20), it suffices to show that

$$\mathbb{E}_I\left(\frac{dP_\pi}{dP_I}\right)^2 \le 1 + o(1), \tag{3.22}$$

since

$$\gamma \ge 1 - \frac{1}{2}\|P_I - P_\pi\|_1 \text{ and } \|P_I - P_\pi\|_1^2 \le \mathbb{E}_I\left(\frac{dP_\pi}{dP_I}\right)^2 - 1.$$

We first begin by showing (3.22), in order to finish the proof of Theorem 3.5. Let,

$$\begin{aligned} H_{n,p} &:= \mathbb{E}_I\left(\frac{dP_\pi}{dP_I}(X_1,\cdots,X_n)\right)^2 \\ &= \mathbb{E}_I\mathbb{E}_{U,V}\left(\frac{\exp\left(-\frac{1}{2}\sum_{k=1}^n X_k^\top((\Sigma_U)^{-1} + (\Sigma_V)^{-1} - 2I)X_k\right)}{(2\pi)^{\frac{np}{2}}\det^{\frac{n}{2}}(\Sigma_U\Sigma_V)}\right). \end{aligned} \tag{3.23}$$

We have

$$H_{n,p} = \mathbb{E}_{U,V}\left(\frac{\det^{-\frac{n}{2}}\left((\Sigma_U)^{-1} + (\Sigma_V)^{-1} - I\right)}{\det^{\frac{n}{2}}(\Sigma_U\Sigma_V)}\right) = \mathbb{E}_{U,V}\left(\det^{-\frac{n}{2}}\left(\Sigma_U + \Sigma_V - \Sigma_U\Sigma_V\right)\right).$$

We define $\Delta_U = \Sigma_U - I$ and note that $\Sigma_U + \Sigma_V - \Sigma_U\Sigma_V = I - \Delta_U\Delta_V$. As the matrix $\Delta_U\Delta_V$ is not necessarily symmetric, we write

$$(I - \Delta_U\Delta_V)(I - \Delta_U\Delta_V)^\top = I - M$$

where $M = M_{U,V} := \Delta_U\Delta_V + \Delta_V\Delta_U - \Delta_U\Delta_V^2\Delta_U$ is symmetric. Moreover, we prove that for all U and $V \in \mathcal{U}$ the eigenvalues of M are in $(-1, 1)$ for all $\alpha > 1/2$ and φ small enough.

Indeed, by Gershgorin's theorem, for each eigenvalue λ_M of M there exists at least one $i \in \{1,\dots,p\}$ such that

$$|\lambda_M - M_{ii}| \le \sum_{j;j\neq i} |M_{ij}|.$$

We can show that $\sum_{j;j\neq i} |M_{ij}| = O(\varphi^{2-\frac{1}{\alpha}})$ and $|M_{ii}| \le O(\varphi^2) + O(\varphi^{4-\frac{1}{\alpha}})$. Thus,

$$H_{n,p} = \mathbb{E}_{U,V}(\det^{-\frac{n}{4}}(I-M)) = \mathbb{E}_{U,V}\exp\Big(-\frac{n}{4}\text{logdet}(I-M)\Big)$$

The Taylor expansion for the logdet of a symmetric matrix writes

$$-\frac{1}{4}\text{logdet}(I-M) = \frac{1}{4}tr(M) + \frac{1}{8}tr(M^2) + O(tr(M^3)).$$

In more details,

$$\begin{aligned}
\frac{1}{4}tr(M) &= \frac{1}{2}tr(\Delta_U\Delta_V) - \frac{1}{4}tr(\Delta_U^2\Delta_V^2) \\
\frac{1}{8}tr(M^2) &= \frac{1}{4}tr(\Delta_U\Delta_V)^2 + \frac{1}{4}tr(\Delta_U^2\Delta_V^2) + \frac{1}{8}tr(\Delta_U\Delta_V^2\Delta_U)^2 \\
&\quad -\frac{1}{4}tr(\Delta_U\Delta_V^2\Delta_U^2\Delta_V) - \frac{1}{4}tr(\Delta_V\Delta_U^2\Delta_V^2\Delta_U)
\end{aligned}$$

Recall that $\forall A,B \in \mathbb{R}^{p\times p}$ we have $\|AB\|_F \le \|A\|_2\|B\|_F$. For all $U,V \in \mathcal{U}$, we use the last inequality and the Cauchy-Schwarz inequality to get

$$tr(\Delta_U\Delta_V^2\Delta_U^2\Delta_V) \le \|\Delta_U\Delta_V\|_F\,\|\Delta_V\Delta_U^2\Delta_V\|_F \le \|\Delta_V\|_2\,\|\Delta_U\|_F\,\|\Delta_V\Delta_U^2\|_2\,\|\Delta_V\|_F \le p\cdot\varphi^{6-\frac{2}{\alpha}},$$

$$tr(\Delta_U\Delta_V^2\Delta_U)^2 = \|\Delta_U\Delta_V^2\Delta_U\|_F^2 \le \|\Delta_U\Delta_V^2\|_2^2\ \|\Delta_U\|_F^2\ \le p\cdot\varphi^{8-\frac{3}{\alpha}}.$$

Finally, using similar arguments we can show that

$$tr(M_{U,V}^3) = O(p\varphi^{6-\frac{2}{\alpha}}).$$

Thus,

$$-\frac{1}{4}\text{logdet}(I-M) = \frac{1}{2}tr(\Delta_U\Delta_V) + \frac{1}{4}tr(\Delta_U\Delta_V)^2 + O(p\varphi^{6-\frac{2}{\alpha}}).$$

Now we develop the terms on the right hand side of the previous equation. We obtain

$$tr(\Delta_U\Delta_V) = \sum_{\substack{1\le i,j\le p\\ 1<|i-j|<T}} u_{ij}v_{ij}\cdot\sigma_{ij}^{*2} = 2\sum_{\substack{1\le i<j\le p\\ 1<|i-j|<T}} u_{ij}v_{ij}\cdot\sigma_{ij}^{*2}$$

and

$$\begin{aligned}
tr(\Delta_U\Delta_V)^2 &= \sum_{\substack{1\le i,j,h,l\le p\\ 1<|i-h|,|h-j|,|i-l|,|l-j|<T}} u_{ih}v_{hj}u_{jl}v_{li}\cdot\sigma_{ih}^{*}\sigma_{hj}^{*}\sigma_{jl}^{*}\sigma_{li}^{*} \\
&= \sum_{\substack{1\le i,l\le p\\ 1<|i-l|<T}} \sigma_{ij}^{*4} + 2\sum_{\substack{1\le i,j,l\le p\\ i<j\\ 1<|i-l|,|l-j|<T}} u_{il}u_{jl}v_{lj}v_{li}\cdot\sigma_{il}^{*2}\sigma_{lj}^{*2} \\
&+ 4\sum_{\substack{1\le i,j,h,l\le p\\ i<j,l<h\\ 1<|i-h|,|h-j|,|i-l|,|l-j|<T}} u_{ih}u_{jl}v_{hj}v_{li}\cdot\sigma_{ih}^{*}\sigma_{jl}^{*}\sigma_{hj}^{*}\sigma_{li}^{*}.
\end{aligned}$$

Now, we can write (3.23) as follows:

$$
\begin{aligned}
H_{n,p} &= \mathbb{E}_{U,V} \exp\Big(-\frac{n \operatorname{logdet}(I-\Delta_U\Delta_V)}{2}\Big) \\
&= \mathbb{E}_{U,V} \exp\Big(n \sum_{\substack{1\le i<j\le p \\ 1<|i-j|<T}} u_{ij}v_{ij}\cdot\sigma_{ij}^{*2} + \frac{n}{2} \sum_{\substack{1\le i,j,l\le p \\ i<j \\ 1<|i-l|,|l-j|<T}} u_{il}u_{jl}v_{lj}v_{li}\cdot\sigma_{il}^{*2}\sigma_{lj}^{*2} \\
&+ n \sum_{\substack{1\le i,j,h,l\le p \\ i<j,l<h \\ 1<|i-h|,|h-j|,|i-l|,|l-j|<T}} u_{ih}u_{jl}v_{hj}v_{li}\cdot\sigma_{ih}^{*}\sigma_{jl}^{*}\sigma_{hj}^{*}\sigma_{li}^{*}\Big) + \frac{n}{4}\sum_{\substack{1\le i,l\le p \\ 1<|i-l|<T}} \sigma_{ij}^{*4} + O(np\varphi^{6-\frac{2}{\alpha}}).
\end{aligned}
$$

We explicit the expected value with respect to the i.i.d Rademacher random variables $\{u_{ij}v_{ij}\}_{i<j}$, $\{u_{il}u_{jl}v_{lj}v_{li}\}_{i<j,l\notin\{i,j\}}$ and $\{u_{ih}u_{jl}v_{hj}v_{li}\}_{i<j,l<h}$ pairwise distinct and independent:

$$
\begin{aligned}
H_{n,p} &= \prod_{\substack{i<j \\ 1<|i-j|<T}} \cosh(n\sigma_{ij}^{*2}) \prod_{\substack{1\le i,j,l\le p \\ i<j \\ 1<|i-l|,|l-j|<T}} \cosh(\frac{n}{2}\sigma_{il}^{*2}\sigma_{lj}^{*2}) \\
&\prod_{\substack{1\le i,j,h,l\le p \\ i<j,l<h \\ 1<|i-h|,|h-j|,|i-l|,|l-j|<T}} \cosh\Big(n\sigma_{ih}^{*}\sigma_{jl}^{*}\sigma_{hj}^{*}\sigma_{li}^{*}\Big) \exp\Big(\frac{n}{2}\sum_{\substack{1\le i,l\le p \\ 1<|i-l|<T}} \sigma_{ij}^{*4} + O(np\varphi^{6-\frac{2}{\alpha}})\Big).
\end{aligned}
$$

We use the inequality $\cosh(x)\le\exp(x^2/2)$ and get

$$
\begin{aligned}
H_{n,p} &\le \exp\Big\{\frac{n^2}{2}\Big(\sum_{\substack{i<j \\ 1<|i-j|<T}} \sigma_{ij}^{*4} + \frac{1}{4}\sum_{\substack{1\le i,j,l\le p \\ i<j \\ 1<|i-l|,|l-j|<T}} \sigma_{il}^{*4}\sigma_{lj}^{*4} + \sum_{\substack{1\le i,j,h,l\le p \\ i<j,l<h \\ 1<|i-h|,|h-j|,|i-l|,|l-j|<T}} \sigma_{ih}^{*2}\sigma_{jl}^{*2}\sigma_{hj}^{*2}\sigma_{li}^{*2}\Big)\Big\} \\
&\cdot \exp\Big(\frac{n}{2}\sum_{\substack{1\le i,l\le p \\ 1<|i-l|<T}} \sigma_{ij}^{*4} + O(np\varphi^{6-\frac{2}{\alpha}})\Big).
\end{aligned}
$$

Or, $\displaystyle\frac{n^2}{2}\sum_{\substack{i<j \\ 1<|i-j|<T}} \sigma_{ij}^{*4} = n^2pb^2(\varphi)$ and since $\varphi\to 0$ we have that

$$
\frac{n^2}{8}\sum_{\substack{1\le i,j,l\le p \\ i\ne j \\ 1<|i-l|,|l-j|<T}} \sigma_{il}^{*4}\sigma_{lj}^{*4} = \frac{n^2}{8}\sum_{\substack{i\ne l \\ 1<|i-l|<T}} \sigma_{il}^{*4} \sum_{\substack{j \\ 1<|l-j|<T}} \sigma_{lj}^{*4} = n^2pb^2(\varphi)\cdot O(\lambda^2 T) = n^2pb^2(\varphi)\cdot o(1)
$$

and

$$
\frac{n^2}{2}\sum_{\substack{1\le i,j,h,l\le p \\ i\ne j,l\ne h \\ 1<|i-h|,|h-j|,|i-l|,|l-j|<T}} \sigma_{ih}^{*2}\sigma_{jl}^{*2}\sigma_{hj}^{*2}\sigma_{li}^{*2} = n^2 O(p\lambda^4T^3) = O(n^2p\varphi^{4+\frac{1}{\alpha}}\cdot\varphi^4) = n^2pb^2(\varphi)\cdot o(1).
$$

Finally, $np\varphi^{6-\frac{2}{\alpha}} = n^2p\varphi^{4+\frac{1}{\alpha}}\cdot\dfrac{\varphi^{2-\frac{3}{\alpha}}}{n} = o(1)$ as soon as $n^2p\varphi^{4+\frac{1}{\alpha}}\to 0$ and $\alpha>3/2$ or $5/8<\alpha<3/2$ and $p<n^{\frac{8\alpha-5}{-2\alpha+3}}$.

As consequence, if $n^2pb^2(\varphi) \to 0$ with the additional conditions on α, n and p given previously, we get

$$\mathbb{E}_I\left(\frac{dP_\pi}{dP_I}\right)^2 \leq \exp\left(n^2pb^2(\varphi)(1+o(1))\right) = 1+o(1),$$

which ends the proof of Theorem 3.5.

Now, we show (3.21) in order to finish the proof of Theorem 3.6. More explicitly,

$$\begin{aligned} L_{n,p} &= \log\frac{f_\pi}{f_I}(X_1,...,X_n) \\ &= \log\mathbb{E}_U\exp\left(-\frac{1}{2}\sum_{k=1}^n X_k^\top((\Sigma_U^*)^{-1}-I)X_k - \frac{n}{2}\log\det(\Sigma_U^*)\right), \end{aligned} \tag{3.24}$$

where U is seen as a randomly chosen matrix with uniform distribution over the set $\mathcal{U}$. Let us denote $\Delta_U = \Sigma_U^* - I$ and recall that proposition 3.9 implies that $\|\Delta_U\| \leq O(1)\varphi^{1-\frac{1}{2\alpha}} = o(1)$ for all $\alpha > 1/2$. We write the following approximations obtained by matrix Taylor expansion:

$$-\frac{1}{2}((\Sigma_U^*)^{-1}-I) = \frac{1}{2}\sum_{l=1}^5(-1)^{l+1}\cdot\Delta_U^l + O(1)\Delta_U^6 \tag{3.25}$$

$$\log\det(\Sigma_U^*) = tr\Big(\sum_{l=1}^5\frac{(-1)^{l+1}}{l}\cdot\Delta_U^l + O(1)\Delta_U^6\Big) \tag{3.26}$$

Note that, $tr(\Delta_U) = 0$ and that $tr(\Delta_U^2) = \|\Sigma^* - I\|_F^2 = 2\sum_{1\leq i<j\leq p}\sigma_{ij}^{*2}$ does not depend on U. Moreover,

$$\begin{aligned} \mathbb{E}_I(\sum_{k=1}^n X_k^\top\Delta_U^6 X_k) = n\cdot tr(\Delta_U^6) &\leq n\cdot\|\Delta_U\|^4\cdot tr(\Delta_U^2) \leq O(1)\cdot n\varphi^{4-\frac{2}{\alpha}}\cdot p\varphi^2 \\ &\leq O(1)\cdot n\sqrt{p}\varphi^{2+\frac{1}{2\alpha}}\cdot\sqrt{p}\varphi^{4-\frac{5}{2\alpha}} \\ &\leq O(1)\cdot u_n\cdot\sqrt{p}\varphi^{2-\frac{1}{2\alpha}}\cdot\varphi^{2-\frac{2}{\alpha}} = o(1) \end{aligned}$$

for all $\alpha > 1$ and when $u_n = O(1)$ and $\sqrt{p}\varphi^{2-\frac{1}{2\alpha}} = O(1)$. Also, for all $\alpha > 1$

$$\mathrm{Var}_I(\sum_{k=1}^n X_k^\top\Delta_U^6 X_k) = 2ntr(\Delta_U^{12}) \leq O(1)n\varphi^{10-\frac{5}{\alpha}}\cdot p\varphi^2 = o(1).$$

In conclusion, we use $Y_n = \mathbb{E}_I(Y_n) + O_P(\sqrt{\mathrm{Var}(Y_n)})$ for any sequence of random variables Y_n, to get

$$\sum_{k=1}^n X_k^\top\Delta_U^6 X_k - ntr(\Delta_U^6) = o_P(1), \quad \text{in } P_I\text{-probability}.$$

We get

$$\begin{aligned} L_{n,p} = \log\mathbb{E}_U\exp\Bigg(&\frac{1}{2}\sum_{k=1}^n X_k^\top\Delta_U X_k - \frac{1}{2}\sum_{k=1}^n X_k^\top\Delta_U^2 X_k + \frac{n}{4}tr(\Delta_U^2) \\ &+ \frac{1}{2}\sum_{l=3}^5(-1)^{l+1}\sum_{k=1}^n X_k^\top\Delta_U^l X_k - \frac{n}{2}\sum_{l=3}^5\frac{(-1)^{l+1}}{l}\cdot tr(\Delta_U^l)\Bigg) + o_P(1). \end{aligned} \tag{3.27}$$

From $l = 3, 4$ and 5, we treat similarly the terms

$$\sum_{k=1}^{n} X_k^\top \Delta_U^l X_k = \mathbb{E}_I(\sum_{k=1}^{n} X_k^\top \Delta_U^l X_k) + O_P\Big(\sqrt{\mathrm{Var}_I(\sum_{k=1}^{n} X_k^\top \Delta_U^l X_k)}\Big)$$
$$= ntr(\Delta_U^l) + O_P(1) \cdot \sqrt{ntr(\Delta_U^{2l})} \tag{3.28}$$

By (3.27), we have $ntr(\Delta_U^6) = o(1)$, similarly we obtain $ntr(\Delta_U^{2l}) = o(1)$ for $l = 4$ and 5. Thus (3.27) becomes :

$$L_{n,p} = \log \mathbb{E}_U \exp \Big(\frac{1}{2}\sum_{k=1}^{n} X_k^\top \Delta_U X_k - \frac{1}{2}\sum_{k=1}^{n} X_k^\top \Delta_U^2 X_k + \frac{n}{4} tr(\Delta_U^2)$$
$$+ \frac{n}{2}\sum_{l=3}^{5}(-1)^{l+1}\Big(1-\frac{1}{l}\Big)\cdot tr(\Delta_U^l)\Big) + o_P(1). \tag{3.29}$$

We have

$$tr(\Delta_U^3) = \sum_{\substack{i\neq j\neq k\\ k\neq i}} u_{ij}u_{jk}u_{ki}\sigma_{ij}^*\sigma_{jk}^*\sigma_{ki}^* = 3! \sum_{i<j<k} u_{ij}u_{jk}u_{ki}\sigma_{ij}^*\sigma_{jk}^*\sigma_{ki}^*$$

and we decompose

$$tr(\Delta_U^4) = \sum_{\substack{i\neq j\neq k\\ i\neq l\neq k}} u_{ij}u_{jk}u_{kl}u_{li}\sigma_{ij}^*\sigma_{jk}^*\sigma_{kl}^*\sigma_{li}^*$$
$$= \sum_{i\neq j}\sigma_{ij}^{*4} + 2\sum_{i\neq j\neq k}\sigma_{ij}^{*2}\sigma_{jk}^{*2} + 4!\sum_{i<j<k<l} u_{ij}u_{jk}u_{kl}u_{li}\sigma_{ij}^*\sigma_{jk}^*\sigma_{kl}^*\sigma_{li}^*.$$

Note that

$$n\sum_{i\neq j}\sigma_{ij}^{*4} = O(np\varphi^{4+\frac{1}{\alpha}}) = O(n\sqrt{p}\varphi^{2+\frac{1}{2\alpha}}\cdot\sqrt{p}\varphi^{2+\frac{1}{2\alpha}}) = o(1)$$
$$2n\sum_{i\neq j\neq k}\sigma_{ij}^{*2}\sigma_{jk}^{*2} = O(np\lambda^2T^2) = O(n\sqrt{p}\varphi^{2+\frac{1}{2\alpha}}\cdot\sqrt{p}\varphi^{2-\frac{1}{2\alpha}}) = O(u_n\cdot\sqrt{p}\varphi^{2-\frac{1}{2\alpha}}) = o(1),$$

if $u_n \asymp 1$ and $\sqrt{p}\varphi^{2-\frac{1}{2\alpha}} \to 0$. As for the last term :

$$tr(\Delta_U^5) = \sum_{\substack{i\neq j\neq k\\ k\neq l\neq v\\ v\neq i}} u_{ij}u_{jk}u_{kl}u_{lv}u_{vi}\sigma_{ij}^*\sigma_{jk}^*\sigma_{kl}^*\sigma_{lv}^*\sigma_{vi}^*$$
$$= 5\sum_{\substack{i\neq j\neq k\\ k\neq l\neq j}} u_{jk}u_{kl}u_{lj}\sigma_{ij}^{*2}\sigma_{jk}^*\sigma_{kl}^*\sigma_{lj}^* + 5\sum_{\substack{i\neq j\neq k\\ k\neq i}} u_{ij}^3u_{jk}u_{ki}\sigma_{ij}^{*3}\sigma_{jk}^*\sigma_{ki}^*$$
$$+ 5!\sum_{i<j<k<l<v} u_{ij}u_{jk}u_{kl}u_{lv}u_{vi}\sigma_{ij}^*\sigma_{jk}^*\sigma_{kl}^*\sigma_{lv}^*\sigma_{vi}^*.$$

The first two terms in the decomposition of $tr(\Delta_U^5)$ group with $tr(\Delta_U^3)$ with extra factor $\sum_{j:|j-i|<T}\sigma_{ij}^{*2} + \sigma_{ij}^{*2} = O(\lambda\cdot T) + O(\lambda) = o(1)$, therefore we ignore these terms in further calculations.

Let us denote by $W_{ij} = \sum_{k=1}^{n} X_{k,i}X_{k,j}$, then $\sum_{k=1}^{n} X_k^\top \Delta_U X_k = \sum_{1\leq i\neq j\leq p} u_{ij}\sigma_{ij}^*, W_{ij}$,

$$\sum_{k=1}^{n} X_k^\top \Delta_U^2 X_k = \sum_{1\leq i,j\leq p}[\Delta_U^2]_{ij}W_{ij} = \sum_{1\leq i\neq j\leq p}\sum_{h\notin\{i,j\}} u_{ih}u_{hj}\sigma_{ih}^*\sigma_{hj}^*W_{ij} + \sum_{i=1}^{p}\sum_{h\neq i}\sigma_{ih}^{*2}W_{ii}$$

and $\frac{n}{4}tr(\Delta_U^2) = \frac{n}{4}\sum_{1\le i\ne h\le p}\sigma_{ih}^{*2}$. Then, from (3.29) we get

$$\begin{aligned}
L_{n,p} &= \log \mathbb{E}_U \exp\Big(\frac{1}{2}\sum_{1\le i\ne j\le p} u_{ij}\sigma_{ij}^* W_{ij} - \frac{1}{2}\sum_{1\le i\ne h\le p}\sigma_{ih}^{*2}\Big(W_{ii}-\frac{n}{2}\Big) \\
&- \frac{1}{2}\sum_{1\le i\ne j\ne h\ne i\le p} u_{ih}u_{hj}\sigma_{ih}^*\sigma_{hj}^* W_{ij} + \frac{n}{2}\sum_{l=3}^{5}(-1)^{l+1}\cdot\frac{l-1}{l}\cdot tr(\Delta_U^l)\Big) + o_P(1) \\
&= \log \mathbb{E}_U \exp\Big(\sum_{1\le i<j\le p} u_{ij}\sigma_{ij}^* W_{ij} - \sum_{\substack{1\le i\ne j\ne h\ne i\le p\\ i<j}} u_{ih}u_{hj}\sigma_{ih}^*\sigma_{hj}^* W_{ij} \\
&+ \frac{n}{2}\sum_{l=3}^{5}(-1)^{l+1}\cdot\frac{(l-1)}{l}\cdot l!\sum_{k_1<k_2<\cdots<k_l} u_{k_1k_2}\cdots u_{k_lk_1}\sigma_{k_1k_2}^*\cdots\sigma_{k_lk_1}^*\Big) \\
&- \frac{1}{2}\sum_{1\le i\ne h\le p}\sigma_{ih}^{*2}\Big(W_{ii}-\frac{n}{2}\Big) + o_P(1).
\end{aligned} \tag{3.30}$$

Now, we explicit the expected value with respect to the i.i.d Rademacher random variables $u_{ij}, u_{ih}u_{hj}, u_{k_1k_2}u_{k_2k_3}u_{k_3k_1}, \ldots$ for all $i<j,h,k_1<k_2<\cdots<k_l$ pairwise distinct. Indeed, products of independent Rademacher random variables are still Rademacher and independent. Thus,

$$\begin{aligned}
L_{n,p} &= \sum_{1\le i<j\le p}\log\cosh(\sigma_{ij}^* W_{ij}) + \sum_{\substack{1\le i<j\le p\\ h\notin\{i,j\}}}\log\cosh(\sigma_{ih}^*\sigma_{hj}^* W_{ij}) \\
&+ \sum_{l=3}^{5}\sum_{k_1<\cdots<k_l}\log\cosh\Big(\frac{n(-1)^{l+1}}{2}\cdot(l-1)\cdot(l-1)!\cdot\sigma_{k_1k_2}^*\cdots\sigma_{k_lk_1}^*\Big) \\
&- \frac{1}{2}\sum_{1\le i\ne h\le p}\sigma_{ih}^{*2}(W_{ii}-\frac{n}{2}) + o_P(1).
\end{aligned} \tag{3.31}$$

We shall use repeatedly the Taylor expansion of $\log\cosh(u) = u^2/2 - (u^4/12)(1+o(1))$ as $u\to 0$. Indeed, $\mathbb{E}_I(W_{ij}) = 0$ and $\mathbb{E}_I(|\sigma_{ij}^* W_{ij}|^2) \le O(1)\cdot\lambda n = O(1)\cdot n^{-\frac{1}{(4\alpha+1)}}p^{-\frac{2\alpha+1}{(4\alpha+1)}} = o(1)$, giving that $|\sigma_{ij}^* W_{ij}| = o_P(1)$. Thus

$$\log\cosh(\sigma_{ij}^* W_{ij}) = \frac{1}{2}(\sigma_{ij}^* W_{ij})^2 - \frac{1}{12}(\sigma_{ij}^* W_{ij})^4(1+o_P(1)). \tag{3.32}$$

Similarly, using the first order Taylor expansion, we get

$$\log\cosh(\sigma_{ih}^*\sigma_{hj}^* W_{ij}) = \frac{1}{2}(\sigma_{ih}^*\sigma_{hj}^* W_{ij})^2(1+o_P(1))$$

and for $l = 3, 4$ and 5,

$$\log\cosh\Big(\frac{n(-1)^{l+1}}{2}\cdot(l-1)\cdot(l-1)!\cdot\sigma_{k_1k_2}^*\cdots\sigma_{k_lk_1}^*\Big) = \frac{n^2\Big((l-1)\cdot(l-1)!\Big)^2}{8}\cdot\sigma_{k_1k_2}^{*2}\cdots\sigma_{k_lk_1}^{*2}(1+o(1))$$

Recall now that $\sigma_{ij}^* = 0$, for all i,j such that $|i-j|\ge T$ and $\sigma_{ij}^{*2}\le\lambda = O(1)\varphi^{2+\frac{1}{\alpha}}$. Then,

$$\mathbb{E}\Big(\sum_{\substack{1\le i<j\le p\\ h\notin\{i,j\}}}\sigma_{ih}^{*2}\sigma_{hj}^{*2}W_{ij}^2\Big) = O(np\lambda^2T^2) = O(np\varphi^4) = O(1)\cdot n\sqrt{p}\varphi^{2+\frac{1}{2\alpha}}\cdot\sqrt{p}\varphi^{2-\frac{1}{2\alpha}} = o(1),$$

as soon as $u_n \asymp 1$ and $\sqrt{p}\varphi^{2-\frac{1}{2\alpha}} \to 0$. In conclusion, as the convergence in $\mathbb{L}_1(P_I)$ implies convergence in P_I probability, we get

$$\sum_{\substack{1\le i<j\le p \\ h\notin\{i,j\}}} \log\cosh(\sigma^*_{ih}\sigma^*_{hj}W_{ij}) \longrightarrow 0 \ \text{ in } P_I \text{ probability.} \tag{3.33}$$

Moreover, for $l = 3, 4$ and 5,

$$n^2 \sum_{k_1<\cdots<k_l} \frac{\Big((l-1)\cdot(l-1)!\Big)^2}{8} \cdot \sigma^{*2}_{k_1k_2}\cdots\sigma^{*2}_{k_lk_1} \quad = \quad O(n^2p\lambda^l T^{l-1}) = O(n^2p\varphi^{2l+\frac{1}{\alpha}}) = o(1). \tag{3.34}$$

Using (3.33) and (3.34), (3.31) gives

$$L_{n,p} = \frac{1}{2}\sum_{1\le i<j\le p}\sigma^{*2}_{ij}W^2_{ij} - \frac{1}{12}\sum_{1\le i<j\le p}\sigma^{*4}_{ij}W^4_{ij}(1+o_P(1)) - \frac{1}{2}\sum_{1\le i<j\le p}\sigma^{*2}_{ij}\Big(W_{ii}-\frac{n}{2}\Big) + o_P(1) \tag{3.35}$$

we further decompose as follows :

$$\frac{1}{2}\sum_{1\le i<j\le p}\sigma^{*2}_{ij}W^2_{ij} = \frac{1}{2}\sum_{k=1}^{n}\sum_{1\le i<j\le p}\sigma^{*2}_{ij}X^2_{k,i}X^2_{k,j} + \frac{1}{2}\sum_{1\le k\ne l\le n}\sum_{1\le i<j\le p}\sigma^{*2}_{ij}X_{k,i}X_{k,j}X_{l,i}X_{l,j}$$

With our definition: $\sigma^{*2}_{ij} = 2\cdot w^*_{ij}\cdot b(\varphi)$, we can write

$$\frac{1}{2}\sum_{1\le k\ne l\le n}\sum_{1\le i<j\le p}\sigma^{*2}_{ij}X_{k,i}X_{k,j}X_{l,i}X_{l,j} = n\sqrt{p}b(\varphi)\cdot(n-1)\sqrt{p}\cdot\widehat{\mathcal{D}}_n$$

and we put $Z_n = (n-1)\sqrt{p}\cdot\widehat{\mathcal{D}}_n$ which has asymptotically standard Gaussian under P_I probability, by Proposition 3.1.

By Proposition 3.10 given in the Section 3.7, we have $\mathbb{E}(W^4_{ij}) = 3n^2(1+o(1))$, then

$$\frac{1}{12}\cdot\mathbb{E}_I\Big(\sum_{1\le i<j\le p}\sigma^{*4}_{ij}W^4_{ij}\Big) = \frac{1}{12}\cdot 3n^2\sum_{1\le i<j\le p}\sigma^{*4}_{ij}(1+o(1)) = \frac{u_n^2}{2}(1+o(1)).$$

Moreover,

$$\begin{aligned}
\mathrm{Var}_I\Big(\sum_{i<j}\sigma^{*4}_{ij}W^4_{ij}\Big) &= \sum_{i<j}\sigma^{*8}_{ij}\mathrm{Var}_I(W^4_{ij}) + \sum\sum\sum_{1\le i<j\ne j'\le p}\sigma^{*4}_{ij}\sigma^{*4}_{ij'}\mathrm{Cov}_I(W^4_{ij},W^4_{ij'}) \\
&= O(n^4p\lambda^4T) + O(n^3p\lambda^4T^2) \\
&= O(n^4p\varphi^{8+\frac{3}{\alpha}}) + O(n^3p\varphi^{8+\frac{2}{\alpha}}) = o(1).
\end{aligned}$$

We deduce that,

$$\frac{1}{12}\sum_{1\le i<j\le p}\sigma^{*4}_{ij}W^4_{ij} = \frac{u_n^2}{2} + o_P(1)$$

Remaining terms in (3.35) can be grouped as follows:

$$\frac{1}{4}\sum_{k=1}^{n}\sum_{1\le i\ne j\le p}\sigma^{*2}_{ij}(X^2_{k,i}-1)(X^2_{k,j}-1) = o_P(1)$$

since the random variable in the previous display is centered and

$$\begin{aligned}\mathbb{E}_I\Big(\sum_{k=1}^n \sum_{1\le i\ne j\le p} \sigma_{ij}^{*2}(X_{k,i}^2-1)(X_{k,j}^2-1)\Big)^2 &= \sum_{k=1}^n \sum_{1\le i\ne j\le p} \sigma_{ij}^{*4}\cdot \mathbb{E}_I(X_{k,i}^2-1)^2\mathbb{E}_I(X_{k,j}^2-1)^2 \\ &= 4n\sum_{1\le i<j\le p}\sigma_{ij}^{*4} = O(npb^2(\varphi)) = o(1),\end{aligned}$$

which concludes the proof of (3.21). □

Proof of Theorem 3.8. The type I error probability tends to 0 as a consequence of the Berry-Essen type inequality in Lemma 3.11 in the Section 3.7 applied to the degenerate U-statistic $\widehat{\mathcal{D}}_{n,p}$. We have that, for some $\varepsilon\in(0,1/2)$ and any $t>0$:

$$\Big|\mathbb{P}_I(\widehat{\mathcal{D}}_{n,r}\le t)-\Phi(n\sqrt{p}\cdot t)\Big| \le 16\varepsilon^{1/2}\exp(-\frac{n^2pt^2}{4}) + O\Big(\frac{1}{n}\Big)+O\Big(\frac{1}{pT_r}\Big) \quad \text{for all } 1\le r\le N.$$

We use the relation $1-\Phi(u)\le(1/u)\exp(-u^2/2)$ for all $u\in\mathbb{R}$, to deduce that

$$\mathbb{P}_I(\widehat{\mathcal{D}}_{n,r}>x)\le\Big(\frac{1}{n\sqrt{p}\cdot x}+16\varepsilon^{1/2}\Big)\exp(-\frac{n^2px^2}{4})+O\Big(\frac{1}{n}\Big)+O\Big(\frac{1}{pT_r}\Big)$$

We use this previous result to show that the type I error probability tends to 0. See that for all $r\in\{1,\cdots,N\}$, $C_{\lambda_r}\ge c(\underline{\alpha},\bar{\alpha})$, where $c(\underline{\alpha},\bar{\alpha})=(2\underline{\alpha}+1)/(2\bar{\alpha}(4\bar{\alpha}+1)^{\frac{1}{2\underline{\alpha}}})$. Thus since $n\sqrt{p}\cdot t_r = C_{\lambda_r}\sqrt{\ln\ln(n\sqrt{p})}$, we obtain that $n\sqrt{p}\cdot t_r\ge c(\underline{\alpha},\bar{\alpha})\sqrt{\ln\ln(n\sqrt{p})} =: t$ for all $r\in\{1,\cdots,N\}$. Recall that $N=\lceil\ln(n\sqrt{p})\rceil$, therefore

$$\begin{aligned}&\mathbb{P}_I(\Delta_{ad}^*=1)=\mathbb{P}_I(\exists r\in\{1,\dots,N\};\widehat{\mathcal{D}}_{n,r}>\mathcal{C}^*t_r)\le\sum_{r=1}^N\mathbb{P}_I(\widehat{\mathcal{D}}_{n,r}>\mathcal{C}^*t_r)\\ \le\ &\sum_{r=1}^N\Big\{\Big(\frac{1}{n\sqrt{p}\cdot\mathcal{C}^*t_r}+16\varepsilon^{1/2}\Big)\exp(-\frac{n^2p\cdot\mathcal{C}^{*2}t_r^2}{4})+O\Big(\frac{1}{n}\Big)+O\Big(\frac{1}{pT_r}\Big)\Big\}\\ \le\ & N\Big(\frac{1}{n\sqrt{p}\cdot c(\underline{\alpha},\bar{\alpha})\mathcal{C}^*t}+16\varepsilon^{1/2}\Big)\exp(-\frac{n^2p\cdot c^2(\underline{\alpha},\bar{\alpha})\mathcal{C}^{*2}t^2}{4})+O\Big(\frac{N}{n}\Big)+O\Big(\frac{1}{p}\Big)\sum_{r=1}^N\frac{1}{T_r}\\ \le\ &\Big(\frac{1}{c(\underline{\alpha},\bar{\alpha})\mathcal{C}^*\sqrt{\ln\ln(n\sqrt{p})}}+16\varepsilon^{1/2}\Big)(\ln(n\sqrt{p}))^{1-(c(\underline{\alpha},\bar{\alpha})\mathcal{C}^*/2)^2}\\ +\ &\frac{O(\ln(n\sqrt{p})}{n}+\frac{O(1)}{p}\sum_{r=1}^N\frac{1}{T_r}.\end{aligned}$$

See that :

$$\frac{1}{p}\sum_{r=1}^N\frac{1}{T_r}=\frac{1}{p}\sum_{r=1}^N(\psi_{\alpha_r})^{\frac{1}{\alpha_r}}=\frac{1}{p}\sum_{r=1}^N\frac{1}{p}\Big(\frac{\rho_{n,p}}{n\sqrt{p}}\Big)^{\frac{2}{4\alpha_r+1}}\le\frac{N}{p}\Big(\frac{\rho_{n,p}}{n\sqrt{p}}\Big)^{\frac{2}{4\bar{\alpha}+1}}=o(1).$$

Moreover if $(\ln p)/n=o(1)$, then $\ln(n\sqrt{p})/n=o(1)$, and if $\mathcal{C}^*\ge 2/c(\underline{\alpha},\bar{\alpha})$ we obtain

$$\mathbb{P}_I(\Delta_{ad}^*=1)=o(1).$$

Now, we move to the type II error probability. Let us consider $\Sigma\in\mathcal{F}(\alpha,L)$ such that $(1/2p)\|\Sigma-I\|_F^2=(1/p)\sum_{i<j}\sigma_{ij}^2\ge(\mathcal{C}\psi_\alpha)^2$ for some $\alpha\in\mathcal{A}$. We defined α_{r_0} as the smallest

point on the grid such that $\alpha < \alpha_{r_0}$. We denote by $\widehat{\mathcal{D}}_{n,r_0}$, t_{r_0}, λ_{r_0}, b_{r_0} and T_{r_0} the test statistic, the threshold and the parameters depending on α_{r_0}. Also we define $C_{T_{r_0}}$, $C_{\lambda_{r_0}}$ and $C_{b_{r_0}}$ the constants define in (3.5) for α_{r_0} instead of α and $L = 1$. We have $C_{b_r} < 1$ and $C_{T_r} > 1$, for all $r \in \{1, \cdots N\}$. The type II error probability is bounded from above as follows, $\forall \alpha \in [\underline{\alpha}, \bar{\alpha}]$ and $\forall \Sigma \in Q(\alpha, L, \mathcal{C}\psi_\alpha)$:

$$\begin{aligned}
\mathbb{P}_\Sigma(\Delta^*_{ad} = 0) &= \mathbb{P}_\Sigma\Big(\forall 1 \le r \le N, \widehat{\mathcal{D}}_{n,r} \le \mathcal{C}^* t_r\Big) \le \mathbb{P}_\Sigma\Big(\widehat{\mathcal{D}}_{n,r_0} \le \mathcal{C}^* t_{r_0}\Big)\\
&\le \mathbb{P}_\Sigma\Big(\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_{n,r_0}) - \widehat{\mathcal{D}}_{n,r_0} \ge \mathbb{E}_\Sigma(\widehat{\mathcal{D}}_{n,r_0}) - \mathcal{C}^* t_{r_0}\Big)
\end{aligned}$$

First we have

$$\begin{aligned}
\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_{n,r_0}) &= \frac{1}{p}\sum_{i<j} w^*_{ij,r_0}\sigma^2_{ij} = \frac{1}{p}\cdot\frac{\lambda_{r_0}}{b_{r_0}}\sum_{i<j}\Big(1 - \Big(\frac{|i-j|}{T_{r_0}}\Big)^{2\alpha_{r_0}}\Big)_+ \sigma^2_{ij}\\
&\ge \frac{1}{p}\cdot\frac{\lambda_{r_0}}{b_{r_0}}\Big(\sum_{i<j}\sigma^2_{ij} - \sum_{\substack{i<j\\ |i-j|>T_{r_0}}}\sigma^2_{ij} - \sum_{\substack{i<j\\ |i-j|<T_{r_0}}}\frac{|i-j|^{2\alpha_{r_0}}}{T^{2\alpha_{r_0}}_{r_0}}\cdot\sigma^2_{ij}\Big)\\
&\ge \frac{\lambda_{r_0}}{b_{r_0}}\Big(\mathcal{C}^2\cdot\psi^2_\alpha - \sum_{i<j}\frac{|i-j|^{2\alpha}}{T^{2\alpha}_{r_0}}\cdot\sigma^2_{ij}\Big) \ge \frac{\lambda_{r_0}}{b_{r_0}}\Big(\mathcal{C}^2\cdot\psi^2_\alpha - L\cdot T^{-2\alpha}_{r_0}\Big)\\
&\ge C_{\lambda_{r_0}}\cdot(C_{b_{r_0}})^{-\frac{1}{2}}(\psi_{\alpha_{r_0}})^{\frac{1}{2\alpha_{r_0}}}\Big(\mathcal{C}^2\cdot\psi^2_\alpha - (C_{T_{r_0}})^{-2\alpha}\cdot(\psi_{\alpha_{r_0}})^{\frac{4\alpha}{2\alpha_{r_0}}}\Big)\\
&\ge C_{\lambda_{r_0}}\Big(\mathcal{C}\cdot(\psi_{\alpha_{r_0}})^{\frac{1}{2\alpha_{r_0}}}\cdot\psi^2_\alpha - (C_{T_{r_0}})^{-2\alpha}\cdot(\psi_{\alpha_{r_0}})^{\frac{4\alpha+1}{2\alpha_{r_0}}}\Big) =: (E_1 - E_2).
\end{aligned}$$

Now we show that, since $\alpha < \alpha_{r_0}$ we have,

$$\begin{aligned}
E_1\cdot t^{-1}_{r_0} &= \mathcal{C}\cdot(\psi_{\alpha_{r_0}})^{\frac{1}{2\alpha_{r_0}}}\cdot(\psi_\alpha)^2\cdot((n\sqrt{p})/\rho_{n,p})\\
&= ((n\sqrt{p})/\rho_{n,p})^{\frac{4(\alpha_{r_0}-\alpha)}{(4\alpha_{r_0}+1)(4\alpha+1)}} \ge \mathcal{C}
\end{aligned}$$

Moreover, use that $0 > \alpha - \alpha_{r_0} \ge -(\bar{\alpha} - \underline{\alpha})/\ln(n\sqrt{p})$, to obtain

$$\begin{aligned}
t_{r_0}\cdot E^{-1}_2 &= (\rho_{n,p}/(n\sqrt{p}))\cdot(C_{T_{r_0}})^{2\alpha}\cdot(\psi_{\alpha_{r_0}})^{-\frac{4\alpha+1}{2\alpha_{r_0}}} \ge ((n\sqrt{p})/\rho_{n,p})^{\frac{4(\alpha-\alpha_{r_0})}{(4\alpha_{r_0}+1)}}\\
&= \exp\Big\{\frac{4(\alpha-\alpha_{r_0})}{4\alpha_{r_0}+1}\cdot\ln((n\sqrt{p})/\rho_{n,p})\Big\}\\
&\ge \exp\Big\{-\frac{4(\bar{\alpha}-\underline{\alpha})}{4\underline{\alpha}+1}(1+o(1))\Big\} \ge C(\underline{\alpha}, \bar{\alpha}).
\end{aligned}$$

We deduce that,

$$\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_{n,r_0}) \ge \Big(\mathcal{C} - \frac{1}{C(\underline{\alpha}, \bar{\alpha})}\Big)\cdot t_{r_0}.$$

Let us denote by $\mathcal{T}_1$ and $\mathcal{T}_2$ the right-hand side termes in (3.7) and (3.8), respectively.

Then by Markov inequality, for $\mathcal{C} - \dfrac{1}{C(\underline{\alpha}, \bar{\alpha})} - \mathcal{C}^* > 0$, we get

$$\begin{aligned}
\mathbb{P}_\Sigma(\Delta^*_{ad} = 0) &\leq \mathbb{P}_\Sigma\Big(|\widehat{\mathcal{D}}_{n,r_0} - \mathbb{E}_\Sigma(\widehat{\mathcal{D}}_{n,r_0})| \geq \mathbb{E}_\Sigma(\widehat{\mathcal{D}}_{n,r_0}) - \mathcal{C}^* t_{r_0} \Big) \\
&\leq \frac{\mathrm{Var}_\Sigma(\widehat{\mathcal{D}}_{n,r_0})}{\Big(\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_{n,r_0}) - \mathcal{C}^* t_{r_0}\Big)^2} \leq \frac{\Big(\mathcal{C} - \frac{1}{C(\underline{\alpha}, \bar{\alpha})}\Big)^2 \mathrm{Var}_\Sigma(\widehat{\mathcal{D}}_{n,r_0})}{\Big(\mathcal{C} - \frac{1}{C(\underline{\alpha}, \bar{\alpha})} - \mathcal{C}^*\Big)^2 \mathbb{E}^2_\Sigma(\widehat{\mathcal{D}}_{n,r_0})} \\
&\leq \frac{\Big(\mathcal{C} - \frac{1}{C(\underline{\alpha}, \bar{\alpha})}\Big)^2 \cdot (\mathcal{T}_1 + (n-1)\mathcal{T}_2)}{n(n-1)p^2 \Big(\mathcal{C} - \frac{1}{C(\underline{\alpha}, \bar{\alpha})} - \mathcal{C}^*\Big)^2 \mathbb{E}^2_\Sigma(\widehat{\mathcal{D}}_{n,r_0})} := F_1 + F_2.
\end{aligned}$$

We use (3.7) to show that F_1 tends to zero.

$$\begin{aligned}
F_1 &:= \frac{\Big(\mathcal{C} - \frac{1}{C(\underline{\alpha}, \bar{\alpha})}\Big)^2 \cdot \mathcal{T}_1}{n(n-1)p^2 \Big(\mathcal{C} - \frac{1}{C(\underline{\alpha}, \bar{\alpha})} - \mathcal{C}^*\Big)^2 \mathbb{E}^2_\Sigma(\widehat{\mathcal{D}}_{n,r_0})} \\
&\leq \frac{1 + o(1)}{n(n-1)p \Big(\mathcal{C} - \frac{1}{C(\underline{\alpha}, \bar{\alpha})} - \mathcal{C}^*\Big)^2 t^2_{r_0}} + \frac{O(T^{\frac{3}{2}}_{r_0} \cdot t_{r_0})}{n(n-1)p t^2_{r_0}} = O(\rho^{-2}_{n,p}) = o(1),
\end{aligned}$$

since $T^{\frac{3}{2}}_{r_0} \cdot t_{r_0} = O\Big((\rho_{n,p}/n\sqrt{p})^{-\frac{3}{4\alpha_{r_0}+1}+1} \Big) = o(1)$ for $\alpha_{r_0} > 1/2$. Similarly we use (3.8) to show that

$$F_2 := \frac{\Big(\mathcal{C} - \frac{1}{C(\underline{\alpha}, \bar{\alpha})}\Big)^2 \cdot \mathcal{T}_2}{np^2 \Big(\mathcal{C} - \frac{1}{C(\underline{\alpha}, \bar{\alpha})} - \mathcal{C}^*\Big)^2 \mathbb{E}^2_\Sigma(\widehat{\mathcal{D}}_{n,r_0})} = o(1).$$

Thus we get, for $\mathcal{C} - \dfrac{1}{C(\underline{\alpha}, \bar{\alpha})} - \mathcal{C}^* > 0$,

$$\sup_{\alpha \in [\underline{\alpha}, \bar{\alpha}]} \sup_{\substack{\Sigma \in \mathcal{F}(\alpha, L); \\ \frac{1}{2p}\|\Sigma - I\|^2_F \geq \mathcal{C}^2 \psi^2_\alpha}} \mathbb{P}_\Sigma(\Delta^*_{ad} = 0) = o(1).$$

□

3.7 Additional proofs

Proof of Proposition 3.1. We recall that under the null hypothesis the coordinates of the vector X_k are independent, so using this fact we have :

$$\begin{aligned}\mathrm{Var}_I(\widehat{\mathcal{D}}_n) &= \frac{2}{n^2(n-1)^2p^2}\mathrm{Var}(\sum_{\substack{i=1\\i<j}}^{p}\sum_{j=1}^{p}\sum_{l=1}^{n}\sum_{\substack{k=1\\k\neq l}}^{n} w_{ij}^* X_{k,i}X_{k,j}X_{l,i}X_{l,j})\\ &= \frac{2}{n(n-1)p^2}\sum_{\substack{i=1\\i<j}}^{p}\sum_{j=1}^{p} w_{ij}^{*2}\mathbb{E}^4(X_{1,i}^2) = \frac{2}{n(n-1)p^2}\sum_{\substack{i=1\\i<j}}^{p}\sum_{j=1}^{p} w_{ij}^{*2} = \frac{1}{n(n-1)p}\end{aligned}$$

For $\Sigma \in Q(\alpha, L, \varphi)$,

$$\begin{aligned}\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n) &= \frac{1}{n(n-1)p}\sum_{\substack{i=1\\i<j}}^{p}\sum_{j=1}^{p}\sum_{l=1}^{n}\sum_{\substack{k=1\\k\neq l}}^{n} w_{ij}^*\mathbb{E}(X_{k,i}X_{k,j}X_{l,i}X_{l,j})\\ &= \frac{1}{p}\sum_{\substack{i=1\\i<j}}^{p}\sum_{j=1}^{p} w_{ij}^*\mathbb{E}(X_{1,i}X_{1,j})\mathbb{E}(X_{2,i}X_{2,j}) = \frac{1}{p}\sum_{\substack{i=1\\i<j}}^{p}\sum_{j=1}^{p} w_{ij}^*\sigma_{ij}^2\end{aligned}$$

Remark that $\widehat{\mathcal{D}}_n - \mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n)$ can be written as the following form

$$\begin{aligned}\widehat{\mathcal{D}}_n - \mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n) &= \frac{1}{n(n-1)p}\sum_{l=1}^{n}\sum_{\substack{k=1\\k\neq l}}^{n}\sum_{\substack{i=1\\i<j}}^{p}\sum_{j=1}^{p} w_{ij}^*(X_{k,i}X_{k,j}-\sigma_{ij})(X_{l,i}X_{l,j}-\sigma_{ij})\\ &\quad+\frac{2}{np}\sum_{k=1}^{n}\sum_{\substack{i=1\\i<j}}^{p}\sum_{j=1}^{p} w_{ij}^*(X_{k,i}X_{k,j}-\sigma_{ij})\sigma_{ij}\end{aligned}\tag{3.36}$$

Then the variance of the estimator $\widehat{\mathcal{D}}_n$ is a sum of two uncorrelated terms

$$\begin{aligned}\mathrm{Var}_\Sigma(\widehat{\mathcal{D}}_n) &= \frac{2}{n(n-1)p^2}\mathbb{E}_\Sigma\{\sum_{\substack{i=1\\i<j}}^{p}\sum_{j=1}^{p} w_{ij}^*(X_{1,i}X_{1,j}-\sigma_{ij})(X_{2,i}X_{2,j}-\sigma_{ij})\}^2\\ &+ \frac{4}{np^2}\mathbb{E}_\Sigma\{\sum_{\substack{i=1\\i<j}}^{p}\sum_{j=1}^{p} w_{ij}^*(X_{k,i}X_{k,j}-\sigma_{ij})\sigma_{ij}\}^2\end{aligned}\tag{3.37}$$

Now we will give an upper bound for the first term on the right-hand side of (3.37). Denote by

$$\begin{aligned}T_1 &= 2\mathbb{E}_\Sigma\{\sum_{\substack{i=1\\i<j}}^{p}\sum_{j=1}^{p} w_{ij}^*(X_{1,i}X_{1,j}-\sigma_{ij})(X_{2,i}X_{2,j}-\sigma_{ij})\}^2\\ &= \frac{1}{2}\sum_{\substack{i=1\\i\neq j}}^{p}\sum_{j=1}^{p}\sum_{\substack{i'=1\\i'\neq j'}}^{p}\sum_{j'=1}^{p} w_{ij}^*w_{i'j'}^*\mathbb{E}_\Sigma^2\{(X_{1,i}X_{1,j}-\sigma_{ij})(X_{1,i'}X_{1,j'}-\sigma_{i'j'})\}\\ &= \frac{1}{2}\sum_{\substack{i=1\\i\neq j}}^{p}\sum_{j=1}^{p}\sum_{\substack{i'=1\\i'\neq j'}}^{p}\sum_{j'=1}^{p} w_{ij}^*w_{i'j'}^*(\sigma_{ii'}\sigma_{jj'}+\sigma_{ij'}\sigma_{i'j})^2\end{aligned}$$

We shall distinguish three terms in the previous sum, that is $(i,j,i',j') \in A_1 \cup A_2 \cup A_3$, where A_1, A_2, A_3 form a partition of the set$\{(i,j,i',j') \in \{1,\dots,p\}^4$ such that $i \neq j, i' \neq j'\}$. More precisely in A_1 we have $(i,j) = (i',j')$ or $(i,j) = (j',i')$, in A_2 we have three different indices ($i = i'$ and $j \neq j'$) or ($j = j'$ and $i \neq i'$) or ($i = j'$ and $j \neq i'$) or ($j = i'$ and $i \neq j'$) and finally in A_3 the indices are pairewise distinct. First, when $(i, j, i', j') \in A_1$, we use that $\mathrm{Var}_\Sigma(X_{1,i}X_{1,j}) = (1 + \sigma_{ij}^2)^2$, to get

$$\begin{aligned} T_{1,1} &= \sum_{\substack{i=1 \\ i \neq j}}^{p} \sum_{j=1}^{p} w_{ij}^{*2}(1+\sigma_{ij}^2)^2 = \sum_{\substack{i=1 \\ i \neq j}}^{p} \sum_{j=1}^{p} w_{ij}^{*2} + \sum_{\substack{i=1 \\ i \neq j}}^{p} \sum_{j=1}^{p} w_{ij}^{*2}(2\sigma_{ij}^2 + \sigma_{ij}^4) \\ &\leq p + 3 \sum_{\substack{i=1 \\ i \neq j}}^{p} \sum_{j=1}^{p} w_{ij}^{*2}\sigma_{ij}^2 \leq p + 6 \cdot p \cdot L \cdot \sup_{i,j} w_{ij}^{*2} \end{aligned} \tag{3.38}$$

and this is $p(1 + o(1))$ since $\sup_{i,j} w_{ij}^{*2} \asymp (1/T) \to 0$. When the indices are in A_2, we have three indices out of four which are equal. We assume $i = i'$, therefore it is sufficient to check that,

$$\begin{aligned} T_{1,2} &= 2 \sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \neq i}}^{p} \sum_{\substack{j'=1 \\ j' \neq i \\ j \neq j'}}^{p} w_{ij}^{*} w_{ij'}^{*} (\sigma_{jj'} + \sigma_{ij}\sigma_{ij'})^2 \\ &\leq 4 \sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \neq i}}^{p} \sum_{\substack{j'=1 \\ j' \neq i \\ j \neq j'}}^{p} w_{ij}^{*} w_{ij'}^{*} \sigma_{jj'}^2 + 4 \sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \neq i}}^{p} \sum_{\substack{j'=1 \\ j' \neq i \\ j \neq j'}}^{p} w_{ij}^{*} w_{ij'}^{*} \sigma_{ij}^2 \sigma_{ij'}^2 \end{aligned}$$

Now let us bound from above the first term of $T_{1,2}$,

$$T_{1,2,1} := \sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \neq i}}^{p} \sum_{\substack{j'=1 \\ j' \neq i \\ j \neq j'}}^{p} w_{ij}^{*} w_{ij'}^{*} \sigma_{jj'}^2 \leq \sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \neq i}}^{p} \sum_{\substack{j'=1 \\ j' \neq i \\ |j-j'| < T}}^{p} w_{ij}^{*} w_{ij'}^{*} \sigma_{jj'}^2 + \sum_{\substack{j=1 \\ j \neq i}}^{p} \sum_{\substack{j'=1 \\ j' \neq i \\ |j-j'| \geq T}}^{p} \frac{|j-j'|^{2\alpha}}{T^{2\alpha}} \sigma_{jj'}^2 \sum_{i=1}^{p} w_{ij}^{*} w_{ij'}^{*} \tag{3.39}$$

Again we will treat each term of $T_{1,2,1}$ separately. We recall that the weights w_{ij}^{*} verify the following properties

$$(w_{ij}^{*} \geq w_{i'j'}^{*} \quad \text{for } |i - j| \leq |i' - j'|) \qquad \text{and} \qquad \sum_{i=1}^{p} w_{ij}^{*} \asymp \sqrt{T}.$$

In the rest of the proof we denote by $k_0(\alpha, L), k_1(\alpha, L), \ldots$ different constants that dependent only on α and/or on L. We have for $\alpha > 1/2$,

$$\begin{aligned}
T_{1,2,1,1} &:= \sum_{i=1}^{p}\sum_{\substack{j=1\\ j\neq i}}^{p}\sum_{\substack{j'=1\\ j'\neq i\\ |j-j'|<T}}^{p} w_{ij}^* w_{ij'}^* \sigma_{jj'}^2 = \sum_{i=1}^{p} \sum_{\substack{j=1\\ j\neq i}}^{p}\sum_{\substack{j'=1\\ j'\neq i\\ |j-j'|\le|i-j|<T}}^{p} w_{ij}^* w_{ij'}^* \sigma_{jj'}^2 + \sum_{i=1}^{p} \sum_{\substack{j=1\\ j\neq i}}^{p}\sum_{\substack{j'=1\\ j'\neq i\\ |i-j|<|j-j'|<T}}^{p} w_{ij}^* w_{ij'}^* \sigma_{jj'}^2 \\
&\le \sum_{\substack{j=1\\ j\neq i}}^{p}\sum_{\substack{j'=1\\ j'\neq i\\ |j-j'|<|i-j|<T}}^{p} w_{jj'}^* \sigma_{jj'}^2 \sum_{i=1}^{p} w_{ij'}^* + \sum_{i=1}^{p} \sum_{\substack{j=1\\ j\neq i}}^{p}\sum_{\substack{j'=1\\ j'\neq i\\ |i-j|<|j-j'|<T}}^{p} \frac{w_{ij}^* w_{ij'}^* |j-j'|^{2\alpha}}{|i-j|^{2\alpha}} \sigma_{jj'}^2 \\
&\le k_0(\alpha, L)\cdot\sqrt{T}\cdot p\cdot \mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n) + (\sup_{i,j} w_{ij}^*)^2 \sum_{j=1}^{p}\sum_{\substack{j'=1\\ j\neq j'}}^{p} |j-j'|^{2\alpha}\sigma_{jj'}^2 (\sum_{i=1}^{p} \frac{1}{|i-j|^{2\alpha}}) \\
&\le k_0(\alpha, L)\cdot\sqrt{T}\cdot p\cdot \mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n) + k_1(\alpha, L)\cdot L\cdot p\cdot (\sup_{i,j} w_{ij}^*)^2 \\
&\le p\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n) O(\sqrt{T}) + o(p).
\end{aligned} \tag{3.40}$$

For the second term in (3.39), where $|j-j'|\ge T$, we use the following bound:

$$\sum_{\substack{i=1\\ i\neq j,j'}}^{p} w_{ij}^* w_{ij'}^* \le \sum_{\substack{i=1\\ i\neq j,j'}}^{p} (w_{ij}^*)^2 \le \frac{1}{2},$$

then we prove that,

$$T_{1,2,1,2} := \sum_{\substack{j=1\\ j\neq i}}^{p}\sum_{\substack{j'=1\\ j'\neq i\\ |j-j'|\ge T}}^{p} \frac{|j-j'|^{2\alpha}}{T^{2\alpha}}\sigma_{jj'}^2 \sum_{i=1}^{p} w_{ij}^* w_{ij'}^* \le \frac{L\cdot p}{T^{2\alpha}} = O(\frac{p}{2T^{2\alpha}}) = o(p). \tag{3.41}$$

Note that $\sup_{i,j}\sigma_{ij}\le 1$. The second term of $T_{1,2}$, is bounded as follows:

$$\begin{aligned}
T_{1,2,2} &:= \sum_{i=1}^{p}\sum_{\substack{j=1\\ j\neq i}}^{p}\sum_{\substack{j'=1\\ j'\neq i\\ j\neq j'}}^{p} w_{ij}^* w_{ij'}^* \sigma_{ij}^2\sigma_{ij'}^2 = \sum_{i=1}^{p}\Big(\sum_{\substack{j=1\\ j\neq i}}^{p} w_{ij}^*\sigma_{ij}^2\Big)\Big(\sum_{\substack{j'=1\\ j'\neq i}}^{p} w_{ij'}^*\sigma_{ij'}^2\Big) \\
&\le (\sup_{i,j} w_{ij}^*)\sup_i\Big(\sum_{\substack{j=1\\ 1\le|j-i|<T}}^{p} \sigma_{ij}^2\Big)\Big(\sum_{i=1}^{p}\sum_{\substack{j'=1\\ j'\neq i}}^{p} w_{ij'}^*\sigma_{ij'}^2\Big) \\
&\le 2L\cdot(\sup_{i,j} w_{ij}^*)\cdot T\cdot p\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n) \le p\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n)\cdot O(\sqrt{T})
\end{aligned} \tag{3.42}$$

As a consequence of (3.40) to (3.42),

$$T_{1,2} \le p\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n)\cdot O(\sqrt{T}) + o(p) \tag{3.43}$$

The last case, where (i,j,i',j') vary in A_3, the indices are pairwise distinct,

$$\begin{aligned}
T_{1,3} &= \sum_{(i,j,i',j')\in A_3} w_{ij}^* w_{i'j'}^* (\sigma_{ii'}\sigma_{jj'} + \sigma_{ij'}\sigma_{i'j})^2 \\
&\le 2\sum_{(i,j,i',j')\in A_3} w_{ij}^* w_{i'j'}^* \sigma_{ii'}^2\sigma_{jj'}^2 + 2\sum_{(i,j,i',j')\in A_3} w_{ij}^* w_{i'j'}^* \sigma_{ij'}^2\sigma_{i'j}^2
\end{aligned}$$

As the two previous terms have the same upper bound, let us deal with the first one say $T_{1,3,1}$. We should distinguish two cases, the first when $|i-i'|<T$ and the second when $|i-i'|\geq T$. We begin by the first case, which in turn will be decomposed into three terms. First,

$$\begin{aligned} T_{1,3,1,1} &:= \sum_{\substack{(i,j,i',j')\in A_3 \\ |i-j|\geq|i-i'|,|i'-j'|\geq|i-i'|}} w^*_{ij}w^*_{i'j'}\sigma^2_{ii'}\sigma^2_{jj'} \leq \sum_{\substack{(i,j,i',j')\in A_3 \\ |i-j|\geq|i-i'|,|i'-j'|\geq|i-i'|}} w^{*2}_{ii'}\sigma^2_{ii'}\sigma^2_{jj'} \\ &\leq (\sup_{ij} w^*_{ij}) \sum_{1\leq i,i'\leq p} w^*_{ii'}\sigma^2_{ii'} \sum_{\substack{1\leq j,j'\leq p \\ 1<|i-j|,|i'-j'|<T}} \sigma^2_{jj'} \leq (\sup_{ij} w^*_{ij}) \cdot T^2 \cdot p \cdot \mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n) \end{aligned} \tag{3.44}$$

Then,

$$\begin{aligned} T_{1,3,1,2} &:= \sum_{\substack{(i,j,i',j')\in A_3 \\ |i-j|<|i-i'|<T,|i'-j'|\geq|i-i'|}} w^*_{ij}w^*_{i'j'}\sigma^2_{ii'}\sigma^2_{jj'} \leq \sum_{\substack{(i,j,i',j')\in A_3 \\ |i-j|<|i-i'|<T,|i'-j'|\geq|j-j'|}} w^*_{ij}w^*_{ii'}\sigma^2_{ii'}\sigma^2_{jj'} \\ &\leq (\sup_{ij} w^*_{ij}) \cdot T^2 \cdot p \cdot \mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n) \leq k_2(\alpha,L)\cdot T\sqrt{T}\cdot p\cdot \mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n) \end{aligned} \tag{3.45}$$

Finally, using Cauchy-Schwarz inequality, we have,

$$\begin{aligned} T_{1,3,1,3} &:= \sum_{\substack{(i,j,i',j')\in A_3 \\ |i-j|<|i-i'|<T,|i'-j'|<|i-i'|<T}} w^*_{ij}w^*_{i'j'}\sigma^2_{ii'}\sigma^2_{jj'} \\ &= \sum_{\substack{(i,j,i',j')\in A_3 \\ |i-j|<|i-i'|<T,|i'-j'|<|j-j'|<T}} w^*_{ij}w^*_{i'j'}\cdot\frac{|i-i'|^{2\alpha}}{|i-j|^\alpha|i'-j'|^\alpha}\cdot\sigma^2_{ii'}\sigma^2_{jj'} \\ &\leq (\sup_{i,j} w^*_{ij})^2 \sum_{i=1}^p \sum_{\substack{i'=1 \\ i'\neq i'}}^p |i-i'|^{2\alpha}\sigma^2_{ii'} \sum_{\substack{1\leq\ j,j'\ \leq p \\ 1\leq|i-j|,|i'-j'|<T}} \frac{\sigma^2_{jj'}}{|i-j|^\alpha|i'-j'|^\alpha} \\ &\leq k_3(\alpha,L)\cdot T^{-1}\cdot 2pL\cdot\max\{1,T^{-2\alpha+2}\} = o(p) \quad \text{for } \alpha>\frac{1}{2}. \end{aligned} \tag{3.46}$$

Now we suppose that we have $|i-i'|>T$, then,

$$\begin{aligned} T_{1,3,2} &:= \sum_{\substack{(i,j,i',j')\in A_3 \\ |i-i'|>T}} w^*_{ij}w^*_{i'j'}\sigma^2_{ii'}\sigma^2_{jj'} = \sum_{\substack{(i,j,i',j')\in A_3 \\ |i-i'|>T}} w^*_{ij}w^*_{i'j'}\frac{|i-i'|^{2\alpha}}{T^{2\alpha}}\sigma^2_{ii'}\sigma^2_{jj'} \\ &\leq \frac{(\sup_{i,j} w^*_{ij})^2}{T^{2\alpha}} \sum_{1\leq i,i'\leq p} |i-i'|^{2\alpha}\sigma^2_{ii'} \sum_{\substack{1\leq j,j'\leq p \\ 1\leq|i-j|,|i'-j'|<T}} \sigma^2_{jj'} \\ &\leq \frac{(\sup_{i,j} w^*_{ij})^2}{T^{2\alpha}}\cdot 2pL\cdot T^2 \leq \frac{k_4(\alpha,L)\cdot p}{T^{2\alpha-1}} = o(p) \quad \text{for } \alpha>\frac{1}{2}. \end{aligned} \tag{3.47}$$

Finally we obtain, from (3.44) to (3.47) :

$$T_{1,3} \leq p\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n)\cdot O(T\sqrt{T}) + o(p). \tag{3.48}$$

Put together (3.38), (3.43) and (3.48) to obtain (3.7). Let us give an upper bound for the second term of (3.37),

$$\begin{aligned}
T_2 &= 4\mathbb{E}_\Sigma\{\sum_{\substack{i=1\\i<j}}^{p}\sum_{j=1}^{p} w^*_{ij}(X_{k,i}X_{k,j}-\sigma_{ij})\sigma_{ij}\}^2 \\
&= \sum_{\substack{i=1\\i\neq j}}^{p}\sum_{j=1}^{p}\sum_{\substack{i'=1\\i'\neq j'}}^{p}\sum_{j'=1}^{p} w^*_{ij}w^*_{i'j'}\sigma_{ij}\sigma_{i'j'}\mathbb{E}_\Sigma(X_{1,i}X_{1,j}-\sigma_{ij})(X_{1,i'}X_{1,j'}-\sigma_{i'j'}) \\
&= \sum_{\substack{i=1\\i\neq j}}^{p}\sum_{j=1}^{p}\sum_{\substack{i'=1\\i'\neq j'}}^{p}\sum_{j'=1}^{p} w^*_{ij}w^*_{i'j'}\sigma_{ij}\sigma_{i'j'}(\sigma^*_{ii'}\sigma^*_{jj'}+\sigma^*_{ij'}\sigma^*_{i'j})
\end{aligned}$$

Proceeding similarly, we shall distinguish three kind of terms. Let us begin by the case when the indices belong to A_1,

$$\begin{aligned}
T_{2,1} &= 2\sum_{\substack{i=1\\i\neq j}}^{p}\sum_{j=1}^{p} w^{*2}_{ij}\sigma^2_{ij}\mathbb{E}_\Sigma[(X_{1,i}X_{1,j}-\sigma_{ij})^2] = 2\sum_{\substack{i=1\\i\neq j}}^{p}\sum_{j=1}^{p} w^{*2}_{ij}\sigma^2_{ij}(1+\sigma^2_{ij}) \\
&\leq 4(\sup_{i,j} w^*_{ij})\sum_{\substack{i=1\\i\neq j}}^{p}\sum_{j=1}^{p} w^*_{ij}\sigma^2_{ij} = 8(\sup_{i,j} w^*_{ij})\cdot p\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n) \\
&= o(1)\cdot p\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n). \qquad (3.49)
\end{aligned}$$

Next, when $(i,j,i',j')\in A_2$,

$$\begin{aligned}
T_{2,2} &= 4\sum_{i=1}^{p}\sum_{\substack{j=1\\j\neq i}}^{p}\sum_{\substack{j'=1\\j'\neq i}}^{p} w^*_{ij}w^*_{ij'}\sigma_{ij}\sigma_{ij'}(\sigma_{jj'}+\sigma_{ij}\sigma_{ij'}) \\
&= 4\sum_{i=1}^{p}\sum_{\substack{j=1\\j\neq i}}^{p}\sum_{\substack{j'=1\\j'\neq i}}^{p} w^*_{ij}w^*_{ij'}\sigma_{ij}\sigma_{ij'}\sigma_{jj'} + 4\sum_{i=1}^{p}\sum_{\substack{j=1\\j\neq i}}^{p}\sum_{\substack{j'=1\\j'\neq i}}^{p} w^*_{ij}w^*_{ij'}\sigma^2_{ij}\sigma^2_{ij'}
\end{aligned}$$

We bound from each term of $T_{2,2}$ separately. Using Cauchy-Schwarz inequality two times we obtain,

$$\begin{aligned}
T_{2,2,1} &:= \sum_{i=1}^{p}\sum_{\substack{j=1\\j\neq i}}^{p}\sum_{\substack{j'=1\\j'\neq i}}^{p} w^*_{ij}w^*_{ij'}\sigma_{ij}\sigma_{ij'}\sigma_{jj'} \leq \sum_{i=1}^{p}\sum_{\substack{j=1\\j\neq i}}^{p} w^*_{ij}\sigma_{ij}\Big(\sum_{\substack{j'=1\\j'\neq i}}^{p} w^{*2}_{ij'}\sigma^2_{ij'}\Big)^{1/2}\Big(\sum_{\substack{j'=1\\j'\neq i}}^{p}\sigma^2_{jj'}\Big)^{1/2} \\
&\leq \Big(\sum_{i=1}^{p}\sum_{\substack{j=1\\j\neq i}}^{p} w^{*2}_{ij}\sigma^2_{ij}\Big)^{1/2}\Big(\sum_{i=1}^{p}\sum_{\substack{j=1\\j\neq i}}^{p}(\sum_{\substack{j'=1\\j'\neq i}}^{p} w^{*2}_{ij'}\sigma^2_{ij'})(\sum_{\substack{j'=1\\j'\neq i}}^{p}\sigma^2_{jj'})\Big)^{1/2} \\
&\leq (\sup_{i,j} w^*_{ij})\cdot p\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n)\cdot O(T) = O(\sqrt{T})\cdot p\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n).
\end{aligned}$$

The second term in $T_{2,2}$ is $T_{1,2,2}$ and therefore,

$$T_{2,2} = O(\sqrt{T})\cdot p\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n). \qquad (3.50)$$

Finally, when $(i,j,i',j') \in A_3$, we have to bound from above

$$T_{2,3} = \sum_{(i,j,i',j')\in A_3} w^*_{ij} w^*_{i'j'} \sigma_{ij}\sigma_{i'j'}\sigma^*_{ii'}\sigma^*_{jj'} + \sum\sum_{(i,j,i',j')\in A_3} w^*_{ij} w^*_{i'j'} \sigma_{ij}\sigma_{i'j'}\sigma^*_{ij'}\sigma^*_{i'j}.$$

These last two terms, in $T_{2,3}$, are treated similarly, so let us deal with :

$$\begin{aligned}
&\sum_{(i,j,i',j')\in A_3} w^*_{ij} w^*_{i'j'} \sigma_{ij}\sigma_{i'j'}\sigma^*_{ii'}\sigma^*_{jj'} \\
\le\ & \sum_j \sum_{i'} \Big(\sum_i w^*_{ij}\sigma^2_{ii'}\Big)^{1/2} \Big(\sum_i w^*_{ij}\sigma^2_{ij}\Big)^{1/2} \Big(\sum_{j'} w^*_{i'j'}\sigma^2_{i'j'}\Big)^{1/2} \Big(\sum_{j'} w^*_{i'j'}\sigma^2_{jj'}\Big)^{1/2} \\
\le\ & \Big(\sum_j \sum_{i'} (\sum_i w^*_{ij}\sigma^2_{ij})(\sum_{j'} w^*_{i'j'}\sigma^2_{i'j'})\Big)^{1/2} \Big(\sum_{(i,j,i',j')\in A_3} w^*_{ij} w^*_{i'j'} \sigma^2_{jj'}\sigma^2_{ii'}\Big)^{1/2} \\
\le\ & p \cdot \mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n) \cdot \Big(\sum_{(i,j,i',j')\in A_3} w^*_{ij} w^*_{i'j'} \sigma^2_{jj'}\sigma^2_{ii'}\Big)^{1/2}
\end{aligned}$$

Using the upper bound of $T_{1,3}$ obtained previously, we have

$$T_{2,3} \le p\sqrt{p} \cdot \Big(\mathbb{E}^{3/2}_\Sigma(\widehat{\mathcal{D}}_n) \cdot O(T^{3/4}) + \mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n) \cdot o(1)\Big) \tag{3.51}$$

Put together (3.49), (3.50) and (3.51) to get (3.8).

The asymptotic normality under the null hypothesis is obvious. □

Proof of Proposition 3.2. We use the decomposition (3.36) in the proof of the Proposition 3.1 and we treat each term separately. Recall that, by our assumptions, $n\sqrt{p}\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n) = O(1)$. Use (3.8) to get

$$\begin{aligned}
&\mathrm{Var}_\Sigma\Big(\frac{2}{\sqrt{p}} \sum_{l=1}^n \sum_{1\le i<j\le p} w^*_{ij}(X_{l,i}X_{l,j} - \sigma_{ij})\sigma_{ij}\Big) \\
\le\ & \frac{n}{p}\Big(p^{3/2}\Big(o(1)\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n) + O(T^{3/4})\mathbb{E}^{3/2}_\Sigma(\widehat{\mathcal{D}}_n)\Big) + p\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n)O(\sqrt{T})\Big) \\
=\ & o(1)n\sqrt{p}\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n) + (n\sqrt{p}\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n))^{3/2}\cdot\frac{O(T^{3/4})}{n^{1/2}p^{1/4}} + n\sqrt{p}\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n)\cdot o(1)
\end{aligned} \tag{3.52}$$

This tends to 0, since $T^3/n^2p = (n^2pb^2(\varphi))^{-1}\cdot\varphi^{4-2/\alpha} = o(1)$, which is true for all $\alpha > 1/2$.

It follows that, for proving the asymptotic normality, it is sufficient to prove the asymptotic normality of

$$n\sqrt{p}\cdot\frac{1}{n(n-1)p}\sum_{1\le k\ne l\le n}\sum_{1\le i<j\le p} w^*_{ij}(X_{k,i}X_{k,j}-\sigma_{ij})(X_{l,i}X_{l,j}-\sigma_{ij}).$$

We study V_n centered, 1-degenerate U-statistic, with symmetric kernel $H_n(X_1, X_2)$ defined as follows

$$\begin{aligned}
V_n &= \sum_{1\le k\ne l\le n} H_n(X_k, X_l), \\
H_n(X_1,X_2) &= \frac{1}{n\sqrt{p}}\sum_{1\le i<j\le p} w^*_{ij}(X_{k,i}X_{k,j}-\sigma_{ij})(X_{l,i}X_{l,j}-\sigma_{ij}).
\end{aligned}$$

We apply Theorem 1 of [50]. Therefore we check that $\mathbb{E}_\Sigma(H_n^2(X_1,X_2)) < +\infty$ and that

$$\frac{\mathbb{E}_\Sigma(G_n^2(X_1,X_2)) + n^{-1}\mathbb{E}_\Sigma(H_n^4(X_1,X_2))}{\mathbb{E}_\Sigma^2(H_n^2(X_1,X_2))} \longrightarrow 0,$$

where $G_n(x,y) := \mathbb{E}_\Sigma(H_n(X_1,x)H_n(X_1,y))$, for $x,y \in \mathbb{R}^p$. We compute

$$G_n(x,y) \;=\; \frac{1}{n^2 p}\sum_{1\le i<j\le p}\;\sum_{1\le i'<j'\le p} w_{ij}^* w_{i'j'}^*(x_ix_j - \sigma_{ij})(y_iy_j - \sigma_{ij})(\sigma_{ii'}\sigma_{jj'} + \sigma_{i'j}\sigma_{ij'}).$$

Since $n\sqrt{p}\cdot\mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n) = O(1)$, and from the inequality (3.7), we have

$$\mathbb{E}_\Sigma(H_n^2(X_1,X_2)) = \frac{1}{2n^2}(1+o(1))\ .$$

In order to prove that $\mathbb{E}_\Sigma(G_n^2(X_1,X_2))/\mathbb{E}_\Sigma^2(H_n^2(X_1,X_2)) = o(1)$, it is sufficient to show that $\mathbb{E}_\Sigma\Big(\sum_{1\le i<j\le p}\sum_{1\le i'<j'\le p} w_{ij}^* w_{i'j'}^*(X_{1,i}X_{1,j}-\sigma_{ij})(X_{2,i'}X_{2,j'}-\sigma_{i'j'})(\sigma_{ii'}\sigma_{jj'}+\sigma_{i'j}\sigma_{ij'})\Big)^2 = o(p^2)$. In fact,

$$\begin{aligned}
&\mathbb{E}_\Sigma\Big(\sum_{1\le i<j\le p}\;\sum_{1\le i'<j'\le p} w_{ij}^* w_{i'j'}^*(X_{1,i}X_{1,j} - \sigma_{ij})(X_{2,i'}X_{2,j'} - \sigma_{i'j'})(\sigma_{ii'}\sigma_{jj'} + \sigma_{i'j}\sigma_{ij'})\Big)^2 \\
&= \sum_{1\le i_1<j_1\le p}\;\sum_{1\le i_1'<j_1'\le p}\;\sum_{1\le i_2<j_2\le p}\;\sum_{1\le i_2'<j_2'\le p} w_{i_1j_1}^* w_{i_1'j_1'}^* w_{i_2j_2}^* w_{i_2'j_2'}^*(\sigma_{i_1i_1'}\sigma_{j_1j_1'} + \sigma_{i_1'j_1}\sigma_{i_1j_1'}) \\
&\quad\cdot(\sigma_{i_2i_2'}\sigma_{j_2j_2'} + \sigma_{i_2'j_2}\sigma_{i_2j_2'})\cdot\mathbb{E}[(X_{1,i_1}X_{1,j_1} - \sigma_{i_1j_1})(X_{1,i_2}X_{1,j_2} - \sigma_{i_2j_2})] \\
&\quad\cdot\mathbb{E}[(X_{2,i_1'}X_{2,j_1'} - \sigma_{i_1'j_1'})(X_{2,i_2'}X_{2,j_2'} - \sigma_{i_2'j_2'})] \\
&= \sum_{1\le i_1<j_1\le p}\;\sum_{1\le i_1'<j_1'\le p}\;\sum_{1\le i_2<j_2\le p}\;\sum_{1\le i_2'<j_2'\le p} w_{i_1j_1}^* w_{i_1'j_1'}^* w_{i_2j_2}^* w_{i_2'j_2'}^*(\sigma_{i_1i_1'}\sigma_{j_1j_1'} + \sigma_{i_1'j_1}\sigma_{i_1j_1'}) \\
&\quad\cdot(\sigma_{i_2i_2'}\sigma_{j_2j_2'} + \sigma_{i_2'j_2}\sigma_{i_2j_2'})(\sigma_{i_1i_2}\sigma_{j_2j_1} + \sigma_{i_1j_2}\sigma_{i_2j_1})(\sigma_{i_1'i_2'}\sigma_{j_2'j_1'} + \sigma_{i_1'j_2'}\sigma_{i_2'j_1'})
\end{aligned} \tag{3.53}$$

To bound from above (3.53), we shall distinguish four cases. The first one is when all couples of indices are equal,

$$\begin{aligned}
\mathcal{G}_1 &:= \sum_{1\le i_1<j_1\le p} w_{i_1j_1}^{*4}(1+\sigma_{i_1j_1}^2)^4 \le (\sup_{i_1,j_1} w_{i_1j_1}^{*2})\cdot(\sup_{i_1,j_1}(1+\sigma_{i_1j_1}^2)^4)\cdot\sum_{1\le i_1<j_1\le p} w_{i_1j_1}^{*2} \\
&\le 8\cdot(\sup_{i_1,j_1} w_{i_1j_1}^{*2})\cdot p = o(p) = o(p^2).
\end{aligned}$$

The second one is when we have two different pairs of couples of indices, which can be obtained by two different combinations of the couples of indices. When we have equal pairs of couples of indices, as for example $(i_1,j_1) = (i_2,j_2)$, $(i_1',j_1') = (i_2',j_2')$ and $(i_1,j_1) \ne (i_1',j_1')$, we get

$$\begin{aligned}
\mathcal{G}_{2,1} &:= \sum_{1\le i_1<j_1\le p}\;\sum_{1\le i_1'<j_1'\le p} w_{i_1j_1}^{*2} w_{i_1'j_1'}^{*2}(\sigma_{i_1i_1'}\sigma_{j_1j_1'} + \sigma_{i_1'j_1}\sigma_{i_1j_1'})^2(1+\sigma_{i_1j_1}^2)(1+\sigma_{i_1'j_1'}^2) \\
&\le (\sup_{i_1,j_1} w_{i_1j_1}^{*2})\cdot(\sup_{i_1,j_1}(1+\sigma_{i_1j_1}^2)^2)\cdot\sum_{1\le i_1<j_1\le p}\;\sum_{1\le i_1'<j_1'\le p} w_{i_1j_1}^* w_{i_1'j_1'}^*(\sigma_{i_1i_1'}\sigma_{j_1j_1'} + \sigma_{i_1'j_1}\sigma_{i_1j_1'})^2 \\
&\le 4\cdot(\sup_{i_1,j_1} w_{i_1j_1}^{*2})\cdot n^2p\cdot\mathbb{E}_\Sigma(H_n^2(X_1,X_2)) = 4\cdot(\sup_{i_1,j_1} w_{i_1j_1}^{*2})\cdot p = o(p^2).
\end{aligned}$$

When we have three couples of indices equal, for example $(i_1,j_1)=(i_2,j_2)=(i_2',j_2')$ and $(i_1,j_1)\neq(i_1',j_1')$, we get

$$\begin{aligned}\mathcal{G}_{2,2} &:= \sum_{1\le i_1<j_1\le p}\sum_{1\le i_1'<j_1'\le p} w^{*3}_{i_1j_1} w^{*}_{i_1'j_1'}(\sigma_{i_1i_1'}\sigma_{j_1j_1'}+\sigma_{i_1'j_1}\sigma_{i_1j_1'})^2(1+\sigma^2_{i_1j_1})(1+\sigma^2_{i_1'j_1'})\\ &\le 4\cdot(\sup_{i_1,j_1} w^{*^2}_{i_1j_1})\cdot n^2p\cdot\mathbb{E}_\Sigma(H_n^2(X_1,X_2))=o(p^2).\end{aligned}$$

For the third case, there are three different couples of pairs of indices, for example, $(i_1,j_1)=(i_2',j_2')$ and $(i_1,j_1)\neq(i_1',j_1')\neq(i_2,j_2)$. Using Cauchy-Schwarz inequality several times we obtain,

$$\begin{aligned}\mathcal{G}_3 &:= \sum_{1\le i_1<j_1\le p}\sum_{1\le i_1'<j_1'\le p}\sum_{1\le i_2<j_2\le p} w^{*}_{i_1j_1}w^{*}_{i_1'j_1'}w^{*2}_{i_2j_2}(\sigma_{i_1i_1'}\sigma_{j_1j_1'}+\sigma_{i_1'j_1}\sigma_{i_1j_1'})\\ &\qquad\cdot(\sigma_{i_1i_2}\sigma_{j_2j_1}+\sigma_{i_1j_2}\sigma_{i_2j_1})(\sigma_{i_1'i_2}\sigma_{j_2j_1'}+\sigma_{i_1'j_2}\sigma_{i_2j_1'})(1+\sigma^2_{i_2,j_2})\\ &\le \sum_{1\le i_1'<j_1'\le p}\sum_{1\le i_2<j_2\le p} w^{*}_{i_1'j_1'}w^{*2}_{i_2j_2}(\sigma_{i_1'i_2}\sigma_{j_2j_1'}+\sigma_{i_1'j_2}\sigma_{i_2j_1'})(1+\sigma^2_{i_2,j_2})\\ &\quad\cdot\Big(\sum_{1\le i_1<j_1\le p} w^{*}_{i_1j_1}(\sigma_{i_1i_1'}\sigma_{j_1j_1'}+\sigma_{i_1'j_1}\sigma_{i_1j_1'})^2\Big)^{1/2}\Big(\sum_{1\le i_1<j_1\le p} w^{*}_{i_1j_1}(\sigma_{i_1i_2}\sigma_{j_2j_1}+\sigma_{i_1j_2}\sigma_{i_2j_1})^2\Big)^{1/2}\\ &\le \sum_{1\le i_2<j_2\le p} w^{*2}_{i_2j_2}(1+\sigma_{i_2,j_2})^2\Big(\sum_{1\le i_1'<j_1'\le p} w^{*}_{i_1'j_1'}(\sigma_{i_1'i_2}\sigma_{j_2j_1'}+\sigma_{i_1'j_2}\sigma_{i_2j_1'})^2\Big)^{1/2}\\ &\quad\cdot\Big(\sum_{1\le i_1'<j_1'\le p}\sum_{1\le i_1<j_1\le p} w^{*}_{i_1'j_1'}w^{*}_{i_1j_1}(\sigma_{i_1i_1'}\sigma_{j_1'j_1}+\sigma_{i_1j_1'}\sigma_{i_1'j_1})^2\Big)^{1/2}\\ &\quad\cdot\Big(\sum_{1\le i_1<j_1\le p} w^{*}_{i_1j_1}(\sigma_{i_1i_2}\sigma_{j_2j_1}+\sigma_{i_1j_2}\sigma_{i_2j_1})^2\Big)^{1/2}.\end{aligned}$$

Moreover, we recognize in these bounds

$$\sum_{i_1'<j_1'}\sum_{i_1<j_1} w^{*}_{i_1'j_1'}w^{*}_{i_1j_1}(\sigma_{i_1i_1'}\sigma_{j_1'j_1}+\sigma_{i_1j_1'}\sigma_{i_1'j_1})^2=n^2p\cdot\mathbb{E}_\Sigma(H_n^2(X_1,X_2))$$

which is $O(p)$. Thus,

$$\begin{aligned}\mathcal{G}_3 &\le \sup_{i_2,j_2}(1+\sigma_{i_2j_2})^2\cdot\Big(\sum_{1\le i_2<j_2\le p}\sum_{1\le i_1'<j_1'\le p} w^{*2}_{i_2j_2}w^{*}_{i_1'j_1'}(\sigma_{i_1'i_2}\sigma_{j_2j_1'}+\sigma_{i_1'j_2}\sigma_{i_2j_1'})^2\Big)^{1/2}\\ &\quad\cdot\Big(n^2p\cdot\mathbb{E}_\Sigma(H_n^2(X_1,X_2))\cdot\sum_{1\le i_2<j_2\le p}\sum_{1\le i_1<j_1\le p} w^{*2}_{i_2j_2}w^{*}_{i_1j_1}(\sigma_{i_1i_2}\sigma_{j_2j_1}+\sigma_{i_1j_2}\sigma_{i_2j_1})^2\Big)^{1/2}\\ &\le 2(\sup_{i_1,j_1} w^{*}_{i_1j_1})\cdot n^3p^{3/2}\cdot\mathbb{E}_\Sigma^{3/2}(H_n^2(X_1,X_2))\le(\sup_{i_1,j_1} w^{*}_{i_1j_1})\cdot p^{3/2}=o(p^{3/2})=o(p^2).\end{aligned}$$

Now we will treat the last case, when the pairs of indices are pairwise distinct, in this case, we have 16 terms to handle. As all terms are treated the same way, let us deal with:

$$\begin{aligned}\mathcal{G}_4 &:= \sum_{1\le i_1<j_1\le p}\sum_{1\le i_1'<j_1'\le p}\sum_{1\le i_2<j_2\le p}\sum_{1\le i_2'<j_2'\le p} w^{*}_{i_1j_1}w^{*}_{i_1'j_1'}w^{*}_{i_2j_2}w^{*}_{i_2'j_2'}\\ &\quad\cdot\sigma_{i_1i_1'}\sigma_{i_2i_2'}\sigma_{j_2j_2'}\sigma_{j_1j_1'}\sigma_{i_1i_2}\sigma_{j_2j_1}\sigma_{i_1'i_2'}\sigma_{j_2'j_1'}\end{aligned}$$

In order to find an upper bound for $\mathcal{G}_4$, we decompose the previous sums, into several sums, similarly to the upper bound of (3.48). That is $(i_1, j_1, i_1', j_1', i_2, j_2, i_2', j_2') \in J_1 \cup J_2 \cup \cdots \cup J_{16}$, where $J_1, \ldots, J_{16}$, form a partition of the set $\{(i_1, j_1, i_1', j_1', i_2, j_2, i_2', j_2') \in \{1, \ldots, p\}^8\}$. Let us define,

$$J_1 := \{(i_1, j_1, i_1', j_1', i_2, j_2, i_2', j_2') \in \{1, \ldots, p\}^8; 1 < |i_1 - i_1'|, |i_1 - i_2|, |i_2 - i_2'|, |i_1' - i_2'| < T)\},$$

$$J_2 := \{(i_1, j_1, i_1', j_1', i_2, j_2, i_2', j_2') \in \{1, \ldots, p\}^8; 1 < |i_1 - i_1'|, |i_1 - i_2|, |i_2 - i_2'| < T, \text{ and } |i_1' - i_2'| > T)\},$$

and so on, for all $J_r, r = 3, \ldots, 16$. To bound from above the sum over J_1, we partition again J_1, $J_1 = J_{1,1} \cup \cdots \cup J_{1,16}$ such that,

$$\begin{aligned} J_{1,1} \quad := \quad & \{(i_1, j_1, i_1', j_1', i_2, j_2, i_2', j_2') \in \{1, \ldots, p\}^8; |i_1 - i_1'| \le |i_1 - j_1|, |i_1' - i_2'| \le |i_1' - j_1'|, \\ & |i_1 - i_2| \le |i_2 - j_2| \text{ and } |i_2 - i_2'| \le |i_2' - j_2'|\}, \end{aligned}$$

and so on, until we get the partition of J_1.

$$\begin{aligned} \mathcal{G}_{4,1} \quad := \quad & \sum_{\substack{1 \le i_1 < j_1 \le p \quad 1 \le i_1' < j_1' \le p \quad 1 \le i_2 < j_2 \le p \quad 1 \le i_2' < j_2' \le p \\ (i_1, j_1, i_1', j_1', i_2, j_2, i_2', j_2') \in J_{1,1}}} w^*_{i_1 j_1} w^*_{i_1' j_1'} w^*_{i_2 j_2} w^*_{i_2' j_2'} \\ & \cdot \sigma_{i_1 i_1'} \sigma_{j_1 j_1'} \sigma_{i_2 i_2'} \sigma_{j_2 j_2'} \sigma_{i_1 i_2} \sigma_{j_2 j_1} \sigma_{i_1' i_2'} \sigma_{j_2' j_1'} \\ \le \quad & \sum_{1 \le i_1, i_1' \le p} \sum_{1 \le i_2, i_2' \le p} w^*_{i_1 i_1'} w^*_{i_1 i_2} w^*_{i_1' i_2'} w^*_{i_2 i_2'} \sigma_{i_1 i_1'} \sigma_{i_2 i_2'} \sigma_{i_1 i_2} \sigma_{i_1' i_2'} \\ & \cdot \sum_{\substack{1 \le j_1, j_1' \le p \quad 1 \le j_2, j_2' \le p \\ 1 < |i_1 - j_1|, |i_1' - j_1'|, |i_2 - j_2|, |i_2' - j_2'| < T}} \sigma_{j_2 j_2'} \sigma_{j_1 j_1'} \sigma_{j_2 j_1} \sigma_{j_2' j_1'} \\ \le \quad & T^4 \cdot (\sup_{i_1, j_1} w^*_{i_1 j_1})^2 \cdot \sum_{1 \le i_1, i_1' \le p} \sum_{1 \le i_2, i_2' \le p} \sqrt{w^*_{i_1 i_1'} w^*_{i_1 i_2} w^*_{i_1' i_2'} w^*_{i_2 i_2'}} \sigma_{i_1 i_1'} \sigma_{i_2 i_2'} \sigma_{i_1 i_2} \sigma_{i_1' i_2'} \\ \le \quad & T^4 \cdot (\sup_{i_1, j_1} w^*_{i_1 j_1})^2 \Big(\sum_{1 \le i_1, i_1' \le p} \sum_{1 \le i_2, i_2' \le p} w^*_{i_1 i_1'} w^*_{i_2 i_2'} \sigma^2_{i_1 i_1'} \sigma^2_{i_2 i_2'} \Big)^{1/2} \\ & \cdot \Big(\sum_{1 \le i_1, i_1' \le p} \sum_{1 \le i_2, i_2' \le p} w^*_{i_1 i_2} w^*_{i_1' i_2'} \sigma^2_{i_1 i_2} \sigma^2_{i_1' i_2'} \Big)^{1/2} \\ \le \quad & T^4 \cdot (\sup_{i_1, j_1} w^*_{i_1 j_1})^2 \cdot p^2 \cdot \mathbb{E}^2_{\Sigma}(\widehat{D}_n) \end{aligned}$$

Again, by our assumption that $n^2 p \cdot \mathbb{E}^2_{\Sigma}(\widehat{D}_n) = O(1)$, we can see that :

$$\mathcal{G}_{4,1} \le \kappa_0(\alpha, L) \cdot T^3 \cdot p^2 \cdot \mathbb{E}^2_{\Sigma}(\widehat{D}_n) = p^2 \cdot O(\frac{T^3}{n^2 p}) = p^2 \cdot o(1)$$

where, from now on, $\kappa_0(\alpha, L), \kappa_1(\alpha, L), \ldots,$ denote constants that depend on α and L. Now, we define $J_{1,2} := \{(i_1, j_1, i_1', j_1', i_2, j_2, i_2', j_2') \in \{1, \ldots, p\}^8$, such that $|i - i'| \le |i - j|$,

$|i'-i'_1| \le |i'-j'|$, $|i-i_1| \le |i_1-j_1|$ and $|i_1-i'_1| > |i'_1-j'_1|\}$, thus we have,

$$\begin{aligned}
\mathcal{G}_{4,2} &:= \sum_{(i_1,j_1,i'_1,j'_1,i_2,j_2,i'_2,j'_2)\in J_{1,2}} \sum\sum\sum w^*_{i_1j_1} w^*_{i'_1j'_1} w^*_{i_2j_2} w^*_{i'_2j'_2} \sigma_{i_1i'_1}\sigma_{i_2i'_2}\sigma_{j_1j'_1}\sigma_{j_2j'_2}\sigma_{i_1i_2}\sigma_{j_2j_1}\sigma_{i'_1i'_2}\sigma_{j'_2j'_1} \\
&\le (\sup_{i_1,j_1} w^*_{i_1j_1})^{5/2} \cdot \sum_{1\le i_1,i'_1\le p}\ \sum_{1\le i_2,i'_2\le p} \sqrt{w^*_{i_1i'_1} w^*_{i_1i_2} w^*_{i'_1i'_2}} \cdot |i_2-i'_2|^{\alpha} \cdot \sigma_{i_1i'_1}\sigma_{i_2i'_2}\sigma_{i_1i_2}\sigma_{i'_1i'_2} \\
&\quad \cdot \sum_{\substack{1\le j_1,j'_1\le p \\ 1<|i_1-j_1|,|i'_1-j'_1|,|i_2-j_2|,|i'_2-j'_2|<T}}\ \sum_{1\le j_2,j'_2\le p} \frac{1}{|i'_2-j'_2|^{\alpha}} \cdot \sigma_{j_2j'_2}\sigma_{j_1j'_1}\sigma_{j_2j_1}\sigma_{j'_2j'_1} \\
&\le (\sup_{i_1,j_1} w^*_{i_1j_1})^{5/2} \cdot \Big(\sum_{1\le i_1,i'_1\le p}\ \sum_{1\le i_2,i'_2\le p} w^*_{i_1i'_1} |i_2-i'_2|^{2\alpha} \sigma^2_{i_1i'_1}\sigma^2_{i_2i'_2} \Big)^{1/2} \\
&\quad \cdot \Big(\sum_{1\le i_1,i'_1\le p}\ \sum_{1\le i_2,i'_2\le p} w^*_{i_1i_2} w^*_{i'_1i'_2} \sigma^2_{i_1i_2}\sigma^2_{i'_1i'_2} \Big)^{1/2} \cdot T^3 \cdot \max\{1, T^{-\alpha+1}\} \\
&\le \sqrt{2L} \cdot (\sup_{i_1,j_1} w^*_{i_1j_1})^{5/2} \cdot T^3 \cdot \max\{1, T^{-\alpha+1}\} \cdot p^2 \cdot \mathbb{E}^{3/2}_{\Sigma}(\widehat{D}_n)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathcal{G}_{4,2} &\le \kappa_1(\alpha, L) \cdot \max\{T^{7/4}, T^{11/4-\alpha}\} \cdot \mathbb{E}^{3/2}_{\Sigma}(\widehat{D}_n) \\
&\le \kappa_1(\alpha, L) \cdot \max\{T^{7/4}, T^{11/4-\alpha}\} \cdot O(\frac{1}{n^{3/2}p^{3/4}}) \\
&= o(1) \quad \text{since } T^3/n^2p \longrightarrow 0 \qquad (3.54)
\end{aligned}$$

Using similar arguments, we can prove that all remaining terms tend to zero. In consequence,

$$\frac{\mathbb{E}_{\Sigma}(G^2_n(X_1,X_2))}{\mathbb{E}^2_{\Sigma}(H^2_n(X_1,X_2))} \longrightarrow 0.$$

Now let us prove that, $\mathbb{E}_{\Sigma}(H^4_n(X_1,X_2))/\mathbb{E}^2_{\Sigma}(H^2_n(X_1,X_2)) = o(n)$,

$$\begin{aligned}
\mathbb{E}_{\Sigma}(H^4_n(X_1,X_2)) &= \frac{1}{n^4p^2} \sum_{i_1<j_1}\sum_{i_2<j_2}\sum_{i_3<j_3}\sum_{i_4<j_4} w^*_{i_1j_1} w^*_{i_2j_2} w^*_{i_3j_3} w^*_{i_4j_4} \\
&\cdot \mathbb{E}^2_{\Sigma}[(X_{1,i_1}X_{1,j_1}-\sigma_{i_1j_1})(X_{1,i_2}X_{1,j_2}-\sigma_{i_2j_2})(X_{1,i_3}X_{1,j_3}-\sigma_{i_3j_3})(X_{1,i_4}X_{1,j_4}-\sigma_{i_4j_4})]
\end{aligned}$$

The above squared expected value is a sum of a large number of terms that are all treated similarly. Let us consider examples of terms containing squared terms and products of terms, respectively. For $\alpha > 1/2$,

$$\begin{aligned}
\mathcal{H}_1 &:= \sum_{i_1<j_1}\sum_{i_2<j_2}\sum_{i_3<j_3}\sum_{i_4<j_4} w^*_{i_1j_1} w^*_{i_2j_2} w^*_{i_3j_3} w^*_{i_4j_4} \sigma^2_{i_1i_2}\sigma^2_{j_1j_2}\sigma^2_{i_3i_4}\sigma^2_{j_3j_4} \\
&\le 4(\sup_{i,j} w^*_{ij})^4 \sum_{i_1=1}^{p}\sum_{i_2=1}^{p} |i_1-i_2|^{2\alpha}\sigma^2_{i_1i_2} \sum_{\substack{j_1=1\\|i_1-j_1|<T}}^{p} \sup_{j_2}\sigma^2_{j_1j_2} \sum_{\substack{j_2=1\\|i_2-j_2|<T}}^{p} \frac{1}{|j_1-j_2|^{2\alpha}} \\
&\quad \cdot \sum_{i_3=1}^{p}\sum_{i_4=1}^{p} |i_3-i_4|^{2\alpha}\sigma^2_{i_3i_4} \sum_{\substack{j_3=1\\|i_3-j_3|<T}}^{p} \sup_{j_4}\sigma^2_{j_3j_4} \sum_{\substack{j_4=1\\|i_4-j_4|<T}}^{p} \frac{1}{|j_3-j_4|^{2\alpha}} \\
&\le 16L^2 \cdot (2\alpha-1)^{-2} \cdot (\sup_{i,j} w^*_{ij})^4 \cdot p^2T^2 \le \kappa_2(\alpha, L) \cdot p^2
\end{aligned}$$

The terms containing no squared values are treated as, e.g.,

$$\mathcal{H}_2 \quad := \quad \sum_{i_1<j_1}\sum_{i_2<j_2}\sum_{i_3<j_3}\sum_{i_4<j_4} w^*_{i_1j_1}w^*_{i_2j_2}w^*_{i_3j_3}w^*_{i_4j_4}\sigma_{i_1i_2}\sigma_{j_1j_2}\sigma_{i_3i_4}\sigma_{j_3j_4}\sigma_{i_1i_3}\sigma_{j_1j_3}\sigma_{i_2i_4}\sigma_{j_2j_4}$$

We can see that $\mathcal{H}_2$ coincides with $\mathcal{G}_{4,2}$. Then we can deduce that ,

$$\frac{\mathbb{E}_\Sigma(H_n^4(X_1,X_2))}{\mathbb{E}^2_\Sigma(H_n^2(X_1,X_2))} = O(1) = o(n).$$

Finally we can apply [50], and we obtain:

$$V_n = \frac{1}{n\sqrt{p}}\sum_{1\le k\ne l\le n}\sum_{1\le i<j\le p} w^*_{ij}(X_{k,i}X_{k,j}-\sigma_{ij})(X_{l,i}X_{l,j}-\sigma_{ij}) \xrightarrow{\mathcal{L}} N(0,1). \tag{3.55}$$

Combining (3.52) and (3.55), we have by Slutsky theorem that:

$$n\sqrt{p}\cdot(\widehat{\mathcal{D}}_n - \mathbb{E}_\Sigma(\widehat{\mathcal{D}}_n)) \xrightarrow{\mathcal{L}} N(0,1).$$

□

Proof of Proposition 3.9 . Let us check the case where $u_{ij}=1$ for all i, j such that $|i-j|\le T$ and the generalization to all U in $\mathcal{U}$ will be obvious. Using Gershgorin's Theorem we get that each eigenvalue of $\Sigma_U^* = [u_{ij}\sigma^*_{ij}]_{1\le i,j\le p}$ lies in one of the disks centered in $\sigma_{ii}=1$ and radius $R_i = \sum_{\substack{j=1\\ j\ne i}}^p |u_{ij}\sigma^*_{ij}| = \sum_{\substack{j=1\\ j\ne i}}^p \sigma^*_{ij}$. We have,

$$\begin{aligned}
\sum_{\substack{j=1\\ j\ne i}}^p \sigma^*_{ij} &= \sqrt{\lambda}\sum_{\substack{j=1\\ j\ne i}}^p \left(1-(\frac{|i-j|}{T})^{2\alpha}\right)_+^{\frac12} \le 2\sqrt{\lambda}\sum_{k=1}^T\left(1-(\frac{k}{T})^{2\alpha}\right)^{\frac12}\\
&\le 2\sqrt{\lambda}\Big(\sum_{k=1}^T(1-(\frac{k}{T})^{2\alpha})\Big)^{\frac12}T^{\frac12} = O(1)T\sqrt{\lambda}\\
&\le O(1)\varphi^{1-\frac{1}{2\alpha}} \to 0 \text{ provided that } \alpha>1/2.
\end{aligned}$$

We deduce that the smallest eigenvalue is bounded from below by

$$\min_{i=1,\dots,p}\lambda_{i,U} \ge \min_i\{\sigma^*_{ii} - \sum_{\substack{j=1\\ j\ne i}}^p\sigma^*_{ij}\} = 1-\max_i\sum_{\substack{j=1\\ j\ne i}}^p\sigma^*_{ij} \ge 1-O(1)\varphi^{1-\frac{1}{2\alpha}}$$

which is strictly positive for $\varphi>0$ small enough. □

Proposition 3.10. *For all $1\le i<j\le p$, W_{ij} is a centered random variable with variance, $Var_I(W_{ij})=n$. Moreover, for $1\le i<j\ne j'\le p$, we have*

$$\mathbb{E}_I(W_{ij}^4) = 3n^2+6n\,,\quad \mathbb{E}_I(W_{ij}^2W_{ij'}^2) = n^2+2n\,,\quad \mathbb{E}_I(W_{ij}^4W_{ij'}^4) = 9(n^4+12n^3+44n^2+48n).$$

Also we have that $\mathbb{E}_I(W_{ij}^8) = 105n^4(1+o(1))$. Note that if we have $i\ne i'$ and $j\ne j'$, then W_{ij}^d and $W_{i'j'}^d$ are not correlated for d finite integer. Moreover, for all $1\le i\le p$, the random variables W_{ii} are such that,

$$\mathbb{E}(W_{ii}) = n\,,\quad \mathbb{E}_I(W_{ii}^2) = n^2+2n\,,\quad \mathbb{E}_I(W_{ii}^4) = n^4+12n^3+44n^2+48n.$$

Proof of Proposition 3.10. To show the results we use lemma 3 and some technical computation of [23].

$$\begin{aligned}
\mathrm{Var}_I(W_{ij}) &= \mathbb{E}_I(W_{ij}^2) = \mathbb{E}_I(X_{\cdot i}^\top X_{\cdot j})^2 = \mathbb{E}_I(tr(X_{\cdot i}X_{\cdot i}^\top X_{\cdot j}X_{\cdot j}^\top)) = tr(I_n^2) = n.\\
\mathbb{E}_I(W_{ij}^4) &= \mathbb{E}_I(X_{\cdot i}^\top X_{\cdot j})^4 = 3tr^2(I_n^2) + 6tr(I_n^4) = 3n^2 + 6n\\
\mathbb{E}_I(W_{ij}^2 W_{ij'}^2) &= \mathbb{E}_I((X_{\cdot i}^\top X_{\cdot j})^2(X_{\cdot i}^\top X_{\cdot j'})^2) = tr^2(I_n^2) + 2tr(I_n^4) = n^2 + 2n\\
\mathbb{E}_I(W_{ij}^4 W_{ij'}^4) &= \mathbb{E}_I((X_{\cdot i}^\top X_{\cdot j})^4(X_{\cdot i}^\top X_{\cdot j'})^4) = \mathbb{E}_I\Big(\mathbb{E}_I\Big((X_{\cdot i}^\top X_{\cdot j})^4(X_{\cdot i}^\top X_{\cdot j'})^4|X_{\cdot i}\Big)\Big).
\end{aligned}$$

Or $\mathbb{E}_I\Big((X_{\cdot i}^\top X_{\cdot j})^4(X_{\cdot i}^\top X_{\cdot j'})^4|X_{\cdot i}\Big) = g(X_{\cdot i})$, where

$$g(x_{\cdot i}) = \mathbb{E}_I\Big((x_{\cdot i}^\top X_{\cdot j})^4(x_{\cdot i}^\top X_{\cdot j'})^4\Big) = \mathbb{E}_I\Big((x_{\cdot i}^\top X_{\cdot j})^4\Big)\mathbb{E}_I\Big((x_{\cdot i}^\top X_{\cdot j'})^4\Big)$$

$$\begin{aligned}
\mathbb{E}_I\Big((x_{\cdot i}^\top X_{\cdot j})^4\Big) &= \mathbb{E}_I\Big(\sum_{k=1}^p x_{k,i}X_{k,j}\Big)^4 = \sum_{k=1}^p x_{k,i}^4\mathbb{E}_I(X_{k,j}^4) + 3\sum\sum_{k_1\neq k_2} x_{k_1,i}^2 x_{k_2,i}^2\mathbb{E}_I(X_{k_1,j}^2)\mathbb{E}_I(X_{k_2,j}^2)\\
&= 3\Big(\sum_{k=1}^p x_{k,i}^2\Big)^2 = 3(x_{\cdot i}^\top x_{\cdot i})^2
\end{aligned}$$

then we obtain that

$$\mathbb{E}_I(W_{ij}^4 W_{ij'}^4) = 9\mathbb{E}_I(X_{\cdot i}^\top X_{\cdot i})^4 = 9(n^4 + 12n^3 + 44n^2 + 48n).$$

Also we have that

$$\begin{aligned}
&\mathbb{E}_I(W_{ij}^8) = \mathbb{E}_I\Big(\sum_{k=1}^n X_{k,i}X_{k,j}\Big)^8 = \sum_{k=1}^n \mathbb{E}_I^2(X_{k,i}^8) + C_8^6\cdot\sum\sum_{k_1\neq k_2}\mathbb{E}_I^2(X_{k_1,i}^6)\cdot\mathbb{E}_I^2(X_{k_2,i}^2)\\
+\ & \frac{C_8^4}{2!}\cdot\sum\sum_{k_1\neq k_2}\mathbb{E}_I^2(X_{k_1,i}^4)\cdot\mathbb{E}_I^2(X_{k_2,i}^4) + \frac{C_8^4\cdot C_4^2}{2!}\cdot\sum\sum\sum_{k_1\neq k_2\neq k_3}\mathbb{E}_I^2(X_{k_1,i}^4)\cdot\mathbb{E}_I^2(X_{k_2,i}^2)\cdot\mathbb{E}_I^2(X_{k_3,i}^2)\\
+\ & \frac{C_8^2\cdot C_6^2\cdot C_4^2}{4!}\sum\sum\sum_{k_1\neq k_2\neq k_3\neq k_4}\mathbb{E}_I^2(X_{k_1,i}^2)\cdot\mathbb{E}_I^2(X_{k_2,i}^2)\cdot\mathbb{E}_I^2(X_{k_3,i}^2)\cdot\mathbb{E}_I^2(X_{k_4,i}^2)\\
=\ & 105^2 n + (28\times 15^2 + 35\times 9)n(n-1) + (210\times 9)n(n-1)(n-2)\\
+\ & 105n(n-1)(n-2)(n-3)\\
=\ & 105n^4 + 1220n^3 + 2100n^2 + 7560n
\end{aligned}$$

We use similar arguments to calculate the moments of W_{ii}. □

Lemme 3.11. *Let $0 < \varepsilon < 1/2$, for any $t > 0$ we have that,*

$$\Big|\mathbb{P}_I(\widehat{\mathcal{D}}_{n,r} \le t) - \Phi(n\sqrt{p}\cdot t)\Big| \le 16\varepsilon^{1/2}\exp(-\frac{n^2pt^2}{4}) + O\Big(\frac{1}{n}\Big) + O\Big(\frac{1}{pT_r}\Big) \quad \textit{for all } 1 \le r \le N.$$

Proof of Lemma 3.11. For each $r \in \{1,\dots,N\}$, $\widehat{\mathcal{D}}_{n,r}$ is a degenerated U-statistic of order 2, and can be written as follows:

$$\widehat{\mathcal{D}}_{n,r} = \sum\sum_{1\le k\neq l\le n} K(X_k, X_l), \quad \text{where } K(X_k,X_l) = \frac{1}{n(n-1)p}\sum\sum_{1\le i<j\le p} w_{ij,r}^* X_{k,i}X_{k,j}X_{l,i}X_{l,j}.$$

Define,

$$Z_k = \frac{1}{\sqrt{\mathrm{Var}_I(\widehat{\mathcal{D}}_{n,r})}} \sum_{l=1}^{k-1} K(X_k, X_l) \quad \text{and } V_n^2 = \sum_{k=2}^{n} \mathbb{E}_I(Z_k^2/\mathcal{F}_{k-1})$$

where $\mathcal{F}_k$ is the σ-field generated by the random variables $\{X_1, \dots, X_k\}$. Moreover, fix $0 < \delta \le 1$, and define

$$J_n = \sum_{k=2}^{n} \mathbb{E}_I(Z_k)^{2+2\delta} + \mathbb{E}_I|V_n^2 - 1|^{1+\delta}.$$

Then by Theorem 3 of [16] we get that, there exists a positive constant k depending only on δ such that for any $0 < \varepsilon < 1/2$ and any real t,

$$\Big|\mathbb{P}_I(\widehat{\mathcal{D}}_{n,r} \le t) - \Phi\Big(\frac{x}{\sqrt{\mathrm{Var}_I(\widehat{\mathcal{D}}_{n,r})}}\Big)\Big| \le 16\varepsilon^{1/2}\exp(-\frac{t^2}{4\mathrm{Var}_I(\widehat{\mathcal{D}}_{n,r})}) + \frac{k}{\varepsilon^{1+\delta}} \cdot J_n.$$

Now, we give upper bounds for $\sum_{k=2}^{n} \mathbb{E}_I(Z_k)^{2+2\delta}$ and $\mathbb{E}_I|V_n^2 - 1|^{1+\delta}$ for $\delta = 1$ and get,

$$\begin{aligned}
\sum_{k=2}^{n} \mathbb{E}_I(Z_k)^4 &= \frac{1}{n^2(n-1)^2p^2} \sum_{k=2}^{n} \mathbb{E}_I\Big(\sum_{l=1}^{k-1} \sum\sum_{1\le i<j\le p} w^*_{ij,r} X_{k,i}X_{k,j}X_{l,i}X_{l,j}\Big)^4 \\
&= \frac{1}{n^2(n-1)^2p^2} \sum_{k=2}^{n} \Big\{(k-1)\Big(3^4 \sum\sum_{1\le i<j\le p} w^{*4}_{ij,r} + 3 \sum\sum_{1\le i<j\le p}\sum\sum_{1\le i'<j'\le p} w^{*2}_{ij,r} w^{*2}_{i'j',r}\Big) \\
&\quad +3(k-1)(k-2)\Big(3^2 \sum\sum_{1\le i<j\le p} w^{*4}_{ij,r} + \sum\sum_{1\le i<j\le p}\sum\sum_{1\le i'<j'\le p} w^{*2}_{ij,r} w^{*2}_{i'j',r}\Big)\Big\} \\
&\le \frac{1}{n^2(n-1)^2p^2}\Big\{\frac{n(n-1)}{2}\Big(\frac{81}{2} \cdot p(\sup_{i,j} w^{*2}_{ij,r}) + \frac{3p^2}{4}\Big) \\
&\qquad + \frac{(n-1)n(2n-1)}{6}\Big(\frac{9}{2} \cdot p(\sup_{i,j} w^{*2}_{ij,r}) + \frac{p^2}{4}\Big)\Big\} \\
&= O\Big(\frac{1}{n}\Big).
\end{aligned}$$

Similarly we can show that $\mathbb{E}_I(V_n^2 - 1)^2 = O\Big(\frac{1}{n}\Big) + O\Big(\frac{1}{pT_r}\Big)$. Thus we obtain the desired result. □

Chapter 4

Adaptive tests for large covariance matrices in presence of missing observations

Abstract.

We observe n independent $p-$dimensional Gaussian vectors with missing coordinates, that is each value (which is assumed standardized) is observed with probability $a > 0$. We investigate the problem of minimax nonparametric testing that the high-dimensional covariance matrix Σ of the underlying Gaussian distribution is the identity matrix, using these partially observed vectors. Here, n and p tend to infinity and $a > 0$ tends to 0, asymptotically.

We assume that Σ belongs to a Sobolev-type ellipsoid with parameter $\alpha > 0$. When α is known, we give asymptotically minimax consistent test procedure and find the minimax separation rates $\tilde{\varphi}_{n,p} = (a^2 n\sqrt{p})^{-\frac{2\alpha}{4\alpha+1}}$, under some additional constraints on n, p and a. We show that, in the particular case of Toeplitz covariance matrices,the minimax separation rates are faster, $\tilde{\phi}_{n,p} = (a^2 np)^{-\frac{2\alpha}{4\alpha+1}}$. We note how the "missingness" parameter a deteriorates the rates with respect to the case of fully observed vectors ($a = 1$).

We also propose adaptive test procedures, that is free of the parameter α in some interval, and show that the loss of rate is $(\ln\ln(a^2 n\sqrt{p}))^{\alpha/(4\alpha+1)}$ and $(\ln\ln(a^2 np))^{\alpha/(4\alpha+1)}$ for Toeplitz covariance matrices, respectively.

4.1 Introduction

Recently, problems related to high-dimensional data became more popular. In particular, in many areas such as genetics, meteorology and others, the generated data sets are high-dimensional and incomplete, in the sense that they contain missing values. In this paper we investigate the problem of testing large covariance matrices from a sample of partially observed vectors.

Let $X_1, \dots, X_n$, be n independent and identically distributed p-vectors following a multivariate normal distribution $\mathcal{N}_p(0, \Sigma)$, where $\Sigma = [\sigma_{ij}]_{1 \le i,j \le p}$ is the normalized covariance matrix, with $\sigma_{ii} = 1$, for all $i = 1$ to p. Let us denote $X_k = (X_{k,1}, \dots, X_{k,p})^\top$ for all $k = 1, \dots, n$. Let $\{\varepsilon_{k,j}\}_{1 \le k \le n, 1 \le j \le p}$ be a sequence of i.i.d. Bernoulli random variables with parameter $a \in (0,1)$, $\mathcal{B}(a)$, and independent from $X_1, \dots, X_n$. We observe n i.i.d. random vectors $Y_1, \dots, Y_n$ such that

$$Y_k = (\varepsilon_{k,1} \cdot X_{k,1}, \dots, \varepsilon_{k,p} \cdot X_{k,p})^\top \quad \text{for all } k = 1, \dots, n.$$

Each component of the vector Y_k is observed with probability equal to a and this is the context of missing observations. We denote by $P_{a,\Sigma}$ the probability distribution of the random vector Y_k when $X_k \sim \mathcal{N}_p(0, \Sigma)$ and $\varepsilon_{k,j} \sim \mathcal{B}(a)$. We also denote by $\mathbb{E}_{a,\Sigma}$ and $\mathrm{Var}_{a,\Sigma}$ the expected value and the variance associated to $P_{a,\Sigma}$. Given the partially observed vectors $Y_1, \dots, Y_n$, we want to test the null hypothesis

$$H_0 : \Sigma = I \tag{4.1}$$

against a composite alternative hypothesis

$$H_1 : \Sigma \in Q(\mathcal{G}(\alpha), \varphi) \tag{4.2}$$

where $\alpha > 0$ and $\mathcal{G}(\alpha)$ is either

$$\mathcal{F}(\alpha) = \{\Sigma > 0, \text{ symmetric}; \frac{1}{p} \sum_{1 \le i < j \le p} \sigma_{ij}^2 |i - j|^{2\alpha} \le 1 \text{ for all } p \ge 1 \text{ and } \sigma_{ii} = 1 \text{ for all } 1 \le i \le p\}$$

in the general case or

$$\mathcal{T}(\alpha) = \{\Sigma > 0, \text{ symmetric}, \Sigma \text{ is Toeplitz }; \sum_{j \ge 1} \sigma_j^2 j^{2\alpha} \le 1 \text{ and } \sigma_0 = 1\}$$

for the case of Toeplitz matrices. Thus, we define the following ℓ_2 ellipsoids with ℓ_2 balls removed:

$$Q(\mathcal{F}(\alpha), \varphi) = \left\{\Sigma \in \mathcal{F}(\alpha) \quad \text{such that } \frac{1}{p} \sum_{1 \le i < j \le p} \sigma_{ij}^2 \ge \varphi^2\right\} \tag{4.3}$$

and

$$Q(\mathcal{T}(\alpha), \varphi) = \left\{\Sigma \in \mathcal{T}(\alpha) \quad \text{such that } \sum_{j \ge 1} \sigma_j^2 \ge \varphi^2\right\} \tag{4.4}$$

Typically, the test procedures depend on the parameter α and it is therefore useful to construct a test procedure that is adaptive to α in some interval. Here we propose minimax and adaptive procedures for testing in the context of missing observations.

The problem of estimating a covariance matrix of partially observed vectors was investigated several times in the literature. The simplest method to deal with missing data is to ignore the missing values and restrict the study to a subset of fully observed variables. This method is not always reliable mainly when the number of missing values is relatively high. Hence, in order to treat this problem, methods based on filling in the missing values were developed, in particular the Expectation-Maximization(EM) algorithm see [82]. Recently, [73] proposed an estimating procedure that does not need imputation of the missing values. Instead, the setup with missing values is treated as an inverse problem. We will also follow this approach for the test problem.

The problem of testing large covariance matrices was considered only in the case of complete data. Out of the large amount of results in the literature on this latter problem, we mention only the most related papers where procedures to test the null hypothesis H_0 in (4.1) are derived. We refer to [4] , [62] and [91], where test procedures based on the likelihood ratio are proposed, and to [67], [86], [29] and [23], where test statistics based on the quadratic loss function $tr(\Sigma - I)^2$ are used. Note that in [18] and [19] asymptotically consistent test procedures where given in order to test (4.1) against (4.2), when the covariance matrices belongs to (4.3) and to (4.4), respectively. They describe the minimax and sharp minimax separation rates. Here, we give the minimax separation rates when assuming that we have partially observed vectors. We describe how the "missingness" parameter a deteriorates the minimax rates in this context. Moreover we develop consistent test procedures free of the class parameter α, via an aggregation procedure of tests.

Missing observations appeared recently in random matrix theory, see [65]. They show that the sequence of the spectral measures of sample covariance matrices with missing observations converge weakly to a sequence of non random measures. Also they studied the limits of the extremes eigenvalues in the same context.

In this paper, we describe the minimax separation rate for testing H_0 given in (4.1) against the composite alternative H_1 in (4.2), when the data contains missing values. For a test procedure Δ we define the type I error probability by $\eta(\Delta) = P_I(\Delta = 1)$, the maximal type II error probability by $\beta(\Delta, Q(\mathcal{G}(\alpha), \varphi)) = \sup_{\Sigma \in Q(\mathcal{G}(\alpha),\varphi)} P_\Sigma(\Delta = 0)$ and the total error probability by

$$\gamma(\Delta, Q(\mathcal{G}(\alpha), \varphi)) = \eta(\Delta) + \beta(\Delta, Q(\mathcal{G}(\alpha), \varphi)).$$

Moreover, we define the minimax total error probability over the class $Q(\mathcal{G}(\alpha), \varphi)$ by

$$\gamma := \inf_{\Delta} \gamma(\Delta, Q(\mathcal{G}(\alpha), \varphi))$$

where the infimum is taken over all possible test procedures. We define the minimax separation rate $\widetilde{\varphi}_\alpha$. On the one hand, we construct a test procedure Λ and derive the conditions on φ for which $\gamma(\Lambda, Q(\alpha,\varphi)) \to 0$. The test Λ will be called asymptotically minimax consistent. On the other hand we give the conditions on φ for which $\gamma \to 1$. The previous conditions together allow us to determinate the minimax separations rate $\widetilde{\varphi}_\alpha$, such that there exists the test Λ with

$$\gamma(\Lambda, Q(\mathcal{G}(\alpha),\varphi)) \to 0 \quad \text{if } \frac{\varphi}{\widetilde{\varphi}_\alpha} \to +\infty,$$

and

$$\gamma = \inf_{\Delta} \gamma(\Delta, Q(\mathcal{G}(\alpha),\varphi)) \to 1 \quad \text{if } \frac{\varphi}{\widetilde{\varphi}_\alpha} \to 0.$$

In other words, when $\varphi >> \widetilde{\varphi}_\alpha$ there exists an asymptotically minimax consistent test procedure and when $\varphi << \widetilde{\varphi}_\alpha$, there is no asymptotically consistent test procedure which can distinguish between the null and the alternative hypothesis.

We also consider the problem of adaptation with respect to the parameter α. To treat this problem we first assume that $\alpha \in A$, for A an interval, and define a larger class of matrices under the alternative than (4.2). The testing problem we are interested in now, is to test H_0 in (4.1) against

$$H_1 : \Sigma \in \bigcup_{\alpha \in A} Q(\mathcal{F}(\alpha), \mathcal{C}\psi_\alpha),$$

where $\psi_\alpha = \rho_{n,p}/\widetilde{\varphi}_\alpha$, and $\widetilde{\varphi}_\alpha$ is the minimax separation rate of testing H_0 given in (4.1) against H_1 in (4.2) for a known α. Our aim is to construct a test procedure Δ_{ad} and to find the loss $\rho_{n,p}$ such that for a large enough constant $\mathcal{C} > 0$:

$$\gamma(\Delta_{ad}, \bigcup_{\alpha \in A} Q(\mathcal{F}(\alpha), \mathcal{C}\psi_\alpha)) \to 0.$$

In this case we say that Δ_{ad} is an asymptotically adaptive consistent test procedure.

The paper is structured as follows: in section 4.2 we solve the case of general covariance matrices in $\mathcal{F}(\alpha)$ and in section 4.3 the particular case of Toeplitz covariance matrices in $\mathcal{T}(\alpha)$. In section 4.2.1, we study the test problem with alternative hypothesis $Q(\mathcal{F}(\alpha),\varphi)$. We construct an asymptotically minimax consistent test procedure based on the data with missing observations and show that the minimax separation rate is

$$\tilde{\varphi}_\alpha(\mathcal{F}) = (a^2 n\sqrt{p})^{-\frac{2\alpha}{4\alpha+1}}.$$

In section 4.2.2, we propose a test procedure adaptive to the unknown parameter α. In section 4.3, we study the problem with alternative hypothesis $Q(\mathcal{T}(\alpha),\varphi)$ and derive analogous results. The minimax separation rate is

$$\tilde{\phi}_\alpha(\mathcal{T}) = (a^2 np)^{-\frac{2\alpha}{4\alpha+1}}.$$

We can view the vectors X_k in this case as a sample of size p from a stationary Gaussian process. However, due to the missing data, this is not true anymore for vectors Y_k. Minimax and adaptive rates of testing are faster by a factor $\sqrt{p}$ over classes $\mathcal{T}(\alpha)$ than over the classes $\mathcal{F}(\alpha)$. The adaptive procedure attains the rates $(\sqrt{\ln\ln(a^2 n\sqrt{p})}/(a^2 n\sqrt{p}))^{2\alpha/(4\alpha+1)}$ and $(\sqrt{\ln\ln(a^2 np)}/(a^2 np))^{2\alpha/(4\alpha+1)}$, respectively. However, the parameter a describing the probability of a missing coordinate appears similarly in both cases. It actually deteriorates the rates with respect to the case $a = 1$ of fully observed data. Proofs are given in section 4.4.

Note that, for the rest of the paper asymptotics will be taken when $n \to +\infty$, $p \to +\infty$ and a is either fix or tends to 0 under further constraints.

4.2 Test for covariance matrices

We want to test from the data with missing coordinates $Y_1, \dots, Y_n$ the null hypothesis (4.1) against the alternative (4.2) that we recall here:

$$H_1 : \Sigma \in Q(\mathcal{F}(\alpha), \varphi)$$

where $Q(\mathcal{F}(\alpha), \varphi)$ is given in (4.3). This testing problem is treated in [18], for the case of fully observed data, which correspond to $a = 1$ in our case. For the sake of clarity, let us recall that in [18], the following weighted U-statistic was studied

$$\widehat{\mathcal{D}}_{n,p} = \frac{1}{n(n-1)p} \sum_{1 \le k \neq l \le n} \sum_{\substack{1 \le i < j \le p \\ |i-j|<m}} w_{ij} X_{k,i} X_{k,j} X_{l,i} X_{l,j}.$$

The test based on $\widehat{\mathcal{D}}_{n,p}$ was shown to achieve minimax and sharp minimax separation rates, i.e. asymptotic equivalents of the type II error and the total error probabilities are also given when $\varphi \asymp \tilde{\varphi}_\alpha$. The weights $\{w_{ij}\}_{1 \le i < j \le p}$ depend on the parameter α and are chosen as solution of the following optimization problem:

$$\begin{aligned} \sup_{\{w_{ij} \ge 0; \sum_{i<j} w_{ij}^2 = \frac{1}{2}\}} \inf_{\Sigma \in Q(\mathcal{F}(\alpha),\varphi)} \mathbb{E}_\Sigma(\widehat{\mathcal{D}}_{n,p}) &= \sup_{\{w_{ij} \ge 0; \sum_{i<j} w_{ij}^2 = \frac{1}{2}\}} \inf_{\Sigma \in Q(\mathcal{F}(\alpha),\varphi)} \frac{1}{p} \sum_{i<j} w_{ij} \sigma_{ij}^2 \\ &= \inf_{\Sigma \in Q(\mathcal{F}(\alpha),\varphi)} \frac{1}{p} \sum_{i<j} \sigma_{ij}^4 =: b(\varphi), \end{aligned} \tag{4.5}$$

where $b(\varphi) \sim C^{\frac{1}{2}}(\alpha) \varphi^{2+\frac{1}{2\alpha}}$ with $C(\alpha) = (2\alpha+1)/((4\alpha+1)^{1+\frac{1}{2\alpha}})$, if $\varphi \to 0$ such that $p\varphi^{1/\alpha} \to +\infty$.

In the next section we introduce a simpler U-statistic for the case of partially observed vectors and give the asymptotic minimax separation rates, then we aggregate these tests in order to construct a procedure free of the parameter α.

4.2.1 Test procedure and minimax separation rate

Let us introduce the asymptotically minimax consistent test procedure with simpler form than $\widehat{\mathcal{D}}_{n,p}$ defined above. For an integer $m \in \mathbb{N}$ large enough, such that it verifies

$$D \le m^\alpha \cdot \varphi \le K^{-2\alpha} \text{ for some constants } D > 1 \text{ and } K > 0, \tag{4.6}$$

we define the following test statistic

$$\widehat{\mathcal{D}}_{n,p,m} = \frac{1}{n(n-1)p} \cdot \frac{1}{\sqrt{2m}} \sum_{1 \le k \ne l \le n} \sum_{\substack{1 \le i < j \le p \\ |i-j| < m}} Y_{k,i} Y_{k,j} Y_{l,i} Y_{l,j} \,. \tag{4.7}$$

Note that, as in [18] we only use m diagonals of the sample covariance matrix $\bar{Y}\bar{Y}^\top$, but the weights are constant and equal to $1/\sqrt{2m}$.

Proposition 4.1. *Under the null hypothesis, the test statistic $\widehat{\mathcal{D}}_{n,p,m}$ in (4.7) is a centered random variable with variance $Var_{a,I}(\widehat{\mathcal{D}}_{n,p,m}) = a^4/(n(n-1)p)$. Moreover,*

$$\frac{n\sqrt{p}}{a^2} \cdot \widehat{\mathcal{D}}_{n,p,m} \to \mathcal{N}(0,1) \quad \text{under } P_I\text{- probability}$$

Under the alternative hypothesis, for all $\Sigma \in \mathcal{F}(\alpha)$ with $\alpha > 1/2$,

$$\mathbb{E}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,T}) = \frac{a^4}{p \cdot \sqrt{2m}} \sum_{\substack{1 \le i < j \le p \\ |i-j| \le m}} \sigma_{ij}^2 \quad \text{and} \quad Var_{a,\Sigma} = \frac{T_1}{n(n-1)p^2} + \frac{T_2}{np^2}$$

where, for $m \to +\infty$ such that $m/p \to 0$ and that (4.6) holds,

$$\begin{aligned} T_1 &\le p \cdot a^4(1 + o(1)) + p \cdot \mathbb{E}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m}) \cdot O(a^2 m\sqrt{m}), \\ T_2 &\le p \cdot \mathbb{E}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m}) \cdot O(a^2\sqrt{m}) + p^{3/2}\Big(\mathbb{E}_{a,\Sigma}^{3/2}(\widehat{\mathcal{D}}_{n,p,m}) O(a^2 m^{3/4}) + \mathbb{E}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m}) \cdot o(a^4)\Big). \end{aligned}$$

Proof of Proposition 4.1. The proof follows the same steps as the proof of Proposition 1 of [18]. We use repeatedly the independence of $(\varepsilon_{k,i})_{k,i}$ and $(X_{k,\cdot})_k$ and obvious properties of the Bernoulli random variables. □

Now, we propose the following test procedure

$$\Delta_m = \Delta_m(t) = \mathbb{1}(\widehat{\mathcal{D}}_{n,p,m} > t), \quad t > 0 \tag{4.8}$$

where $\widehat{\mathcal{D}}_{n,p,m}$ is the test statistic defined in (4.7).

Theorem 4.2.

Upper bound: let $m \to +\infty$ such that $m/p \to 0$ and that (4.6) holds. If $\alpha > 1/2$ and if

$$\varphi \to 0 \;\text{ and }\; a^2 n\sqrt{p}\,\varphi^{2+\frac{1}{2\alpha}} \to +\infty$$

the test procedure defined in (4.8) with $t \le c \cdot a^4 \varphi^{2+\frac{1}{2\alpha}}$ for some constant $c < K(1-D^{-2})/\sqrt{2}$ and with $n\sqrt{p}\,t/a^2 \to +\infty$, is asymptotically minimax consistent, that is

$$\gamma(\Delta_m(t), Q(\mathcal{F}(\alpha), \varphi)) \to 0.$$

Lower bound: if $\alpha > 1/2$ and if

$$a^2 n \to +\infty\,, \quad p = o(1) \cdot (a^2 n)^{4\alpha - 1} \quad \text{and} \quad a^2 n\sqrt{p}\,\varphi^{2+\frac{1}{2\alpha}} \to 0$$

then

$$\gamma = \inf_{\Delta} \gamma(\Delta, Q(\mathcal{F}(\alpha), \varphi)) \to 1$$

Proof of Theorem 4.2. The proof is given in section 4.4. □

From the previous theorem we deduce that the minimax separation rate is given by:

$$\widetilde{\varphi}_\alpha = \left(a^2 n\sqrt{p}\right)^{-\frac{2\alpha}{4\alpha+1}}$$

Thus the separation $\widetilde{\varphi}_\alpha$ obtained for the observations with missing values is slower by the a^2 factor than the separation rate obtained in the case of fully observed vectors.

Note that the conditions on t, the threshold of $\Delta_m(t)$ in (4.8), are compatible. Indeed, $a^2/(n\sqrt{p}) \ll c \cdot a^4 \varphi^{2+1/(2\alpha)}$ is equivalent to our assumption that $a^2 n\sqrt{p}\,\varphi^{2+1/(2\alpha)} \to \infty$.

4.2.2 Adaptation

In this section we construct an asymptotically adaptive consistent test procedure Δ_{ad} free of the parameter $\alpha \in A := [\alpha_*, \alpha^*_{n,p}] \subset]1/2, +\infty[$, with $\alpha^*_{n,p} \to +\infty$ and $\alpha^*_{n,p} = o(1)\ln(a^2 n\sqrt{p})$, to test H_0 given in (4.1) against the large alternative

$$H_1 : \Sigma \in \bigcup_{\alpha\in A} \left\{\mathcal{F}(\alpha)\ ;\ \frac{1}{2}\sum_{i<j} \sigma_{ij}^2 \geq (\mathcal{C}\Phi_\alpha)^2\right\},$$

where $\mathcal{C} > 0$ is some constant and

$$\Phi_\alpha = \left(\frac{\sqrt{\ln\ln(a^2 n\sqrt{p})}}{a^2 n\sqrt{p}}\right)^{\frac{2\alpha}{4\alpha+1}}.$$

For each $\alpha \in [\alpha_*, \alpha^*_{n,p}]$, there exists $l \in \mathbb{N}^*$ such that

$$2^{l-1} \leq (\Phi_\alpha)^{-\frac{1}{\alpha}} < 2^l, \quad \text{it suffices to take } l \sim \frac{\frac{2}{4\alpha+1}\ln(a^2 n\sqrt{p})}{\ln(2)}.$$

Let $L_*, L^* \in \mathbb{N}^*$ be defined by

$$L_* = \left(\frac{2}{(4\alpha^*_{n,p}+1)\ln 2}\right)\ln(a^2 n\sqrt{p}) \quad \text{and} \quad L^* = \left(\frac{2}{(4\alpha_*+1)\ln 2}\right)\ln(a^2 n\sqrt{p}).$$

We see that L_* a,d L^* tend to infinity. We define the adaptive test procedure as follows

$$\Delta_{ad} = \max_{L_*\leq l\leq L^*} \mathbb{1}(\mathcal{D}_{n,p,2^l} > t_l), \tag{4.9}$$

where $\mathcal{D}_{n,p,2^l}$ is the test statistic defined in (4.7), with m replaced by 2^l.

Theorem 4.3. *The test procedure Δ_{ad} defined in (4.9), with $t_l = a^2\dfrac{\sqrt{\mathcal{C}^*\ln l}}{n\sqrt{p}}$, verifies :*
Type I error probability : $\eta(\Delta_{ad}) \to 0$, for $\mathcal{C}^* > 4$.
Type II error probability : if

$$a^2 n\sqrt{p} \to +\infty\,, 2^{L^*}/p \to 0\,, \ln(a^2 n\sqrt{p})/n \to 0 \quad \textit{and } \mathcal{C}^2 > 1 + 4\sqrt{\mathcal{C}^*}$$

we get

$$\beta(\Delta_{ad}\,, \cup_{\alpha\in A} Q(\mathcal{F}(\alpha), \mathcal{C}\Phi_\alpha) \to 0.$$

Note that the condition $2^{L^*}/p \to 0$ is equivalent to $a^2 n \ll p^{2\alpha_*}$.

Proof of Theorem 4.3. The proof of this theorem is similar the the proof of the Theorem 4.7 which is given in section 4.4. □

4.3 Toeplitz covariance matrices

In this section we assume that the covariance matrix Σ is Toeplitz. In this case, we are interested to test (4.1) against the following alternative

$$H_1 : \Sigma \in Q(\mathcal{T}(\alpha), \phi) \tag{4.10}$$

where $Q(\mathcal{T}(\alpha), \phi)$ is defined in (4.4) for ϕ instead of φ. This testing problem is treated in [19], for the particular case $a = 1$, where a weighted U-statistic $\widehat{\mathcal{A}}_{n,p}$ of order 2 is used to construct an asymptotically consistent test procedure that achieve the sharp separation rates. Similarly to the previous setup, we construct an asymptotically consistent test procedure with constant weights. Recall that in [18] the weights are defined as solution of the following optimization problem:

$$\sup_{\{w_j \geq 0; \sum_j w_j^2 = \frac{1}{2}\}} \inf_{\Sigma \in Q(\mathcal{T}(\alpha), \phi)} \sum_{j \geq 1} w_j \sigma_j^2 = \inf_{\Sigma \in Q(\mathcal{T}(\alpha), \phi)} \sum_{j=1}^{p-1} \sigma_j^4 = C^{1/2}(\alpha) \phi^{2 + \frac{1}{2\alpha}} \tag{4.11}$$

Remark that the optimization problems given in (4.5) and (4.11) have the same solution when $\phi \to 0$ such that $p\phi^{1/\alpha} \to +\infty$.

4.3.1 Test procedure and separation rates

Take $m \in \mathbb{N}$ such that $m \to +\infty$ and m verifies (4.6) for ϕ instead of φ, we define the following test statistic:

$$\widehat{\mathcal{A}}_{n,p,m} = \frac{1}{n(n-1)(p-m)^2} \cdot \frac{1}{\sqrt{2m}} \sum_{1 \leq k \neq l \leq n} \sum_{j=1}^{m} \sum_{m+1 \leq i_1, i_2 \leq p} Y_{k,i_1} Y_{k,i_1 - j} Y_{l,i_2} Y_{l,i_2 - j} \tag{4.12}$$

The main difference between the two test statistic $\widehat{\mathcal{D}}_{n,p,m}$ and $\widehat{\mathcal{A}}_{n,p,m}$ is that in this latter we take into consideration the fact that, we have repeated information on the same diagonal elements.

Now, we give bounds on the moments of this test statistic :

Proposition 4.4. *Under the null hypothesis $\widehat{\mathcal{A}}_{n,p,m}$ is a centered random variable whose variance is $Var_{a,I}(\widehat{\mathcal{A}}_{n,p,m}) = a^4/(n(n-1)(p-m)^2)$. Moreover, we have that $(n(p-m)/a^2) \cdot \widehat{\mathcal{A}}_{n,p,m}/ \to \mathcal{N}(0,1)$. Under the alternative hypothesis, for all $\Sigma \in \mathcal{T}(\alpha)$,*

$$\mathbb{E}_{a,\Sigma}(\widehat{\mathcal{A}}_{n,p,m}) = (a^4/\sqrt{2m}) \sum_{j=1}^{m} \sigma_j^2 \text{ and } Var_{a,\Sigma} = \frac{R_1}{n(n-1)(p-m)^4} + \frac{R_2}{n(p-m)^2},$$

where

$$\begin{aligned} R_1 &\leq (p-m)^2 \cdot \{a^4(1+o(1)) + \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_{n,p,m}) \cdot (O(a^2 m) + O(a^3 m^{3/2-2\alpha})) \\ &+ \mathbb{E}^2_\Sigma(\widehat{\mathcal{A}}_{n,p,m}) \cdot O(m^2)\} \\ R_2 &\leq (p-m) \cdot \{a^2 \cdot \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_{n,p,m}) \cdot o(1) + \mathbb{E}^{3/2}_\Sigma(\widehat{\mathcal{A}}_{n,p,m}) \cdot (O(a \cdot m^{1/4}) + O(a^2 m^{3/4-\alpha})) \\ &+ \mathbb{E}^2_\Sigma(\widehat{\mathcal{A}}_{n,p,m}) \cdot O(m)\}. \end{aligned}$$

It is easy to show that, since m verifies (4.6), we have for all $\Sigma \in \mathcal{T}(\alpha)$

$$\mathbb{E}_{a,\Sigma}(\widehat{\mathcal{A}}_{n,p,m}) \geq a^4 B \cdot \phi^{2+\frac{1}{2\alpha}}$$

where B is given in (4.16).

To test (4.1) against (4.10), we define the following test procedure based on the statistic defined in (4.12) :

$$\Delta^T_m = \Delta^T_m(t) = \mathbb{1}(\widehat{\mathcal{A}}_{n,p,m} > t), \quad t > 0. \tag{4.13}$$

Theorem 4.5.

Upper bound: let $m \to +\infty$, such that $m/p \to 0$ and that (4.6) holds. If $\alpha > 1/4$ and if

$$\phi \to 0 \text{ and } a^2 np\phi^{2+\frac{1}{2\alpha}} \to +\infty$$

the test procedure defined in (4.8) with $t \leq \kappa \cdot a^4 \phi^{2+\frac{1}{2\alpha}}$ for some constant $\kappa \leq B$ such that $npt/a^2 \to +\infty$ is consistent, that is $\gamma(\Delta^T_m(t), Q(\mathcal{T}(\alpha), \phi)) \to 0$.

Lower bound: if $\alpha > 1/2$ and if

$$a^2 np \to +\infty \quad \text{and } a^2 np\phi^{2+\frac{1}{2\alpha}} \to 0$$

then

$$\gamma = \inf_\Delta \gamma(\Delta, Q(\mathcal{T}(\alpha), \phi)) \to 1.$$

The main consequence of Theorem 4.5 is that the separation rate is given as follows :

$$\widetilde{\phi} = \left(a^2 np\right)^{-\frac{2\alpha}{4\alpha+1}}$$

Proof of Theorem 4.5. The proof follows the same steps as the proof of Theorem 4.2, we therefore omit it. The most significant difference is that in order to show the lower bound in this case, we consider a sub-class of Toeplitz matrices:

$$Q_T = \{\Sigma_U : [\Sigma_U]_{ij} = \mathbb{1}_{(i=j)} + u_{|i-j|}\sigma \mathbb{1}_{(|1<|i-j|<T)} \text{ for all } 1 \leq i,j \leq p\ ,\ U \in \mathcal{U}\},$$

where σ and T are defined in (4.17) and where

$$\mathcal{U} = \{U = [u_{|i-j|}]_{1 \leq |i-j| \leq p-1} : u_{ii} = 0, \forall i \text{ and } u_{|i-j|} \pm 1 \cdot \mathbb{1}_{(|i-j|<T)}, \text{ for } i \neq j\}.$$

Indeed, the signs are randomized but constant on each diagonal. We re-write the terms of $L_{n,p}$ taking into consideration the fact that the matrices are Toeplitz see, for example, the proof of lower bound in [19]. □

Remarque 4.6. *Remark that the conditions on m imply that m is of order of $\phi^{-\frac{1}{\alpha}}$ in the case of Toeplitz covariance matrices and of order of $\varphi^{-\frac{1}{\alpha}}$ in the case of general covariance matrices.*

4.3.2 Adaptation

In this section, it is always assumed that the covariance matrices are Toeplitz. Our goal is to construct a consistent test procedure independent of the parameter $\alpha \in A := [\alpha_*, \alpha^*_{n,p}] \subset]1/4, +\infty[$, such that $\alpha^*_{n,p} \to +\infty$ and $\alpha^*_{n,p} = o(1)\ln(a^2np)$, to test H_0 given in (4.1) against the large alternative

$$H_1 : \Sigma \in \bigcup_{\alpha \in A} \Big\{ \mathcal{T}(\alpha) \ ; \ \sum_{j=1}^{p-1} \sigma_j^2 \geq (\mathcal{C}\psi_\alpha)^2 \Big\}, \tag{4.14}$$

where $\mathcal{C} > 0$ is some constant and

$$\psi_\alpha = \left(\frac{\sqrt{\ln\ln(a^2np)}}{a^2np} \right)^{\frac{2\alpha}{4\alpha+1}}.$$

First, see that $\forall \alpha \in [\alpha_*, \alpha^*_{n,p}]$, $\exists l \in \mathbb{N}^*$ such that

$$2^{l-1} \leq (\psi_\alpha)^{-\frac{1}{\alpha}} < 2^l, \quad \text{it suffices to take } l \sim \frac{\frac{2}{4\alpha+1}\ln(a^2np)}{\ln(2)}$$

Let L_*, $L^* \in \mathbb{N}^*$ be defined by

$$L_* = \Big(\frac{2}{(4\alpha^*_{n,p} + 1)\ln 2} \Big) \ln(a^2np) \quad \text{and} \quad L^* = \Big(\frac{2}{(4\alpha_* + 1)\ln 2} \Big) \ln(a^2np)$$

We aggregate tests for all given values of l from L_* to L^* giving the following test procedure free of the parameter α:

$$\Delta_{ad} = \max_{L_* \leq l \leq L^*} \mathbb{1}(\mathcal{A}_{n,p,2^l} > t_l), \tag{4.15}$$

where $\mathcal{A}_{n,p,2^l}$ is the test statistic defined in (4.12), with m replaced by 2^l.

Theorem 4.7. *The test procedure Δ_{ad} defined in (4.15), with $t_l = a^2 \dfrac{\sqrt{\mathcal{C}^* \ln l}}{n(p - 2^l)}$, verifies :*
Type I error probability : $\eta(\Delta_{ad}) \to 0$, for $\mathcal{C}^* > 4$.
Type II error probability : if

$$a^2np \to +\infty\,, 2^{L^*}/p \to 0\,, ln(a^2np)/n \to 0 \quad \textit{and } \mathcal{C}^2 \geq 1 + 4\sqrt{\mathcal{C}^*}$$

we get

$$\beta(\Delta_{ad}, \cup_{\alpha \in A} Q(\mathcal{T}(\alpha), \mathcal{C}\psi_\alpha) \to 0.$$

Proof of Theorem 4.7. The proof is given in section 4.4. □

Remark that the condition $2^{L^*}/p$ gives that $a^2n \ll p^{2\alpha_* - \frac{1}{2}}$ and $\ln(a^2np)/n \to 0$ implies that $a^2np \ll e^n$. Together, these conditions are mild as they give $a^2np \ll \min\{p^{2\alpha_* + \frac{1}{2}}, e^n\}$.

4.4 Proofs

Proof of Theorem 4.2. **Upper bound** We use the asymptotic normality of $\widehat{\mathcal{D}}_{n,p,m}$ to show that, the type I error probability

$$\eta(\Delta_m(t)) = P_{a,I}(\Delta_m(t) = 1) = P_{a,I}(\widehat{\mathcal{D}}_{n,p,m} > t) = \Phi\Big(- \frac{n\sqrt{p}t}{a^2}\Big) + o(1) = o(1)$$

as soon as $n\sqrt{p}t/a^2 \to +\infty$. In order to control the maximal type II error probability, we use the Markov inequality to get that for all Σ in $Q(\mathcal{F}(\alpha), \varphi)$:

$$\begin{aligned} P_{a,\Sigma}(\Delta_m(t) = 0) &= P_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m} < t) \le P_{a,\Sigma}\Big(|\widehat{\mathcal{D}}_{n,p,m} - \mathbb{E}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m})| < \mathbb{E}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m}) - t\Big) \\ &\le \frac{\mathrm{Var}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m})}{(\mathbb{E}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m}) - t)^2} = \frac{T_1 + (n-1)T_2}{n(n-1)p^2(\mathbb{E}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m}) - t)^2}, \end{aligned}$$

for t properly chosen. In order to bound the previous quantity uniformly in Σ over $Q(\mathcal{F}(\alpha), \varphi)$ we need to control

$$\inf_{\Sigma \in Q(\mathcal{F}(\alpha),\varphi)} \mathbb{E}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m}) = \inf_{\Sigma \in Q(\mathcal{F}(\alpha),\varphi)} \frac{a^4}{p \cdot \sqrt{2m}} \sum_{\substack{1 \le i < j \le p \\ |i-j| < m}} \sigma_{ij}^2 .$$

For all $\Sigma \in Q(\mathcal{F}(\alpha), \varphi)$, we have

$$\begin{aligned} \mathbb{E}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m}) &= \frac{a^4}{p \cdot \sqrt{2m}} \sum_{\substack{1 \le i < j \le p \\ |i-j| < m}} \sigma_{ij}^2 = \frac{a^4}{p \cdot \sqrt{2m}} \Big(\sum_{1 \le i < j \le p} \sigma_{ij}^2 - \sum_{\substack{1 \le i < j \le p \\ |i-j| \ge m}} \sigma_{ij}^2 \Big) \\ &\ge \frac{a^4}{\sqrt{2m}} \Big(\varphi^2 - \frac{1}{p} \sum_{i<j} \frac{|i-j|^{2\alpha}}{m^{2\alpha}} \sigma_{ij}^2 \Big) \ge \frac{a^4 \varphi^2}{\sqrt{2m}} \Big(1 - \frac{1}{m^{2\alpha} \varphi^2} \Big). \end{aligned}$$

We use (4.6) to get that, for all $\Sigma \in Q(\mathcal{F}(\alpha), \varphi)$

$$\mathbb{E}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m}) \ge \frac{a^4 K}{\sqrt{2}} \cdot \varphi^{2+\frac{1}{2\alpha}} \Big(1 - \frac{1}{D^2} \Big) =: a^4 B \cdot \varphi^{2+\frac{1}{2\alpha}}, \text{ where } B = \frac{K}{\sqrt{2}}(1 - D^{-2}). \quad (4.16)$$

Therefore, take $t \le c \cdot a^4 \varphi^{2+\frac{1}{2\alpha}}$ for $c < B$ and use (4.16) to obtain that

$$\begin{aligned} \frac{T_1}{n(n-1)p^2(\mathbb{E}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m}) - t)^2} &\le \frac{1 + o(1)}{a^4 n(n-1)p\varphi^{4+\frac{1}{\alpha}}(B-c)^2} + \frac{a^2 \cdot O(m\sqrt{m})}{a^4 n(n-1)p\varphi^{2+\frac{1}{2\alpha}}(1 - c/B)^2} \\ &= o(1), \end{aligned}$$

if $a^4 n(n-1)p\varphi^{4+1/\alpha} \to +\infty$ and for all $\alpha > 1/2$. Indeed, $a^2 \cdot m\sqrt{m}\varphi^{2+1/(2\alpha)} \asymp a^2 \varphi^{2-\frac{1}{\alpha}} = o(1)$. Similarly we show that under the previous conditions the term $T_2/np^2(\mathbb{E}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m}) - t)^2$ tends to 0.

Lower bound To show the lower bound we first restrict the class $Q(\mathcal{F}(\alpha), L)$ to the class

$$Q := \{\Sigma_U : [\Sigma_U]_{ij} = \mathbb{1}_{(i=j)} + u_{ij}\sigma\mathbb{1}_{(|1<|i-j|<T)} \text{ for all } 1 \le i, j \le p \ , \ U \in \mathcal{U}\},$$

where

$$\sigma = \varphi^{1+\frac{1}{2\alpha}}, \quad T \asymp \varphi^{-\frac{1}{\alpha}}, \tag{4.17}$$

and

$$\mathcal{U} = \{U = [u_{ij}]_{1\le i,j\le p} : u_{ii} = 0, \forall i \text{ and } u_{ij} = u_{ji} = \pm 1 \cdot \mathbb{1}_{(|i-j|<T)}, \text{ for } i \neq j\},$$

Denote by $\varepsilon_k = (\varepsilon_{k,1}, \ldots, \varepsilon_{k,p})^\top$ the random vector with i.i.d. entries $\varepsilon_{k,i} \sim \mathcal{B}(a)$, for all $1 \le k \le n$. Moreover denote by P_ε and by P_{ε_k} the distributions of $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ and of ε_k, respectively. Recall that the observations $Y_1, \cdots, Y_n$ verify $Y_k = \varepsilon_k * X_k$ for all $1 \le k \le n$, where $*$ designate the Schur product.

Under the null hypothesis $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(0, I)$, thus the conditional random vectors $Y_k|\varepsilon_k$, are independent Gaussian vectors such that, for all $1 \le k \le n$, $Y_k|\varepsilon_k \sim \mathcal{N}(0, I * (\varepsilon_k \varepsilon_k^\top))$. We denote respectively by P_I and by $P_I^{(\varepsilon)}$ the distributions of $(Y_1, \ldots, Y_n)$ and of $(Y_1, \ldots, Y_n)|(\varepsilon_1, \ldots, \varepsilon_n)$ under the null hypothesis. Under the alternative hypothesis, for $X_1, \ldots, X_n \sim \mathcal{N}(0, \Sigma_U)$, we get that the conditional random vectors $Y_k|\varepsilon_k$ are independent Gaussian vectors such that $Y_k|\varepsilon_k \sim \mathcal{N}(0, \Sigma_U * (\varepsilon_k \varepsilon_k^\top))$ for all $1 \le k \le n$, where

$$(\Sigma_U * (\varepsilon_k \varepsilon_k^\top))_{ij} = \begin{cases} \varepsilon_{ik} & \text{for } i = j \\ \varepsilon_{ik}\varepsilon_{jk} \cdot \sigma & \text{if } 1 < |i-j| < T \\ 0 & \text{otherwise.} \end{cases}$$

We denote by $P_U = P_{\Sigma_U}$ and $P_U^{(\varepsilon)} = P_{\Sigma_U}^{(\varepsilon)}$ the distributions of $(Y_1, \ldots, Y_n)$ and of the conditional distribution $(Y_1, \ldots, Y_n)|(\varepsilon_1, \ldots, \varepsilon_n)$ respectively, when $X_1, \ldots, X_n \sim \mathcal{N}(0, \Sigma_U)$.

We define the average distribution over Q by

$$P_\pi = \frac{1}{2^{p(T-1)/2}} \sum_{U \in \mathcal{U}} P_U.$$

It is known (see [58]) that the minimax total error probability satisfies

$$\gamma \ge 1 - \frac{1}{2}\|P_I - P_\pi\|_1$$

In order to prove that $\gamma \longrightarrow 1$, we bound from above the L_1 distance by the Kullback-Leibler divergence (see [90])

$$\|P_I - P_\pi\|_1^2 \le \frac{1}{2} \cdot K(P_I, P_\pi), \quad \text{where } K(P_I, P_\pi) := \mathbb{E}_I \log\Big(\frac{dP_I}{dP_\pi}\Big).$$

Therefore, to complete the proof, it is sufficient to show that $K(P_I, P_\pi) \to 0$. In order to prove this we use the conditional likelihoods as follows:

$$K(P_I, P_\pi) \;=\; \mathbb{E}_I \log\Big(\frac{d(P_\varepsilon \otimes P_I^{(\varepsilon)})}{d(P_\varepsilon \otimes P_\pi^{(\varepsilon)})}\Big) = \mathbb{E}_\varepsilon \mathbb{E}_I^{(\varepsilon)} \log\Big(\frac{dP_I^{(\varepsilon)}}{dP_\pi^{(\varepsilon)}}\Big).$$

Let $\varepsilon(w)$ be a realization of ε, we denote by $S_k \in \{1,\dots,p\}$ the support of $\varepsilon_k(w)$ that is $\varepsilon_{k,i}(w) = 1$ if and only if $i \in S_k$. Also we denote by $d_k = \mathrm{Card}(S_k)$, $\Sigma_U^{\varepsilon_k(w)}$ the positive matrix $\in \mathbb{R}^{d_k \times d_k}$, defined as the sub-matrix of Σ_U obtained by removing all the i-th rows and columns corresponding to $i \notin S_k$ and $X_{\varepsilon_k(w)}$ the sub vector of X_k of dimension d_k in which we retain the coordinate with indices in S_k. Thus,

$$L((Y_1,\dots,Y_n)|\varepsilon(w)) := \mathbb{E}_I^{(\varepsilon(w))} \log\Big(\frac{dP_I^{(\varepsilon(w))}}{dP_\pi^{(\varepsilon(w))}}(Y_1,\dots,Y_n)\Big)$$
$$= \mathbb{E}_I^{(\varepsilon(w))}\Bigg(-\log \mathbb{E}_U \exp\Big(-\frac{1}{2}\sum_{k=1}^n \Big(X_{\varepsilon_k(w)}^\top((\Sigma_U^{\varepsilon_k(w)})^{-1} - I_{\varepsilon_k})X_{\varepsilon_k(w)} + \log\det(\Sigma_U^{\varepsilon_k(w)})\Big)\Big)\Bigg)$$

Therefore we have

$$\mathbb{E}_\varepsilon\Big(L((Y_1,\dots,Y_n)|\varepsilon)\Big)$$
$$= \mathbb{E}_\varepsilon\mathbb{E}_I^{(\varepsilon)}\Bigg(-\log \mathbb{E}_U \exp\Big(-\frac{1}{2}\sum_{k=1}^n \Big(X_{\varepsilon_k}^\top((\Sigma_U^{\varepsilon_k})^{-1} - I_{\varepsilon_k})X_{\varepsilon_k} + \log\det(\Sigma_U^{\varepsilon_k})\Big)\Big)\Bigg)$$

Denote by

$$L_{n,p} := \log \mathbb{E}_U \exp\Big(-\frac{1}{2}\sum_{k=1}^n \Big(X_{\varepsilon_k}^\top((\Sigma_U^{\varepsilon_k})^{-1} - I_{\kappa_k})X_{\varepsilon_k} + \log\det(\Sigma_U^{\varepsilon_k})\Big)\Big) \qquad (4.18)$$

We define $\Delta_U^{\varepsilon_k} = \Sigma_U^{\varepsilon_k} - I^{\varepsilon_k}$, for all $U \in \mathcal{U}$ and any realization of ε_k, we have $tr(\Delta_U^{\varepsilon_k}) = 0$ and $\|\Delta_U^{\varepsilon_k}\| = O(\varphi^{1-\frac{1}{2\alpha}})$ which is $o(1)$, as soon as $\varphi \to 0$ and $\alpha > 1/2$. In fact, by the Gershgorin's theorem we have

$$\|\Delta_U^{\varepsilon_k}\| \le \max_{i\in S_k} \sum_{\substack{j\in S_k\\ j\neq i}} |(\Delta_U^{\varepsilon_k})_{ij}| \le \max_{i\in S_k} \sum_{\substack{j\in S_k\\ 1<|i-j|<T}} |u_{ij}\sigma| \le 2T\cdot\sigma = O(\varphi^{1-\frac{1}{2\alpha}}).$$

For all $x \in [-\frac{1}{2}, +\frac{1}{2}]$ we have the following inequalities

$$\begin{aligned} x - x^2 + x^3 - 2x^4 &\le -\Big(\frac{1}{1+x} - 1\Big) \le x - x^2 + x^3 \\ -x + \frac{x^2}{2} - \frac{x^3}{3} &\le -\log(1+x) \le -x + \frac{x^2}{2} - \frac{x^3}{3} + \frac{x^4}{2}. \end{aligned}$$

Applying these inequalities to the eigenvalues of $\Delta_U^{\varepsilon_k}$ we get

$$\Delta_U^{\varepsilon_k} - (\Delta_U^{\varepsilon_k})^2 + (\Delta_U^{\varepsilon_k})^3 - 2(\Delta_U^{\varepsilon_k})^4 \le -((\Sigma_U^{\varepsilon_k})^{-1} - I^{\varepsilon_k}) \le \Delta_U^{\varepsilon_k} - (\Delta_U^{\varepsilon_k})^2 + (\Delta_U^{\varepsilon_k})^3$$

$$\frac{1}{2}tr(\Delta_U^{\varepsilon_k})^2 - \frac{1}{3}tr(\Delta_U^{\varepsilon_k})^3 \le -\log\det(\Sigma_U^{\varepsilon_k}) \le \frac{1}{2}tr(\Delta_U^{\varepsilon_k})^2 - \frac{1}{3}tr(\Delta_U^{\varepsilon_k})^3 + \frac{1}{2}tr(\Delta_U^{\varepsilon_k})^4,$$

for φ small enough such that $\|\Delta_U^{\varepsilon_k}\| \le 1/2$. Thus we can bound $L_{n,p}$, $\underline{L}_{n,p} \le L_{n,p} \le \bar{L}_{n,p}$

where

$$
\begin{aligned}
\underline{L}_{n,p} \quad &:= \quad \log \mathbb{E}_U \exp\Big(\frac{1}{2}\sum_{k=1}^{n} X_{\varepsilon_k}^{\top}(\Delta_U^{\varepsilon_k} - (\Delta_U^{\varepsilon_k})^2 + (\Delta_U^{\varepsilon_k})^3 - 2(\Delta_U^{\varepsilon_k})^4)X_{\varepsilon_k} \\
&\qquad + \frac{1}{2}\sum_{k=1}^{n}\Big(\frac{1}{2}tr(\Delta_U^{\varepsilon_k})^2 - \frac{1}{3}tr(\Delta_U^{\varepsilon_k})^3\Big)\Big), \text{ and} \\
\bar{L}_{n,p} \quad &:= \quad \log \mathbb{E}_U \exp\Big(\frac{1}{2}\sum_{k=1}^{n} X_{\varepsilon_k}^{\top}(\Delta_U^{\varepsilon_k} - (\Delta_U^{\varepsilon_k})^2 + (\Delta_U^{\varepsilon_k})^3)X_{\varepsilon_k} \\
&\qquad + \frac{1}{2}\sum_{k=1}^{n}\Big(\frac{1}{2}tr(\Delta_U^{\varepsilon_k})^2 - \frac{1}{3}tr(\Delta_U^{\varepsilon_k})^3 + \frac{1}{2}tr(\Delta_U^{\varepsilon_k})^4\Big)\Big).
\end{aligned}
$$

Now we develop the terms of $\bar{L}_{n,p}$

$$
tr(\Delta_U^{\varepsilon_k})^2 = 2\sigma^2 \sum_{\substack{i<j \\ 1<|i-j|<T}} \varepsilon_{k,i}\varepsilon_{k,j}, \qquad tr(\Delta_U^{\varepsilon_k})^3 = 3!\sigma^3 \sum_{\substack{i<i_1<i_2 \\ 1<|i-i_1|,|i_1-i_2|,|i_2-i|<T}} u_{ii_1}u_{i_1i_2}u_{i_2i}\varepsilon_{k,i}\varepsilon_{k,i_1}\varepsilon_{k,i_2}.
$$

and

$$
\begin{aligned}
tr(\Delta_U^{\varepsilon_k})^4 \quad &= \quad \sigma^4 \sum_{\substack{i,i_1,i_2,i_3 \\ 1<|i-i_1|,\dots,|i_3-i|<T}} u_{ii_1}u_{i_1i_2}u_{i_2i_3}u_{i_3i}\varepsilon_{k,i}\varepsilon_{k,i_1}\varepsilon_{k,i_2}\varepsilon_{k,i_3} \\
&= \quad \sigma^4 \sum_{1<|i-i_1|<T} \varepsilon_{k,i}\varepsilon_{k,i_1} + 2\sigma^4 \sum_{\substack{i,i_1,i_2 \\ 1<|i-i_1|,|i_1-i_2|<T}} \varepsilon_{k,i}\varepsilon_{k,i_1}\varepsilon_{k,i_2} \\
&+ \quad 4!\ \sigma^4 \sum_{\substack{i<i_1<i_2<i_3 \\ 1<|i-i_1|,\dots,|i_3-i|<T}} u_{ii_1}u_{i_1i_2}u_{i_2i_3}u_{i_3i}\varepsilon_{k,i}\varepsilon_{k,i_1}\varepsilon_{k,i_2}\varepsilon_{k,i_3}
\end{aligned}
$$

Moreover, we have (recall that $u_{ij}^2 = 1$ and $\varepsilon_{ij}^2 = \varepsilon_{ij}$)

$$
\begin{aligned}
\sum_{k=1}^{n} X_{\varepsilon_k}^{\top}\Delta_U^{\varepsilon_k}X_{\varepsilon_k} \quad &= \quad 2\sigma \cdot \sum_{\substack{i<j \\ 1<|i-j|<T}} u_{ij}\sum_{k=1}^{n}\varepsilon_{k,i}\varepsilon_{k,j}X_{k,i}X_{k,j} \\
\sum_{k=1}^{n} X_{\varepsilon_k}^{\top}(\Delta_U^{\varepsilon_k})^2 X_{\varepsilon_k} \quad &= \quad \sigma^2 \sum_{k=1}^{n}\sum_{i,j} X_{\varepsilon_{k,i}}X_{\varepsilon_{k,j}} \sum_{\substack{i_1 \\ 1<|i_1-i|,|i_1-j|<T}} u_{ii_1}u_{i_1j}\varepsilon_{k,i}\varepsilon_{k,i_1}\varepsilon_{k,j} \\
&= \quad \sigma^2 \sum_{k=1}^{n}\sum_{i} X_{\varepsilon_{k,i}}^2 \sum_{\substack{i_1 \\ 1<|i_1-i|<T}} \varepsilon_{k,i}\varepsilon_{k,i_1} \\
&+ \quad 2\sigma^2 \sum_{k=1}^{n}\sum_{i<j} X_{\varepsilon_{k,i}}X_{\varepsilon_{k,j}} \sum_{\substack{i_1 \\ 1<|i_1-i|,|i_1-j|<T}} u_{ii_1}u_{i_1j}\varepsilon_{k,i}\varepsilon_{k,i_1}\varepsilon_{k,j}\ ,
\end{aligned}
$$

and

$$\begin{aligned}
\sum_{k=1}^{n} X_{\varepsilon_k}^{\top} (\Delta_U^{\varepsilon_k})^3 X_{\varepsilon_k} &= \sigma^3 \sum_{k=1}^{n} \sum_{i,j} X_{\varepsilon_{k,i}} X_{\varepsilon_{k,j}} \sum_{\substack{i_1,i_2 \\ 1<|i-i_1|,|i_1-i_2|,|i_2-j|<T}} u_{ii_1}u_{i_1i_2}u_{i_2j}\varepsilon_{k,i}\varepsilon_{k,i_1}\varepsilon_{k,i_2}\varepsilon_{k,j} \\
&= 2\sigma^3 \sum_{k=1}^{n} \sum_{i} X_{\varepsilon_{k,i}}^2 \sum_{\substack{i_1<i_2 \\ 1<|i-i_1|,|i_1-i_2|,|i_2-i|<T}} u_{ii_1}u_{i_1i_2}u_{i_2i}\varepsilon_{k,i}\varepsilon_{k,i_1}\varepsilon_{k,i_2} \\
&+ 2\sigma^3 \sum_{k=1}^{n} \sum_{\substack{i<j \\ 1<|i-j|<T}} X_{\varepsilon_{k,i}} X_{\varepsilon_{k,j}} \Big(u_{ij}^3 \varepsilon_{k,i}\varepsilon_{k,j} + 2 \sum_{\substack{i_1 \\ 1<|i_1-i|<T}} u_{ij}\varepsilon_{k,i}\varepsilon_{k,i_1}\varepsilon_{k,j} \Big) \\
&+ 2\sigma^3 \sum_{k=1}^{n} \sum_{i<j} X_{\varepsilon_{k,i}} X_{\varepsilon_{k,j}} \sum_{\substack{j\neq i_1\neq i_2\neq i \\ 1<|i-i_1|,|i_1-i_2|,|i_2-j|<T}} u_{ii_1}u_{i_1i_2}u_{i_2j}\varepsilon_{k,i}\varepsilon_{k,i_1}\varepsilon_{k,i_2}\varepsilon_{k,j}
\end{aligned}$$

In consequence, $\bar{L}_{n,p}$ can be written as follows:

$$\begin{aligned}
\bar{L}_{n,p} &= \log \mathbb{E}_U \exp\Big\{ \sum_{\substack{i<j \\ 1<|i-j|<T}} u_{ij} \sum_{k=1}^{n} \varepsilon_{k,i}\varepsilon_{k,j} X_{k,i} X_{k,j} \Big(\sigma + \sigma^3 (1 + 2 \sum_{\substack{i_1 \\ 1<|i_1-i|<T}} \varepsilon_{k,i_1}) \Big) \\
&- \sum_{\substack{i,i_1,j \\ i<j \\ 1<|i_1-i|,|i_1-j|<T}} u_{ii_1}u_{i_1j} \sum_{k=1}^{n} X_{\varepsilon_{k,i}} X_{\varepsilon_{k,j}} \varepsilon_{k,i}\varepsilon_{k,i_1}\varepsilon_{k,j}\, \sigma^2 \\
&+ \sum_{\substack{i<i_1<i_2 \\ 1<|i-i_1|,|i_1-i_2|,|i_2-i|<T}} u_{ii_1}u_{i_1i_2}u_{i_2i} \sum_{k=1}^{n} \varepsilon_{k,i}\varepsilon_{k,i_1}\varepsilon_{k,i_2} \sigma^3 \Big(3X_{\varepsilon_{k,i}}^2 - 1 \Big) \\
&+ \sum_{\substack{i,,i_1,i_2,j \\ i<j \\ j\neq i_1\neq i_2\neq i \\ 1<|i-i_1|,|i_1-i_2|,|i_2-j|<T}} u_{ii_1}u_{i_1i_2}u_{i_2j} \sum_{k=1}^{n} X_{\varepsilon_{k,i}} X_{\varepsilon_{k,j}} \varepsilon_{k,i}\varepsilon_{k,i_1}\varepsilon_{k,i_2}\varepsilon_{k,j}\, \sigma^3 \\
&+ 6 \sum_{\substack{i<i_1<i_2<i_3 \\ 1<|i-i_1|,\ldots,|i_3-i|<T}} u_{ii_1}u_{i_1i_2}u_{i_2i_3}u_{i_3i} \sum_{k=1}^{n} \varepsilon_{k,i}\varepsilon_{k,i_1}\varepsilon_{k,i_2}\varepsilon_{k,i_3}\, \sigma^4 \Big\} \\
&+ \frac{\sigma^2}{4} \sum_{\substack{i<j \\ 1<|i-j|<T}} \sum_{k=1}^{n} \varepsilon_{k,i}\varepsilon_{k,j} - \frac{\sigma^2}{2} \sum_{\substack{i<i_1 \\ 1<|i_1-i|<T}} \sum_{k=1}^{n} X_{\varepsilon_{k,i}}^2 \varepsilon_{k,i}\varepsilon_{k,i_1} \\
&+ \frac{\sigma^4}{4} \sum_{\substack{i,i_1 \\ 1<|i-i_1|<T}} \sum_{k=1}^{n} \varepsilon_{k,i}\varepsilon_{k,i_1} + \frac{\sigma^4}{2} \sum_{\substack{i,i_1,i_2 \\ 1<|i-i_1|,\ldots,|i_2-i|<T}} \sum_{k=1}^{n} \varepsilon_{k,i}\varepsilon_{k,i_1}\varepsilon_{k,i_2}
\end{aligned}$$

We have that $\{u_{ij}\}_{1\leq i<j\leq p}$ is a sequence of i.i.d Rademacher random variables. Note that sequences composed of finite products of i.i.d Rademacher random variables, for example the sequences $\{u_{ir}u_{rj}\}_{\substack{1<i\neq r\neq j\leq p \\ i<j}}$ and $\{u_{ir}u_{rs}u_{sj}\}_{\substack{1\leq i\neq r\neq s\neq j\leq p \\ i<j}}$, form sequences of i.i.d Rademacher random variables. Moreover they are mutually independent and independent from the initial sequence $\{u_{ij}\}_{i<j}$. Now we explicit in $\bar{L}_{n,p}$ the expected value with respect to the i.i.d Rademacher random variables and get

$$
\begin{aligned}
\bar{L}_{n,p} &= \sum_{\substack{i<j\\ 1<|i-j|<T}} \log\cosh\Big(\sum_{k=1}^{n}\varepsilon_{k,i}\varepsilon_{k,j}X_{k,i}X_{k,j}\Big(\sigma+\sigma^3(1+2\sum_{\substack{i_1\\ 1<|i_1-i|<T}}\varepsilon_{k,i_1})\Big)\Big)\\
&+ \frac{\sigma^2}{2}\sum_{\substack{i<j\\ 1<|i-j|<T}}\sum_{k=1}^{n}\varepsilon_{k,i}\varepsilon_{k,j} - \sigma^2\sum_{\substack{i<i_1\\ 1<|i_1-i|<T}}\sum_{k=1}^{n}X^2_{\varepsilon_{k,i}}\varepsilon_{k,i}\varepsilon_{k,i_1}\\
&+ \sum_{\substack{i,i_1,j\\ i<j\\ 1<|i_1-i|,|i_1-j|<T}} \log\cosh\Big(\sum_{k=1}^{n}X_{\varepsilon_{k,i}}X_{\varepsilon_{k,j}}\varepsilon_{k,i}\varepsilon_{k,i_1}\varepsilon_{k,j}\,\sigma^2\Big)\\
&+ \sum_{\substack{i<i_1<i_2\\ 1<|i-i_1|,|i_1-i_2|,|i_2-i|<T}} \log\cosh\Big(3\sigma^3\sum_{k=1}^{n}X^2_{\varepsilon_{k,i}}\varepsilon_{k,i}\varepsilon_{k,i_1}\varepsilon_{k,i_2}-\sigma^3\Big)\\
&+ \sum_{\substack{i,i_1,i_2,j\\ i<j\\ j\neq i_1\neq i_2\neq i\\ 1<|i-i_1|,|i_1-i_2|,|i_2-j|<T}} \log\cosh\Big(\sum_{k=1}^{n}X_{\varepsilon_{k,i}}X_{\varepsilon_{k,j}}\varepsilon_{k,i}\varepsilon_{k,i_1}\varepsilon_{k,i_2}\varepsilon_{k,j}\,\sigma^3\Big)\\
&+ \sum_{\substack{i<i_1<i_2<i_3\\ 1<|i-i_1|,\dots,|i_3-i|<T}} \log\cosh\Big(6\sum_{k=1}^{n}\varepsilon_{k,i}\varepsilon_{k,i_1}\varepsilon_{k,i_2}\varepsilon_{k,i_3}\,\sigma^4\Big) + \frac{\sigma^4}{4}\sum_{\substack{i,i_1\\ 1<|i-i_1|<T}}\sum_{k=1}^{n}\varepsilon_{k,i}\varepsilon_{k,i_1}\\
&+ \frac{\sigma^4}{2}\sum_{\substack{i,i_1,i_2\\ 1<|i-i_1|,\dots,|i_2-i|<T}}\sum_{k=1}^{n}\varepsilon_{k,i}\varepsilon_{k,i_1}\varepsilon_{k,i_2}.
\end{aligned}
$$

We use the inequality $\frac{x^2}{2}-\frac{x^4}{12}\le\log\cosh(x)\le\frac{x^2}{2}$ for all $x\in\mathbb{R}$. Thus,

$$
\begin{aligned}
\bar{L}_{n,p,1} &:= \sum_{\substack{i<j\\ 1<|i-j|<T}} \log\cosh\Big(\sum_{k=1}^{n}\varepsilon_{k,i}\varepsilon_{k,j}X_{k,i}X_{k,j}\Big(\sigma+\sigma^3(1+2\sum_{\substack{i_1\\ 1<|i_1-i|<T}}\varepsilon_{k,i_1})\Big)\Big)\\
&+ \frac{\sigma^2}{2}\sum_{\substack{i<j\\ 1<|i-j|<T}}\sum_{k=1}^{n}\varepsilon_{k,i}\varepsilon_{k,j} - \sigma^2\sum_{\substack{i<i_1\\ 1<|i_1-i|<T}}\sum_{k=1}^{n}X^2_{\varepsilon_{k,i}}\varepsilon_{k,i}\varepsilon_{k,i_1}\\
&\le \frac{1}{2}\sum_{\substack{i<j\\ 1<|i-j|<T}}\Big(\sum_{k=1}^{n}\varepsilon_{k,i}\varepsilon_{k,j}X_{k,i}X_{k,j}\Big(\sigma+\sigma^3(1+2T)\Big)^2\\
&+ \frac{\sigma^2}{2}\sum_{\substack{i<j\\ 1<|i-j|<T}}\sum_{k=1}^{n}\varepsilon_{k,i}\varepsilon_{k,j} - \sigma^2\sum_{\substack{i<i_1\\ 1<|i_1-i|<T}}\sum_{k=1}^{n}X^2_{\varepsilon_{k,i}}\varepsilon_{k,i}\varepsilon_{k,i_1}. \qquad (4.19)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathbb{E}_\varepsilon\mathbb{E}_I^{(\varepsilon)}\Big(\bar{L}_{n,p,1}\Big) &\le \frac{na^2}{2}\sum_{\substack{i<j\\ 1<|i-j|<T}}\Big(\sigma+\sigma^3(1+2T)\Big)^2 + \frac{\sigma^2}{2}\sum_{\substack{i<j\\ 1<|i-j|<T}}na^2 - \sigma^2\sum_{\substack{i<i_1\\ 1<|i_1-i|<T}}na^2\\
&= \frac{na^2}{2}\sum_{\substack{i<j\\ 1<|i-j|<T}}\Big(2\sigma^4(1+2T)+\sigma^6(1+2T)^2\Big) = O(na^2pT^2\sigma^4)+O(na^2pT^3\sigma^6)\\
&= O(a^2np\varphi^4)+O(a^2np\varphi^6) = O\Big((a^2n)^{\frac{-4\alpha+1}{4\alpha+1}}p^{\frac{1}{4\alpha+1}}\Big) = o(1)\,, \qquad (4.20)
\end{aligned}
$$

as soon as $p = o(1)(a^2n)^{4\alpha-1}$. Similarly, we show that

$$\begin{aligned}
\mathbb{E}_\varepsilon\mathbb{E}_I^{(\varepsilon)}\Big(\bar{L}_{n,p,1}\Big) \geq & -\frac{1}{12}\,\mathbb{E}_\varepsilon\mathbb{E}_I^{(\varepsilon)} \sum_{\substack{i<j\\ 1<|i-j|<T}} \Big(\sum_{k=1}^n \varepsilon_{k,i}\varepsilon_{k,j}X_{k,i}X_{k,j}\Big(\sigma+\sigma^3(1+2\sum_{\substack{i_1\\ 1<|i_1-i|<T}}\varepsilon_{k,i})\Big)\Big)^4 + o(1)\\
= & -\frac{a^2n}{12}\cdot\mathbb{E}_I^{(\varepsilon_1)}(X_{k,i}^4X_{k,j}^4)\sum_{\substack{i<j\\ 1<|i-j|<T}}\Big(\sigma+\sigma^3(1+2T)\Big)^4 + o(1)\\
& - \frac{a^4n(n-1)}{12}\cdot 3\mathbb{E}_I^{(\varepsilon_1)}(X_{1,i}^2X_{1,j}^2)\mathbb{E}_I^{(\varepsilon_2)}(X_{2,i}^2X_{2,j}^2)\sum_{\substack{i<j\\ 1<|i-j|<T}}\Big(\sigma+\sigma^3(1+2T)\Big)^4.
\end{aligned}$$

See that the first term was already bounded from above in the previous display and that

$$\begin{aligned}
& \frac{a^4n(n-1)}{12}\cdot 3\mathbb{E}_I^{(\varepsilon_1)}(X_{1,i}^2X_{1,j}^2)\mathbb{E}_I^{(\varepsilon_2)}(X_{2,i}^2X_{2,j}^2)\sum_{\substack{i<j\\ 1<|i-j|<T}}\Big(\sigma+\sigma^3(1+2T)\Big)^4\\
\leq\ & O(a^4n^2)\cdot\Big(pT\sigma^4+pT^2\sigma^6+pT^3\sigma^8+pT^4\sigma^{10}+pT^5\sigma^{12}\Big) = O(a^4n^2p\varphi^{4+\frac{1}{\alpha}}) = o(1),
\end{aligned}$$

as soon as $a^4n^2p\varphi^{4+\frac{1}{\alpha}} \to 0$ and $\alpha > 1/2$. We deduce that $\mathbb{E}_\varepsilon\mathbb{E}_I^{(\varepsilon)}\Big(\bar{L}_{n,p,1}\Big) \geq o(1)$. As consequence $\mathbb{E}_\varepsilon\mathbb{E}_I^{(\varepsilon)}\Big(\bar{L}_{n,p,1}\Big) = o(1)$. Now we treat the second term of $\bar{L}_{n,p}$:

$$\begin{aligned}
\bar{L}_{n,p,2} &:= \sum_{\substack{i,i_1,j\\ i<j\\ 1<|i_1-i|,|i_1-j|<T}} \log\cosh\Big(\sum_{k=1}^n X_{\varepsilon_{k,i}}X_{\varepsilon_{k,j}}\varepsilon_{k,i}\varepsilon_{k,i_1}\varepsilon_{k,j}\,\sigma^2\Big)\\
&\leq \sum_{\substack{i,i_1,j\\ i<j\\ 1<|i_1-i|,|i_1-j|<T}} \Big(\sum_{k=1}^n X_{\varepsilon_{k,i}}X_{\varepsilon_{k,j}}\varepsilon_{k,i}\varepsilon_{k,i_1}\varepsilon_{k,j}\,\sigma^2\Big)^2.
\end{aligned}$$

So,

$$\begin{aligned}
\mathbb{E}_\varepsilon\mathbb{E}_I^{(\varepsilon)}\Big(\bar{L}_{n,p,2}\Big) &\leq \mathbb{E}_\varepsilon\mathbb{E}_I^{(\varepsilon)} \sum_{\substack{i,i_1,j\\ i<j\\ 1<|i_1-i|,|i_1-j|<T}} \Big(\sum_{k=1}^n X_{\varepsilon_{k,i}}X_{\varepsilon_{k,j}}\varepsilon_{k,i}\varepsilon_{k,i_1}\varepsilon_{k,j}\,\sigma^2\Big)^2\\
&= \sum_{\substack{i,i_1,j\\ i<j\\ 1<|i_1-i|,|i_1-j|<T}} na^3\sigma^4 \leq a^3npT\sigma^4 = O(a^3np\varphi^{4+\frac{1}{\alpha}}) = o(1).
\end{aligned}$$

Using the bound from below of $\log\cosh$ inequality, we show that $\mathbb{E}_\varepsilon\mathbb{E}_I^{(\varepsilon)}\Big(\bar{L}_{n,p,2}\Big)$ is bounded from below by a quantity that tends to zero. Therefore we get $\mathbb{E}_\varepsilon\mathbb{E}_I^{(\varepsilon)}\Big(\bar{L}_{n,p,2}\Big) = o(1)$. In a similar way we show that the expected value of the remaining terms with $\log\cosh$ in $\bar{L}_{n,p}$ tend to 0. Finally we have

$$\begin{aligned}
& \mathbb{E}_\varepsilon\mathbb{E}_I^{(\varepsilon)}\Big(\frac{\sigma^4}{4}\sum_{\substack{i,i_1\\ 1<|i-i_1|<T}}\sum_{k=1}^n\varepsilon_{k,i}\varepsilon_{k,i_1} + \frac{\sigma^4}{2}\sum_{\substack{i,i_1,i_2\\ 1<|i-i_1|,\ldots,|i_2-i|<T}}\sum_{k=1}^n\varepsilon_{k,i}\varepsilon_{k,i_1}\varepsilon_{k,i_2}\Big)\\
=\ & O(a^2\sigma^4pTn) + O(a^3\sigma^4pT^2n) = O(a^2np\varphi^{4+\frac{1}{\alpha}}) + O(a^3np\varphi^4) = o(1), \qquad (4.21)
\end{aligned}$$

under the previous conditions. As consequence if $p = o(1)(a^2 n)^{4\alpha-1}$ and if $a^4 n^2 p \varphi^{4+\frac{1}{\alpha}} \to 0$, then

$$\mathbb{E}_\varepsilon \mathbb{E}_I^{(\varepsilon)}\Big(- \bar{L}_{n,p}\Big) = o(1)\,. \tag{4.22}$$

To achieve the proof, we show in a similar way that $\mathbb{E}_\varepsilon \mathbb{E}_I^{(\varepsilon)}\Big(- \underline{L}_{n,p}\Big) = o(1)$. □

Proof of Theorem 4.7. To control the type I error probability, we derive an inequality of Berry-Essen type for $\mathcal{A}_{n,p,2^l}$. For any fixed l in $\mathbb{N}^*$ we denote by $v_{n,p,l} := \mathrm{Var}_{a,I}(\mathcal{A}_{n,p,2^l})$, which gives $v_{n,p,l} \sim a^4/(n^2(p-2^l)^2)$ by Proposition 4.4. Next, we rewrite $\mathcal{A}_{n,p,2^l}$ as follows :

$$\mathcal{A}_{n,p,2^l} = \sum_{1\le k<\ell\le n} H(Y_k, Y_\ell)$$

where,

$$H(Y_k, Y_\ell) = \frac{\sqrt{2}}{n(n-1)(p-2^l)^2} \cdot \frac{1}{\sqrt{2^l}} \sum_{j=1}^{2^l} \sum_{2^l+1\le i_1,i_2\le p} Y_{k,i_1} Y_{k,i_1-j} Y_{\ell,i_2} Y_{\ell,i_2-j}.$$

For $2 \le k, h \le n$, define

$$Z_k = \frac{1}{\sqrt{v_{n,p,l}}} \sum_{\ell=1}^{k-1} H(Y_k, Y_\ell), \quad \text{and} \quad S_h = \sum_{k=2}^{h} Z_k\,.$$

Remark that $\{S_h\}_{h\ge2}$ is a centered martingale with respect to the filtration $\{\mathcal{F}_h\}_{h\ge2}$ where $\mathcal{F}_h$ is the σ-field generated by the random vectors $\{X_1, \dots, X_h\}$. Note that $\mathcal{A}_{n,p,2^l} = \sqrt{v_{n,p,l}} \cdot S_n$ and let $V_n^2 = \sum_{k=2}^n \mathbb{E}_{a,I}(Z_k^2/\mathcal{F}_{k-1})$. We fix $0 < \delta \le 1$ and define

$$J_n = \sum_{k=2}^{n} \mathbb{E}_{a,I}(Z_k)^{2+2\delta} + \mathbb{E}_{a,I}|V_n^2 - 1|^{1+\delta}.$$

We use the Skorokhod representation and Lemma 3.3 in [51] to obtain that, for any $0 < \varepsilon < 1/2$ and any $x \in \mathbb{R}$, there exists a positive constant C depending only on δ such that

$$\begin{aligned} |P_{a,I}(\mathcal{A}_{n,p,2^l} \le x) - \Phi\Big(\frac{x}{\sqrt{v_{n,p,l}}}\Big)| &= \Big|P_{a,I}\Big(S_n \le \frac{x}{\sqrt{v_{n,p,l}}}\Big) - \Phi\Big(\frac{x}{\sqrt{v_{n,p,l}}}\Big)\Big| \\ &\le 16\varepsilon^{1/2} \exp\Big(-\frac{x^2}{4v_{n,p,l}}\Big) + C\cdot\varepsilon^{-1-\delta} J_n. \end{aligned}$$

Then using that $1 - \Phi(u) \le (1/u)\exp(-u^2/2)$ for all $u > 0$, we obtain

$$\begin{aligned} P_{a,I}(\mathcal{A}_{n,p,2^l} > x) &\le \Big(1 - \Phi\Big(\frac{x}{\sqrt{v_{n,p,l}}}\Big)\Big) + 16\varepsilon^{1/2} \exp\Big(-\frac{x^2}{4v_{n,p,l}}\Big) + C\cdot\varepsilon^{-1-\delta} J_n \\ &\le \Big(\frac{1}{x/\sqrt{v_{n,p,l}}} + 16\varepsilon^{1/2}\Big) \exp\Big(-\frac{x^2}{4v_{n,p,l}}\Big) + C\cdot\varepsilon^{-1-\delta} J_n. \end{aligned} \tag{4.23}$$

Choose $\delta = 1$, then

$$J_n = \sum_{k=2}^{n} \mathbb{E}_{a,I}(Z_k)^4 + \mathbb{E}_{a,I}|V_n^2 - 1|^2.$$

We can show that

$$\sum_{k=2}^{n} \mathbb{E}_{a,I}(Z_k)^4 = O\Big(\frac{1}{n}\Big) \quad \text{and} \quad \mathbb{E}_{a,I}|V_n^2 - 1|^2 = O\Big(\frac{1}{n}\Big) + O\Big(\frac{1}{2^l}\Big) \tag{4.24}$$

Take $t_l = a^2 \dfrac{\sqrt{\mathcal{C}^* \ln l}}{n(p-2^l)}$, we use (4.23) and (4.24) to bound from above the type I error probability:

$$\begin{aligned}
P_{a,I}(\Delta_{ad} = 1) &= P_{a,I}(\exists l \in \mathbb{N}, L_* \le l \le L^*; \mathcal{A}_{n,p,2^l} > t_l) \le \sum_{L_* \le l \le L^*} P_{a,I}(\mathcal{A}_{n,p,2^l} > t_l) \\
&\le \sum_{L_* \le l \le L^*} \Bigg(\Big(\frac{a^2}{n(p-2^l)t_l} + 16\varepsilon^{1/2}\Big) \exp\Big(-\frac{t_l^2}{4v_{n,p}}\Big) + \frac{O(1)}{\varepsilon^2}\Big(\frac{1}{n} + \frac{1}{2^l}\Big)\Bigg) \\
&\le \sum_{l \ge L_*} \Big(\frac{1}{\sqrt{\mathcal{C}^* \ln l}} + 16\varepsilon^{1/2}\Big) \exp\Big(-\frac{\mathcal{C}^* \ln l}{4}\Big) + O(1)\frac{L^* - L_*}{n\varepsilon^2} + \frac{O(1)}{\varepsilon^2} \sum_{l \ge L_*} \frac{1}{2^l} \\
&\le \sum_{l \ge L_*} \Big(\frac{1}{\sqrt{\mathcal{C}^* \ln l}} + 16\varepsilon^{1/2}\Big) l^{-\mathcal{C}^*/4} + +O(1)\frac{L^*}{n\varepsilon^2} + \frac{O(1)2^{-L_*}}{\varepsilon^2} = o(1),
\end{aligned}$$

for $\mathcal{C}^* > 4$ and since L_* and L^* both tend to infinity, such that $\ln(a^2 n\sqrt{p})/n$ tends to 0.

Now, we control the type II error probability. Assume that $\Sigma \in \mathcal{T}(\alpha)$ and that α is such that there exists $l_0 \in \{L_*, L^*\}$ such that $2^{l_0 - 1} \le (\psi_\alpha)^{-\frac{1}{\alpha}} < 2^{l_0}$, thus

$$\begin{aligned}
\mathbb{E}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,2^{l_0}}) &= \frac{a^4}{\sqrt{2 \cdot 2^{l_0}}}\Big(\sum_{1 \le j < p} \sigma_j^2 - \sum_{2^{l_0} < j < p} \sigma_j^2\Big) \\
&\ge \frac{a^4}{2(\psi_\alpha)^{-\frac{1}{2\alpha}}}\Big(\mathcal{C}^2 \psi_\alpha^2 - \sum_j \frac{j^{2\alpha}}{(2^{l_0})^{2\alpha}} \sigma_j^2\Big) \ge (\psi_\alpha)^{2 + \frac{1}{2\alpha}} \cdot \frac{a^4}{2}\Big(\mathcal{C}^2 - 1\Big).
\end{aligned}$$

We assumed that $a^2 np(\psi_\alpha)^{2+\frac{1}{2\alpha}} = \sqrt{\ln\ln(a^2 np)}$. Moreover, we have

$$t_{l_0} \le \frac{a^2\sqrt{\mathcal{C}^* \ln L^*}}{n(p - 2^{l_0})} \le \frac{a^2\sqrt{\mathcal{C}^* \ln\ln(a^2 np)}}{n(p - 2^{l_0})} \le 2\sqrt{\mathcal{C}^*} a^4 (\psi_\alpha)^{2+\frac{1}{2\alpha}}.$$

Thus, we have $\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}}) - t_{l_0} \ge a^4(\psi_\alpha)^{2+1/(2\alpha)}(\mathcal{C}^2 - 1 - 4\sqrt{\mathcal{C}^*})/2$ by our assumption that $\mathcal{C}^2 > 1 + 4\sqrt{\mathcal{C}^*}$. Therefore we get

$$\begin{aligned}
P_{a,\Sigma}(\Delta_{ad} = 0) &= P_{a,\Sigma}(\forall l \in \{L_*, L^*\}; \mathcal{A}_{n,p,2^l} < t_l) \le P_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}} < t_{l_0}) \\
&\le P_{a,\Sigma}(|\mathcal{A}_{n,p,2^{l_0}} - \mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}})| > \mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}}) - t_{l_0}).
\end{aligned}$$

Now, we use Markov inequality and get :

$$\begin{aligned}
P_{a,\Sigma}(\Delta_{ad} = 0) &\le \frac{\mathrm{Var}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}})}{(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}}) - t_{l_0})^2} \\
&\le \frac{R_1 + (n-1)(p - 2^{l_0})^2 R_2}{n(n-1)(p - 2^{l_0})^4(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}}) - t_{l_0})^2}.
\end{aligned} \tag{4.25}$$

First, we bound from above the first term of (4.25), using Proposition 4.4

$$\begin{aligned} S_1 &:= \frac{R_1}{n(n-1)(p-2^{l_0})^4(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}})-t_{l_0})^2} \\ &= \frac{a^4(1+o(1))+\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_{n,p,2^{l_0}})\cdot(O(a^2\sqrt{2^{l_0}})+O(a^3(2^{l_0})^{3/2-2\alpha}))}{n(n-1)(p-2^{l_0})^2(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}})-t_{l_0})^2} \\ &+ \frac{\mathbb{E}^2_\Sigma(\widehat{\mathcal{A}}_{n,p,2^{l_0}})\cdot O(m^2/a)}{n(n-1)(p-2^{l_0})^2(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}})-t_{l_0})^2} \end{aligned}$$

We decompose S_1 as sum of three terms: the first one

$$\begin{aligned} S_{1,1} &:= \frac{a^4(1+o(1))}{n(n-1)(p-2^{l_0})^2(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}})-t_{l_0})^2} \\ &\le \frac{a^4(1+o(1))}{n(n-1)(p-2^{l_0})^2a^8(\psi_\alpha)^{4+\frac{1}{\alpha}}\left(\mathcal{C}^2-1-4\sqrt{\mathcal{C}^*}\right)^2} \\ &= O\Big(\frac{1}{\ln\ln(a^2np)}\Big) = o(1)\,, \text{ as soon as } a^2np\to+\infty. \end{aligned}$$

Now we show that the second term of S_1 also tends to 0. Recall that $2^{l_0}\asymp(\psi_\alpha)^{-\frac{1}{\alpha}}$, therefore

$$\begin{aligned} S_{1,2} &:= \frac{O(a^2\,2^{l_0})+O(a^3(2^{l_0})^{3/2-2\alpha})}{n(n-1)(p-2^{l_0})^2\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}})(1-t_{l_0}/\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}}))^2} \\ &\le \frac{(O(a^2\,2^{l_0})+O(a^3(2^{l_0})^{3/2-2\alpha}))}{n(n-1)(p-2^{l_0})^2\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}})\left(1-\frac{4\sqrt{\mathcal{C}^*}}{\mathcal{C}^2-1}\right)^2} \\ &\le \frac{O(2^{l_0})+O(a(2^{l_0})^{3/2-2\alpha})}{n(n-1)(p-2^{l_0})^2a^2(\psi_\alpha)^{2+\frac{1}{2\alpha}}\left(1-\frac{4\sqrt{\mathcal{C}^*}}{\mathcal{C}^2-1}\right)^2} \\ &\le \frac{O(\sqrt{2^{l_0}}\cdot(\psi_\alpha)^{2+\frac{1}{2\alpha}})+O((2^{l_0})^{3/2-2\alpha}\cdot(\psi_\alpha)^{2+\frac{1}{2\alpha}})}{\ln\ln(a^2np)} = o(1). \end{aligned}$$

since $2^{l_0}\cdot(\psi_\alpha)^{2+\frac{1}{2\alpha}}\asymp(\psi_\alpha)^{2-\frac{1}{2\alpha}}=o(1)$ and $(2^{l_0})^{3/2-2\alpha}\cdot(\psi_\alpha)^{2+\frac{1}{2\alpha}}\asymp(\psi_\alpha)^{4-\frac{1}{\alpha}}=o(1)$ for all $\alpha>1/4$. Finally,

$$\begin{aligned} S_{1,3} &:= \frac{\mathbb{E}^2_\Sigma(\widehat{\mathcal{A}}_{n,p,2^{l_0}})\cdot O((2^{l_0})^2)}{n(n-1)(p-2^{l_0})^2(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}})-t_{l_0})^2} \\ &= \frac{O((2^{l_0})^2)}{n(n-1)p^2} = o(1). \end{aligned}$$

Now, we bound from above the second term of (4.25).

$$\begin{aligned} S_2 &= \frac{R_2}{n(p-2^{l_0})^2(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}})-t_{l_0})^2} \\ &= \frac{a^2\cdot\mathbb{E}_{\Sigma}(\widehat{\mathcal{A}}_{n,p,2^{l_0}})\cdot o(1)}{n(p-2^{l_0})(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}})-t_{l_0})^2} + \frac{\mathbb{E}^2_{\Sigma}(\widehat{\mathcal{A}}_{n,p,2^{l_0}})\cdot O(2^{l_0})}{n(p-2^{l_0})(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}})-t_{l_0})^2} \\ &+ \frac{\mathbb{E}^{3/2}_{\Sigma}(\widehat{\mathcal{A}}_{n,p,2^{l_0}})\cdot(O(a\cdot(2^{l_0})^{1/4})+O(a^2(2^{l_0})^{3/4-\alpha}))}{n(p-2^{l_0})(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}})-t_{l_0})^2}. \end{aligned}$$

We bound from above each term of S_2. For the first term,

$$\begin{aligned} S_{2,1} &:= \frac{a^2\cdot\mathbb{E}_{\Sigma}(\widehat{\mathcal{A}}_{n,p,2^{l_0}})\cdot o(1)}{n(p-2^{l_0})(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}})-t_{l_0})^2} \le \frac{o(1)}{n(p-2^{l_0})a^2(\psi_\alpha)^{2+\frac{1}{2\alpha}}} \\ &= \frac{o(1)}{\sqrt{\ln\ln(a^2np)}} = o(1). \end{aligned}$$

For the second term we have,

$$S_{2,2} := \frac{\mathbb{E}^2_{\Sigma}(\widehat{\mathcal{A}}_{n,p,2^{l_0}})\cdot O(2^{l_0})}{n(p-2^{l_0})(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}})-t_{l_0})^2} \le \frac{O(2^{l_0})}{np} = o(1)$$

Finally for the last term,

$$\begin{aligned} S_{2,3} &:= \frac{\mathbb{E}^{3/2}_{\Sigma}(\widehat{\mathcal{A}}_{n,p,2^{l_0}})\cdot(O(a\cdot(2^{l_0})^{1/4})+O(a^2(2^{l_0})^{3/4-\alpha}))}{n(p-2^{l_0})(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}})-t_{l_0})^2} \\ &\le \frac{O((2^{l_0})^{1/4})+O(a\cdot(2^{l_0})^{3/4-\alpha})}{n(p-2^{l_0})a\psi_\alpha^{1+\frac{1}{4\alpha}}} \\ &\le \frac{O((2^{l_0})^{1/4})}{\sqrt{n(p-2^{l_0})}\,(\ln\ln(a^2np))^{\frac{1}{4}}} + \frac{O(a^2\cdot\psi_\alpha^{1+\frac{1}{4\alpha}}\cdot(2^{l_0})^{3/4-\alpha})}{\sqrt{\ln\ln(a^2np)}} = o(1), \end{aligned}$$

as $a^2\cdot\psi^{1+\frac{1}{4\alpha}}\cdot(2^{l_0})^{3/4-\alpha}\asymp\psi_\alpha^{2-\frac{1}{2\alpha}} = o(1)$. □

Bibliography

[1] T. W. Anderson. *An introduction to multivariate statistical analysis.* Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, third edition, 2003.

[2] G. Aneiros and Ph. Vieu. Variable selection in infinite-dimensional problems. *Statist. Probab. Lett.*, 94:12–20, 2014.

[3] F. Autin and C. Pouet. Adaptive test on components of densities mixture. *Math. Methods Statist.*, 21(2):93–108, 2012.

[4] Z. Bai, D. Jiang, J.-F Yao, and S. Zheng. Corrections to lrt on large-dimensional covariance matrix by rmt. *The Annals of Statistics*, 37(6B):3822–3840, 12 2009.

[5] Y. Baraud. Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*, 8(5):577–606, 2002.

[6] Y. Baraud, S. Huet, and B. Laurent. Testing convex hypotheses on the mean of a Gaussian vector. Application to testing qualitative hypotheses on a regression function. *Ann. Statist.*, 33(1):214–257, 2005.

[7] Q. Berthet and Ph. Rigollet. Optimal detection of sparse principal components in high dimension. *Ann. Statist.*, 41(4):1780–1815, 2013.

[8] P. Bianchi, M. Debbah, M. Maida, and J. Najim. Performance of statistical tests for single-source detection using random matrix theory. *IEEE Trans. Inform. Theory*, 57(4):2400–2419, 2011.

[9] Peter J. Bickel and Elizaveta Levina. Regularized estimation of large covariance matrices. *Ann. Statist.*, 36(1):199–227, 2008.

[10] P.J. Bickel and E. Levina. Covariance regularization by thresholding. *Ann. Statist.*, 36(6):2577–2604, 2008.

[11] H. Böhm and R. von Sachs. Structural shrinkage of nonparametric spectral estimators for multivariate time series. *Electron. J. Stat.*, 2:696–721, 2008.

[12] H. Böhm and R. von Sachs. Shrinkage estimation in the frequency domain of multivariate time series. *J. Multivariate Anal.*, 100(5):913–935, 2009.

[13] E.G. Bongiorno, E. Salinelli, A. Goia, and Ph. Vieu. *Contributions in infinite-dimensional statistics and related topics:*. Esculapio, 2014.

[14] C. Butucea and G. Gayraud. Sharp detection of smooth signals in a high-dimensional sparse matrix with indirect observations. *ArXiv e-prints*, jan 2013.

[15] C. Butucea, C. Matias, and C. Pouet. Adaptivity in convolution models with partially known noise distribution. *Electron. J. Stat.*, 2:897–915, 2008.

[16] C. Butucea, C. Matias, and C. Pouet. Adaptive goodness-of-fit testing from indirect observations. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45(2):352–372, 2009.

[17] C. Butucea and K. Meziani. Quadratic functional estimation in inverse problems. *Stat. Methodol.*, 8(1):31–41, 2011.

[18] C. Butucea and R. Zgheib. Sharp minimax tests for large covariance matrices and adaptation. *Electronic Journal of Statistics, tentatively accepted*, 2015.

[19] C. Butucea and R. Zgheib. Sharp minimax tests for large toeplitz covariance matrices with repeated observations. *Journal of Multivariate Analysis, accepted*, 2015.

[20] C. Butucea and R. Zgheib. Adaptive test for large covariance matrices with missing observations. *ArXiv e-prints*, February 2016.

[21] T. Cai, Z. Ren, and H. Zhou. Optimal rates of convergence for estimating toeplitz covariance matrices. *Probab. Theory Relat. Fields*, 156:101–143, 2013.

[22] T. T. Cai and T. Jiang. Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.*, 39(3):1496–1525, 2011.

[23] T. T. Cai and Z. Ma. Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli*, 19(5B):2359–2388, 11 2013.

[24] T. T. Cai, Z. Ren, and H. H. Zhou. Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation. *Electron. J. Statist.*, 10(1):1–59, 2016.

[25] T. T. Cai and M. Yuan. Adaptive covariance matrix estimation through block thresholding. *Ann. Statist.*, 40(4):2014–2042, 2012.

[26] T. T. Cai, C.-H Zhang, and H. H. Zhou. Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.*, 38(4):2118–2144, 2010.

[27] T. T. Cai and H. H. Zhou. Minimax estimation of large covariance matrices under ℓ_1-norm. *Statist. Sinica*, 22(4):1319–1349, 2012.

[28] K. Chen, K. Chen, H.-G. Müller, and J.-L. Wang. Stringing high-dimensional data for functional analysis. *J. Amer. Statist. Assoc.*, 106(493):275–284, 2011.

[29] S. X. Chen, L.-X. Zhang, and P.-S. Zhong. Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.*, 105(490):810–819, 2010.

[30] L. Comminges and A. S. Dalalyan. Minimax testing of a composite null hypothesis defined via a quadratic functional in the model of regression. *Electron. J. Stat.*, 7:146–190, 2013.

[31] F. Comte. Adaptive estimation of the spectrum of a stationary gaussian sequence. *Bernoulli*, 7(2):pp. 267–298, 2001.

[32] A. Cuevas. A partial overview of the theory of statistics with functional data. *J. Statist. Plann. Inference*, 147:1–23, 2014.

[33] L. Dümbgen and V. G. Spokoiny. Multiscale testing of qualitative hypotheses. *Ann. Statist.*, 29(1):124–152, 2001.

[34] M. S. Ermakov. Asymptotically minimax criteria for testing nonparametric hypotheses on the density of a distribution. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)*, 166(Issled. po Mat. Statist. 8):44–53, 187, 1988.

[35] M. S. Ermakov. Minimax detection of a signal in Gaussian white noise. *Teor. Veroyatnost. i Primenen.*, 35(4):704–715, 1990.

[36] M. S. Ermakov. Minimax nonparametric testing of hypotheses on a distribution density. *Teor. Veroyatnost. i Primenen.*, 39(3):488–512, 1994.

[37] M. S. Ermakov. A minimax test for hypotheses on a spectral density. *Journal of Mathematical Science*, 68(4):475–483, 1994.

[38] M. Fiecas and R. von Sachs. Data-driven shrinkage of the spectral density matrix of a high-dimensional time series. *Electron. J. Stat.*, 8(2):2975–3003, 2014.

[39] T. J. Fisher. On testing for an identity covariance matrix when the dimensionality equals or exceeds the sample size. *J. Statist. Plann. Inference*, 142(1):312–326, 2012.

[40] T. J. Fisher and C. M. Gallagher. New weighted portmanteau statistics for time series goodness of fit testing. *Journal of the American Statistical Association*, 107(498):777–787, 2012.

[41] T. J. Fisher, X. Sun, and C.M. Gallagher. A new test for sphericity of the covariance matrix for high dimensional data. *Journal of Multivariate Analysis*, 101(10):2554 – 2570, 2010.

[42] M. Fromont and B. Laurent. Adaptive goodness-of-fit tests in a density model. *Ann. Statist.*, 34(2):680–720, 2006.

[43] G. Gayraud and C. Pouet. Adaptive minimax testing in the discrete regression scheme. *Probab. Theory Related Fields*, 133(4):531–558, 2005.

[44] G. Gayraud and Ch. Pouet. Minimax testing composite null hypotheses in the discrete regression scheme. *Math. Methods Statist.*, 10(4):375–394 (2002), 2001. Meeting on Mathematical Statistics (Marseille, 2000).

[45] E. Giné and R. Nickl. *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2015.

[46] G. Golubev. Nonparametric estimation of smooth spectral densities of gaussian stationary sequences. *Theory of Probability & Its Applications*, 38(4):630–639, 1994.

[47] G.K. Golubev, M. Nussbaum, and H.H. Zhou. Asymptotic equivalence of spectral density estimation and gaussian white noise. *The Annals of Statistics*, 38:181–214, 2010.

[48] Y. K. Golubev, B. Y. Levit, and A. B. Tsybakov. Asymptotically efficient estimation of analytic functions in gaussian noise. *Bernoulli*, 2(2):167–181, 06 1996.

[49] A. K. Gupta and T. Bodnar. An exact test about the covariance matrix. *Journal of Multivariate Analysis*, 125(0):176 – 189, 2014.

[50] P. Hall. Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.*, 14(1):1–16, 1984.

[51] P. Hall and C. C. Heyde. *Martingale limit theory and its application*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1980. Probability and Mathematical Statistics.

[52] Yu. Ingster and I. Suslina. On estimation and detection of smooth functions of many variables. *Math. Methods Statist.*, 14(3):299–331, 2005.

[53] Yu. I. Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives. I. *Math. Methods Statist.*, 2(2):85–114, 1993.

[54] Yu. I. Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives. II. *Math. Methods Statist.*, 2(3):171–189, 1993.

[55] Yu. I. Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives. III. *Math. Methods Statist.*, 2(4):249–268, 1993.

[56] Yu. I. Ingster and T. Sapatinas. Minimax goodness-of-fit testing in multivariate nonparametric regression. *Math. Methods Statist.*, 18(3):241–269, 2009.

[57] Yu. I. Ingster and N. Stepanova. Estimation and detection of functions from weighted tensor product spaces. *Math. Methods Statist.*, 18(4):310–340, 2009.

[58] Yu. I. Ingster and I. A. Suslina. *Nonparametric goodness-of-fit testing under Gaussian models*, volume 169 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 2003.

[59] Yu. I. Ingster and I. A. Suslina. On estimation and detection of a function of infinitely many variables. *Journal of Mathematical Sciences*, 139(3):6548–6561, 2006.

[60] Yuri Ingster and Natalia Stepanova. Estimation and detection of functions from anisotropic Sobolev classes. *Electron. J. Stat.*, 5:484–506, 2011.

[61] V. G. Spokoiny J. L. Horowitz. An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, 69(3):599–631, 2001.

[62] D. Jiang, T. Jiang, and F. Yang. Likelihood ratio tests for covariance matrices of high-dimensional normal distributions. *J. Statist. Plann. Inference*, 142(8):2241–2256, 2012.

[63] T. Jiang. The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.*, 14(2):865–880, 2004.

[64] I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29(2):295–327, 2001.

[65] K. Jurczak and A. Rohde. Spectral analysis of high-dimensional sample covariance matrices with missing observations. *ArXiv e-prints*, July 2015.

[66] B. Laurent, J.-M. Loubes, and C. Marteau. Non asymptotic minimax rates of testing in signal detection with heterogeneous variances. *Electron. J. Stat.*, 6:91–122, 2012.

[67] O. Ledoit and M. Wolf. Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.*, 30(4):1081–1102, 2002.

[68] E. L. Lehmann and Joseph P. Romano. *Testing statistical hypotheses*. Springer Texts in Statistics. Springer, New York, third edition, 2005.

[69] O. V. Lepski and V.G. Spokoiny. Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative. *Bernoulli*, 5(2):333–358, 1999.

[70] O. V. Lepski and A. B. Tsybakov. Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probab. Theory Related Fields*, 117(1):17–48, 2000.

[71] O. V. Lepskiĭ. Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Teor. Veroyatnost. i Primenen.*, 36(4):645–659, 1991.

[72] O. V. Lepskiĭ. Asymptotically minimax adaptive estimation. II. Schemes without optimal adaptation. Adaptive estimates. *Teor. Veroyatnost. i Primenen.*, 37(3):468–481, 1992.

[73] K. Lounici. High-dimensional covariance matrix estimation with missing observations. *Bernoulli*, 20(3):1029–1058, 2014.

[74] C. Marteau and T. Sapatinas. A unified treatment for non-asymptotic and asymptotic approaches to minimax signal detection. *ArXiv e-prints*, jun 2014.

[75] John W. Mauchly. Significance test for sphericity of a normal n-variate distribution. *Ann. Math. Statistics*, 11:204–209, 1940.

[76] Robb J. Muirhead. *Aspects of multivariate statistical theory*. John Wiley & Sons, Inc., New York, 1982. Wiley Series in Probability and Mathematical Statistics.

[77] H. Nagao. On some test criteria for covariance matrix. *The Annals of Statistics*, 1(4):700–709, 07 1973.

[78] M. H. Neumann. Spectral density estimation via nonlinear wavelet methods for stationary non-gaussian time series. *Journal of Time Series Analysis*, 17(6):601–633, 1996.

[79] C. Pouet. An asymptotically optimal test for a parametric set of regression functions against a non-parametric alternative. *J. Statist. Plann. Inference*, 98(1-2):177–189, 2001.

[80] Christophe Pouet. Test asymptotiquement minimax pour une hypothèse nulle composite dans le modèle de densité. *C. R. Math. Acad. Sci. Paris*, 334(10):913–916, 2002.

[81] Y. Qiu and S. X. Chen. Test for bandedness of high-dimensional covariance matrices and bandwidth estimation. *Ann. Statist.*, 40(3):1285–1314, 06 2012.

[82] T. Schneider. Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *J. Climate*, 14:853–871, 2001.

[83] A. N. Shiryaev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1996. Translated from the first (1980) Russian edition by R. P. Boas.

[84] Ph. Soulier. Adaptive estimation of the spectral density of a weakly or strongly dependent Gaussian process. *Math. Methods Statist.*, 10(3):331–354, 2001. Meeting on Mathematical Statistics (Marseille, 2000).

[85] V. G. Spokoiny. Adaptive hypothesis testing using wavelets. *Ann. Statist.*, 24(6):2477–2498, 1996.

[86] M. S. Srivastava. Some tests concerning the covariance matrix in high dimensional data. *J. Japan Statist. Soc.*, 35(2):251–272, 2005.

[87] M. S. Srivastava. Some tests criteria for the covariance matrix with fewer observations than the dimension. *Acta Comment. Univ. Tartu. Math.*, (10):77–93, 2006.

[88] M. S. Srivastava, Tõnu Kollo, and Dietrich von Rosen. Some tests for the covariance matrix with fewer observations than the dimension under non-normality. *J. Multivariate Anal.*, 102(6):1090–1103, 2011.

[89] M. S. Srivastava, H. Yanagihara, and T. Kubokawa. Tests for covariance matrices in high dimension with less sample size. *Journal of Multivariate Analysis*, 130(0):289 – 309, 2014.

[90] A. B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

[91] C. Wang, L. Cao, and B. Miao. Asymptotic power of likelihood ratio tests for high dimensional data. *ArXiv e-prints*, feb 2013.

[92] H. Xiao and W. Biao Wu. Simultaneous Inference of Covariances. *ArXiv e-prints*, sep 2011.

[93] H. Xiao and W.B. Wu. Asymptotic theory for maximum deviations of sample covariance matrix estimation. *Stochastic Processes and their Applications*, 123:2899–2920, 2013.