Statistical Applications of Random Matrix Theory: Comparison of Two Populations

Thèse N° 9597

Présentée le 20 septembre 2019

à la Faculté des sciences de base
Chaire de statistique appliquée
Programme doctoral en mathématiques

pour l'obtention du grade de Docteur ès Sciences

par

Rémy MARIÉTAN

Acceptée sur proposition du jury

Prof. J. Krieger, président du jury
Prof. S. Morgenthaler, directeur de thèse
Prof. C. Mazza, rapporteur
Prof. M. G. Genton, rapporteur
Prof. A. C. Davison, rapporteur

2019

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Acknowledgements

Il y a de nombreuses personnes qui m'ont aidé de manière directe ou indirecte à réaliser ma thèse et que je souhaite remercier.

Premièrement, j'ai eu la chance de réaliser mon travail de doctorat avec un homme extraordinaire, Stephan Morgenthaler. En plus d'être un brillant professeur d'EPFL qui m'a éclairé dans le long tunnel sombre qu'est la thèse, Stephan est aussi un grand philanthrope, passionné et ouvert à la discussion. J'ai eu beaucoup de plaisir à le côtoyer durant ces années. Grâce à lui, j'ai pu développer mes connaissances des statistiques au-delà de mon sujet de thèse en participant entre autres à ses cours plus poussés.

Je voudrais aussi remercier mon jury de thèse, Anthony Davison, Christian Mazza et Marc Genton qui ont eu le courage de parcourir ce long ouvrage pour en améliorer le contenu. Je remercie également le président de ce jury, Joachim Krieger.

Durant ces années, l'épanouissement statistique a été rendu possible grâce à mes collègues de STAT/STAP/SMAT. Je tiens à remercier particulièrement Nadia Kaiser et Anna Dietler. En plus de me sauver la vie au niveau administratif, elles ont toujours été disponibles, joyeuses et motivantes.

Au début de mon doctorat, j'ai eu la chance de rencontrer deux doctorants incroyables qui ont vraiment illuminé mon début de thèse, Marie-Hélène et Raphaël Gérard Théodore Michel Marie. Je les remercie pour nos soirées, nos verres et tous ces bons moments.

Je tiens particulièrement à remercier Jacques Saliba, ce grand probabiliste plein de bonnes idées, pour toutes nos pauses de midi et cafés. Sa bonne humeur a vraiment apaisé les moments les plus difficiles de la thèse. Moments difficiles qui ne duraient pas longtemps, car il suffisait alors de demander à Yoav ! La solution se trouvait alors dans nos discussions très passionnantes de votations ou de mesure... Je remercie donc très chaleureusement ces deux collègues qui m'ont accompagné tout au long de ce doctorat.

The members of the Morgenthaler lab made my PhD enjoyable. I really appreciated Wen's cheerfulness when we worked together. Puis, je suis très heureux d'avoir rencontré Djalel. Nos conversations, que ce soit en matrices aléatoires, tests multiples, politique ou religion, ont toujours été passionnantes et j'ai beaucoup apprécié ces moments. Finally I would like to thank my office mate the last year of my PhD, Helen, who cured my loneliness in the office. She was so kind and funny that I was "forced" to pratice my English so I could enjoy talking with her. I am also professionally grateful to her for carefully correcting my English in this thesis (including these few lines...).

Durant mes études et mon début de doctorat, j'ai eu la chance de côtoyer un voisin et ami, Kader. Grâce à son coté chaleureux et méditerranéen, je me suis enfin senti bien, loin du Valais. Je remercie aussi mon autre voisin, André-Maurice. J'ai particulièrement apprécié sa bienveillance et sa sagesse lors de nos repas de midi en dehors de l'EPFL.

Toutes ces années à l'EPFL ont été rendues meilleures grâce à mes amis valaisans sans qui mon Bachelor et mon Master auraient été beaucoup plus ennuyeux. Je remercie en particulier Jean-Marc, Xavier, Elina, Jérome, Gäetan, Aurélie et Sebastien qui n'ont que très rarement reculé devant une bière à Sat.

Je suis aussi reconnaissant envers mon parrain de confirmation, Jo, qui, en plus de me soutenir dans mes choix académiques, m'a proposé des discussions très intéressantes et stimulantes dans d'autres domaines que les mathématiques. Merci aussi à ces points magiques de guérison, Xiaguan E7 et Fengchi VB20.

Durant ce long parcours, mes weekends à Genève ont été illuminés par ma "belle" famille. Merci à Maria et Eusebio pour leur accueil toujours aussi chaleureux et leur joie de vivre. Je remercie aussi les inépuisables, drôles et motivants Yannick et Victoria qui s'avèrent être d'incroyables organisateurs spécialistes en communication et comptabilité ! Grâce à eux, j'ai passé de très bons moments à Genève, cette capitale romande que j'ai découverte sous un très beau visage.

Tout au long de mes études, j'ai toujours été soutenu par ma famille dans les moments les plus critiques. Sans eux, je n'aurais très certainement pas entrepris un voyage académique aussi long. Un grand merci à mes incroyables parents, Sylvaine et Philippe, pour cela. Ils m'ont entre autres empêché de faire l'erreur d'abandonner à plusieurs reprises : après deux semaines de collège, deux semaines de première année à l'EPFL, deux semaines de deuxième année,... Merci aussi à Délia et Michel (à qui je souhaite de bons vœux de mariage) pour, entre autres, leur capacité inépuisable à organiser des activités les weekends, les anniversaires, les repas de fêtes,... Finalement merci à mes grands parents, Edouard et Marie-Louise, pour leurs filets d'agneau et riz du samedi, mais aussi leurs encouragements lors de mes études.

Enfin durant ces études, j'ai accidentellement rencontré un petit soleil qui a illuminé ma vie. Ce soleil a écouté l'intégralité de ma thèse : chaque beau résultat, pour ensuite se réjouir avec moi, mais aussi, chaque difficulté et passage stressant, pour ensuite les évaporer par sa plussoyance. Merci à Sabrina pour ses encouragements, sa patience, sa joie et sa motivation à entreprendre des activités telles que la salsa, des spectacles ou des voyages qui ont dissous tout moment d'ombre qui aurait pu surgir durant la thèse.

Lausanne, 2019-05-23 R. M.

Abstract

During the last twenty years, Random matrix theory (RMT) has produced numerous results that allow a better understanding of large random matrices. These advances have enabled interesting applications in the domain of communication. Although this theory can contribute to many other domains such as brain imaging or genetic research, its has been rarely applied.
The main barrier to the adoption of RMT may be the lack of concrete statistical results from probabilistic Random matrix theory. Indeed, direct generalisation of classical multivariate theory to high dimensional assumptions is often difficult and the proposed procedures often assume strong hypotheses on the data matrix such as normality or overly restrictive independence conditions on the data.

This thesis proposes a statistical procedure for testing the equality of two independent estimated covariance matrices when the number of potentially dependent data vectors is large and proportional to the size of the vectors corresponding to the number of observed variables.
Although the existing theory builds a very good intuition of the behaviour of these matrices, it does not provide enough results to build a satisfactory test for both the power and the robustness.
Hence, inspired by spike models, we define the residual spikes and prove many theorems describing the behaviour of many statistics using eigenvectors and eigenvalues in very general cases. For example in the two central theorems of this thesis, the Invariant Angle Theorem and the Invariant Dot Product Theorem.

Using numerous generalisations of the theory, this thesis finally proposes a description of the behaviour of a statistic under a null hypothesis. This statistic allows the user to test the equality of two populations, but also other null hypotheses such as the independence of two sets of variables.
Finally, the robustness of the procedure is demonstrated for different classes of models and criteria for evaluating robustness are proposed to the reader.

Therefore, the major contribution of this thesis is to propose a methodology both easy to apply and having good properties. Secondly, a large number of theoretical results are demonstrated and could be easily used to build other applications.

Keywords: Dependent data, eigenvalue, eigenvector, equality test of two covariance matrices, high dimension, random matrix theory, residual spike, spike model.

Résumé

Ces vingt dernières années, la théorie des matrices aléatoires (RMT) a produit de nombreux résultats permettant une meilleure compréhension de grandes matrices aléatoires. Ces avancées ont permis des applications intéressantes dans le domaine de la communication. Bien que cette théorie puisse contribuer à beaucoup d'autres domaines comme en imagerie cérébrale ou dans la recherche en génétique, son application est restée très modérée.
La raison principale de cette difficulté à s'implanter pourrait être le manque de résultats statistiques concrets qui s'inspirent de la théorie probabiliste des matrices aléatoires. En effet la généralisation directe en haute dimension des résultats de statistiques multivariées classiques s'avère souvent difficile et les procédures créées supposent de fortes hypothèses sur la matrices de données comme la normalité des entrées ou des indépendances exigeantes.

Cette thèse propose une procédure statistique pour tester l'égalité entre deux matrices de covariances estimées indépendantes quand les données sont de dimension large et que le nombre de données, potentiellement dépendantes, est proportionnel au nombre de variables observées.
Bien que la théorie crée une très bonne intuition du comportement de ces matrices, elle ne procure cependant pas suffisamment de résultats pour construire un test satisfaisant aussi bien au niveau de sa puissance que de sa robustesse.
De ce fait, en s'inspirant des modèles spikes, nous définissons les spikes résiduels et démontrons plusieurs théorèmes étudiant différentes statistiques construites à partir de vecteurs propres et de valeurs propres dans des cas très généraux. Citons par exemple les deux théorèmes centraux de cette thèse, le Théorème de l'angle invariant et le Théorème du produit scalaire invariant.

Basé sur de nombreuses généralisations de la théorie, cette thèse propose finalement une description du comportement d'une statistique sous une hypothèse nulle. L'utilisateur pourra alors tester l'égalité entre deux populations, mais aussi d'autres hypothèses telles que l'indépendance entre deux jeux de variables.
Finalement, la robustesse de la méthode est démontrée dans différentes classes de modèles et des critères évaluant la robustesse sont proposés à l'utilisateur.

Ainsi cette thèse propose principalement une méthode facile à utiliser et ayant de bonnes propriétés. Dans un second temps, un grand nombre de résultats théoriques sont démontrés et leur utilisation pourrait facilement être détournée pour construire d'autres applications.

Mots clés: Données dépendantes, haute dimension,modèle spike, test d'égalité sur des matrices de covariances, théorie des matrices aléatoires, spike résiduel, valeur propre, vecteur propre.

Contents

Chapter 1

Introduction to Random matrix theory

In this chapter we introduce the main results of Random matrix theory (RMT) and then focus on the problem of comparing two random matrices in order to decide whether they really differ. Our results will be useful when comparing two large estimated covariance matrices.
As we will see, RMT provides many strong theoretical results but has only a few applications to date. Although this thesis does not use all the results of this chapter, they provide important context for our subsequent results.

1.1 Notation

Let $X_1, ..., X_n$ be centred i.i.d random vectors of size m and covariance Σ_m. We study the estimated covariance matrix of these vectors

$$\hat{\Sigma} = \hat{\Sigma}_{(m,n)} = \frac{1}{n}\sum_{i=1}^{n} X_i X_i^t = \frac{1}{n}\mathbf{X}\mathbf{X}^t,$$

where $\mathbf{X} = (X_1, ..., X_n)$ is an $m \times n$ random matrix with columns X_i.
RMT concerns this random matrix; or more precisely its eigenvalues and eigenvectors when both n and m tend to infinity. When m is finite and n tends to infinity the behaviour of the random matrix is well known and presented in the books of Mardia, Kent, and Bibby [1979], Muirhead [2005] and Anderson [2003] (or its original version Anderson [1958]).
When both m and n tend to infinity in such a way that $m/n \to c$, the behaviour is more complex, but many results of interest are known. Anderson, Guionnet, and Zeitouni [2009], Tao [2012] and more recently Bose [2018] contain comprehensive introductions to RMT and Bai and Silverstein [2010] covers the case of empirical (estimated) covariance matrices. Before introducing the results we present some concepts and notations.

Definition 1.1.1.
Let $A_m \in \mathbb{R}^{m\times m}$ be a matrix with eigenvalues $\lambda_1, \lambda_2, ..., \lambda_m \in [\tilde{a}_m, \tilde{b}_m] \subset \mathbb{R}$.

1. The cumulative spectrum of A_m, F_{A_m}, is defined on $\mathbb{R}$ as

$$F_{A_m}(\lambda) = \frac{1}{m}\sum_{i=1}^{m} \mathbf{1}_{\{\lambda_i \leqslant \lambda\}}.$$

2. Suppose that $A_1, A_2, A_3, ...$ is a sequence of matrices with cumulative spectrum, $F_{A_1}, F_{A_2}, F_{A_3}, ...$ If the sequence of the cumulative spectra converges weakly in probability to a bounded cumulative function, $F_A(\lambda)$, on $[\tilde{a}, \tilde{b}] \subset \mathbb{R}$,

$$F_{A_m} \to F_A,$$

then we call F_A the asymptotic cumulative spectrum.

3. If F_A is differentiable, then we define the spectral asymptotic density
$$f_A(\lambda) = F'_A(\lambda).$$

Notation 1.1.1.

1. The ratio of m and n is denoted $c_m = m/n$ and tends to $c \in]0, +\infty[$.
2. We will often use Σ instead of Σ_m if it is clear from the context what is meant. Note that $\Sigma_m \in \mathbb{R}^{m\times m}$. We define $F_{\Sigma_m}(\lambda)$ to be the spectrum of Σ_m and $F_\Sigma(\lambda)$ to be the limit of $F_{\Sigma_m}(\lambda)$.
3. The empirical covariance matrix is then $\hat{\Sigma}_{(m,n)} = \hat{\Sigma}_m$ and its spectrum is $F_{\hat{\Sigma}_{(m,n)}}(x)$ which tends to $F_{\hat{\Sigma}}(x)$. Thus, $F'_{\hat{\Sigma}}(x) = f_{\hat{\Sigma}}(x)$ is the density corresponding to the limit of the empirical spectral distribution.
4. For $i = 1, 2, ..., m$, $\lambda_i(A)$ denotes the i^{th} largest eigenvalue of the square matrix $A \in \mathbb{R}^{m\times m}$.
5. We will denote by $\lambda_{\Sigma_m,i}$ and $u_{\Sigma_m,i}$ $(i = 1, 2, ..., m)$ the eigenvalues and eigenvectors of non-random Σ_m. When no confusion is possible we use the simpler notation, λ_i and u_i.
6. The empirical eigenvalues and eigenvectors of $\hat{\Sigma}$ are $\hat{\lambda}_{\hat{\Sigma},i}$ and $\hat{u}_{\hat{\Sigma},i}$ and depend of course on n and m. When no confusion is possible we use the simpler notation, $\hat{\lambda}_i$ and $\hat{u}_i$.
7. If a random matrix $\mathbf{X} \in \mathbb{R}^{m\times n}$ is not symmetric, $\hat{\lambda}_{\mathbf{X},i}$ and $\hat{u}_{\mathbf{X},i}$ denote the i^{th} eigenvalues and eigenvectors of $\frac{1}{n}\mathbf{X}\mathbf{X}^t$.
8. We use σ^2 for $\frac{\text{Trace}(\Sigma)}{m}$. We will often standardise Σ by σ^2 and still call the resulting matrix Σ.
9. We denote by $\Sigma^{1/2}$ the positive definite matrix in $\mathbb{R}^{m\times m}$ such that $\Sigma^{1/2}\Sigma^{1/2} = \Sigma \in \mathbb{R}^{m\times m}$.

1.2 Independent case

In this section we assume $\mathbf{X} = (X_1, X_2, ..., X_n)$ is a real random matrix of size $m \times n$ such that for $i = 1, 2, ..., m$ and $j = 1, 2, ..., n$,
$$\begin{aligned}
&\mathbf{X}_{i,j} \text{ are independent,}\\
&\mathrm{E}\left[\mathbf{X}_{i,j}\right] = 0,\\
&\mathrm{Var}\left(\mathbf{X}_{i,j}\right) = \sigma^2.
\end{aligned}$$

Many interesting results are known for this scenario, in particular:

1. the behaviour of the asymptotic empirical spectrum,
2. the asymptotic distribution of the largest empirical eigenvalue,
3. the behaviour of the empirical eigenvectors.

1. Spectrum: The most famous result of RMT gives the limiting density of the spectrum of $\hat{\Sigma}$.

Result 1.2.1. *(Marcenko-Pastur)*
The spectral density of $\hat{\Sigma}_{(m,n)} = \frac{1}{n}\mathbf{X}\mathbf{X}^t$ converges to a ***Marcenko-Pastur*** *form,*
$$f_{MP}(x) = \left(1 - \frac{1}{c}\right)^+ \delta_0(x) + \frac{\sqrt{(b-x)(x-a)}}{2\pi\sigma^2 xc}\mathbf{1}_{[a,b]}(x),$$
where $\lim\limits_{m,n\to\infty} \frac{m}{n} = c \in]0, \infty[$, $a = \sigma^2\left((1-\sqrt{c})^+\right)^2$, $b = \sigma^2(1+\sqrt{c})^2$, *$\delta_0$ is the Dirac function at 0 and*
$$(x)^+ = \begin{cases} x \text{ if } x \geqslant 0,\\ 0 \text{ if } x < 0.\end{cases}$$

The Dirac function in 0 is necessary in the degenerate case when $m > n$ or $c > 1$.
In Figure 1.1 we see an example of Marcenko-Pastur density. The empirical spectrum is asymptotically a bulk on the interval $[a, b]$ drawn in red in the figure.
This result was proven by Marchenko and Pastur [1967].

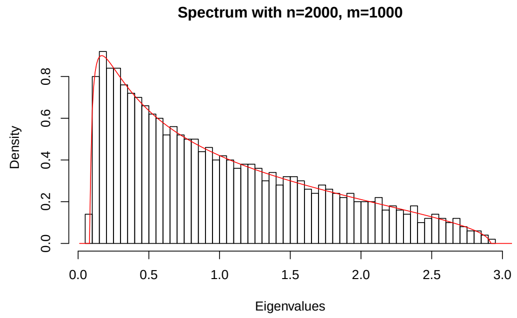


Figure 1.1 – Density of Marcenko-Pastur compared with the histogram of the eigenvalues of simulated random matrix when $\sigma^2 = 1$.

2. Largest empirical eigenvalue: The second result focuses on the largest eigenvalue of $\hat{\Sigma}_{(m,n)}$ and provides its asymptotic distribution.

Result 1.2.2. *(Tracy-Widom)*
If all entries $\mathbf{X}_{i,j}$ are such that enough moments exist and their support is always strictly more than two points, then the largest eigenvalue of the random matrix $\hat{\Sigma}_{(m,n)} = \frac{1}{n}\mathbf{X}\mathbf{X}^t$ asymptotically follows a ***Tracy-Widom*** (TW) *distribution,*

$$\frac{n^{2/3}}{\Theta_m}\left(\lambda_{\max}\left(\hat{\Sigma}_{(m,n)}\right) - \sigma^2(1+\sqrt{c})^2\right) \xrightarrow[n,m\to\infty]{D} \mathrm{TW},$$

where $\lim\limits_{m,n\to\infty} \frac{m}{n} = c \in \,]0,\infty[$ *and* $\Theta_m = \sigma^2(1+\sqrt{c})\left(\frac{1}{\sqrt{c}}+1\right)^{1/3}$.

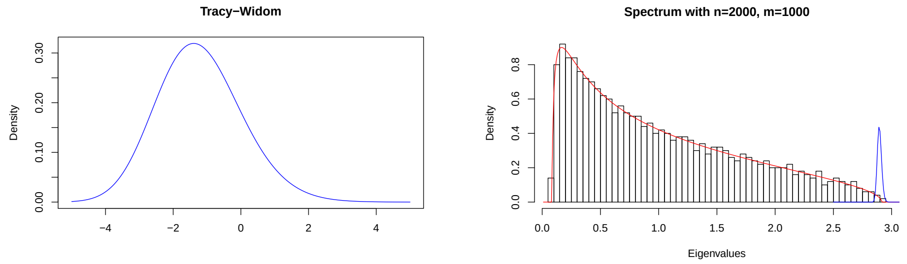
Figure 1.2 shows the link between Marcenko-Pastur and Tracy-Widom.
This result was shown by Pillai and Yin [2014]. The additional moment condition is discussed by Wang [2012] following Theorem 2.2. Moreover, this last paper extends the result to the smallest eigenvalue when $c < 1$. An interesting extension of this result to correlation matrices is provided by Pillai and Yin [2012]. The original result was introduced by Johnstone [2001] for Gaussian entries.

3. Eigenvectors: The last result of this section helps us to understand the behaviour of the empirical eigenvectors in the independent case.

Result 1.2.3. *(Eigenvector)*
The eigenvectors $\hat{u}_1, \hat{u}_2, ..., \hat{u}_m$ of $\hat{\Sigma}_{(m,n)} = \frac{1}{n}\mathbf{X}\mathbf{X}^t$ are uniformly distributed.
Assuming $\lim\limits_{m,n\to\infty} \frac{m}{n} = c \in \,]0,\infty[$, *then for a deterministic unit vector $w = (w_1, w_2, ..., w_m)$ we have*

$$\begin{aligned}\forall s, \quad \langle w, \hat{u}_s\rangle^2 &= \frac{1}{m}\sum_{i=1}^{m} w_i^2 \Theta(w,s) \quad \textit{and} \quad \Theta(w,s) \xrightarrow[n,m\to\infty]{D} \chi_1^2, \\ \langle w, \hat{u}_s\rangle^2 &\xrightarrow[n,m\to\infty]{P} 0.\end{aligned}$$

Moreover, if the columns of $\mathbf{X}$ are spherical, then $(\hat{u}_1, \hat{u}_2, ..., \hat{u}_m)$ follows the unit invariant Haar distribution.



a b

Figure 1.2 – a: Density of Tracy-Widom. b: Density of Marcenko-Pastur with $\sigma^2 = 1$ and non-scaled density of the maximal eigenvalue.

These results mean that an estimated eigenvector $\hat{u}_i$ carries asymptotically no information. This conclusion is not surprising because the eigenvectors u_i of $\Sigma_m = \sigma^2 \mathrm{I}_m$ are not uniquely determined. These results and many others are proved in Bloemendal, Knowles, Yau, and Yin [2016]. The result for spherical entries is proved in Muirhead [2005].

1.3 Prerequisite tools

The assumption $\Sigma_m = \mathrm{I}_m$ can be generalized to more complicated structures and is investigated in the next section. We next define some useful mathematical tools.
First we define the Stieltjes transform and its inverse.

Definition 1.3.1. (Stieltjes transform)
Assume F is a measure supported on $[a, b] \in \mathbb{R}$. The **Stieltjes transform** of the measure F evaluated in $z \in \mathbb{C} \setminus [a, b]$ is defined as

$$m_F(z) = \int \frac{1}{z - x} dF(x).$$

If $f = F'$ exists,

$$m_f(z) = \int \frac{1}{z - x} f(x) dx.$$

If f is continuous we have the inverse formula

$$f(x) = \lim_{y \to 0+} \frac{m_f(x - iy) - m_f(x + iy)}{2i\pi},$$

where $z = x + iy$ and $y > 0$.

Related to this is the T-transform.

Definition 1.3.2. (T-tranform)
Assume F is a measure supported on $[a, b] \in \mathbb{R}$. The **T-transform** of the measure F evaluated in $z \in \mathbb{C} \setminus [a, b]$ is defined as

$$T_F(z) = \int \frac{x}{z - x} dF(x).$$

If $f = F'$ exists,

$$T_f(z) = \int \frac{x}{z - x} f(x) dx.$$

We can easily see that

$$T_f(z) = z m_f(z) - 1.$$

1.4 Deformed Marcenko-Pastur distribution

In this section we suppose that F_{Σ_m} and $F_\Sigma = \lim_{m\to\infty} F_{\Sigma_m}$ (weak convergence) are known. Moreover, the support of F_{Σ_m} is bounded.
Let $\mathbf{X} = (X_1, X_2, ..., X_n)$ and $\tilde{\mathbf{X}} = \left(\tilde{X}_1, \tilde{X}_2, ..., \tilde{X}_n\right)$ be $m \times n$ matrices such that for $i = 1, 2, ..., m$ and $j = 1, 2, ..., n$,

$$\mathbf{X} = \Sigma_m^{1/2}\tilde{\mathbf{X}},$$

and

Σ_m is independent of $\tilde{\mathbf{X}}$ with cumulative spectrum F_{Σ_m},
$\tilde{\mathbf{X}}_{i,j}$ are independent,
$\mathrm{E}\left[\tilde{\mathbf{X}}_{i,j}\right] = 0$,
$\mathrm{Var}\left[\tilde{\mathbf{X}}_{i,j}\right] = 1$.

Again, many results of interest have been demonstrated by RMT concerning the spectrum of $\mathbf{X}$:

1. the asymptotic empirical spectral distribution, $F_{\hat{\Sigma}}$,
2. the behaviour of the maximal eigenvalue.

Properties of the eigenvectors are more difficult to obtain in this case.

1. Spectrum: The most famous result of RMT can be generalized and expresses the spectral measure as a function of its Stieltjes transform:

Result 1.4.1. *(Deformed Marcenko-Pastur)*
Assume $F_{\hat{\Sigma}_{(m,n)}}$ is the spectral distribution of $\hat{\Sigma}_{(m,n)} = \frac{1}{n}\mathbf{X}\mathbf{X}^t = \frac{1}{n}\Sigma_m^{1/2}\tilde{\mathbf{X}}\tilde{\mathbf{X}}^t\Sigma_m^{1/2}$, where $\tilde{\mathbf{X}}$ is defined above.
*If $\lim_{m,n\to\infty} \frac{m}{n} = c \in]0, \infty[$ and the spectrum of Σ_m is F_{Σ_m}, which tends to a limit F_Σ that satisfies the Carleman sufficiency condition, then $F_{\hat{\Sigma}_{(m,n)}}$ tends to the **deformed Marcenko-Pastur** distribution $F_{\hat{\Sigma}}$ whose Stieltjes transform satisfies:*

$$m_{F_{\hat{\Sigma}}}(z) = \frac{1}{-z + c\int \frac{t}{t m_{F_{\hat{\Sigma}}}(z)+1} dF_\Sigma(t)}.$$

If $F_{\hat{\Sigma}}$ is smooth enough, the inverse transform formula gives $f_{\hat{\Sigma}}$ as a function of F_Σ.

The inverse formula leads to the deformed Marcenko-Pastur distribution. The result was first shown by Marchenko and Pastur [1967] for diagonal Σ and then extended by Silverstein [1995] to non-diagonal Σ.

Example 1.4.1.1.

1. If $f_\Sigma(\lambda) = \delta_1(\lambda)$, i.e., the Dirac delta function at 1, then $m_{f_{\hat{\Sigma}}}(z)$ is the Stieltjes transform of the Marcenko-Pastur distribution. Therefore, the inverse formula leads to the M-P density.
2. If f_Σ is itself an M-P density, that is, $f_\Sigma(\lambda) = f_{MP}(\lambda)$, then the Stieltjes transform has to be computed numerically and the inverse formula leads to a deformed Marcenko-Pastur distribution as shown in Figure 1.3.

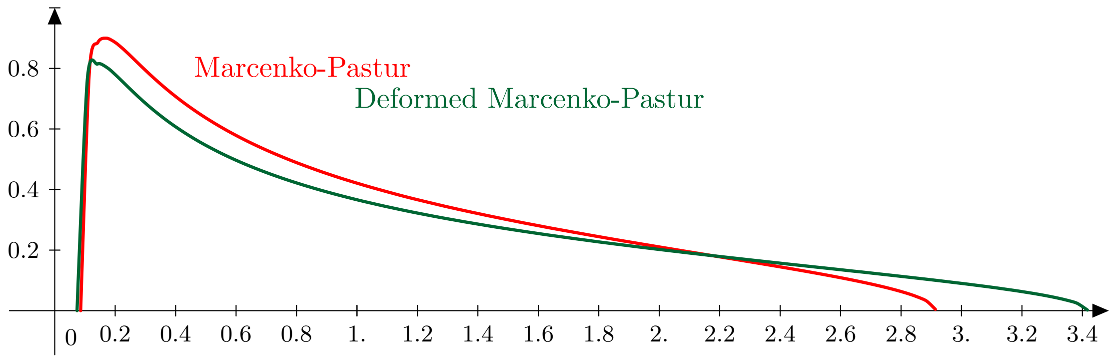


Figure 1.3 – Plot of a deformed Marcenko-Pastur density.

2. Largest empirical eigenvalue: The extension of the largest eigenvalue distribution to more general Σ leads to the following result.

Result 1.4.2. *(Tracy-Widom for Perturbed Marcenko-Pastur)*
Suppose $\hat{\Sigma}_{(m,n)}$ is defined as in Result 1.4.1 and $\lim_{m,n\to\infty} \frac{m}{n} = c \in]0,\infty[$. Under the additional conditions (C1) on $F_{\Sigma_m} \to F_\Sigma$ and (C2) defined below, the largest eigenvalue of $\hat{\Sigma}_{(m,n)}$ is such that

$$\frac{n^{2/3}}{\sigma}\left(\lambda_{\max}\left(\hat{\Sigma}_{(m,n)}\right) - \mu\right) \underset{n,m\to\infty}{\xrightarrow{D}} \mathrm{TW},$$

where TW *is a Tracy-Widom distribution and*

$$\begin{aligned}
\mu &= \frac{1}{k}\left(1 + c\int \frac{\lambda k}{1-\lambda k} dF_{\Sigma_m}(\lambda)\right), \\
\sigma^3 &= \frac{1}{k^3}\left(1 + c\int \left(\frac{\lambda k}{1-\lambda k}\right)^3 dF_{\Sigma_m}(\lambda)\right), \\
k = k_m &\in [0, 1/\lambda_{\max}) : \int \left(\frac{\lambda k}{1-\lambda k}\right)^2 dF_{\Sigma_m}(\lambda) = \frac{1}{c}.
\end{aligned}$$

Condition (C1) on F_{Σ_m} can be written as

$$\lim_{n,m\to\infty} \sup(\lambda_{\max,m} k_m) < 1,$$

where $\lambda_{\max,m} = \sup\{\lambda | F_{\Sigma_m}(\lambda) > 0\}$.
The condition (C2) demands that the entries of the matrix $\tilde{\mathbf{X}}$ have uniformly subexponential tails.

Figure 1.4 is a schematic drawing of the situation where $f_\Sigma(\lambda) = \delta_1(\lambda)$ or $f_\Sigma(\lambda) = f_{MP}(\lambda)$.
This powerful result, first proven for normal entries by El Karoui [2007], gives a condition which allows F_{Σ_m} to be replaced by F_Σ in the formula. The condition assumes convergence of the edges of the support of F_{Σ_m} and F_Σ.
The result was later extended to entries assuming a uniform subexponential distribution by Bao, Pan, and Zhou [2015].

The extension of the independent theory to the deformed theory could be used when the matrix Σ_m is clearly not I_m and standardisation of the data would be too complex.

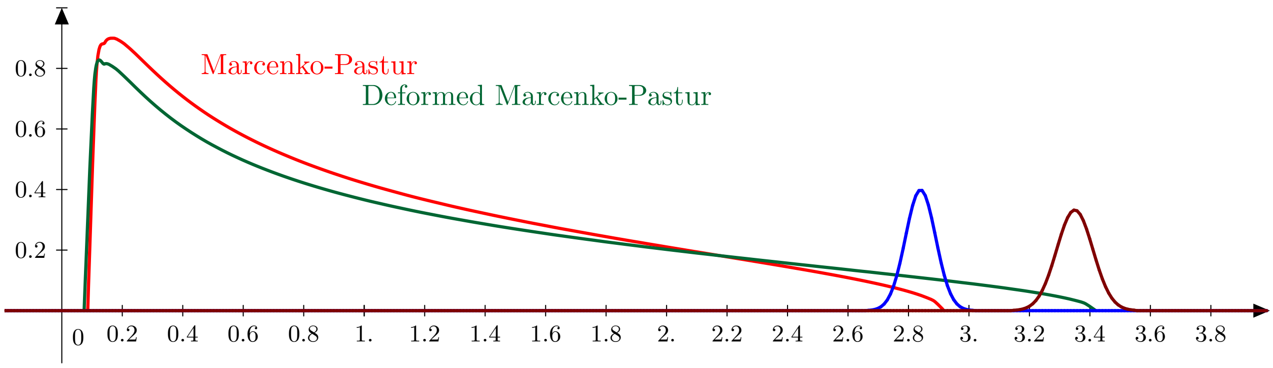


Figure 1.4 – Schema of a deformed Marcenko-Pastur density and Tracy-Widom density (not scaled).

1.5 Spiked model under the Marcenko-Pastur spectrum

RMT provides powerful tools based on the spiked model introduced by Baik and Silverstein [2005]. This model is based on a finite perturbation and offers tools to estimate the true covariance matrix. First, we will introduce the model. Then we will present some useful results for estimating the unknown parameters such as the asymptotic convergence of the eigenvalues or the asymptotic distribution of the eigenvalues. Finally, two results will highlight the difficulty of estimating the eigenvectors.

In this section we suppose that Σ_m is a finite perturbation

$$\Sigma_m = \mathrm{I}_m + \sum_{i=1}^{k} (\theta_i - 1) u_i u_i^t,$$

where $u_1, ..., u_k$ are unit orthogonal vectors and $\theta_i > 0$. To simplify somewhat we suppose $\theta_i \neq \theta_j$ $\forall i \neq j$. The results can be generalized to

$$\Sigma_m = \sigma^2 \mathrm{I}_m + \sum_{i=1}^{k} (\theta_i - \sigma^2) u_i u_i^t = \sigma^2 \left(\mathrm{I}_m + \sum_{i=1}^{k} \left(\theta_i/\sigma^2 - 1\right) u_i u_i^t \right)$$

by looking only at $\mathrm{I}_m + \sum_{i=1}^{k} \left(\theta_i/\sigma^2 - 1\right) u_i u_i^t$ with an accompanying estimate of σ^2.

In this section we present some results about the spectrum, the largest eigenvalue and the eigenvectors.

1. Spectrum: The first result shows the robustness of the spectrum against finite perturbations.

Result 1.5.1.
Suppose $\tilde{\Sigma}$ is a random matrix of size $m \times n$ with bounded spectral distribution $F_{\tilde{\Sigma}}$. We define $\hat{\Sigma} = \Sigma_m^{1/2} \tilde{\Sigma} \Sigma_m^{1/2}$ with spectrum $F_{\hat{\Sigma}}$. If Σ_m is a finite perturbation then

$$\int g(\lambda)\, dF_{\tilde{\Sigma}}(\lambda) = \int g(\lambda)\, dF_{\hat{\Sigma}}(\lambda) + O_p\left(\frac{1}{m}\right),$$

for g bounded on $S = \left\{x | F_{\hat{\Sigma}} > 0 \text{ or } F_{\tilde{\Sigma}} > 0\right\}$.

This robustness is a direct consequence of Cauchy's interlacing law.

2. Largest empirical eigenvalue: Two important results about the limit and limit distribution are contained in the literature.
The first result was shown by Baik and Silverstein [2005] and provides the asymptotic limit of the largest eigenvalues.

Result 1.5.2. *(Eigenvalue detection)*
Suppose $\tilde{\mathbf{X}}$ is a random matrix of dimension $m \times n$ with independent entries such that for $i = 1, 2, .., m$ and $j = 1, 2, ..., n$,

$$\mathrm{E}\left[\tilde{\mathbf{X}}_{i,j}\right] = 0,\ \mathrm{Var}\left(\tilde{\mathbf{X}}_{i,j}\right) = 1,\ \mathrm{E}\left[\left|\tilde{\mathbf{X}}_{i,j}\right|^4\right] < \infty.$$

Then we define $\lim_{m,n\to\infty} \frac{m}{n} = c \in]0, \infty[$ and $\hat{\Sigma}_{(m,n)} = \frac{1}{n}\mathbf{X}\mathbf{X}^t$ where $\mathbf{X} = \Sigma_m^{1/2}\tilde{\mathbf{X}}$.
If $\theta_i > 1 + \sqrt{c}$ then there exists $\lambda_{\theta_i}\left(\hat{\Sigma}_{(m,n)}\right)$ such that

$$\lambda_{\theta_i}\left(\hat{\Sigma}_{(m,n)}\right) \xrightarrow[n,m\to\infty]{a.s} \theta_i\left(1 + \frac{c}{\theta_i - 1}\right).$$

If $\forall i,\ \theta_i < 1 + \sqrt{c}$ then

$$\lambda_{\max}\left(\hat{\Sigma}_{(m,n)}\right) \xrightarrow[n,m\to\infty]{a.s} \left(1 + \sqrt{c}\right)^2.$$

If θ_i is too small we lose the empirical eigenvalue in the Marcenko-Pastur bulk. However, if θ_i is large enough, the largest estimated eigenvalues are biased estimators of the true eigenvalues and can be debiased to obtain consistent estimators

$$\hat{\theta}_i \left| \lambda_{\theta_i}\left(\hat{\Sigma}_{(m,n)}\right) = \hat{\theta}_i\left(1 + \frac{c}{\hat{\theta}_i - 1}\right)\right..$$

Assuming Gaussian entries, the previous result can be extended to provide the distribution.

Result 1.5.3. *(Eigenvalue distribution)*
Suppose $c \in]0, \infty[$, $\lim_{m,n\to\infty} \frac{m}{n} - c = o\left(1/\sqrt{m}\right)$ and $\mathbf{X} = \Sigma_m^{1/2}\tilde{\mathbf{X}}$, where $\tilde{\mathbf{X}}_{i,j} \overset{i.i.d.}{\sim} \mathbf{N}(0, 1)$.
If $\theta_i > 1 + \sqrt{c}$ then using the same notation as Result 1.5.2

$$\sqrt{n}\left(\lambda_{\theta_i}\left(\hat{\Sigma}_{(m,n)}\right) - \nu_i\right) \xrightarrow[n,m\to\infty]{D} \mathrm{Normal}(0, \gamma_i),$$

where $\nu_i = \theta_i\left(1 + \frac{c}{\theta_i - 1}\right)$ and $\gamma_i = 2\theta_i^2\left(1 - \frac{c}{(\theta_i - 1)^2}\right)$.
If $\theta_i < 1 + \sqrt{c}$ then we obtain a Tracy-Widom distribution as shown in Result 1.4.2.

The schema in the Figure 1.5 shows the behaviour of the largest eigenvalue.

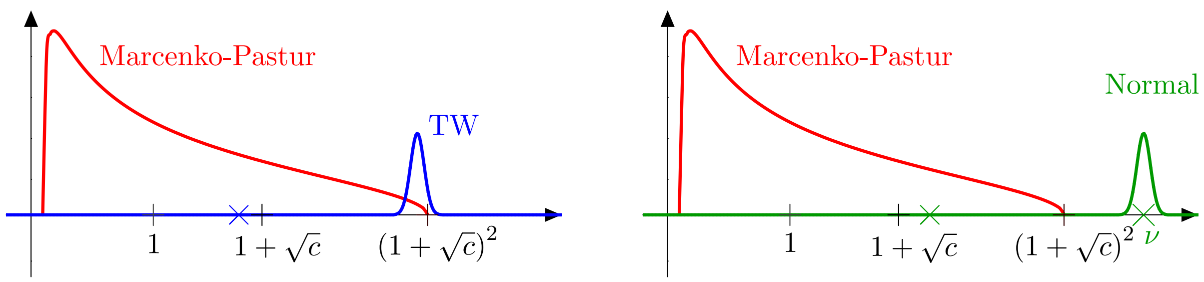


Figure 1.5 – Schema of the impact of the transition point $1 + \sqrt{c}$ on the distribution of the empirical eigenvalue. The cross represents the true eigenvalue and the green curve represents the distribution of the estimated eigenvalues (Spike). The blue curve represents the distribution of Tracy-Widom that is not impacted by small perturbations.

This result was extended by Paul [2007] to understand the distribution of the estimated eigenvalues when the entries $\tilde{X}_{i,j} \overset{i.i.d.}{\sim} \mathbf{N}(0,1)$. In the complex non degenerate Gaussian case the same result is provided by Baik, Ben Arous, and Péché [2005]. They also showed the second part of the result when the θ_i are small. Later this extension of the Tracy-Widom distribution was shown to be a corollary of a more general result by Bao, Pan, and Zhou [2015].
The previous papers do not prove the result for the degenerate case when $c > 1$. This more general case has been covered in Benaych-Georges, Guionnet, and Maida [2010] who also extend it to equal eigenvalues.
Finally note that the investigation of perturbation is older than RMT. For example, this topic is treated by Kato [1995] for small perturbations of covariance operators.

3. Eigenvectors: Some results about the asymptotic behaviour of eigenvectors exist. We first present the almost sure convergence of the dot product and then some results about the distribution of the eigenvectors.

Result 1.5.4. *(Eigenvector bias)*
Suppose $\lim_{m,n\to\infty} \frac{m}{n} = c \in]0,\infty[$ *and* $\tilde{\mathbf{X}}$ *a random matrix of size* $m \times n$ *with independent entries such that for* $i = 1, 2, ..., m$ *and* $j = 1, 2, ..., m$,

$$\mathrm{E}\left[\tilde{\mathbf{X}}_{i,j}\right] = 0,\ \mathrm{Var}\left(\tilde{\mathbf{X}}_{i,j}\right) = 1.$$

Moreover, for $p \in \mathbb{N}$, *there is a* C_p *such that for all entries of* $\tilde{\mathbf{X}}$

$$\mathrm{E}\left[\left|\tilde{\mathbf{X}}_{i,j}\right|^p\right] < C_p.$$

Then we define $\mathbf{X} = \Sigma_m^{1/2}\tilde{\mathbf{X}}$ *and* $\hat{\Sigma}_{(m,n)} = \frac{1}{n}\mathbf{X}\mathbf{X}^t$.
If $\theta_i > 1 + \sqrt{c}$ *then*

$$\left|\left\langle u_i, \hat{u}_{\hat{\lambda}_{\theta_i}} \right\rangle\right| \xrightarrow[n,m\to\infty]{a.s} \sqrt{\frac{1 - \frac{c}{(\theta_i - 1)^2}}{1 + \frac{c}{\theta_i - 1}}},$$

where $\hat{u}_{\hat{\lambda}_{\theta_i}}$ *is the empirical eigenvector of* $\hat{\Sigma}_{(m,n)}$ *corresponding to the empirical eigenvalue* $\lambda_{\theta_i}\left(\hat{\Sigma}_{(m,n)}\right)$ *described in the previous result 1.5.3.*
If $\theta_i < 1 + \sqrt{c}$ *then*

$$\forall j,\ \langle u_i, \hat{u}_j \rangle \xrightarrow[n,m\to\infty]{a.s} 0,$$

where $\hat{u}_j$ *is the* j^{th} *empirical eigenvector of* $\hat{\Sigma}_{(m,n)}$. *More precisely for all* j *such that* $\theta_j > 1 + \sqrt{c}$,

$$m\left\langle u_i, \hat{u}_{\hat{\lambda}_{\theta_j}} \right\rangle = \frac{c\theta_j}{\left(\theta_i - 1 - \sqrt{c}\right)^2}\ \Theta(u_i, j)\ \text{ and } \Theta(u_i, j) \xrightarrow[n,m\to\infty]{a.s} \chi_1^2.$$

Remark 1.5.4.1.
If c tends to 0 then the empirical eigenvector is a consistent estimator of the true eigenvector. However, if c is different from 0, an asymptotic angle between the true and the estimated eigenvector appears as shown in Figure 1.6. Moreover, if θ_i is too small, the eigenvector of the perturbation cannot be identified, whereas if θ_i is large the angle tends again to 0.

A weaker result was proven for Gaussian entries by Paul [2007] and as a particular case of Benaych-Georges and Rao [2009]. The convergence results based on these conditions are provided as a particular case in the paper of Bloemendal, Knowles, Yau, and Yin [2016].
Assuming Gaussian entries the previous result was extended by Paul [2007] to provide some interesting distributions.

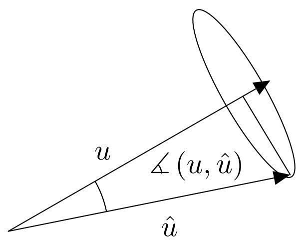


Figure 1.6 – Angle between the true eigenvector u and the estimated eigenvector $\hat{u}$.

Result 1.5.5. *(Eigenvector moment)*
Suppose $c \in]0, \infty[$, $\lim_{m,n\to\infty} \frac{m}{n} - c = o\left(\frac{1}{\sqrt{m}}\right)$, $\mathbf{X} = \Sigma_m^{1/2}\tilde{\mathbf{X}}$, *where* $\tilde{\mathbf{X}}_{i,j} \overset{i.i.d.}{\sim} \mathbf{N}(0,1)$ *and* $\hat{\Sigma}_{(m,n)} = \frac{1}{n}\mathbf{X}\mathbf{X}^t$.
Without loss of generality assume $(u_1, ..., u_k) = (e_1, ..., e_k)$.
If $\theta_1 > 1 + \sqrt{c}$ *then we define*

$$\begin{aligned} \tilde{u}_{\hat{\lambda}_{\theta_1}} &= \frac{\left(\hat{u}_{\hat{\lambda}_{\theta_1},1}, \hat{u}_{\hat{\lambda}_{\theta_1},2}, ..., \hat{u}_{\hat{\lambda}_{\theta_1},k}\right)}{\sqrt{\sum_{i=1}^k \hat{u}^2_{\hat{\lambda}_{\theta_1},i}}}, \\ \tilde{\tilde{u}}_{\hat{\lambda}_{\theta_1}} &= \frac{(\hat{u}_{\hat{\lambda}_{\theta_1},k+1}, \hat{u}_{\hat{\lambda}_{\theta_1},k+2}, ..., \hat{u}_{\hat{\lambda}_{\theta_1},m})}{\sqrt{\sum_{i=k+1}^m \hat{u}^2_{\hat{\lambda}_{\theta_1},i}}}, \end{aligned}$$

where $\hat{u}_{\hat{\lambda}_{\theta_1}}$ *is the empirical eigenvector of* $\hat{\Sigma}_{(m,n)}$ *corresponding to the empirical eigenvalue* $\lambda_{\theta_1}\left(\hat{\Sigma}_{(m,n)}\right)$ *described in the previous result 1.5.3. Then*

1.
$$\sqrt{n}(\tilde{u}_{\hat{\lambda}_{\theta_1}} - e_1) \xrightarrow[n\to\infty]{} \mathbf{N}\left(\vec{0}_k, \tilde{\Sigma}\right),$$
$$\text{where } \tilde{\Sigma} = \frac{1}{1 - \frac{c}{(\theta_1-1)^2}} \sum_{i=2}^k \frac{\theta_1\theta_i}{(\theta_1-\theta_i)^2} e_i e_i^t.$$

2. *The vector* $\tilde{\tilde{u}}_{\hat{\lambda}_{\theta_1}}$ *follows the invariant Haar distribution and is independent of* $\sum_{i=k+1}^m \hat{u}^2_{\hat{\lambda}_{\theta_1},i}$.

1.6 Spiked model under the deformed Marcenko-Pastur spectrum

This section is essentially based on Benaych-Georges and Rao [2009] for the convergence result and Benaych-Georges, Guionnet, and Maida [2010] for the distributions. We suppose that the structure of the random matrix can be understood through its spectrum and a finite perturbation.
The empirical covariance $\hat{\Sigma}_{(m,n)}$ has the following form:

$$\hat{\Sigma}_{(m,n)} = \frac{1}{n}\mathbf{X}\mathbf{X}^t, \ \mathbf{X} = \Sigma_m^{1/2}\tilde{\mathbf{X}}.$$

Therefore, if $\tilde{\Sigma}_{(m,n)} = \frac{1}{n}\tilde{\mathbf{X}}\mathbf{X}^t$,

$$\hat{\Sigma}_{(m,n)} = \Sigma_m^{1/2}\tilde{\Sigma}_{(m,n)}\Sigma_m^{1/2}.$$

We assume the spectrum of $\tilde{\Sigma}_{(m,n)}$, $F_{\tilde{\Sigma}_{(m,n)}} \xrightarrow[n\to\infty]{weakly} F_{\tilde{\Sigma}}$ with support $[a,b]$ and for k finite, $\Sigma_m = \mathrm{I}_m + \sum_{i=1}^{k}(\theta_i - 1)u_i u_i^t$, where $\theta_i \neq \theta_j$ if $i \neq j$. In order to study the eigenvalues we add some randomness to Σ_m by supposing $u_1, ..., u_k$ follows the same distribution as the k first columns of an invariant Haar random matrix. Equivalently in the normal case, we can assume without loss of generality $u_i = e_i$ and $\tilde{\Sigma}_{(m,n)} = O_m \tilde{\Lambda}_{(m,n)} O_m^t$ with O_m following a unit invariant Haar distribution of size m.

Remark 1.6.1.
The condition on u_i means that the result is valid if the eigenvectors of $\tilde{\Sigma}_m$ are independent of the u_i.

1. Spectrum: The spectrum is robust to a finite perturbation as in Result 1.5.1.

2. Largest empirical eigenvalue: As Result 1.5.2 we derive a detection condition by looking at the almost sure convergence of the empirical eigenvalues.

Result 1.6.1. *(Eigenvalue detection)*
Suppose $\hat{\Sigma}_{(m,n)}$ is defined as above and the support of $f_{\tilde{\Sigma}}$ is $[a,b]$, then two limits of eigenvalues can be described using the same notation as in Result 1.5.2 and the T-Transform defined in Definition 1.3.2. If $\theta_i > 1 + 1/T_{f_{\tilde{\Sigma}}}(b^+)$ then there exists $\lambda_{\theta_i}\left(\hat{\Sigma}_{(m,n)}\right)$ such that

$$\lambda_{\theta_i}\left(\hat{\Sigma}_{(m,n)}\right) \xrightarrow[n,m\to\infty]{a.s} T_{f_{\tilde{\Sigma}}}^{-1}\left(\frac{1}{\theta_i - 1}\right).$$

If $\theta_i < 1 + 1/T_{f_{\tilde{\Sigma}}}(b^+)$ and $\theta_1 > \theta_2 > ... > \theta_k$ then

$$\lambda_i\left(\hat{\Sigma}_{(m,n)}\right) \xrightarrow[n,m\to\infty]{a.s} b.$$

This convergence has been shown by Benaych-Georges and Rao [2009], whereas the distribution of the eigenvalues is described by Benaych-Georges, Guionnet, and Maida [2010] in Theorem 3.2. under Assumption 1.2 and Hypotheses 1.1 and 3.1. presented in the paper.

Result 1.6.2. *(Eigenvalue distribution)*
Under the conditions introduced by Benaych-Georges, Guionnet, and Maida [2010], the limit distribution of the eigenvalues can be described using the same notation as 1.6.1. If $\theta_i > 1 + 1/T_{f_{\tilde{\Sigma}}}(b^+)$ then

$$\sqrt{n}\left(\lambda_{\theta_i}\left(\hat{\Sigma}_{(m,n)}\right) - \nu_i\right) \xrightarrow[n,m\to\infty]{D} \mathrm{Normal}(0, \gamma_i),$$

where $\nu_i = T_{f_{\tilde{\Sigma}}}^{-1}\left(\frac{1}{\theta_i - 1}\right)$ and $\gamma_i = \frac{-m'_{f_{\tilde{\Sigma}}}(\nu_i) - \frac{1}{\theta_i}}{\left(m'_{f_{\tilde{\Sigma}}}(\nu_i)\right)^2}$.

In this thesis we will only consider the case where $\theta_i \neq \theta_j$ if $i \neq j$. The result of Benaych-Georges, Guionnet, and Maida [2010] is more general and considers the case with equal eigenvalues.

3. Eigenvectors: The asymptotic limit of the dot product provided in Result 1.5.4 is generalized in Benaych-Georges and Rao [2009].

Result 1.6.3. *(Eigenvector bias)*
Using the same notation as Results 1.6.1 and 1.5.4: If $\theta_i > 1 + 1/T_{f_{\tilde{\Sigma}}}(b^+)$ then

$$|\left\langle u_i, \hat{u}_{\lambda_{\theta_i}} \right\rangle| \xrightarrow[n,m\to\infty]{a.s} \sqrt{-\frac{\theta_i}{(\theta_i - 1)^2 \nu_i T'_{f_{\tilde{\Sigma}}}(\nu_i)}}.$$

If $\theta_i < 1 + 1/T_{f_{\tilde{\Sigma}}}(b^+)$ *then for all* j,

$$\langle u_i, \hat{u}_j \rangle \xrightarrow[n,m\to\infty]{a.s} 0.$$

All the results of this section are generalisations of the previous section. The second part of 1.6.3 is weaker than the previous result 1.5.4.

1.7 Linear spectral statistics

This section presents a result of RMT using the whole spectrum of a matrix. This result leads to an interesting application in statistical testing.
We assume that $\mathbf{X}$ and $\tilde{\mathbf{X}}$ satisfy

$$\mathbf{X} = \Sigma_m^{1/2}\tilde{\mathbf{X}},$$

where for $i = 1, 2, ..., m$ and $j = 1, 2, ..., n$,

$$\begin{aligned} &\tilde{\mathbf{X}}_{i,j} \text{ are independent}, \\ &\mathrm{E}\left[\tilde{\mathbf{X}}_{i,j}\right] = 0, \\ &\mathrm{Var}\left(\tilde{\mathbf{X}}_{i,j}\right) = 1, \\ &\mathrm{E}\left[\tilde{\mathbf{X}}_{i,j}^4\right] = 3. \end{aligned}$$

For example $\tilde{\mathbf{X}}_{i,j}$ are i.i.d. $\mathbf{N}(0,1)$. Using these conditions on the matrix, Bai and Silverstein [2004] provide a central limit theorem for linear spectral statistics.

Result 1.7.1. *(Linear Spectral Statistics)*
Suppose $c \in \,]0,\infty[$, $\lim_{m,n\to\infty} \frac{m}{n} = c$ *and* $\hat{\Sigma}_{(m,n)} = \frac{1}{n}\mathbf{X}\mathbf{X}^t = \frac{1}{n}\Sigma_m^{1/2}\tilde{\mathbf{X}}\tilde{\mathbf{X}}\Sigma_m^{1/2}$, *where* $\tilde{X}$ *is defined above. Moreover, we assume that*

- f_m *is the spectral density induced by the spectrum cumulative distribution* F_m *of* $\hat{\Sigma}_m$. *Note that* f_m *is not an empirical distribution but the true non-asymptotic distribution of the spectrum.* ($F_m \neq F_{\hat{\Sigma}_m}$),
- *when* m *tends to ininifty,* F_{Σ_m} *tends to* F_Σ *and* f_{Σ_m} *tends to* f_Σ,
- $\mathrm{Trace}\left(\Sigma_m\right)$ *is bounded in* m.

Then

$$m\left(\frac{1}{m}\sum_{i=1}^{m} g_1\left(\lambda_i\left(\hat{\Sigma}_{(m,n)}\right)\right) - \int g_1(\lambda) f_m(\lambda)\, d\lambda \right.$$
$$\left. ,..., \frac{1}{m}\sum_{i=1}^{m} g_k\left(\lambda_i\left(\hat{\Sigma}_{(m,n)}\right)\right) - \int g_k(\lambda) f_m(\lambda)\, d\lambda \right) \xrightarrow[m,n\to\infty]{D} \mathbf{N}\left(\mu_{\vec{g},f_\Sigma,c}, \Sigma_{\vec{g},f_\Sigma,c}\right),$$

where

$$\begin{aligned} \mu_{\vec{g},f,s} &= -\frac{1}{2\pi i}\int_{\mathcal{C}} g_s(z) \frac{c\int \frac{m_\phi(z)^3 t^2}{\left(1+tm_\phi(z)\right)^3} f_\Sigma(t)dt}{\left(1 - c\int \frac{m_\phi(z)^2 t^2}{\left(1+tm_\phi(z)\right)^2} f_\Sigma(z)d(z)\right)^2} dz, \\ \Sigma_{\vec{g},f,s,r} &= -\frac{1}{2\pi^2}\int_{\mathcal{C}_1}\int_{\mathcal{C}_2} \frac{g_s(z_1)g_r(z_2)}{\left(m_\phi(z_1) - m_\phi(z_2)\right)^2} m'_\phi(z_1)m'_\phi(z_2)dz_1dz_2, \\ m_\phi(z) &= -\frac{1-c}{z} m_{f_\Sigma}(z), \end{aligned}$$

where m_{f_Σ}, m_ϕ *are the Stieltjes transforms of* f_Σ *and* ϕ. *We note that* $\phi = (1-c)\mathbf{1}_{[0,\infty[} + cf_\Sigma$.

The interesting fact we exploit in this thesis is the order $1/m$, which shows that estimating an asymptotic spectral statistic by the empirical one leads to an error of size $1/m$.

1.8 Statistical issues

RMT offers powerful applied tools. The aim of this thesis will be to detect differences between two random matrices by looking at the largest estimated eigenvalues (spikes). We will consider two such applications. The first one builds a test to detect a perturbation in a covariance matrix and the second one tests the equality between two groups by comparing two Wishart random matrices.

1.8.1 Spike model detection

Result 1.2.2 shows that the largest eigenvalue of random matrix with i.i.d. entries tends to a Tracy-Widom distribution. Based on this, Bianchi, Debbah, Maida, and Najim [2011] constructed a test to detect a signal surrounded by white noise. The approach was first developed in Wax and Kailath [1985] without using RMT.

Result 1.8.1.
Assume that $\mathbf{X}$ *is a random matrix of dimension* $(m \times n)$ *and the hypothesis of test are*

$$\begin{aligned} \mathrm{H}_0 &: \mathbf{X} = \tilde{\mathbf{X}}, \\ \mathrm{H}_1 &: \mathbf{X} = P^{1/2}\tilde{\mathbf{X}}, \end{aligned}$$

where $\tilde{\mathbf{X}}$ *has i.i.d. entries and* $P = \mathrm{I}_m + (\theta - 1)uu^t$ *with* $\theta > 1$.
The test statistic

$$S = \frac{\lambda_{\max}\left(\frac{1}{n}\mathbf{X}\mathbf{X}^t\right)}{\mathrm{Trace}\left(\frac{1}{n}\mathbf{X}\mathbf{X}^t\right)}$$

allows us to control the level α *of false discovery.*
Under H_0,

$$\frac{n^{2/3}}{\Theta_m}\left(S - (1+\sqrt{c})^2\right) \xrightarrow[n,m\to\infty]{D} \mathrm{TW},$$

where $\lim_{m,n\to\infty} \frac{m}{n} = c \in \,]0,\infty[$ *and* $\Theta_m = (1+\sqrt{c})\left(\frac{1}{\sqrt{c}}+1\right)^{1/3}$.
Under H_1,

$$\begin{aligned} &\text{If } \theta > 1+\sqrt{c} \text{ then } S \xrightarrow[n,m\to\infty]{P} \theta\left(1+\frac{c}{\theta_1 - 1}\right), \\ &\text{If } \theta < 1+\sqrt{c} \text{ then } S \xrightarrow[n,m\to\infty]{P} (1+\sqrt{c})^2. \end{aligned}$$

The paper motivates our work in several aspects. The Tracy-Widom distribution statistics occur when applying a generalized likelihood-ratio test. The procedure we will propose in this thesis can be seen as a modified generalized likelihood-ratio test. Moreover, filtering the noise by looking at spikes leads to more powerful tests when m is large and in this work we also attempt to detect differences by looking at spikes.

Spike detection can also be used to filter an image containing noise as in Ray et al. [2012]. Another interesting application filters the noise in order to determine the number of endmembers in a image Cawse-Nicholson et al. [2010]. Finally, we still cite Mestre and Lagunas [2008] and Vallet et al. [2011] for the application of random matrix to the multiple signal classification (MUSIC) algorithm using m antennas that detect the source of a signal. Nevertheless this last application exploits random matrices in the complex plane.

1.8.2 Extended Application

The test of Bianchi, Debbah, Maida, and Najim [2011] could be extended using Result 1.4.2 to more general matrices. Doing so, would, however require the estimation of the limiting spectrum associated with F_{Σ_m}. This problem is discussed in El Karoui [2008].

1.8.3 Difference between two random matrices

Bai, Jiang, Yao, and Zheng [2009] use the linear spectral statistic result 1.7.1 to test the equality between two covariance matrices.

Result 1.8.2. *Assume* $\mathbf{X}$ *and* $\mathbf{Y}$ *are random matrices of dimension* $m \times n_X$ *and* $m \times n_Y$ *respectively, such that*

$$\mathbf{X} = \Sigma_X \tilde{\mathbf{X}} \text{ and } \mathbf{Y} = \Sigma_Y \tilde{\mathbf{Y}},$$

where the entries of $\tilde{\mathbf{X}}$ *and* $\tilde{\mathbf{Y}}$ *are i.i.d.* $\mathbf{N}(0,1)$. *The hypothesis to be tested are*

$$\begin{aligned} \mathrm{H}_0 &: \Sigma_X = \Sigma_Y, \\ \mathrm{H}_1 &: \Sigma_X \neq \Sigma_Y. \end{aligned}$$

We define

$$\begin{aligned} \hat{\Sigma}_X &= \frac{1}{n_X}\mathbf{X}\mathbf{X}^t \text{ and } \hat{\Sigma}_Y = \frac{1}{n_Y}\mathbf{Y}\mathbf{Y}^t, \\ c_X &= \frac{m}{n_X},\ c_Y = \frac{m}{n_Y} \text{ and } N = n_X + n_Y. \end{aligned}$$

The test statistic,

$$S = -\frac{2\log\left(\frac{|\hat{\Sigma}_X|^{n_X/2}|\hat{\Sigma}_Y|^{n_Y/2}}{|c_X\hat{\Sigma}_X + c_Y\hat{\Sigma}_Y|^{N/2}}\right)}{Nm},$$

allows us to control the level α *of false rejections asymptotically based on the following limit. Under* H_0,

$$m(S - \mu_S) \xrightarrow[n,m\to\infty]{D} \mathbf{N}\left(\tilde{\mu}_S, \sigma_S^2\right),$$

where

$$\begin{aligned} \mu_S &= -\frac{c_X + c_Y - c_Xc_Y}{c_Xc_Y}\log(c_X + c_Y - c_Xc_Y) + \frac{c_X + c_Y - c_Xc_Y}{c_Xc_Y}\log(c_X + c_Y) \\ &\quad + \frac{c_X(1-c_Y)}{c_Y(c_X+c_Y)}\log(1-c_Y) + \frac{c_Y(1-c_X)}{c_X(c_X+c_Y)}\log(1-c_X), \\ \tilde{\mu}_S &= \frac{1}{2}\left[\log\left(\frac{c_X + c_Y - c_Xc_Y}{c_X+c_Y}\right) - \frac{c_X}{c_X+c_Y}\log(1-c_Y) - \frac{c_Y}{c_X+c_Y}\log(1-c_X)\right], \\ \sigma_S^2 &= -\frac{2c_Y^2}{(c_X+c_Y)^2}\log(1-c_X) - \frac{2c_X^2}{(c_X+c_Y)^2}\log(1-c_Y) - 2\log\left(\frac{c_X+c_Y}{c_X+c_Y-c_Xc_Y}\right). \end{aligned}$$

It turns out that this test is powerful in detecting differences affecting the entire spectrum. However, for differences of finite rank, the power is weak. Moreover, the test is not robust under non-Gaussian entries. The procedure we will develop in this thesis will improve these two aspects.

1.9 Conclusion

The aim of this thesis is to exploit theoretical results to construct a powerful robust procedure for comparing two estimated covariance matrices.
We are not convinced that looking at the entire spectrum is the best approach to detect perturbations. Therefore, we choose to study spikes in general cases. Our work can be seen as a mixture of

- Spike model detection, because we focus on the spike 1.8.1,
- The extended application of El Karoui [2008], because we assume general spectra,
- The difference procedure presented in 1.8.3.

The strength of our procedure will be:

- its robustness,
- its power when the difference is a complex finite perturbation.

On the other hand, the main weakness is

- the lack of power when the difference is sparse and explained by a few parameters.

Therefore, this work is not intended as a replacement of the usual procedure dealing with random matrices, but instead proposes an alternative when the data or the differences are complex.

Chapter 2

Introduction to residual spikes and our model

We are interested in comparing the random covariance matrices of two groups of observations X and Y.
This chapter begins with a small introduction to define the matrices of our study in Section 2.1. Then Section 2.2 defines residual spikes and uses them to detect differences between two covariance matrices.

2.1 Residual spike in a particular case

In this section, we assume a simple normal model and introduce the concept of residual eigenvalues.

2.1.1 The matrices and their estimators

Let $X_1, ..., X_n$ and $Y_1, ..., Y_n$ be i.i.d Normal$(0, \Sigma_m)$ where $\Sigma_m = \mathrm{I}_m + \sum_{i=1}^{k}(\theta_i - 1)u_i u_i^t$ with k finite. If $\mathrm{Var}\,(X_1) = \Sigma_X \neq \Sigma_Y = \mathrm{Var}\,(Y_1)$, we write $u_{X,i}$, $\theta_{X,i}$, $u_{Y,i}$ and $\theta_{Y,i}$ for the parameters of the model. The usual estimates of the covariances are called $\hat{\Sigma}_X$ and $\hat{\Sigma}_Y$. To avoid some difficulties we assume normality of the data. Without loss of generality we can assume $u_i = e_i$, i.e., the canonical basis vectors. In this thesis we will usually assume spherical data. The non-spherical case is investigated in the section about robustness, Section 4.2.
Suppose that the perturbations are **detectable** as defined in Definition 5.1.2. In this particular case, this means that $\min_{i=1,2,...k} \theta_i > 1 + \sqrt{\frac{m}{n}}$. If this is the case, we can filter our matrices to obtain

$$\hat{\hat{\Sigma}}_X = \mathrm{I}_m + \sum_{i=1}^{k}(\hat{\hat{\theta}}_{X,i} - 1)\hat{u}_{X,i}\hat{u}_{X,i}^t,$$
$$\hat{\hat{\Sigma}}_Y = \mathrm{I}_m + \sum_{i=1}^{k}(\hat{\hat{\theta}}_{Y,i} - 1)\hat{u}_{Y,i}\hat{u}_{Y,i}^t,$$

where $\hat{\hat{\theta}}_{X,i}$ is an asymptotic unbiased estimator of θ_i using Result 1.5.2. Indeed, because

$$\hat{\theta}_{X,i} \to \theta_i \left(1 + \frac{c}{\theta_i + 1}\right),$$

we have that

$$\hat{\hat{\theta}}_{X,i} \text{ such that } \hat{\theta}_{X,i} = \hat{\hat{\theta}}_{X,i} \left(1 + \frac{c}{\hat{\hat{\theta}}_{X,i} + 1}\right)$$

is asymptotically unbiased.
In order to detect the difference between the matrices, we are led to look at the eigenvalues of

$$\hat{\hat{\Sigma}}_X^{-1/2}\hat{\Sigma}_Y\hat{\hat{\Sigma}}_X^{-1/2} \text{ and } \hat{\hat{\Sigma}}_Y^{-1/2}\hat{\Sigma}_X\hat{\hat{\Sigma}}_Y^{-1/2} \tag{2.1}$$

or

$$\hat{\hat{\Sigma}}_X^{-1/2}\hat{\hat{\Sigma}}_Y\hat{\hat{\Sigma}}_X^{-1/2} \text{ and } \hat{\hat{\Sigma}}_Y^{-1/2}\hat{\hat{\Sigma}}_X\hat{\hat{\Sigma}}_Y^{-1/2}. \tag{2.2}$$

We name these matrices the type 1 matrix and the type 2 matrix respectively. The principal aim of our thesis is to understand the second type.

Remark 2.1.1.

1. We assume $u_i = e_i$ without loss of generality because a rotation does not change the eigenvalues of matrices (2.1) and (2.2). Moreover, if the data are normal, the distributions of the matrices are invariant by rotation. We could also just assume a spherical distribution for the data.
   As in Benaych-Georges and Rao [2009], the spherical distribution is not necessary and for many results, assuming u_i to be random invariant by rotation is equivalent. This is, however, not always the case as we will see in Section 4.2.
2. The results of this chapter assume spherical data such that the spectrum of the estimated covariance matrix asymptotically follows asymptotically the Marcenko-Pastur distribution. Many results are still valid for more general spectra.

2.1.2 Introduction of the residual spike

If $\Sigma_Y = \Sigma_X = \Sigma$ then by looking at

$$\hat{\hat{\Sigma}}_X^{-1/2}\hat{\Sigma}_Y\hat{\hat{\Sigma}}_X^{-1/2}$$

the distribution of the spectrum follows a Marcenko-Pastur distribution. However, because of the non-consistency of the eigenvectors presented in Chapter 1, we observe some **residual spikes** that we define later. These spikes are shown in Figure 2.1, which essentially summarizes why we need to investigate the eigenvalues of this matrix.
Indeed even if the two random matrices are based on the same matrix, we see some spikes outside the bulk. This observation is worse in the fourth plot because four spikes fall outside the bulk even if there is actually no difference!

How can we distinguish the spikes indicative of a true difference from the residual spikes?

This question is difficult and as we will see, the weaker the assumptions on the data, the more difficult it is to obtain results. In order to provide a partial answer, we first define the notion of residual.

Definition 2.1.1.
The **residual spikes** of type 1 are the isolated eigenvalues of

$$\hat{\hat{\Sigma}}_X^{-1/2}\hat{\Sigma}_Y\hat{\hat{\Sigma}}_X^{-1/2}.$$

The **residual spikes** of type 2 are the eigenvalues of

$$\hat{\hat{\Sigma}}_X^{-1/2}\hat{\hat{\Sigma}}_Y\hat{\hat{\Sigma}}_X^{-1/2}$$

that do not converge toward 1.
The **residual zone** of type 1 (or 2) is the interval where a residual spike of type 1 (or 2) falls asymptotically.

This thesis studies these residual spikes when the true covariance matrices of the two groups are the same perturbation of order k, $\Sigma_k = \Sigma_X = \Sigma_X$. In this null case, we provide a bound for the largest residual spike. Consequently, we are not able to distinguish a true small spike from a residual one. However, when a spike is asymptotically larger than the bound, then this spike is likely not a residual spike. This philosophy is explained in Figure 2.2. All the eigenvalues lying in what we call the residual zone are potentially not real differences. However, when an eigenvalue is larger, this spike expresses a true difference.

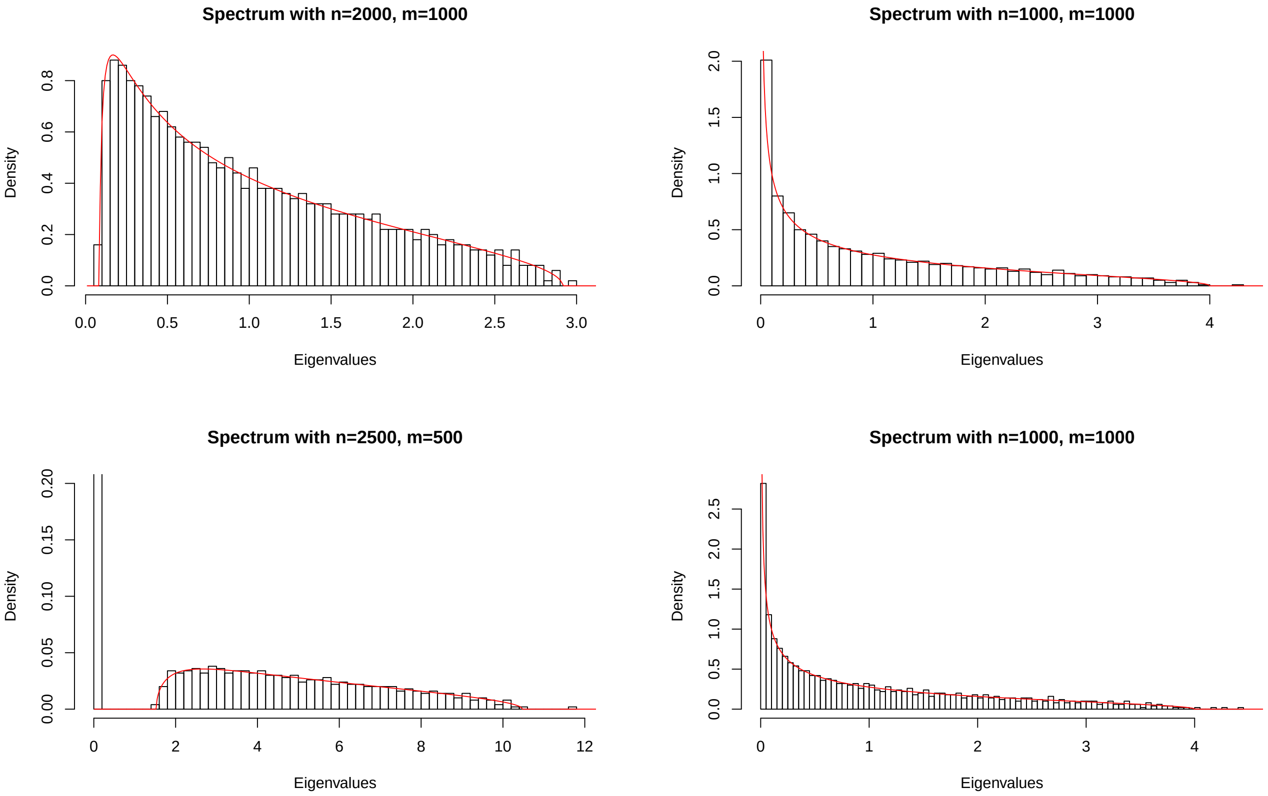


Figure 2.1 – Example of residual spikes of $\hat{\hat{\Sigma}}_X^{-1/2}\hat{\Sigma}_Y\hat{\hat{\Sigma}}_X^{-1/2}$ when $\theta = 10$ for the first three figures and $\theta_{1,2,3,4} = 10, 15, 20, 25$ for the last figure. The residual spikes are the largest isolated eigenvalues.

As shown in Figure 2.2, the matrices of type 2,

$$\hat{\hat{\Sigma}}_X^{-1/2}\hat{\hat{\Sigma}}_Y\hat{\hat{\Sigma}}_X^{-1/2},$$

also lead to residual spikes when $\Sigma_Y = \Sigma_X = \Sigma$. These spikes do not converge to 1. In this thesis, we will focus on this second type. Although the first type seems easier to study in term of convergence in probability when the perturbation is of order 1, this is no longer the case in more complex situations. Moreover, investigation of robustness is easier for the second type. Finally, the eigenvector of the residual spike is not studied in detail in this thesis, however the author thinks that this task is simpler for type 2 matrices.

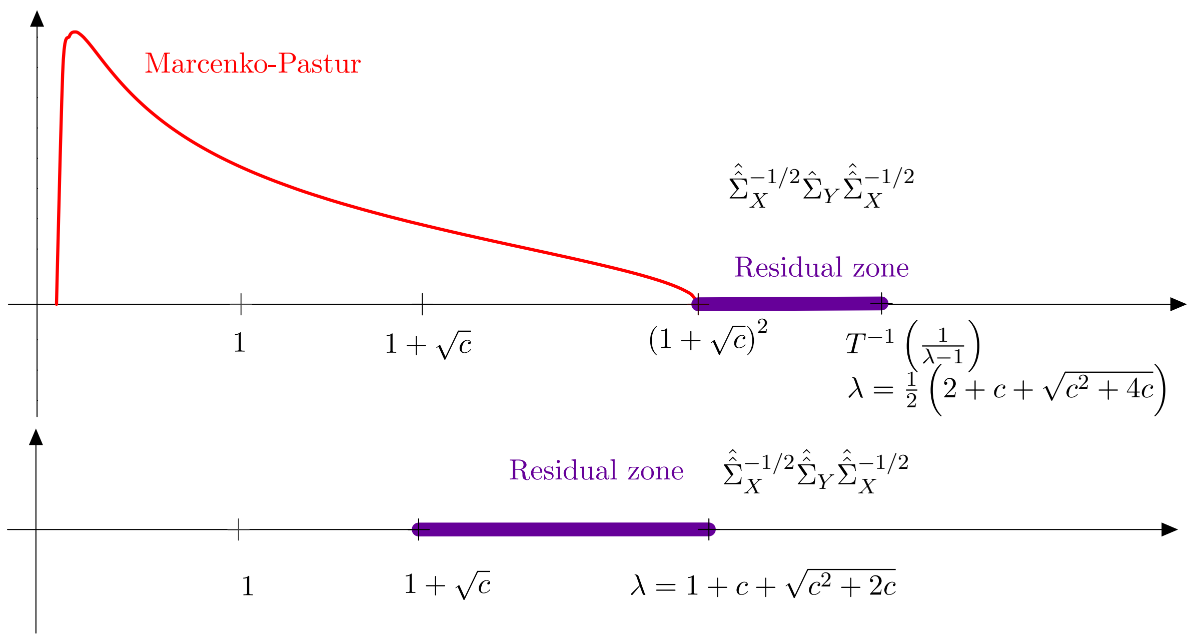


Figure 2.2 – Residual zones of $\hat{\hat{\Sigma}}_X^{-1/2}\hat{\Sigma}_Y\hat{\hat{\Sigma}}_X^{-1/2}$ and $\hat{\hat{\Sigma}}_X^{-1/2}\hat{\hat{\Sigma}}_Y\hat{\hat{\Sigma}}_X^{-1/2}$.

2.2 Introduction of the model, the test and the hypothesis

In the previous section, we introduced the residual spike when the data are very regular. In order to generalise this work to a broader set of conditions, we introduce more general random covariance matrices.

2.2.1 Introduction of the model

Let $X_1, ..., X_{n_X}$ and $Y_1, ..., Y_{n_Y}$ be our set of data. In this thesis, we investigate the eigen-structure of the usual covariance matrix estimators

$$\hat{\Sigma}_X = \frac{1}{n_X}\sum_{i=1}^{n_X} X_i X_i^t \text{ and } \hat{\Sigma}_Y = \frac{1}{n_Y}\sum_{i=1}^{n_Y} Y_i Y_i^t.$$

In Section 4.2, we will relax some assumptions on the data. For the moment, however, we assume the following to hold.

Assumption 2.2.1.
Let W_X and W_Y be random matrices such that

$$W_X = O_X \Lambda_X O_X \text{ and } W_Y = O_Y \Lambda_Y O_Y,$$

where

O_X, O_Y are unit orthonormal invariant and independent random matrices,
Λ_X, Λ_Y are diagonal bounded matrices and independent of O_X, O_Y,
$\text{Trace}\,(W_X) = 1$ and $\text{Trace}\,(W_Y) = 1$.

Assume $P_X = \mathrm{I}_m + \sum_{i=1}^{k}(\theta_{X,i}-1)e_i e_i^t$ and $P_Y = \mathrm{I}_m + \sum_{i=1}^{k}(\theta_{Y,i}-1)e_i e_i^t$ with k finite, $\theta_{X,i} \neq \theta_{X,j}$ and $\theta_{Y,i} \neq \theta_{Y,j}$ if $i \neq j$. Then

$$\hat{\Sigma}_X = P_X^{1/2} W_X P_X^{1/2} \text{ and } \hat{\Sigma}_Y = P_Y^{1/2} W_Y P_Y^{1/2}.$$

Remark 2.2.1.

1. We say that $\mathbf{X} \in \mathbb{R}^m \times \mathbb{R}^{n_X}$ and $\mathbf{Y} \in \mathbb{R}^m \times \mathbb{R}^{n_Y}$ respect this assumption if
$$\hat{\Sigma}_X = \frac{1}{n_X}\mathbf{X}\mathbf{X}^t \text{ and } \hat{\Sigma}_Y = \frac{1}{n_Y}\mathbf{Y}\mathbf{Y}^t$$
respect the assumption.

2. Because O_X and O_Y are independent and invariant by rotation, the assumption of canonical perturbations P_X and P_Y is assumed without loss of generality as in Benaych-Georges and Rao [2009].

3. In the theoretical part, we will assume that Λ_X and Λ_Y are observed. However, in practice we will estimate them by estimators that do not affect the asymptotic result. Moreover, it can be shown that in many cases the spectra are independent of O_X and O_Y.

4. If $X_1, ..., X_{n_X}$ are i.i.d Normal$(0, \Sigma_m)$ where $\Sigma_m = \mathrm{I}_m + \sum_{i=1}^k (\theta_i - 1)e_i e_i^t$ with k finite, then
$$\hat{\Sigma}_X = \Sigma_m^{1/2} W_X \Sigma_m^{1/2}$$
and
$$W_X = O_X \Lambda_X O_X,$$
where Λ_X is asymptotically the Marcenko-Pastur spectrum.

5. If $P_X = P_Y$ we use the notation P and $\theta_{X,i} = \theta_{Y,i} = \theta_i$.

Remark 2.2.2. In this thesis we use the usual estimator of covariance. Nevertheless an interesting generalisation would consider robust covariance estimators and will be treated in a future work.

The stronger results of this thesis will often assume large eigenvalues. However, we always try to give a result for general eigenvalues. The main weakness of our proofs is the assumption of non-equality of the eigenvalues of the perturbation. Although we are not able to prove it, we are convinced that our Main Theorem is still valid without this assumption. The following are possible assumptions about the perturbations:

Assumption 2.2.2.

(A1) $\frac{\theta}{\sqrt{m}} \to \infty$, as $n, m \to \infty$.

(A2) $\theta \equiv \theta(m) \to \infty$.

(A3) $\theta_i = p_i \theta$, where p_i is fixed different from 1.

(A4) For $i = 1, ..., k_\infty,\ \theta_i = p_i\theta,\ \theta \to \infty$ according to (A1) or (A2),
For $i = k_\infty + 1, ..., k,\ \theta_i = p_i \theta_0$.
For all $i \neq j$, $p_i/p_j \not\to 1$.

(A5) No condition on the eigenvalues and the following conjecture is true.

Conjecture 2.2.1.
Theorem 5.8.1 part 3 can be improved such that for all $s = 1, 2, ..., k$,
$$\sum_{i=k+1}^{m} \hat{u}_{i,s}^2 = O_p\left(\frac{1}{\theta_s}\right).$$

Remark 2.2.1.1.

1. The conjecture is trivially true for $k = 1$ and we can prove it for $k = 2$ using the tools of this thesis.
2. The Main Theorem and a majority of the results assume (A4). Under assumption (A5), many important theorems are easily shown, with the exception of the Main Theorem. Despite this fact, the author is convinced that assuming (A5) is sufficient.

Throughout the thesis we always work under the assumption (A4).

2.2.2 Introduction of the procedure

We want to compare two estimated random covariance matrices $\hat{\Sigma}_X$ and $\hat{\Sigma}_Y$ obtained from n_X and n_Y observed random vectors of size m. We assumed $n_X > n_Y$.
First we define the filtered estimator of a covariance matrix.

Definition 2.2.1.
Suppose $\hat{\Sigma}$ respects Assumption 2.2.1.
The **unbiased estimators of** θ_i for $i = 1, 2, ..., k$ are defined as

$$\hat{\hat{\theta}}_i = 1 + \frac{1}{\frac{1}{m-k}\sum_{j=k+1}^m \frac{\hat{\lambda}_{\hat{\Sigma},j}}{\hat{\theta}_i - \hat{\lambda}_{\hat{\Sigma},j}}}.$$

Suppose that for $i = 1, 2, ..., k$, $\hat{u}_i$ is the corresponding eigenvectors of $\hat{\theta}_i$.
The **filtered estimated covariance matrix** is defined as

$$\hat{\hat{\Sigma}} = \mathrm{I}_m + \sum_{i=1}^k (\hat{\hat{\theta}}_i - 1)\hat{u}_i\hat{u}_i^t.$$

Remark 2.2.3.

1. We will often use a theoretical unbiased estimator in proofs,

$$\hat{\hat{\theta}}_i = 1 + \frac{1}{\frac{1}{m}\sum_{j=1}^m \frac{\hat{\lambda}_{W,j}}{\hat{\theta}_i - \hat{\lambda}_{W,j}}}.$$

   Although we can show that these unbiased estimators are asymptotically equivalent under Assumption 2.2.1, we will always specify which estimator is used.

We can now introduce the procedure investigated in this thesis.

1. If all the perturbations are detectable as in Result 1.6.1 part 1 or Definition 5.1.2, their k largest eigenvalues are $\hat{\theta}_{X,1}, \hat{\theta}_{X,2}, ..., \hat{\theta}_{X,k}$ and $\hat{\theta}_{Y,1}, \hat{\theta}_{Y,2}, ..., \hat{\theta}_{Y,k}$. Using an empirical inverse T-transform, we obtain the asymptotic unbiased estimators:

$$\hat{\hat{\theta}}_{X,i} = 1 + \frac{1}{\frac{1}{m}\sum_{j=1}^m \frac{\hat{\lambda}_{W_X,j}}{\hat{\theta}_{X,i} - \hat{\lambda}_{W_X,j}}} \text{ and } \hat{\hat{\theta}}_{Y,i} = 1 + \frac{1}{\frac{1}{m}\sum_{j=1}^m \frac{\hat{\lambda}_{W_Y,j}}{\hat{\theta}_{Y,i} - \hat{\lambda}_{W_Y,j}}},$$

   where $\hat{\lambda}_{W_X,j}$ and $\hat{\lambda}_{W_Y,j}$ are the eigenvalues of W_X and W_Y respectively.

   Remark 2.2.4.

In practice we do not know W_X and W_Y and we thus replace the previous estimator by

$$\hat{\hat{\theta}}_{X,i} = 1 + \frac{1}{\frac{1}{m-k}\sum_{j=k+1}^{m} \frac{\hat{\lambda}_{\hat{\Sigma}_X,j}}{\hat{\theta}_{X,i}-\hat{\lambda}_{\hat{\Sigma}_X,j}}} \text{ and } \hat{\hat{\theta}}_{Y,i} = 1 + \frac{1}{\frac{1}{m-k}\sum_{j=k+1}^{m} \frac{\hat{\lambda}_{\hat{\Sigma}_Y,j}}{\hat{\theta}_{Y,i}-\hat{\lambda}_{\hat{\Sigma}_Y,j}}},$$

where $\hat{\lambda}_{\hat{\Sigma}_X,j}$ and $\hat{\lambda}_{\hat{\Sigma}_Y,j}$ are the j^{th} ordered eigenvalues of $\hat{\Sigma}_X$ and $\hat{\Sigma}_Y$ respectively.
This modification does not change the asymptotic result, as we will show in Section 4.4.

2. Using the unbiased estimators of the eigenvalues and the estimator of the eigenvectors, we build the filtered estimators of the covariance matrices:

$$\hat{\hat{\Sigma}}_X = \mathrm{I}_m + \sum_{i=1}^{k}(\hat{\hat{\theta}}_{X,i} - 1)\hat{u}_{X,i}\hat{u}_{X,i}^t \text{ and } \hat{\hat{\Sigma}}_Y = \mathrm{I}_m + \sum_{i=1}^{k}(\hat{\hat{\theta}}_{Y,i} - 1)\hat{u}_{Y,i}\hat{u}_{Y,i}^t,$$

where $\hat{u}_{X,i}$ and $\hat{u}_{Y,i}$ are the eigenvectors of $\hat{\Sigma}_X$ and $\hat{\Sigma}_Y$ respectively.

3. Finally, we look at the statistics

$$\lambda_{\min}\left(\hat{\hat{\Sigma}}_X^{-1/2}\hat{\hat{\Sigma}}_Y\hat{\hat{\Sigma}}_X^{-1/2}\right) \text{ and } \lambda_{\max}\left(\hat{\hat{\Sigma}}_X^{-1/2}\hat{\hat{\Sigma}}_Y\hat{\hat{\Sigma}}_X^{-1/2}\right).$$

These statistics provide a very powerful and robust test for the equality of the perturbations P_X and P_Y.

Remark 2.2.5.
More precisely, the simulations of Chapter 8 seem to show that in order to maintain conservativeness when $n_X > n_Y$ are not very large, we should

- Estimate $\lambda_{\min}\left(\hat{\hat{\Sigma}}_X^{-1/2}\hat{\hat{\Sigma}}_Y\hat{\hat{\Sigma}}_X^{-1/2}\right)$ using Theorem 3.1.1,
- Estimate $\lambda_{\max}\left(\hat{\hat{\Sigma}}_X^{-1/2}\hat{\hat{\Sigma}}_Y\hat{\hat{\Sigma}}_X^{-1/2}\right)$ using Theorem 3.1.1 on $\lambda_{\min}\left(\hat{\hat{\Sigma}}_Y^{-1/2}\hat{\hat{\Sigma}}_X\hat{\hat{\Sigma}}_Y^{-1/2}\right)$ and then invert the estimate.

When n_X is large and n_Y is not very large, we could simply apply Theorem 3.1.1 to $\hat{\hat{\Sigma}}_X^{-1/2}\hat{\hat{\Sigma}}_Y\hat{\hat{\Sigma}}_X^{-1/2}$.
Finally, when both n_X and n_Y are large, both methods are equivalent.

2.2.3 The test

Under Assumption 2.2.1, we can test

$$H_0 : P = P_X = P_Y,$$
$$H_1 : P_X \neq P_Y.$$

Under H_0 :

$$V_{\min} \leqslant \lambda_{\min}\left(\hat{\hat{\Sigma}}_X^{-1/2}\hat{\hat{\Sigma}}_Y\hat{\hat{\Sigma}}_X^{-1/2}\right) \text{ and } \lambda_{\max}\left(\hat{\hat{\Sigma}}_X^{-1/2}\hat{\hat{\Sigma}}_Y\hat{\hat{\Sigma}}_X^{-1/2}\right) \leqslant V_{\max},$$

where $V_{\max}$ and $V_{\min}$ are random variables depending on m and the spectra of W_X and W_Y. These random variables are introduced in Section 3.1.
In practice, we observe the extreme eigenvalues $\hat{\lambda}_{\max,obs}$ and $\hat{\lambda}_{\min,obs}$. Knowing the distribution of $V_{\max}$ and $V_{\min}$, the test rejects H_0 if either $P\left(V_{\max} > \hat{\lambda}_{\max,obs}\right)$ or $P\left(V_{\min} < \hat{\lambda}_{\min,obs}\right)$ is small.

Remark 2.2.6.
This test is similar to the generalised Neyman-Pearson test of equality of variances under the assumption of normality. In this case, we first replace the maximum likelihood estimator by the filter estimator to build a more powerful test. Then a determinant statistic is replaced by the extreme eigenvalues to obtain a more robust test.
As an alternative, another test using the joint distribution of the residual spikes under the normality assumption would improve the power, but lacks robustness. This procedure is not considered any further in this thesis.

2.2.4 Assumptions on θ

Obviously under H_0, $\lambda_{\max}\left(\hat{\hat{\Sigma}}_X^{1/2}\hat{\hat{\Sigma}}_Y\hat{\hat{\Sigma}}_X^{1/2}\right)$ is a function of $\theta_1, ..., \theta_k$. In Section 4.2, we propose two criteria 4.2.5 and 4.2.8 to check if

$$\lambda_{\max}\left(\hat{\hat{\Sigma}}_X^{-1/2}\hat{\hat{\Sigma}}_Y\hat{\hat{\Sigma}}_X^{-1/2}\right) \leqslant \lim_{\theta\to\infty}\lambda_{\max}\left(\hat{\hat{\Sigma}}_X^{-1/2}\hat{\hat{\Sigma}}_Y\hat{\hat{\Sigma}}_X^{-1/2}\right) = V_{\max},$$
$$\lambda_{\min}\left(\hat{\hat{\Sigma}}_X^{-1/2}\hat{\hat{\Sigma}}_Y\hat{\hat{\Sigma}}_X^{-1/2}\right) \geqslant \lim_{\theta\to\infty}\lambda_{\min}\left(\hat{\hat{\Sigma}}_X^{-1/2}\hat{\hat{\Sigma}}_Y\hat{\hat{\Sigma}}_X^{-1/2}\right) = V_{\min},$$

where $\theta \to \infty$ means all the θ_i tend to infinity. The previous inequalities assume that the extreme spikes appear when the eigenvalues of the perturbation are large. We call this the **worst case**. Because we focus on the worst case scenario under H_0, we will often assume $\theta_i \to \infty$ or $\frac{\theta_i}{\sqrt{n}} \to \infty$ as in Assumption 2.2.2.

Remark 2.2.7.
This thesis will always investigate perturbations with $\theta_i > 1$ and we may well ask, what happens if there exist some small perturbations $\theta_i < 1$?
Some easy arguments show that we can neglect them in the estimation of the filtered estimator without changing the procedure. Moreover, in the degenerate case, these values are not detectable. However, if we really want to take care of the small eigenvalues, the procedure is still valid. In this case we will assume that the small eigenvalues tend to 0 instead of infinity.

Chapter 3

The Main Theorem

3.1 Main Theorem

In this section, we present the main Theorem of the thesis concerning the distribution of extreme residual spikes under appropriate assumptions. This theorem is the basis for testing the difference between two covariance matrices as presented in Section 2.2.3. More precisely, it characterises the asymptotic distributions of $V_{\max}$ and $V_{\min}$.
In order to obtain useful results, assumptions that will be further discussed and relaxed in Section 4.2 are needed. The conclusion of the next section says that applying our testing procedure is in many cases conservative, that is, the level of the test is controlled.

The Theorem is divided into two parts and remarks.

1. The first part provides asymptotic results for perturbations of order 1.

2. The second part provides the main result of this thesis. It gives a link between the largest/smallest residual spike and the largest/smallest eigenvalue of a small random matrix of size $k \times k$.

3. Most of the remarks extend the main result to particular cases. One of the remarks provides an algorithm to generate the distribution of the residual spikes when n_X or n_Y are small.
   The most important remark discusses the assumptions needed for asymptotic normality. Indeed, normality is obtained under Assumption 3.1.1 defined below and we will discuss what the result would be without this assumption. The results using this assumption appear in red in the theorem.

In order to prove normality, we need an assumption.

Assumption 3.1.1.
Let W_X and W_Y respect Assumption 2.2.1 such that for $s, t = 1, 2, ..., k$

$$W_{X,s,t}, W_{Y,s,t}, \left(W_X{}^2\right)_{s,t}, \left(W_Y{}^2\right)_{s,t} \text{ and } \frac{1}{m}\sum_{i=k+1}^{m} W_{X,s,i} W_{Y,t,i}$$

are asymptotically jointly normal.

Remark 3.1.1.

1. This is most likely true as a consequence of Assumption 2.2.1.

2. This assumption leads to normality and independence in the Main Theorem and its impact is highlighted in red.

Theorem 3.1.1.
Suppose $W_X, W_Y \in \mathbb{R}^{m\times m}$ respect Assumptions 2.2.1 and 3.1.1 and for $i = 1, 2, ..., k$, θ_i respects Assumptions 2.2.2 (A3) and (A1).

1. *Let $\tilde{P}_i = \mathrm{I}_m + (\theta_i - 1)e_i e_i^t \in \mathbb{R}^{m\times m}$ and define*

$$\hat{\Sigma}_{X,\tilde{P}_i} = \tilde{P}_i^{1/2} W_X \tilde{P}_i^{1/2} \text{ and } \hat{\Sigma}_{Y,\tilde{P}_i} = \tilde{P}_i^{1/2} W_Y \tilde{P}_i^{1/2}.$$

The induced filtered estimators become

$$\hat{\hat{\Sigma}}_{X,\tilde{P}_i} = \mathrm{I}_m + (\hat{\hat{\theta}}_{\hat{\Sigma}_{X,\tilde{P}_i}} - 1)\hat{u}_{\hat{\Sigma}_{X,\tilde{P}_i}}\hat{u}^t_{\hat{\Sigma}_{X,\tilde{P}_i}} \text{ and } \hat{\hat{\Sigma}}_{Y,\tilde{P}_i} = \mathrm{I}_m + (\hat{\hat{\theta}}_{\hat{\Sigma}_{Y,\tilde{P}_i}} - 1)\hat{u}_{\hat{\Sigma}_{Y,\tilde{P}_i}}\hat{u}^t_{\hat{\Sigma}_{Y,\tilde{P}_i}},$$

where $\hat{\hat{\theta}}_{\hat{\Sigma}_{X,\tilde{P}_i}}$ and $\hat{\hat{\theta}}_{\hat{\Sigma}_{Y,\tilde{P}_i}}$ are the unbiased estimators of the largest eigenvalues of $\hat{\Sigma}_{X,\tilde{P}_i}$ and $\hat{\Sigma}_{Y,\tilde{P}_i}$ respectively, as defined in Definition 2.2.1.

Then, conditioning on the spectra of W_X and W_Y,

$$\sqrt{m}\frac{\left(\lambda_{\max}\left(\hat{\hat{\Sigma}}^{-1/2}_{X,\tilde{P}_i}\hat{\hat{\Sigma}}_{Y,\tilde{P}_i}\hat{\hat{\Sigma}}^{-1/2}_{X,\tilde{P}_i}\right) - \lambda^+\right)}{\sigma^+} \sim \mathbf{N}(0,1) + o_{p;m}(1),$$

where

$$\lambda^+ = \sqrt{M_2^2 - 1} + M_2,$$

$$\begin{aligned}
\sigma^{+2} = {} & \frac{1}{(M_{2,X} + M_{2,Y} - 2)(M_{2,X} + M_{2,Y} + 2)} \\
& \Big(9M_{2,X}^4 M_{2,Y} + 4M_{2,X}^3 M_{2,Y}^2 + 4M_{2,X}^3 M_{2,Y} + 2M_{2,X}^3 M_{3,Y} - 2M_{2,X}^2 M_{2,Y}^3 \\
& +4M_{2,X}^2 M_{2,Y}^2 - 11M_{2,X}^2 M_{2,Y} - 8M_{3,X} M_{2,X}^2 M_{2,Y} + 2M_{2,X}^2 M_{2,Y} M_{3,Y} \\
& -2M_{2,X}^2 M_{3,Y} + M_{2,X}^2 M_{4,Y} + 4M_{2,X} M_{2,Y}^3 + M_{2,X} M_{2,Y}^2 + 4M_{2,X} M_{2,Y} \\
& -4M_{3,X} M_{2,X} M_{2,Y}^2 - 4M_{3,X} M_{2,X} M_{2,Y} - 2M_{2,X} M_{2,Y}^2 M_{3,Y} - 4M_{2,X} M_{2,Y} M_{3,Y} \\
& -6M_{2,X} M_{3,Y} + 2M_{4,X} M_{2,X} M_{2,Y} + 2M_{2,X} M_{2,Y} M_{4,Y} - 2M_{3,X} M_{2,Y}^2 \\
& +2M_{3,X} M_{2,Y} + M_{4,X} M_{2,Y}^2 + 4M_{2,X}^5 + 2M_{2,X}^4 - 4M_{3,X} M_{2,X}^3 - 13M_{2,X}^3 \\
& -2M_{3,X} M_{2,X}^2 + M_{4,X} M_{2,X}^2 - 2M_{2,X}^2 + 10M_{3,X} M_{2,X} + 4M_{2,X} + 4M_{3,X} \\
& -2M_{4,X} + M_{2,Y}^5 + 2M_{2,Y}^4 - M_{2,Y}^3 - 2M_{2,Y}^2 + 4M_{2,Y} - 2M_{2,Y}^3 M_{3,Y} \\
& -2M_{2,Y}^2 M_{3,Y} + 2M_{2,Y} M_{3,Y} + 4M_{3,Y} + M_{2,Y}^2 M_{4,Y} - 2M_{4,Y} - 4\Big) \\
& + \frac{1}{\sqrt{(M_{2,X} + M_{2,Y} - 2)(M_{2,X} + M_{2,Y} + 2)}} \\
& \Big(5M_{2,X}^3 M_{2,Y} - M_{2,X}^2 M_{2,Y}^2 + 2M_{2,X}^2 M_{2,Y} + 2M_{2,X}^2 M_{3,Y} - M_{2,X} M_{2,Y}^3 \\
& +2M_{2,X} M_{2,Y}^2 - 4M_{2,X} M_{2,Y} - 4M_{3,X} M_{2,X} M_{2,Y} - 2M_{2,X} M_{3,Y} + M_{2,X} M_{4,Y} \\
& -2M_{3,X} M_{2,Y} + M_{4,X} M_{2,Y} + 4M_{2,X}^4 + 2M_{2,X}^3 - 4M_{3,X} M_{2,X}^2 - 5M_{2,X}^2 \\
& -2M_{3,X} M_{2,X} + M_{4,X} M_{2,X} + 2M_{2,X} + 2M_{3,X} + M_{2,Y}^4 + 2M_{2,Y}^3 + M_{2,Y}^2 \\
& +2M_{2,Y} - 2M_{2,Y}^2 M_{3,Y} - 2M_{2,Y} M_{3,Y} - 2M_{3,Y} + M_{2,Y} M_{4,Y}\Big),
\end{aligned}$$

$$M_{s,X} = \frac{1}{m}\sum_{i=1}^{m} \hat{\lambda}^s_{W_X,i},$$
$$M_{s,Y} = \frac{1}{m}\sum_{i=1}^{m} \hat{\lambda}^s_{W_Y,i},$$
$$M_s = \frac{M_{s,X}+M_{s,Y}}{2}.$$

Moreover,

$$\sqrt{m}\frac{\left(\lambda_{\min}\left(\hat{\hat{\Sigma}}^{-1/2}_{X,\tilde{P}_i}\hat{\hat{\Sigma}}_{Y,\tilde{P}_i}\hat{\hat{\Sigma}}^{-1/2}_{X,\tilde{P}_i}\right)-\lambda^-\right)}{\sigma^-} \sim \mathbf{N}(0,1) + o_{p;m}(1),$$

where

$$\lambda^- = -\sqrt{M_2^2-1}+M_2,$$
$$\sigma^{-2} = \left(\lambda^-\right)^4\sigma^{+2}.$$

2. *Let* $P_k = \mathrm{I}_m + \sum_{i=1}^k(\theta_i-1)e_ie_i^t\mathbb{R}^{m\times m}$ *and define*

$$\hat{\Sigma}_{X,P_k} = P_k^{1/2}W_XP_k^{1/2} \text{ and } \hat{\Sigma}_{Y,P_k} = P_k^{1/2}W_YP_k^{1/2}.$$

The induced filtered estimators become

$$\hat{\hat{\Sigma}}_{X,P_k} = \mathrm{I}_m + \sum_{i=1}^k(\hat{\hat{\theta}}_{\hat{\Sigma}_{X,P_k},i}-1)\hat{u}_{\hat{\Sigma}_{X,P_k},i}\hat{u}^t_{\hat{\Sigma}_{X,P_k},i} \text{ and } \hat{\hat{\Sigma}}_{Y,P_k} = \mathrm{I}_m + \sum_{i=1}^k(\hat{\hat{\theta}}_{\hat{\Sigma}_{Y,P_k},i}-1)\hat{u}_{\hat{\Sigma}_{Y,P_k},i}\hat{u}^t_{\hat{\Sigma}_{Y,P_k},i},$$

where $\hat{\hat{\theta}}_{\hat{\Sigma}_{X,P_k},i}$ *and* $\hat{\hat{\theta}}_{\hat{\Sigma}_{Y,P_k},i}$ *are the unbiased estimators of the* $i^{\,th}$ *largest eigenvalue of* $\hat{\Sigma}_{X,P_k}$ *and* $\hat{\Sigma}_{Y,P_k}$ *respectively.*
Then, conditioning on the spectra of W_X *and* W_Y*,*

$$\lambda_{\max}\left(\hat{\hat{\Sigma}}^{-1/2}_{X,P_k}\hat{\hat{\Sigma}}_{Y,P_k}\hat{\hat{\Sigma}}^{-1/2}_{X,P_k}\right) = \lambda_{\max}\left(H^+\right)+1+O_p\left(\frac{1}{m}\right)+O_p\left(\frac{1}{\theta\sqrt{m}}\right),$$
$$\lambda_{\min}\left(\hat{\hat{\Sigma}}^{-1/2}_{X,P_k}\hat{\hat{\Sigma}}_{Y,P_k}\hat{\hat{\Sigma}}^{-1/2}_{X,P_k}\right) = \lambda_{\max}\left(H^-\right)+1+O_p\left(\frac{1}{m}\right)+O_p\left(\frac{1}{\theta\sqrt{m}}\right),$$

where

$$H^{\pm} = \zeta^{\pm}_{\infty}\begin{pmatrix} \hat{\zeta}^{\pm}_1/\zeta^{\pm}_{\infty} & w^{\pm}_{1,2} & w^{\pm}_{1,3} & \cdots & w^{\pm}_{1,k} \\ w^{\pm}_{2,1} & \hat{\zeta}^{\pm}_2/\zeta^{\pm}_{\infty} & w^{\pm}_{2,3} & \cdots & w^{\pm}_{2,k} \\ w^{\pm}_{3,1} & w^{\pm}_{3,2} & \hat{\zeta}^{\pm}_3/\zeta^{\pm}_{\infty} & \cdots & w^{\pm}_{3,k} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ w^{\pm}_{k,1} & w^{\pm}_{k,2} & w^{\pm}_{k,3} & \cdots & \hat{\zeta}^{\pm}_k/\zeta^{\pm}_{\infty} \end{pmatrix},$$

and

$$\hat{\zeta}^+_i = \lambda_{\max}\left(\hat{\hat{\Sigma}}^{1/2}_{X,\tilde{P}_i}\hat{\hat{\Sigma}}_{Y,\tilde{P}_i}\hat{\hat{\Sigma}}^{1/2}_{X,\tilde{P}_i}\right)-1,$$
$$\hat{\zeta}^-_i = \lambda_{\min}\left(\hat{\hat{\Sigma}}^{1/2}_{X,\tilde{P}_i}\hat{\hat{\Sigma}}_{Y,\tilde{P}_i}\hat{\hat{\Sigma}}^{1/2}_{X,\tilde{P}_i}\right)-1,$$
$$\zeta^{\pm}_{\infty} = \lim_{m\to\infty}\hat{\zeta}^{\pm}_i = \lambda^{\pm}-1,$$
$$w^{\pm}_{i,j} \sim \mathbf{N}\left(0,\frac{1}{m}\frac{2(M_{2,X}-1)(M_{2,Y}-1)+B^{\pm}_X+B^{\pm}_Y}{\left((\zeta^{\pm}_{\infty}-2M_2+1)^2+2(M_2-1)\right)^2}\right)+o_p\left(\frac{1}{\sqrt{m}}\right),$$

$$B_X^+ = \left(1 - M_2 + 2M_{2,X} + \sqrt{M_2^2 - 1}\right)^2 (M_{2,X} - 1)$$
$$+2\left(-1 + M_2 - 2M_{2,x} - \sqrt{M_2^2 - 1}\right)(M_{3,X} - M_{2,X}) + (M_{4,X} - M_{2,X}^2),$$
$$B_Y^+ = \left(1 + M_2 + M_{2,Y} - M_{2,X} - \sqrt{M_2^2 - 1}\right)^2 (M_{2,Y} - 1)$$
$$+2\left(-1 - M_2 - M_{2,Y} - M_{2,X} - \sqrt{M_2^2 - 1}\right)(M_{3,Y} - M_{2,Y}) + (M_{4,Y} - M_{2,Y}^2),$$
$$B_X^- = \left(1 - M_2 + 2M_{2,X} - \sqrt{M_2^2 - 1}\right)^2 (M_{2,X} - 1)$$
$$+2\left(-1 + M_2 - 2M_{2,x} + \sqrt{M_2^2 - 1}\right)(M_{3,X} - M_{2,X}) + (M_{4,X} - M_{2,X}^2),$$
$$B_Y^- = \left(1 + M_2 + M_{2,Y} - M_{2,X} + \sqrt{M_2^2 - 1}\right)^2 (M_{2,Y} - 1)$$
$$+2\left(-1 - M_2 - M_{2,Y} + M_{2,X} - \sqrt{M_2^2 - 1}\right)(M_{3,Y} - M_{2,Y}) + (M_{4,Y} - M_{2,Y}^2).$$

The matrices H^+ and H^- are strongly correlated. However, within a matrix, all the entries are jointly independent.

Remark 3.1.2.

1. Without Assumption 3.1.1, the red parts of the theorem are weaker. The entries $w_{i,j}$ have the same two first moments but not necessarily the asymptotic normal distribution. Then, the entries of $H^\pm$ are just uncorrelated instead of being independent.

2. If the spectra follow Marcenko-Pastur distributions, then

$$c = \frac{c_X + c_Y}{2},$$
$$\lambda^+ = c + \sqrt{c(c+2)} + 1,$$
$$\sigma^{+2} = c_X^3 + c_X^2 c_Y + 3c_X^2 + 4c_X c_Y - c_X + c_Y^2 + c_Y$$
$$+\frac{\left(8c_X + 2c_X^2 + \left(c_X^3 + 5c_X^2 + c_X^2 c_Y + 4c_X c_Y + 5c_X + 3c_Y + c_Y^2\right)\sqrt{c(c+2)}\right.}{c+2},$$
$$w_{i,j}^+ \sim \mathbf{N}\left(0, \frac{\sigma_w^2}{m}\right),$$
$$\sigma_w^2 = \frac{2c_X\left(\sqrt{c(c+2)} + 2\right) + 2\boldsymbol{c}_Y\left(-\sqrt{c(c+2)} + 2\right) + c_X^2 + c_Y^2}{4c\left(-\sqrt{c(c+2)} + c + 2\right)^2}.$$

3. If c_X tends to 0, then

$$\sigma^{+2} = \Big(M_{2,Y}^5 + 2M_{2,Y}^4 - 2M_{3,Y}M_{2,Y}^3 + M_{2,Y}^3 - 4M_{3,Y}M_{2,Y}^2 + M_{4,Y}M_{2,Y}^2 + 2M_{2,Y}^2$$
$$+2M_{4,Y}M_{2,Y} + 2M_{2,Y} - 2M_{3,Y} - M_{4,Y} - 2\Big) \Big/ \Big((M_{2,Y} - 1)(M_{2,Y} + 3)\Big)$$
$$+\Big(M_{2,Y}^4 + M_{2,Y}^3 - 2M_{3,Y}M_{2,Y}^2 + 2M_{2,Y}^2 - 2M_{3,Y}M_{2,Y} + M_{4,Y}M_{2,Y}$$
$$-2M_{3,Y} + M_{4,Y}\Big) \Big/ \sqrt{(M_{2,Y} - 1)(M_{2,Y} + 3)}.$$

4. When m is not large enough, the normality assumption of $\hat{\lambda}_i$ is not respected. In this case, and in particular if k is large, it could be profitable to estimate the order 1 residual spike $(\tilde{P}_i)$ with the following algorithm.

   (a) Let $\hat{\lambda}_{W_X,i}$, $\hat{\lambda}_{W_Y,i}$ be the eigenvalues of W_X and W_Y respectively.

   (b) Generate u_x and u_y, two independent uniform unit vectors of size m.

   (c) Generate Z, a standard normal independent of u_x and u_y.

   (d) We define

$$W_{x,1,1} = \sum_{i=1}^{m} \hat{\lambda}_{W_X,i} u_{x,i}^2 \text{ and } W_{x,1,1}^2 = \sum_{i=1}^{m} \hat{\lambda}_{W_X,i} u_{x,i}^2,$$
$$W_{y,2,2} = \sum_{i=1}^{m} \hat{\lambda}_{W_Y,i} u_{y,i}^2 \text{ and } W_{y,2,2}^2 = \sum_{i=1}^{m} \hat{\lambda}_{W_Y,i} u_{y,i}^2.$$

   Assuming 2.2.1, the statistics $W_{x,1,1}$ and $W_{x,1,1}^2$ follow the distribution of the first entry of W_X and ${W_X}^2$ respectively.

   (e) Construct

$$\theta_x = \theta \ W_{x,1,1} \text{ and } \alpha_x^2 = 1 + 1/\theta - \theta/\theta_x^2 \ W_{x,1,1}^2,$$
$$\theta_y = \theta \ W_{y,2,2} \text{ and } \alpha_y^2 = 1 + 1/\theta - \theta/\theta_y^2 \ W_{y,2,2}^2,$$
$$\alpha^2 = \alpha_x^2\alpha_y^2 + 2\sqrt{\alpha_x^2\alpha_y^2} Z \frac{\sqrt{(1-\alpha_x^2)(1-\alpha_y^2)}}{m}.$$

   (f) Finally, because θ is large enough

$$\lambda_{\max}\left(\hat{\hat{\Sigma}}_{X,\tilde{P}_i}^{1/2} \hat{\hat{\Sigma}}_{Y,\tilde{P}_i} \hat{\hat{\Sigma}}_{X,\tilde{P}_i}^{1/2}\right) \sim \frac{\theta_y\left(1-\alpha^2\right) + 1 + \frac{\theta_y}{\theta_x} + \sqrt{-4\frac{\theta_y}{\theta_x} + \left(\theta_y\left(1-\alpha^2\right) + \frac{\theta_y}{\theta_x} + 1\right)^2}}{2},$$
$$\lambda_{\min}\left(\hat{\hat{\Sigma}}_{X,\tilde{P}_i}^{1/2} \hat{\hat{\Sigma}}_{Y,\tilde{P}_i} \hat{\hat{\Sigma}}_{X,\tilde{P}_i}^{1/2}\right) \sim \frac{\theta_y\left(1-\alpha^2\right) + 1 + \frac{\theta_y}{\theta_x} - \sqrt{-4\frac{\theta_y}{\theta_x} + \left(\theta_y\left(1-\alpha^2\right) + \frac{\theta_y}{\theta_x} + 1\right)^2}}{2}$$

   In practice the spectra of W_X and W_Y are not observed and will be replaced by the $m-k$ smallest eigenvalues of $\hat{\Sigma}_{X,P_k}$ and $\hat{\Sigma}_{Y,P_k}$.

5. Assuming that we would like to use Monte Carlo methods to estimate the distribution, we should first estimate the eigenvalues of the covariance matrices.
   Without the theorem, the loops of the simulation generating the residual spikes generate $O\left(m^2\right)$ elements.
   Using the theorem, the loops generate k^2 elements.
   Finally, using the previous algorithm, the loops generate $O(m)$ elements.

(Proof page 168.)

3.1.1 Discussion and simulation

The above theorem gives the limiting distribution of the test statistic of the Section 2.2.3, $V_{\max}$ and $V_{\min}$. When the perturbation is of order k, Theorem 3.1.1 reduces the size of the matrix needed for estimating eigenvalues from $m \times m$ to $k \times k$.
The theorem is asymptotic in m, θ and fixed value of c_X, c_Y, $1/c_X$, $1/c_Y$ and k finite. Moreover, the

eigenvalues of the perturbation are all distinct.
By simulation, it appears that some assumptions are actually not necessary:

- The result holds if some eigenvalues θ_i and θ_j are equal for $i \neq j$.
- The result does work if n_X and n_Y are very large and m is not so large. In this case, $1/c_X$ and $1/c_Y$ are very large. However, because m is small, the normality is not achieved and we should use the algorithm introduced in Remark 3.1.2. In particular the result holds for c_X tending to 0.
- If $k = 1$, then the result works even for relatively large c_X and c_Y.

Some other assumptions are very important:

- If c_X and c_Y are very large, then convergence to the limit is slow.
- If k is large, then the asymptotic result gives poor results and can give approximations far from the true distributions of the residual spikes even for c_X and c_Y reasonably larger than 1.

Some simulations

1. Assume $\mathbf{X} \in \mathbb{R}^{m \times n_X}$ and $\mathbf{Y} \in \mathbb{R}^{m \times n_Y}$ with $\mathbf{X} = (X_1, X_2, ..., X_{n_X})$ and $\mathbf{Y} = (Y_1, Y_2, ..., Y_{n_Y})$.

$$X_i \sim \mathbf{N}_m\left(\vec{0}, \sigma^2 \mathrm{I}_m\right) \text{ and } X_{i+1} = \rho X_i + \sqrt{1-\rho^2}\ \epsilon_{X,i+1}, \text{ with } \epsilon_{X,i+1} \overset{i.i.d}{\sim} \mathbf{N}_m\left(\vec{0}, \sigma^2 \mathrm{I}_m\right),$$
$$Y_i \sim \mathbf{N}_m\left(\vec{0}, \sigma^2 \mathrm{I}_m\right) \text{ and } Y_{i+1} = \rho Y_i + \sqrt{1-\rho^2}\ \epsilon_{Y,i+1}, \text{ with } \epsilon_{Y,i+1} \overset{i.i.d}{\sim} \mathbf{N}_m\left(\vec{0}, \sigma^2 \mathrm{I}_m\right)$$

Let $P_X = \mathrm{I}_m + \sum_{i=1}^k (\theta_{X,i} - 1) u_{X,i} u_{X,i}^t$ and $P_Y = \mathrm{I}_m + \sum_{i=1}^k (\theta_{Y,i} - 1) u_{Y,i} u_{Y,i}^t$ be two perturbations in $\mathbb{R}^{m \times m}$. Then,

$$\mathbf{X}_P = P_X^{1/2} \mathbf{X} \text{ and } \mathbf{Y}_P = P_Y^{1/2} \mathbf{Y},$$
$$\hat{\Sigma}_X = \frac{\mathbf{X}_P^t \mathbf{X}_P}{n_X} \text{ and } \hat{\Sigma}_Y = \frac{\mathbf{Y}_P^t \mathbf{Y}_P}{n_Y}.$$

As proposed in Theorem 3.1.1, we build the filtered estimators $\hat{\hat{\Sigma}}_X$ and $\hat{\hat{\Sigma}}_Y$ defined in 2.2.1 .
We assume $P_X = P_Y = P$ and that with high probability we have

$$\lambda_{\max}\left(\hat{\hat{\Sigma}}_{X,P_k}^{-1/2} \hat{\hat{\Sigma}}_{Y,P_k} \hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\right) \leqslant \lambda_{\max}\left(H^+\right) + 1,$$
$$\lambda_{\min}\left(\hat{\hat{\Sigma}}_{X,P_k}^{-1/2} \hat{\hat{\Sigma}}_{Y,P_k} \hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\right) \leqslant \lambda_{\min}\left(H^-\right) + 1.$$

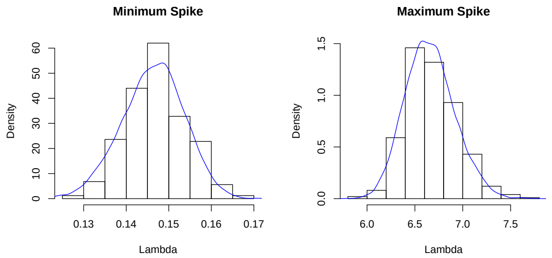
In particular, by the Main Theorem, if the eigenvalues are all large enough, then we have asymptotic equality. Later in Section 4.2, we will propose a criterion to ensure that if the eigenvalues are small, we obtain strict inequalities.
The following simulations compare in different scenarios the distributions of the empirical extreme residual spikes with the estimation obtained by $H^{\pm}$ of Theorem 3.1.1.

Equality of the eigenvalues:
What happens when some eigenvalues of the perturbation are the same?
The following simulations compare our estimate using $H^{\pm}$ of Theorem 3.1.1 with 10′000 replicates with a time consuming empirical simulation using only 500 replicates.

$\rho = 0.5$	$c_X = 0.5$	$c_Y = 2$
$m = 1000$	$n_X = 2'000$	$n_Y = 500$

$k = 4,\ \vec{\theta} = (15'000, 5'000, 2'000, 500).$

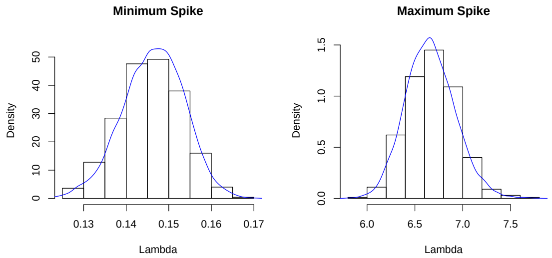
Figure 3.1 – Distributions of the residual spikes and their estimation in blue.



$\rho = 0.5$	$c_X = 0.5$	$c_Y = 2$
$m = 1000$	$n_X = 2'000$	$n_Y = 500$

$k = 4,\ \vec{\theta} = (5'000, 5'000, 5'000, 5'000).$

Figure 3.2 – Distributions of the residual spikes and their estimate in blue.



Figures 3.1 and 3.2 compare the estimated to the real distributions. They are close even when the eigenvalues are all the same in Figure 3.2.

The choice of k:
In practice, we need to determine k and this could be very difficult. In the next simulations, we choose the wrong k. Figure 3.3 shows the spectrum of $\hat{\Sigma}_Y$ of the first simulation. We clearly see four perturbations on the first plot, but it is difficult to argue the exact number of perturbations as we see in the second plot enlarging the limit zone near the bulk.

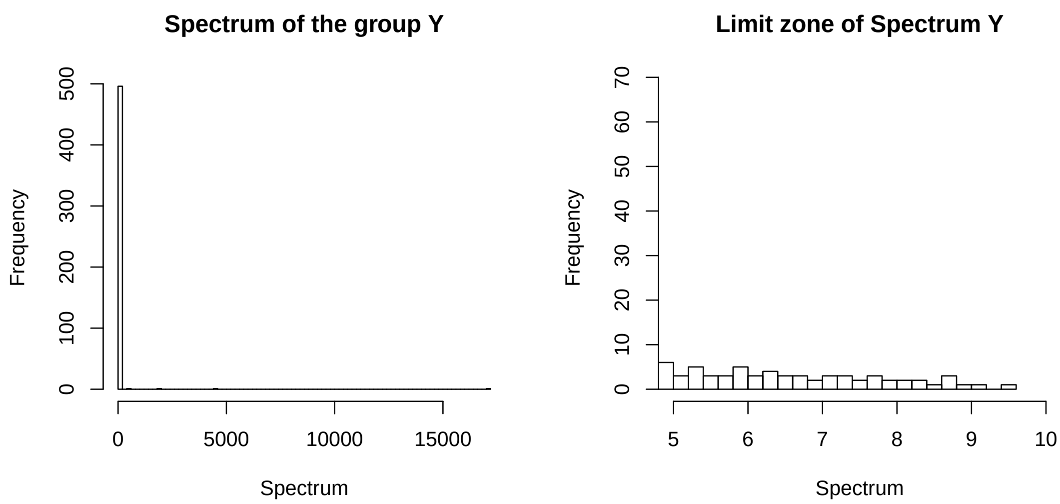
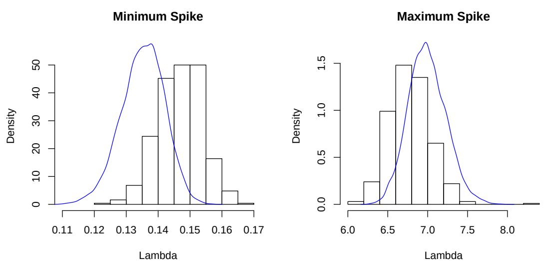


Figure 3.3 – Spectrum of $\hat{\Sigma}_Y$ and zoom on the right side of the bulk.

$\rho = 0.5$	$c_X = 0.5$	$c_Y = 2$
$m = 1000$	$n_X = 2'000$	$n_Y = 500$

$k = 4,\ \vec{\theta} = (15'000, 5'000, 2'000, 500).$
$k_{est} = 7$

Figure 3.4 – Distributions of the residual spikes and their estimation in blue.



$\rho = 0.5$	$c_X = 0.5$	$c_Y = 2$
$m = 1000$	$n_X = 2'000$	$n_Y = 500$

$$k = 4,\ \vec{\theta} = (15'000, 5'000, 2'000, 500).$$
$$k_{est} = 3$$

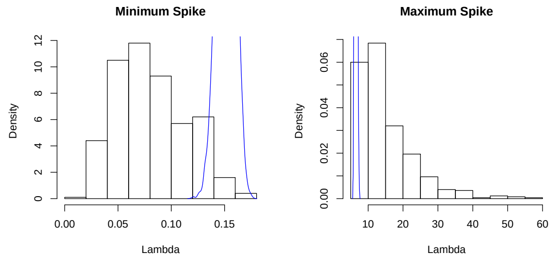


Figure 3.5 – Distributions of the residual spikes and their estimation in blue.

$\rho = 0.5$	$c_X = 0.5$	$c_Y = 2$
$m = 1000$	$n_X = 2'000$	$n_Y = 500$

$$k = 4,\ \vec{\theta} = (15'000, 3'000, 8, 6).$$
$$k_{est} = 2$$

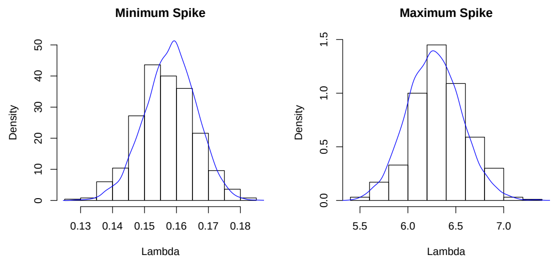


Figure 3.6 – Distributions of the residual spikes and their estimation in blue.

Figure 3.4 overestimates the number of perturbations. In this case, although the convergence does not work, the inequality remains valid, as overestimating the number of perturbations leads to a conservative test.
What happens if we underestimate the number of perturbations? Figure 3.5 neglects a large perturbation with $\theta_4 = 500$. This perturbation is large and in this case, the inequality does not work. This leads to a non-conservative test!
However, as we see in Figure 3.6, if we neglect small perturbations $\theta_3 = 8$, $\theta_4 = 6$, then the convergence in distribution remains true.
We can argue that a wise selection of k will not neglect large perturbations. Therefore, assuming this, our method is resistant to mistakes in the estimation of k due to small perturbations.

Fluctuation of k:
The next simulations present the robustness of the result as a function of k.

$\rho = 0.5$	$c_X = 0.5$	$c_Y = 2$
$m = 1000$	$n_X = 2'000$	$n_Y = 500$

$$k = 1,\ \vec{\theta} = (5'000).$$

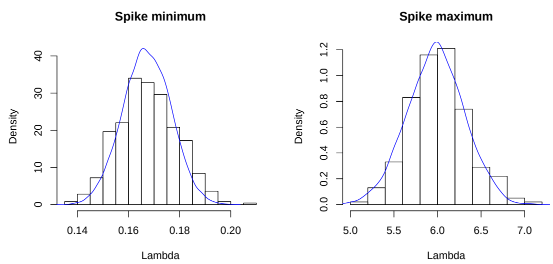


Figure 3.7 – Distributions of the residual spikes and their estimation in blue.

$\rho = 0.5$	$c_X = 0.5$	$c_Y = 2$
$m = 1000$	$n_X = 2'000$	$n_Y = 500$

$$k = 8,\ \vec{\theta} = (5'000, 5'000, ..., 5'000).$$

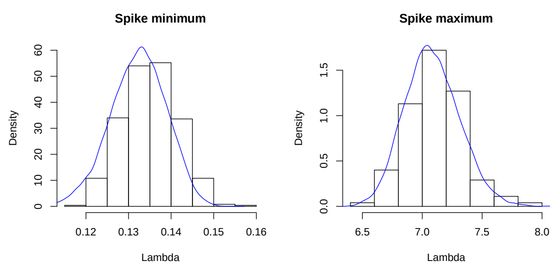


Figure 3.8 – Distributions of the residual spikes and their estimation in blue.

$\rho = 0.5$	$c_X = 0.5$	$c_Y = 2$
$m = 1000$	$n_X = 2'000$	$n_Y = 500$

$k = 15,\ \vec{\theta} = (5'000, 5'000, ..., 5'000)$.

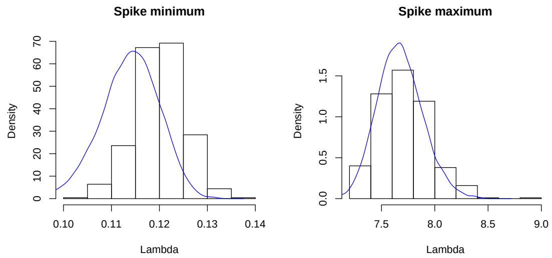


Figure 3.9 – Distributions of the residual spikes and their estimation in blue.

The Main Theorem creates a natural test tending to be more conservative when k grows. When $k = 1$ the convergence is very good as we see in Figure 3.7. Moreover, Figure 3.8 shows that when $k = 8$ our estimation is still good despite a small conservative bias in the distribution of the minimum. However, when $k = 15$ in Figure 3.9, the distributions are quite different. Luckily, the procedure remains conservative.
The lack of precision is due to k. Indeed in all our results we always neglect errors of size $O_p(k/n_Y)$ and these quantities are not necessarily small when n_Y is not large enough.

Fluctuation of c_X and c_Y:
In the following simulations we change c_X and c_Y.

$\rho = 0.5$	$c_X = 0.1$	$c_Y = 0.2$
$m = 100$	$n_X = 1'000$	$n_Y = 500$

$k = 4,\ \vec{\theta} = (15'000, 5'000, 2'000, 500)$.

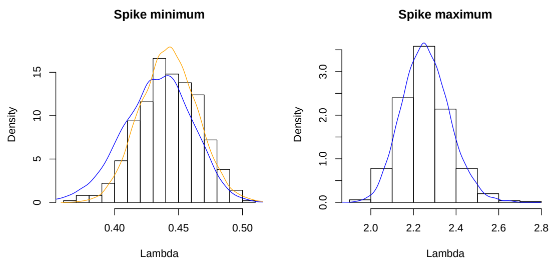


Figure 3.10 – Distributions of the residual spikes and their estimations in blue and orange (Remark 3.1.3).

$\rho = 0.5$	$c_X = 0.25$	$c_Y = 5$
$m = 500$	$n_X = 2'000$	$n_Y = 100$

$k = 4,\ \vec{\theta} = (15'000, 5'000, 2'000, 500)$.

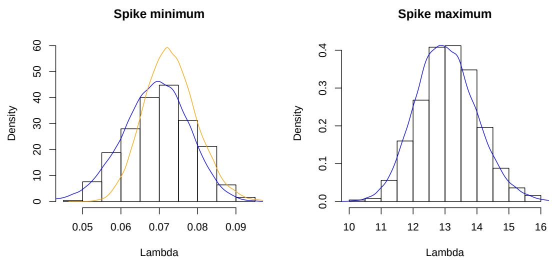


Figure 3.11 – Distributions of the residual spikes and their estimations in blue and orange (Remark 3.1.3).

$\rho = 0.5$	$c_X = 0.2$	$c_Y = 0.1$
$m = 1000$	$n_X = 200$	$n_Y = 100$

$k = 4,\ \vec{\theta} = (15'000, 5'000, 2'000, 500)$.

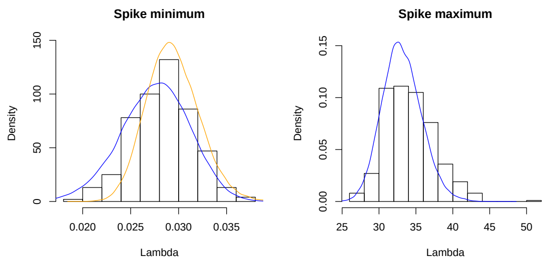


Figure 3.12 – Distributions of the residual spikes and their estimations in blue and orange (Remark 3.1.3).

$\rho = 0.5$	$c_X = 0.2$	$c_Y = 0.1$
$m = 3000$	$n_X = 600$	$n_Y = 300$

$k = 4,\ \vec{\theta} = (15'000, 5'000, 2'000, 500)$.

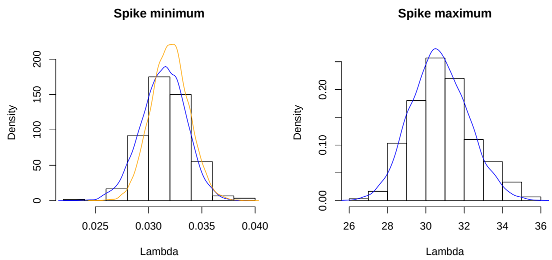


Figure 3.13 – Distribution of the residual spikes and their estimations in blue and orange (Remark 3.1.3).

The minimum spike seems to be well estimated in the different regimes in Figure 3.11, 3.12 and 3.13. However, we underestimate the true distribution in Figure 3.10, where n_X and n_Y are larger than m. Therefore, our procedure gives a conservative test.
On the other hand, the maximum spike is well estimated in Figure 3.10 and 3.11. When n_X and n_Y are small, Figure 3.12 underestimates the true distribution. This error would lead to non-conservative test. Nevertheless, the asymptotic theory is confirmed by Figure 3.12 keeping the same ratio c_X and c_Y and a larger sample.

Remark 3.1.3.
The Figures 3.10, 3.11, 3.12 and 3.13 present an orange curve that estimates the minimum residual spike quite well. There are two ways to estimate the minimum residual spike.

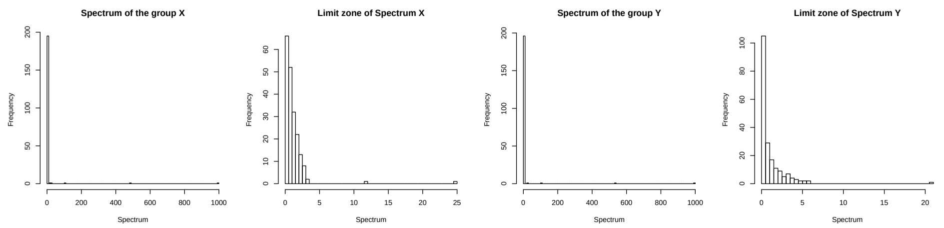
(a) First, we estimate the minimum eigenvalue of $\hat{\hat{\Sigma}}_X^{1/2}\hat{\hat{\Sigma}}_Y\hat{\hat{\Sigma}}_X^{1/2}$ using the Main Theorem 3.1.1. This leads to the blue curve.

(b) The second way estimates and inverts the largest eigenvalue of $\hat{\hat{\Sigma}}_Y^{1/2}\hat{\hat{\Sigma}}_X\hat{\hat{\Sigma}}_Y^{1/2}$. This leads to the orange curve.

We might be surprised to see a difference between these two curves. However, when m, n_X and n_Y tend to infinity then the estimations converge to the same distribution.
Our simulations tend to show that, although the orange curve is closer to the real distribution in Figure 3.10, it tends to overestimate the minimum eigenvalue in other cases as in Figure 3.11, 3.12 and 3.13.
These imprecisions are probably mainly due to the inversion of one of the matrices. Therefore,

We highly recommend choosing $\mathbf{n_X > n_Y}$!

2. In the previous simulations, we compare our estimates to an empirical distribution computed easily by Monte Carlo methods. However, the complexity of the algorithm is equal to the number of loops times m^2. This approach tends to be too long in high dimensions. Therefore, it becomes necessary to use the dimension reduction of Theorem 3.1.1.
The next simulation introduces a concrete case. We compute the singular value of two data matrices $\mathbf{X}$ and $\mathbf{Y}$. Looking at the spectra on Figure 3.14, we sensibly choose $k_{est} = 5$. A procedure to choose k is proposed in Section 4.4. Then, we compute the largest observed residual spikes, $\hat{\lambda}_{\max,obs}$ and $\hat{\lambda}_{\min,obs}$.



$\rho = 0.5$	$c_X = 1/3$	$c_Y = 4$
$m = 20'000$	$n_X = 60'000$	$n_Y = 5'000$

$k_{est} = 5$, $\hat{\lambda}_{\max,obs} = 9.31$, $\hat{\lambda}_{\min,obs} = 0.09$.

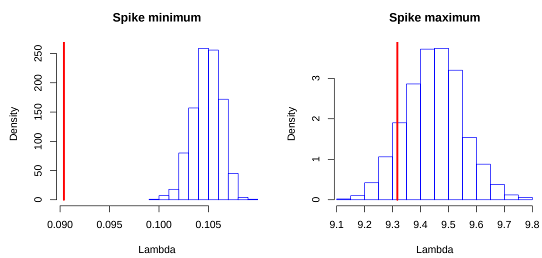


Figure 3.14 – Estimated distribution of the residual spikes and observed values in red

In Figure, 3.14 we can directly compare the observed statistics in red , $\hat{\lambda}_{\max,obs}$ and $\hat{\lambda}_{\min,obs}$ with

the estimated distribution of $\lambda_{\max}\left(\hat{\hat{\Sigma}}_X^{1/2}\hat{\hat{\Sigma}}_Y\hat{\hat{\Sigma}}_X^{1/2}\right)$ and $\lambda_{\min}\left(\hat{\hat{\Sigma}}_X^{1/2}\hat{\hat{\Sigma}}_Y\hat{\hat{\Sigma}}_X^{1/2}\right)$.
In this case, we would reject H_0 because one residual spike is too small. Therefore, the perturbations are different and the difference is in the eigenvector of $\hat{\lambda}_{\min,obs}$.

Some other simulations are presented in the Chapter 8.

Chapter 4

Data, robustness and application

The Main Theorem 3.1.1 provides the distributions of the extreme residual spikes under some conditions, which are:

- The perturbation, $P = \mathrm{I}_m + \sum_{i=1}^{m}(\theta_i - 1)u_i u_i^t$ respects Assumptions 2.2.2(A1) and (A3): The eigenvalues of the perturbation are proportional but not equal and tend to infinity faster than $\sqrt{m}$.

$$\theta_i/\sqrt{m} \to \infty,$$
$$\theta_i = p_i\theta,\ p_i \neq p_j \text{ if } i \neq j \text{ and } p_i \text{ is fixed.}$$

- The covariance matrices $\hat{\Sigma}_X$ and $\hat{\Sigma}_Y$ built from the data respect Assumption 2.2.1. If the perturbation P_k is known and cancelled, then the resulting matrices, $W_X = P^{-1/2}\hat{\Sigma}_X P^{-1/2}$ and W_Y of the two groups have independent unit invariant orthonormal eigenvectors, O_X and O_Y, and their bounded spectra, Λ_X and Λ_Y, are either fixed or independent of the eigenvectors.

$$\hat{\Sigma}_X = P_k^{1/2} W_X P_k^{1/2} \text{ and } \hat{\Sigma}_Y = P_k^{1/2} W_Y P_k^{1/2},$$

  where

$$W_X = O_X \Lambda_X O_X \text{ and } W_Y = O_Y \Lambda_Y O_Y,$$
$$P_k = \mathrm{I}_m + \sum_{i=1}^{k}(\theta_i - 1)e_i e_i^t,$$

  with k finite, $\theta_{X,i} \neq \theta_{X,j}$ if $i \neq j$ and

$$O_X, O_Y \text{ are unit orthonormal invariant and independent random matrices,}$$
$$\Lambda_X, \Lambda_Y \text{ are bounded diagonal matrices.}$$

Remark 4.0.1.
In Assumption 2.2.1, the perturbations P_X and P_Y can be different. Nevertheless, in this section, we study the behaviour of our procedure when the perturbations are the same, $P_k = P_X = P_Y$.

These assumptions allow us to prove some results and to give an explicit form to the residual spike. This chapter shows that applying our model to data that do not satisfy some of the conditions often leads to a conservative test. For example, we can assume the weaker condition 2.2.2 (A4) instead of 2.2.2(A1) and (A3). Another example relaxes the assumptions of 2.2.1 that O_X and O_Y are invariant or independent.
Therefore, we argue that despite the restrictiveness of the assumptions, our procedure is valid in broader circumstances.

Unluckily, we are not able to prove rigorously our method protects against wrong discoveries. Moreover, we can even build counterexamples where the conservativeness of the model fails. For example we can chose Λ_X and Λ_Y such that the largest residual spike occurs for small perturbations!
Despite this counterexample, studying estimated covariance matrices with our model is conservative in most practical applications. Moreover, we provide a criterion that partially checks this assumption.

This chapter is divided into three sections. First, we introduce some typical data. Then, we study the "robustness" of our procedure through some classes of models that we define later. In particular, this study introduces some criteria to check whether the largest residual spike is due to a large perturbations. Finally, the last section solves some practical difficulties and proposes tools using the residual spike to solve well-known problems.

Notation 4.0.1.
In order to distinguish different types of data derived from random matrices, we use colours:

- The data $\mathbf{X} \in \mathbb{R}^m \times \mathbb{R}^{n_X}$ such that $\mathrm{Var}(\mathbf{X}_{\cdot,i}) = P_k^{1/2}\Sigma P_k^{1/2}$ and $\mathrm{Var}(\mathbf{X}_{i,\cdot}) = \Psi$, where $\Sigma \in \mathbb{R}^m \times \mathbb{R}^m$ and $\Psi \in \mathbb{R}^{n_X} \times \mathbb{R}^{n_X}$ are covariance matrices,
- The data without the perturbation, $\mathbf{X} = P_k^{-1/2}\mathbf{X}$,
- The data without correlation between the rows (spatial), $\mathbf{X} = \Sigma^{-1/2}P_k^{-1/2}\mathbf{X}$,
- The data without any correlation (spatial-temporal) $\mathbf{X} = \Sigma^{-1/2}P_k^{-1/2}\mathbf{X}\Psi^{-1/2}$.

4.1 Data

First, we introduce data satisfying the assumptions of the Main Theorem 3.1.1. Then, we explain the difficulties that could arise in practice with spatial-temporal data. This consideration leads to more realistic models.

4.1.1 Ideal data

In order to respect the conditions of the Main Theorem 3.1.1, we assume that data $\mathbf{X} \in \mathbb{R}^{m\times n_X}$ and $\mathbf{Y} \in \mathbb{R}^{m\times n_Y}$ are such that:

$$\mathbf{X} = P_k^{1/2}\mathbf{X} \text{ and } \mathbf{Y} = P_k^{1/2}\mathbf{Y},$$

where

$$P_k = \mathrm{I}_m + \sum_{i=1}^{k}(\theta_i - 1)u_iu_i^t,$$

$$\text{with } \frac{\theta_i}{\sqrt{m}} \to \infty \text{ and for } i \neq j = 1,2,...,k,\ \exists\delta > 0 \text{ such that } \frac{\theta_i}{\theta_j} = p_{i,j} \notin [1-\delta, 1+\delta],$$

$\mathbf{X}$ and $\mathbf{Y}$ are independent and invariant by rotation with bounded spectra.

This section introduces data that create random matrices respecting Assumptions 2.2.1 and 2.2.2(A1),(A3).

Remark 4.1.1.

1. Because $\frac{\theta_i}{\sqrt{m}} \to \infty$, it is not necessary to assume bounded spectra.
2. Usually the literature assumes independent columns of $\mathbf{X}$. In this sense, our model is more general. Benaych-Georges et al. [2010] prove some results under similar conditions. However, many existing results treat $\mathbf{X}$ with independent entries. In this sense, our result is less general.
   Note that comparing assumptions with equivalent results would be fairer. For example,

we can probably show that the residual spikes converge in probability to some value under weaker conditions.

4.1.2 Spatial-temporal correlation and fluctuation

In order to explain the particularities of our model, we need to explain the two types of correlations or fluctuations that we can add and are likely to meet in practice.
Suppose our matrix of data ${\color{blue}\mathbf{X}}$ is of size $m \times n_X$. The spatial size is m and represents the features. The temporal size is n_X and represents the number of patients or replicates of the spatial vectors. Ideally, the data are a finite perturbation of independent spatial vectors with independent entries:

$$ {\color{blue}\mathbf{X}} = P_k^{1/2}\mathbf{X}, $$

where $\mathbf{X} \in \mathbb{R}^m \times \mathbb{R}^{n_X}$ has independent entries (often Gaussian). The following schema explains this simple situation.

$$
\begin{array}{cc}
 & \xrightarrow{\text{Independence}} \\
P_k \Big\downarrow & \begin{pmatrix} {\color{blue}\mathbf{X}}_{1,1} & {\color{blue}\mathbf{X}}_{1,2} & \cdots & {\color{blue}\mathbf{X}}_{1,n_X} \\ {\color{blue}\mathbf{X}}_{2,1} & {\color{blue}\mathbf{X}}_{2,2} & \cdots & {\color{blue}\mathbf{X}}_{2,n_X} \\ \vdots & \vdots & \ddots & \vdots \\ {\color{blue}\mathbf{X}}_{m,1} & {\color{blue}\mathbf{X}}_{m,2} & \cdots & {\color{blue}\mathbf{X}}_{m,n_X} \end{pmatrix}
\end{array}
\qquad
\begin{array}{cc}
 & \xrightarrow{\text{Independence}} \\
\text{Independence} \Big\downarrow & \begin{pmatrix} \mathbf{X}_{1,1} & \mathbf{X}_{1,2} & \cdots & \mathbf{X}_{1,n_X} \\ \mathbf{X}_{2,1} & \mathbf{X}_{2,2} & \cdots & \mathbf{X}_{2,n_X} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}_{m,1} & \mathbf{X}_{m,2} & \cdots & \mathbf{X}_{m,n_X} \end{pmatrix}
\end{array}
$$

However, in practice we always meet more realistic data :

- The correlation of the spatial vectors is usually more complex than a finite perturbation.
- The distribution can fluctuate in time, with different fluctuations in each row.
- The temporal vectors can be correlated.

$$
\begin{array}{cc}
 & \xrightarrow{\text{Fluctuation of the distribution + Correlation}} \\
\text{Correlation } \Sigma \text{ and } P_k \Big\downarrow & \begin{pmatrix} {\color{blue}\mathbf{X}}_{1,1} & {\color{blue}\mathbf{X}}_{1,2} & \cdots & {\color{blue}\mathbf{X}}_{1,n_X} \\ {\color{blue}\mathbf{X}}_{2,1} & {\color{blue}\mathbf{X}}_{2,2} & \cdots & {\color{blue}\mathbf{X}}_{2,n_X} \\ \vdots & \vdots & \ddots & \vdots \\ {\color{blue}\mathbf{X}}_{m,1} & {\color{blue}\mathbf{X}}_{m,2} & \cdots & {\color{blue}\mathbf{X}}_{m,n_X} \end{pmatrix}
\end{array}
$$

It is clear that without any additional assumptions on this second matrix, it is not possible to prove interesting results. Our goal is to add the smallest number of conditions possible and to argue that they make sense in practice.

4.1.3 Introduction of a particular model

We present a practical case involving the comparison between two random matrices in brain imaging: A study investigates the behaviour of $n = n_X + n_Y$ brains of patients, where each brain is represented as a spatial vector of size m. The first group contains n_X healthy patients and the second one contains n_Y patients suffering from a disorder.
It is expected that the brains behave differently in the two groups. What could be a good model for such data?

Model 1: The first model is based on one image per patient and independence between the patients. However, the vectors derived from image contain strongly correlated entries and the structure of correlation can fluctuate from patient to patient. We define $\mathbf{X} \in \mathbb{R}^{m \times n_X}$ such that

$$\begin{aligned}
\mathbf{X} &= P^{1/2}\mathbf{X}, \\
\mathbf{X} &= \Sigma^{1/2}\mathbf{X}\Psi_X^{1/2}, \\
&\quad \Psi_{X,i,i} \overset{i.i.d.}{\sim} \mathrm{RV}\left(1, \sigma_\Psi^2\right), \; \Psi_{X,i,j} = 0,
\end{aligned}$$

where $\mathbf{X}$ is a matrix of independent normal random variables. With this model the image of the i^{th} patient, $\mathbf{X}_{\cdot,i}$ is

$$\begin{aligned}
\mathbf{X}_{\cdot,i} &= P^{1/2}\mathbf{X}_{\cdot,i}, \\
\mathbf{X}_{\cdot,i} &= \Sigma^{1/2}\mathbf{X}_{\cdot,i}\Psi_{X,i,i}^{1/2}, \\
&\quad \mathbf{X}_{\cdot,i}|\Psi_{X,i,i} \overset{i.i.d}{\sim} \mathbf{N}\left(0, \Psi_{X,i,i}\Sigma\right).
\end{aligned}$$

Model 2: The brain images are usually repeated for each patient, in order to increase precision, that is, each patient provides many correlated images in function of the time. In this case, we need to extend model 1 to $\mathbf{X} \in \mathbb{R}^{m \times n_X}$ such that,

$$\begin{aligned}
\mathbf{X} &= P^{1/2}\mathbf{X}, \\
\mathbf{X} &= \Sigma^{1/2}\mathbf{X}\tilde{\Psi}_X^{1/2}, \\
&\quad \tilde{\Psi}_X = \Psi^{1/2}\mathrm{R}\Psi^{1/2}, \\
&\quad\quad \Psi_{X,i,i} \sim \mathrm{RV}\left(1, \sigma_\Psi^2\right), \; \Psi_{X,i,j} = 0,
\end{aligned}$$

where R is a correlation matrix and $\mathbf{X}$ is a matrix of independent normal random variables.

General model: The most general model that we hope to evaluate with our test is

$$\begin{aligned}
\mathbf{X} &= P^{1/2}\mathbf{X}, \text{ where } P = \mathrm{I}_m + \sum_{i=1}^{k}(\theta_i - 1)u_i u_i^t, \\
\mathbf{X} &= \left(\mathbf{X}_{\cdot,1}, \mathbf{X}_{\cdot,2}, ..., \mathbf{X}_{\cdot,n_X}\right), \\
&\quad \mathbf{X}_{\cdot,1}, \mathbf{X}_{\cdot,2}, ..., \mathbf{X}_{\cdot,n_X} \sim \mathrm{RV}\left(0, \Sigma\right) \text{ (not necessarily the same distribution.)} \\
&\quad \mathbf{X}_{1,\cdot}, \mathbf{X}_{2,\cdot}, ..., \mathbf{X}_{m,\cdot} \text{ are dependent random vectors. (temporal correlation.)}
\end{aligned}$$

General fluctuations can modify the images through time, dependence can occur between the images, distribution of the entries of $\mathbf{X}$ are free, and spatial dependence (more complex than a covariance) can exist.
This model would be very general; however, we need to assume some hypotheses to extend existing theory.
Comparison with our model: Our model assumes invariance by rotation of $\mathbf{X}$. Therefore, we can only treat the model 2 for $\Sigma = \mathrm{I}_m$ without any condition on $\tilde{\Psi}_X$. In Section 4.4, we see that temporal finite perturbations are difficult to cancel; therefore, we should assume that the spectrum of $\tilde{\Psi}_X$ is compact enough in the sense that it does not create isolated estimated eigenvalues outside from the spectrum.

What happens if we use the methods derived from the simple model in situations where the true model is more general?

Remark 4.1.2.
In practice, we will observe a spectrum, which most probably will be deformed.
The general model assumes that the deformation is due to complex behaviour of the random

matrix such as:

- Spatial dependence, temporal fluctuations of the density and temporal correlations possibly different in each row.

On the other hand, our model assumes that all the deformation is due to:

- Spatially uncorrelated dependence due to the rotation invariance, and identical temporal correlations and fluctuations of the density in the rows.

$$\begin{array}{cc} & \xrightarrow{\text{Fluctuation of the distribution + Correlation}} \\ P_k \Big\downarrow & \begin{pmatrix} \mathbf{X}_{1,1} & \mathbf{X}_{1,2} & \cdots & \mathbf{X}_{1,n_X} \\ \mathbf{X}_{2,1} & \mathbf{X}_{2,2} & \cdots & \mathbf{X}_{2,n_X} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}_{m,1} & \mathbf{X}_{m,2} & \cdots & \mathbf{X}_{m,n_X} \end{pmatrix} \end{array} \qquad \begin{array}{cc} & \xrightarrow{\text{Fluctuation of the distribution + Correlation}} \\ \text{Invariant by rotation} \Big\downarrow & \begin{pmatrix} \mathbf{X}_{1,1} & \mathbf{X}_{1,2} & \cdots & \mathbf{X}_{1,n_X} \\ \mathbf{X}_{2,1} & \mathbf{X}_{2,2} & \cdots & \mathbf{X}_{2,n_X} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}_{m,1} & \mathbf{X}_{m,2} & \cdots & \mathbf{X}_{m,n_X} \end{pmatrix} \end{array}$$

Our model seems to be too restrictive to treat such data. Nevertheless, the next section will show that it is robust to a range of violations of the assumptions. For example we show that the location of the residual spike of our model is robust to the presence of Σ when θ is large and $k = 1$.

Example 4.1.2.1.
If $\mathbf{X}_{\cdot,1}, ..., \mathbf{X}_{\cdot,n_X}, \mathbf{Y}_{\cdot,1}, ..., \mathbf{Y}_{\cdot,n_Y} \overset{i.i.d.}{\sim} \mathbf{N}(0,\Sigma)$ or if $\mathbf{X}_{\cdot,1}, ..., \mathbf{X}_{\cdot,n_X}, \mathbf{Y}_{\cdot,1}, ..., \mathbf{Y}_{\cdot,n_Y} \overset{i.i.d.}{\sim} \mathbf{N}(0,\mathrm{I}_m)$, then the expectation of the resulting residual spike when $k = 1$ is asymptotically (in m and θ) the same under weak conditions on Σ.

4.2 Robustness

This section investigates the robustness of our model under perturbations of order 1 for four classes of models. First, we define what kind of robustness we hope to achieve. Then, we introduce the four classes. Each classes is, then, investigated by means of a theorem of robustness. The next section discusses the remaining ambiguities of our investigation and the difficulties to prove perfect robustness. Then, all the results are summarised in a table. Finally a small section argues for a stronger claim of robustness of our procedure based on simulations.

4.2.1 Definition of robustness

Robustness of a model means that its results will remain accurate under some modifications. In this thesis, we argue that our model is robust because it is still conservative under some perturbations of the model.
Before defining the robustness, we define the **moments in probability**.

Definition 4.2.1.
If X_n is such that

$$\frac{X_n - \mu_n}{\sigma_n} \xrightarrow{D} \mathrm{RV}\left(0,1\right),$$

with $\sigma_n^2 \overset{n\to\infty}{\longrightarrow} 0$, then

$$\begin{aligned} \mathrm{E}_{p(n)}\left[X_n\right] &= \mu_n + o_n(\mu_n\sigma_n), \\ \mathrm{Var}_{p(n)}\left[X_n\right] &= \sigma_n^2 + o_n(\sigma_n^2) \end{aligned}$$

are called the **expectation and the variance in probability**.

Remark 4.2.1.1.

1. In particular if $X_n \sim \mathbf{N}\left(\mu, \frac{\sigma^2}{n}\right) + o_p\left(\frac{1}{\sqrt{n}}\right)$, then $\mathrm{E}_{p(n)}\left[X_n\right] = \mu$ and $\mathrm{Var}_{p(n)}\left[X_n\right] = \frac{\sigma^2}{n}$.

2. When the errors o_p do not affect the two first moments, i.e.,
$$\mathrm{E}\left[Z_n + o_p\left(\frac{1}{\sqrt{n}}\right)\right] = \mathrm{E}\left[Z_n\right] + o_n\left(\frac{1}{\sqrt{n}}\right) \text{ and } \mathrm{Var}\left(Z_n + o_p\left(\frac{1}{\sqrt{n}}\right)\right) = \mathrm{Var}\left[Z_n\right] + o_n\left(\frac{1}{n}\right),$$
then the moments in probability coincides with the usual two first moments.

3. If $X_n \sim \mathbf{N}\left(2, \frac{1}{n}\right) + o_p\left(\frac{1}{n}\right)$ and $\tilde{X}_n \sim \mathbf{N}\left(1, \frac{1}{n}\right) + o_p\left(\frac{1}{n}\right)$, then
$$\mathrm{E}_{p(n)}[X_n] = 2 > 1 = \mathrm{E}_{p(n)}[\tilde{X}_n] \text{ and } \mathrm{Var}_{p(n)}\left(X_n\right) = \mathrm{Var}_{p(n)}\left(\tilde{X}_n\right) = \frac{1}{n}$$
but we could have
$$\mathrm{E}[X_n] < \mathrm{E}[\tilde{X}_n].$$

4. We could relax the condition $\sigma_n \to 0$. However, for the purpose of this section it is convenient and not restrictive.

Next we define two sorts of robustness, the **strong robustness** and the **weak robustness**.

Definition 4.2.2.
Suppose that a procedure using n replicates leads to a test statistic T_n and a test will reject its hypothesis if the random variable T_n exceeds a bound.

- Assuming a null model, the random statistic is $T_{0,n}$.
- Assuming a model $\mathcal{C}_i$ in $\mathcal{C}$ a class of models, the statistic is $T_{\mathcal{C}_i,n}$.

1. A null model is **asymptotically strongly robust on the right** in a class $\mathcal{C}$ of models if:
For all models $\mathcal{C}_i$ in $\mathcal{C}$, there exists ϵ_n such that
$$\mathrm{E}_{p(n)}\left[T_{\mathcal{C}_i,n}\right] \leqslant \mathrm{E}_{p(n)}\left[T_{0,n}\right] + \epsilon_n,$$
$$\lim_{n\to\infty} \frac{\epsilon_n^2}{\mathrm{Var}_{p(n)}\left(T_{\mathcal{C}_i,n}\right)} = \lim_{n\to\infty} \frac{\epsilon_n^2}{\mathrm{Var}_{p(n)}\left(T_{0,n}\right)} = 0.$$
Moreover,
$$\text{if } \exists\, \tilde{\epsilon}_n \text{ such that } \mathrm{E}_{p(n)}\left[T_{\mathcal{C}_i,n}\right] = \mathrm{E}_{p(n)}\left[T_{0,n}\right] + \tilde{\epsilon}_n \text{ and}$$
$$\max\left(\frac{\tilde{\epsilon}_n^2}{\mathrm{Var}_{p(n)}\left(T_{\mathcal{C}_i,n}\right)}, \frac{\tilde{\epsilon}_n^2}{\mathrm{Var}_{p(n)}\left(T_{0,n}\right)}\right) \underset{n\to\infty}{\not\to} \infty,$$
$$\text{then } \frac{\mathrm{Var}_{p(n)}\left(T_{\mathcal{C}_i,n}\right)}{\mathrm{Var}_{p(n)}\left(T_{0,n}\right)} = 1 + o_n(1).$$

2. A null model is **asymptotically weakly robust on the right** in a class $\mathcal{C}$ of models if
For all models $\mathcal{C}_i$ in $\mathcal{C}$, there exists ϵ_n with $\lim_{n\to\infty} \epsilon_n = 0$ such that
$$\mathrm{E}_{p(n)}\left[T_{\mathcal{C}_i,n}\right] \leqslant \mathrm{E}_{p(n)}\left[T_{0,n}\right] + \epsilon_n.$$

3. The definition of **asymptotically strongly robust on the left** and **asymptotically weakly robust on the left** are obtained by substituting
$$\mathrm{E}_{p(n)}\left[T_{\mathcal{C}_i,n}\right] \leqslant \mathrm{E}_{p(n)}\left[T_{0,n}\right] + \epsilon_n$$
in the previous definitions by
$$\mathrm{E}_{p(n)}\left[T_{\mathcal{C}_i,n}\right] \geqslant \mathrm{E}_{p(n)}\left[T_{0,n}\right] + \epsilon_n.$$

4.2.2 Introduction to different classes

Our model is confronted with four classes:

- Class $\mathcal{C}_A$ is very general and extends our model to very general spatial structures, Σ, possibly linked to the direction of the perturbation, u. Moreover, the temporal fluctuations and correlations in the direction u can be a little different from the other directions. The main restriction of this class is the asymptotic assumption that θ, the eigenvalue of the perturbation, tends to infinity.

- Class $\mathcal{C}_B$ generalises our model to perturbations with any θ.

- Class $\mathcal{C}_C$ assumes that $\mathbf{X} = P^{-1/2}\mathbf{X}$ have i.i.d. symmetric entries. Moreover, no assumption is made on θ.

- Class $\mathcal{C}_D$ makes no assumption on θ or Σ. However, the perturbation is invariant by rotation.

These classes are inspired by the previous Section 4.1. Naturally, they do not cover all the possible realities; however, controlling these classes is a good first step to validate our procedure.
Recall that our model assumes that $\mathbf{X}$ and $\mathbf{Y}$ are two independent identical random matrices invariant by rotation of size $m \times n_X$ and $m \times n_Y$ respectively. Then, we assume our data, $\mathbf{X}$ and $\mathbf{Y}$, to be such that,

$$\mathbf{X} = P^{1/2}\mathbf{X} \text{ and } \mathbf{Y} = P^{1/2}\mathbf{Y}, \text{ where } P = \mathrm{I}_m + (\theta - 1)uu^t \text{ and } \frac{\theta}{\sqrt{m}} \to \infty.$$

The residual spike obtained from this model observing m features is $\lambda_{0,m}$ such that,

$$\sqrt{m}\frac{\lambda_{0,m} - \mathrm{E}_{p(m)}\left[\lambda_{0,m}\right]}{\sqrt{\mathrm{Var}_{p(m)}\left(\lambda_{0,m}\right)}} \sim \mathbf{N}\left(0,1\right) + o_{p;m}(1).$$

In the Main result we showed that $\mu_{\lambda_0} = \mu_{\lambda_0}\left(M_2\right)$, where $M_2 = M_{2,X} + M_{2,Y}$ and $M_{2,X}, M_{2,Y}$ are the second moments of the spectra of $\mathbf{X}$ and $\mathbf{Y}$.

Class A

In this part, we define the class $\mathcal{C}_A$ and introduce a theorem proving the robustness of our model in this class.

Definition 4.2.3.
For $\Sigma \in \mathbb{R}^{m\times m}$, we define the class $\mathcal{C}_A(\Sigma)$ of models assuming random matrices $\mathbf{X} \in \mathbb{R}^{m\times n_X}$ such that

$$\begin{aligned}
&\mathbf{X} = P^{1/2}\mathbf{X},\\
&\quad P = \mathrm{I}_m + (\theta-1)e_1e_1^t \text{ with } \theta \to \infty,\\
&\quad \mathbf{X} = \Sigma^{1/2}\mathbf{X},\\
&\qquad \mathrm{Trace}\left(\Sigma\right) = m,\ \lambda_{\max}\left(\Sigma\right) = O_p\left(1\right),\\
&\qquad \mathbf{X}_{1,\cdot} \sim \mathcal{L}_1,\ \mathbf{X}_{2,\cdot}, ..., \mathbf{X}_{m,\cdot} \sim \mathcal{L},\\
&\qquad\quad \mathrm{Var}\left(\mathbf{X}_{1,\cdot}\right) = \Psi_{1,X}, \mathrm{Var}\left(\mathbf{X}_{2,\cdot}\right) = \mathrm{Var}\left(\mathbf{X}_{3,\cdot}\right) = ... = \mathrm{Var}\left(\mathbf{X}_{m,\cdot}\right) = \Psi_X,\\
&\qquad\quad \text{where } \lambda_{\max}\left(\Psi_X\right) \text{ and } \lambda_{\max}\left(\Psi_{1,X}\right) \text{ are } O_p(1) \text{ and}\\
&\qquad\quad \Delta_X = \Psi_{1,X} - \Psi_X \text{ is such that } \mathrm{Trace}\left(\Delta_X^2\right) = o\left(n_X\right),\\
&\qquad \mathbf{X}_{\cdot,i} \text{ are vectors with i.i.d. independent entries},\\
&\quad \text{If } W_X = \frac{1}{n_X}\mathbf{X}\mathbf{X}^t,\\
&\qquad \lambda_{\max}\left(W_X\right) = O_p(1),\ \mathrm{Var}\left(W_{X,1,1}\right) = O\left(1/m\right) \text{ and } \mathrm{Var}\left(\left(W_X^2\right)_{1,1}\right) = o_m\left(1\right).\\
&\quad \text{If } W_X = \frac{1}{n_X}\mathbf{X}\mathbf{X}^t,\ \lambda_{\max}\left(W_X\right) = O_p(1).
\end{aligned}$$

Finally we assume $m/n_X = c_X$ is a fixed constant. (Or $m/n_X - c_X = o(1/m^{1/2})$).

Remark 4.2.3.1.
The condition $\Delta = \Psi_1 - \Psi$ and $\mathrm{Trace}\left(\Delta^2\right) = o\left(n_X\right)$ can be replaced by

$$\Psi_1 = \Psi + \sum_{r=1}^{R} \eta_r v_r v_r^t,$$

where $R = o(n_X)$. This means that Δ has rank $o(n_X)$ and implies the original weaker condition because

$$||\Delta||_F = \mathrm{Trace}\left(\Delta^2\right)^{1/2} \leqslant \sqrt{\mathrm{rank}\left(\Delta\right)}\lambda_{\max}\left(\Delta\right),$$

and $\lambda_{\max}\left(\Delta\right)$ is finite. ($\lambda_{\max}\left(\Psi_1\right)$ and $\lambda_{\max}\left(\Psi\right)$ are finite.)

The following theorem investigates the consequences of applying our model to data from class $\mathcal{C}_A$.

Theorem 4.2.1.
Assume $\mathbf{X} \in \mathbb{R}^{m\times n_X}$ *and* $\mathbf{Y} \in \mathbb{R}^{m\times n_Y}$ *are random matrices from class* $\mathcal{C}_A(\Sigma)$ *defined in 4.2.3. By definition of* $\mathcal{C}_A$*,*

$$\mathbf{X} = P^{1/2}\Sigma^{1/2}\mathbf{X} \text{ and } \mathbf{Y} = P^{1/2}\Sigma^{1/2}\mathbf{Y}.$$

We define

$$\tilde{\mathbf{X}} = P^{1/2}\tilde{\mathbf{X}} \text{ and } \tilde{\mathbf{Y}} = P^{1/2}\tilde{\mathbf{Y}},$$

such that $\tilde{\mathbf{X}}$ *and* $\tilde{\mathbf{Y}}$ *are invariant by rotation with spectra equal to the spectra of* $\mathbf{X}$ *and* $\mathbf{Y}$ *respectively. Moreover, we define* $\lambda_{A,m}$ *and* $\lambda_{0,m}$ *as the largest residual spikes obtained using* $\left(\mathbf{X},\mathbf{Y}\right)$ *and* $\left(\tilde{\mathbf{X}},\tilde{\mathbf{Y}}\right)$ *respectively.*
Then,

1. *The new data* $\tilde{\mathbf{X}}$ *and* $\tilde{\mathbf{Y}}$ *satisfy the conditions of our model.*

2. *Our model is asymptotically weakly robust on the right,*

$$\mathrm{E}_{p(\min(m,\theta))}\left[\lambda_{A,m}\right] \leqslant \mathrm{E}_{p(\min(m,\theta))}\left[\lambda_{0,m}\right] \text{ with equality when } \Sigma = \mathrm{I}_m.$$

3. *If* $\mathrm{Trace}\left(\Sigma\right) - 1 > d$*, where* d *is a positive constant, our model is asymptotically strongly robust on the right,*

$$\begin{gathered}\mathrm{E}_{p(\min(m,\theta))}\left[\lambda_{A,m}\right] + D(d) < \mathrm{E}_{p(\min(m,\theta))}\left[\lambda_{0,m}\right], \\ \text{where } D(d) \text{ is positive and depends on } d, \\ \lim_{m,\theta\to\infty} \mathrm{Var}_{p(\min(m,\theta))}\left(\lambda_{A,m}\right) = \lim_{m,\theta\to\infty} \mathrm{Var}_{p(\min(m,\theta))}\left(\lambda_{0,m}\right) = 0.\end{gathered}$$

Remark 4.2.1.1.

1. The theorem shows that applying the procedure based on our base model to data generated by model A leads to conservative tests.

2. This theorem can be extended to the minimum residual spike to show robustness on the left.

(Proof page 201)

Class B

In this part, we define the class $\mathcal{C}_B$ and an additional criterion. A theorem shows the robustness of our base model applied to data from $\mathcal{C}_B$ if in addition the criterion is satisfied.

Definition 4.2.4.
We define the class $\mathcal{C}_B$ of models assuming random matrices $\mathbf{X} \in \mathbb{R}^{m \times n_X}$ such that

$$\begin{aligned} &\mathbf{X} = P^{1/2}\mathbf{X}, \\ &\text{where } P = \mathrm{I}_m + (\theta - 1)uu^t, \\ &\qquad \mathbf{X} \text{ is invariant by rotation.} \end{aligned}$$

Finally we assume $m/n_X = c_X$ is a fixed constant. (Or $m/n_X - c_X = o(1/m^{1/2})$).

The models generated by class $\mathcal{C}_B$ are close to our base model but θ can be finite. Thus, we need to define a criterion to ensure that the largest residual spike increases with θ.

Definition 4.2.5.
Let $S_{\mathbf{X}}$ and $S_{\mathbf{Y}}$ be the spectra of $\frac{\mathbf{X}\mathbf{X}^t}{n_X}$ and $\frac{\mathbf{Y}\mathbf{Y}^t}{n_Y}$, where $\mathbf{X} = P_k^{1/2}\mathbf{X}$ and $\mathbf{Y} = P_k^{1/2}\mathbf{Y}$. We define the curve

$$\mu_\lambda(\theta, S_{\mathbf{X}}, S_{\mathbf{Y}}) = \frac{1}{2}\left(\theta + \alpha^2 - \theta\alpha^2 + \frac{1 + (\theta-1)\alpha^2 + \sqrt{-4\theta^2 + (1+\theta^2 - (\theta-1)^2\alpha^2)^2}}{\theta}\right), \quad (4.1)$$

where

$$\begin{aligned} \alpha = \alpha_{\mathbf{X}}\alpha_{\mathbf{Y}}, \quad \alpha_{\mathbf{X}}^2 &= \frac{m\theta}{(\theta-1)^2\hat{\theta}_{\mathbf{X}} \sum_{i=k+1}^m \frac{\hat{\lambda}_{\mathbf{X},i}}{(\hat{\theta}_{\mathbf{X}} - \hat{\lambda}_{\mathbf{X},i})^2}}, \quad \alpha_{\mathbf{Y}}^2 = \frac{m\theta}{(\theta-1)^2\hat{\theta}_{\mathbf{Y}} \sum_{i=k+1}^m \frac{\hat{\lambda}_{\mathbf{Y},i}}{(\hat{\theta}_{\mathbf{Y}} - \hat{\lambda}_{\mathbf{Y},i})^2}}, \\ \hat{\theta}_{\mathbf{X}} \,\Big|\, \frac{1}{\theta-1} &= \frac{1}{m}\sum_{i=k+1}^m \frac{\hat{\lambda}_{\mathbf{X},i}}{(\hat{\theta}_{\mathbf{X}} - \hat{\lambda}_{\mathbf{X},i})}, \quad \hat{\theta}_{\mathbf{Y}} \,\Big|\, \frac{1}{\theta-1} = \frac{1}{m}\sum_{i=k+1}^m \frac{\hat{\lambda}_{\mathbf{Y},i}}{(\hat{\theta}_{\mathbf{Y}} - \hat{\lambda}_{\mathbf{Y},i})}, \end{aligned}$$

and $\hat{\lambda}_{\mathbf{X},i} = \lambda_i\left(\frac{\mathbf{X}\mathbf{X}^t}{n_X}\right) \in S_{\mathbf{X}}$ and $\hat{\lambda}_{\mathbf{Y},i} = \lambda_i\left(\frac{\mathbf{Y}\mathbf{Y}^t}{n_Y}\right) \in S_{\mathbf{Y}}$.
The criterion is satisfied if the curve is monotone increasing.

Remark 4.2.5.1.
We can use this as a criterion to argue that the expectation of a residual spike is monotone increasing in θ. However, when θ is large compared to m, this estimator fails and we should use an asymptotic estimator of α based on:

$$\begin{aligned} \alpha_{\mathbf{X}}^2 &= 1 + \frac{1}{\theta}\left(1 - \frac{1}{m}\sum_{i=k+1}^m \hat{\lambda}_{\mathbf{X},i}^2\right) \\ &\quad + \frac{1}{\theta^2}\left(1 + \frac{2}{m}\sum_{i=k+1}^m \hat{\lambda}_{\mathbf{X},i}^2 + \frac{3}{m^2}\left(\sum_{i=k+1}^m \hat{\lambda}_{\mathbf{X},i}^2\right)^2 - \frac{2}{m}\sum_{i=k+1}^m \hat{\lambda}_{\mathbf{X},i}^3\right) + O_p\left(\frac{1}{\theta^3}\right), \\ \theta_{\mathbf{X}} &= (\theta - 1) + \frac{1}{m}\sum_{i=k+1}^m \hat{\lambda}_{\mathbf{X},i}^2 + O_p\left(\frac{1}{\theta^2}\right). \end{aligned}$$

Using this approximation, the estimated curve criterion 4.1 makes an error of $O_p\left(1/\theta^2\right)$.

Theorem 4.2.2.
Assume $\mathbf{X} \in \mathbb{R}^{m \times n_X}$ and $\mathbf{Y} \in \mathbb{R}^{m \times n_Y}$ are random matrices from class $\mathcal{C}_B$ defined in 4.2.4. By definition of $\mathcal{C}_B$,

$$\mathbf{X} = P^{1/2}\mathbf{X} \text{ and } \mathbf{Y} = P^{1/2}\mathbf{Y},$$

where $P = \mathrm{I}_m + (\theta - 1)uu^t$*. We define*

$$\tilde{\mathbf{X}} = \tilde{P}^{1/2}\mathbf{X} \text{ and } \tilde{\mathbf{Y}} = \tilde{P}^{1/2}\mathbf{Y},$$

such that $\tilde{P} = \mathrm{I}_m + (\tilde{\theta} - 1)uu^t$ *and* $\lim \frac{\tilde{\theta}}{\sqrt{m}} = \infty$*. Moreover, we define* $\lambda_{B,m}$ *and* $\lambda_{0,m}$ *as the largest residual spikes obtained using* $\left(\mathbf{X}, \mathbf{Y}\right)$ *and* $\left(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}\right)$ *respectively.*
If the criterion defined in 4.2.5 is such that

$$\forall \theta > 0,\ \mu_\lambda(\theta, S_{\mathbf{X}}, S_{\mathbf{Y}}) < \lim_{\theta \to \infty} \mu_\lambda(\theta, S_{\mathbf{X}}, S_{\mathbf{Y}}) + o(1/\sqrt{m}),$$

then

1. *If the variances in probability of* $\hat{\theta}_X$*,* $\hat{\theta}_Y$*,* $\langle \hat{u}_X, \hat{u}_Y \rangle$ *are* $O\left(1/m\right)$*, our model is asymptotically strongly robust on the right,*

$$\mathrm{E}_{p(m)}\left[\lambda_{B,m}\right] \leqslant \mathrm{E}_{p(m)}\left[\lambda_{0,m}\right].$$

   The equality occurs when

$$\lim_{m \to \infty} \frac{\mathrm{Var}_{p(m)}\left(\lambda_{B,m}\right)}{\mathrm{Var}_{p(m)}\left(\lambda_{0,m}\right)} = 1.$$

2. *If the variances in probability of* $\hat{\theta}_X$*,* $\hat{\theta}_Y$*,* $\langle \hat{u}_X, \hat{u}_Y \rangle$ *are* $o_m\left(1\right)$*, our model is asymptotically weakly robust on the right,*

$$\mathrm{E}_{p(m)}\left[\lambda_{B,m}\right] \leqslant \mathrm{E}_{p(m)}\left[\lambda_{0,m}\right].$$

Remark 4.2.2.1.

1. If the criterion is satisfied, then applying the Main Theorem to data generated by model B leads to conservative tests.
2. This theorem can be extended to the minimum residual spike to show robustness on the left.

(Proof page 202)

Class C

In this part we define the class $\mathcal{C}_C$. Then, we introduce a theorem proving the robustness of our model applied to data from $\mathcal{C}_C$ which satisfies the criterion 4.2.5.

Definition 4.2.6.
We define the class $\mathcal{C}_C$ of models assuming random matrices $\mathbf{X} \in \mathbb{R}^{m \times n_X}$ such that

$$\begin{aligned} \mathbf{X} &= P^{1/2}\mathbf{X}, \\ &P = \mathrm{I}_m + (\theta - 1)e_1 e_1^t, \\ &\mathbf{X}_{1,\cdot}, \mathbf{X}_{2,\cdot}, ..., \mathbf{X}_{m,\cdot} \sim \mathcal{L}, \\ &\mathbf{X}_{\cdot,i} \text{ are vectors with i.i.d. independent symmetric entries.} \end{aligned}$$

Finally we assume $m/n_X = c_X$ is a fixed constant. (Or $m/n_X - c_X = o(1/m^{1/2})$).

Theorem 4.2.3.
Assume $\mathbf{X} \in \mathbb{R}^{m\times n_X}$ *and* $\mathbf{Y} \in \mathbb{R}^{m\times n_Y}$ *are random matrices from class* $\mathcal{C}_C$ *defined in 4.2.6. By definition of* $\mathcal{C}_B$*,*

$$\mathbf{X} = P^{1/2}\mathbf{X} \text{ and } \mathbf{Y} = P^{1/2}\mathbf{Y},$$

where $P = \mathrm{I}_m + (\theta-1)uu^t$*. We define*

$$\tilde{\mathbf{X}} = \tilde{P}^{1/2}\tilde{\mathbf{X}} \text{ and } \tilde{\mathbf{Y}} = \tilde{P}^{1/2}\tilde{\mathbf{Y}},$$

such that $\tilde{P} = \mathrm{I}_m + (\tilde{\theta}-1)uu^t$, $\lim \frac{\tilde{\theta}}{\sqrt{m}} = \infty$ *and* $\tilde{\mathbf{X}}$ *(respectively* $\tilde{\mathbf{Y}}$*) is invariant by rotation with spectrum equal to the spectrum of* $\mathbf{X}$ *(respectively* $\mathbf{Y}$*). Moreover, we define* $\lambda_{C,m}$ *and* $\lambda_{0,m}$ *as the largest residual spikes obtained using* $(\mathbf{X}, \mathbf{Y})$ *and* $(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}})$ *respectively.*
If the criterion defined in 4.2.5 is such that

$$\forall \theta > 0,\ \mu_\lambda(\theta, S_{\mathbf{X}}, S_{\mathbf{Y}}) < \lim_{\theta\to\infty} \mu_\lambda(\theta, S_{\mathbf{X}}, S_{\mathbf{Y}}) + o(1/\sqrt{m})$$

and if the variance in probability of $\hat{\theta}_X$*,* $\hat{\theta}_Y$*,* $\langle \hat{u}_X, \hat{u}_Y\rangle$ *are* $o_m(1)$*, our model is asymptotically weakly robust on the right,*

$$\mathrm{E}_{p(m)}\left[\lambda_{C,m}\right] \leqslant \mathrm{E}_{p(m)}\left[\lambda_{0,m}\right].$$

Remark 4.2.3.1.

1. We can show that assuming data of the classes $\mathcal{C}_B$ and $\mathcal{C}_C$ with same spectrum leads to
$$\mathrm{E}_{p(m)}\left[\lambda_{C,m}(\theta)\right] \leqslant \mathrm{E}_{p(m)}\left[\lambda_{B,m}(\theta)\right].$$
2. If θ is large, $\mathrm{E}_{p(m)}\left[\lambda_{C,m}\right] = \mathrm{E}_{p(m)}\left[\lambda_{0,m}\right]$ but we have no information about the variance.
3. If the criterion is satisfied, then applying the Main Theorem to data generated by model C leads to conservative tests.
4. This theorem can be extended to the minimum residual spike to show robustness on the left.

(Proof page 203)

Class D

In this part we define the class $\mathcal{C}_D$ and a criterion. A theorem shows the robustness of our model applied to data from $\mathcal{C}_D$ satisfying the criterion 4.2.8.

Definition 4.2.7.
We define the class $\mathcal{C}_D(u)$ of random matrices, where $u \in \mathbb{R}^m$. We say that $\mathbf{X} \in \mathbb{R}^{m\times n_X}$ and $\mathbf{Y} \in \mathbb{R}^{m\times n_Y}$ are in $\mathcal{C}_D(u)$ if

$$\begin{aligned}\mathbf{X} = P^{1/2}\Sigma^{1/2}\mathbf{X},\ \text{and}\ \mathbf{Y} = P^{1/2}\Sigma^{1/2}\mathbf{Y},\\ P = \mathrm{I}_m + (\theta-1)uu^t,\ \text{where } u \text{ is a unit uniform random vector},\\ \mathrm{Var}\left(\mathbf{X}_{\cdot,i}\right) = \mathrm{I}_m \text{ for } i = 1,2,...,n_X,\\ \mathrm{Var}\left(\mathbf{Y}_{\cdot,i}\right) = \mathrm{I}_m \text{ for } i = 1,2,...,n_Y.\end{aligned}$$

Finally we assume $m/n_X = c_X$ and $m/n_Y = c_Y$ are fixed constants. (Or $m/n_X - c_X = o(1/m^{1/2})$ and $m/n_Y - c_Y = o(1/m^{1/2})$).

We define a second criterion for small θ.

Definition 4.2.8.
Suppose $\left(S_{\mathbf{X}}, \hat{U}_{\mathbf{X}}\right)$ and $\left(S_{\mathbf{Y}}, \hat{U}_{\mathbf{Y}}\right)$ are the spectra and the eigenvectors of $\mathbf{X}$ and $\mathbf{Y}$, two random matrices deformed by a perturbation P of order k in direction $u_1, u_2, ..., u_k$. We apply the same random rotation $U = (u_1, u_2, ..., u_k, u_{k+1}, ..., u_m)^t$ to $\hat{U}_{\mathbf{X}}$ and $\hat{U}_{\mathbf{Y}}$ and rename them $\hat{U}_{\mathbf{X}}$ and $\hat{U}_{\mathbf{Y}}$. We define W_X and W_Y as

$$W_X\left(S_{\mathbf{X}}, \hat{U}_{\mathbf{X}}\right) = \sum_{i=k+1}^{m} \hat{\lambda}_{\mathbf{X},i} \hat{u}_{\mathbf{X},i} \hat{u}_{\mathbf{X},i}^t,$$
$$W_Y\left(S_{\mathbf{Y}}, \hat{U}_{\mathbf{Y}}\right) = \sum_{i=k+1}^{m} \hat{\lambda}_{\mathbf{Y},i} \hat{u}_{\mathbf{Y},i} \hat{u}_{\mathbf{Y},i}^t,$$

where $\hat{\lambda}_{\mathbf{X},i}$ and $\hat{u}_{\mathbf{X},i}$ are the i^{th} eigenvalue and eigenvector of $\frac{1}{n_X}\mathbf{X}\mathbf{X}^t$.
Then, we define for θ,

$$Q_{\mathbf{X}}\left(\theta, S_{\mathbf{X}}, \hat{U}_{\mathbf{X}}\right) = \sum_{i=k+1}^{m} \frac{\hat{\lambda}_{\mathbf{X},i}}{\tilde{\theta}_{\mathbf{X}} - \hat{\lambda}_{\mathbf{X},i}} \hat{u}_{\mathbf{X},i} \hat{u}_{\mathbf{X},i}^t, \qquad Q_{\mathbf{Y}}\left(\theta, S_{\mathbf{Y}}, \hat{U}_{\mathbf{Y}}\right) = \sum_{i=k+1}^{m} \frac{\hat{\lambda}_{\mathbf{Y},i}}{\tilde{\theta}_{\mathbf{Y}} - \hat{\lambda}_{\mathbf{Y},i}} \hat{u}_{\mathbf{Y},i} \hat{u}_{\mathbf{Y},i}^t,$$
$$\frac{1}{\theta - 1} = \sum_{i=k+1}^{m} \frac{\hat{\lambda}_{\mathbf{X},i}}{\tilde{\theta}_{\mathbf{X}} - \hat{\lambda}_{\mathbf{X},i}} \hat{u}_{\mathbf{X},1,i}^2, \qquad \frac{1}{\theta - 1} = \sum_{i=k+1}^{m} \frac{\hat{\lambda}_{\mathbf{Y},i}}{\tilde{\theta}_{\mathbf{Y}} - \hat{\lambda}_{\mathbf{Y},i}} \hat{u}_{\mathbf{Y},1,i}^2.$$

The matrices $Q_{\mathbf{X}} \equiv Q_{\mathbf{X}}\left(\theta, S_{\mathbf{X}}, \hat{U}_{\mathbf{X}}\right)$ and $Q_{\mathbf{Y}} \equiv Q_{\mathbf{Y}}\left(\theta, S_{\mathbf{Y}}, \hat{U}_{\mathbf{Y}}\right)$ are in $\mathrm{Mat}((m-k) \times (m-k))$. Then, we define

$$T_{r,s}\left(\theta, S_{\mathbf{X}}, S_{\mathbf{Y}}, \hat{U}_{\mathbf{X}}, \hat{U}_{\mathbf{Y}}\right) = \frac{Q_{\mathbf{X},r,s}}{\sqrt{\sum_{t \neq r} (Q_{\mathbf{X},r,t})^2}} \frac{Q_{\mathbf{Y},r,s}}{\sqrt{\sum_{t \neq r} (Q_{\mathbf{Y},r,t})^2}},$$
$$T\left(\theta, S_{\mathbf{X}}, S_{\mathbf{Y}}, \hat{U}_{\mathbf{X}}, \hat{U}_{\mathbf{Y}}\right) = \frac{1}{m-k} \sum_{s}^{m-k} \sum_{r \neq s}^{m-k} T_{r,s}\left(\theta, S_{\mathbf{X}}, S_{\mathbf{Y}}, \hat{U}_{\mathbf{X}}, \hat{U}_{\mathbf{Y}}\right).$$

1. We define $T\left(\theta, S_{\mathbf{X}}, S_{\mathbf{Y}}, \hat{U}_{\mathbf{X}}, \hat{U}_{\mathbf{Y}}\right)$ as the first part of the second criterion for small θ.

2. We define the general criterion for small θ as $\mu_\lambda^G(\theta, S_X, S_Y, \hat{U}_{\mathbf{X}}, \hat{U}_{\mathbf{Y}})$ constructed from 4.2.5 by replacing the estimation of α by

$$\alpha = \alpha_X \alpha_Y + T\sqrt{1 - \alpha_X^2}\sqrt{1 - \alpha_Y^2}.$$

Remark 4.2.8.1.

1. Assuming our model, $T = O_p(1/m^{1/2})$. Moreover, it seems that in practice T tends to be positive. Therefore, the general criterion is bounded by the first criterion 4.2.5,

$$\mu_\lambda \geqslant \mu_\lambda^G.$$

   If this hypothesis holds, then the general criterion would always be useless. However, we are not able to prove such a result.

2. In practice we could replace T by the simpler estimator:

$$T = \frac{1}{d} \sum_{s}^{d} \sum_{r \neq s}^{m-k} T_{r,s},$$

   where d is smaller than $m - k$. If T tends to be strictly larger or smaller than 0, then a finite d is asymptotically sufficient to detect it.

3. This criterion check case for consistency is necessary because $\mathbf{X}$ and $\mathbf{Y}$ are no longer independent! Indeed the random u is the same in the two matrices.

Theorem 4.2.4.
Assume $\mathbf{X} \in \mathbb{R}^{m \times n_X}$ and $\mathbf{Y} \in \mathbb{R}^{m \times n_Y}$ are random matrices from class $\mathcal{C}_D(u)$ defined in 4.2.7 for $u \in \mathbb{R}^m$. By definition of $\mathcal{C}_D$,

$$\mathbf{X} = P^{1/2}\mathbf{X} = P^{1/2}\Sigma^{1/2}\mathbf{X} \text{ and } \mathbf{Y} = P^{1/2}\mathbf{Y} = P^{1/2}\Sigma^{1/2}\mathbf{Y},$$

where $P = \mathrm{I}_m + (\theta - 1)uu^t$. We define

$$\tilde{\mathbf{X}} = \tilde{P}^{1/2}\tilde{\mathbf{X}} \text{ and } \tilde{\mathbf{Y}} = \tilde{P}^{1/2}\tilde{\mathbf{Y}},$$

such that $\tilde{P} = \mathrm{I}_m + (\tilde{\theta} - 1)uu^t$, $\lim \frac{\tilde{\theta}}{\sqrt{m}} = \infty$ and $\tilde{\mathbf{X}}$ (respectively $\tilde{\mathbf{Y}}$) is invariant by rotation with spectrum equal to the spectrum of $\mathbf{X}$ (respectively $\mathbf{Y}$). Moreover, we define $\lambda_{D,m}$ and $\lambda_{0,m}$ as the largest residual spikes obtained using $(\mathbf{X}, \mathbf{Y})$ and $(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}})$ respectively.
If the criterion defined in 4.2.5 is such that

$$\forall \theta > 0,\ \mu_\lambda\left(\theta, S_{\mathbf{X}}, S_{\mathbf{Y}}, \hat{U}_{\mathbf{X}}, \hat{U}_{\mathbf{Y}}\right) < \lim_{\theta \to \infty} \mu_\lambda\left(\theta, S_{\mathbf{X}}, S_{\mathbf{Y}}, \hat{U}_{\mathbf{X}}, \hat{U}_{\mathbf{Y}}\right) + o(1/\sqrt{m})$$

and the variance of $\langle \hat{u}_X, \hat{u}_Y \rangle$ is $o_m(1)$, then

1. *If the criterion 4.2.8 is such that $T = 0$, then our model is asymptotically weakly robust on the right,*

$$\mathrm{E}_{p(m)}\left[\lambda_{D,m}\right] \leqslant \mathrm{E}_{p(m)}\left[\lambda_{0,m}\right].$$

2. *If the criterion of 4.2.8 is such that $T > d$ for d a positive constant, then our model is asymptotically strongly robust on the right,*

$$\mathrm{E}_{p(m)}\left[\lambda_{D,m}\right] + D(d) < \mathrm{E}_{p(m)}\left[\lambda_{0,m}\right] + o_m(1), \text{ where } D(d) \text{ is a positive function of } d,$$
$$\lim_{m\to\infty} \mathrm{Var}_{p(m)}\left(\lambda_{D,m}\right) = \lim_{m\to\infty} \mathrm{Var}_{p(m)}\left(\lambda_{0,m}\right) = 0.$$

Remark 4.2.4.1.

1. If $\mathrm{E}_{p(m)}\left[\lambda_{D,m}\right] = \mathrm{E}_{p(m)}\left[\lambda_{0,m}\right]$ when θ tends to infinity, the asymptotic variances are not necessarily the same. We recall that models letting θ tend to infinity are in class $\mathcal{C}_A$. Therefore, assuming $\Sigma \neq \mathrm{I}_m$ leads to conservative tests.

2. Simulations seem to show that for all θ, when $\Sigma \neq \mathrm{I}_m$, we usually have

$$\mathrm{E}_{p(m)}\left[\lambda_{D,m}\right] < \mathrm{E}_{p(m)}\left[\lambda_{0,m}\right].$$

3. If the criterion is respected and the variance of the residual spike tends to 0, then applying the Main Theorem to data generated by model D leads to conservative tests.

4. This theorem can be extended to the minimum residual spike to show robustness on the left.

(Proof page 204)

4.2.3 Uncertainties

In this section, we discuss the shortcomings of the robustness investigation.

1. The robustness is proven for perturbations of order 1.
   (a) $\mathcal{C}_A$ can probably be extended to finite perturbation of order k with some assumptions on Σ.
   (b) Some additional work is needed to check if criteria 4.2.5 and 4.2.8 still work for perturbation of order k in class $\mathcal{C}_B$ and $\mathcal{C}_D$.
   It is easy to show that if the data are perturbations of order k, then the criteria still checks if the residual spike of a perturbation of order 1 is increasing. However, the criteria do not check if a perturbation of order k creates an increasing residual spike as a function of $(\theta_1, \theta_2, ..., \theta_k)$.
   This extension is not in the thesis but can probably be proven using its theorems. Indeed when θ is small many theorems have a weak form that usually allows the computation of expectations or bounds on the expectation.
   (c) The Main Theorem 3.1.1 assumes 2.2.2(A1) and (A3). We do not show that data satisfying only 2.2.2(A1) and (A4) would lead to a conservative test. It is reasonable to think that small perturbations are not as problematic as large one. Moreover, it seems that all the material to prove such a result is provided in this thesis and we refer to the proof of the Main Theorem and Lemma 6.2.4 to convince the reader.
2. We have proved weak robustness in a large number of scenarios. However, if θ is small and the perturbation is not random (for example linked to Σ as in class $\mathcal{C}_A$), we are not able to prove robustness.
3. The weak robustness only assumes that we are conservative in expectation. Assuming that the variance tends to 0, this robustness is sufficient. However, the variance does not necessarily tend to 0. Moreover, even if the variance tends to 0, a finite number of data could easily create a residual spike with smaller expectation but larger variance such that the quantile 0.95 is enlarged.

4.2.4 Summary

This section summarises in Table 4.1 all the classes where our test is conservative with their weaknesses and their strengths. The weak points are in red and the strong one are in blue. The black comments are considered as interesting generalisations or results. Nevertheless, because another class is more general or provides more powerful results, these comments are not in blue.

	θ	Σ	Temporal	u	Distribution	Criteria	Robustness
Our model	$\frac{\theta}{\sqrt{m}} \to \infty$	$\Sigma = \mathrm{I}_m$	A unique temporal fluctuation trough the rows	No condition on u	Invariant distributions	No	
$\mathcal{C}_A$	$\theta \to \infty$	Weak conditions on Σ	Small temporal differences in direction u	No condition on u	Weak conditions on the distributions	No	Weak, Strong if $\Sigma \neq \mathrm{I}_m$
$\mathcal{C}_B$	$\forall \theta > 1$	$\Sigma = \mathrm{I}_m$	A unique temporal fluctuation trough the rows	No condition on u	Invariant distributions	Simple criterion	Strong
$\mathcal{C}_C$	$\forall \theta > 1$	$\Sigma = \mathrm{I}_m$	A unique temporal fluctuation trough the rows	No condition on u	Symmetric distributions	Simple criterion	Weak
$\mathcal{C}_D$	$\forall \theta > 1$	No condition on Σ	Temporal differences in all directions are allowed	u is uniform	No condition on the distributions	Heavy criterion	Weak if $T = 0$, Strong if $T > 0$

Table 4.1 – Summary of all the classes.

Remark 4.2.1.

1. The Table 4.1 shows that the most interesting classes are $\mathcal{C}_A$ and $\mathcal{C}_D$. The temporal cases

of $\mathcal{C}_A$ just says that the temporal fluctuation, Ψ_1, in the direction of the perturbation has to be closed to the fluctuation, Ψ, in the other directions such that Trace $\left((\Psi_1 - \Psi)^2\right) = o\,(n_X)$ as in the remark of Definition 4.2.3.

2. If $\Sigma = \mathrm{I}_m$, then $\mathcal{C}_A$ is still weakly robust. If $T < 0$, then $\mathcal{C}_D$ is no longer robust!

3. Although we write the invariant distributions in red in the table, this condition is already more general than the condition of Gaussianity.

4.2.5 Simulation

Chapter 8 simulates many scenarios and confirms the good behaviour of our model.
Assuming $\mathbf{X}$ with $\mathbf{X}_{i,j}$ i.i.d. random variable with finite fourth moment, the main conclusions of these simulations are:

1. For $\mathbf{X} = P_k^{1/2}\mathbf{X}$, the procedure still seems weakly robust and "nearly" strongly robust. Moreover, the first criteria 4.2.5 seems to be always satisfied.

2. The procedure appears to be strongly robust for $\mathbf{X} = P_k^{1/2}\Sigma_m^{1/2}\mathbf{X}$.

3. The criterion 4.2.8 for Class $\mathcal{C}_D$ with general k seems to be always respected.

These conclusions hold for autoregressive $X_{i,\cdot}$. The problems appearing in the simulations are:

1. It is not so clear that the procedure remains strongly robust with $\mathbf{X} = P_k^{1/2}\mathbf{X}$ and θ_i large. The simulations only show that using the quantile 0.95 leads to conservative procedures.

2. We can create spectra such that the largest residual spike appears for small θ. Moreover, this construction is feasible for data that are invariant by rotation.
   However, such an example is based on data with a complex temporal structure leading to a spectrum with isolated eigenvalues.

4.2.6 Conclusion

This Chapter argues that, despite strong underlying assumptions, our procedure is a good choice to treat all kinds of data and protects against wrong discoveries.
As seen in the part uncertainties 4.2.3, our four classes do not cover all possibilities. However, it seems that data with large perturbations lead to conservative tests. If this theoretical part does not convince the reader, simulations of Chapter 8 demonstrate the conservative properties of the test applied to other data.
In conclusion, the procedure developed in this thesis is an intuitive test constructed on the basis of a model that is difficult to treat with traditional tools. Moreover, this test has good robustness properties and can easily be applied in practice without risking of mistake in the level of false rejections.

4.3 Power

Our procedure seems to have interesting robustness properties. However, we have not yet investigated its power.
A good test must necessarily have an asymptotic power of 1 when n_X and n_Y tend to infinity for fixed m. Moreover, it would be interesting to reach a power 1 when m also increases with constant ratio, $c_X = \frac{m}{n_X}$ and $c_Y = \frac{m}{n_Y}$.
First, we define the **continuity property**.

Definition 4.3.1.
Let $P = \mathrm{I}_m + \sum_{i=1}^{k}(\theta_i - 1)u_i u_i^t$. We say that the perturbation, $\theta_i u_i u_i^t$ for $i = 1, 2, ..., k$, satisfies the **continuity property** if $\theta_i \propto m$.

Remark 4.3.0.1.
The property of continuity is intuitive for imaging. Assuming that we increase the number of pixels, then a perturbation affects more pixels and increases with m.
If we call $h_m = \sqrt{\theta_m} u_m$, then $h_m h_m^t$ is the perturbation applied to the matrix. For m large we define h_0 such that, $\theta_m = ||h_m||^2 = m||h_0||^2$. Then, we can use an argument of continuity to argue that $\theta_{2m} \approx 2\theta_m$.

$$h_{2m} = (h_{m,1}, \tilde{h}_{m,1}, h_{m,2}, \tilde{h}_{m,2}, ..., h_{m,m}, \tilde{h}_{m,m}),$$

with

$$\begin{aligned} h_m &= (h_{m,1}, h_{m,2}, ..., h_{m,m}), \\ \tilde{h}_m &= (\tilde{h}_{m,1}, \tilde{h}_{m,2}, ..., \tilde{h}_{m,m}) \end{aligned}$$

and by continuity we can reasonably assume that

$$||\tilde{h}_m||^2 = ||h_m||^2 + o(||h||^2).$$

Finally

$$\theta_{2m} = ||h_{2m}||^2 = (2 + o(1))||h_m||^2 = 2\theta_m + o(\theta_m) = 2m||h_0||^2 + o(m).$$

The same computation holds for $\theta_{3m}, \theta_{4m}...$
Therefore, assuming $\theta \propto m$ seems plausible in practice.

Theorem 4.3.1.
Suppose $\mathbf{X} = P^{1/2}\mathbf{X} \in \mathbb{R}^{m \times n_X}$ *and* $\mathbf{Y} = P^{1/2}\mathbf{Y} \in \mathbb{R}^{m \times n_Y}$ *are data such that for* $\mathbf{X} \in \mathbb{R}^{m \times n_X}$ *and* $\mathbf{Y} \in \mathbb{R}^{m \times n_Y}$ *2.2.1 and furthermore* $P \in \mathbb{R}^{m \times m}$ *satisfies Assumption 2.2.2(A4). We test*

$$\begin{aligned} &\mathrm{H}_0 : P_X = P_Y, \\ &\mathrm{H}_1 : P_Y = P_X + (\lambda - 1)vv^t, \end{aligned}$$

where P_X *is a finite perturbation such that*

$$P_X = \mathrm{I}_m + \sum_{i=1}^{k} (\theta_i - 1) u_i u_i^t.$$

We define the ***Power*** *as*

$$\beta_m = P_m \{\mathrm{H}_0 \textit{ rejected } \}.$$

1. *Assuming the continuity properties defined in 4.3.1, then the* ***Power*** *tends to* 1 *in the following two cases:*
   (a) $\langle v, u_i \rangle = 0$ *for* $i = 1, 2, ..., k$,
   (b) $\sum_{i=1}^k \langle v, u_i \rangle^2 \neq 1$,
2. *Assuming the classical multivariate assumption,* $n_X, n_Y \to \infty$ *with* $\frac{m}{n_X}, \frac{m}{n_Y} \to 0$ *and a fixed dependence between the columns (temporal structure), then the* ***Power*** *tends to* 1 *in the two cases defined above.*
3. *Assuming* $m, n_X, n_Y \to \infty$ *with* $c_X = \frac{m}{n_X}$, $c_Y = \frac{m}{n_Y}$, *then*
   (a) *If* $\langle v, u_i \rangle = 0$ *for* $i = 1, 2, ..., k$, *and*

$$\lambda > M_2 + \sqrt{M_2^2 - 1},$$

where

$$M_2 = \frac{M_{2,X} + M_{2,Y}}{2}, \ M_{2,X} = \frac{1}{m}\sum_{i=1}^{m} \hat{\lambda}^2_{\mathbf{X},i} \text{ and } M_{2,Y} = \frac{1}{m}\sum_{i=1}^{m} \hat{\lambda}^2_{\mathbf{Y},i},$$

the ***Power*** *tends to* 1.

(Proof page 208)
This shows that despite the good robustness properties, our procedure keeps a good asymptotic power.

4.4 Application of the model

When applying the procedure to real data, we typically want to achieve the following:

- Estimation of the spectra and their moments,
- Standardisation the data,
- Estimation of k, the order of the perturbation,
- Avoiding the effects of temporal finite perturbation which can lead to a failed test.

Our main result provides a tool to deal with two well-known problems:

- A first group of size n_X looks at m features. What should be the size of the second group in order to obtain results?
- How should the number of principal components be selected and justified in a principal component analysis?

This section will study this problem theoretically and the simulation in Chapter 8 presents some concrete cases of spectrum estimations, standardisation of the data, and control of temporal perturbations. All the routines for the analysis are proposed in our R package "RMTResidualSpike".

4.4.1 Spectrum estimation

Assume data $\mathbf{X} \in \mathbb{R}^{m\times n_X}$ and $\mathbf{Y} \in \mathbb{R}^{m\times n_y}$ respect Assumption 2.2.1 such that

$$\mathbf{X} = P_k^{1/2}\mathbf{X} \text{ and } \mathbf{Y} = P_k^{1/2}\mathbf{Y},$$

where $\mathbf{X}$, $\mathbf{Y}$ are invariant by rotation random matrices with compact bounded spectra and $P_k = \mathrm{I}_m + \sum_{i=1}^{k}(\theta_i - 1)u_i u_i^t$ has distinct eigenvalues whose k_X and k_Y respectively are detectable in $\mathbf{X}$ and $\mathbf{Y}$. We build the estimated covariance matrices,

$$\hat{\Sigma}_{\mathbf{X}} = \frac{\mathbf{X}\mathbf{X}^t}{n_X} \text{ and } \hat{\Sigma}_{\mathbf{Y}} = \frac{\mathbf{Y}\mathbf{Y}^t}{n_Y}.$$

In case of difficulties in estimating k_X and k_Y the last remark of Theorem 4.4.1 recommends underestimating them for the estimation of the spectra.
Theorem 3.1.1 provides the variance and the expectation of random variables leading to the extreme residual spikes:

$$\lambda^{\pm}, \ \sigma^{\pm 2} \text{ and } \sigma_w^{\pm 2}.$$

However, these values are expressed in function of the four first moments of the spectra of W_X and W_Y where

$$W_X = \frac{\mathbf{X}\mathbf{X}^t}{n_X} \text{ and } W_Y = \frac{\mathbf{Y}\mathbf{Y}^t}{n_Y}.$$

Theorem 4.4.1 suggests estimating the moments with

$$\begin{aligned}
M_{s,\hat{\Sigma}_{\mathbf{X}}}(k_X) &= \frac{1}{m-k_X}\sum_{i=k_X+1}^{m}\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right)^s = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{\lambda}_{W_X,i}\right)^s + O_p\left(\frac{1}{m}\right) = M_{s,W_X} + O_p\left(\frac{1}{m}\right),\\
M_{s,\hat{\Sigma}_{\mathbf{Y}}}(k_Y) &= \frac{1}{m-k_Y}\sum_{i=k_Y+1}^{m}\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{Y}},i}\right)^s = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{\lambda}_{W_X,i}\right)^s + O_p\left(\frac{1}{m}\right) = M_{s,W_Y} + O_p\left(\frac{1}{m}\right),
\end{aligned}$$

where $\hat{\lambda}_{W_X,i}$ is the i^{th} eigenvalue of W_X and $\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}$ is the i^{th} eigenvalue of $\hat{\Sigma}_{\mathbf{X}}$. These estimators provide good results in practice as we will see in the simulation section.
Nevertheless the estimation of $\lambda^{\pm}$ could lose conservative properties when m is not large enough, k is large and the extreme eigenvalues of W_X are far from 1. This problem is partially solved by using a bounded estimation of $M_{2,X}$ and $M_{2,Y}$. This second estimator is provided in the second part of Theorem 4.4.1,

$$\begin{aligned}
M_{2,W_X} + O_p\left(\frac{1}{m}\right) &= \frac{\sum_{i=k_X+1}^{m}\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right)^2 + 2k_X\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}\right)^2}{(m-k_X)M_{1,\hat{\Sigma}_{\mathbf{X}}}(k_X) + 2k_X\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}} \geqslant M_{2,W_X} + O_p\left(\frac{1}{m^{5/3}}\right),\\
M_{2,W_Y} + O_p\left(\frac{1}{m}\right) &= \frac{\sum_{i=k_Y+1}^{m}\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{Y}},i}\right)^2 + 2k_Y\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{Y}},k_Y+1}\right)^2}{(m-k_Y)M_{1,\hat{\Sigma}_{\mathbf{Y}}}(k_Y) + 2k_Y\hat{\lambda}_{\hat{\Sigma}_{\mathbf{Y}},k_Y+1}} \geqslant M_{2,W_Y} + O_p\left(\frac{1}{m^{5/3}}\right).
\end{aligned}$$

Using the previous notation, we present the theorems of this section:

Theorem 4.4.1.
Using the notation of Section 4.4.1 and assuming ϕ is an increasing function on the support of the spectrum, the following holds.

1. *If the values $\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}$ for $i = k_X, k_X+1, ..., k$ are finite ($k \geqslant k_X$),*

$$0 < \frac{1}{m}\sum_{i=1}^{m}\phi\left(\hat{\lambda}_{W_X,i}\right) - \frac{1}{m-k_X}\sum_{i=k_X+1}^{m}\phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right) \leqslant O_p\left(\frac{1}{m}\right).$$

In particular if $\rho \in \mathbb{R}$ is outside the spectrum support, for $s_1, s_2 \geqslant 0$,

$$M_{s_1,s_2,W_X}(\rho) = \frac{1}{m}\sum_{i=1}^{m}\frac{\left(\hat{\lambda}_{W_X,i}\right)^{s_1}}{\left(\rho - \hat{\lambda}_{W_X,i}\right)^{s_2}} + O_p(1/m) = M_{s_1,s_2,\hat{\Sigma}_{\mathbf{X}}}(\rho) + O_p(1/m).$$

More precisely

$$\begin{aligned}
&\frac{k_X\sum_{i=k_X+1}^{m}\phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right)}{m(m-k_X)}\\
&\qquad\leqslant \frac{1}{m}\sum_{i=1}^{m}\phi\left(\hat{\lambda}_{W_X,i}\right) - \frac{1}{m-k_X}\sum_{i=k_X+1}^{m}\phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right)\\
&\qquad\qquad\leqslant \frac{2k\phi\left(\hat{\lambda}_{W_X,1}\right) - \sum_{i=k_X+1}^{k}\phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right)}{m} + \frac{k_X\sum_{i=k_X+1}^{m}\phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right)}{m(m-k_X)}.
\end{aligned}$$

2. *The estimator can be improved when $\phi(x) = x^2$ using another estimator. We assume that the*

variance of the entries of the data is σ^2 *and* $\sigma^2 - M_{1,\hat{\Sigma}_{\mathbf{X}}} = O_p\left(\frac{1}{mn_X}\right)$*, then*

$$\frac{1}{\sigma^2}\frac{1}{m}\sum_{i=1}^{m}\left(\hat{\lambda}_{W_X,i}\right)^2 - \frac{\sum_{i=k_X+1}^{m}\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}(k_X)\right)^2 + 2k_X\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}\right)^2}{(m-k_X)M_{1,\hat{\Sigma}_{\mathbf{X}}} + 2k_X\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}}$$
$$\leqslant \frac{2k\left(\hat{\lambda}_{W_X,1}\right)^2 - \sum_{i=k_X+1}^{k}\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right)^2 - 2k_X\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}\right)^2}{\sigma^2 m}$$
$$+\frac{\left(\sum_{i=k_X+1}^{m}\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right)^2\right)}{m\sigma^2 M_{1,\hat{\Sigma}_{\mathbf{X}}}}$$
$$\left(M_{1,\hat{\Sigma}_{\mathbf{X}}} - \sigma^2 + \frac{-\sum_{i=k_X+1}^{k}\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i} + 2k_X\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1} - (k+1)M_{1,\hat{\Sigma}_{\mathbf{X}}}}{m}\right) + O_p\left(1/m^2\right).$$

Remark 4.4.1.1.

1. The first inequality shows that the error of the intuitive estimator is of order $1/m$. Moreover, we always underestimate the true value and this underestimation can lead to a non-conservative test. Nevertheless, asymptotically, this estimation is enough and simulations show very good performance with reasonably large m.

2. We propose the second estimator to improve the conservative properties of the first one. The first estimator underestimates $M_{2,X}$ with an error of size

$$\frac{2k\phi\left(\hat{\lambda}_{W_X,1}\right) - \sum_{i=k_X+1}^{k}\phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right)}{m} = \frac{2k\left(\hat{\lambda}_{W_X,1}\right)^2 - \sum_{i=k_X+1}^{k}\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right)^2}{m}.$$

   This estimator has also an error of order $1/m$ and tends to bound $M_{2,X}$. However, assuming $\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}$ close to $\hat{\lambda}_{W_X,1}$, the terms that contribute to the underestimation are

$$\frac{2k_X\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}}{m} \text{ and } \frac{(k-k_X)\left(\hat{\lambda}_{W_X,1}\right)^2}{\sigma^2 m}.$$

   The error is of order $1/m$, but the numerator is smaller than $k\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}\right)^2$.

3. We do not treat the case $k_X > k$ because,

   - If $k_X \leqslant k$, the estimators are $O_p\left(1/m\right)$ when the missing perturbations are small. Moreover, underestimation of k leads to overestimation of M_2
   - If $k_X > k$, we tend to underestimate M_2 and the procedure loses the property of conservativeness.

   Before concluding that choosing $k_X < k$ is better, we should remember that this scenario creates errors on the statistics used for the Main Theorem 3.1.1. A better answer could be:

   Choose $\mathbf{k_X < k}$ to estimate the spectrum and $\mathbf{k_X > k}$ for the rest of the procedure.

   Of course all this discussion can be neglected when m is large enough.

(Proof page 211.)

Simulations: In this paragraph we investigate the consequences of Theorem 4.4.1 on the Main Theorem 3.1.1. All the simulations are on page 222.
We recall that Theorem 4.4.1 proposed two estimators, the intuitive estimator and the corrected one which improves the properties of conservativeness.

- The main conclusion of the simulations is the better performance of the intuitive estimator when m is sufficiently large. This does not contradict the theorem that says that both estimators commit an error of order $1/m$ and the corrected one tends to overestimate $M_{2,X}$.

- The simulations directly compare the estimators in the estimation of the residual spikes for known and unknown k. This leads to the second main conclusion. Although we recommend underestimation of k_X and k_X to favour conservativeness, we find in simulation that neither a small overestimation nor a small underestimation substantially affect the test.

- Finally we see that asymptotically both estimators provide good results, even though for small m, large k, and $n_X, n_Y >> m$, the intuitive estimator tends to underestimate the residual spike.

4.4.2 Standardisation

Raw data $\color{red}{\mathbf{X}}$ and $\color{red}{\mathbf{Y}}$ are not centred and scaled. In order to use our procedure we need to standardise our data. The aim of this section is to argue that some standardisations can be applied without perturbing the procedure.
In this section we suppose k as known and defined some colour code and operators for the data $\mathbf{X} \in \mathbb{R}^{m \times n_x}$.

$$\begin{aligned}
\mu(\mathbf{X}) &= \frac{1}{mn_X} \sum_{i,j} \mathbf{X}_{i,j} \in \mathbb{R}, \\
\mu_c(\mathbf{X}) &= \left(\mu\left(\mathbf{X}_{1,1:n_X}\right), \mu\left(\mathbf{X}_{2,1:n_X}\right), ..., \mu\left(\mathbf{X}_{m,1:n_X}\right)\right) \in \mathbb{R}^m, \\
\mu_r(\mathbf{X}) &= \left(\mu\left(\mathbf{X}_{1:m,1}\right), \mu\left(\mathbf{X}_{1:m,2}\right), ..., \mu\left(\mathbf{X}_{1:m,n_X}\right)\right) \in \mathbb{R}^{n_X}, \\
\sigma(\mathbf{X}) &= \sqrt{\frac{1}{mn_X} \sum_{i,j} \mathbf{X}_{i,j}^2} \in \mathbb{R}, \\
\sigma_c(\mathbf{X}) &= \left(\sigma\left(\mathbf{X}_{1,1:n_X}\right), \sigma\left(\mathbf{X}_{2,1:n_X}\right), ..., \sigma\left(\mathbf{X}_{m,1:n_X}\right)\right) \in \mathbb{R}^m, \\
\sigma_r(\mathbf{X}) &= \left(\sigma\left(\mathbf{X}_{1:m,1}\right), \sigma\left(\mathbf{X}_{1:m,2}\right), ..., \sigma\left(\mathbf{X}_{1:m,n_X}\right)\right) \in \mathbb{R}^{n_X}, \\
\mathcal{P}_k^{-1}(\mathbf{X}) &= U_{1:m,k+1:m} \Lambda_{k+1:m,1:n_X} V_{1:n_X,1:n_X}^t \text{ where we use the svd } \mathbf{X} = U\Lambda V^t.
\end{aligned}$$

In this section, we use red, as in $\color{red}{\mathbf{X}}$, to denote the raw data, and blue as in $\color{blue}{\mathbf{X}}$, the processed data ready for our procedure. The other colours allow us to distinguish the data as it moves through the steps of the standardisation process.

The different procedures of the standardisation processes are as follows.

Centring the column vectors:

We can subtract from column the column average in order to simplify the computation of covariance matrix estimator.

$$\begin{aligned}
&\color{blue}{\mathbf{X}} = \color{red}{\mathbf{X}} - \left(\mu_c(\color{red}{\mathbf{X}})^t, \mu_c(\color{red}{\mathbf{X}})^t, ..., \mu_c(\color{red}{\mathbf{X}})^t\right) = \color{red}{\mathbf{X}} - \frac{1}{n_X}\mathbf{1}\mathbf{1}^t\color{red}{\mathbf{X}} \in \mathbb{R}^{m \times n_X}, \\
&\color{blue}{\mathbf{Y}} = \color{red}{\mathbf{Y}} - \left(\mu_c(\color{red}{\mathbf{Y}})^t, \mu_c(\color{red}{\mathbf{Y}})^t, ..., \mu_c(\color{red}{\mathbf{Y}})^t\right) \in \mathbb{R}^{m \times n_Y}, \\
&\hat{\Sigma}_{\color{blue}{\mathbf{X}}} = \color{blue}{\mathbf{X}}\color{blue}{\mathbf{X}}^t/n_X \text{ and } \hat{\Sigma}_{\color{blue}{\mathbf{Y}}} = \color{blue}{\mathbf{Y}}\color{blue}{\mathbf{Y}}^t/n_Y.
\end{aligned}$$

Justification: Assume

$${\color{red}\mathbf{X}} = P_k^{1/2}\mathbf{X} + M \text{ or } {\color{red}\mathbf{X}}_{\cdot,i} = P_k^{1/2}\mathbf{X}_{\cdot,i} + \vec{\mu},$$

where $M = (\vec{\mu}, \vec{\mu}, ..., \vec{\mu})$ is a matrix of size $m \times n_X$ and $\mathbf{X}$ is invariant by rotation.
Then, the matrix $\hat{\Sigma}_{{\color{blue}\mathbf{X}}}$ is invariant by rotation because

$$\begin{aligned}
{\color{blue}\mathbf{X}}_{\cdot,i} &= {\color{red}\mathbf{X}}_{\cdot,i} - \mu_c({\color{red}\mathbf{X}})^t \\
&= ({\color{red}\mathbf{X}}_{\cdot,i} - \vec{\mu}) - (\mu_c({\color{red}\mathbf{X}}) - \vec{\mu}) \\
&= \left(P_k^{1/2}\mathbf{X}_{\cdot,i}\right) - \frac{1}{n_X}\sum_{j=1}^{n_X}({\color{red}\mathbf{X}}_{\cdot,j} - \vec{\mu}) \\
&= \left(P_k^{1/2}\mathbf{X}_{\cdot,i}\right) - \frac{1}{n_X}\sum_{j=1}^{n_X} P_k^{1/2}\mathbf{X}_{\cdot,j}.
\end{aligned}$$

Moreover,

$$\begin{aligned}
{\color{blue}\mathbf{X}} &= {\color{red}\mathbf{X}}\left(\mathrm{I}_{n_X} - \frac{1}{n_X}\mathbf{1}\mathbf{1}^t\right)^{1/2} \\
&= P_k^{1/2}\mathbf{X}\left(\mathrm{I}_{n_X} - \frac{1}{n_X}\mathbf{1}\mathbf{1}^t\right)^{1/2}.
\end{aligned}$$

The right perturbation does not affect the global spectrum and the spike of $P_k\mathbf{X}$ except by adding a null eigenvalue.

Adaptation of the method: We cannot directly apply the Main Theorem 3.1.1. Indeed the spectrum of the matrix $\mathbf{X}$ is not standardised. This generalisation is not proposed because we can just replace $M_{s,X}$ and $M_{s,Y}$ by

$$\frac{M_{s,X}}{\sigma^s} \text{ and } \frac{M_{s,Y}}{\sigma^s},$$

where $\sigma^2 = \mathrm{E}\left[\mathrm{Trace}\left(\frac{\mathbf{X}\mathbf{X}^t}{n_X}\right)\right] = \mathrm{E}\left[\mathrm{Trace}\left(\frac{\mathbf{Y}\mathbf{Y}^t}{n_Y}\right)\right]$.
This value can be estimated by Theorem 4.4.1 assuming $\sigma_X \neq \sigma_Y$ or $\sigma_X \neq \sigma_Y$.
In the first case we recommend to use

$$\begin{aligned}
\hat{\sigma}_{{\color{blue}\mathbf{X}}} &= M_{1,\hat{\Sigma}_{{\color{blue}\mathbf{X}}}}, \\
\hat{\sigma}_{{\color{blue}\mathbf{Y}}} &= M_{1,\hat{\Sigma}_{{\color{blue}\mathbf{Y}}}}.
\end{aligned}$$

In the second scenario we could choose,

$$\hat{\sigma}_{{\color{blue}\mathbf{X}},{\color{blue}\mathbf{Y}}} = \frac{n_X M_{1,\hat{\Sigma}_{{\color{blue}\mathbf{X}}}} + n_Y M_{1,\hat{\Sigma}_{{\color{blue}\mathbf{Y}}}}}{n_X + n_Y}.$$

We recall that a second estimator is given by Theorem 4.4.1 and the choice of k should be an underestimate of the correct value to protect conservatism.

Centring and rescaling the column vectors:

Instead of rescaling the matrices, the above procedure rescales the spectra when using the Main

Theorem. We could also rescale the data.

$$\begin{aligned}
\mathbf{X} &= \mathbf{X} - \left(\mu_c(\mathbf{X})^t, \mu_c(\mathbf{X})^t, ..., \mu_c(\mathbf{X})^t\right) \in \mathbb{R}^{m \times n_X}, \\
\mathbf{X} &= \mathbf{X}/\sigma\left(\mathcal{P}_k^{-1}(\mathbf{X})\right), \\
\hat{\Sigma}_{\mathbf{X}} &= \mathbf{X}\mathbf{X}^t/n_X.
\end{aligned}$$

Justification: It is clear that if the procedure works for $\mathbf{X}$ and $\mathbf{Y}$, then it also works asymptotically for $\mathbf{X}$ and $\mathbf{Y}$ despite the fact that a small mistake on k creates an error of order $1/m$.

Adaptation of the method: In this case we can directly apply the Main Theorem by replacing $M_{s,X}$ and $M_{s,Y}$ by their estimators introduced in Theorem 4.4.1.
When m is not large enough and in case of uncertainty on k, we recommend to overestimate it a little in $\sigma\left(\mathcal{P}_k^{-1}(\mathbf{X})\right)$ and to underestimate it when using Theorem 4.4.1.

Centring and rescaling of Subgroups column vectors:

If the n_X samples are divided in $\tilde{n}_X$ subgroups, then we could standardise each subgroup separately. We use the notation $\mathbf{X}_{G_s}$ to look at the data matrix created by the subgroup s.

$$\begin{aligned}
\mathbf{X}_{G_s} &= \mathbf{X}_{G_s} - \left(\mu_c(\mathbf{X}_{G_s})^t, \mu_c(\mathbf{X}_{G_s})^t, ..., \mu_c(\mathbf{X}_{G_s})^t\right) \in \mathbb{R}^{m \times n_X}, \\
\mathbf{X}_{G_s} &= \mathbf{X}_{G_s}/\sigma\left(\mathcal{P}_k^{-1}(\mathbf{X}_{G_s})\right).
\end{aligned}$$

Then, we could also rescale the whole matrix by the estimated standard error.

Adaptation of the method: In this case we can directly apply the Main Theorem by replacing $M_{s,X}$ and $M_{s,Y}$ by their estimators introduced in Theorem 4.4.1.
When k is unknown we should overestimate it a little in $\sigma\left(\mathcal{P}_k^{-1}(\mathbf{X}_{G_s})\right)$ and underestimate it when using Theorem 4.4.1.

Remark 4.4.1.
The procedure does not work for all partitions of n_X and n_Y in subgroups. We are only able to treat the data if at least one of the following conditions holds:

- all the sizes of the subgroups are of same size,
- the subgroups do not create isolated temporal spikes presented in Section 4.4.3.

In other cases, the subgroups could create temporal spikes that we can not distinguish from the spatial spikes.

Centring and rescaling the row and the column vectors:

Some temporal structure interfering with the study could be minimised by a stronger standardisation. However, this kind of processing creates weaker, but not necessarily less useful tests.

$$\begin{aligned}
\mathbf{X} &= \mathbf{X} - \left(\mu_c(\mathbf{X})^t, \mu_c(\mathbf{X})^t, ..., \mu_c(\mathbf{X})^t\right) \in \mathbb{R}^{m \times n_X}, \\
\mathbf{X} &= \mathbf{X} - \begin{pmatrix} \mu_r(\mathbf{X}) \\ \mu_r(\mathbf{X}) \\ \vdots \\ \mu_r(\mathbf{X}) \end{pmatrix} \in \mathbb{R}^{m \times n_X}, \\
\mathbf{X} &= \mathbf{X}/\sigma\left(\mathcal{P}_k^{-1}(\mathbf{X})\right).
\end{aligned}$$

Justification: Assume

$$\mathbf{X} = P_k^{1/2}\mathbf{X} + M + \tilde{M} \text{ or } \mathbf{X}_{\cdot,i} = P_k^{1/2}\mathbf{X}_{\cdot,i} + \vec{\mu} + \eta_i,$$

where $M = (\vec{\mu}, \vec{\mu}, ..., \vec{\mu})$ and $\tilde{M}^t = (\vec{\eta}, \vec{\eta}, ..., \vec{\eta})^t$ are matrices of size $m \times n_X$ and $\mathbf{X}$ is invariant by rotation. Therefore, $\vec{\mu} = (\mu_1, \mu_2, ..., \mu_m)$ and $\vec{\eta} = (\eta_1, \eta_2, ..., \eta_{n_X})$.
The matrix $\hat{\Sigma}_{\mathbf{X}}$ is invariant by rotation because

$$\begin{aligned}
\mathbf{X}_{\cdot,i} &= \mathbf{X}_{\cdot,i} - \mu_c(\mathbf{X})^t \\
&= \left(\mathbf{X}_{\cdot,i} - \vec{\mu} - \eta_i \vec{\mathbf{1}}_m\right) - \left(\mu_c(\mathbf{X}) - \vec{\mu} - \eta_i \vec{\mathbf{1}}_m\right) \\
&= \left(P_k^{1/2}\mathbf{X}_{\cdot,i}\right) - \frac{1}{n_X}\sum_{j=1}^{n_X}\left(\mathbf{X}_{\cdot,j} - \vec{\mu} - \eta_i \vec{\mathbf{1}}_m\right) \\
&= \left(P_k^{1/2}\mathbf{X}_{\cdot,i}\right) - \frac{1}{n_X}\sum_{j=1}^{n_X}\left(P_k^{1/2}\mathbf{X}_{\cdot,j} + \eta_j \vec{\mathbf{1}}_m - \eta_i \vec{\mathbf{1}}_m\right) \\
&= \left(P_k^{1/2}\mathbf{X}_{\cdot,i}\right) - \frac{1}{n_X}\sum_{j=1}^{n_X}\left(P_k^{1/2}\mathbf{X}_{\cdot,j}\right) - \bar{\eta}\vec{\mathbf{1}}_m + \eta_i \vec{\mathbf{1}}_m,
\end{aligned}$$

where $\vec{\mathbf{1}}_m = (1, 1, ..., 1) \in \mathbb{R}^m$.
Then,

$$\begin{aligned}
\mathbf{X}_{t,i} &= \mathbf{X}_{t,i} - \mu_r(\mathbf{X})_i \\
&= \left(P_k^{1/2}\mathbf{X}_{\cdot,i}\right)_t - \frac{1}{n_X}\sum_{j=1}^{n_X}\left(P_k^{1/2}\mathbf{X}_{\cdot,j}\right)_t - \bar{\eta} + \eta_i \\
&\qquad - \sum_{h=1}^{m}\left(P_k^{1/2}\mathbf{X}_{\cdot,i}\right)_h + \frac{1}{mn_X}\sum_{h=1}^{m}\sum_{j=1}^{n_X}\left(P_k^{1/2}\mathbf{X}_{\cdot,j}\right)_h + (\bar{\eta} - \eta_i) \\
&= \left(P_k^{1/2}\mathbf{X}_{\cdot,i}\right)_t - \frac{1}{n_X}\sum_{j=1}^{n_X}\left(P_k^{1/2}\mathbf{X}_{\cdot,j}\right)_t - \sum_{h=1}^{m}\left(P_k^{1/2}\mathbf{X}_{\cdot,i}\right)_h + \frac{1}{mn_X}\sum_{h=1}^{m}\sum_{j=1}^{n_X}\left(P_k^{1/2}\mathbf{X}_{\cdot,j}\right)_h.
\end{aligned}$$

Therefore

$$\begin{aligned}
\mathbf{X} &= \left(\mathrm{I}_m - \frac{1}{m}\mathbf{1}\mathbf{1}^t\right)^{1/2} \mathbf{X} \left(\mathrm{I}_{n_X} - \frac{1}{n_X}\mathbf{1}\mathbf{1}^t\right)^{1/2} \\
&= \left(\mathrm{I}_m - \frac{1}{m}\mathbf{1}\mathbf{1}^t\right)^{1/2} P_k^{1/2}\mathbf{X} \left(\mathrm{I}_{n_X} - \frac{1}{n_X}\mathbf{1}\mathbf{1}^t\right)^{1/2} \\
&= H_m^{1/2} P_k^{1/2} \mathbf{X} H_{n_X}^{1/2}.
\end{aligned}$$

The temporal perturbation H_{n_X} just adds a null eigenvalue; however, the spatial perturbation is projected on a subspace without the direction $\mathbf{1}$ by H_m. We define

$$\tilde{P}_k = H_m^{1/2} P_k H_m^{1/2}$$

such that

$$\begin{aligned}
\hat{\Sigma}_{\mathbf{X}} &= H_m^{1/2} P_k^{1/2} \mathbf{X} H_{n_X} \mathbf{X}^t P_k^{1/2} H_m^{1/2} \\
&= O\Lambda U^t \mathbf{X} H_{n_X} \mathbf{X}^t U \Lambda O^t \\
&= O\Lambda O^t O U^t \mathbf{X} H_{n_X} \mathbf{X}^t U O^t O \Lambda O^t \\
&= \tilde{P}_k^{1/2} O U^t \mathbf{X} H_{n_X} \mathbf{X}^t U O^t \tilde{P}_k^{1/2} \\
&\sim \tilde{P}_k^{1/2} \mathbf{X} H_{n_X} \mathbf{X}^t \tilde{P}_k^{1/2},
\end{aligned}$$

where the last line is obtained by invariance under orthogonal transformation and $H_m^{1/2}P_k^{1/2} = O\Lambda U^t$ is a singular value decomposition.

Remark 4.4.2.
If $\mathbf{X}$ and $\mathbf{Y}$ are not invariant by rotation, this transformation is still allowed. Because

$$\begin{aligned} \hat{\Sigma}_{\mathbf{X}} &= \tilde{P}_k^{1/2} O U^t \mathbf{X} H_{n_X} \mathbf{X}^t U O^t \tilde{P}_k^{1/2} = \tilde{P}_k^{1/2} \tilde{\mathbf{X}} H_{n_X} \tilde{\mathbf{X}}^t \tilde{P}_k^{1/2} \\ \hat{\Sigma}_{\mathbf{Y}} &= \tilde{P}_k^{1/2} O U^t \mathbf{Y} H_{n_Y} \mathbf{Y}^t U O^t \tilde{P}_k^{1/2} = \tilde{P}_k^{1/2} \tilde{\mathbf{Y}} H_{n_Y} \tilde{\mathbf{Y}}^t \tilde{P}_k^{1/2}, \end{aligned}$$

the distributions of $\mathbf{X}$ and $\tilde{\mathbf{X}}$ are different but $\tilde{\mathbf{X}}$ has still same distribution as $\tilde{\mathbf{Y}}$.

Therefore,

$$\begin{aligned} \mathbf{X} &= \tilde{P}_k^{\ 1/2} \mathbf{X} H_{n_X}^{1/2} = \tilde{\tilde{P}}_k^{1/2} H_m^{1/2} \mathbf{X} H_{n_X}^{1/2} \\ \mathbf{Y} &= \tilde{P}_k^{\ 1/2} \mathbf{Y} H_{n_Y}^{1/2} = \tilde{\tilde{P}}_k^{1/2} H_m^{1/2} \mathbf{Y} H_{n_Y}^{1/2} \end{aligned}$$

where $\tilde{P}_k$ is a finite perturbation with a null eigenvalue and $\tilde{\tilde{P}}_k$ replaces the null eigenvalue of $\tilde{P}_k$ by 1.
The resulting matrices are ready for our procedure. We still look at a finite spatial perturbation after adding two null eigenvalues. The spectra bulk of the matrices were asymptotically unaffected by the standardisation process. However, the eigenvectors of the perturbation are strongly modified by the projection cancelling the direction $\mathbf{1}$. Therefore, if the difference is exclusively in this direction, the procedure will not detect it.

Adaptation of the method: As for the previous processes, we can use Theorem 4.4.1 to estimate the parameters of the Main Theorem 3.1.1. The choice of k has an impact when m is not large enough and was discussed previously.

Centring and rescaling of subgroups row and column vectors:

Some temporal structure could be divided in $\tilde{n}_X$ groups. In this case we would like to standardise each subgroup separately. We use the notation $\mathbf{X}_{G_s}$ to look at the data matrix created by the group s.

$$\begin{aligned} \mathbf{X}_{G_s} &= \mathbf{X}_{G_s} - \left(\mu_c(\mathbf{X}_{G_s})^t, \mu_c(\mathbf{X}_{G_s})^t, ..., \mu_c(\mathbf{X}_{G_s})^t\right) \in \mathbb{R}^{m \times n_X}, \\ \mathbf{X}_{G_s} &= \mathbf{X}_{G_s} - \begin{pmatrix} \mu_r(\mathbf{X}_{G_s}) \\ \mu_r(\mathbf{X}_{G_s}) \\ \vdots \\ \mu_r(\mathbf{X}_{G_s}) \end{pmatrix} \in \mathbb{R}^{m \times n_X}, \\ \mathbf{X}_{G_s} &= \mathbf{X}_{G_s} / \sigma\left(P_k^{-1}(\mathbf{X}_{G_s})\right). \end{aligned}$$

The problem of data pre-processing is created by the effect that the preprocessing has on the distribution of the data. Can we still justify our model ?
Through the description of all the processes, we see that if the data can be expressed as a transformation of $\mathbf{X}$ invariant by rotation, then we can invert the transformation to keep the invariance properties.
More generally if we can express the data as a transformation of $\mathbf{X}$ such that the procedure is justified for $\mathbf{X}$, then the procedure is valid for the pre-processed data applying the inverse of the transformation.

Simulations: We comment on the effect of the pre-processing on page 8.
The five processes introduced in Section 4.4.2 provide good results when applying them to appropriate data. If the data are adapted for the procedure, it seems that we can apply the processing and build

conservative tests.
However, the two last standardisations, which center the rows, loose all power in the direction $\vec{\mathbf{1}}$ by applying a projection. We think that this loss a power is not important for two reasons. First, differences in this direction are directly found by other methods. Second, it is important in addition to cancel temporal structures as well as possible in order to decrease large temporal perturbations.

4.4.3 Temporal perturbation

The strongest point of our method is its ability to deal with temporal structures. As shown previously, temporal structures do not affect our result as long as we are able to distinguish temporal spikes from spatial spikes. In practice, this is typically not the case! Temporal and spatial perturbations are merged in the spectrum and treating a temporal spike as a spatial spike leads to mistakes. In this section we call $\mathbf{X}$, the data before preprocessing, $\mathbf{X}$ the transformed data suitable for our procedure and $\mathbf{X}$ the data without any finite perturbations.
On the one hand, dealing with compact temporal structures is our strength. These structures do not create isolated estimated eigenvalues. On the other hand finite isolated temporal perturbations are our weakest point as we explain below. We want to test

$$
\begin{aligned}
H_0 &: P_X = P_Y, \\
H_1 &: P_X \neq P_Y.
\end{aligned}
$$

This null hypothesis is wrong if the data create temporal finite perturbations. It is also the case even if these perturbations are the same in the two groups. The correct hypotheses would be,

$$
\begin{aligned}
\mathbf{X} &= P_X^{1/2}\mathbf{X}P_{X,T}^{1/2}, \\
\mathbf{Y} &= P_Y^{1/2}\mathbf{Y}P_{Y,T}^{1/2}
\end{aligned}
$$

such that $\mathbf{X}$ and $\mathbf{Y}$ have no isolated spike and we test

$$
\begin{aligned}
\mathrm{H}_0 &: P_X = P_Y \text{ and } P_{X,T} = \mathrm{I}_{n_x},\ P_{Y,T} = \mathrm{I}_{n_X}, \\
\mathrm{H}_1 &: P_X \neq P_Y \text{ or } P_{X,T} \neq \mathrm{I}_{n_x} \text{ or } P_{Y,T} \neq \mathrm{I}_{n_Y}.
\end{aligned}
$$

Therefore, H_0 could be rejected because of temporal perturbations.

In this section we proposed a heuristic method to filter temporal perturbations. In this context, the data under H_0 are such that

$$
\mathbf{X} = P^{1/2}\mathbf{X}P_{X,T}^{1/2},
$$

and create k either temporal or spatial spikes.
In order to treat the temporal perturbation we proposed the following algorithm based on the singular value decomposition.

Algorithm:

1. For $i = 1, 2, ..., n_X$, compute R_i such that

   $$\mathbf{X}_{1:m,-i} = R_i \Lambda_i V_i.$$

   Define

   $$\mathbf{X} = \left(\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_{n_X}\right),$$

   where $\mathbf{X}_i = R_i^t \mathbf{X}_{1:m,i}$.

2. Compute

   $$\mathbf{X} = \mathbf{X}\hat{\psi}^{-1/2},$$

where

$$\begin{aligned}
\hat{\psi} &= (m-k)\left(\mathrm{I}_{n_X} + \sum_{i=1}^{k} \left(\lambda\left(W_{-k}\right)-1\right) u\left(W_{-k}\right) u\left(W_{-k}\right)^t\right), \\
W_{-k} &= \frac{\mathbf{X}^t_{k+1:m,1:n_X}\mathbf{X}_{k+1:m,1:n_X}}{m-k} \in \mathbb{R}^{n_X \times n_X}.
\end{aligned}$$

3. Finally we call $\mathbf{X}$ the temporal filtered matrix,

$$\mathbf{X}_{1:m,i} = R_i \mathbf{X}_{1:m,i}.$$

Justification:

The following algorithm is heuristic. It works for data $\mathbf{X}$ that are invariant under rotations or are such that the singular values allow for spatial-temporal separability. We have not formally proved it, but we present a convincing heuristic argument for this case.
If the temporal and spatial perturbations correspond to different eigenvalues, then the rotation, R_i in the first part of the algorithm sends all the spatial perturbations on e_1 to e_k but does not affect the temporal perturbation.
The reason for this temporal robustness to the rotation R_i is the lack of precision of the spatial eigenvectors corresponding to temporal spikes. More precisely, let λ_t be an eigenvalue corresponding to a temporal perturbation in direction $v_t \in \mathbb{R}^{n_X}$ and denote by $\hat{u}_t \in \mathbb{R}^m$ a spatial noisy estimated eigenvector. Because λ_t is mainly temporal, the estimation of the same perturbation without the i^{th} column ($n_X - 1$ samples) provides an other noisy $\hat{u}_{-i,t}$ with $\langle \hat{u}_{-i,t}, \hat{u}_t \rangle$ tending to 0.
Luckily for us this noisy estimator allows us to invert the temporal perturbation without touching the spatial perturbation. Indeed in the second step we estimate the temporal perturbation, $\hat{\psi}$, with only $m-k$ rows and this estimation tends to the correct perturbation.
Therefore, we do not affect the spatial perturbation and neutralise to some degree the temporal perturbation.
The question is whether we neutralize enough of the temporal perturbation ?
By simulation, the answer seems to be yes if the temporal perturbation is not too large.

Simulations:

The simulations on page 8.7 indicate the following:

- If there is no temporal perturbation, then the algorithm does not affect the existing spikes. The procedure creates k temporal spikes that do not change the asymptotic behaviour of the residual spike.
- Assuming a data matrix $\mathbf{X}$ that is invariant by rotation and a finite temporal perturbation, then the algorithm deletes the temporal spikes without touching the spatial perturbation.
- Intuitively we would expect worse simulations for temporal perturbations of order larger than $o(m)$. However our simulations still show good conservative properties in this case.

Remark 4.4.3.
Another temporal filtered procedure consists of rescaling each column by its variance. This process will delete canonical temporal perturbations. We will not pursue this procedure any further in this thesis.

Summary:

In conclusion, if the data $\mathbf{X}$ and $\mathbf{Y}$ are suspected to have temporal perturbations, we should use the algorithm of this section. Alternatively, the user can simply apply the usual procedure and check in the end if the residual eigenvalue is due to a temporal eigenvector. Therefore, instead of looking at the eigenstructure of

$$\hat{\hat{\Sigma}}_X^{-1/2}\hat{\hat{\Sigma}}_Y\hat{\hat{\Sigma}}_X^{-1/2},$$

we look at the eigenstructure of

$$\hat{\hat{\Sigma}}_X^{-1/2}\left(\sum_{i=k+1}^{m} \hat{u}_{Y,i}\hat{v}_{Y,i}^t + \sum_{i=1}^{k}(\hat{\hat{\theta}}_{Y,i}-1)\hat{u}_{Y,i}\hat{v}_{Y,i}^t\right),$$

where $\hat{v}_{Y,i}$ is the i^{th} singular vector of **Y**. If the square of extreme residual eigenvalue of the previous matrix is too large compared to the theory, we can then investigate its corresponding temporal eigenvector.

4.4.4 The problem of selecting the number of components

The method of this section is completely heuristic and could probably be improved. In the section of the Main Theorem 3.1, we showed by simulation that an error on k does not lead to a less conservative test. If we add too many perturbation we build a very conservative test and on the other hand if we miss some small perturbations we still build a conservative test when m is large.

How to choose k?

We recall that we look at

$$\lambda_{\max}\left(\hat{\hat{\Sigma}}_X^{-1}\hat{\hat{\Sigma}}_Y\right) \text{ and } \lambda_{\min}\left(\hat{\hat{\Sigma}}_X^{-1}\hat{\hat{\Sigma}}_Y\right),$$

where $\hat{\hat{\Sigma}}_X$ and $\hat{\hat{\Sigma}}_Y$ are the filtered estimators of $\hat{\Sigma}_X$ and $\hat{\Sigma}_Y$. We reject if the value is outside an interval.
The estimated covariance matrices are assumed to be finite perturbations of W_X and W_Y,

$$\hat{\Sigma}_X = P_X^{1/2}W_XP_X^{1/2} \text{ and } \hat{\Sigma}_Y = P_Y^{1/2}W_YP_Y^{1/2},$$

where P_X and P_Y are finite perturbations of order k_X and k_Y respectively.
The method of this section selects the numbers k_X and k_Y respecting the three following points:

- The algorithm tends to choose the correct k_X and k_Y asymptotically (if easily detectable).
- The procedure is conservative.
- If a perturbation is only detectable in the group X, the procedure must not reject this perturbation.

We propose the following method. We define the interval,

$$\begin{aligned}&I(k_x, k_y, S_{\hat{\Sigma}_X,-k_x}, S_{\hat{\Sigma}_Y,-k_Y}, \beta)\\ &= \left[\text{Quantile}_\beta\left(\hat{\lambda}_{\min,\max(k_X,k_Y),S_{\hat{\Sigma}_X,-k_x},S_{\hat{\Sigma}_Y,-k_Y}}\right), \text{Quantile}_{1-\beta}\left(\hat{\lambda}_{\max,\max(k_X,k_Y),S_{\hat{\Sigma}_X,-k_X},S_{\hat{\Sigma}_Y,-k_Y}}\right)\right],\end{aligned}$$

where $\hat{\lambda}_{\min,\max(k_X,k_Y),S_{\hat{\Sigma}_X,-k_X},S_{\hat{\Sigma}_Y,-k_Y}}$ is the random residual spike defined in the Main Theorem 3.1.1 when the perturbation is of order $k = \max(k_X, k_Y)$. The estimated spectra are $S_{\hat{\Sigma}_X,-k_X}$ and $S_{\hat{\Sigma}_Y,-k_Y}$, the spectra of respectively $\hat{\Sigma}_X$ and $\hat{\Sigma}_Y$ without respectively the k_X, k_Y largest eigenvalues.

The idea of the algorithm is the following :
The matrices $\hat{\Sigma}_X$ and $\hat{\Sigma}_Y$ have some biased perturbations. It seems obvious that if an unbiased observed eigenvalue falls inside the interval, $I(k_X, k_Y, S_{\hat{\Sigma}_X,-k_X}, S_{\hat{\Sigma}_Y,-k_Y}, \beta)$ for β close to 1, this value

will not create an interesting residual spike and can be neglected. Indeed, it will create a residual spike much smaller than our threshold.

We look at different values of k_X and k_Y in $\mathcal{K}$ chosen by the user.

1. For $k_X, k_Y \in \mathcal{K}$, define $S_{\hat{\Sigma}_X,-k_X}$ and $S_{\hat{\Sigma}_Y,-k_Y}$ the estimated spectra of $\hat{\Sigma}_X$ and $\hat{\Sigma}_Y$ without respectively the k_X and the k_Y largest eigenvalues.

2. Compute two tables with entries (k_X, k_Y) equal to

   - 1 if the k_Y^{th} unbiased eigenvalue of $\hat{\Sigma}_Y$, $\hat{\hat{\theta}}_{Y,k_Y}$ is larger than $I\left(k_X, k_Y, S_{\hat{\Sigma}_X,-k_X}, S_{\hat{\Sigma}_Y,-k_Y}\right)$,
   - 0 else,

   and the second

   - 1 if the k_X^{th} unbiased eigenvalues of $\hat{\Sigma}_X$ is such that $1/\hat{\hat{\theta}}_{X,k_X}$ is smaller than $I\left(k_X, k_Y, S_{\hat{\Sigma}_X,-k_X}, S_{\hat{\Sigma}_Y,-k_Y}\right)$,
   - 0 else.

   For example

k_X \ k_Y	1	2	3	4	5
1	0	0	0	0	0
2	1	1	0	0	0
3	1	1	1	0	0
4	1	1	1	0	0
5	1	1	1	0	0

Selection of k_Y

k_Y \ k_X	1	2	3	4	5
1	0	0	0	0	0
2	1	1	0	0	0
3	1	1	1	0	0
4	1	1	1	1	0
5	1	1	1	1	0

Selection of k_X

3. Finally we choose the largest k_X such that the k_X^{th} column contains a 1 in the selection table of k_X. The second table chooses k_Y similarly. In the example, we set $k_X = 4$ and $k_Y = 3$.

Remark 4.4.4.
When a perturbation θ is detectable only in X but not in Y, the test should not reject the equality. Indeed the worst possible case occurs when n_X tends to infinity and the k^{th} perturbation θ_k is not detectable in Y. In this case, if we want to avoid detecting a difference in θ_k, we need

$$\theta_k << \frac{1 + M_{2,Y}}{2} + \sqrt{\frac{(1 + M_{2,Y})^2}{4} - 1}.$$

Assuming Wishart distributions for W_X and W_Y then the spectra tend to Marcenko-Pastur distributions and $M_{2,Y} = 1 + c_Y$. In this case the property is straightforward:

- θ_k is not detectable in Y because $\theta_k < 1 + \sqrt{c_Y}$ by 1.5.2,
- θ is detectable as a difference if $\theta > 1 + \frac{c_Y}{2} + \sqrt{c_Y + \frac{c_Y^2}{4}} > 1 + \sqrt{c_Y} > \theta_k$.

Therefore, the algorithm satisfies the condition in this case.
However, we were not able to prove this property for general spectra. Therefore, when $\frac{n_X}{n_Y}$ is large, we recommend checking if the unbiased eigenvalues are detectable in both groups. The following could be applied:

Suppose that n_X is larger than n_Y, we test if $\hat{\hat{\theta}}_{X,k}$ could have been detected in Y. We solve

$$\frac{1}{\hat{\hat{\theta}}_{X,k}-1} = \frac{1}{m-k_Y}\sum_{i=k_Y+1}^{m} \frac{\hat{\lambda}_{\hat{\Sigma}_Y,i}}{\rho_Y - \hat{\lambda}_{\hat{\Sigma}_Y,i}}$$

to find ρ_Y. Then, we could apply a test to see whether ρ_Y is strictly larger than $\hat{\lambda}_{\hat{\Sigma}_Y,k_Y+1}$ or just look whether

$$\rho_Y - \hat{\lambda}_{\hat{\Sigma}_Y,k_Y+1} >> \hat{\lambda}_{\hat{\Sigma}_Y,k_Y+1} - \hat{\lambda}_{\hat{\Sigma}_Y,k_Y+2},$$

where the difference should be important in practice.
When the difference does not seem significant enough, we should just neglect this perturbation.

4.4.5 Other applications

The method developed in this thesis can also be used in different situations.

Estimation of the size of the second group

Suppose a study has already collected n_X data on healthy patients. The analysis seems to show a perturbation of order k. The researchers need to estimate a minimum number $n_Y \leqslant n_X$ of patients for the second group. The Main result 3.1.1 can be used to estimate this number.
We know the distribution of the residual spike under H_0. Moreover, if an additional perturbation $(\eta-1)u_\eta u_\eta^t$ orthogonal to the other is applied to the group Y, the reader can show that

$$\lambda_{\max}\left(\hat{\hat{\Sigma}}_X^{-1}\hat{\hat{\Sigma}}_Y\right) \underset{Asy}{\overset{\mathrm{H}_1}{\sim}} \hat{\hat{\eta}},$$

1. First we choose some group sizes n_Y that we would like to test. For example $n_y = 100, 300, 800$.

2. We need to estimate the spectrum of Y for different values n_Y by the spectrum of $X[1:m, 1:n_Y]$ without the k largest eigenvalues.

3. We use the Main Theorem to estimate the distribution of the residual spikes under H_0. The spectrum of X is known, the spectrum of Y is replaced by its estimator and k is replaced by $k+1$.

4. For n_Y and a fixed α, we can plot the power in function of η,

$$\beta_\alpha(\eta) = P_\eta\left\{\hat{\hat{\eta}} \geqslant q\lambda_{\max,k+1}(1-\alpha)\right\},$$

   where $q\lambda_{\max,k+1}(1-\alpha)$ is the $(1-\alpha)-$quantile of the largest residual spike assuming a perturbation of order $k+1$. This value is easily computed by the Main Theorem 3.1.1.
   This allows us to choose n_Y such that interesting η can be detected with good probability.

Remark 4.4.5.
This heuristic way to estimate the power can be criticised for many reasons such as for example:

- The model is conservative and leads to an overestimation of n_Y,

- The estimation is slow because we compute the eigenvalues of huge matrices for each size n_Y.

Estimation of the number of principal components

The Main Theorem 3.1.1 can help to solve the famous problem of selecting the number of principal components.
In this case, we have only one matrix $\mathbf{Y} = P_k^{1/2}\mathbf{Y}$. The idea of the selection criterion is to assume that the i^{th} component with unbiased eigenvalues $\hat{\hat{\theta}}_i$ is interesting if

$$\hat{\hat{\theta}} > q\lambda_{\max,i}(1-\alpha),$$

$q\lambda_{\max,i}(1-\alpha)$ is the $(1-\alpha)-$quantile of the largest residual spike assuming perturbation of order i and S_X is a spectrum of 1.
The distribution of this largest eigenvalues is computed in the remark of the Main Theorem 3.1.1 or by assuming $M_{s,X} = 1$ for $s = 1, 2, 3, 4$.

This method considers that $\hat{\hat{\theta}}_i$ creates a new principal component if, applying our procedure to $\mathbf{Y} = P_i^{1/2}\mathbf{Y}$ and $\mathbf{X}_\infty = P_{i-1}^{1/2}\mathbf{X}_\infty$, a sample of size $n_X \to \infty$ with $P_i - P_{i-1} = (\theta_i - 1)u_i u_i^t$, would detect the difference between perturbations.

Of course if the s^{th} component with $s > r$ is accepted as principal component, then the r^{th} is also accepted.

Remark 4.4.6.

1. The same idea could be used to compute

$$P\left\{\hat{\hat{\theta}}_i > \lambda_{\max}\left(P_{i-1}^{-1}\hat{\hat{\Sigma}}_{Y,P_{i-1}}\right)\Big|\,\hat{\hat{\theta}}_i\right\},$$

where P_{i-1} is the perturbation containing the $i-1$ largest perturbations of P_k.

2. We let the user decide how many component have to be tested. But we recall that our test only allows a finite number of principal components and we recommend $k < 15$.

3. We could imagine a less restrictive criterion that would compare each principal component to the residual spike created by a perturbation of order 1. Indeed in our criteria, the k^{th} component is significant if it is larger than the residual spike induced by perturbation of order k.

Testing a particular covariance matrix

Suppose that the data $\mathbf{X} \in \mathbb{R}^{m\times n}$ have centred normal entries with covariance Σ. We divide the columns of $\mathbf{X}$ into two groups of size m_1 and m_2 and write $\Sigma = \begin{pmatrix}\Sigma_{1,1} & \Sigma 1,2\\ \Sigma_{2,1} & \Sigma_{2,2}\end{pmatrix}$ with $\Sigma_{1,1} \in \mathbb{R}^{m_1\times m_1}$, $\Sigma_{2,2} \in \mathbb{R}^{m_2\times m_2}$ and $\Sigma_{1,2} = \Sigma_{2,1}^t \in \mathbb{R}^{m_1\times m_2}$.
The books of Muirhead [2005], Anderson [2003] and Mardia et al. [1979] proposed a test for

$$\mathrm{H}_0 \ : \ \Sigma_{1,2} = 0.$$

Using the ratio of the likelihood, the classical theory proposed to use the test statistic,

$$T = \frac{\left|\hat{\Sigma}_{m_1+1:m,m_1+1:m} - \hat{\Sigma}_{m_1+1:m,1:m_1}\hat{\Sigma}_{1:m_1,1:m_1}\hat{\Sigma}_{1:m_1,m_1+1:m}\right|}{\left|\hat{\Sigma}_{m_1+1:m,m_1+1:m}\right|}.$$

Using the independence of

$$W_1 = n\left(\hat{\Sigma}_{m_1+1:m,m_1+1:m} - \hat{\Sigma}_{m_1+1:m,1:m_1}\hat{\Sigma}_{1:m_1,1:m_1}\hat{\Sigma}_{1:m_1,m_1+1:m}\right) \sim \text{Wishart}\left(\Sigma_{2,2}, df = n-1-m_1\right)$$

and

$$W_2 = n\left(\hat{\Sigma}_{m_1+1:m,1:m_1}\hat{\Sigma}_{1:m_1,1:m_1}\hat{\Sigma}_{1:m_1,m_1+1:m}\right) \sim \text{Wishart}\left(\Sigma_{2,2}, df = m_1\right),$$

this random variable is a Wilks.
This statistic is looking at the difference between two independent Wishart random matrices, W_1 and W_2. We easily see that

$$T = \frac{1}{\left|\mathrm{I}_{m_2} + W_1^{-1}W_2\right|}.$$

As is usual for a test from the classical theory, this test is weak when the differences are finite perturbations and affects only a few eigenvalues.
We propose to directly test equality between W_1 and W_2 using the Main Theorem 3.1.1. On one hand, this new test will be more powerful in some situations. On the other hand, some investigations of the robustness are still necessary for non normal entries.

Chapter 5

Theorems

5.1 Introduction

In this chapter, we introduce the main theorems of this thesis :

- The convergence of the eigenvalues, the angle and the double angle.
- The joint distribution of the asymptotic eigenvalues and angles for perturbation of order 1.
- Invariant Eigenvalue Theorem.
- Invariant Angle Theorem.
- Invariant Dot Product Theorem.
- Invariant Component Theorem.
- The Double Invariant Angle Theorem. (Corollary)

The name "invariant" comes from **invariant** statistics as functions of k, the order of the perturbation.

Definition 5.1.1.
Suppose W is a random matrix. Moreover, define $P_1 = \mathrm{I}_m + (\theta_1 - 1)u_1u_1^t$ and $P_k = \mathrm{I}_m + \sum_{i=1}^{k}(\theta_i - 1)u_iu_i^t$ to be perturbations of orders 1 and k respectively. We say that a statistic $T(W_m, P_1)$ is **invariant** with respect to k, if $T(W_m, P_k)$ is such that

$$T(W_m, P_k) = T(W_m, P_1) + \epsilon_m, \text{ where } \max\left(\frac{\epsilon_m}{\mathrm{E}\left[T(W, P_1)\right]}, \frac{\epsilon_m^2}{\mathrm{Var}\left(T(W, P_1)\right)}\right) \to 0.$$

Before introducing the theorems of this thesis, we need to clarify the notation.

Notation 5.1.1.
Although we use a precise notation to enunciate the theorems, the proofs often use a simpler notation when no confusion is possible. This difference is always specified at the beginning of the proofs.

- If W is a symmetric random matrix, we denote by $\left(\hat{\lambda}_{W,i}, \hat{u}_{W,i}\right)$ its i^{th} eigenvalues and eigenvectors. If a random matrix $\mathbf{X} \in \mathbb{R}^{m\times n}$ is not symmetric, $\hat{\lambda}_{\mathbf{X},i}$ and $\hat{u}_{\mathbf{X},i}$ denote the i^{th} eigenvalues and eigenvectors of $\frac{1}{n}\mathbf{X}\mathbf{X}^t$.
- A finite perturbation of order k is denoted by $P_k = \mathrm{I}_m + \sum_{i=1}^{k}(\theta_i - 1)u_iu_i^t$. Similarly, a perturbation in the direction u_i is denoted by $\tilde{P}_i = \mathrm{I}_m + (\theta_i - 1)u_iu_i^t$.

- We denote by $W \in \mathbb{R}^{m\times m}$ an invariant by rotation random matrix as defined in Assumption 2.2.1. Moreover, the estimated covariance matrix is $\hat{\Sigma} = P_k^{1/2} W P_k^{1/2}$.
When comparing two groups, we use W_X, W_Y and $\hat{\Sigma}_X$, $\hat{\Sigma}_Y$.

- When we consider only one group, $\hat{\Sigma}_{P_r} = P_r^{1/2} W P_r^{1/2}$ is the perturbation of order r of the matrix W and:
  - $\hat{u}_{P_r,i}$ is its i^{th} eigenvector.
  - $\hat{u}_{P_r,i,j}$ is the j^{th} component of the i eigenvector.
  - $\hat{\lambda}_{P_r,i}$ is its i^{th} eigenvalue. If $\theta_1 > \theta_2 > ... > \theta_r$, then for $i = 1,2,...,r$ we use also the notation $\hat{\theta}_{P_r,i} = \hat{\lambda}_{P_r,i}$. We call these eigenvalues the spikes. When $r = k$, we just use the simpler notation $\hat{\theta}_i = \hat{\theta}_{P_k,i}$.
  - $\hat{\alpha}^2_{P_r,i} = \sum_{j=1}^r \langle \hat{u}_{P_r,i}, u_j\rangle^2$ is called the **general angle**.

  With this notation, we have $\hat{\Sigma} = \hat{\Sigma}_{P_k} = P_k^{1/2} W P_k^{1/2}$.

- When we look at two groups X and Y, we use a notation similar to the above. The perturbation of order r of the matrices W_X and W_Y are $\hat{\Sigma}_{X,P_r} = P_r^{1/2} W_X P_r^{1/2}$ and $\hat{\Sigma}_{Y,P_r} = P_r^{1/2} W_Y P_r^{1/2}$ respectively. Then, we define for the group $\hat{\Sigma}_{X,P_r}$ (and similarly for $\hat{\Sigma}_{Y,P_r}$):
  - $\hat{u}_{\hat{\Sigma}_{X,P_r},i}$ is its i^{th} eigenvector.
  - $\hat{u}_{\hat{\Sigma}_{X,P_r},i,j}$ is the j^{th} component of the i eigenvector.
  - $\hat{\lambda}_{\hat{\Sigma}_{X,P_r},i}$ is its i^{th} eigenvalue. If $\theta_1 > \theta_2 > ... > \theta_r$, then for $i = 1,2,...,r$ we use the notation $\hat{\theta}_{\hat{\Sigma}_{X,P_r},i} = \hat{\lambda}_{\hat{\Sigma}_{X,P_r},i}$. When $r = k$, we just use the simpler notation $\hat{\theta}_{X,i} = \hat{\theta}_{\hat{\Sigma}_{X,P_k},i}$
  - $\hat{\alpha}^2_{\hat{\Sigma}_{X,P_r},i} = \sum_{j=1}^r \left\langle \hat{u}_{\hat{\Sigma}_{X,P_r},i}, u_j \right\rangle^2$.
  - $\hat{\alpha}^2_{X,Y,P_r,i} = \sum_{j=1}^r \left\langle \hat{u}_{\hat{\Sigma}_{X,P_r},i}, \hat{u}_{\hat{\Sigma}_{Y,P_r},j} \right\rangle^2$ is the **double angle** and, when no confusion is possible, we use the simpler notation $\hat{\alpha}^2_{P_r,i}$. When this simpler notation is used, it is stated explicitly.

- The theorems can assume a sign convention
$$\hat{u}_{P_s,i,i} > 0, \text{ for } s = 1,2,...,k \text{ and } i = 1,2,...,s,$$
as in Theorem 5.7.1 or 5.8.1. On the other hand, some theorems assume the convention
$$\hat{u}_{P_s,i,s} > 0, \text{ for } s = 1,2,...,k \text{ and } i = 1,2,...,s,$$
as in Theorem 5.11.1.
Other theorems are not affected by this convention and do not specify it precisely. Nevertheless, the convention will be given in the proofs when confusion is possible.

- We define the function $M_{s_1,s_2,X}(\rho_X)$, $M_{s_1,s_2,Y}(\rho_Y)$ and $M_{s_1,s_2}(\rho_X,\rho_Y)$ as
$$\begin{aligned} M_{s_1,s_2,X}(\rho_X) &= \frac{1}{m}\sum_{i=1}^m \frac{\hat{\lambda}^{s_1}_{W_X,i}}{\left(\rho_X - \hat{\lambda}_{W_X,i}\right)^2}, \\ M_{s_1,s_2,Y}(\rho_Y) &= \frac{1}{m}\sum_{i=1}^m \frac{\hat{\lambda}^{s_1}_{W_Y,i}}{\left(\rho_Y - \hat{\lambda}_{W_Y,i}\right)^2}, \\ M_{s_1,s_2}(\rho_X,\rho_Y) &= \frac{M_{s_1,s_2,X}(\rho_X) + M_{s_1,s_2,Y}(\rho_Y)}{2}. \end{aligned}$$

In particular, when $s_2 = 0$, we use $M_{s_1,X} = M_{s_1,0,X}$. When we only study one group, we use the simpler notation $M_{s_1,s_2}(\rho)$ when no confusion is possible.

- We use three transforms:
  - T_{f_S} is used for the T-transform defined in Definition 1.3.2 using f_S, but if no confusion is possible we call it T.
  - $T_{W,u}(z) = \sum_{i=1}^{m} \frac{\hat{\lambda}_{W,i}}{z-\hat{\lambda}_{W,i}} \langle \hat{u}_{W,i}, u\rangle^2$ is the transform in direction u using the random matrix W.
  - $\hat{T}_{\hat{\Sigma}_X}(z) = \frac{1}{m}\sum_{i=k+1}^{m} \frac{\hat{\lambda}_{\hat{\Sigma}_X,i}}{z-\hat{\lambda}_{\hat{\Sigma}_X,i}}$, and $\hat{T}_{W_X}(z) = \frac{1}{m}\sum_{i=1}^{m} \frac{\hat{\lambda}_{W_X,i}}{z-\hat{\lambda}_{W_X,i}}$ are the estimated transforms using $\hat{\Sigma}_X$ and W respectively.
- Finally, we define some convenient notations for the eigenvectors. Assuming $\hat{u}_i$ the i^{th} eigenvector of size m with j^{th} entry $\hat{u}_{i,j}$, then
  - $\hat{u}_{\cdot,j} = (\hat{u}_{1,j}, \hat{u}_{2,j}, ..., \hat{u}_{m,j})$,
  - $\hat{u}_{r:s,j} = (\hat{u}_{r,j}, \hat{u}_{r+1,j}, ..., \hat{u}_{s,j})$,
  - $\sum \hat{u}^2_{q:r,s:t} = \sum_{i=q}^{r}\sum_{j=s}^{t} \hat{u}^2_{i,j}$.

Assumptions

In this section, we assume two sets of data, $X_1, ..., X_{n_X}$ and $Y_1, ..., Y_{n_Y}$ satisfying Assumption 2.2.1. This means that without loss of generality the perturbations are canonical. We build $\hat{\hat{\Sigma}}_X$ and $\hat{\hat{\Sigma}}_Y$,

$$\hat{\hat{\Sigma}}_X = \mathrm{I}_m + \sum_{i=1}^{k}(\hat{\hat{\theta}}_{X,i} - 1)\hat{u}_{X,i}\hat{u}^t_{X,i},$$
$$\hat{\hat{\Sigma}}_Y = \mathrm{I}_m + \sum_{i=1}^{k}(\hat{\hat{\theta}}_{Y,i} - 1)\hat{u}_{Y,i}\hat{u}^t_{Y,i},$$

where

$$\begin{aligned}\hat{\hat{\theta}}_{X,i} &= 1 + \frac{1}{\hat{T}_{\hat{\Sigma}_X}\left(\hat{\theta}_{X,i}\right)} \\ &= 1 + \frac{m}{\sum_{i=k+1}^{m} \frac{\hat{\lambda}_{\hat{\Sigma}_X,i}}{\hat{\theta}_{X,i}-\hat{\lambda}_{\hat{\Sigma}_X,i}}},\end{aligned}$$

where the $\hat{\lambda}_{\hat{\Sigma}_X,i}$ are the eigenvalues of $\hat{\Sigma}_X$.
The aim of all the following theorems is to understand the joint behaviour of all the unbiased eigenvalues, $\hat{\hat{\theta}}_{X,i}$, and all the possible dot products between the eigenvectors.
This task is difficult and we must assume **detectable** perturbations of order k satisfying Assumption 2.2.2(A4),

$$\begin{aligned}&\theta_i \neq \theta_j \text{ if } i \neq j,\\ &\text{For } i = 1, 2, ..., k_1,\ \theta_i = p_i\theta \text{ with } \theta_i \to \infty,\\ &\text{For } i = k_1 + 1, k_1 + 2, ..., k,\ \theta_i \text{ are finite.}\end{aligned}$$

These assumptions allow us to prove the theorems, although they are not necessary for all the theorems. First, assumption $(\theta_i \neq \theta_j)$ is necessary for the second moment results of many theorems, but seems to be useless in the Main Theorem 3.1.1.

Then, the finite proportionality p_i does not seem necessary even though many arguments in all the proofs use the fact that p_i is fixed when θ_i is large.
Finally, a **detectable** perturbation is just a perturbation that we can see in the spectrum of our data. More precisely,

Definition 5.1.2.

1. We assume that a perturbation $P = \mathrm{I}_m + (\theta - 1)uu^t$ is **detectable** in $\hat{\Sigma} = P^{1/2}WP^{1/2}$ if the perturbation creates a new largest isolated eigenvalue, $\hat{\theta}$.

2. We say that a finite perturbation of order k is **detectable** if it creates k largest eigenvalues well separated from the spectrum of W.

In the following theorems, we always assume that the perturbations are detectable. Nevertheless, with a small amount of additional calculation, we could extend all the results to the existence of a finite number of non-detectable perturbations.

5.2 Convergence of eigenvalue, angle and double angle

In this section, we study the convergence of the random variable $\hat{\hat{\theta}}_X$ and the angle between the eigenvectors. The parts $2.a$ and $2.b$ are proven in Benaych-Georges and Rao [2009] which also provides the main idea of the theorem. We only proves convergence results for perturbations of order $k = 1$, $P = P_1$, although we express eigenvectors and eigenvalues of a matrix $P^{1/2}WP^{1/2}$ as a function of the eigenstructure of W in general.

Theorem 5.2.1.
In this theorem, $P = \mathrm{I}_m + (\theta - 1)uu^t$ is a finite perturbation of order 1.

1. *Suppose W is a symmetric matrix with eigenvalues $\hat{\lambda}_{W,i} \geqslant 0$ and eigenvectors $\hat{u}_{W,i}$ for $i = 1, 2, ..., m$ (without any additional assumptions on the eigenstructure).*
For $i = 1, 2, ..., m$, we define $\tilde{u}_{P,i}$ and $\hat{\lambda}_{P,i}$ such that

$$WP\tilde{u}_{P,i} = \hat{\lambda}_{P,i}\tilde{u}_{P,i},$$

and the usual $\hat{u}_{P,i}$ such that if $\hat{\Sigma}_P = P^{1/2}WP^{1/2}$, then

$$\hat{\Sigma}_P\hat{u}_{P,i} = P^{1/2}WP^{1/2}\hat{u}_{P,i} = \hat{\lambda}_{P,i}\hat{u}_{P,i}.$$

- *The eigenvalues $\hat{\lambda}_{P,s}$ are such that for $s = 1, 2, ..., m$,*

$$\sum_{i=1}^{m} \frac{\hat{\lambda}_{W,i}}{\hat{\lambda}_{\hat{\Sigma}_P,s} - \hat{\lambda}_{W,i}} \langle \hat{u}_{W,i}, u\rangle^2 = \frac{1}{\theta_k - 1}.$$

- *The eigenvectors $\tilde{u}_{P,s}$ are such that*

$$\langle \tilde{u}_{P,s}, v\rangle^2 = \frac{\left(\sum_{i=1}^{m} \frac{\hat{\lambda}_{W,i}}{\hat{\lambda}_{P,s}-\hat{\lambda}_{W,i}} \langle \hat{u}_{W,i}, v\rangle \langle \hat{u}_{W,i}, u\rangle\right)^2}{\sum_{i=1}^{m} \frac{\hat{\lambda}_{W,i}^2}{(\hat{\lambda}_{P,s}-\hat{\lambda}_{W,i})^2} \langle \hat{u}_{W,i}, u\rangle^2}.$$

In particular if $v = u$,

$$\langle \tilde{u}_{P,s}, u\rangle^2 = \frac{1}{(\theta - 1)^2 \left(\sum_{i=1}^{m} \frac{\hat{\lambda}_{W,i}^2}{(\hat{\lambda}_{P,s}-\hat{\lambda}_{W,i})^2} \langle \hat{u}_{W,i}, u\rangle^2\right)}.$$

Moreover,

$$\hat{u}_{P,s} = \frac{P^{1/2}\tilde{u}_{P,s}}{\sqrt{1+(\theta-1)\left\langle \tilde{u}_{P,s}, u\right\rangle^2}}.$$

Therefore, for u and v such that $\langle v, u\rangle = 0$,

$$\begin{aligned}
\left\langle \hat{u}_{P,s}, u\right\rangle^2 &= \frac{\theta\left\langle \tilde{u}_{P,s}, u\right\rangle^2}{1+(\theta-1)\left\langle \tilde{u}_{P,s}, u\right\rangle^2} = -\frac{\theta}{(\theta-1)^2\hat{\lambda}_{P,s}T'_{W,u}(\hat{\lambda}_{P,s})},\\
\left\langle \hat{u}_{P,s}, v\right\rangle^2 &= \frac{\left\langle \tilde{u}_{P,s}, u\right\rangle^2}{1+(\theta-1)\left\langle \tilde{u}_{P,s}, u\right\rangle^2},
\end{aligned}$$

where $T_{W,u}(z) = \sum_{i=1}^m \frac{\hat{\lambda}_{W,i}}{z-\hat{\lambda}_{W,i}}\left\langle \hat{u}_{W,i}, u\right\rangle^2$.

2. *Suppose that W_X, W_Y and $P = P_X = P_Y$ satisfy Assumption 2.2.1. Moreover, suppose that θ is large enough to create detectable spikes, $\hat{\theta}_X$ and $\hat{\theta}_Y$, in the matrices $\hat{\Sigma}_X = P^{1/2}W_XP^{1/2}$ and $\hat{\Sigma}_Y = P^{1/2}W_YP^{1/2}$. Then,*

$$\begin{aligned}
&a)\quad \hat{\hat{\theta}}_X, \hat{\hat{\theta}}_Y \xrightarrow[n,m\to\infty]{P} \theta,\\
&b)\quad \left\langle \hat{u}_X, u\right\rangle - \alpha_X, \left\langle \hat{u}_Y, u\right\rangle - \alpha_Y \xrightarrow[n,m\to\infty]{P} 0,\\
&c)\quad \left\langle \hat{u}_X, \hat{u}_Y\right\rangle - \alpha_X\alpha_Y \xrightarrow[n,m\to\infty]{P} 0,
\end{aligned}$$

where

$$\begin{aligned}
&\hat{\theta}_X \xrightarrow[n,m\to\infty]{P} \rho_X,\\
&\hat{\hat{\theta}}_X = 1 + \frac{1}{\hat{T}_{\hat{\Sigma}_X}\left(\hat{\theta}_X\right)} = 1 + \frac{m}{\sum_{i=k+1}^m \frac{\hat{\lambda}_{\hat{\Sigma}_X,i}}{\hat{\theta}_X - \hat{\lambda}_{\hat{\Sigma}_X,i}}},\\
&\alpha_X^2 = -\frac{\theta}{(\theta-1)^2\rho_X\hat{T}'_{W_X}(\rho_X)},\\
&\alpha_Y^2 = -\frac{\theta}{(\theta-1)^2\rho_Y\hat{T}'_{W_Y}(\rho_Y)},
\end{aligned}$$

$\hat{\lambda}_{\hat{\Sigma}_X,i}$ *and* $\hat{\lambda}_{\hat{\Sigma}_Y,i}$ *are the eigenvalues of respectively* $\hat{\Sigma}_X$ *and* $\hat{\Sigma}_Y$.

Remark 5.2.1.1.
If the spectra of W_X and W_Y tend fast enough to the Marcenko-Pastur distribution with parameter c_X and c_Y respectively, then

$$\alpha_X^2 = \frac{1 - \frac{c_X}{(\theta-1)^2}}{1 + \frac{c_X}{\theta-1}},$$

$\hat{\hat{\theta}}_X$ is such that $\hat{\theta}_X = \hat{\hat{\theta}}_X\left(1 + \frac{c_X}{\hat{\hat{\theta}}_X - 1}\right)$, *and*

$$\lim_{m\to\infty}\hat{\theta}_X = \theta\left(1 + \frac{c_X}{\theta-1}\right).$$

(Proof page 103)

The second part of Theorem 5.2.1 2 is very surprising! We already knew that the eigenvectors are not consistent. We show in the proof that the difference between $\hat{u}_X$ and $\hat{u}_Y$ is larger than the difference between $\hat{u}_X$ and u or $\hat{u}_Y$ and u.
The consequences of this Theorem are simple:

We can not conclude that two estimated eigenvectors are different even if they are asymptotically different! There is always a bias !

5.3 Asymptotic distribution of the eigenvalue and the angle

Suppose that you observe a perturbation of order $k=1$ applied to two random matrices $\mathbf{X} \in \mathbb{R}^m \times \mathbb{R}^{n_X}$ and $\mathbf{Y} \in \mathbb{R}^m \times \mathbb{R}^{n_Y}$ such that Assumption 2.2.1 is respected. We investigate the distribution of $\hat{\hat{\theta}}_X$, $\hat{\hat{\theta}}_Y$, $\langle \hat{u}_X, u\rangle^2$ and $\langle \hat{u}_X, \hat{u}_Y\rangle^2$.

Theorem 5.3.1.
Suppose W_X and W_Y satisfy 2.2.1 with $P = P_X = P_Y$, a detectable perturbation of order $k=1$. Moreover, we assume that the spectra of of W_X and W_Y are known. We defined

$$\begin{aligned} \hat{\Sigma}_X &= P^{1/2} W_X P^{1/2}, \\ \hat{\Sigma}_Y &= P^{1/2} W_Y P^{1/2}, \\ P &= \mathrm{I}_m + (\theta - 1) u u^t, \end{aligned}$$

where u is fixed and the eigenvalues of W_X and W_Y are $\hat{\lambda}_{W_X,i}$ and $\hat{\lambda}_{W_Y,i}$ for $i = 1, 2, ..., m$,respectively. We construct the unbiased estimators of θ,

$$\hat{\hat{\theta}}_X \left| \frac{1}{\hat{\hat{\theta}}_X - 1} = \frac{1}{m} \sum_{i=1}^m \frac{\hat{\lambda}_{W_X,i}}{\hat{\theta}_X - \hat{\lambda}_{W_X,i}} \right. \quad \textit{and} \quad \hat{\hat{\theta}}_Y \left| \frac{1}{\hat{\hat{\theta}}_Y - 1} = \frac{1}{m} \sum_{i=1}^m \frac{\hat{\lambda}_{W_Y,i}}{\hat{\theta}_Y - \hat{\lambda}_{W_Y,i}} \right.$$

where $\hat{\theta}_X = \hat{\lambda}_{\hat{\Sigma}_X,1}$ and $\hat{\theta}_Y = \hat{\lambda}_{\hat{\Sigma}_Y,1}$ are the largest eigenvalues of respectively $\hat{\Sigma}_X$ and $\hat{\Sigma}_Y$ corresponding to the eigenvectors $\hat{u}_X = \hat{u}_{\hat{\Sigma}_X,1}$ and $\hat{u}_Y = \hat{u}_{\hat{\Sigma}_Y,1}$.
We can also construct these estimators using the $m-1$ smaller eigenvalues of $\hat{\Sigma}_X$ and $\hat{\Sigma}_Y$ instead of $\hat{\lambda}_{W_X,i}$ and $\hat{\lambda}_{W_Y,i}$.

1. *If $\frac{\theta}{\sqrt{m}} \to 0$, we define*

$$M_{s,r,X} \equiv M_{s,r,X}(\rho_X) = \frac{1}{m} \sum_{i=1}^m \frac{\hat{\lambda}^s_{W_X,i}}{(\rho_X - \hat{\lambda}_{W_X,i})^r}, \; M_{s,r,Y} \equiv M_{s,r,Y}(\rho_Y) = \frac{1}{m} \sum_{i=1}^m \frac{\hat{\lambda}^s_{W_X,i}}{(\rho_Y - \hat{\lambda}_{W_X,i})^r},$$

where we assume

$$\rho_X = \mathrm{E}\left[\hat{\theta}_X\right] + o\left(\frac{\theta}{\sqrt{m}}\right), \; \rho_Y = \mathrm{E}\left[\hat{\theta}_Y\right] + o\left(\frac{\theta}{\sqrt{m}}\right).$$

and convergence rate of $(\hat{\theta}_X, \hat{\theta}_Y)$ to (ρ_X, ρ_Y) in $O_p\left(\theta/\sqrt{m}\right)$,

$$\begin{pmatrix} \hat{\hat{\theta}}_X \\ \langle \hat{u}_X, u\rangle^2 \end{pmatrix} \sim \mathbf{N}\left(\begin{pmatrix} \theta \\ \alpha_X^2 \end{pmatrix}, \frac{1}{m} \begin{pmatrix} \sigma^2_{\theta,X} & \sigma_{\theta,\alpha^2,X} \\ \sigma_{\theta,\alpha^2,X} & \sigma^2_{\alpha^2,X} \end{pmatrix} \right) + \begin{pmatrix} o_p\left(\frac{\theta}{\sqrt{m}}\right) \\ o_p\left(\frac{1}{\theta\sqrt{m}}\right) \end{pmatrix},$$

$$\begin{pmatrix} \hat{\hat{\theta}}_X \\ \hat{\hat{\theta}}_Y \\ \langle \hat{u}_X, \hat{u}_Y\rangle^2 \end{pmatrix} \sim \mathbf{N}\left(\begin{pmatrix} \theta \\ \theta \\ \alpha_{X,Y}^2 \end{pmatrix}, \frac{1}{m} \begin{pmatrix} \sigma^2_{\theta,X} & 0 & \sigma_{\theta,\alpha^2,X} \\ 0 & \sigma^2_{\theta,Y} & \sigma_{\theta,\alpha^2,Y} \\ \sigma_{\theta,\alpha^2,X} & \sigma_{\theta,\alpha^2,Y} & \sigma^2_{\alpha^2,X,Y} \end{pmatrix} \right) + \begin{pmatrix} o_p\left(\frac{\theta}{\sqrt{m}}\right) \\ o_p\left(\frac{\theta}{\sqrt{m}}\right) \\ o_p\left(\frac{1}{\theta\sqrt{m}}\right) \end{pmatrix},$$

where

$$\alpha_X^2 = \frac{\theta}{(\theta-1)^2} \frac{1}{\rho_X M_{1,2,X}},$$

$$\alpha_{X,Y}^2 = \frac{\theta^2}{(\theta-1)^4} \frac{1}{\rho_X \rho_Y M_{1,2,X} M_{1,2,Y}},$$

$$\sigma_{\theta,X}^2 = \frac{2\left(M_{2,2,X} - M_{1,1,X}^2\right)}{M_{1,1,X}^4},$$

$$\begin{aligned}\sigma_{\alpha^2,X}^2 = \frac{2\theta^2}{((\theta-1)\rho_X M_{1,2,X})^4} \Bigg(& \rho_X^2 \left(M_{2,4,X} - M_{1,2,X}^2\right) + \left(2\rho_X \frac{M_{1,3,X}}{M_{1,2,X}} - 1\right)^2 \left(M_{2,2,X} - M_{1,1,X}^2\right) \\ & -2\rho_X \left(2\rho_X \frac{M_{1,3,X}}{M_{1,2,X}} - 1\right)\left(M_{2,3,X} - \frac{M_{1,1,X}}{M_{1,2,X}}\right)\Bigg),\end{aligned}$$

$$\begin{aligned}\sigma_{\theta,\alpha^2,X} = \frac{2\theta}{M_{1,1,X}^2 M_{1,2,X}^3 \rho_X^2 (-1+\theta)^2} \Big(& M_{1,1,X} M_{1,2,X}^2 \rho_X + 2M_{1,3,X} M_{2,2,X} \rho_X \\ & + M_{1,1,X}^2 (M_{1,2,X} - 2M_{1,3,X}\rho_X) - M_{1,2,X}(M_{2,2,X} + M_{2,3,X}\rho_X)\Big),\end{aligned}$$

$$\sigma_{\alpha^2,X,Y}^2 = \sigma_{\alpha^2,X}^2 \alpha_Y^4 + \sigma_{\alpha^2,Y}^2 \alpha_X^4 + 4\alpha_{X,Y}^2 (1-\alpha_X^2)(1-\alpha_X^2).$$

2. *If* $\frac{\theta}{\sqrt{m}} \to \infty$ *(Assumption 2.2.2* (A1)*), then we can simplify the formulas. We define*

$$M_{r,X} = \frac{1}{m} \sum_{i=1}^{m} \hat{\lambda}_{W_X,i}^r \quad \textit{and} \quad M_{r,Y} = \frac{1}{m} \sum_{i=1}^{m} \hat{\lambda}_{W_Y,i}^r.$$

Using this notation,

$$\begin{aligned}\begin{pmatrix} \hat{\hat{\theta}}_X \\ \langle \hat{u}_X, u\rangle^2 \end{pmatrix} \sim \mathbf{N}\Bigg(& \begin{pmatrix} \theta + O_p(1) \\ 1 + \frac{1-M_{2,X}}{\theta} + O_p\left(\frac{1}{\theta^2}\right) \end{pmatrix}, \\ & \frac{1}{m}\begin{pmatrix} 2\theta^2(1-M_{2,X}) & 2\left(2M_{2,X}^2 - M_{2,X} - M_{3,X}\right) \\ 2\left(2M_{2,X}^2 - M_{2,X} - M_{3,X}\right) & \frac{2}{\theta^2}\left(4M_{2,X}^3 - M_{2,X}^2 - 4M_{2,X}M_{3,X} + M_{4,X}\right) \end{pmatrix}\Bigg) + \begin{pmatrix} o_p\left(\frac{\theta}{\sqrt{m}}\right) \\ o_p\left(\frac{1}{\theta\sqrt{m}}\right) \end{pmatrix},\end{aligned}$$

$$\begin{aligned}\begin{pmatrix} \hat{\hat{\theta}}_X \\ \hat{\hat{\theta}}_Y \\ \langle \hat{u}_X, \hat{u}_Y\rangle^2 \end{pmatrix} \sim \mathbf{N}\Bigg(& \begin{pmatrix} \theta + O_p(1) \\ \theta + O_p(1) \\ 1 + \frac{2-M_{2,X}-M_{2,Y}}{\theta} + O_p\left(\frac{1}{\theta^2}\right) \end{pmatrix}, \\ & \frac{1}{m}\begin{pmatrix} 2\theta^2(1-M_{2,X}) & 0 & 2\left(2M_{2,X}^2 - M_{2,X} - M_{3,X}\right) \\ 0 & 2\theta^2(1-M_{2,Y}) & 2\left(2M_{2,Y}^2 - M_{2,Y} - M_{3,Y}\right) \\ 2\left(2M_{2,X}^2 - M_{2,X} - M_{3,X}\right) & 2\left(2M_{2,Y}^2 - M_{2,Y} - M_{3,Y}\right) & \frac{S}{\theta^2} \end{pmatrix}\Bigg) + \begin{pmatrix} o_p\left(\frac{\theta}{\sqrt{m}}\right) \\ o_p\left(\frac{\theta}{\sqrt{m}}\right) \\ o_p\left(\frac{1}{\theta\sqrt{m}}\right) \end{pmatrix},\end{aligned}$$

$$S = 2\left(4M_{2,X}^3 - M_{2,X}^2 - 4M_{2,X}M_{3,X} + M_{4,X}\right) + 2\left(4M_{2,Y}^3 - M_{2,Y}^2 - 4M_{2,Y}M_{3,Y} + M_{4,Y}\right) + 4(M_{2,Y}-1)(M_{2,X}-1).$$

Moreover, the asymptotic distributions of $\hat{\theta}_X$ *and* $\hat{\hat{\theta}}_X$ *are the same.*

3. *If* $\frac{\theta}{\sqrt{m}} \to d$*, a finite constant, then a mixture of the two first scenarios describes the first two moments of the joint distribution.*
   The formula of the second moment is asymptotically the same as the variance formula when $\frac{\theta}{\sqrt{m}} \to \infty$.
   The formula of the first moment is asymptotically the same as the expectation formula when $\frac{\theta}{\sqrt{m}} \to 0$.

4. *The random variables can be expressed as functions of invariant unit random statistics of the form:*

$$\overset{u}{M}_{r,s,X}(\rho) = \sum_{i=1}^{m} \frac{\hat{\lambda}_{W_X,i}^r}{\left(\rho - \hat{\lambda}_{W_X,i}\right)^s} \langle \hat{u}_{W_X,i}, u\rangle^2.$$

(Assuming a canonical perturbation leads to a simpler formula)

- *Exact distributions:*

$$\hat{\theta}_X \left| \frac{1}{\theta - 1} = \overset{u}{M}_{1,1,X}\left(\hat{\theta}_X\right), \right.$$
$$\langle \hat{u}_X, e_1 \rangle^2 = \frac{\theta}{(\theta-1)^2} \frac{1}{\hat{\theta}_X \overset{u}{M}_{1,2,X}\left(\hat{\theta}_X\right)}.$$

Moreover,

$$\begin{aligned}
\langle \hat{u}_X, \hat{u}_Y \rangle &= \langle \hat{u}_X, e_1 \rangle \langle \hat{u}_Y, e_1 \rangle + \sqrt{1 - \langle \hat{u}_X, e_1 \rangle^2} \sqrt{1 - \langle \hat{u}_Y, e_1 \rangle^2} Z, \\
\sum_{i=2}^{m} \hat{u}_{X,1} \hat{u}_{Y,1} &= \sqrt{1 - \langle \hat{u}_X, e_1 \rangle^2} \sqrt{1 - \langle \hat{u}_Y, e_1 \rangle^2} Z, \\
Z &\sim \mathbf{N}\left(0, \frac{1}{m}\right) + O_p\left(\frac{1}{m}\right),
\end{aligned}$$

where Z is independent of $\langle \hat{u}_X, e_1 \rangle$, $\langle \hat{u}_Y, e_1 \rangle$, $\hat{\theta}_X$ and $\hat{\theta}_Y$. In order to get the exact distribution, we should replace Z by $\sum_{i=1}^{m-1} v_i \tilde{v}_i$ where $v_i \tilde{v}_i$ are independent unit invariant random vectors.

- *Approximations:*

$$\begin{aligned}
\hat{\theta}_X &= \rho + \frac{\left(\overset{u}{M}_{1,1,X}(\rho) - M_{1,1,X}(\rho)\right)}{M_{1,2,X}(\rho)} + O_p\left(\frac{\theta}{m}\right) \\
&= \theta \overset{u}{M}_{1,X} + O_p(1), \\
\hat{\hat{\theta}}_X &= \theta + (\theta-1)^2 \left(\overset{u}{M}_{1,1,X}(\rho) - M_{1,1,X}(\rho_X)\right) + O_p\left(\frac{\theta}{m}\right).
\end{aligned}$$

We provide three methods of estimation of the angle in order to estimate it for all θ.

$$\begin{aligned}
\langle \hat{u}_X, e_1 \rangle^2 &= \frac{\theta}{(\theta-1)^2} \left(\frac{1}{\rho_X M_{1,2,X}(\rho_X)} + \left(\frac{2 M_{1,3,X}(\rho_X)}{M_{1,2,X}(\rho_X)} - \frac{1}{\rho_X} \right) \frac{\overset{u}{M}_{1,1,X}(\rho_X) - \frac{1}{\theta-1}}{\left(M_{1,2,X}(\rho_X)\right)^2} \right. \\
&\qquad \left. - \frac{\overset{u}{M}_{1,2,X}(\rho_X) - M_{1,2,X}(\rho_X)}{\rho_X \left(M_{1,2,X}(\rho_X)\right)^2} \right) + O_p\left(\frac{1}{m}\right), \\
&= 1 + \frac{1}{\theta}\left(1 - \overset{u}{M}_{2,X} + 2 M_{2,X}\left(\overset{u}{M}_{1,X} - 1\right)\right) \\
&\qquad + \frac{1}{\theta^2}\left(1 - 2\overset{u}{M}_{2,X} + 3\overset{u}{M}{}^2_{2,X} - 2\overset{u}{M}_{3,X}\right) + O_p\left(\frac{1}{\theta^3}\right) + O_p\left(\frac{1}{\theta m}\right), \\
&= 1 + \frac{1}{\theta} - \frac{1}{\theta}\overset{u}{M}_{2,X} + \frac{2}{\theta} M_{2,X}\left(\overset{u}{M}_{1,X} - 1\right) + O_p\left(\frac{1}{\theta^2}\right) + O_p\left(\frac{1}{\theta m}\right)
\end{aligned}$$

Finally, the double angle is such that

$$\langle \hat{u}_X, \hat{u}_Y \rangle = \langle \hat{u}_X, e_1 \rangle \langle \hat{u}_Y, e_1 \rangle + \frac{\sqrt{M_{2,X} - 1}\sqrt{M_{2,Y} - 1}}{\theta} Z + O_p\left(\frac{1}{\theta^2 \sqrt{m}}\right).$$

Remark 5.3.1.1.
If the spectra of W_X and W_Y tend sufficiently fast to a Marcenko-Pastur distribution of

parameter c we have,

$$\begin{pmatrix} \hat{\hat{\theta}}_X \\ \langle \hat{u}_X, u_0 \rangle^2 \end{pmatrix} \overset{Asy}{\sim} \mathbf{N}\left(\begin{pmatrix} \theta \\ \frac{1-\frac{c}{(\theta-1)^2}}{1+\frac{c}{\theta-1}} \end{pmatrix}, \frac{1}{m} \begin{pmatrix} -\frac{2c(\theta-1)^2\theta^2}{c-(\theta-1)^2} & -\frac{2c^2(\theta-1)\theta^3}{(c-(\theta-1)^2)(c+\theta-1)^2} \\ -\frac{2c^2(\theta-1)\theta^3}{(c-(\theta-1)^2)(c+\theta-1)^2} & -\frac{2c^2\theta^2\left(c^2+(\theta(\theta+2)-2)c+(\theta-1)^2\right)}{(c-(\theta-1)^2)(c+\theta-1)^4} \end{pmatrix} \right)$$

and

$$\begin{pmatrix} \hat{\hat{\theta}}_X \\ \hat{\hat{\theta}}_Y \\ \langle \hat{u}_X, \hat{u}_Y \rangle^2 \end{pmatrix} \overset{Asy}{\sim} \mathbf{N}\left(\begin{pmatrix} \theta \\ \theta \\ \left(\frac{1-\frac{c}{(\theta-1)^2}}{1+\frac{c}{\theta-1}}\right)^2 \end{pmatrix}, \right.$$

$$\left. \frac{1}{m} \begin{pmatrix} -\frac{2c(\theta-1)^2\theta^2}{c-(\theta-1)^2} & 0 & -\frac{2c^2(\theta-1)\theta^3}{\left(c-(\theta-1)^2\right)(c+\theta-1)^2} \\ 0 & -\frac{2c(\theta-1)^2\theta^2}{c-(\theta-1)^2} & -\frac{2c^2(\theta-1)\theta^3}{\left(c-(\theta-1)^2\right)(c+\theta-1)^2} \\ -\frac{2c^2(\theta-1)\theta^3}{\left(c-(\theta-1)^2\right)(c+\theta-1)^2} & -\frac{2c^2(\theta-1)\theta^3}{\left(c-(\theta-1)^2\right)(c+\theta-1)^2} & \frac{4c^2\theta^2\left(c-(\theta-1)^2\right)^2\left(c^3+4c^2(\theta-1)+c(\theta-1)(\theta(\theta+5)-5)+2(\theta-1)^3\right)}{(\theta-1)^4(c+\theta-1)^7} \end{pmatrix} \right).$$

If θ tends to infinity, then

$$\begin{pmatrix} \hat{\hat{\theta}}_X \\ \langle \hat{u}_X, u \rangle^2 \end{pmatrix} \overset{Asy}{\sim} \mathbf{N}\left(\begin{pmatrix} \theta \\ \frac{1-\frac{c}{(\theta-1)^2}}{1+\frac{c}{\theta-1}} \end{pmatrix}, \frac{1}{m} \begin{pmatrix} 2c\theta^2 & 2c^2 \\ 2c^2 & \frac{2c^2(c+1)}{\theta^2} \end{pmatrix} \right).$$

Moreover,

$$\begin{pmatrix} \hat{\hat{\theta}}_X \\ \hat{\hat{\theta}}_Y \\ \langle \hat{u}_X, \hat{u}_X \rangle^2 \end{pmatrix} \overset{Asy}{\sim} \mathbf{N}\left(\begin{pmatrix} \theta \\ \theta \\ \frac{1-\frac{c}{(\theta-1)^2}}{1+\frac{c}{\theta-1}} \end{pmatrix}, \frac{1}{m} \begin{pmatrix} 2c\theta^2 & 0 & 2c^2 \\ 0 & 2c\theta^2 & 2c^2 \\ 2c^2 & 2c^2 & \frac{4c^2(c+2)}{\theta^2} \end{pmatrix} \right).$$

(Proof page 110)

5.4 Invariant Eigenvalue Theorem

The previous section provides results for perturbations of order 1. This section shows the invariant behaviour of the estimated eigenvalue as a function of the orders of the perturbation.

Theorem 5.4.1. *(Invariant Eigenvalue Theorem)*
Suppose that W respects Assumption 2.2.1 and

$$\tilde{P}_s = \mathrm{I}_m + (\theta_s - 1)e_s e_s^t, \text{ for } s = 1, 2, ..., k,$$
$$P_k = \mathrm{I}_m + \sum_{i=1}^{k} (\theta_i - 1)e_i e_i^t \text{ satisfies 2.2.2 (A4),}$$

where $\theta_1 > \theta_2 > ... > \theta_k$. We define

$$\hat{\Sigma}_{\tilde{P}_s} = \tilde{P}_s^{1/2} W \tilde{P}_s^{1/2},$$
$$\hat{\Sigma}_{P_k} = P_k^{1/2} W P_k^{1/2}.$$

Moreover, for $s = 1, 2, ..., k$, we define

$$\hat{u}_{\tilde{P}_s,1}, \hat{\theta}_{\tilde{P}_s,1} \quad s.t. \quad \hat{\Sigma}_{\tilde{P}_1} \hat{u}_{\tilde{P}_s,1} = \hat{\theta}_{\tilde{P}_s,1} \hat{u}_{\tilde{P}_s,1},$$
$$\hat{u}_{P_k,s}, \hat{\theta}_{P_k,s} \quad s.t. \quad \hat{\Sigma}_{P_k} \hat{u}_{P_k,s} = \hat{\theta}_{P_k,s} \hat{u}_{P_k,s},$$

where $\hat{\theta}_{\tilde{P}_s,1} = \hat{\lambda}_{\hat{\Sigma}_{\tilde{P}_s},1}$ and $\hat{\theta}_{P_k,s} = \hat{\lambda}_{\hat{\Sigma}_{P_k},s}$.

1. *Then, for $s > 1$,*

$$\boxed{\hat{\theta}_{P_k,s} - \hat{\theta}_{\tilde{P}_s,1} \sim \frac{\theta_s}{m}}$$

and

$$\boxed{\hat{\theta}_{P_k,1} - \hat{\theta}_{\tilde{P}_1,1} \sim \frac{\theta_2}{m} + \frac{\theta_2}{m^{3/2}},}$$

where θ_1 and θ_2 are the largest and the second largest eigenvalue, respectively.
The distribution of $\hat{\theta}_{P_k,s}$ is therefore asymptotically the same as the distribution of $\hat{\theta}_{\tilde{P}_s,1}$ studied in Theorem 5.3.1.

2. *More precisely we define for $r, s \in \{1, 2, ..., k\}$ with $r \neq s$,*

$$P_{-r} = \mathrm{I}_m + \sum_{\substack{i=1 \\ i \neq r}}^{k} (\theta_i - 1)\, e_i e_i^t.$$

- *If $\theta_s > \theta_r$, then*

$$\hat{\theta}_{P_k,s} - \hat{\theta}_{P_{-r},s} = -\frac{\hat{\theta}_{P_{-r},s}\hat{\theta}_{P_k,s}(\theta_r - 1)}{\theta_r - 1 - \hat{\theta}_{P_k,s}} \hat{u}^2_{P_{-r},s,r} + O_p\left(\frac{1}{m}\right) + O_p\left(\frac{\theta_r}{m^{3/2}}\right).$$

- *If $\theta_s < \theta_r$, then*

$$\hat{\theta}_{P_k,s} - \hat{\theta}_{P_{-r},s-1} = -\frac{\hat{\theta}_{P_{-r},s-1}\hat{\theta}_{P_k,s}(\theta_r - 1)}{\theta_r - 1 - \hat{\theta}_{P_k,s}} \hat{u}^2_{P_{-r},s-1,r} + O_p\left(\frac{1}{m}\right) + O_p\left(\frac{\theta_s}{m^{3/2}}\right).$$

Remark 5.4.1.1.
In this work, we are interested by the unbiased estimation of $\hat{\hat{\theta}}_{P_k,1}$. The invariance of $\hat{\hat{\theta}}_{P_k,1}$ is a direct consequence of the theorem. Moreover, Theorem 5.3.1 provides the distribution of $\hat{\hat{\theta}}_{P_1,1}$.

(Proof page 121 for $k = 2$ and 123 for all k.)

5.5 Invariant Angle Theorem

In this section, we introduce the most important theorem of this thesis, the Invariant Angle Theorem. The usual angle between two vectors is $\langle u, v\rangle$. Here we define the general angle between a vector and a subspace of dimension k as $\sum_{i=1}^{k} \langle u, v_i\rangle^2$, where $(v_1, ..., v_k)$ is a orthonormal basis of the subspace. This generalization of the angle used with the correct subspace leads to an invariance.

Theorem 5.5.1. *(Invariant Angle Theorem)*
Suppose that W satisfies Assumption 2.2.1 and

$$\tilde{P}_s = \mathrm{I}_m + (\theta_s - 1)e_s e_s^t, \text{ for } s = 1, 2, ..., k,$$
$$P_k = \mathrm{I}_m + \sum_{i=1}^{k} (\theta_i - 1)e_i e_i^t \text{ respects 2.2.2 (A4)},$$

where $\theta_1 > \theta_2 > ... > \theta_k$. We define

$$\hat{\Sigma}_{\tilde{P}_s} = \tilde{P}_s^{1/2} W \tilde{P}_s^{1/2},$$
$$\hat{\Sigma}_{P_k} = P_k^{1/2} W P_k^{1/2}.$$

Moreover, for $s = 1, 2, ..., k$, we define

$$\hat{u}_{\tilde{P}_s,1}, \hat{\theta}_{\tilde{P}_s,1} \quad s.t. \quad \hat{\Sigma}_{\tilde{P}_s} \hat{u}_{\tilde{P}_s,1} = \hat{\theta}_{\tilde{P}_s,1} \hat{u}_{\tilde{P}_s,1},$$
$$\hat{u}_{P_k,s}, \hat{\theta}_{P_k,s} \quad s.t. \quad \hat{\Sigma}_{P_k} \hat{u}_{P_k,s} = \hat{\theta}_{P_k,s} \hat{u}_{P_k,s},$$

where $\hat{\theta}_{\tilde{P}_s,1} = \hat{\lambda}_{\hat{\Sigma}_{\tilde{P}_s,1}}$ and $\hat{\theta}_{P_k,s} = \hat{\lambda}_{\hat{\Sigma}_{P_k},s}$.

1. Then,

$$\boxed{\sum_{i=1}^{k} \hat{u}^2_{P_k,s,i} = \hat{u}^2_{\tilde{P}_s,1,s} + O_p\left(\frac{1}{\theta_s m}\right).}$$

Therefore, the distribution of $\sum_{i=1}^{k} \hat{u}^2_{P_k,s,i}$ is asymptotically the same as the distribution of $\hat{u}^2_{\tilde{P}_s,1,s}$ studied in Theorem 5.3.1.

2. Moreover,

$$\hat{u}^2_{P_k,s,s} = \hat{u}^2_{\tilde{P}_s,1,s} + O_p\left(\frac{1}{m}\right).$$

Remark 5.5.1.1.

1. If

$$\hat{u}^2_{P_1,1,1} \sim \mathbf{N}\left(\alpha^2, \frac{\sigma^2_{\alpha^2}}{\theta_1^2 m}\right) + o_p\left(\frac{1}{\theta_1\sqrt{m}}\right),$$

then

$$\sum_{i=1}^{k} \hat{u}^2_{P_k,1,i} \sim \mathbf{N}\left(\alpha^2, \frac{\sigma^2_{\alpha^2}}{\theta_1^2 m}\right) + o_p\left(\frac{1}{\theta_1\sqrt{m}}\right),$$

where $\alpha^2 = 1 - \frac{M_2-1}{\theta_1} + O_p\left(\frac{1}{\theta_1^2}\right) < 1$ can be computed more precisely as in Theorem 5.3.1.

2. Assuming a Marcenko-Pastur spectrum, $\alpha^2 = \frac{1-\frac{c}{(\theta_1-1)^2}}{1+\frac{c}{\theta_1-1}}$ and $\sigma^2_{\alpha^2} = 2c^2(c+1) + o_\theta(1)$. In particular if $\frac{\theta_1}{\sqrt{m}}$ is large, then $\alpha^2 \approx 1 - c/\theta_1$,

3. In the general case, if $\frac{\theta_1}{\sqrt{m}}$ is large,

$$\alpha^2 \approx 1 + \frac{1 - M_{2,X}}{\theta_1} \text{ and } \sigma^2_{\alpha^2} \approx 2\left(4M^3_{2,X} - M^2_{2,X} - 4M_{2,X}M_{3,X} + M_{4,X}\right).$$

(Proof page 132)

5.6 Asymptotic distribution of the dot product

In this section, we compute the distribution of a dot product used to compute the distributions of the residual spikes.

Theorem 5.6.1. *(Dot Product Theorem)*
Suppose that W respects Assumption 2.2.1 and $P_2 = \mathrm{I}_m + \sum_{i=1}^{2}(\theta_i - 1)e_i e_i^t$ with $\theta_1 > \theta_2$. We define

$$\hat{\Sigma}_{P_2} = P_2^{1/2} W P_2^{1/2} \text{ and } \hat{\Sigma}_{P_1} = P_1^{1/2} W P_1^{1/2}.$$

Moreover, for $s, k = 1, 2$ and $s \leqslant k$, we define

$$\hat{u}_{P_k,s}, \hat{\theta}_{P_k,s} \quad s.t. \quad \hat{\Sigma}_{P_k}\hat{u}_{P_k,s} = \hat{\theta}_{P_k,s}\hat{u}_{P_k,s},$$

where $\hat{\theta}_{P_k,s} = \hat{\lambda}_{\hat{\Sigma}_{P_k},s}$. Finally the present theorem uses the convention:

$$\text{For } s = 1, 2, ..., k \text{ and } i = 1, 2, ..., s,\ \hat{u}_{P_s,i,i} > 0.$$

1. *Assuming that Assumptions 2.2.2(A2) and (A3) ($\theta_i = p_i\theta \to \infty$) hold,*

$$\begin{aligned}\sum_{s=3}^{m} \hat{u}_{P_2,1,s}\hat{u}_{P_2,2,s} &= \hat{u}_{P_2,1,2}\left(\frac{1}{\theta_1}-\frac{1}{\theta_2}\right) - \frac{1}{\theta_2^{1/2}}\sum_{j>1}^{m}\hat{\lambda}_{P_1,j}\hat{u}_{P_1,j,1}\hat{u}_{P_1,j,2} \\ &\quad + O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{1/2}m}\right) + O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{3/2}m^{1/2}}\right) \\ &= \frac{-\left(1+M_2\right)W_{1,2} + \left(W^2\right)_{1,2}}{\sqrt{\theta_1\theta_2}} + O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{1/2}m}\right) + O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{3/2}m^{1/2}}\right).\end{aligned}$$

Thus, we can estimate the distribution conditioned on the spectrum of W,

$$\begin{aligned}\sum_{s=3}^{m} \hat{u}_{P_2,1,s}\hat{u}_{P_k,2,s} &\sim \mathbf{N}\left(0, \frac{\left(1+M_2\right)^2\left(M_2-1\right) + \left(M_4-(M_2)^2\right) - 2\left(1+M_2\right)\left(M_3-M_2\right)}{\theta_1\theta_2 m}\right) \\ &\quad + O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{1/2}m}\right) + O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{3/2}m^{1/2}}\right).\end{aligned}$$

2. *If $\theta_1 \to \infty$ and θ_2 is finite, then*

$$\sum_{s=3}^{m} \hat{u}_{P_2,1,s}\hat{u}_{P_2,2,s} = O_p\left(\frac{1}{\sqrt{\theta_1 m}}\right).$$

3. *If θ_1 and θ_2 are finite, then*

$$\sum_{s=3}^{m} \hat{u}_{P_k,1,s}\hat{u}_{P_k,2,s} = O_p\left(\frac{1}{\sqrt{m}}\right).$$

Remark 5.6.1.1.

1. We can easily show

$$\begin{aligned}&\hat{u}_{P_2,1,2}\left(\frac{1}{\theta_1}-\frac{1}{\theta_2}\right)\delta + \sum_{s=3}^{m}\hat{u}_{P_2,1,s}\hat{u}_{P_2,2,s} \\ &= \frac{-\left(\delta+M_2\right)W_{1,2} + \left(W^2\right)_{1,2}}{\sqrt{\theta_1\theta_2}} + O_p\left(\frac{1}{\theta m}\right) + O_p\left(\frac{1}{\theta^2 m^{1/2}}\right) \\ &\sim \mathbf{N}\left(0, \frac{\left(\delta+M_2\right)^2\left(M_2-1\right) + \left(M_4-(M_2)^2\right) - 2\left(\delta+M_2\right)\left(M_3-M_2\right)}{\theta_1\theta_2 m}\right) \\ &\quad + O_p\left(\frac{1}{\theta m}\right) + O_p\left(\frac{1}{\theta^2 m^{1/2}}\right).\end{aligned}$$

2. If W is a standard Wishart random matrix, then Assumptions 2.2.2(A2) and (A3) lead to a Marcenko-Pastur spectrum and

$$\sum_{s=3}^{m} \hat{u}_{P_2,1,s}\hat{u}_{P_2,2,s} \sim \mathbf{N}\left(0, \frac{(1-\alpha_1^2)(1-\alpha_2^2)}{m}\right) + o_p\left(\frac{1}{\theta\sqrt{m}}\right),$$

where $\alpha_s^2 = \lim_{m\to\infty} \sum_{i=1}^{2} \langle \hat{u}_{P_2,s}, u_i\rangle^2$.

(Proof page 142)

5.7 Invariant Dot Product Theorem

Theorem 5.7.1. *(Invariant Dot Product Theorem)*
Suppose that W satisfies Assumption 2.2.1 and

$$P_{s,r} = \mathrm{I}_m + \sum_{i=s,r}^{2} (\theta_i - 1) e_i e_i^t$$

$$P_k = \mathrm{I}_m + \sum_{i=1}^{k} (\theta_i - 1) e_i e_i^t \text{ respects 2.2.2 (A4),}$$

where $\theta_1 > \theta_2 > ... > \theta_k$. We define

$$\hat{\Sigma}_{P_{s,r}} = P_{s,r}^{1/2} W P_{s,r}^{1/2},$$
$$\hat{\Sigma}_{P_k} = P_k^{1/2} W P_k^{1/2}.$$

Moreover, for $s, r = 1, 2, ..., k$ with $s \neq r$, we define

$$\begin{aligned} \hat{u}_{P_{s,r},1}, \hat{\theta}_{P_{s,r},1} \quad & s.t. \quad \hat{\Sigma}_{P_{s,r}} \hat{u}_{P_{s,r},1} = \hat{\theta}_{P_{s,r},1} \hat{u}_{P_{s,r},1}, \\ \hat{u}_{P_k,s}, \hat{\theta}_{P_k,s} \quad & s.t. \quad \hat{\Sigma}_{P_k} \hat{u}_{P_k,s} = \hat{\theta}_{P_k,s} \hat{u}_{P_k,s}, \end{aligned}$$

where $\hat{\theta}_{P_{s,r},1} = \hat{\lambda}_{\hat{\Sigma}_{P_{s,r},1}}$ and $\hat{\theta}_{P_k,s} = \hat{\lambda}_{\hat{\Sigma}_{P_k},s}$.
Assuming the convention:

$$\text{For } s = 1, 2, ..., k \text{ and } i = 1, 2, ..., s,\ \hat{u}_{P_s,i,i} > 0.$$

leads to

$$\boxed{\sum_{\substack{i=1 \\ i \neq s,r}}^{m} \hat{u}_{P_{s,r},1,i} \hat{u}_{P_{s,r},2,i} = \sum_{i=k+1}^{m} \hat{u}_{P_k,s,i} \hat{u}_{P_k,r,i} + O_p\left(\frac{1}{\sqrt{\theta_s \theta_r} m}\right).}$$

(Proof page 150)

5.8 Component Distribution Theorem

Theorem 5.8.1.
Suppose Assumption 2.2.1 holds with canonical P and 2.2.2(A4). We define:

$$U = \begin{pmatrix} \hat{u}_{P_k,1}^t \\ \hat{u}_{P_k,2}^t \\ \vdots \\ \hat{u}_{P_k,m}^t \end{pmatrix} = \begin{pmatrix} \hat{u}_{P_k,1:k,1:k} & \hat{u}_{P_k,1:k,k+1:m} \\ \hat{u}_{P_k,k+1:m,1:k} & \hat{u}_{P_k,k+1:m,k+1:m}. \end{pmatrix}$$

To simplify the result we assume the sign convention,

$$\text{For } s = 1, 2, ..., k \text{ and } i = 1, 2, ..., s,\ \hat{u}_{P_s,i,i} > 0.$$

1. *Without loss of generality on the k first components, the k^{th} element of the first eigenvector is*

$$\begin{aligned} \hat{u}_{P_k,1,k} &= \frac{\sqrt{\theta_k}\theta_1}{|\theta_k - \theta_1|} \hat{u}_{P_{k-1},1,k} + O_p\left(\frac{\min(\theta_1, \theta_k)}{\theta_1^{1/2}\theta_k^{1/2} m}\right) + O_p\left(\frac{1}{\sqrt{\theta_1 \theta_k m}}\right) \\ &= \frac{\theta_1 \sqrt{\theta_k}}{|\theta_k - \theta_1|} \frac{1}{m} \sqrt{1 - \hat{\alpha}_1^2}\ Z + O_p\left(\frac{\min(\theta_1, \theta_k)}{\theta_1^{1/2}\theta_k^{1/2} m}\right) + O_p\left(\frac{1}{\sqrt{\theta_1 \theta_k m}}\right) \\ &= \frac{\sqrt{\theta_1 \theta_k}}{|\theta_k - \theta_1|} \frac{1}{\sqrt{m}} \sqrt{M_2 - 1}\ Z + O_p\left(\frac{\min(\theta_1, \theta_k)}{\theta_1^{1/2}\theta_k^{1/2} m}\right) + O_p\left(\frac{1}{\sqrt{\theta_1 \theta_k m}}\right), \end{aligned}$$

where Z is a standard normal and $M_2 = \frac{1}{m}\sum_{i=1}^m \hat{\lambda}_{W,i}^2$ is obtained by conditioning on the spectrum.

- *Therefore knowing the spectrum and assuming $\theta_1, \theta_k \to \infty$,*

$$\hat{u}_{P_k,1,k} \overset{Asy}{\sim} \mathbf{N}\left(0, \frac{\theta_1\theta_k}{|\theta_1 - \theta_k|}\frac{M_2 - 1}{m}\right).$$

- *If $\theta_1 \to \infty$ and θ_k is finite,*

$$\hat{u}_{P_k,1,k} = O_p\left(\frac{1}{\sqrt{\theta_1 m}}\right).$$

- *Therefore, assuming θ_1 and θ_k are finite,*

$$\hat{u}_{P_k,1,k} = O_p\left(\frac{1}{\sqrt{m}}\right).$$

This result holds for any component $\hat{u}_{P_k,s,t}$ where $s \neq t \in \{1, 2, ..., k\}$.

Remark 5.8.1.1.
The sign of $\hat{u}_{P_k,1,k}$ obtained by construction using Theorem 5.11.1 is always positive. By convention ($\hat{u}_{P_k,i,i} > 0$, for $i = 1, 2, ..., k$), we multiply by $\text{sign}\left(\hat{u}_{P_k,1,1}\right)$ obtained in the construction. Therefore, the remark of Theorem 5.11.1 describes the sign of the component assuming the convention.

$$P\left\{\text{sign}\left(\hat{u}_{P_k,1,k}\right) = \text{sign}\left(\left(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1}\right)\hat{u}_{P_{k-1},1,k}\hat{u}_{P_{k-1},1,1}\right)\right\} = 1 + O\left(\frac{1}{m}\right).$$

2. *For $s = 1, ..., k$, the vector $\frac{\hat{u}_{s,k+1:m}}{\sqrt{1-\hat{\alpha}_s^2}}$, where $\hat{\alpha}_s^2 = \sum_{i=1}^k \hat{u}_{i,s}^2$, is unit invariant by rotation. Moreover, for $j > k$,*

$$\hat{u}_{j,s} \sim \mathbf{N}\left(0, \frac{1 - \alpha_s^2}{m}\right),$$

where α_s^2 is the limit of $\hat{\alpha}_s^2$.
Moreover, the columns of $U^t[k+1:m, k+1:m]$ are invariant by rotation.

3. *Assuming $P_k = \mathrm{I}_m + \sum_{i=1}^k (\theta_i - 1)\epsilon_i\epsilon_i^t$ is such that*

$$\begin{aligned}&\theta_1, \theta_2, ..., \theta_{k_1} \text{ are proportional, and}\\&\theta_{k_1+1}, \theta_{k_1+2}, ..., \theta_k \text{ are proportional,}\end{aligned}$$

then

$$\begin{aligned}\sum \hat{u}_{k+1:m,1}^2 &< \sum \hat{u}_{k+1:m,1:k_1}^2\\ &\sim \mathrm{RV}\left(O\left(\frac{1}{\theta_1}\right), O\left(\frac{1}{\theta_1^2 m}\right)\right) + O_p\left(\frac{\min(\theta_1, \theta_k)}{\max(\theta_1, \theta_k)m}\right).\end{aligned}$$

Moreover, if P satisfies Assumption 2.2.2(A4) with $\min\left(\frac{\theta_1}{\theta_k}, \frac{\theta_k}{\theta_1}\right) \to 0$, then

$$\sum \hat{u}_{k+1:m,1}^2 \sim \mathrm{RV}\left(O\left(\frac{1}{\theta_1}\right), O\left(\frac{1}{\theta_1^2 m}\right)\right) + O_p\left(\frac{1}{\theta_1 m}\right).$$

(Proof page 122 for Part 1 with $k = 2$ and 120 for Part 2 and 3 for $k = 1$. For general k, proof page 130 for Part 1 and 136 for Part 2 and 3)

5.9 Invariant Component Theorem

Theorem 5.9.1. *(Invariant Component Theorem)*
Suppose that W satisfies Assumption 2.2.1 and

$$P_{s,r} = \mathrm{I}_m + \sum_{i=s,r}^{2} (\theta_i - 1) e_i e_i^t$$

$$P_k = \mathrm{I}_m + \sum_{i=1}^{k} (\theta_i - 1) e_i e_i^t \text{ respects 2.2.2 (A4),}$$

where $\theta_1 > \theta_2 > ... > \theta_k$. We define

$$\hat{\Sigma}_{P_{s,r}} = P_{s,r}^{1/2} W P_{s,r}^{1/2},$$
$$\hat{\Sigma}_{P_k} = P_k^{1/2} W P_k^{1/2}.$$

Moreover, for $s, r = 1, 2, ..., k$ with $s \neq r$, we define

$$\hat{u}_{P_{s,r},1}, \hat{\theta}_{P_{s,r},1} \quad s.t. \quad \hat{\Sigma}_{P_{s,r}} \hat{u}_{P_{s,r},1} = \hat{\theta}_{P_{s,r},1} \hat{u}_{P_{s,r},1},$$
$$\hat{u}_{P_k,s}, \hat{\theta}_{P_k,s} \quad s.t. \quad \hat{\Sigma}_{P_k} \hat{u}_{P_k,s} = \hat{\theta}_{P_k,s} \hat{u}_{P_k,s},$$

where $\hat{\theta}_{P_{s,r},1} = \hat{\lambda}_{\hat{\Sigma}_{P_{s,r},1}}$ and $\hat{\theta}_{P_k,s} = \hat{\lambda}_{\hat{\Sigma}_{P_k},s}$.
Therefore, for $\theta_s < \theta_r$,

$$\hat{u}_{P_k,r,s} = \hat{u}_{P_{r,s},1,s} + O_p\left(\frac{\theta_s^{1/2}}{\theta_r^{1/2} m}\right),$$

and

$$\hat{u}_{P_k,s,r} = \hat{u}_{P_{r,s},2,r} + O_p\left(\frac{\theta_s^{1/2}}{\theta_r^{1/2} m}\right).$$

(The proof of this theorem is similar to the other theorems using 5.11.1. Because it is not used in this thesis, it is left to the reader.)

5.10 Invariant Double Angle Theorem

The Double Angle Theorem and its Invariant Theorem are two main contributions of this thesis to Random matrix theory.

Corollary 5.10.1.
Suppose W_X and W_Y satisfies Assumption 2.2.1 and

$$\tilde{P}_s = \mathrm{I}_m + (\theta_s - 1) e_s e_s^t, \text{ for } s = 1, 2, ..., k,$$

$$P_k = \mathrm{I}_m + \sum_{i=1}^{k} (\theta_i - 1) e_i e_i^t \text{ respects 2.2.2 (A4),}$$

where $\theta_1 > \theta_2 > ... > \theta_k$. We define

$$\hat{\Sigma}_{X,\tilde{P}_s} = \tilde{P}_s^{1/2} W_X \tilde{P}_s^{1/2} \text{ and } \hat{\Sigma}_{X,\tilde{P}_s} = \tilde{P}_s^{1/2} W_Y \tilde{P}_s^{1/2},$$
$$\hat{\Sigma}_{X,P_k} = P_k^{1/2} W_X P_k^{1/2} \text{ and } \hat{\Sigma}_{Y,P_k} = P_k^{1/2} W_Y P_k^{1/2}.$$

Moreover, for $s = 1, ..., k$, we define

$$\begin{aligned}
&\hat{u}_{\hat{\Sigma}_{X,\tilde{P}_s},1}, \hat{\theta}_{\hat{\Sigma}_{X,\tilde{P}_s},1} \quad s.t. \quad \hat{\Sigma}_{X,\tilde{P}_s}\hat{u}_{\hat{\Sigma}_{X,\tilde{P}_s},1} = \hat{\theta}_{\hat{\Sigma}_{X,\tilde{P}_s},1}\hat{u}_{\hat{\Sigma}_{X,\tilde{P}_s},1},\\
&\hat{u}_{\hat{\Sigma}_{X,P_k},s}, \hat{\theta}_{\hat{\Sigma}_{X,P_k},s} \quad s.t. \quad \hat{\Sigma}_{X,P_k}\hat{u}_{\hat{\Sigma}_{X,P_k},s} = \hat{\theta}_{\hat{\Sigma}_{X,P_k},s}\hat{u}_{\hat{\Sigma}_{X,P_k},s},
\end{aligned}$$

where $\hat{\theta}_{\hat{\Sigma}_{X,\tilde{P}_s},1} = \hat{\lambda}_{\hat{\Sigma}_{X,\tilde{P}_s,1}}$ and $\hat{\theta}_{\hat{\Sigma}_{X,P_k},s} = \hat{\lambda}_{\hat{\Sigma}_{X,P_k},s}$. The statistics of the group Y are defined in analogous manner.
Then,

$$\begin{aligned}
\boxed{\left\langle \hat{u}_{\hat{\Sigma}_{X,\tilde{P}_s},1}, \hat{u}_{\hat{\Sigma}_{Y,\tilde{P}_s},1} \right\rangle^2 = \sum_{i=1}^{k} \left\langle \hat{u}_{\hat{\Sigma}_{X,P_k},s}, \hat{u}_{\hat{\Sigma}_{Y,P_k},s} \right\rangle^2 + O_p\left(\frac{1}{\theta_s m}\right)}\\
= \sum_{i=1}^{k+\epsilon} \left\langle \hat{u}_{\hat{\Sigma}_{X,P_k},s}, \hat{u}_{\hat{\Sigma}_{Y,P_k},i} \right\rangle^2 + O_p\left(\frac{1}{\theta_s m}\right),
\end{aligned}$$

where ϵ is a small integer.

Remark 5.10.0.1.

1. The procedure of the proof shows an interesting invariant:
   Assuming the sign convention $\hat{u}_{P_s,i,i} > 0$ for $s = 1, 2, ..., k$ and $i = 1, 2, ..., s$,
   $$\sum_{i=k+1}^{m} \hat{u}_{P_k,1,i}\hat{\hat{u}}_{P_k,1,i} = \sum_{i=k}^{m} \hat{u}_{P_{k-1},1,i}\hat{\hat{u}}_{P_{k-1},1,i} + O_p\left(\frac{1}{\theta_1 m}\right).$$
2. The distribution of $\left\langle \hat{u}_{\hat{\Sigma}_{X,P_1},1}, \hat{u}_{\hat{\Sigma}_{Y,P_1},1} \right\rangle^2$ is computed in Theorem 5.3.1.
3. An error of ϵ principal components does not affect the asymptotic distribution of the general double angle. This property allows us to construct a robust test.

(Proof page 156)

5.11 Tool Theorems

This section introduces three theorems. Two theorems are key of most of the proofs and the last one is used to prove the Main Theorem 3.1.1.
The first theorem gives a characterization of the eigenvalues and the eigenvectors of a random matrix. It provides more precise results than Theorem 5.2.1.
The second theorem gives the distribution of statistics using unit orthonormal random vectors.
The third theorem reveals an important quantity to be a functions of the invariant statistics.

5.11.1 Characterization of eigenstructure

Theorem 5.11.1. *(Characterization of eigenstructure)*
Using the same notation as in the Invariant Theorem (5.5.1, 5.4.1) and under Assumption 2.2.1 and 2.2.2(A4), we can compute the eigenvalues and the components of interest of the eigenvector of $\hat{\Sigma}_{P_k}$. Using assumption 2.2.1, we can without loss of generality suppose the canonical form for the perturbation P_k.

- *Eigenvalues :*

$$\underbrace{\sum_{i=k}^{m} \frac{\hat{\lambda}_{P_{k-1},i}}{\hat{\theta}_{P_k,s} - \hat{\lambda}_{P_{k-1},i}} \hat{u}^2_{P_{k-1},i,k}}_{(a) O_p\left(\frac{1}{\theta_s}\right)} + \underbrace{\frac{\hat{\theta}_{P_{k-1},s}}{\hat{\theta}_{P_k,s} - \hat{\theta}_{P_{k-1},s}} \hat{u}^2_{P_{k-1},s,k}}_{(b) \sim \left(\frac{\theta_k - \theta_s}{\theta_s \theta_k}\right)} + \underbrace{\sum_{\substack{i=1 \\ i \neq s}}^{k-1} \frac{\hat{\theta}_{P_{k-1},i}}{\hat{\theta}_{P_k,s} - \hat{\theta}_{P_{k-1},i}} \hat{u}^2_{P_{k-1},i,k}}_{(c) O_p\left(\frac{1}{\theta_s m}\right)} = \frac{1}{\theta_k - 1},$$

for $s = 1, 2, ..., k$.

Remark 5.11.1.1.

1. The formula is still valid without Assumption 2.2.1. However, the size estimations of each term are not necessarily accurate anmore.
2. If we do not assume canonical perturbations, then the formula is longer but the structure remains essentially the same. Assuming Assumption 2.2.1 leads matrices that are invariant under rotations. Elementary linear algebra methods extend the result to any perturbations.

- *Eigenvectors :*
We define $\tilde{u}_{P_k,i}$ *such that* $WP_k\tilde{u}_{P_k,i} = \hat{\theta}_{P_k,i}\tilde{u}_{P_k,i}$ *and* $\hat{u}_{P_k,i}$ *such that* $P_k^{1/2}WP_k^{1/2}\hat{u}_{P_k,i} = \hat{\theta}_{P_k,i}\hat{u}_{P_k,i}$. *To simplify notation we assume that* θ_i *corresponds to* $\hat{\theta}_{P_k,i}$. *This notation is explained in 7.1.1 and allows to describe the k first eigenvectors more efficiently.*

$$\left\langle \tilde{u}_{P_k,1}, e_1 \right\rangle^2$$
$$= \frac{\left(\overbrace{\sum_{i=k}^{m} \frac{\hat{\lambda}_{P_{k-1},i}}{\hat{\theta}_{P_k,1} - \hat{\lambda}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,1} \hat{u}_{P_{k-1},i,k}}^{(a) O_p\left(\frac{1}{\theta_1^{3/2}\sqrt{m}}\right)} + \overbrace{\frac{\hat{\theta}_{P_{k-1},1}}{\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1}} \hat{u}_{P_{k-1},1,1} \hat{u}_{P_{k-1},1,k}}^{(b) \sim \frac{\sqrt{\theta_1 m}}{\min(\theta_1,\theta_k)}} + \overbrace{\sum_{i=2}^{k-1} \frac{\hat{\theta}_{P_{k-1},i}}{\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,1} \hat{u}_{P_{k-1},i,k}}^{(c) O_p\left(\frac{1}{\theta_1^{1/2} m}\right)} \right)^2}{\underbrace{\sum_{i=k}^{m} \frac{\hat{\lambda}^2_{P_{k-1},i}}{(\hat{\theta}_{P_k,1} - \hat{\lambda}_{P_{k-1},i})^2} \hat{u}^2_{P_{k-1},i,k}}_{(d) O_p\left(\frac{1}{\theta_1^2}\right)} + \underbrace{\frac{\hat{\theta}^2_{P_{k-1},1}}{(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1})^2} \hat{u}^2_{P_{k-1},1,k}}_{(e) \sim \frac{\theta_1 m}{\min(\theta_1,\theta_k)^2}} + \underbrace{\sum_{i=2}^{k-1} \frac{\hat{\theta}^2_{P_{k-1},i}}{(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},i})^2} \hat{u}^2_{P_{k-1},i,k}}_{(f) O_p\left(\frac{1}{\theta_1 m}\right)}},$$

$$\left\langle \tilde{u}_{P_k,1}, e_k \right\rangle^2 = \frac{1}{D_1(\theta_k - 1)^2} (g),$$

$$\left\langle \tilde{u}_{P_k,1}, e_s \right\rangle^2$$
$$= \frac{1}{D_1} \left(\overbrace{\sum_{i=k}^{m} \frac{\hat{\lambda}_{P_{k-1},i}}{\hat{\theta}_{P_k,1} - \hat{\lambda}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,s} \hat{u}_{P_{k-1},i,k}}^{(h) O_p\left(\frac{1}{\theta_s^{1/2}\theta_1\sqrt{m}}\right)} + \overbrace{\frac{\hat{\theta}_{P_{k-1},1}}{\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1}} \hat{u}_{P_{k-1},1,s} \hat{u}_{P_{k-1},1,k}}^{(i) \sim \frac{\min(\theta_1,\theta_s)}{\sqrt{\theta_s}\min(\theta_1,\theta_k)}} \right.$$
$$\left. + \overbrace{\sum_{i=2,\neq s}^{k-1} \frac{\hat{\theta}_{P_{k-1},i}}{\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,s} \hat{u}_{P_{k-1},i,k}}^{(j) O_p\left(\max_{i\neq 1,s}\left(\frac{\min(\theta_1,\theta_i)\min(\theta_s,\theta_i)}{\sqrt{\theta_s}\theta_1\theta_i\sqrt{m}}\right)\right)} + \overbrace{\frac{\hat{\theta}_{P_{k-1},s}}{\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},s}} \hat{u}_{P_{k-1},s,s} \hat{u}_{P_{k-1},s,k}}^{(k) O_p\left(\frac{\min(\theta_1,\theta_s)}{\sqrt{\theta_s}\theta_1\sqrt{m}}\right)} \right)^2.$$

Finally,

$$\hat{u}_{P_k,1} = \frac{\left(\tilde{u}_{P_k,1,1}, \tilde{u}_{P_k,1,2}, ..., \sqrt{\theta_k}\tilde{u}_{P_k,1,k}, ..., \tilde{u}_{P_k,}\right)}{\underbrace{\sqrt{1 + (\theta_k - 1)\,\tilde{u}^2_{P_k 1,k}}}_{1 + O_p\left(\frac{\min(\theta_1,\theta_k)}{\max(\theta_1,\theta_k) m}\right)}},$$

where $\sqrt{1+(\theta-1)\,\tilde{u}^2_{P_k1,k}}$ *is the norm of* $P_k^{1/2}\tilde{u}_{P_k,1}$ *that we will call* N_1.

Remark 5.11.1.2.

1. By construction, the sign of $\hat{u}_{P_k,1,k}$ is always positive. This is, however, not the case of $\hat{u}_{P_{k-1},i,i}$. We can show that :
$$P\left\{\operatorname{sign}\left(\hat{u}_{P_k,1,1}\right)=\operatorname{sign}\left(\left(\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},1}\right)\hat{u}_{P_{k-1},1,1}\hat{u}_{P_{k-1},1,k}\right)\right\}\underset{m\to\infty}{\to}1.$$
Moreover, the convergence to 1 is of order $1/m$. If θ_1 tends to infinity, then
$$P\left\{\operatorname{sign}\left(\hat{u}_{P_k,1,1}\right)=\operatorname{sign}\left(\left(\theta_1-\theta_k\right)\hat{u}_{P_{k-1},1,k}\right)\right\}\underset{m,\theta_1\to\infty}{\to}1.$$
Therefore, if we use a convention such as $\operatorname{sign}\left(\hat{u}_{P_k,i,i}\right)>0$ for $i=1,...,k-1$, then the sign of $\hat{u}_{P_k,1,k}$ is distributed as a Bernoulli with parameter $1/2$.

2. Without loss of generality, the other eigenvector $\hat{u}_{P_k,r}$ for $r=1,2,...,k-1$ can be computed by the same formula thanks to the notation linking the estimated eigenvector to the eigenvalue θ_i.
However, the formula does not work for the vector $\hat{u}_{P_k,k}$. Applying a different order of perturbation shows that similar formulas exist for $\hat{u}_{P_k,k}$. (If the perturbation in e_1 is applied at the end for example.)
This observation leads to a problem in the proofs of the Dot Product Theorems 5.6.1 and 5.7.1. Deeper investigations are necessary to understand the two eigenvectors when $k=2$.
$$\begin{aligned}
D_2 &= \underbrace{\sum_{i=2}^{m}\frac{\hat{\lambda}^2_{P_1,i}}{(\hat{\theta}_{P_2,2}-\hat{\lambda}_{P_1,i})^2}\hat{u}^2_{P_1,i,2}}_{O_p\left(\frac{1}{\theta_2^2}\right)}+\underbrace{\frac{\hat{\theta}^2_{P_1,1}}{(\hat{\theta}_{P_2,2}-\hat{\theta}_{P_1,1})^2}\hat{u}^2_{P_1,1,2}}_{O_p\left(\frac{\theta_1}{(\theta_2-\theta_1)^2m}\right)},\\
N_2^2 &= 1+\frac{1}{(\theta_2-1)D_2},\\
N_2D_2 &= D_2+\frac{1}{\theta_2-1}\\
&= \frac{1}{\theta_2-1}+O_p\left(\frac{1}{\theta_2^2}\right)+O_p\left(\frac{\theta_1}{(\theta_2-\theta_1)m}\right).
\end{aligned}$$
Furthermore, the theorem must investigate the $m-k$ noisy components of the eigenvectors. For $r=1,2$ and $s=3,4,...,m$,
$$\hat{u}_{P_2,r,s}=\frac{\sum_{i=1}^{m}\frac{\hat{\lambda}_{P_1,i}}{\hat{\theta}_{P_2,r}-\hat{\lambda}_{P_1,i}}\hat{u}_{P_1,i,s}\hat{u}_{P_1,i,2}}{\sqrt{D_r}N_r}.$$
The estimations using this last formula are difficult. When we investigate these components, it is profitable to look at
$$\hat{u}_{P_2,1,t}\Big/\sqrt{\sum_{s=3}^{m}\hat{u}^2_{P_2,1,s}}\text{ and }\hat{u}_{P_2,2,t}\Big/\sqrt{\sum_{s=3}^{m}\hat{u}^2_{P_2,2,s}}$$
for $t=3,4,...,m$.

3. If the perturbation is not canonical, then we can apply a rotation U, such that $Uu_s=\epsilon_s$, and replace $\hat{u}_{P_{k-1},i}$ by $U^t\hat{u}_{P_{k-1},i}$. Then, $\langle\tilde{u}_{P_k,1},e_s\rangle^2$ is replaced by $\langle\tilde{u}_{P_k,1},u_s\rangle^2$.

(Proof page 122 for $k = 2$ and 125 for all k.)

5.11.2 Unit invariant vector statistic

Theorem 5.11.2.
Let W be a random matrix with spectrum, $\hat{\lambda}_{W,1}, \hat{\lambda}_{W,2}, ..., \hat{\lambda}_{W,m}$ such that its trace is 1. We denote by u_{p_1} and u_{p_2}, two orthonormal invariant random vectors independent of the eigenvalues of W. We set

$$\vec{B}_m(\rho, \vec{s}, \vec{r}, \vec{p}) = \sqrt{m}\left(\begin{pmatrix} \sum_{i=1}^m \frac{\hat{\lambda}_{W,i}^{s_1}}{\left(\rho-\hat{\lambda}_{W,i}\right)^{s_2}} u_{p_1,i}u_{p_2,i} \\ \sum_{i=1}^m \frac{\hat{\lambda}_{W,i}^{r_1}}{\left(\rho-\hat{\lambda}_{W,i}\right)^{r_2}} u_{p_1,i}u_{p_2,i} \end{pmatrix} - \begin{pmatrix} M_{s_1,s_2} \\ M_{r_1,r_2} \end{pmatrix} \mathbf{1}_{p_1=p_2}\right),$$

where $\vec{s} = (s_1, s_2)$, $\vec{r} = (r_1, r_2)$ and $\vec{p} = (p_1, p_2)$ with indices $1 \leqslant p_1 \leqslant p_2 \leqslant m$ and $s_1, s_2, r_1, r_2 \in \mathbf{N}$. Conditioning on the spectrum, if $p = p_1 = p_2$,

$$\vec{B}_m(\rho, \vec{s}, \vec{r}, \vec{p}) \sim \mathbf{N}\left(\vec{0}, \begin{pmatrix} 2\left(M_{2s_1,2s_2} - M_{s_1,s_1}^2\right) & 2\left(M_{s_1+r_1,s_2+r_2} - M_{s_1,s_2}M_{r_1,r_2}\right) \\ 2\left(M_{s_1+r_1,s_2+r_2} - M_{s_1,s_2}M_{r_1,r_2}\right) & 2\left(M_{2r_1,2r_2} - M_{r_1,r_1}^2\right) \end{pmatrix}\right) + o_{p;m}(1),$$

where $M_{s,r} = M_{s,r}(\rho) = \frac{1}{m}\sum_{i=1}^m \frac{\hat{\lambda}_{W,i}^s}{\left(\rho-\hat{\lambda}_{W,i}\right)^r}$.
Moreover, for $p_1 \neq p_2$,

$$\vec{B}_m(\rho, \vec{s}, \vec{r}, \vec{p}) \sim \mathbf{N}\left(\vec{0}, \begin{pmatrix} M_{2s_1,2s_2} - M_{s_1,s_1}^2 & M_{s_1+r_1,s_2+r_2} - M_{s_1,s_2}M_{r_1,r_2} \\ M_{s_1+r_1,s_2+r_2} - M_{s_1,s_2}M_{r_1,r_2} & M_{2r_1,2r_2} - M_{r_1,r_1}^2 \end{pmatrix}\right) + o_{p;m}(1),$$

In particular, with the notation $M_{s,0} = M_s = \frac{1}{m}\sum_{i=1}^m \hat{\lambda}_{W,i}^s$,

$$\sqrt{m}\left(\begin{pmatrix} \sum_{i=1}^m \hat{\lambda}_{W,i}u_{p,i}^2 \\ \sum_{i=1}^m \hat{\lambda}_{W,i}^2 u_{p,i}^2 \end{pmatrix} - \begin{pmatrix} 1 \\ M_2 \end{pmatrix}\right) \sim \mathbf{N}\left(\vec{0}, \begin{pmatrix} 2(M_2 - 1) & 2(M_3 - M_2) \\ 2(M_3 - M_2) & 2\left(M_4 - M_2^2\right) \end{pmatrix}\right) + o_{p;m}(1),$$

and

$$\sqrt{m}\left(\begin{pmatrix} \sum_{i=1}^m \hat{\lambda}_{W,i}u_{p_1,i}u_{p_2,i} \\ \sum_{i=1}^m \hat{\lambda}_{W,i}^2 u_{p_1,i}u_{p_2,i} \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) \sim \mathbf{N}\left(\vec{0}, \begin{pmatrix} M_2 - 1 & M_3 - M_2 \\ M_3 - M_2 & M_4 - M_2^2 \end{pmatrix}\right) + o_{p;m}(1).$$

Finally if we look at K bivariate normal random variables :

$$\mathbf{B}_m(\vec{\rho}, \mathbf{s}, \mathbf{r}, \mathbf{p}) = \left(\vec{B}_m(\rho_1, \vec{s}_1, \vec{r}_1, \vec{p}_1), \vec{B}_m(\rho_2, \vec{s}_2, \vec{r}_2, \vec{p}_2), ..., \vec{B}_m(\rho_K, \vec{s}_K, \vec{r}_K, \vec{p}_K)\right),$$

where ,$\vec{p}_i \neq \vec{p}_j$ if $i \neq j$. Then, $\tilde{\mathbf{B}}_m(\vec{\rho}, \mathbf{s}, \mathbf{r}, \mathbf{p})$ tends to a multivariate Normal. Moreover, all the bivariate elements are asymptotically independent.

Remark 5.11.2.1.

1. In order to obtain this simple result we assume the trace of W is one. This property can easily be obtained by rescaling the matrix by the trace.

2. The condition of independence between eigenvectors and eigenvalues of W is strong but is an automatic consequence if the eigenvectors are Haar distributed. Moreover, if the fluctuation of the spectra is small, the result is still valid without conditioning on the spectra. For example for independent vector data, we can easily show that the result still holds using Bai and Silverstein [2010] (page 259-261, Theorem 9.10).

3. If the convergence of the spectrum to f_S is fast, then we can define $M_i = \int \lambda^i f_S(\lambda)d\lambda$.

4. If the data are such that the spectrum of W tends to Marcenko-Pastur distribution

sufficiently fast, then f_S can be replaced by $f_{MP,c}$ and we obtain

$$\sqrt{m}\left(\begin{pmatrix}\sum_{i=1}^m \hat{\lambda}_{W,i}\hat{u}_i^2 \\ \sum_{i=1}^m \hat{\lambda}_{W,i}^2\hat{u}_i^2\end{pmatrix} - \begin{pmatrix}1 \\ 1+c\end{pmatrix}\right) \underset{m\to\infty}{\to} \mathbf{N}\left(\vec{0}, \begin{pmatrix}2c & 2c(2+c) \\ 2c(2+c) & 2c(c+1)(c+4)\end{pmatrix}\right).$$

Nevertheless the result is wrong if we only assume this spectral distribution! The independence with invariant eigenvectors is necessary.

(Proof page 98)

5.11.3 Double dot product

Theorem 5.11.3.
Suppose W_X and W_Y satisfies Assumption 2.2.1 and $P_k = \mathrm{I}_m + \sum_{i=1}^k(\theta_i - 1)e_ie_i^t$ satisfies 2.2.2 (A4), where $\theta_1 > \theta_2 > ... > \theta_k$. We set

$$\hat{\Sigma}_X = \hat{\Sigma}_{X,P_k} = P_k^{1/2}W_XP_k^{1/2} \text{ and } \hat{\Sigma}_{Y,P_k} = P_k^{1/2}W_YP_k^{1/2}.$$

and for $s = 1, ..., k$,

$$\begin{aligned} \hat{u}_{\hat{\Sigma}_X,s}, \hat{\theta}_{\hat{\Sigma}_X,s} \quad & s.t. \quad \hat{\Sigma}_X\hat{u}_{\hat{\Sigma}_X,s} = \hat{\theta}_{\hat{\Sigma}_X,s}\hat{u}_{\hat{\Sigma}_X,s}, \\ \hat{u}_{\hat{\Sigma}_Y,s}, \hat{\theta}_{\hat{\Sigma}_Y,s} \quad & s.t. \quad \hat{\Sigma}_Y\hat{u}_{\hat{\Sigma}_Y,s} = \hat{\theta}_{\hat{\Sigma}_Y,s}\hat{u}_{\hat{\Sigma}_Y,s}, \end{aligned}$$

where $\hat{\theta}_{\hat{\Sigma}_Y,s} = \hat{\lambda}_{\hat{\Sigma}_Y,s}$ and $\hat{\theta}_{\hat{\Sigma}_X,s} = \hat{\lambda}_{\hat{\Sigma}_X,s}$. To simplify the result we assume the sign convention:

For $s = 1, 2, ..., k$ and $i = 1, 2, ..., s$, $\hat{u}_{\hat{\Sigma}_X,i,i} > 0$, $\hat{u}_{\hat{\Sigma}_Y,i,i} > 0$.

Finally, we define

$$\tilde{u}_s = \hat{U}_X^t\hat{u}_{\hat{\Sigma}_Y,s},$$

where,

$$\hat{U}_X = (v_1, v_2, \cdots, v_m) = \left(\hat{u}_{\hat{\Sigma}_X,1}, \hat{u}_{\hat{\Sigma}_X,2}, \cdots \hat{u}_{\hat{\Sigma}_X,k}, v_{k+1}, v_{k+2}, \cdots, v_m\right),$$

where the vectors $v_{k+1}, ..., v_m$ are chosen such that the matrix $\hat{U}_X$ is orthonormal. Then,

- *If $\theta_j, \theta_t \to \infty$:*

$$\begin{aligned} \sum_{i=k+1}^m \tilde{u}_{j,i}\tilde{u}_{t,i} = & \sum_{i=k+1}^m \hat{u}_{\hat{\Sigma}_Y,j,i}\hat{u}_{\hat{\Sigma}_Y,t,i} + \sum_{i=k+1}^m \hat{u}_{\hat{\Sigma}_X,j,i}\hat{u}_{\hat{\Sigma}_X,t,i} - \sum_{i=k+1}^m \hat{u}_{\hat{\Sigma}_X,j,i}\hat{u}_{\hat{\Sigma}_Y,t,i} \\ & - \sum_{i=k+1}^m \hat{u}_{\hat{\Sigma}_Y,j,i}\hat{u}_{\hat{\Sigma}_X,t,i} - \left(\hat{u}_{\hat{\Sigma}_X,t,j} + \hat{u}_{\hat{\Sigma}_Y,j,t}\right)\left(\hat{\alpha}^2_{\hat{\Sigma}_X,j} - \hat{\alpha}^2_{\hat{\Sigma}_X,t}\right) \\ & + O_p\left(\frac{1}{\theta_1 m}\right) + O_p\left(\frac{1}{\theta_1^2\sqrt{m}}\right), \end{aligned}$$

  where $\hat{\alpha}^2_{\hat{\Sigma}_X,t} = \sum_{i=1}^k \hat{u}^2_{\hat{\Sigma}_X,t,i}$.

- *If $\theta_j \to \infty$ and θ_t is finite:*

$$\sum_{i=k+1}^m \tilde{u}_{j,i}\tilde{u}_{t,i} = O_p\left(\frac{1}{\sqrt{m}\sqrt{\theta_1}}\right).$$

- *If θ_j and θ_t are finite:*

$$\sum_{i=k+1}^{m} \tilde{u}_{j,i}\tilde{u}_{t,i} = O_p\left(\frac{1}{\sqrt{m}}\right).$$

Moreover, for $s = 1, ..., k$, $t = 2, ..., k$ and $j = k+1, ...m$,

$$\begin{aligned}
\sum_{i=1}^{k} \tilde{u}_{s,i}^2 &= \sum_{i=1}^{k} \left\langle \hat{u}_{\hat{\Sigma}_X,i}, \hat{u}_{\hat{\Sigma}_Y,s} \right\rangle^2, \\
\tilde{u}_{s,s} &= \hat{u}_{\hat{\Sigma}_X,s,s}\hat{u}_{\hat{\Sigma}_Y,s,s} + O_p\left(\frac{1}{m}\right) + O_p\left(\frac{1}{\theta_s^{1/2} m^{1/2}}\right), \\
\tilde{u}_{s,t} &= \hat{u}_{\hat{\Sigma}_X,t,s} + \hat{u}_{\hat{\Sigma}_X,s,t} + O_p\left(\frac{\sqrt{\min(\theta_s,\theta_t)}}{m\sqrt{\max(\theta_s,\theta_t)}}\right) + O_p\left(\frac{1}{\theta_t m^{1/2}}\right), \\
\tilde{u}_{t,s} &= O_p\left(\frac{\sqrt{\min(\theta_s,\theta_t)}}{m\sqrt{\max(\theta_s,\theta_t)}}\right) + O_p\left(\frac{1}{\theta_s m^{1/2}}\right), \\
\tilde{u}_{s,j} &= \hat{u}_{\hat{\Sigma}_Y,s,j} - \hat{u}_{\hat{\Sigma}_X,s,j}\left\langle \hat{u}_{\hat{\Sigma}_Y,j}, \hat{u}_{\hat{\Sigma}_X,j} \right\rangle + O_p\left(\frac{1}{\theta_s^{1/2} m}\right).
\end{aligned}$$

(Proof page 163)

Chapter 6

List of lemmas

6.1 Lemmas for Invariant Dot Product Theorem

This section introduces a Lemma used in the proof of the Dot Product Theorem 5.6.1.

Lemma 6.1.1.
Assuming W and $\hat{\Sigma}_{P_1}$ as in Theorem 5.6.1, then by construction of the eigenvectors using Theorem 5.11.1,

$$\begin{aligned}
\hat{u}_{P_1,1,2} &= \frac{W_{1,2}}{\sqrt{\theta_1}W_{1,1}} - \frac{W_{1,2}}{\theta_1^{3/2}}\left(-1/2 + 3/2M_2\right) + \frac{\left(W^2\right)_{1,2}}{\theta_1^{3/2}} + O_p\left(\frac{1}{\theta_1^{3/2}m}\right) + O_p\left(\frac{1}{\theta_1^{5/2}m^{1/2}}\right) \\
&= \frac{W_{1,2}}{\sqrt{\theta_1}} + O_p\left(\frac{1}{\theta_1^{1/2}m}\right) + O_p\left(\frac{1}{\theta_1^{3/2}m^{1/2}}\right), \\
\sum_{i=2}^m \hat{\lambda}_{P_1,i}^2 \hat{u}_{P_1,i,2}^2 &= W_{2,2} + O_p\left(\frac{1}{m}\right), \\
\sum_{i=2}^m \hat{\lambda}_{P_1,i} \hat{u}_{P_1,i,1} \hat{u}_{P_1,i,2} &= W_{1,2}\frac{M_2}{\sqrt{\theta_1}} - \left(W^2\right)_{1,2}\frac{1}{\sqrt{\theta_1}} + O_p\left(\frac{1}{\theta_1^{1/2}m}\right) + O_p\left(\frac{1}{\theta_1^{3/2}m^{1/2}}\right).
\end{aligned}$$

Remark 6.1.1.1.
Because the perturbation is of order 1, the two sign conventions defined in 5.1.1 are the same.

(Proof page 138)

6.2 Lemmas for the Main Theorem

This section introduces some results from linear algebra used in the proof of the Main Theorem 3.1.1.

Lemma 6.2.1.
Suppose

$$D = \left(\mathrm{I}_m + (\theta-1)u_X u_X^t\right)^{-1/2}\left(\mathrm{I}_m + (\theta-1)u_Y u_Y^t\right)\left(\mathrm{I}_m + (\theta-1)u_X u_X^t\right)^{-1/2}.$$

The eigenvalues of D are 1 *and*

$$\lambda(D) = -\frac{1}{2\theta}\left(-1 + \alpha^2 - 2\alpha^2\theta - \theta^2(1-\alpha^2) \pm \sqrt{-4\theta^2 + [1 + \theta^2 - (-1+\theta)^2\alpha^2]^2}\right),$$

where $\alpha^2 = \langle u_X, u_Y\rangle^2$.
Moreover, if

$$D_2 = \left(\mathrm{I}_m + (\theta_X-1)u_X u_X^t\right)^{-1/2}\left(\mathrm{I}_m + (\theta_Y-1)u_Y u_Y^t\right)\left(\mathrm{I}_m + (\theta_X-1)u_X u_X^t\right)^{-1/2}.$$

The eigenvalues of D_2 are 1 *and*

$$\lambda\left(D_2\right) = \frac{1}{2}\left(\theta_Y+\alpha^2-\theta_Y\alpha^2+\frac{1+(\theta_Y-1)\alpha^2\pm\sqrt{-4\theta_Y\theta_X+\left(1+\theta_Y\theta_X-(\theta_Y-1)(\theta_X-1)\alpha^2\right)^2}}{\theta_X}\right),$$

where $\alpha^2=\langle u_X,u_Y\rangle^2$.

(Proof page 176)

Lemma 6.2.2.
Suppose $w_1,...,w_k\in\mathbb{R}^m$ and $\lambda_1,...,\lambda_k\in\mathbb{R}^*$, then if the function $\lambda()$ provides non-trivial eigenvalues,

$$\lambda\left(\sum_{i=1}^{k}\lambda_i w_i w_i^t\right)=\lambda\left(H\right),$$

where

$$H=\begin{pmatrix}\lambda_1 & \sqrt{\lambda_1\lambda_2}\,\langle w_1,w_2\rangle & \sqrt{\lambda_1\lambda_3}\,\langle w_1,w_3\rangle & \cdots & \sqrt{\lambda_k\lambda_2}\,\langle w_1,w_k\rangle\\ \sqrt{\lambda_2\lambda_1}\,\langle w_2,w_1\rangle & \lambda_2 & \sqrt{\lambda_2\lambda_3}\,\langle w_2,w_3\rangle & \cdots & \sqrt{\lambda_2\lambda_k}\,\langle w_2,w_k\rangle\\ \sqrt{\lambda_3\lambda_1}\,\langle w_3,w_1\rangle & \sqrt{\lambda_3\lambda_2}\,\langle w_3,w_2\rangle & \lambda_3 & \cdots & \sqrt{\lambda_3\lambda_k}\,\langle w_3,w_k\rangle\\ \vdots & \vdots & \ddots & \ddots & \vdots\\ \sqrt{\lambda_k\lambda_1}\,\langle w_k,w_1\rangle & \sqrt{\lambda_k\lambda_2}\,\langle w_k,w_2\rangle & \sqrt{\lambda_k\lambda_3}\,\langle w_k,w_3\rangle & \cdots & \lambda_k\end{pmatrix}.$$

(Proof page 177)

Lemma 6.2.3.
Suppose $e_1,w\in\mathbf{R}^m$ and $a,b\in\mathbf{R}$. Then, if $||w||=1$, the two ($\pm$) non-trivial eigenvalues and eigenvectors are

$$\begin{aligned}\lambda^{\pm}\left(ae_1e_1^t+bww^t\right) &= \frac{1}{2}\left(a+b\pm\sqrt{4abw_1^2+(a-b)^2}\right),\\ u^{\pm}\left(ae_1e_1^t+bww^t\right) &= \frac{1}{\mathrm{Norm}^{\pm}}\left(\frac{\lambda^{\pm}\left(ae_1e_1^t+bww^t\right)+b\left(w1^2-1\right)}{bw_1},w_2,w_3,w_4,...,w_m\right),\\ \left(\mathrm{Norm}^{\pm}\right)^2 &= \frac{\left(\lambda^{\pm}\left(ae_1e_1^t+bww^t\right)+b\left(w1^2-1\right)\right)^2}{b^2w_1^2}+1-w_1^2.\end{aligned}$$

If $||w||\neq 1$,

$$\begin{aligned}\lambda^{\pm}\left(ae_1e_1^t+ww^t\right) &= \frac{1}{2}\left(\pm\sqrt{(a+||w||^2)^2-4a\left(||w||^2-w_1^2\right)}+a+||w||^2\right),\\ u^{\pm}\left(ae_1e_1^t+ww^t\right) &= \frac{1}{\mathrm{Norm}^{\pm}}\left(\frac{\lambda^{\pm}\left(ae_1e_1^t+ww^t\right)-||w||^2+w_1^2}{w_1},w_2,w_3,w_4,...,w_m\right),\\ \left(\mathrm{Norm}^{\pm}\right)^2 &= \frac{\left(\lambda^{\pm}\left(ae_1e_1^t+ww^t\right)-||w||^2+w_1^2\right)^2}{w_1^2}+||w||^2-w_1^2.\end{aligned}$$

(Proof page 178)

Lemma 6.2.4.
Suppose $u_1, ..., u_k \in \mathbb{R}^m$ are orthonormal and $\lambda_1 > ... > \lambda_k \in \mathbb{R}^+$ where k is finite. Suppose $v \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^+$ such that $\langle u_i, v\rangle = O_p\left(1/\sqrt{m}\right)$ and $\mu - \lambda_1 < d < 0$ for a fixed d, then

$$\lambda_{\max}\left(\sum_{i=1}^{k} \lambda_i u_i u_i^t + \mu v v^t\right) = \lambda_1 + O_p\left(\frac{1}{m}\right).$$

Moreover, if $\mu - \lambda_k > d_2 > 0$ for a fixed d_2,

$$\lambda_{\min}\left(\sum_{i=1}^{k} \lambda_i u_i u_i^t + \mu v v^t\right) = \lambda_k + O_p\left(\frac{1}{m}\right).$$

(Proof page 179)

6.3 Lemmas for robustness

In this section we introduce lemmas that characterise the residual spike when θ is large. These results are used to prove properties of robustness of our methods. In particular we use the moment in probability E_p and Var_p defined in 4.2.1.

Lemma 6.3.1.
Suppose $\mathbf{X} \in \mathbb{R}^{m\times n_X}$ and $\mathbf{Y} \in \mathbb{R}^{m\times n_Y}$ are members of the same class, $\mathcal{C}_A(\Sigma)$, $\mathcal{C}_B$, $\mathcal{C}_C$ or $\mathcal{C}_D(u)$. Moreover, define $W_X = \frac{1}{n_X}\mathbf{X}\mathbf{X}^t$ and $\hat{\Sigma}_{\mathbf{X}} = \frac{1}{n_X}\mathbf{X}\mathbf{X}^t$, where $\mathbf{X} = P^{1/2}\mathbf{X}$, and P is a finite perturbation of order 1. (Without loss of generality we can assume a canonical perturbation, but $\hat{u}_{\mathbf{X},1:m,1}$ is not necessarily uniform. In particular the following formulas are only true for $\mathcal{C}_D(u)$ after a rotation to change u in e_1.)
The largest residual spike obtained using $\mathbf{X}$ and $\mathbf{Y}$ is

$$\lambda_{m,\mathcal{C}} = \frac{1}{2}\left(\hat{\hat{\theta}}_Y(1-\hat{\alpha}^2) + 1 + \frac{\hat{\hat{\theta}}_Y}{\hat{\hat{\theta}}_X} + \sqrt{-4\frac{\hat{\hat{\theta}}_Y}{\hat{\hat{\theta}}_X} + \left(\hat{\hat{\theta}}_Y(1-\hat{\alpha}^2) + \frac{\hat{\hat{\theta}}_Y}{\hat{\hat{\theta}}_X} + 1\right)^2}\right) + O_p\left(\frac{1}{\theta}\right),$$

where

$$\begin{aligned}
\frac{1}{\theta-1} &= \sum_{i=1}^{m} \frac{\hat{\lambda}_{W_X,i}}{\hat{\theta}_X - \hat{\lambda}_{W_X,i}} \hat{u}^2_{W_X,i,1},\\
\frac{1}{\hat{\hat{\theta}}_X - 1} &= \frac{1}{m}\sum_{i=1}^{m} \frac{\hat{\lambda}_{W_X,i}}{\hat{\theta}_X - \hat{\lambda}_{W_X,i}},\\
\hat{\hat{\theta}}_X &= \theta \sum_{i=1}^{m} \hat{\lambda}_{W_X,i}\hat{u}^2_{W_X,1,i} + O_p(1),\\
\hat{\alpha}^2_X &= \frac{\theta}{(\theta-1)}\left(1 - \frac{\theta-1}{\hat{\theta}_X^2}\sum_{i=1}^{m} \hat{\lambda}^2_{W_X,i}\hat{u}^2_{W_X,i,1}\right) + O_p\left(\frac{1}{\theta^2}\right),\\
\hat{\alpha} &= \hat{\alpha}_X\hat{\alpha}_Y + \sum_{i=2}^{m} \hat{u}_{\hat{\Sigma}_{\mathbf{X}},i}\hat{u}_{\hat{\Sigma}_{\mathbf{Y}},i} + O_p\left(\frac{1}{\theta^2}\right),
\end{aligned}$$

$\hat{u}_{W_X,1,i}$ is the first entry of the i^{th} eigenvector of W_X and $\hat{\theta}_X = \hat{\theta}_{\hat{\Sigma}_{\mathbf{X}},1}$ is the largest eigenvalue of $\hat{\Sigma}_{\mathbf{X}}$.

Moreover for class $\mathcal{C}_A(\Sigma)$, $\mathcal{C}_B$ and $\mathcal{C}_C$,

$$\begin{aligned}\frac{\hat{\hat{\theta}}_Y}{\hat{\hat{\theta}}_X} &= 1 + O_p\left(\frac{1}{\theta}\right) + \left(\frac{1}{m^{1/2}}\right), \\ \hat{\hat{\theta}}_Y(1-\hat{\alpha}^2) &= \frac{1}{\Sigma_{1,1}}\left(\left({W_X}^2\right)_{1,1} + \left({W_Y}^2\right)_{1,1}\right) - \frac{2}{\Sigma_{1,1}}\left(\Sigma^2\right)_{1,1} + O_p\left(\frac{1}{\theta}\right) + O_p\left(\frac{1}{m^{1/2}}\right).\end{aligned}$$

(Proof page 191)

The next lemmas are used to prove the robustness of our model in the class defined in 4.2.3.

Lemma 6.3.2.
In the class $\mathcal{C}_A(\Sigma)$ defined in Definition 4.2.3, the limit of the expectation of the residual spike is invariant of Σ. This means that:
Suppose $\mathbf{X}_\Sigma, \mathbf{Y}_\Sigma \in \mathcal{C}_A(\Sigma)$ and $\mathbf{X}_{\mathrm{I}_m}, \mathbf{Y}_{\mathrm{I}_m} \in \mathcal{C}_A(\mathrm{I}_m)$ such that

$$\begin{aligned}\mathbf{X}_\Sigma &= P^{1/2}\Sigma^{1/2}\mathbf{X} \in \mathbb{R}^{m\times n_X} \text{ and } \mathbf{Y}_\Sigma = P^{1/2}\Sigma^{1/2}\mathbf{Y} \in \mathbb{R}^{m\times n_Y}, \\ \mathbf{X}_{\mathrm{I}_m} &= P^{1/2}\mathbf{X} \in \mathbb{R}^{m\times n_X} \text{ and } \mathbf{Y}_{\mathrm{I}_m} = P^{1/2}\mathbf{Y} \in \mathbb{R}^{m\times n_Y}.\end{aligned}$$

Let $\lambda_{m,\Sigma}$ and λ_{m,I_m} be the resulting residual spikes using $\left(\mathbf{X}_\Sigma, \mathbf{Y}_\Sigma\right)$ and $\left(\mathbf{X}_{\mathrm{I}_m}, \mathbf{Y}_{\mathrm{I}_m}\right)$ respectively. We have

$$\mathrm{E}_{p(\min(m,\theta))}\left[\lambda_{m,\Sigma}\right] = \mathrm{E}_{p(\min(m,\theta))}\left[\lambda_{m,\mathrm{I}_m}\right].$$

(Proof page 193)

Lemma 6.3.3.
We define in the class $\mathcal{C}_A(\mathrm{I}_m)$, $\mathbf{X}_{\mathrm{I}_m,\mathcal{L}_X} = P^{1/2}\mathbf{X}_{\mathcal{L}_X} \in \mathbb{R}^{m\times n_X}$ and $\mathbf{Y}_{\mathrm{I}_m,\mathcal{L}_Y} = P^{1/2}\mathbf{Y}_{\mathcal{L}_Y} \in \mathbb{R}^{m\times n_Y}$. Moreover, $\mathbf{X}_{\mathcal{L}_X}$ and $\mathbf{Y}_{\mathcal{L}_Y}$ have temporal structure such that

$$\begin{aligned}\mathbf{X}_{\mathcal{L}_X,1,\cdot}, \mathbf{X}_{\mathcal{L}_X,2,\cdot}, \ldots, \mathbf{X}_{\mathcal{L}_X,m,\cdot} \sim \mathcal{L}_X, \\ \mathbf{Y}_{\mathcal{L}_Y,1,\cdot}, \mathbf{Y}_{\mathcal{L}_Y,2,\cdot}, \ldots, \mathbf{Y}_{\mathcal{L}_Y,m,\cdot} \sim \mathcal{L}_Y.\end{aligned}$$

We define $W_{X,\mathcal{L}_X} = \frac{1}{n_X}\mathbf{X}_{\mathcal{L}_X}\mathbf{X}_{\mathcal{L}_X}^t$ and $W_{Y,\mathcal{L}_Y} = \frac{1}{n_Y}\mathbf{Y}_{\mathcal{L}_Y}\mathbf{Y}_{\mathcal{L}_Y}^t$.
If the spectra of $\left(W_{X,\mathcal{L}_1}, W_{Y,\mathcal{L}_3}\right)$ and $\left(W_{X,\mathcal{L}_2}, W_{Y,\mathcal{L}_4}\right)$ are rescaled and the second moments of the spectra are the same, i.e.

$$\begin{aligned}\mathrm{E}\left[\sum_{i=1}^m \hat{\lambda}_{W_{X,\mathcal{L}_1},i}\right] = \mathrm{E}\left[\sum_{i=1}^m \hat{\lambda}_{W_{X,\mathcal{L}_2},i}\right] = 1 \text{ and } \mathrm{E}\left[\sum_{i=1}^m \hat{\lambda}^2_{W_{X,\mathcal{L}_1},i}\right] = \mathrm{E}\left[\sum_{i=1}^m \hat{\lambda}^2_{W_{X,\mathcal{L}_2},i}\right], \\ \mathrm{E}\left[\sum_{i=1}^m \hat{\lambda}_{W_{Y,\mathcal{L}_3},i}\right] = \mathrm{E}\left[\sum_{i=1}^m \hat{\lambda}_{W_{Y,\mathcal{L}_4},i}\right] = 1 \text{ and } \mathrm{E}\left[\sum_{i=1}^m \hat{\lambda}^2_{W_{Y,\mathcal{L}_3},i}\right] = \mathrm{E}\left[\sum_{i=1}^m \hat{\lambda}^2_{W_{Y,\mathcal{L}_4},i}\right],\end{aligned}$$

then

$$\mathrm{E}_{p(\min(m,\theta))}\left[\lambda_{m,\mathcal{L}_1,\mathcal{L}_3}\right] = \mathrm{E}_{p(\min(m,\theta))}\left[\lambda_{m,\mathcal{L}_2,\mathcal{L}_4}\right],$$

where $\lambda_{m,\mathcal{L}_1,\mathcal{L}_3}$ and $\lambda_{m,\mathcal{L}_2,\mathcal{L}_4}$ are the resulting residual spikes using $\left(\mathbf{X}_{\mathrm{I}_m,\mathcal{L}_1}, \mathbf{Y}_{\mathrm{I}_m,\mathcal{L}_3}\right)$ and $\left(\mathbf{X}_{\mathrm{I}_m,\mathcal{L}_2}, \mathbf{Y}_{\mathrm{I}_m,\mathcal{L}_4}\right)$ respectively.

(Proof page 196)

Lemma 6.3.4.
Suppose that $\mathbf{X}_{\mathcal{L}_X}, \mathbf{X}_{\mathcal{L}^*_X} \in \mathbb{R}^{m\times n_X}$ *and* $\mathbf{Y}_{\mathcal{L}_Y}, \mathbf{Y}_{\mathcal{L}^*_Y} \in \mathbb{R}^{m\times n_Y}$ *are in* $\mathcal{C}_A(\mathrm{I}_m)$ *such that,*

$$\begin{aligned}\mathbf{X}_{\mathcal{L}_X} &= P^{1/2}\mathbf{X}_{\mathcal{L}_X},\ \mathbf{Y}_{\mathcal{L}_Y} = P^{1/2}\mathbf{Y}_{\mathcal{L}_Y},\\ \mathbf{X}_{\mathcal{L}^*_X} &= P^{1/2}\mathbf{X}_{\mathcal{L}^*_X},\ \mathbf{Y}_{\mathcal{L}^*_Y} = P^{1/2}\mathbf{Y}_{\mathcal{L}^*_Y},\end{aligned}$$

where $P = \mathrm{I}_m + (\theta-1)e_1e_1^t$.
Moreover, $\left(\mathbf{X}_{\mathcal{L}_X}, \mathbf{Y}_{\mathcal{L}_Y}\right)$ *and* $\left(\mathbf{X}_{\mathcal{L}^*_X}, \mathbf{Y}_{\mathcal{L}^*_Y}\right)$ *have temporal structure such that*

$$\begin{aligned}&\mathbf{X}_{\mathcal{L}_X,1,\cdot}, \mathbf{X}_{\mathcal{L}_X,2,\cdot}, ..., \mathbf{X}_{\mathcal{L}_X,m,\cdot} \sim \mathcal{L}_X,\\ &\mathbf{Y}_{\mathcal{L}_Y,1,\cdot}, \mathbf{Y}_{\mathcal{L}_Y,2,\cdot}, ..., \mathbf{Y}_{\mathcal{L}_Y,m,\cdot} \sim \mathcal{L}_Y,\\ &\mathbf{X}_{\mathcal{L}^*_X,1,\cdot} \sim \mathcal{L}_{X,1},\\ &\mathbf{X}_{\mathcal{L}^*_X,2,\cdot}, \mathbf{X}_{\mathcal{L}^*_X,3,\cdot}, ..., \mathbf{X}_{\mathcal{L}^*_X,m,\cdot} \sim \mathcal{L}_X,\\ &\mathbf{Y}_{\mathcal{L}^*_Y,1,\cdot} \sim \mathcal{L}_{Y,1},\\ &\mathbf{Y}_{\mathcal{L}^*_Y,2,\cdot}, \mathbf{Y}_{\mathcal{L}^*_Y,3,\cdot}, ..., \mathbf{Y}_{\mathcal{L}^*_Y,m,\cdot} \sim \mathcal{L}_Y.\end{aligned}$$

In particular, $\mathrm{Cov}\left(\mathbf{X}_{\mathcal{L}_X,1,\cdot}\right) = \Psi_{1,X} = \Psi_X$, $\mathrm{Cov}\left(\mathbf{Y}_{\mathcal{L}_Y,1,\cdot}\right) = \Psi_{1,Y} = \Psi_Y$, $\mathrm{Cov}\left(\mathbf{X}_{\mathcal{L}^*,1,\cdot}\right) = \Psi^*_{1,X}$, $\mathrm{Cov}\left(\mathbf{Y}_{\mathcal{L}^*,1,\cdot}\right) = \Psi^*_{1,Y}$, $\Delta_X = \Psi_X - \Psi^*_{1,X}$ *and* $\Delta_Y = \Psi_Y - \Psi^*_{1,Y}$ *respect the conditions of the class* $\mathcal{C}_A(\mathrm{I}_m)$.
Therefore we have

$$\mathrm{E}_{p(\min(m,\theta))}\left[\lambda_{m,\mathcal{L}_X,\mathcal{L}_Y}\right] = \mathrm{E}_{p(\min(m,\theta))}\left[\lambda_{m,\mathcal{L}^*_X,\mathcal{L}^*_Y}\right],$$

where $\lambda_{m,\mathcal{L}_X,\mathcal{L}_Y}$ *and* $\lambda_{m,\mathcal{L}^*_X,\mathcal{L}^*_Y}$ *are the resulting residual spikes using* $\left(\mathbf{X}_{\mathcal{L}_X}, \mathbf{Y}_{\mathcal{L}_Y}\right)$ *and* $\left(\mathbf{X}_{\mathcal{L}^*_X}, \mathbf{Y}_{\mathcal{L}^*_Y}\right)$ *respectively.*

(Proof page 198)

Lemma 6.3.5.
Suppose that $\mathbf{X} = P^{1/2}\mathbf{X} \in \mathbb{R}^{m\times n_X}$ *and* $\mathbf{Y} = P^{1/2}\mathbf{Y} \in \mathbb{R}^{m\times n_Y}$ *are two random matrices in class* $\mathcal{C}_A(\Sigma)$ *such that*

$$\begin{aligned}&\frac{1}{m}\sum_{i=1}^m \hat{\lambda}_{W_X,i} = \frac{1}{m}\sum_{i=1}^m \hat{\lambda}_{W_Y,i} = 1,\\ &\frac{1}{m}\sum_{i=1}^m \hat{\lambda}^2_{W_X,i} = M_{2,X} \text{ and } \frac{1}{m}\sum_{i=1}^m \hat{\lambda}^2_{W_Y,i} = M_{2,Y},\\ &M_2 = \frac{M_{2,X}+M_{2,Y}}{2}.\end{aligned}$$

We define $\lambda_{m,\mathcal{C}_A(\Sigma)}$ *as the resulting residual spike between* $\mathbf{X}$ *and* $\mathbf{Y}$.
The Main Theorem 3.1.1 says that knowing the spectra, the estimator of the expectation of residual spike assuming our model is

$$\lambda^+(M_2) = M_2 + \sqrt{M_2^2 - 1}.$$

This expectation of $\lambda_{m,\mathcal{C}_A(\Sigma)}$ *is conservative in the sense of the second robustness defined in 4.2.2. This means that the following results are true when* n_X, n_Y *and* θ *tends to infinity.*

1: If $\Sigma = \mathrm{I}_m$, *then* $\lambda^+(M_2) = M_2 + \sqrt{M_2^2-1} = \mathrm{E}_{p(\min(m,\theta))}\left[\lambda_{m,\mathcal{C}_A(\Sigma)}\right]$.

2: If $\Sigma \neq \mathrm{I}_m$, *then* $\lambda^+(M_2) = M_2 + \sqrt{M_2^2-1} > \mathrm{E}_{p(\min(m,\theta))}\left[\lambda_{m,\mathcal{C}_A(\Sigma)}\right]$.

(Proof page 199)

Chapter 7

Proofs

7.1 Proofs of the main theorems

In this section, we prove the theorems of the Chapter 5.

1. The Unit invariant vector statistic Theorem 5.11.2 and the characterisation of the eigenstructure 5.2.1,
2. The distributions of statistics when the perturbation is of order 1, Theorem 5.3.1,
3. Most of the Invariant Theorems using a circle induction procedure,
4. The Dot Product Theorem 5.6.1 and its invariant 5.7.1,
5. The double invariant angle Corollary 5.10.1.

7.1.1 Unit invariant vector statistic

Theorem 5.11.2.
Let W be a random matrix with spectrum, $\hat\lambda_{W,1}, \hat\lambda_{W,2}, ..., \hat\lambda_{W,m}$ such that its trace is 1. We denote by u_{p_1} and u_{p_2}, two orthonormal invariant random vectors independent of the eigenvalues of W. We set

$$\vec{B}_m\left(\rho, \vec{s}, \vec{r}, \vec{p}\right) = \sqrt{m}\left(\begin{pmatrix} \sum_{i=1}^m \frac{\hat\lambda_{W,i}^{s_1}}{\left(\rho - \hat\lambda_{W,i}\right)^{s_2}} u_{p_1,i} u_{p_2,i} \\ \sum_{i=1}^m \frac{\hat\lambda_{W,i}^{r_1}}{\left(\rho - \hat\lambda_{W,i}\right)^{r_2}} u_{p_1,i} u_{p_2,i} \end{pmatrix} - \begin{pmatrix} M_{s_1,s_2} \\ M_{r_1,r_2} \end{pmatrix} \mathbf{1}_{p_1=p_2}\right),$$

where $\vec{s} = (s_1, s_2)$, $\vec{r} = (r_1, r_2)$ and $\vec{p} = (p_1, p_2)$ with indices $1 \leqslant p_1 \leqslant p_2 \leqslant m$ and $s_1, s_2, r_1, r_2 \in \mathbf{N}$. Conditioning on the spectrum, if $p = p_1 = p_2$,

$$\vec{B}_m\left(\rho, \vec{s}, \vec{r}, \vec{p}\right) \sim \mathbf{N}\left(\vec{0}, \begin{pmatrix} 2\left(M_{2s_1,2s_2} - M_{s_1,s_1}^2\right) & 2\left(M_{s_1+r_1,s_2+r_2} - M_{s_1,s_2}M_{r_1,r_2}\right) \\ 2\left(M_{s_1+r_1,s_2+r_2} - M_{s_1,s_2}M_{r_1,r_2}\right) & 2\left(M_{2r_1,2r_2} - M_{r_1,r_1}^2\right) \end{pmatrix}\right) + o_{p;m}(1),$$

where $M_{s,r} = M_{s,r}(\rho) = \frac{1}{m}\sum_{i=1}^m \frac{\hat\lambda_{W,i}^{s}}{\left(\rho - \hat\lambda_{W,i}\right)^{r}}$.
Moreover, for $p_1 \neq p_2$,

$$\vec{B}_m\left(\rho, \vec{s}, \vec{r}, \vec{p}\right) \sim \mathbf{N}\left(\vec{0}, \begin{pmatrix} M_{2s_1,2s_2} - M_{s_1,s_1}^2 & M_{s_1+r_1,s_2+r_2} - M_{s_1,s_2}M_{r_1,r_2} \\ M_{s_1+r_1,s_2+r_2} - M_{s_1,s_2}M_{r_1,r_2} & M_{2r_1,2r_2} - M_{r_1,r_1}^2 \end{pmatrix}\right) + o_{p;m}(1),$$

In particular, with the notation $M_{s,0} = M_s = \frac{1}{m}\sum_{i=1}^m \hat\lambda_{W,i}^s$,

$$\sqrt{m}\left(\begin{pmatrix} \sum_{i=1}^m \hat\lambda_{W,i} u_{p,i}^2 \\ \sum_{i=1}^m \hat\lambda_{W,i}^2 u_{p,i}^2 \end{pmatrix} - \begin{pmatrix} 1 \\ M_2 \end{pmatrix}\right) \sim \mathbf{N}\left(\vec{0}, \begin{pmatrix} 2\left(M_2 - 1\right) & 2\left(M_3 - M_2\right) \\ 2\left(M_3 - M_2\right) & 2\left(M_4 - M_2^2\right) \end{pmatrix}\right) + o_{p;m}(1),$$

and

$$\sqrt{m}\left(\begin{pmatrix}\sum_{i=1}^m \hat{\lambda}_{W,i}u_{p_1,i}u_{p_2,i}\\ \sum_{i=1}^m \hat{\lambda}_{W,i}^2 u_{p_1,i}u_{p_2,i}\end{pmatrix}-\begin{pmatrix}0\\0\end{pmatrix}\right)\sim \mathbf{N}\left(\vec{0},\begin{pmatrix}M_2-1 & M_3-M_2\\ M_3-M_2 & M_4-M_2^2\end{pmatrix}\right)+o_{p;m}(1).$$

Finally if we look at K bivariate normal random variables :

$$\mathbf{B}_m(\vec{\rho},\mathbf{s},\mathbf{r},\mathbf{p})=\left(\vec{B}_m(\rho_1,\vec{s}_1,\vec{r}_1,\vec{p}_1),\vec{B}_m(\rho_2,\vec{s}_2,\vec{r}_2,\vec{p}_2),...,\vec{B}_m(\rho_K,\vec{s}_K,\vec{r}_K,\vec{p}_K)\right),$$

where ,$\vec{p}_i\neq\vec{p}_j$ if $i\neq j$. Then, $\tilde{\mathbf{B}}_m(\vec{\rho},\mathbf{s},\mathbf{r},\mathbf{p})$ tends to a multivariate Normal. Moreover, all the bivariate elements are asymptotically independent.

Remark 5.11.2.1.

1. In order to obtain this simple result we assume the trace of W is one. This property can easily be obtained by rescaling the matrix by the trace.
2. The condition of independence between eigenvectors and eigenvalues of W is strong but is an automatic consequence if the eigenvectors are Haar distributed. Moreover, if the fluctuation of the spectra is small, the result is still valid without conditioning on the spectra. For example for independent vector data, we can easily show that the result still holds using Bai and Silverstein [2010] (page 259-261, Theorem 9.10).
3. If the convergence of the spectrum to f_S is fast, then we can define $M_i=\int\lambda^i f_S(\lambda)d\lambda$.
4. If the data are such that the spectrum of W tends to Marcenko-Pastur distribution sufficiently fast, then f_S can be replaced by $f_{MP,c}$ and we obtain

$$\sqrt{m}\left(\begin{pmatrix}\sum_{i=1}^m \hat{\lambda}_{W,i}\hat{u}_i^2\\ \sum_{i=1}^m \hat{\lambda}_{W,i}^2\hat{u}_i^2\end{pmatrix}-\begin{pmatrix}1\\1+c\end{pmatrix}\right)\underset{m\to\infty}{\to}\mathbf{N}\left(\vec{0},\begin{pmatrix}2c & 2c(2+c)\\ 2c(2+c) & 2c(c+1)(c+4)\end{pmatrix}\right).$$

   Nevertheless the result is wrong if we only assume this spectral distribution! The independence with invariant eigenvectors is necessary.

(Page 87)

Proof. **Theorem 5.11.2**

The proof is divided into three steps. First, we compute the two first moments of the statistics using the same eigenvectors. Then, we show the asymptotic joint normality. Finally, we show the asymptotic independence of the statistics using at least one different eigenvector. In each step of the proof, we condition on the spectrum of W.

- For any non-random g function in $\mathbb{R}$,

$$\begin{aligned}\mathrm{E}\left[\sum_{i=1}^m g(\hat{\lambda}_{W,i})u_{1,i}^2\right] &= \frac{1}{m}\sum_{i=1}^m g(\hat{\lambda}_{W,i}),\\ \mathrm{E}\left[\sum_{i=1}^m g(\hat{\lambda}_{W,i})u_{1,i}u_{2,i}\right] &= 0\end{aligned}$$

  This proves the formulas for the first moments. We can now compute the covariance matrix knowing that for unit invariant eigenvectors, $\operatorname{Cov}\left(u_{1,i}^2, u_{1,j}^2\right) = -2/(m^2(m-1))$. Therefore

$$\begin{aligned}
&\operatorname{Var}\left(\sum_{i=1}^m \frac{\hat{\lambda}_{W,i}^{s_1}}{\left(\rho-\hat{\lambda}_{W,i}\right)^{s_2}} u_{1,i}^2\right) \\
&\qquad = \sum_{i=1}^m \frac{\hat{\lambda}_{W,i}^{2s_1}}{\left(\rho-\hat{\lambda}_{W,i}\right)^{2s_2}} \frac{2}{m^2} - \sum_{\substack{i,j=1\\ i\neq j}}^m \left(\frac{\hat{\lambda}_{W,i}^{s_1}}{\left(\rho-\hat{\lambda}_{W,i}\right)^{s_2}} \frac{\hat{\lambda}_{W,j}^{s_1}}{\left(\rho-\hat{\lambda}_{W,j}\right)^{s_2}}\right) \frac{2}{m^2(m-1)} \\
&\qquad = \frac{2M_{2s_1,2s_2}(\rho)}{m} - \frac{2M_{s_1,s_2}(\rho)^2}{m-1} + \frac{2M_{2s_1,2s_2}(\rho)}{m(m-1)} \\
&\qquad = \frac{2}{m}\left(M_{2s_1,2s_2}(\rho) - M_{s_1,s_2}(\rho)^2\right) + O_p\left(\frac{1}{m^2}\right)
\end{aligned}$$

and

$$\begin{aligned}
&\operatorname{Cov}\left(\sum_{i=1}^m \frac{\hat{\lambda}_{W,i}^{s_1}}{\left(\rho-\hat{\lambda}_{W,i}\right)^{s_2}} u_{1,i}^2, \sum_{j=1}^m \frac{\hat{\lambda}_{W,i}^{r_1}}{\left(\rho-\hat{\lambda}_{W,i}\right)^{r_2}} u_{1,j}^2\right) \\
&\qquad = \sum_{i=1}^m \frac{\hat{\lambda}_{W,i}^{s_1+r_1}}{\left(\rho-\hat{\lambda}_{W,i}\right)^{s_2+r_2}} \frac{2}{m^2} - \sum_{\substack{i,j=1\\ i\neq j}}^m \frac{\hat{\lambda}_{W,i}^{s_1}}{\left(\rho-\hat{\lambda}_{W,i}\right)^{s_2}} \frac{\hat{\lambda}_{W,i}^{r_1}}{\left(\rho-\hat{\lambda}_{W,i}\right)^{r_2}} \frac{2}{(m-1)m^2} \\
&\qquad = \frac{2M_{s_1+r_1,s_2+r_2}(\rho)}{m} - \frac{2M_{s_1,s_2}(\rho)M_{r_1,r_2}(\rho)}{m-1} + \frac{2M_{s_1+r_1,s_2+r_2}(\rho)}{(m-1)m} \\
&\qquad = \frac{2}{m}\left(M_{s_1+r_1,s_2+r_2}(\rho) - M_{s_1,s_2}(\rho)M_{r_1,r_2}(\rho)\right) + O_p\left(\frac{1}{m^2}\right)
\end{aligned}$$

Similarly, we can show that

$$\operatorname{Var}\left(\sum_{i=1}^m \frac{\hat{\lambda}_{W,i}^{s_1}}{\left(\rho-\hat{\lambda}_{W,i}\right)^{s_2}} u_{1,i}u_{2,i}\right) = \frac{M_{2s_1,2s_2}(\rho) - M_{s_1,s_2}(\rho)^2}{m} + O_p\left(\frac{1}{m^2}\right)$$

and

$$\begin{aligned}
&\operatorname{Cov}\left(\sum_{i=1}^m \frac{\hat{\lambda}_{W,i}^{s_1}}{\left(\rho-\hat{\lambda}_{W,i}\right)^{s_2}} u_{1,i}u_{2,i}, \sum_{j=1}^m \frac{\hat{\lambda}_{W,i}^{r_1}}{\left(\rho-\hat{\lambda}_{W,i}\right)^{r_2}} u_{1,j}u_{2,i}\right) \\
&\qquad\qquad = \frac{M_{s_1+r_1,s_2+r_2}(\rho) - M_{s_1,s_2}(\rho)M_{r_1,r_2}(\rho)}{m} + O_p\left(\frac{1}{m^2}\right).
\end{aligned}$$

- In order to prove the joint normality we show that any projections of a vector with unit statistics components built using $u_1, u_2, ..., u_k$ are Normal with an error tending to 0. This vector contains all the elements of the following matrix:

$$\left(\sqrt{m}\sum_{i=1}^m a_{r,s}\left(\hat{\lambda}_{W,i}\right)\left(u_{s,i}u_{r,i} - \mathbf{1}_{r=s}\frac{1}{m}\right)\right)_{s,r},$$

where $a_{r,s}\left(\hat{\lambda}_{W,i}\right)$ are constant depending on $\hat{\lambda}_{W,i}$.
We start with p_2 independent Normal random vectors of size m, $v_1, v_2, ..., v_{p_2}$ ($p_2 \geqslant p_1$), where

$\text{Var}(v_i) = \frac{1}{m}\text{I}_m$. Then, we apply Gramm-Schmidt.

$$\begin{aligned} u_s &= \frac{v_s - \sum_{i=1}^{s-1}\left(\langle v_s, v_i\rangle + O_p\left(\frac{1}{m}\right)\right)\frac{v_i}{||v_i||^2 + O_p\left(\frac{1}{m}\right)}}{||v_s|| + O_p\left(\frac{1}{m}\right)} \\ &= \sum_{i=1}^{s} b_{s,i} v_i. \end{aligned}$$

Moreover, the central limit theorem of Bentkus [2005] shows that for any $t_1, t_2 = 1, 2, ...k$,

$$\sqrt{m}\sum_{i=1}^{m} a_{r,s}\left(\hat{\lambda}_{W,i}\right)\left(v_{t_1,i}v_{t_2,i} - \mathbf{1}_{t_1=t_2}\frac{1}{m}\right) \tag{7.1}$$

can be jointly estimated by a centred multivariate Normal when k is finite.
In particular using Slutsky's Theorem we show that

$$\forall s, t = 1, 2, ..., k,\ \ s \neq t,\ \ \sqrt{m}\ b_{s,t} \tag{7.2}$$

tends to centred joint Normal random variable with variance of size $O_p(1)$. Moreover,

$$\forall s = 1, 2, ..., k,\ \ \sqrt{m}\,(b_{s,s} - 1) \tag{7.3}$$

tends to a centred joint normal random variable with variance of size $O_p(1)$.

Finally we show that any projections of the unit statistic are Normal. Without loss of generality we keep the constant $a_{r,s}\left(\hat{\lambda}_{W,i}\right)$.

$$\begin{aligned} &\sum_{s,r=1}^{k}\sqrt{m}\sum_{i=1}^{m} a_{r,s}\left(\hat{\lambda}_{W,i}\right)\left(u_{s,i}u_{r,i} - \mathbf{1}_{s=r}\frac{1}{m}\right) \\ &= \sum_{s,r=1}^{k}\sum_{t_1=1}^{s}\sum_{t_2=1}^{r} b_{s,t_1}b_{r,t_2}\sqrt{m}\sum_{i=1}^{m} a_{r,s}\left(\hat{\lambda}_{W,i}\right) v_{t_1,i}v_{t_2,i} - \frac{1}{m}\sum_{s=1}^{k}\sqrt{m}\sum_{i=1}^{m} a_{s,s}\left(\hat{\lambda}_{W,i}\right) \\ &= \sum_{s,r=1}^{k}\sum_{t_1=1}^{s}\sum_{t_2=1}^{r} b_{s,t_1}b_{r,t_2}\sqrt{m}\sum_{i=1}^{m} a_{r,s}\left(\hat{\lambda}_{W,i}\right)\left(v_{t_1,i}v_{t_2,i} - \mathbf{1}_{t_1=t_2}\frac{1}{m}\right) \\ &\quad + \sum_{s,r=1}^{k}\sum_{t=1}^{\min(r,s)} b_{s,t}b_{r,t}\sqrt{m}\frac{1}{m}\sum_{i=1}^{m} a_{r,s}\left(\hat{\lambda}_{W,i}\right) - \sum_{s=1}^{k}\sqrt{m}\frac{1}{m}\sum_{i=1}^{m} a_{s,s}\left(\hat{\lambda}_{W,i}\right) \\ &= \sum_{s,r=1}^{k}\sum_{t_1=1}^{s}\sum_{t_2=1}^{r} \underbrace{b_{s,t_1}b_{r,t_2}\sqrt{m}\sum_{i=1}^{m} a_{r,s}\left(\hat{\lambda}_{W,i}\right)\left(v_{t_1,i}v_{t_2,i} - \mathbf{1}_{t_1=t_2}\frac{1}{m}\right)}_{A_1} \\ &\quad \underbrace{+ \sum_{\substack{s,r=1 \\ s\neq r}}^{k}\sum_{t=1}^{\min(r,s)} b_{s,t}b_{r,t}\sqrt{m}\frac{1}{m}\sum_{i=1}^{m} a_{r,s}\left(\hat{\lambda}_{W,i}\right)}_{A_2} \\ &\quad \underbrace{-\sum_{s=1}^{k}\sqrt{m}\left(1 - b_{s,s}^2\right)\frac{1}{m}\sum_{i=1}^{m} a_{s,s}\left(\hat{\lambda}_{W,i}\right)}_{A_3} \end{aligned}$$

We show asymptotic normality of the three elements using Slutsky's Theorem:

$$\begin{aligned}
A_1 &= b_{s,t_1} b_{r,t_2} \sqrt{m} \sum_{i=1}^{m} a_{r,s}\left(\hat{\lambda}_{W,i}\right)\left(v_{t_1,i} v_{t_2,i} - \mathbf{1}_{t_1=t_2}\frac{1}{m}\right) \\
&= \mathrm{E}\left[b_{s,t_1} b_{r,t_2}\right] \sqrt{m} \sum_{i=1}^{m} a_{r,s}\left(\hat{\lambda}_{W,i}\right)\left(v_{t_1,i} v_{t_2,i} - \mathbf{1}_{t_1=t_2}\frac{1}{m}\right) + O_p\left(\frac{1}{\sqrt{m}}\right).
\end{aligned}$$

We recognise the equation 7.1 on the right and we know that we can neglect all the terms when $s \neq t_1$ or $r \neq t_2$. Indeed in these cases by equation 7.2, $b_{s,t_1} b_{r,t_2} = O_p\left(1/\sqrt{m}\right)$ and the terms become of order $O_p\left(1/\sqrt{m}\right)$.

$$\begin{aligned}
A_2 &= \sum_{\substack{s,r=1 \\ s\neq r}}^{k} \sum_{t=1}^{\min(r,s)} b_{s,t} b_{r,t} \sqrt{m} \frac{1}{m} \sum_{i=1}^{m} a_{r,s}\left(\hat{\lambda}_{W,i}\right) \\
&= \sum_{\substack{s,r=1 \\ s\neq r}}^{k} b_{s,\min(r,s)} b_{r,\min(r,s)} \sqrt{m} \frac{1}{m} \sum_{i=1}^{m} a_{r,s}\left(\hat{\lambda}_{W,i}\right) + \sum_{\substack{s,r=1 \\ s\neq r}}^{k} \sum_{t=1}^{\min(r,s)-1} b_{s,t} b_{r,t} \sqrt{m} \frac{1}{m} \sum_{i=1}^{m} a_{r,s}\left(\hat{\lambda}_{W,i}\right) \\
&= \sum_{\substack{s,r=1 \\ s\neq r}}^{k} \mathrm{E}\left[b_{\min(r,s),\min(r,s)}\right] \sqrt{m} b_{\max(r,s),\min(r,s)} \frac{1}{m} \sum_{i=1}^{m} a_{r,s}\left(\hat{\lambda}_{W,i}\right) + O_p\left(\frac{1}{\sqrt{m}}\right) \\
&= \sum_{\substack{s,r=1 \\ s\neq r}}^{k} \sqrt{m} b_{\max(r,s),\min(r,s)} \frac{1}{m} \sum_{i=1}^{m} a_{r,s}\left(\hat{\lambda}_{W,i}\right) + O_p\left(\frac{1}{\sqrt{m}}\right).
\end{aligned}$$

We recognise the equation 7.2 multiplied by a constant.

$$A_3 = \sum_{s=1}^{k} \sqrt{m}\left(1 - b_{s,s}^2\right) \frac{1}{m} \sum_{i=1}^{m} a_{s,s}\left(\hat{\lambda}_{W,i}\right).$$

We recognise the equation 7.3 multiplied by a constant. Therefore, we obtain :

$$\begin{aligned}
&\sum_{s,r=1}^{k} \sqrt{m} \sum_{i=1}^{m} a_{r,s}\left(\hat{\lambda}_{W,i}\right)\left(u_{s,i} u_{r,i} - \mathbf{1}_{s=r}\frac{1}{m}\right) \\
&\quad = \sum_{s,r=1}^{k} \sqrt{m} \sum_{i=1}^{m} a_{r,s}\left(\hat{\lambda}_{W,i}\right)\left(v_{s,i} v_{r,i} - \mathbf{1}_{s=r}\frac{1}{m}\right) + \sum_{\substack{s,r=1 \\ s\neq r}}^{k} \sqrt{m} b_{\min(s,t),\max(s,t)} \frac{1}{m} \sum_{i=1}^{m} a_{r,s}\left(\hat{\lambda}_{W,i}\right) \\
&\qquad + \sum_{s=1}^{k} 2\sqrt{m}\left(b_{s,s} - 1\right)) \frac{1}{m} \sum_{i=1}^{m} a_{s,s}\left(\hat{\lambda}_{W,i}\right).
\end{aligned}$$

Because all the quantities in the finite three sums are approximately jointly normal, then the quantity is approximately normal with an error $o_{p;m}(1)$.

- Assuming $s_i \leqslant r_i$, if $(s_1, r_1) \neq (s_2, r_2)$, then by invariance under rotation,

$$\mathrm{Cov}\left(\sum_{i=1}^{m} a_{r_1,s_1}\left(\hat{\lambda}_{W,i}\right)\left(u_{s_1,i} u_{r_1,i} - \mathbf{1}_{r_1=s_1}\frac{1}{m}\right), \sum_{i=1}^{m} a_{r_2,s_2}\left(\hat{\lambda}_{W,i}\right)\left(u_{s_2,i} u_{r_2,i} - \mathbf{1}_{r_2=s_2}\frac{1}{m}\right)\right) = 0.$$

Therefore, when $(s_1, r_1) \neq (s_2, r_2)$, the resulting joint Normal statistics are asymptotically independent. However, when $(s_1, r_1) = (s_2, r_2)$, then the statistics are jointly Normal and

correlated.
Note that when $s_1 = r_1$ and $s_2 = r_2$ we use the linearity of the covariance and the fact that $\mathrm{Cov}\left(\sum_{i=1}^m a_i u_{1,i}^2, \sum_{i=1}^m \tilde{a}_i u_{2,i}^2\right) = \mathrm{Cov}\left(\sum_{i=1}^m a_i u_{1,i}^2, \sum_{i=1}^m \tilde{a}_i \frac{1-u_{1,i}^2}{m-1}\right)$.

$\square$

7.1.2 Convergence and general characterization

Theorem 5.2.1.
In this theorem, $P = \mathrm{I}_m + (\theta-1)uu^t$ is a finite perturbation of order 1.

1. *Suppose W is a symmetric matrix with eigenvalues $\hat{\lambda}_{W,i} \geqslant 0$ and eigenvectors $\hat{u}_{W,i}$ for $i = 1, 2, ..., m$ (without any additional assumptions on the eigenstructure).*
   For $i = 1, 2, ..., m$, we define $\tilde{u}_{P,i}$ and $\hat{\lambda}_{P,i}$ such that

$$W P \tilde{u}_{P,i} = \hat{\lambda}_{P,i} \tilde{u}_{P,i},$$

   and the usual $\hat{u}_{P,i}$ such that if $\hat{\Sigma}_P = P^{1/2} W P^{1/2}$, then

$$\hat{\Sigma}_P \hat{u}_{P,i} = P^{1/2} W P^{1/2} \hat{u}_{P,i} = \hat{\lambda}_{P,i} \hat{u}_{P,i}.$$

   - *The eigenvalues $\hat{\lambda}_{P,s}$ are such that for $s = 1, 2, ..., m$,*

$$\sum_{i=1}^m \frac{\hat{\lambda}_{W,i}}{\hat{\lambda}_{\hat{\Sigma}_P,s} - \hat{\lambda}_{W,i}} \langle \hat{u}_{W,i}, u \rangle^2 = \frac{1}{\theta_k - 1}.$$

   - *The eigenvectors $\tilde{u}_{P,s}$ are such that*

$$\langle \tilde{u}_{P,s}, v \rangle^2 = \frac{\left(\sum_{i=1}^m \frac{\hat{\lambda}_{W,i}}{\hat{\lambda}_{P,s} - \hat{\lambda}_{W,i}} \langle \hat{u}_{W,i}, v \rangle \langle \hat{u}_{W,i}, u \rangle\right)^2}{\sum_{i=1}^m \frac{\hat{\lambda}_{W,i}^2}{(\hat{\lambda}_{P,s} - \hat{\lambda}_{W,i})^2} \langle \hat{u}_{W,i}, u \rangle^2}.$$

     In particular if $v = u$,

$$\langle \tilde{u}_{P,s}, u \rangle^2 = \frac{1}{(\theta-1)^2 \left(\sum_{i=1}^m \frac{\hat{\lambda}_{W,i}^2}{(\hat{\lambda}_{P,s} - \hat{\lambda}_{W,i})^2} \langle \hat{u}_{W,i}, u \rangle^2\right)}.$$

     Moreover,

$$\hat{u}_{P,s} = \frac{P^{1/2} \tilde{u}_{P,s}}{\sqrt{1 + (\theta-1) \langle \tilde{u}_{P,s}, u \rangle^2}}.$$

     Therefore, for u and v such that $\langle v, u \rangle = 0$,

$$\langle \hat{u}_{P,s}, u \rangle^2 = \frac{\theta \langle \tilde{u}_{P,s}, u \rangle^2}{1 + (\theta-1) \langle \tilde{u}_{P,s}, u \rangle^2} = -\frac{\theta}{(\theta-1)^2 \hat{\lambda}_{P,s} T'_{W,u}(\hat{\lambda}_{P,s})},$$

$$\langle \hat{u}_{P,s}, v \rangle^2 = \frac{\langle \tilde{u}_{P,s}, u \rangle^2}{1 + (\theta-1) \langle \tilde{u}_{P,s}, u \rangle^2},$$

     where $T_{W,u}(z) = \sum_{i=1}^m \frac{\hat{\lambda}_{W,i}}{z - \hat{\lambda}_{W,i}} \langle \hat{u}_{W,i}, u \rangle^2$.

2. *Suppose that W_X, W_Y and $P = P_X = P_Y$ satisfy Assumption 2.2.1. Moreover, suppose that θ is large enough to create detectable spikes, $\hat{\theta}_X$ and $\hat{\theta}_Y$, in the matrices $\hat{\Sigma}_X = P^{1/2}W_XP^{1/2}$ and $\hat{\Sigma}_Y = P^{1/2}W_YP^{1/2}$. Then,*

$$
\begin{aligned}
&a) \quad \hat{\hat{\theta}}_X, \hat{\hat{\theta}}_Y \xrightarrow[n,m\to\infty]{P} \theta,\\
&b) \quad \langle \hat{u}_X, u\rangle - \alpha_X, \langle \hat{u}_Y, u\rangle - \alpha_Y \xrightarrow[n,m\to\infty]{P} 0,\\
&c) \quad \langle \hat{u}_X, \hat{u}_Y\rangle - \alpha_X\alpha_Y \xrightarrow[n,m\to\infty]{P} 0,
\end{aligned}
$$

where

$$
\begin{aligned}
&\hat{\theta}_X \xrightarrow[n,m\to\infty]{P} \rho_X,\\
&\hat{\hat{\theta}}_X = 1 + \frac{1}{\hat{T}_{\hat{\Sigma}_X}\left(\hat{\theta}_X\right)} = 1 + \frac{m}{\sum_{i=k+1}^m \frac{\hat{\lambda}_{\hat{\Sigma}_X,i}}{\hat{\theta}_X - \hat{\lambda}_{\hat{\Sigma}_X,i}}},\\
&\alpha_X^2 = -\frac{\theta}{(\theta-1)^2\rho_X\hat{T}'_{W_X}(\rho_X)},\\
&\alpha_Y^2 = -\frac{\theta}{(\theta-1)^2\rho_Y\hat{T}'_{W_Y}(\rho_Y)},
\end{aligned}
$$

$\hat{\lambda}_{\hat{\Sigma}_X,i}$ and $\hat{\lambda}_{\hat{\Sigma}_Y,i}$ are the eigenvalues of respectively $\hat{\Sigma}_X$ and $\hat{\Sigma}_Y$.

Remark 5.2.1.1.
If the spectra of W_X and W_Y tend fast enough to the Marcenko-Pastur distribution with parameter c_X and c_Y respectively, then

$$
\begin{aligned}
&\alpha_X^2 = \frac{1 - \frac{c_X}{(\theta-1)^2}}{1 + \frac{c_X}{\theta-1}},\\
&\hat{\hat{\theta}}_X \text{ is such that } \hat{\theta}_X = \hat{\hat{\theta}}_X\left(1 + \frac{c_X}{\hat{\hat{\theta}}_X - 1}\right), \textit{and}\\
&\lim_{m\to\infty} \hat{\theta}_X = \theta\left(1 + \frac{c_X}{\theta-1}\right).
\end{aligned}
$$

(Page 72)

Proof. **Theorem 5.2.1**

In the proof of this theorem, we use two transforms:

- $T_{W_X,u}(z) = \sum_{i=1}^m \frac{\hat{\lambda}_{W_X,i}}{z-\hat{\lambda}_{W_X,i}} \langle \hat{u}_{W_X,i}, u\rangle^2$,
- $\hat{T}_{W_X}(z) = \frac{1}{m}\sum_{i=1}^m \frac{\hat{\lambda}_{W_X,i}}{z-\hat{\lambda}_{W_X,i}}$.

1. In order to study $P^{1/2}WP^{1/2}$, where $P = \mathrm{I} + (\theta-1)uu^t$, we define the diagonalisation $W = \hat{U}_W^t\Lambda_W\hat{U}_W$.

Eigenvalues: The matrix of interest has the same eigenvalues as WP. Thus, for z an eigenvalue of $P^{1/2}WP^{1/2}$,

$$\begin{aligned}
0 &= \det\left(z\mathrm{I}_m - P^{1/2}WP^{1/2}\right) \\
&= \det\left(z\mathrm{I}_m - W\left(\mathrm{I}_m + (\theta-1)uu^t\right)\right) \\
&= \det\left(z\mathrm{I}_m - \hat{U}_W^t\Lambda_W\hat{U}_W\left(\mathrm{I}_m + (\theta-1)uu^t\right)\right) \\
&= \det\left(z\mathrm{I}_m - \Lambda_W\left(\mathrm{I}_m + (\theta-1)\left(\hat{U}_W^t u\right)\left(\hat{U}_W^t u\right)^t\right)\right) \\
&= \det\left(z\mathrm{I}_m - \Lambda_W\right)\det\left(\mathrm{I}_m - (z\mathrm{I}_m - \Lambda_W)^{-1}\Lambda_W\left((\theta-1)\left(\hat{U}_W^t u\right)\left(\hat{U}_W^t u\right)^t\right)\right).
\end{aligned}$$

If z is not an eigenvalue of W but an eigenvalue of WP, then

$$\det\left(z\mathrm{I}_m - \Lambda_W\right) \neq 0.$$

Therefore,

$$\begin{aligned}
&\det\left(\mathrm{I}_m - (z\mathrm{I} - \Lambda_W)^{-1}\Lambda_W\left((\theta-1)\left(\hat{U}_W^t u\right)\left(\hat{U}_W^t u\right)^t\right)\right) = 0 \\
&\Rightarrow \mathrm{Trace}\left((z\mathrm{I}_m - \Lambda_W)^{-1}\Lambda_W\left((\theta-1)\left(\hat{U}_W^t u\right)\left(\hat{U}_W^t u\right)^t\right)\right) = 1 \\
&\Rightarrow \mathrm{Trace}\left((z\mathrm{I}_m - \Lambda_W)^{-1}\Lambda_W\left(\left(\hat{U}_W^t u\right)\left(\hat{U}_W^t u\right)^t\right)\right) = \frac{1}{\theta-1} \\
&\Rightarrow \sum_{i=1}^m \frac{\hat{\lambda}_{W,i}}{z - \hat{\lambda}_{W,i}}\left\langle \hat{u}_{W,i}, u\right\rangle^2 = \frac{1}{\theta-1}.
\end{aligned}$$

In our notation, we can replace z by the eigenvalues of $P^{1/2}WP^{1/2}$, $\hat{\lambda}_{P,s}$ for $s = 1, 2, ..., m$.

Eigenvectors: We first study $\tilde{u}_{P,s}$ for $s = 1, 2, ..., m$, the eigenvectors of WP. Because $\hat{\lambda}_{P,s}$ is the eigenvalue of WP corresponding to $\tilde{u}_{P,s}$,

$$\hat{\lambda}_{P,s}\tilde{u}_{P,s} = WP\tilde{u}_{P,s} = W(\mathrm{I}_m + \bar{P})\tilde{u}_{P,s} = (W + W\bar{P})\tilde{u}_{P,s}.$$

Therefore, we have

$$\tilde{u}_{P,s} = \left((\theta-1)u^t\tilde{u}_{P,s}\right)\left(\hat{\lambda}_{P,s}\mathrm{I}_m - W\right)^{-1}Wu.$$

Using the fact that $\tilde{u}_{P,s}$ is proportional to $\left(\hat{\lambda}_{P,s}\mathrm{I}_m - W\right)^{-1}Wu$ and its norm is unit, we get

$$\tilde{u}_{P,s} = \frac{\left(\hat{\lambda}_{P,s}\mathrm{I}_m - W\right)^{-1}W}{\sqrt{u^tW\left(\hat{\lambda}_{P,s}\mathrm{I}_m - W\right)^{-2}Wu}}u,$$

which leads to most of the results about eigenvectors.
Assuming $v \in \mathbf{R}^m$, then

$$\begin{aligned}
\left\langle \tilde{u}_{P,s}, v\right\rangle^2 &= \frac{\left(v^tW\left(\hat{\lambda}_{P,s}\mathrm{I}_m - W\right)^{-1}u\right)^2}{u^tW\left(\hat{\lambda}_{P,s}\mathrm{I}_m - W\right)^{-2}W_Xu} \\
&= \frac{\left(\sum_{i=1}^m \frac{\hat{\lambda}_{W,i}}{\hat{\lambda}_{P,s}-\hat{\lambda}_{W,i}}\left\langle \hat{u}_{W,i}, v\right\rangle\left\langle \hat{u}_{W,i}, u\right\rangle\right)^2}{\sum_{i=1}^m \frac{\hat{\lambda}_{W,i}^2}{(\hat{\lambda}_{P,s}-\hat{\lambda}_{W,i})^2}\left\langle \hat{u}_{W,i}, u\right\rangle^2}.
\end{aligned}$$

In particular, when $v = u$, $u^t W_X(\hat{\lambda}_{P,s} I - W_X)^{-1} u = 1/(\theta - 1)$ by the previous part on eigenvalues.
In order to obtain a more elegant formula, we need to study the denominator and its "pseudo"-transform:

$$\int \frac{x^2}{(z-x)^2} f(x)dx = -\int \frac{x}{(z-x)} f(x)dx + z \int \frac{x}{(z-x)^2} f(x)dx = -T(z) - zT'(z),$$

where T is defined in Definition 1.3.2. Therefore,

$$\begin{aligned} \langle \tilde{u}_{P,s}, u \rangle^2 &= \frac{T_{W,u}(\hat{\lambda}_{P,s})^2}{-T_{W,u}(\hat{\lambda}_{P,s}) - \hat{\lambda}_{P,s} T'_{W,u}(\hat{\lambda}_{P,s})} \\ &= -\frac{1}{(\theta-1)^2 \hat{\lambda}_{P,s} T'_{W,u}(\hat{\lambda}_{P,s}) + (\theta - 1)}. \end{aligned}$$

Then, we focus on $\hat{u}_{P,s}$. Some simple results of linear algebra show that,

$$\hat{u}_{P,s} = \frac{P^{1/2}\tilde{u}_{P,s}}{\sqrt{\tilde{u}^t_{P,s} P \tilde{u}_{P,s}}}.$$

Using the fact that $\tilde{u}^t_{P,s} P \tilde{u}_{P,s} = 1 + (\theta - 1) \langle \tilde{u}_{P,s}, u \rangle^2$, the angle can be written,

$$\begin{aligned} \langle \hat{u}_{P,s}, v \rangle^2 &= \frac{\left\langle P^{1/2}\tilde{u}_{P,s}, v \right\rangle^2}{\tilde{u}^t_{P,s} P \tilde{u}_{P,s}} \\ &= \frac{\left\langle \left(\mathrm{I}_m + (\sqrt{\theta} - 1)uu^t \right) \tilde{u}_{P,s}, v \right\rangle^2}{1 + (\theta - 1) \langle \tilde{u}_{P,s}, u \rangle^2} \\ &= \frac{\left(\langle \tilde{u}_{P,s}, v \rangle + \left(\sqrt{\theta} - 1 \right) \langle \tilde{u}_{P,s}, u \rangle \langle u, v \rangle \right)^2}{1 + (\theta - 1) \langle \tilde{u}_{P,s}, u \rangle^2}. \end{aligned}$$

For u and v such that $\langle v, u \rangle = 0$, we easily obtain,

$$\begin{aligned} \langle \hat{u}_{P,s}, u \rangle^2 &= \frac{\theta \langle \tilde{u}_{P,s}, u \rangle^2}{1 + (\theta - 1) \langle \tilde{u}_{P,s}, u \rangle^2} \\ &= -\frac{\theta}{(\theta - 1)^2 \hat{\lambda}_{P,s} T'_{W,u}(\hat{\lambda}_{P,s})}, \\ \langle \hat{u}_{P,s}, v \rangle^2 &= \frac{\langle \tilde{u}_{P,s}, u \rangle^2}{1 + (\theta - 1) \langle \tilde{u}_{P,s}, u \rangle^2}. \end{aligned}$$

2. The second part of the theorem is a direct consequence of the first part. Because W_X and W_Y satisfy assumption 2.2.1 and P is assumed detectable, then $\hat{\lambda}_{\hat{\Sigma}_X,1} = \hat{\theta}_X$ and $\hat{\lambda}_{\hat{\Sigma}_Y,1} = \hat{\theta}_Y$ are the spikes.

   (a) We can write

   $$\frac{1}{\theta - 1} = T_{W,u}(\hat{\theta}_X) = \sum_{i=1}^{m} \frac{\hat{\lambda}_{W_X,i}}{\hat{\theta}_X - \hat{\lambda}_{W_X,i}} \langle \hat{u}_{W,i}, u \rangle^2,$$

   where $\langle \hat{u}_{W_X,i}, u \rangle^2 = w_i$ creates an unit uniform vector w independent of the spectrum of W_X.
   Because

   $$\frac{1}{\theta - 1} - \frac{1}{\hat{\hat{\theta}}_X - 1} = T_{W,u}(\hat{\theta}_X) - \frac{1}{m} \sum_{i=1}^{m} \frac{\hat{\lambda}_{\hat{\Sigma}_X,i}}{\hat{\theta}_X - \hat{\lambda}_{\hat{\Sigma}_X,i}} \xrightarrow[n,m \to \infty]{P} 0,$$

we obtain $\hat{\hat{\theta}}_X \xrightarrow[n,m\to\infty]{P} \theta$. The convergence is obtained using Theorem 4.4.1 to replace $\hat{\lambda}_{W_X,i}$ by $\hat{\lambda}_{\hat{\Sigma}_X,i}$.

(b) Moreover, assuming that $\hat{\theta}_X \xrightarrow[n,m\to\infty]{P} \rho_X$,

$$\langle \hat{u}_X, u\rangle^2 - \alpha_X^2 \quad = \quad -\frac{\theta}{(\theta-1)^2\hat{\theta}_X T'_{W,u}(\hat{\theta}_X)} + \frac{\theta}{(\theta-1)^2\rho_X T'_{W_X}(\rho_X)} \xrightarrow[n,m\to\infty]{P} 0.$$

(c) Finally, a minor result from linear algebra proves the limit of the double angle. If $W = O^t\Lambda O$, with invariant uniform O independent of the bounded spectrum Λ, then for two fixed vectors $v, u \in \mathbf{R}^m$,

$$v^tWu = \tilde{v}^t\Lambda\tilde{u} = \sum_{i=1}^m \hat{\lambda}_{W,i}\tilde{v}_i\tilde{u}_i = \langle u, v\rangle \operatorname{Trace}(W)/m + O_p(1/m).$$

This result is easily proven because $\tilde{v}, \tilde{u}$ are uniform, $\langle\tilde{v},\tilde{u}\rangle = \langle v, u\rangle$ and we easily show that

$$\begin{aligned}
&\mathrm{E}[\tilde{v}_i\tilde{u}_i] = \mathrm{E}\left[\sum_{s,r=1}^m O_{i,s}O_{i,r}v_sv_r\right] = \langle v, u\rangle/m,\\
&\operatorname{Var}\left(\sum_{i=1}^m \tilde{v}_i\tilde{u}_i\right) = 0 \text{ and } \operatorname{Var}(\tilde{v}_i\tilde{u}_i) = O_p\left(1/m^2\right),\\
&\operatorname{Cov}(\tilde{v}_i\tilde{u}_i, \tilde{v}_j\tilde{u}_j) = O_p\left(1/m^3\right).
\end{aligned}$$

Using the notation $\hat{\theta}_X = \hat{\lambda}_{\hat{\Sigma}_X,1}$ and $\hat{\theta}_Y = \hat{\lambda}_{\hat{\Sigma}_Y,1}$ for the estimated eigenvalues, we can combine the first part and this small linear algebra result to obtain

$$\begin{aligned}
\langle\tilde{u}_X,\tilde{u}_Y\rangle^2 &= \frac{\left(u^tW_X(\hat{\theta}_X\mathrm{I}_m - W_X)^{-1}(\hat{\theta}_Y\mathrm{I}_m - W_Y)^{-1}W_Yu\right)^2}{u^tW_X(\hat{\theta}_XI - W_X)^{-2}W_XW_Y(\hat{\theta}_YI - W_Y)^{-2}W_Yu}\\
&= \frac{\left(u^tW_X(\hat{\theta}_X\mathrm{I}_m - W_X)^{-1}(\hat{\theta}_Y\mathrm{I}_m - W_Y)^{-1}W_Yu\right)^2}{\left(-T_{W,u}\left(\hat{\theta}_X\right) - \hat{\theta}_XT'_{W,u}\left(\hat{\theta}_X\right)\right)\left(-T_{W,u}\left(\hat{\theta}_Y\right) - \hat{\theta}_YT'_{W,u}\left(\hat{\theta}_Y\right)\right)}\\
&= \frac{\left(\operatorname{Trace}\left(W_X(\hat{\theta}_X\mathrm{I}_m - W_X)^{-1}\right)u^t(\hat{\theta}_Y\mathrm{I}_m - W_Y)^{-1}W_Yu + O_p\left(\frac{1}{m\theta}\right)\right)^2}{\left(-T_{W,u}\left(\hat{\theta}_X\right) - \hat{\theta}_XT'_{W,u}\left(\hat{\theta}_X\right)\right)\left(-T_{W,u}\left(\hat{\theta}_Y\right) - \hat{\theta}_YT'_{W,u}\left(\hat{\theta}_Y\right)\right)}\\
&= \langle u,\tilde{u}_X\rangle^2\langle u,\tilde{u}_Y\rangle^2 + O_p\left(\frac{1}{m\theta}\right).
\end{aligned}$$

As in the first part, we use the relation between $\hat{u}_X$ and $\tilde{u}_X$ to get

$$\begin{aligned}
\langle\hat{u}_X,\hat{u}_Y\rangle^2 &= \frac{\left\langle\left(\mathrm{I}_m+\bar{P}\right)^{1/2}\tilde{u}_X, \left(\mathrm{I}_m+\bar{P}\right)^{1/2}\tilde{u}_Y\right\rangle^2}{\left(\tilde{u}_X^t(\mathrm{I}_m+\bar{P})\tilde{u}_X\right)\left(\tilde{u}_Y^t(\mathrm{I}_m+\bar{P})\tilde{u}_Y\right)}\\
&= \frac{\left(\langle\tilde{u}_X,u\rangle\langle u,\tilde{u}_Y\rangle\theta + \overbrace{\left(\langle\tilde{u}_X,\tilde{u}_Y\rangle - \langle\tilde{u}_X,u\rangle\langle u,\tilde{u}_Y\rangle\right)}^{\to 0}\right)^2}{\left(1+(\theta-1)\langle\tilde{u}_X,u\rangle^2\right)\left(1+(\theta-1)\langle\tilde{u}_Y,u\rangle^2\right)}.
\end{aligned}$$

This leads to the result,

$$\langle\hat{u}_X,\hat{u}_Y\rangle^2 - \alpha_X^2\alpha_Y^2 \xrightarrow[n,m\to\infty]{P} 0.$$

$\square$

7.1.3 Asymptotic distribution of the eigenvalue and angle

Theorem 5.3.1.
Suppose W_X and W_Y satisfy 2.2.1 with $P = P_X = P_Y$, a detectable perturbation of order $k = 1$. Moreover, we assume that the spectra of of W_X and W_Y are known. We defined

$$\begin{aligned}\hat{\Sigma}_X &= P^{1/2} W_X P^{1/2},\\ \hat{\Sigma}_Y &= P^{1/2} W_Y P^{1/2},\\ P &= \mathrm{I}_m + (\theta - 1) u u^t,\end{aligned}$$

where u is fixed and the eigenvalues of W_X and W_Y are $\hat{\lambda}_{W_X,i}$ and $\hat{\lambda}_{W_Y,i}$ for $i = 1, 2, ..., m$,respectively. We construct the unbiased estimators of θ,

$$\hat{\hat{\theta}}_X \left| \frac{1}{\hat{\hat{\theta}}_X - 1} = \frac{1}{m}\sum_{i=1}^m \frac{\hat{\lambda}_{W_X,i}}{\hat{\theta}_X - \hat{\lambda}_{W_X,i}} \right. \quad \text{and} \quad \hat{\hat{\theta}}_Y \left| \frac{1}{\hat{\hat{\theta}}_Y - 1} = \frac{1}{m}\sum_{i=1}^m \frac{\hat{\lambda}_{W_Y,i}}{\hat{\theta}_Y - \hat{\lambda}_{W_Y,i}} \right.$$

where $\hat{\theta}_X = \hat{\lambda}_{\hat{\Sigma}_X,1}$ and $\hat{\theta}_Y = \hat{\lambda}_{\hat{\Sigma}_Y,1}$ are the largest eigenvalues of respectively $\hat{\Sigma}_X$ and $\hat{\Sigma}_Y$ corresponding to the eigenvectors $\hat{u}_X = \hat{u}_{\hat{\Sigma}_X,1}$ and $\hat{u}_Y = \hat{u}_{\hat{\Sigma}_Y,1}$.
We can also construct these estimators using the $m-1$ smaller eigenvalues of $\hat{\Sigma}_X$ and $\hat{\Sigma}_Y$ instead of $\hat{\lambda}_{W_X,i}$ and $\hat{\lambda}_{W_Y,i}$.

1. *If $\frac{\theta}{\sqrt{m}} \to 0$, we define*

$$M_{s,r,X} \equiv M_{s,r,X}(\rho_X) = \frac{1}{m}\sum_{i=1}^m \frac{\hat{\lambda}^s_{W_X,i}}{(\rho_X - \hat{\lambda}_{W_X,i})^r}, \; M_{s,r,Y} \equiv M_{s,r,Y}(\rho_Y) = \frac{1}{m}\sum_{i=1}^m \frac{\hat{\lambda}^s_{W_X,i}}{(\rho_Y - \hat{\lambda}_{W_X,i})^r},$$

where we assume

$$\rho_X = \mathrm{E}\left[\hat{\theta}_X\right] + o\left(\frac{\theta}{\sqrt{m}}\right), \; \rho_Y = \mathrm{E}\left[\hat{\theta}_Y\right] + o\left(\frac{\theta}{\sqrt{m}}\right).$$

and convergence rate of $(\hat{\theta}_X, \hat{\theta}_Y)$ to (ρ_X, ρ_Y) in $O_p(\theta/\sqrt{m})$,

$$\begin{pmatrix} \hat{\hat{\theta}}_X \\ \langle \hat{u}_X, u\rangle^2 \end{pmatrix} \sim \mathbf{N}\left(\begin{pmatrix} \theta \\ \alpha_X^2 \end{pmatrix}, \frac{1}{m}\begin{pmatrix} \sigma^2_{\theta,X} & \sigma_{\theta,\alpha^2,X} \\ \sigma_{\theta,\alpha^2,X} & \sigma^2_{\alpha^2,X}\end{pmatrix}\right) + \begin{pmatrix} o_p\left(\frac{\theta}{\sqrt{m}}\right) \\ o_p\left(\frac{1}{\theta\sqrt{m}}\right)\end{pmatrix},$$

$$\begin{pmatrix} \hat{\hat{\theta}}_X \\ \hat{\hat{\theta}}_Y \\ \langle \hat{u}_X, \hat{u}_Y\rangle^2 \end{pmatrix} \sim \mathbf{N}\left(\begin{pmatrix} \theta \\ \theta \\ \alpha_{X,Y}^2 \end{pmatrix}, \frac{1}{m}\begin{pmatrix} \sigma^2_{\theta,X} & 0 & \sigma_{\theta,\alpha^2,X} \\ 0 & \sigma^2_{\theta,Y} & \sigma_{\theta,\alpha^2,Y} \\ \sigma_{\theta,\alpha^2,X} & \sigma_{\theta,\alpha^2,Y} & \sigma^2_{\alpha^2,X,Y}\end{pmatrix}\right) + \begin{pmatrix} o_p\left(\frac{\theta}{\sqrt{m}}\right) \\ o_p\left(\frac{\theta}{\sqrt{m}}\right) \\ o_p\left(\frac{1}{\theta\sqrt{m}}\right)\end{pmatrix},$$

where

$$\begin{aligned}\alpha_X^2 &= \frac{\theta}{(\theta-1)^2}\frac{1}{\rho_X M_{1,2,X}},\\ \alpha_{X,Y}^2 &= \frac{\theta^2}{(\theta-1)^4}\frac{1}{\rho_X\rho_Y M_{1,2,X}M_{1,2,Y}},\\ \sigma^2_{\theta,X} &= \frac{2\left(M_{2,2,X} - M^2_{1,1,X}\right)}{M^4_{1,1,X}},\end{aligned}$$

$$\sigma^2_{\alpha^2,X} = \frac{2\theta^2}{\left((\theta-1)\rho_X M_{1,2,X}\right)^4}\left(\rho_X^2\left(M_{2,4,X}-M_{1,2,X}^2\right)+\left(2\rho_X\frac{M_{1,3,X}}{M_{1,2,X}}-1\right)^2\left(M_{2,2,X}-M_{1,1,X}^2\right)\right.$$
$$\left.-2\rho_X\left(2\rho_X\frac{M_{1,3,X}}{M_{1,2,X}}-1\right)\left(M_{2,3,X}-\frac{M_{1,1,X}}{M_{1,2,X}}\right)\right),$$
$$\sigma_{\theta,\alpha^2,X} = \frac{2\theta}{M_{1,1,X}^2 M_{1,2,X}^3\rho_X^2(-1+\theta)^2}\Big(M_{1,1,X}M_{1,2,X}^2\rho_X+2M_{1,3,X}M_{2,2,X}\rho_X$$
$$+M_{1,1,X}^2(M_{1,2,X}-2M_{1,3,X}\rho_X)-M_{1,2,X}(M_{2,2,X}+M_{2,3,X}\rho_X)\Big),$$
$$\sigma^2_{\alpha^2,X,Y} = \sigma^2_{\alpha^2,X}\alpha_Y^4+\sigma^2_{\alpha^2,Y}\alpha_X^4+4\alpha_{X,Y}^2(1-\alpha_X^2)(1-\alpha_X^2).$$

2. *If $\frac{\theta}{\sqrt{m}}\to\infty$ (Assumption 2.2.2 (A1)), then we can simplify the formulas. We define*

$$M_{r,X}=\frac{1}{m}\sum_{i=1}^m\hat\lambda_{W_X,i}^r \quad \textit{and} \quad M_{r,Y}=\frac{1}{m}\sum_{i=1}^m\hat\lambda_{W_Y,i}^r.$$

Using this notation,

$$\begin{pmatrix}\hat{\hat\theta}_X\\ \langle\hat u_X,u\rangle^2\end{pmatrix}\sim\mathbf{N}\left(\begin{pmatrix}\theta+O_p(1)\\ 1+\frac{1-M_{2,X}}{\theta}+O_p\left(\frac{1}{\theta^2}\right)\end{pmatrix},\right.$$
$$\left.\frac{1}{m}\begin{pmatrix}2\theta^2(1-M_{2,X}) & 2\left(2M_{2,X}^2-M_{2,X}-M_{3,X}\right)\\ 2\left(2M_{2,X}^2-M_{2,X}-M_{3,X}\right) & \frac{2}{\theta^2}\left(4M_{2,X}^3-M_{2,X}^2-4M_{2,X}M_{3,X}+M_{4,X}\right)\end{pmatrix}\right)+\begin{pmatrix}o_p\left(\frac{\theta}{\sqrt{m}}\right)\\ o_p\left(\frac{1}{\theta\sqrt{m}}\right)\end{pmatrix},$$

$$\begin{pmatrix}\hat{\hat\theta}_X\\ \hat{\hat\theta}_Y\\ \langle\hat u_X,\hat u_Y\rangle^2\end{pmatrix}\sim\mathbf{N}\left(\begin{pmatrix}\theta+O_p(1)\\ \theta+O_p(1)\\ 1+\frac{2-M_{2,X}-M_{2,Y}}{\theta}+O_p\left(\frac{1}{\theta^2}\right)\end{pmatrix},\right.$$
$$\left.\frac{1}{m}\begin{pmatrix}2\theta^2(1-M_{2,X}) & 0 & 2\left(2M_{2,X}^2-M_{2,X}-M_{3,X}\right)\\ 0 & 2\theta^2(1-M_{2,Y}) & 2\left(2M_{2,Y}^2-M_{2,Y}-M_{3,Y}\right)\\ 2\left(2M_{2,X}^2-M_{2,X}-M_{3,X}\right) & 2\left(2M_{2,Y}^2-M_{2,Y}-M_{3,Y}\right) & \frac{S}{\theta^2}\end{pmatrix}\right)+\begin{pmatrix}o_p\left(\frac{\theta}{\sqrt{m}}\right)\\ o_p\left(\frac{\theta}{\sqrt{m}}\right)\\ o_p\left(\frac{1}{\theta\sqrt{m}}\right)\end{pmatrix},$$
$$S=2\left(4M_{2,X}^3-M_{2,X}^2-4M_{2,X}M_{3,X}+M_{4,X}\right)+2\left(4M_{2,Y}^3-M_{2,Y}^2-4M_{2,Y}M_{3,Y}+M_{4,Y}\right)+4(M_{2,Y}-1)(M_{2,X}-1).$$

Moreover, the asymptotic distributions of $\hat\theta_X$ and $\hat{\hat\theta}_X$ are the same.

3. *If $\frac{\theta}{\sqrt{m}}\to d$, a finite constant, then a mixture of the two first scenarios describes the first two moments of the joint distribution.*
The formula of the second moment is asymptotically the same as the variance formula when $\frac{\theta}{\sqrt{m}}\to\infty$.
The formula of the first moment is asymptotically the same as the expectation formula when $\frac{\theta}{\sqrt{m}}\to 0$.

4. *The random variables can be expressed as functions of invariant unit random statistics of the form:*

$$\overset{u}{M}_{r,s,X}(\rho) \quad = \quad \sum_{i=1}^m\frac{\hat\lambda_{W_X,i}^r}{\left(\rho-\hat\lambda_{W_X,i}\right)^s}\langle\hat u_{W_X,i},u\rangle^2.$$

(Assuming a canonical perturbation leads to a simpler formula)

- *Exact distributions:*

$$\hat\theta_X\left|\frac{1}{\theta-1}=\overset{u}{M}_{1,1,X}\left(\hat\theta_X\right),\right.$$
$$\langle\hat u_X,e_1\rangle^2=\frac{\theta}{(\theta-1)^2}\frac{1}{\hat\theta_X\overset{u}{M}_{1,2,X}\left(\hat\theta_X\right)}.$$

Moreover,

$$\begin{aligned}
\langle \hat{u}_X, \hat{u}_Y \rangle &= \langle \hat{u}_X, e_1 \rangle \langle \hat{u}_Y, e_1 \rangle + \sqrt{1 - \langle \hat{u}_X, e_1 \rangle^2} \sqrt{1 - \langle \hat{u}_Y, e_1 \rangle^2} Z, \\
\sum_{i=2}^{m} \hat{u}_{X,1} \hat{u}_{Y,1} &= \sqrt{1 - \langle \hat{u}_X, e_1 \rangle^2} \sqrt{1 - \langle \hat{u}_Y, e_1 \rangle^2} Z, \\
Z &\sim \mathbf{N}\left(0, \frac{1}{m}\right) + O_p\left(\frac{1}{m}\right),
\end{aligned}$$

where Z is independent of $\langle \hat{u}_X, e_1 \rangle$, $\langle \hat{u}_Y, e_1 \rangle$, $\hat{\theta}_X$ and $\hat{\theta}_Y$. In order to get the exact distribution, we should replace Z by $\sum_{i=1}^{m-1} v_i \tilde{v}_i$ where $v_i \tilde{v}_i$ are independent unit invariant random vectors.

- *Approximations:*

$$\begin{aligned}
\hat{\theta}_X &= \rho + \frac{\left(\overset{u}{M}_{1,1,X}(\rho) - M_{1,1,X}(\rho)\right)}{M_{1,2,X}(\rho)} + O_p\left(\frac{\theta}{m}\right) \\
&= \theta \overset{u}{M}_{1,X} + O_p(1), \\
\hat{\hat{\theta}}_X &= \theta + (\theta - 1)^2 \left(\overset{u}{M}_{1,1,X}(\rho) - M_{1,1,X}(\rho_X)\right) + O_p\left(\frac{\theta}{m}\right).
\end{aligned}$$

We provide three methods of estimation of the angle in order to estimate it for all θ.

$$\begin{aligned}
\langle \hat{u}_X, e_1 \rangle^2 &= \frac{\theta}{(\theta-1)^2} \left(\frac{1}{\rho_X M_{1,2,X}(\rho_X)} + \left(\frac{2M_{1,3,X}(\rho_X)}{M_{1,2,X}(\rho_X)} - \frac{1}{\rho_X} \right) \frac{\overset{u}{M}_{1,1,X}(\rho_X) - \frac{1}{\theta-1}}{(M_{1,2,X}(\rho_X))^2} \right. \\
&\qquad \left. - \frac{\overset{u}{M}_{1,2,X}(\rho_X) - M_{1,2,X}(\rho_X)}{\rho_X (M_{1,2,X}(\rho_X))^2} \right) + O_p\left(\frac{1}{m}\right), \\
&= 1 + \frac{1}{\theta}\left(1 - \overset{u}{M}_{2,X} + 2M_{2,X}\left(\overset{u}{M}_{1,X} - 1\right)\right) \\
&\qquad + \frac{1}{\theta^2}\left(1 - 2\overset{u}{M}_{2,X} + 3\overset{u}{M}{}^2_{2,X} - 2\overset{u}{M}_{3,X}\right) + O_p\left(\frac{1}{\theta^3}\right) + O_p\left(\frac{1}{\theta m}\right), \\
&= 1 + \frac{1}{\theta} - \frac{1}{\theta}\overset{u}{M}_{2,X} + \frac{2}{\theta} M_{2,X}\left(\overset{u}{M}_{1,X} - 1\right) + O_p\left(\frac{1}{\theta^2}\right) + O_p\left(\frac{1}{\theta m}\right)
\end{aligned}$$

Finally, the double angle is such that

$$\langle \hat{u}_X, \hat{u}_Y \rangle = \langle \hat{u}_X, e_1 \rangle \langle \hat{u}_Y, e_1 \rangle + \frac{\sqrt{M_{2,X} - 1}\sqrt{M_{2,Y} - 1}}{\theta} Z + O_p\left(\frac{1}{\theta^2 \sqrt{m}}\right).$$

Remark 5.3.1.1.
If the spectra of W_X and W_Y tend sufficiently fast to a Marcenko-Pastur distribution of parameter c we have,

$$\begin{pmatrix} \hat{\hat{\theta}}_X \\ \langle \hat{u}_X, u_0 \rangle^2 \end{pmatrix} \overset{Asy}{\sim} \mathbf{N}\left(\begin{pmatrix} \theta \\ \frac{1 - \frac{c}{(\theta-1)^2}}{1 + \frac{c}{\theta-1}} \end{pmatrix}, \frac{1}{m} \begin{pmatrix} -\frac{2c(\theta-1)^2\theta^2}{c-(\theta-1)^2} & -\frac{2c^2(\theta-1)\theta^3}{(c-(\theta-1)^2)(c+\theta-1)^2} \\ -\frac{2c^2(\theta-1)\theta^3}{(c-(\theta-1)^2)(c+\theta-1)^2} & -\frac{2c^2\theta^2\left(c^2+(\theta(\theta+2)-2)c+(\theta-1)^2\right)}{(c-(\theta-1)^2)(c+\theta-1)^4} \end{pmatrix} \right)$$

and

$$\begin{pmatrix} \hat{\hat{\theta}}_X \\ \hat{\hat{\theta}}_Y \\ \langle \hat{u}_X, \hat{u}_Y \rangle^2 \end{pmatrix} \overset{Asy}{\sim} \mathbf{N} \left(\begin{pmatrix} \theta \\ \theta \\ \left(\frac{1 - \frac{c}{(\theta-1)^2}}{1 + \frac{c}{\theta-1}} \right)^2 \end{pmatrix}, \right.$$
$$\left. \frac{1}{m} \begin{pmatrix} -\frac{2c(\theta-1)^2\theta^2}{c-(\theta-1)^2} & 0 & -\frac{2c^2(\theta-1)\theta^3}{\left(c-(\theta-1)^2\right)(c+\theta-1)^2} \\ 0 & -\frac{2c(\theta-1)^2\theta^2}{c-(\theta-1)^2} & -\frac{2c^2(\theta-1)\theta^3}{\left(c-(\theta-1)^2\right)(c+\theta-1)^2} \\ -\frac{2c^2(\theta-1)\theta^3}{\left(c-(\theta-1)^2\right)(c+\theta-1)^2} & -\frac{2c^2(\theta-1)\theta^3}{\left(c-(\theta-1)^2\right)(c+\theta-1)^2} & \frac{4c^2\theta^2\left(c-(\theta-1)^2\right)^2\left(c^3+4c^2(\theta-1)+c(\theta-1)(\theta(\theta+5)-5)+2(\theta-1)^3\right)}{(\theta-1)^4(c+\theta-1)^7} \end{pmatrix} \right).$$

If θ tends to infinity, then

$$\begin{pmatrix} \hat{\hat{\theta}}_X \\ \langle \hat{u}_X, u \rangle^2 \end{pmatrix} \overset{Asy}{\sim} \mathbf{N} \left(\begin{pmatrix} \theta \\ \frac{1 - \frac{c}{(\theta-1)^2}}{1 + \frac{c}{\theta-1}} \end{pmatrix}, \frac{1}{m} \begin{pmatrix} 2c\theta^2 & 2c^2 \\ 2c^2 & \frac{2c^2(c+1)}{\theta^2} \end{pmatrix} \right).$$

Moreover,

$$\begin{pmatrix} \hat{\hat{\theta}}_X \\ \hat{\hat{\theta}}_Y \\ \langle \hat{u}_X, \hat{u}_X \rangle^2 \end{pmatrix} \overset{Asy}{\sim} \mathbf{N} \left(\begin{pmatrix} \theta \\ \theta \\ \frac{1 - \frac{c}{(\theta-1)^2}}{1 + \frac{c}{\theta-1}} \end{pmatrix}, \frac{1}{m} \begin{pmatrix} 2c\theta^2 & 0 & 2c^2 \\ 0 & 2c\theta^2 & 2c^2 \\ 2c^2 & 2c^2 & \frac{4c^2(c+2)}{\theta^2} \end{pmatrix} \right).$$

(Page 74)

Proof. **Theorem 5.3.1**

The proof of the theorem is divided into six parts. First, we simplify the problem and introduce the notation. Then, we recall the characterisation of eigenvalues and eigenvectors of perturbations of order 1. This allows us to demonstrate the fourth part of the theorem divided into three points: eigenvalues $\left(\hat{\hat{\theta}}_X\right)$, angle $\left(\langle \hat{u}_X, u \rangle^2\right)$ and double angle $\left(\langle \hat{u}_X, \hat{u}_Y \rangle^2\right)$. Finally, the three first parts of the theorem are just consequences of the fourth part. The computation of the particular case assuming a fast convergence of the spectrum to Marcenko-Pastur distribution is left to the reader.

Prerequisite discussion Using the invariant by rotation property, we can assume without loss of generality that P is canonical. Therefore, $u = \epsilon_1$.
In this proof, we use the following notation

$$\begin{aligned} \overset{u}{M}_{r,s,X}(\rho) &= \sum_{i=1}^{m} \frac{\hat{\lambda}_{W_X,i}^r}{\left(\rho - \hat{\lambda}_{W_X,i}\right)^s} \hat{u}_{W_X,i,1}^2, \\ M_{r,s,X}(\rho) &= \sum_{i=1}^{m} \frac{\hat{\lambda}_{W_X,i}^r}{\left(\rho - \hat{\lambda}_{W_X,i}\right)^s}, \\ M_{r,s,\hat{\Sigma}_X}(\rho) &= \sum_{i=2}^{m} \frac{\hat{\lambda}_{\hat{\Sigma}_X,i}^r}{\left(\rho - \hat{\lambda}_{\hat{\Sigma}_X,i}\right)^s}. \end{aligned}$$

In particular, $\overset{u}{M}_{1,1,X}(\rho) = T_{W_X,\epsilon_1}(\rho)$ and $M_{1,1,X}(\rho) = T_{W_X}(\rho)$.
In Theorem 4.4.1, we show that $M_{r,s,X}(\rho)$ and $M_{r,s,\hat{\Sigma}_X}(\rho)$ are asymptotically the same. Therefore, both transforms lead to the same asymptotic unbiased estimator $\hat{\hat{\theta}}_X$ and we can prove results only for $M_{r,s,X}(\rho)$.

Characterisation: Using characterisations presented in Benaych-Georges and Rao [2009],

$$\begin{aligned}
\theta &= 1 + \frac{1}{\overset{u}{M}_{1,1,X}(\hat{\theta}_X)} \\
&= 1 + \frac{1}{\sum_{i=1}^{m} \frac{\hat{\lambda}_{W_X,i}}{\hat{\theta}_X - \hat{\hat{\lambda}}_{W_X,i}} \hat{u}^2_{W_X,i,1}}
\end{aligned}$$

and

$$\langle \hat{u}_X, u \rangle^2 = \frac{\theta}{(\theta-1)^2 \hat{\theta}_X \overset{u}{M}_{1,2,X}(\hat{\theta}_X)}.$$

This result is generalised in Theorem 5.2.1 and its proof can be found in page 103.

Unit invariant vector statistics for eigenvalues: Assuming that $\mathrm{E}\left[\hat{\theta}_X\right] = \rho_X$ and the rate of convergence is $\theta/\sqrt{m}$, then

$$\frac{1}{\theta-1} = \sum_{i=1}^{m} \left(\frac{\hat{\lambda}_{W_X,i}}{\rho_X - \hat{\lambda}_{W_X,i}} \hat{u}^2_{W_X,i,1} - \frac{\hat{\lambda}_{W_X,i}}{\left(\rho_X - \hat{\lambda}_{W_X,i}\right)^2} \hat{u}^2_{W_X,i,1} \left(\hat{\theta}_X - \rho_X\right) \right.$$
$$\left. + \frac{\hat{\lambda}_{W_X,i}}{\left(\tilde{\theta}_X - \hat{\lambda}_{W_X,i}\right)^3} \hat{u}^2_{W_X,i,1} \left(\hat{\theta}_X - \rho_X\right)^2 \right).$$

Therefore,

$$\begin{aligned}
\hat{\theta}_X &= \rho_X + \frac{\overset{u}{M}_{1,1,X}(\rho_X) - \frac{1}{\theta-1}}{\overset{u}{M}_{1,2,X}(\rho_X)} + \frac{\overset{u}{M}_{1,3,X}(\tilde{\theta}_X)}{\overset{u}{M}_{1,2,X}(\tilde{\theta}_X)} \left(\hat{\theta}_X - \rho_X\right)^2, \text{ where } \tilde{\theta}_X \sim \rho_X, \\
&= \rho_X + \frac{\overset{u}{M}_{1,1,X}(\rho) - \frac{1}{\theta-1}}{M_{1,2,X}(\rho)} - \frac{\overset{u}{M}_{1,1,X}(\rho_X) - \frac{1}{\theta-1}}{M_{1,2,X}(\rho_X)^2} \left(\overset{u}{M}_{1,2,X}(\rho) - M_{1,2,X}(\rho_X)\right) + O_p\left(\frac{\theta}{m}\right).
\end{aligned}$$

We show that $\overset{u}{M}_{1,1,X}(\rho_X) - \frac{1}{\theta-1} = O_p\left(\frac{1}{\sqrt{m}\theta}\right)$.

$$\begin{aligned}
\overset{u}{M}_{1,1,X}(\rho_X) - \frac{1}{\theta-1} &= \overset{u}{M}_{1,1,X}(\rho_X) - \overset{u}{M}_{1,1,X}(\hat{\theta}_X) \\
&= \overset{u}{M}_{1,2,X}(\rho_X)\left(\hat{\theta}_X - \rho_X\right) + O_p\left(\frac{1}{m\theta}\right) \\
&= O_p\left(\frac{1}{\sqrt{m}\theta}\right).
\end{aligned}$$

Moreover, a similar argument shows that $\left(\overset{u}{M}_{1,2,X}(\rho_X) - M_{1,2,X}(\rho_X)\right) = O_p\left(\frac{1}{\sqrt{m}\theta^2}\right)$.
This leads to the first important equation:

$$\hat{\theta}_X = \rho_X + \frac{\overset{u}{M}_{1,1,X}(\rho_X) - \frac{1}{\theta-1}}{M_{1,2,X}(\rho_X)} + O_p\left(\frac{\theta}{m}\right) \tag{7.4}$$

Then, we study the unbiased eigenvalue,

$$\begin{aligned}
\hat{\hat{\theta}}_X - 1 &= \frac{1}{M_{1,2,X}(\hat{\theta}_X)} \\
&= \frac{m}{\sum_{i=1}^m \frac{\hat{\lambda}_{W_X,i}}{\hat{\theta}_X - \hat{\lambda}_{W_X,i}}} \\
&= \frac{1}{M_{1,1,X}(\rho_X) - M_{1,2,X}(\rho_X)\left(\hat{\theta}_X - \rho_X\right) + O_p\left(\frac{1}{\theta m}\right)} \\
&= \frac{1}{M_{1,1,X}(\rho_X) - \overset{u}{M}_{1,1,X}(\rho_X) + \frac{1}{\theta-1}} + O_p\left(\frac{\theta}{m}\right) \\
&= \theta - 1 - (\theta-1)^2\left(M_{1,1,X}(\rho_X) - \overset{u}{M}_{1,1,X}(\rho_X)\right) + O_p\left(\frac{\theta}{m}\right).
\end{aligned}$$

This leads to the second important equation:

$$\hat{\hat{\theta}}_X = \theta + (\theta-1)^2\left(\overset{u}{M}_{1,1,X}(\rho_X) - M_{1,1,X}(\rho_X)\right) + O_p\left(\frac{\theta}{m}\right). \tag{7.5}$$

When θ is large we can use

$$\frac{1}{\theta-1} = \frac{1}{\hat{\theta}_X}\sum_{i=1}^m \hat{\lambda}_{W_X,i}\hat{u}^2_{W_X,i,1} + \frac{1}{\hat{\theta}_X^2}\sum_{i=1}^m \hat{\lambda}^2_{W_X,i}\hat{u}^2_{W_X,i,1} + O_p\left(\frac{1}{\theta^3}\right)$$

to show

$$\begin{aligned}
\hat{\theta}_X &= (\theta-1)\overset{u}{M}_{1,X} + \frac{\overset{u}{M}_{2,X}}{M_{1,X}} + O_p\left(\frac{1}{m}\right) + O_p\left(\frac{1}{\theta}\right) && (7.6)\\
&= \theta\overset{u}{M}_{1,X} + O_p(1). && (7.7)
\end{aligned}$$

Unit invariant vector statistics for angles: We propose 3 Taylor expansions of the angle.

1: Expansion of $\hat{\theta}_X$ around ρ_X. Applying a Taylor expansion around ρ_X as computed previously for $\hat{\theta}$ (7.4) and a similar one for $\overset{u}{M}_{1,2,X}(\hat{\theta}_X)$ lead to

$$\begin{aligned}
\langle \hat{u}_X, u\rangle^2 &= \frac{\theta}{(\theta-1)^2\hat{\theta}_X\overset{u}{M}_{1,2,X}(\hat{\theta}_X)} \\
&= \frac{\theta}{(\theta-1)^2}\left(\frac{1}{\rho_X M_{1,2,X}(\rho_X)} + \left(\frac{2M_{1,3,X}(\rho_X)}{M_{1,2,X}(\rho_X)} - \frac{1}{\rho_X}\right)\frac{\overset{u}{M}_{1,1,X}(\rho_X) - \frac{1}{\theta-1}}{\left(M_{1,2,X}(\rho_X)\right)^2}\right. \\
&\qquad \left. - \frac{\overset{u}{M}_{1,2,X}(\rho_X) - M_{1,2,X}(\rho_X)}{\rho_X\left(M_{1,2,X}(\rho_X)\right)^2}\right) + O_p\left(\frac{1}{m}\right).
\end{aligned}$$

2: Expansion of order 2 of the transform when θ is large. Using the fact that θ is large with (7.6) and that $\overset{u}{M}_{1,1}(\hat{\theta}_X) = 1/(\theta-1)$ leads to a different estimation,

$$\begin{aligned}
\langle \hat{u}_X, u\rangle^2 &= 1 + \frac{1}{\theta} - \frac{\theta}{\hat{\theta}_X^2}\overset{u}{M}_{2,X} + O_p\left(\frac{1}{\theta^2}\right) \\
&= 1 + \frac{1}{\theta} - \frac{1}{\theta}\frac{\overset{u}{M}_{2,X}}{\left(\overset{u}{M}_{1,X}\right)^2} + O_p\left(\frac{1}{\theta^2}\right) \\
&= 1 + \frac{1}{\theta} - \frac{1}{\theta}\overset{u}{M}_{2,X} + \frac{2}{\theta}M_{2,X}\left(\overset{u}{M}_{1,X} - 1\right) + O_p\left(\frac{1}{\theta^2}\right) + O_p\left(\frac{1}{\theta m}\right).
\end{aligned}$$

3: Expansion of order 3 of the transform when θ is large. Using better estimations of $\hat{\theta}_X$ (7.7) and $\overset{u}{M}_{1,2,X}(\hat{\theta}_X)$ when θ is large leads to

$$\begin{aligned}\langle \hat{u}_X, u\rangle^2 &= 1+\frac{1}{\theta}\left(1-\overset{u}{M}_{2,X}+2M_{2,X}\left(\overset{u}{M}_{1,X}-1\right)\right)\\ &\quad +\frac{1}{\theta^2}\left(1-2\overset{u}{M}_{2,X}+3\overset{u}{M}{}_{2,X}^{2}-2\overset{u}{M}_{3,X}\right)+O_p\left(\frac{1}{\theta^3}\right)+O_p\left(\frac{1}{\theta m}\right).\end{aligned}$$

The study of the angle $\langle \hat{u}_X, u\rangle^2$ is separated in many cases when θ grows,

θ finite		$\frac{\theta^3}{\sqrt{m}} \to$ constant		$\frac{\theta^2}{\sqrt{m}} \to$ constant		$\frac{\theta^2}{m} \to$ constant	
A	B	C	D	E	F	G	H

A: A delta method using the first expansion of the angle leads to the result. In this case the standard error is of order $1/\sqrt{m}$; therefore, the error of order $1/m$ is negligible.

B − >F: Studying the first expansion shows that the error is not important. Indeed assuming that $\theta \to \infty$ allows an expansion for large θ. This shows

$$\mathrm{Var}\left(\langle \hat{u}_X, u\rangle^2\right) \sim \frac{1}{\theta^2 m}.$$

Moreover, $1-\mathrm{E}\left[\langle \hat{u}_X, u\rangle^2\right]$ is of order $1/\theta$. Assuming $\epsilon \sim 1/m$ is the error of the first expansion, then

$$\begin{aligned}\frac{\epsilon}{1/\theta} &\sim \frac{\frac{1}{m}}{\frac{1}{\theta}}=\frac{\theta}{m}\to 0,\\ \frac{\epsilon}{\sqrt{\mathrm{Var}\left(\langle \hat{u}_X, u\rangle^2\right)}} &\sim \frac{\frac{1}{m}}{\frac{1}{\theta\sqrt{m}}}=\frac{\theta}{\sqrt{m}}\to 0.\end{aligned}$$

D-E-F: The previous arguments are still valid in these cases. Moreover, we can also show the same result using the third expansion. Therefore, the estimations are asymptotically the same.

G: This scenario is the most difficult and only the third approximation works in this case. However,

- The expectations obtained by expansions 1 and 3 are asymptotically the same.
- The variances obtained by expansions 2 and 3 are asymptotically the same.

The equivalence between the variance is direct, but the expectation requires some computations left to the reader.

H: Using expansion 2 leads to a negligible error compared to the standard error and the expectation. This study provides the behaviour of the statistic for all θ

- If $\frac{\theta}{\sqrt{m}} \to 0$, then we use the first expansion.
- If $\frac{\theta}{\sqrt{m}} \to \infty$, then we use the second expansion.
- If $\frac{\theta}{\sqrt{m}} \to d$, where d is a fixed constant, then we use the first expansion to estimate the expectation and the second expansion to estimate the variance.

Unit invariant vector statistics for double angle: Assuming that $\hat{u}_X$ and $\hat{u}_Y$ are the first eigenvectors of $\hat{\Sigma}_X$ and $\hat{\Sigma}_Y$ respectively, then without loss of generality, if $u = e_1$,

$$\begin{aligned}
\langle \hat{u}_X, \hat{u}_Y \rangle &= \hat{u}_{X,1}\hat{u}_{Y,1} + \sum_{i=2}^{m} \hat{u}_{X,i}\hat{u}_{Y,i} \\
&= \hat{u}_{X,1}\hat{u}_{Y,1} + \sqrt{1-\hat{u}_{X,1}^2}\sqrt{1-\hat{u}_{Y,1}^2}\,Z/\sqrt{m} \\
&= \hat{u}_{X,1}\hat{u}_{Y,1} + \sqrt{1-\alpha_X^2}\sqrt{1-\alpha_Y^2}\,Z/\sqrt{m},
\end{aligned}$$

where α_X^2 and α_Y^2 are the limits of $\hat{u}_{X,1}^2$ and $\hat{u}_{Y,1}^2$ respectively. Because the eigenvectors are invariant by rotations that do not affect the first component, we can easily show that Z is independent of $\hat{u}_{X,1}$ and $\hat{u}_{Y,1}$. Moreover, Z is the scalar product of two independent unit invariant vectors of size $m-1$ and can be estimated by a standard Normal divided by $\sqrt{m}$.

Joint distribution: Using Theorem 4.4.1 and applying a simple delta method to the estimation of $\hat{\theta}_X$, $\hat{\theta}_Y$, $\langle \hat{u}_X, \hat{u}_Y \rangle$, $\langle \hat{u}_X, u \rangle$, $\langle \hat{u}_Y, u \rangle$ and Z leads directly to the joint distribution.
The computations are lengthy but straightforward. Therefore, it is left to the reader, who simply needs to separate the three rates of $\theta/\sqrt{m}$ as previously.

□

7.1.4 Induction proof of Invariant Theorems

In this section, we prove some invariant theorems all at once by induction. The procedure is summarize in Figure 7.1. First we initialize the induction in pink. Then, the induction assumes the proven results in the grey part and tries to prove the blue, red and green parts.

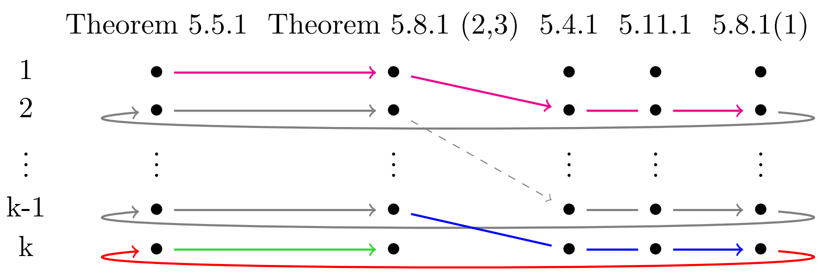


Figure 7.1 – Proof procedure.

Theorem 5.5.1. *(Invariant Angle Theorem)*
Suppose that W satisfies Assumption 2.2.1 and

$$\tilde{P}_s = \mathrm{I}_m + (\theta_s - 1)e_s e_s^t, \ \textit{for } s = 1, 2, ..., k,$$

$$P_k = \mathrm{I}_m + \sum_{i=1}^{k}(\theta_i - 1)e_i e_i^t \ \textit{respects 2.2.2 (A4)},$$

where $\theta_1 > \theta_2 > ... > \theta_k$. We define

$$\begin{aligned}
\hat{\Sigma}_{\tilde{P}_s} &= \tilde{P}_s^{1/2} W \tilde{P}_s^{1/2}, \\
\hat{\Sigma}_{P_k} &= P_k^{1/2} W P_k^{1/2}.
\end{aligned}$$

Moreover, for $s = 1, 2, ..., k$, we define

$$\hat{u}_{\tilde{P}_s,1}, \hat{\theta}_{\tilde{P}_s,1} \quad s.t. \quad \hat{\Sigma}_{\tilde{P}_s} \hat{u}_{\tilde{P}_s,1} = \hat{\theta}_{\tilde{P}_s,1} \hat{u}_{\tilde{P}_s,1},$$
$$\hat{u}_{P_k,s}, \hat{\theta}_{P_k,s} \quad s.t. \quad \hat{\Sigma}_{P_k} \hat{u}_{P_k,s} = \hat{\theta}_{P_k,s} \hat{u}_{P_k,s},$$

where $\hat{\theta}_{\tilde{P}_s,1} = \hat{\lambda}_{\hat{\Sigma}_{\tilde{P}_s,1}}$ *and* $\hat{\theta}_{P_k,s} = \hat{\lambda}_{\hat{\Sigma}_{P_k},s}$.

1. *Then,*

$$\boxed{\sum_{i=1}^{k} \hat{u}^2_{P_k,s,i} = \hat{u}^2_{\tilde{P}_s,1,s} + O_p\left(\frac{1}{\theta_s m}\right).}$$

   Therefore, the distribution of $\sum_{i=1}^{k} \hat{u}^2_{P_k,s,i}$ *is asymptotically the same as the distribution of* $\hat{u}^2_{\tilde{P}_s,1,s}$ *studied in Theorem 5.3.1.*

2. *Moreover,*

$$\hat{u}^2_{P_k,s,s} = \hat{u}^2_{\tilde{P}_s,1,s} + O_p\left(\frac{1}{m}\right).$$

Remark 5.5.1.1.

1. If

$$\hat{u}^2_{P_1,1,1} \sim \mathbf{N}\left(\alpha^2, \frac{\sigma^2_{\alpha^2}}{\theta_1^2 m}\right) + o_p\left(\frac{1}{\theta_1\sqrt{m}}\right),$$

   then

$$\sum_{i=1}^{k} \hat{u}^2_{P_k,1,i} \sim \mathbf{N}\left(\alpha^2, \frac{\sigma^2_{\alpha^2}}{\theta_1^2 m}\right) + o_p\left(\frac{1}{\theta_1\sqrt{m}}\right),$$

   where $\alpha^2 = 1 - \frac{M_2-1}{\theta_1} + O_p\left(\frac{1}{\theta_1^2}\right) < 1$ can be computed more precisely as in Theorem 5.3.1.

2. Assuming a Marcenko-Pastur spectrum, $\alpha^2 = \frac{1-\frac{c}{(\theta_1-1)^2}}{1+\frac{c}{\theta_1-1}}$ and $\sigma^2_{\alpha^2} = 2c^2(c+1) + o_\theta(1)$. In particular if $\frac{\theta_1}{\sqrt{m}}$ is large, then $\alpha^2 \approx 1 - c/\theta_1$,

3. In the general case, if $\frac{\theta_1}{\sqrt{m}}$ is large,

$$\alpha^2 \approx 1 + \frac{1 - M_{2,X}}{\theta_1} \text{ and } \sigma^2_{\alpha^2} \approx 2\left(4M^3_{2,X} - M^2_{2,X} - 4M_{2,X}M_{3,X} + M_{4,X}\right).$$

(Page 78)

Theorem 5.8.1.
Suppose Assumption 2.2.1 holds with canonical P and 2.2.2(A4). We define:

$$U = \begin{pmatrix} \hat{u}^t_{P_k,1} \\ \hat{u}^t_{P_k,2} \\ \vdots \\ \hat{u}^t_{P_k,m} \end{pmatrix} = \begin{pmatrix} \hat{u}_{P_k,1:k,1:k} & \hat{u}_{P_k,1:k,k+1:m} \\ \hat{u}_{P_k,k+1:m,1:k} & \hat{u}_{P_k,k+1:m,k+1:m}. \end{pmatrix}$$

To simplify the result we assume the sign convention,

For $s = 1, 2, ..., k$ and $i = 1, 2, ..., s$, $\hat{u}_{P_s,i,i} > 0$.

1. *Without loss of generality on the k first components, the k^{th} element of the first eigenvector is*

$$\begin{aligned}
\hat{u}_{P_k,1,k} &= \frac{\sqrt{\theta_k}\theta_1}{|\theta_k-\theta_1|}\hat{u}_{P_{k-1},1,k} + O_p\left(\frac{\min(\theta_1,\theta_k)}{\theta_1^{1/2}\theta_k^{1/2}m}\right) + O_p\left(\frac{1}{\sqrt{\theta_1\theta_k m}}\right) \\
&= \frac{\theta_1\sqrt{\theta_k}}{|\theta_k-\theta_1|}\frac{1}{m}\sqrt{1-\hat{\alpha}_1^2}\ Z + O_p\left(\frac{\min(\theta_1,\theta_k)}{\theta_1^{1/2}\theta_k^{1/2}m}\right) + O_p\left(\frac{1}{\sqrt{\theta_1\theta_k m}}\right) \\
&= \frac{\sqrt{\theta_1\theta_k}}{|\theta_k-\theta_1|}\frac{1}{\sqrt{m}}\sqrt{M_2-1}\ Z + O_p\left(\frac{\min(\theta_1,\theta_k)}{\theta_1^{1/2}\theta_k^{1/2}m}\right) + O_p\left(\frac{1}{\sqrt{\theta_1\theta_k m}}\right),
\end{aligned}$$

where Z is a standard normal and $M_2 = \frac{1}{m}\sum_{i=1}^m \hat{\lambda}_{W,i}^2$ is obtained by conditioning on the spectrum.

- *Therefore knowing the spectrum and assuming $\theta_1, \theta_k \to \infty$,*

$$\hat{u}_{P_k,1,k} \overset{Asy}{\sim} \mathbf{N}\left(0, \frac{\theta_1\theta_k}{|\theta_1-\theta_k|}\frac{M_2-1}{m}\right).$$

- *If $\theta_1 \to \infty$ and θ_k is finite,*

$$\hat{u}_{P_k,1,k} = O_p\left(\frac{1}{\sqrt{\theta_1 m}}\right).$$

- *Therefore, assuming θ_1 and θ_k are finite,*

$$\hat{u}_{P_k,1,k} = O_p\left(\frac{1}{\sqrt{m}}\right).$$

This result holds for any component $\hat{u}_{P_k,s,t}$ where $s \neq t \in \{1,2,...,k\}$.

Remark 5.8.1.1.
The sign of $\hat{u}_{P_k,1,k}$ obtained by construction using Theorem 5.11.1 is always positive. By convention ($\hat{u}_{P_k,i,i} > 0$, for $i = 1,2,...,k$), we multiply by $\text{sign}\left(\hat{u}_{P_k,1,1}\right)$ obtained in the construction. Therefore, the remark of Theorem 5.11.1 describes the sign of the component assuming the convention.

$$P\left\{\text{sign}\left(\hat{u}_{P_k,1,k}\right) = \text{sign}\left(\left(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1}\right)\hat{u}_{P_{k-1},1,k}\hat{u}_{P_{k-1},1,1}\right)\right\} = 1 + O\left(\frac{1}{m}\right).$$

2. *For $s = 1,...,k$, the vector $\frac{\hat{u}_{s,k+1:m}}{\sqrt{1-\hat{\alpha}_s^2}}$, where $\hat{\alpha}_s^2 = \sum_{i=1}^k \hat{u}_{i,s}^2$, is unit invariant by rotation. Moreover, for $j > k$,*

$$\hat{u}_{j,s} \sim \mathbf{N}\left(0, \frac{1-\alpha_s^2}{m}\right),$$

where α_s^2 is the limit of $\hat{\alpha}_s^2$.
Moreover, the columns of $U^t[k+1:m, k+1:m]$ are invariant by rotation.

3. *Assuming $P_k = \mathrm{I}_m + \sum_{i=1}^k (\theta_i - 1)\epsilon_i\epsilon_i^t$ is such that*

$$\theta_1, \theta_2, ..., \theta_{k_1} \text{ are proportional, and}$$
$$\theta_{k_1+1}, \theta_{k_1+2}, ..., \theta_k \text{ are proportional,}$$

then

$$\begin{aligned}\sum \hat{u}^2_{k+1:m,1} &< \sum \hat{u}^2_{k+1:m,1:k_1} \\ &\sim \mathrm{RV}\left(O\left(\frac{1}{\theta_1}\right), O\left(\frac{1}{\theta_1^2 m}\right)\right) + O_p\left(\frac{\min(\theta_1,\theta_k)}{\max(\theta_1,\theta_k)m}\right).\end{aligned}$$

Moreover, if P satisfies Assumption 2.2.2(A4) with $\min\left(\frac{\theta_1}{\theta_k}, \frac{\theta_k}{\theta_1}\right) \to 0$, *then*

$$\sum \hat{u}^2_{k+1:m,1} \sim \mathrm{RV}\left(O\left(\frac{1}{\theta_1}\right), O\left(\frac{1}{\theta_1^2 m}\right)\right) + O_p\left(\frac{1}{\theta_1 m}\right).$$

(Page 81)

Theorem 5.4.1. *(Invariant Eigenvalue Theorem)*
Suppose that W respects Assumption 2.2.1 and

$$\begin{aligned}\tilde{P}_s &= \mathrm{I}_m + (\theta_s - 1)e_s e_s^t, \ \textit{for } s = 1, 2, ..., k, \\ P_k &= \mathrm{I}_m + \sum_{i=1}^{k}(\theta_i - 1)e_i e_i^t \ \textit{satisfies 2.2.2 (A4)},\end{aligned}$$

where $\theta_1 > \theta_2 > ... > \theta_k$. We define

$$\begin{aligned}\hat{\Sigma}_{\tilde{P}_s} &= \tilde{P}_s^{1/2} W \tilde{P}_s^{1/2}, \\ \hat{\Sigma}_{P_k} &= P_k^{1/2} W P_k^{1/2}.\end{aligned}$$

Moreover, for $s = 1, 2, ..., k$, we define

$$\begin{aligned}&\hat{u}_{\tilde{P}_s,1}, \hat{\theta}_{\tilde{P}_s,1} \quad s.t. \quad \hat{\Sigma}_{\tilde{P}_1}\hat{u}_{\tilde{P}_s,1} = \hat{\theta}_{\tilde{P}_s,1}\hat{u}_{\tilde{P}_s,1}, \\ &\hat{u}_{P_k,s}, \hat{\theta}_{P_k,s} \quad s.t. \quad \hat{\Sigma}_{P_k}\hat{u}_{P_k,s} = \hat{\theta}_{P_k,s}\hat{u}_{P_k,s},\end{aligned}$$

where $\hat{\theta}_{\tilde{P}_s,1} = \hat{\lambda}_{\hat{\Sigma}_{\tilde{P}_s,1}}$ and $\hat{\theta}_{P_k,s} = \hat{\lambda}_{\hat{\Sigma}_{P_k},s}$.

1. *Then, for $s > 1$,*

$$\boxed{\hat{\theta}_{P_k,s} - \hat{\theta}_{\tilde{P}_s,1} \sim \frac{\theta_s}{m}}$$

and

$$\boxed{\hat{\theta}_{P_k,1} - \hat{\theta}_{\tilde{P}_1,1} \sim \frac{\theta_2}{m} + \frac{\theta_2}{m^{3/2}},}$$

where θ_1 and θ_2 are the largest and the second largest eigenvalue, respectively. The distribution of $\hat{\theta}_{P_k,s}$ is therefore asymptotically the same as the distribution of $\hat{\theta}_{\tilde{P}_s,1}$ studied in Theorem 5.3.1.

2. *More precisely we define for $r, s \in \{1, 2, ..., k\}$ with $r \neq s$,*

$$P_{-r} = \mathrm{I}_m + \sum_{\substack{i=1\\ i\neq r}}^{k} (\theta_i - 1)\, e_i e_i^t.$$

- *If $\theta_s > \theta_r$, then*

$$\hat{\theta}_{P_k,s} - \hat{\theta}_{P_{-r},s} = -\frac{\hat{\theta}_{P_{-r},s}\hat{\theta}_{P_k,s}(\theta_r - 1)}{\theta_r - 1 - \hat{\theta}_{P_k,s}}\hat{u}^2_{P_{-r},s,r} + O_p\left(\frac{1}{m}\right) + O_p\left(\frac{\theta_r}{m^{3/2}}\right).$$

- *If $\theta_s < \theta_r$, then*

$$\hat{\theta}_{P_k,s} - \hat{\theta}_{P_{-r},s-1} = -\frac{\hat{\theta}_{P_{-r},s-1}\hat{\theta}_{P_k,s}(\theta_r - 1)}{\theta_r - 1 - \hat{\theta}_{P_k,s}}\hat{u}^2_{P_{-r},s-1,r} + O_p\left(\frac{1}{m}\right) + O_p\left(\frac{\theta_s}{m^{3/2}}\right).$$

Remark 5.4.1.1.
In this work, we are interested by the unbiased estimation of $\hat{\hat{\theta}}_{P_k,1}$. The invariance of $\hat{\hat{\theta}}_{P_k,1}$ is a direct consequence of the theorem. Moreover, Theorem 5.3.1 provides the distribution of $\hat{\hat{\theta}}_{P_1,1}$.

(Page 77)

Theorem 5.11.1. *(Characterization of eigenstructure)*
Using the same notation as in the Invariant Theorem (5.5.1, 5.4.1) and under Assumption 2.2.1 and 2.2.2(A4), we can compute the eigenvalues and the components of interest of the eigenvector of $\hat{\Sigma}_{P_k}$. Using assumption 2.2.1, we can without loss of generality suppose the canonical form for the perturbation P_k.

- *Eigenvalues :*

$$\underbrace{\sum_{i=k}^{m} \frac{\hat{\lambda}_{P_{k-1},i}}{\hat{\theta}_{P_k,s} - \hat{\lambda}_{P_{k-1},i}} \hat{u}^2_{P_{k-1},i,k}}_{(a) O_p\left(\frac{1}{\theta_s}\right)} + \underbrace{\frac{\hat{\theta}_{P_{k-1},s}}{\hat{\theta}_{P_k,s} - \hat{\theta}_{P_{k-1},s}} \hat{u}^2_{P_{k-1},s,k}}_{(b) \sim \left(\frac{\theta_k - \theta_s}{\theta_s \theta_k}\right)} + \underbrace{\sum_{\substack{i=1 \\ i \neq s}}^{k-1} \frac{\hat{\theta}_{P_{k-1},i}}{\hat{\theta}_{P_k,s} - \hat{\theta}_{P_{k-1},i}} \hat{u}^2_{P_{k-1},i,k}}_{(c) O_p\left(\frac{1}{\theta_s m}\right)} = \frac{1}{\theta_k - 1},$$

for $s = 1, 2, ..., k$.

Remark 5.11.1.1.

1. The formula is still valid without Assumption 2.2.1. However, the size estimations of each term are not necessarily accurate anmore.
2. If we do not assume canonical perturbations, then the formula is longer but the structure remains essentially the same. Assuming Assumption 2.2.1 leads matrices that are invariant under rotations. Elementary linear algebra methods extend the result to any perturbations.

- *Eigenvectors :*
We define $\tilde{u}_{P_k,i}$ such that $W P_k \tilde{u}_{P_k,i} = \hat{\theta}_{P_k,i} \tilde{u}_{P_k,i}$ and $\hat{u}_{P_k,i}$ such that $P_k^{1/2} W P_k^{1/2} \hat{u}_{P_k,i} = \hat{\theta}_{P_k,i} \hat{u}_{P_k,i}$. To simplify notation we assume that θ_i corresponds to $\hat{\theta}_{P_k,i}$. This notation is explained in 7.1.1 and allows to describe the k first eigenvectors more efficiently.

$$\left\langle \tilde{u}_{P_k,1}, e_1 \right\rangle^2 = \frac{\left(\overbrace{\sum_{i=k}^{m} \frac{\hat{\lambda}_{P_{k-1},i}}{\hat{\theta}_{P_k,1} - \hat{\lambda}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,1} \hat{u}_{P_{k-1},i,k}}^{(a) O_p\left(\frac{1}{\theta_1^{3/2}\sqrt{m}}\right)} + \overbrace{\frac{\hat{\theta}_{P_{k-1},1}}{\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1}} \hat{u}_{P_{k-1},1,1} \hat{u}_{P_{k-1},1,k}}^{(b) \sim \frac{\sqrt{\theta_1 m}}{\min(\theta_1,\theta_k)}} + \overbrace{\sum_{i=2}^{k-1} \frac{\hat{\theta}_{P_{k-1},i}}{\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,1} \hat{u}_{P_{k-1},i,k}}^{(c) O_p\left(\frac{1}{\theta_1^{1/2} m}\right)} \right)^2}{\underbrace{\sum_{i=k}^{m} \frac{\hat{\lambda}^2_{P_{k-1},i}}{(\hat{\theta}_{P_k,1} - \hat{\lambda}_{P_{k-1},i})^2} \hat{u}^2_{P_{k-1},i,k}}_{(d) O_p\left(\frac{1}{\theta_1^2}\right)} + \underbrace{\frac{\hat{\theta}^2_{P_{k-1},1}}{(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1})^2} \hat{u}^2_{P_{k-1},1,k}}_{(e) \sim \frac{\theta_1 m}{\min(\theta_1,\theta_k)^2}} + \underbrace{\sum_{i=2}^{k-1} \frac{\hat{\theta}^2_{P_{k-1},i}}{(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},i})^2} \hat{u}^2_{P_{k-1},i,k}}_{(f) O_p\left(\frac{1}{\theta_1 m}\right)}},$$

$$\left\langle \tilde{u}_{P_k,1}, e_k \right\rangle^2 = \frac{1}{D_1(\theta_k - 1)^2} (g),$$

$$\left\langle \tilde{u}_{P_k,1}, e_s \right\rangle^2$$

$$= \frac{1}{D_1} \left(\overbrace{\sum_{i=k}^{m} \frac{\hat{\lambda}_{P_{k-1},i}}{\hat{\theta}_{P_k,1} - \hat{\lambda}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,s} \hat{u}_{P_{k-1},i,k}}^{(h) O_p\left(\frac{1}{\theta_s^{1/2}\theta_1\sqrt{m}}\right)} + \overbrace{\frac{\hat{\theta}_{P_{k-1},1}}{\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1}} \hat{u}_{P_{k-1},1,s} \hat{u}_{P_{k-1},1,k}}^{(i) \sim \frac{\min(\theta_1,\theta_s)}{\sqrt{\theta_s}\min(\theta_1,\theta_k)}} \right.$$

$$\left. + \overbrace{\sum_{i=2,\neq s}^{k-1} \frac{\hat{\theta}_{P_{k-1},i}}{\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,s} \hat{u}_{P_{k-1},i,k}}^{(j) O_p\left(\max_{i\neq 1,s}\left(\frac{\min(\theta_1,\theta_i)\min(\theta_s,\theta_i)}{\sqrt{\theta_s}\theta_1\theta_i\sqrt{m}}\right)\right)} + \overbrace{\frac{\hat{\theta}_{P_{k-1},s}}{\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},s}} \hat{u}_{P_{k-1},s,s} \hat{u}_{P_{k-1},s,k}}^{(k) O_p\left(\frac{\min(\theta_1,\theta_s)}{\sqrt{\theta_s}\theta_1\sqrt{m}}\right)} \right)^2 .$$

Finally,

$$\hat{u}_{P_k,1} = \frac{\left(\tilde{u}_{P_k,1,1}, \tilde{u}_{P_k,1,2}, ..., \sqrt{\theta_k}\tilde{u}_{P_k,1,k}, ..., \tilde{u}_{P_k,}\right)}{\underbrace{\sqrt{1 + (\theta_k - 1)\,\tilde{u}^2_{P_k 1,k}}}_{1+O_p\left(\frac{\min(\theta_1,\theta_k)}{\max(\theta_1,\theta_k)m}\right)}},$$

where $\sqrt{1 + (\theta - 1)\,\tilde{u}^2_{P_k 1,k}}$ *is the norm of* $P_k^{1/2}\tilde{u}_{P_k,1}$ *that we will call* N_1*.*

Remark 5.11.1.2.

1. By construction, the sign of $\hat{u}_{P_k,1,k}$ is always positive. This is, however, not the case of $\hat{u}_{P_{k-1},i,i}$. We can show that :
$$P\left\{\mathrm{sign}\left(\hat{u}_{P_k,1,1}\right) = \mathrm{sign}\left(\left(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1}\right)\hat{u}_{P_{k-1},1,1}\hat{u}_{P_{k-1},1,k}\right)\right\} \underset{m\to\infty}{\rightarrow} 1.$$
Moreover, the convergence to 1 is of order $1/m$. If θ_1 tends to infinity, then
$$P\left\{\mathrm{sign}\left(\hat{u}_{P_k,1,1}\right) = \mathrm{sign}\left((\theta_1 - \theta_k)\,\hat{u}_{P_{k-1},1,k}\right)\right\} \underset{m,\theta_1\to\infty}{\rightarrow} 1.$$
Therefore, if we use a convention such as $\mathrm{sign}\left(\hat{u}_{P_k,i,i}\right) > 0$ for $i = 1, ..., k-1$, then the sign of $\hat{u}_{P_k,1,k}$ is distributed as a Bernoulli with parameter $1/2$.

2. Without loss of generality, the other eigenvector $\hat{u}_{P_k,r}$ for $r = 1, 2, ..., k-1$ can be computed by the same formula thanks to the notation linking the estimated eigenvector to the eigenvalue θ_i.
However, the formula does not work for the vector $\hat{u}_{P_k,k}$. Applying a different order of perturbation shows that similar formulas exist for $\hat{u}_{P_k,k}$. (If the perturbation in e_1 is applied at the end for example.)
This observation leads to a problem in the proofs of the Dot Product Theorems 5.6.1 and 5.7.1. Deeper investigations are necessary to understand the two eigenvectors when $k = 2$.
$$\begin{aligned}
D_2 &= \underbrace{\sum_{i=2}^{m} \frac{\hat{\lambda}^2_{P_1,i}}{(\hat{\theta}_{P_2,2} - \hat{\lambda}_{P_1,i})^2}\hat{u}^2_{P_1,i,2}}_{O_p\left(\frac{1}{\theta_2^2}\right)} + \underbrace{\frac{\hat{\theta}^2_{P_1,1}}{(\hat{\theta}_{P_2,2} - \hat{\theta}_{P_1,1})^2}\hat{u}^2_{P_1,1,2}}_{O_p\left(\frac{\theta_1}{(\theta_2-\theta_1)^2 m}\right)}, \\
N_2^2 &= 1 + \frac{1}{(\theta_2 - 1)D_2}, \\
N_2 D_2 &= D_2 + \frac{1}{\theta_2 - 1} \\
&= \frac{1}{\theta_2 - 1} + O_p\left(\frac{1}{\theta_2^2}\right) + O_p\left(\frac{\theta_1}{(\theta_2 - \theta_1)m}\right).
\end{aligned}$$

Furthermore, the theorem must investigate the $m-k$ noisy components of the eigenvectors. For $r = 1, 2$ and $s = 3, 4, ..., m$,

$$\hat{u}_{P_2,r,s} = \frac{\sum_{i=1}^{m} \frac{\hat{\lambda}_{P_1,i}}{\hat{\theta}_{P_2,r}-\hat{\lambda}_{P_1,i}} \hat{u}_{P_1,i,s}\hat{u}_{P_1,i,2}}{\sqrt{D_r}N_r}.$$

The estimations using this last formula are difficult. When we investigate these components, it is profitable to look at

$$\hat{u}_{P_2,1,t}/\sqrt{\sum_{s=3}^{m} \hat{u}^2_{P_2,1,s}} \text{ and } \hat{u}_{P_2,2,t}/\sqrt{\sum_{s=3}^{m} \hat{u}^2_{P_2,2,s}}$$

for $t = 3, 4, ..., m$.

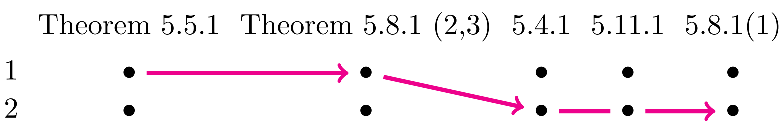
3. If the perturbation is not canonical, then we can apply a rotation U, such that $Uu_s = \epsilon_s$, and replace $\hat{u}_{P_{k-1},i}$ by $U^t\hat{u}_{P_{k-1},i}$. Then, $\langle \tilde{u}_{P_k,1}, e_s\rangle^2$ is replaced by $\langle \tilde{u}_{P_k,1}, u_s\rangle^2$.

(Page 84)

Proof

Pink

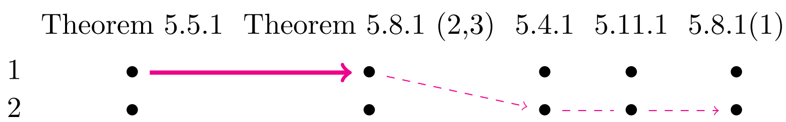
First, we show the initialization pink part.



The Invariant Theorem 5.5.1 is trivially true for perturbations of order $k = 1$. We directly prove Theorem 5.8.1 (2,3) for k=1.

Proof. **Theorem 5.8.1 (2,3), k = 1**

We prove the theorem for $k = 1$. In the following picture we can assume the first result for $k = 1$ is proven.



We define

$$U = \begin{pmatrix} \hat{u}^t_{P_k,1} \\ \hat{u}^t_{P_k,2} \\ \vdots \\ \hat{u}^t_{P_k,m} \end{pmatrix} = \begin{pmatrix} \hat{u}_{P_k,1:k,1:k} & \hat{u}_{P_k,1:k,k+1:m} \\ \hat{u}_{P_k,k+1:m,1:k} & \hat{u}_{P_k,k+1:m,k+1:m}. \end{pmatrix}$$

and

$$O_1 = \begin{pmatrix} \mathrm{I}_1 & 0 \\ 0 & O_{m-1} \end{pmatrix},$$

where O_{m-1} is a rotation matrix.

2. Assuming a canonical $P_1 = \mathrm{I}_m + (\theta_1 - 1)e_1 e_1^t$, we know that $\hat{\Sigma} \sim P_1^{1/2} W P_1^{1/2}$ and $O_1 \hat{\Sigma} O_1^t$ follow the same distribution under Assumption 2.2.1. Although the eigenvectors change, they still follow the same distribution, $O_1 U^t \sim U^t$. Therefore, $\hat{u}_{i,(k+1):m}$ is rotationally invariant and $\mathrm{Corr}\left(\hat{u}_{i,j_1}, \hat{u}_{i,j_2}\right) = \delta_{j_1}(j_2)$.
We can show that knowing the first line of the matrix, then $\hat{u}_{P_1,i,2:m}/||\hat{u}_{P_1,i,2:m}||$ is unit uniform for $i = 1, 2, ..., m$. Therefore, these statistics are independent (not jointly) of the first line. Uniformity of $\hat{u}_{P_1,i,2:m}$ implies for $s = 2, 3, ..., m$,

$$\sqrt{m}\frac{\hat{u}_{P_1,1,s}}{||\hat{u}_{P_1,1,2:m}||} = \sqrt{m}\frac{\hat{u}_{P_1,1,s}}{\sqrt{1-\hat{\alpha}^2_{P_1,1}}} \sim \mathbf{N}\left(0,1\right) + o_p\left(1\right).$$

By Slutsky's Theorem and the distribution of the angle for $k = 1$, Theorem 5.3.1,

$$\hat{u}_{P_1,1,s} \sim \mathbf{N}\left(0, \frac{1-\alpha_1^2}{m}\right) + o_p\left(\frac{1}{\sqrt{m}}\right),$$

where α_1^2 is the limit of the angle and can be approximated by $1 - \frac{M_2-1}{\theta_1} + O_p\left(\frac{1}{\theta_1^2}\right) < 1$.

3. Using the distribution of $\hat{\alpha}^2_{P_1,1,1}$ given in Theorem 5.3.1,

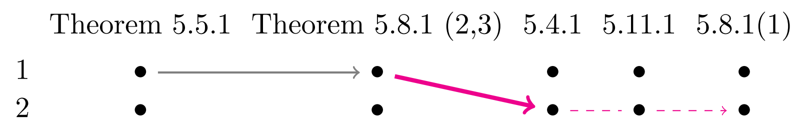
$$\sum \hat{u}^2_{2:m,1} = 1 - \hat{\alpha}^2_{P_1,1,1} \sim \mathrm{RV}\left(O\left(\frac{1}{\theta_1}\right), O\left(\frac{1}{\theta_1^2 m}\right)\right).$$

□

Then, we prove the Invariant Angle Theorem for the eigenvalues, Theorem 5.4.1 for k=2.

Proof. **Theorem 5.4.1, k = 2**

We prove the theorem for $k = 2$. In the following picture we can assume the grey results as proven.



Without loss of generality, we prove the invariance of $\hat{\theta}_{P_1,1}$. For simplicity, we assume $\theta_1 > \theta_2$ but this assumption is only used to simplify notation. Each step can be done assuming $\theta_1 < \theta_2$. Using Theorem 5.2.1 and the canonical perturbation $\tilde{P}_2$ lead to

$$\sum_{i=2}^{m} \frac{\hat{\lambda}_{P_1,i}}{\hat{\theta}_{P_2,1} - \hat{\lambda}_{P_1,i}} \hat{u}^2_{P_1,i,2} + \frac{\hat{\theta}_{P_1,1}}{\hat{\theta}_{P_2,1} - \hat{\theta}_{P_1,1}} \hat{u}^2_{P_1,1,2} = \frac{1}{\theta_2 - 1}.$$

Therefore,

$$\begin{aligned}
\frac{\hat{\theta}_{P_1,1}}{\hat{\theta}_{P_2,1} - \hat{\theta}_{P_1,1}} \hat{u}^2_{P_1,1,2} &= -\sum_{i=2}^{m} \frac{\hat{\lambda}_{P_1,i}}{\hat{\theta}_{P_2,1} - \hat{\lambda}_{P_1,i}} \hat{u}^2_{P_1,i,2} + \frac{1}{\theta_2 - 1} \\
&= -\frac{1}{\hat{\theta}_{P_2,1}} \sum_{i=2}^{m} \hat{\lambda}_{P_1,i} \hat{u}^2_{P_1,i,2} + \frac{1}{\theta_2 - 1} + O_p\left(\frac{1}{\theta_1^2}\right) \\
&\stackrel{1*}{=} -\frac{1}{\hat{\theta}_{P_2,1}}\left(1 + O_p\left(\frac{1}{\sqrt{m}}\right)\right) + \frac{1}{\theta_2 - 1} + O_p\left(\frac{1}{\theta_1^2}\right) \\
&= -\frac{\theta_2 - 1 - \hat{\theta}_{P_2,1}}{\hat{\theta}_{P_2,1}(\theta_2 - 1)} + O_p\left(\frac{1}{\theta_1^2}\right) + O_p\left(\frac{1}{\theta_1\sqrt{m}}\right),
\end{aligned}$$

where 1* is true because

$$\begin{aligned}
\sum_{i=2}^{m} \hat{\lambda}_{P_1,i} \hat{u}_{P_1,i,2}^2 &= \sum_{i=1}^{m} \hat{\lambda}_{P_1,i} \hat{u}_{P_1,1,2}^2 - \hat{\theta}_{P_1,1} \hat{u}_{P_1,1,2}^2 \\
&= \hat{\Sigma}_{P_1,2,2} - \hat{\theta}_{P_1,1} \hat{u}_{P_1,1,2}^2 \\
&= W_{2,2} - \hat{\theta}_{P_1,1} \hat{u}_{P_1,1,2}^2 \\
&= 1 + O_p\left(\frac{1}{\sqrt{m}}\right).
\end{aligned}$$

The last line is obtained using the fact that the canonical perturbation P_1 does not affect $W_{2:m,2:m}$. Moreover, W respects assumption 2.2.1; therefore, $W_{2,2} = 1 + O_p\left(1/\sqrt{m}\right)$. On the other hand the second term $\hat{\theta}_{P_1,1} \hat{u}_{P_1,1,2}^2 = O_p\left(1/m\right)$ by Theorem 5.8.1(2) for $k = 1$.

Thus, by Theorem 5.8.1(2),

$$\begin{aligned}
\left(1 + O_p\left(\frac{\theta_2}{\theta_1(\theta_2 - \theta_1)}\right) + O_p\left(\frac{\theta_2}{\sqrt{m}(\theta_2 - \theta_1)}\right)\right)\left(\hat{\theta}_{P_2,1} - \hat{\theta}_{P_1,1}\right) &= -\frac{\hat{\theta}_{P_1,1}\hat{\theta}_{P_2,1}\left(\theta_2 - 1\right)}{\theta_2 - 1 - \hat{\theta}_{P_2,1}} \hat{u}_{P_1,1,2}^2 \\
&= O_p\left(\frac{\theta_1\theta_2}{m(\theta_2 - \theta_1)}\right).
\end{aligned}$$

We note that even without Assumption 2.2.2(A4),

$$\hat{\theta}_{P_2,1} - \hat{\theta}_{P_1,1} \sim \frac{\min\left(\theta_1, \theta_2\right)}{m}.$$

More precisely we can write

$$\hat{\theta}_{P_2,1} - \hat{\theta}_{P_1,1} = -\frac{\hat{\theta}_{P_1,1}\hat{\theta}_{P_2,1}\left(\theta_2 - 1\right)}{\theta_2 - 1 - \hat{\theta}_{P_2,1}} \hat{u}_{P_1,1,2}^2 + O_p\left(\frac{1}{m}\right) + O_p\left(\frac{\min(\theta_1, \theta_2)}{m^{3/2}}\right).$$

Each step of the computation can be done for $\hat{\theta}_{\tilde{P}_2,1} - \hat{\theta}_{P_2,2}$. Therefore, for $s \neq t \in \{1, 2\}$ we obtain the general result

$$\begin{aligned}
\left(1 + O_p\left(\frac{\theta_t}{\theta_s(\theta_t - \theta_s)}\right) + O_p\left(\frac{\theta_t}{\sqrt{m}(\theta_t - \theta_s)}\right)\right)\left(\hat{\theta}_{P_2,s} - \hat{\theta}_{\tilde{P}_s,1}\right) &= -\frac{\hat{\theta}_{\tilde{P}_s,1}\hat{\theta}_{P_2,s}\left(\theta_t - 1\right)}{\theta_t - 1 - \hat{\theta}_{P_2,s}} \hat{u}_{\tilde{P}_s,1,t}^2 \\
&\sim \frac{\theta_1\theta_2}{m(\theta_2 - \theta_1)}.
\end{aligned}$$

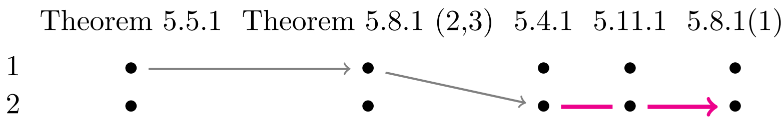
This leads to

$$\begin{aligned}
\hat{\theta}_{P_2,s} - \hat{\theta}_{\tilde{P}_s,1} &= -\frac{\hat{\theta}_{\tilde{P}_s,1}\hat{\theta}_{P_2,s}\left(\theta_t - 1\right)}{\theta_t - 1 - \hat{\theta}_{P_2,s}} \hat{u}_{\tilde{P}_s,1,t}^2 + O_p\left(\frac{1}{m}\right) + O_p\left(\frac{\min(\theta_1, \theta_2)}{m^{3/2}}\right) \\
&\sim \frac{\min\left(\theta_1, \theta_2\right)}{m}.
\end{aligned}$$

□

Proof. **Theorem 5.11.1, and 5.8.1(1), k = 2**

We prove the theorems for $k = 2$. In the following picture we can assume the grey results as proven.

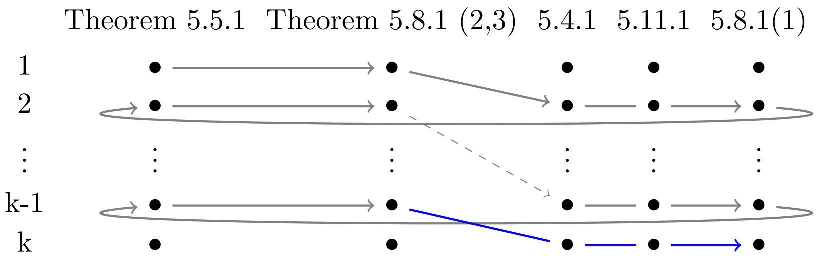


These proofs are exactly the same when the perturbation is of order k. Therefore, we will do it only once in pages 130 and 125. As we will see the proof of these theorems uses only the grey results and the proof of Theorem 5.8.1(1) for k uses Theorem 5.11.1 for k. Moreover, although the proof of Theorem 5.11.1 for $k > 2$ uses Theorem 5.8.1(1) for $k-1$, the initializing part $k = 2$ does not need Theorem 5.8.1(1).

□

Blue

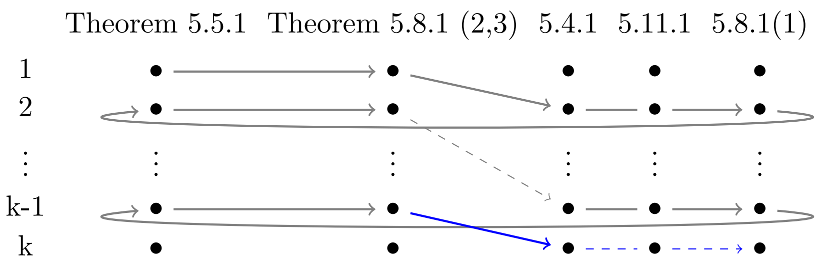
In this section, we assume all the results true for $k-1$. These results appear in grey in the following picture. We want to prove 5.4.1, 5.11.1 and 5.8.1(1) for k.



First, we prove the Invariant Eigenvalue Theorem.

Proof. **Theorem 5.4.1**

To prove this result we can assume the grey results in the following picture as proven.



The proof for k is the same as the proof for $k = 2$ with a small negligible error. We present the proof for $\hat{\theta}_{P_k,s} - \hat{\theta}_{P_{k-1},s}$, where the last added perturbation is of order θ_k and $\theta_1 > \theta_2 > ... > \theta_k$. Similar computations can be done to prove the result when the last added perturbation is of order θ_r, $r \neq s$.

By using Theorem 5.2.1, 5.8.1(1) for $k-1$ and using the fact that the p_i are different in Assumption 2.2.2(A4),

$$\sum_{i=k}^{m} \frac{\hat{\lambda}_{P_{k-1},i}}{\hat{\theta}_{P_k,s} - \hat{\lambda}_{P_{k-1},i}} \hat{u}^2_{P_{k-1},i,k} + \frac{\hat{\theta}_{P_{k-1},s}}{\hat{\theta}_{P_k,s} - \hat{\theta}_{P_{k-1},s}} \hat{u}^2_{P_{k-1},s,k} + \sum_{\substack{i=1\\ i\neq s}}^{k} \frac{\hat{\theta}_{P_{k-1},i}}{\hat{\theta}_{P_k,s} - \hat{\theta}_{P_{k-1},i}} \hat{u}^2_{P_{k-1},i,k} = \frac{1}{\theta_k - 1}$$

$$\Rightarrow \sum_{i=k-1}^{m} \frac{\hat{\lambda}_{P_{k-1},i}}{\hat{\theta}_{P_k,s} - \hat{\lambda}_{P_{k-1},i}} \hat{u}^2_{P_{k-1},i,k} + \frac{\hat{\theta}_{P_{k-1},s}}{\hat{\theta}_{P_k,s} - \hat{\theta}_{P_{k-1},s}} \hat{u}^2_{P_{k-1},s,k} + O_p\left(\frac{1}{m \min\limits_{\substack{i=1,2,3,\dots,k-1\\ i\neq s}} (\theta_s - \theta_i)} \right) = \frac{1}{\theta_k - 1}$$

Therefore,

$$\begin{aligned}
\frac{\hat\theta_{P_{k-1},s}}{\hat\theta_{P_k,s}-\hat\theta_{P_{k-1},s}}\hat u^2_{P_{k-1},s,k} &= -\sum_{i=k}^m \frac{\hat\lambda_{P_{k-1},i}}{\hat\theta_{P_k,s}-\hat\lambda_{P_{k-1},i}}\hat u^2_{P_{k-1},i,k}+\frac{1}{\theta_k-1}+O_p\left(\frac{1}{m\theta_s}\right)\\
&\overset{1*}{=} -\frac{1}{\hat\theta_{P_k,s}}\left(1+O_p\left(\frac{1}{\sqrt{m}}\right)\right)+\frac{1}{\theta_k-1}+O_p\left(\frac{1}{m\theta_s}\right)+O_p\left(\frac{1}{\theta_s^2}\right)\\
&= -\frac{\theta_k-1-\hat\theta_{P_k,s}}{\hat\theta_{P_k,s}(\theta_k-1)}+O_p\left(\frac{1}{m\theta_s}\right)+O_p\left(\frac{1}{\sqrt{m}\theta_s}\right)+O_p\left(\frac{1}{\theta_s^2}\right),
\end{aligned}$$

where 1* is true because

$$\begin{aligned}
\sum_{i=k}^m \hat\lambda_{P_{k-1},i}\hat u^2_{P_{k-1},i,k} &= \sum_{i=1}^m \hat\lambda_{P_{k-1},i}\hat u^2_{P_{k-1},i,k}-\sum_{i=1}^{k-1}\hat\theta_{P_{k-1},i}\hat u^2_{P_{k-1},i,k}\\
&= \hat\Sigma_{P_{k-1},k,k}-\sum_{i=1}^{k-1}\hat\theta_{P_{k-1},i}\hat u^2_{P_{k-1},i,k}\\
&= W_{k,k}-\sum_{i=1}^{k-1}\hat\theta_{P_{k-1},i}\hat u^2_{P_{k-1},i,k}\\
&= 1+O_p\left(\frac{1}{\sqrt{m}}\right).
\end{aligned}$$

The last line is obtained because the canonical perturbation P_{k-1} does not affect $W_{k:m,k:m}$. Moreover, W respects Assumption 2.2.1; therefore, $W_{k,k}=1+O_p\left(1/\sqrt{m}\right)$. On the other hand the second term $\sum_{i=1}^{k-1}\hat\theta_{P_{k-1},i}\hat u^2_{P_{k-1},i,k}=O_p\left(1/m\right)$ by Theorem 5.8.1(2) for $k-1$.

Thus, by Theorem 5.8.1(2) for $k-1$,

$$\begin{aligned}
&\left(1+O_p\left(\frac{\theta_k}{\theta_s(\theta_k-\theta_s)}\right)+O_p\left(\frac{\theta_k}{\sqrt{m}(\theta_k-\theta_s)}\right)+O_p\left(\frac{\theta_s\theta_k}{m(\theta_k-\theta_s)(\theta_{k-1}-\theta_s)}\right)\right)\left(\hat\theta_{P_k,s}-\hat\theta_{P_{k-1},s}\right)\\
&\quad= -\frac{\hat\theta_{P_{k-1},s}\hat\theta_{P_k,s}(\theta_k-1)}{\theta_k-1-\hat\theta_{P_k,s}}\hat u^2_{P_{k-1},s,k}\\
&\quad= O_p\left(\frac{\theta_s\theta_k}{m(\theta_k-\theta_s)}\right).
\end{aligned}$$

and

$$\hat\theta_{P_k,s}-\hat\theta_{P_{k-1},s}\sim\frac{\min\left(\theta_s,\theta_k\right)}{m}.$$

More precisely we can write

$$\hat\theta_{P_k,s}-\hat\theta_{P_{k-1},s} = -\frac{\hat\theta_{P_{k-1},s}\hat\theta_{P_k,s}(\theta_k-1)}{\theta_k-1-\hat\theta_{P_k,s}}\hat u^2_{P_{k-1},s,k}+O_p\left(\frac{1}{m}\right)+O_p\left(\frac{\min(\theta_s,\theta_k)}{m^{3/2}}\right).$$

The min function can be simplified in our case $\theta_k<\theta_s$; however the above notation is more generalisable.
Each step of the computation can be done assuming that the last applied perturbation is θ_r instead of θ_k for $r=1,2,...,k$. Moreover, in this case, similar computations lead to $\hat\theta_{P_k,s}-\hat\theta_{P_{-r},s}$ where $s=1,2,...,k$, $s\neq r$. We define the notation

$$P_{-r}=\mathrm{I}_m+\sum_{\substack{i=1\\ i\neq r}}^k(\theta_i-1)e_ie_i^t.$$

Therefore, for $s\neq r\in\{1,2,...,k\}$ we obtain the general result.

- If $\theta_s > \theta_r$, then

$$\begin{aligned}\hat{\theta}_{P_k,s} - \hat{\theta}_{P_{-r},s} &= -\frac{\hat{\theta}_{P_{-r},s}\hat{\theta}_{P_k,s}(\theta_r - 1)}{\theta_r - 1 - \hat{\theta}_{P_k,s}}\hat{u}^2_{P_{-r},s,r} + O_p\left(\frac{1}{m}\right) + O_p\left(\frac{\theta_s}{m^{3/2}}\right)\\ &\sim \frac{\theta_r}{m}.\end{aligned}$$

- If $\theta_s < \theta_r$, then

$$\begin{aligned}\hat{\theta}_{P_k,s} - \hat{\theta}_{P_{-r},s-1} &= -\frac{\hat{\theta}_{P_{-r},s-1}\hat{\theta}_{P_k,s}(\theta_r - 1)}{\theta_r - 1 - \hat{\theta}_{P_k,s}}\hat{u}^2_{P_{-r},s-1,r} + O_p\left(\frac{1}{m}\right) + O_p\left(\frac{\theta_s}{m^{3/2}}\right)\\ &\sim \frac{\theta_s}{m}.\end{aligned}$$

Finally, we obtain for $s > 1$,

$$\hat{\theta}_{P_k,s} - \hat{\theta}_{\tilde{P}_s,1} \sim \frac{\theta_s}{m}$$

and for $s = 1$,

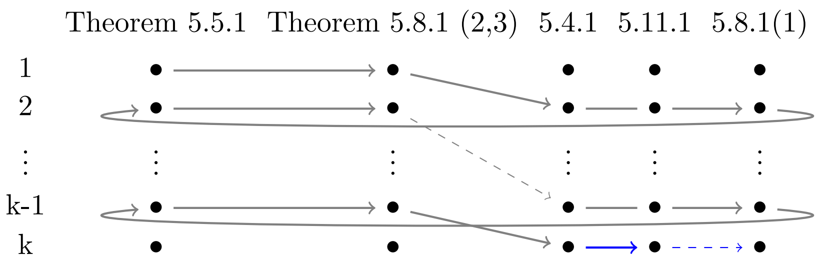
$$\hat{\theta}_{P_k,1} - \hat{\theta}_{\tilde{P}_1,1} \sim \frac{\theta_2}{m} + O_p\left(\frac{\theta_1}{m^{3/2}}\right).$$

□

Now, we prove the characterization of eigenvalues and eigenvectors.

Proof. **Theorem 5.11.1**

To prove this result we can assume the grey results in the following picture as proven.



The initialisation of the induction, $k = 2$, is easily proven using Theorem 5.2.1, 5.4.1 for $k = 2$ and 5.8.1 for $k = 1$. Therefore, we directly prove it for k.
Assuming Assumption 2.2.2(A4) means that we have two groups of eigenvalues composing the perturbation. The first group is finite with bounded eigenvalues and the second group has proportional eigenvalues tending to infinity.
In order to do a general proof we need to discuss the notation.

Notation 7.1.1.

- Usually we assume $\theta_1 > \theta_2 > ... > \theta_k$ such that $\hat{\theta}_{P_k,s}$, the s^{th} largest eigenvalue of $\hat{\Sigma}_{P_k}$ corresponds to θ_s.
  In this proof we relax the order $\theta_1 > \theta 2 > ... > \theta_k$ to do a general proof. The order of θ_s in the eigenvalues $\theta_1, \theta_2, ..., \theta_t$, $t \geqslant s$ is $\text{rank}_t(\theta_s) = r_{t,s}$. Therefore, assuming a perturbation P_t, θ_s corresponds to the $r^{\text{th}}_{t,s}$ largest eigenvalue of $\hat{\Sigma}_{P_t}$. In order to use simple notation, we again call this corresponding estimated eigenvalue, $\hat{\theta}_{P_r,s}$.

Moreover, we change the notation for the eigenvector. In this theorem, for $i = 1, 2, ..., r$, $\hat{u}_{P_r,s}$ is the eigenvector corresponding to $\hat{\theta}_{P_r,s}$.

- We assume two groups of eigenvalues of size k_1 and $k - k_1$ such that these groups respect Assumption 2.2.2(A4). Moreover, θ_1 is supposed to be in the first group. We say that the groups are of order θ_1 and θ respectively, such that only one of them tends to infinity.

Using this new notation we can without loss of generality build the proof for $\hat{u}_{P_k,1}$. However, θ_1 is not the largest eigenvalue anymore.

(a),(h) By Cauchy-Schwarz and using $\rho_1 = \mathrm{E}\left[\hat{\theta}_{P_k,1}\right]$,

$$\left|\sum_{i=k}^{m} \frac{\hat{\lambda}_{P_{k-1},i}}{\hat{\theta}_{P_k,1} - \hat{\lambda}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,1}\hat{u}_{P_{k-1},i,k}\right| \leqslant \left|\sum_{i=k}^{m} \frac{\hat{\lambda}_{P_{k-1},i}}{\rho_1 - \hat{\lambda}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,1}\hat{u}_{P_{k-1},i,k}\right| + O_p\left(\frac{1}{\theta_1^{3/2} m^{1/2}}\right).$$

Some prerequisite results are easily proven using theorems for $k - 1$:

$$\begin{aligned}
&\hat{u}_{P_{k-1},i,1} \text{ s.t. } \sum_{i=k}^{m} \hat{u}^2_{P_{k-1},i,1} = O_p\left(\frac{1}{\theta_1}\right) \text{ (By Theorem 5.8.1 part 3)},\\
&\hat{u}_{P_{k-1},i,k} \sim \mathrm{RV}\left(0, 1/\mathrm{m}\right),\\
&\mathrm{E}\left[\hat{\lambda}_{P_{k-1},i}\hat{u}_{P_{k-1},i,1}\hat{u}_{P_{k-1},i,k}\right] = 0, \text{ (By invariance under rotation)},\\
&\mathrm{Var}\left(\sum_{i=k}^{m} \hat{u}_{P_{k-1},i,1}\hat{u}_{P_{k-1},i,k}\right) = \mathrm{Var}\left(\sum_{i=1}^{k-1} \hat{u}_{P_{k-1},i,1}\hat{u}_{P_{k-1},i,k}\right) = O\left(\frac{1}{\theta_1 m}\right).
\end{aligned}$$

This leads to

$$\begin{aligned}
\sum_{i=k}^{m} \mathrm{Var}\left(\hat{u}_{P_{k-1},i,1}\hat{u}_{P_{k-1},i,k}\right) &= \sum_{i=k}^{m} \mathrm{E}\left[\hat{u}^2_{P_{k-1},i,1}\hat{u}^2_{P_{k-1},i,k}\right]\\
&= \frac{1}{m-k+1}\sum_{i=k}^{m} \mathrm{E}\left[\hat{u}^2_{P_{k-1},i,1}\left(1 - \sum_{s=1}^{k-1}\hat{u}^2_{P_{k-1},i,s}\right)\right]\\
&= O\left(\frac{1}{\theta_1 m}\right).
\end{aligned}$$

In order to obtain the order size, we use the last part of Theorem 5.8.1. Either the perturbation in direction e_1 is finite and the result follows directly, or the perturbation tends to infinity and we can separate the perturbations into two groups, one finite and the other one tending to infinity. The last result of Theorem 5.8.1 leads to the estimation.

$$\begin{aligned}
&\mathrm{Var}\left(\sum_{i=k}^{m} \frac{\hat{\lambda}_{P_{k-1},i}}{\rho_1 - \hat{\lambda}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,1}\hat{u}_{P_{k-1},i,k}\right)\\
&= \sum_{i=k}^{m} \mathrm{Var}\left(\frac{\hat{\lambda}_{P_{k-1},i}}{\rho_1 - \hat{\lambda}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,1}\hat{u}_{P_{k-1},i,k}\right)\\
&\qquad + \sum_{i\neq j=k}^{m} \mathrm{Cov}\left(\frac{\hat{\lambda}_{P_{k-1},i}}{\rho_1 - \hat{\lambda}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,1}\hat{u}_{P_{k-1},i,k}, \frac{\hat{\lambda}_{P_{k-1},j}}{\rho_1 - \hat{\lambda}_{P_{k-1},j}} \hat{u}_{P_{k-1},j,1}\hat{u}_{P_{k-1},j,k}\right)\\
&= A + B.
\end{aligned}$$

The parts A and B are studied separately. By Assumption 2.2.1, $\hat{\lambda}_{P_{k-1},k}$ is bounded by a constant λ.

$$\begin{aligned}
A &= \sum_{i=k}^{m} \mathrm{Var}\left(\frac{\hat{\lambda}_{P_{k-1},i}}{\rho_1 - \hat{\lambda}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,1}\hat{u}_{P_{k-1},i,k}\right) \\
&= \sum_{i=k}^{m} \mathrm{E}\left[\left(\frac{\hat{\lambda}_{P_{k-1},i}}{\rho_1 - \hat{\lambda}_{P_{k-1},i}}\right)^2 \hat{u}^2_{P_{k-1},i,1}\hat{u}^2_{P_{k-1},i,k}\right] \\
&\leqslant \left(\frac{\lambda}{\rho_1-\lambda}\right)^2 \sum_{i=k}^{m} \mathrm{E}\left[\hat{u}^2_{P_{k-1},i,1}\hat{u}^2_{P_{k-1},i,k}\right] \\
&= \left(\frac{\lambda}{\rho_1-\lambda}\right)^2 \sum_{i=k}^{m} \mathrm{Var}\left(\hat{u}_{P_{k-1},i,1}\hat{u}_{P_{k-1},i,k}\right) \\
&= O\left(\frac{1}{\theta_1^3 m}\right).
\end{aligned}$$

$$\begin{aligned}
|B| &= \left|\sum_{i\neq j=k}^{m} \mathrm{Cov}\left(\frac{\hat{\lambda}_{P_{k-1},i}}{\rho_1 - \hat{\lambda}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,1}\hat{u}_{P_{k-1},i,k}, \frac{\hat{\lambda}_{P_{k-1},j}}{\rho_1 - \hat{\lambda}_{P_{k-1},j}} \hat{u}_{P_{k-1},j,1}\hat{u}_{P_{k-1},j,k}\right)\right| \\
&= \left|\sum_{i\neq j=k}^{m} \left(\mathrm{E}\left[\frac{\hat{\lambda}_{P_{k-1},i}}{\rho_1 - \hat{\lambda}_{P_{k-1},i}}\frac{\hat{\lambda}_{P_{k-1},j}}{\rho_1 - \hat{\lambda}_{P_{k-1},j}} \hat{u}_{P_{k-1},i,1}\hat{u}_{P_{k-1},j,1}\hat{u}_{P_{k-1},i,k}\hat{u}_{P_{k-1},j,k}\right] - 0\right)\right| \\
&= \left|\sum_{i\neq j=k}^{m} \frac{1}{m-k+1}\mathrm{E}_p\left[\frac{\hat{\lambda}_{P_{k-1},i}}{\rho_1 - \hat{\lambda}_{P_{k-1},i}}\frac{\hat{\lambda}_{P_{k-1},j}}{\rho_1 - \hat{\lambda}_{P_{k-1},j}} \hat{u}_{P_{k-1},i,1}\hat{u}_{P_{k-1},j,1}\sum_{r=k}^{m}\hat{u}_{P_{k-1},i,r}\hat{u}_{P_{k-1},j,r}\right]\right| \\
&= \left|\sum_{i\neq j=k}^{m} \frac{1}{m-k+1}\mathrm{E}\left[\frac{\hat{\lambda}_{P_{k-1},i}}{\rho_1 - \hat{\lambda}_{P_{k-1},i}}\frac{\hat{\lambda}_{P_{k-1},j}}{\rho_1 - \hat{\lambda}_{P_{k-1},j}} \hat{u}_{P_{k-1},i,1}\hat{u}_{P_{k-1},j,1}\sum_{r=1}^{k-1}\hat{u}_{P_{k-1},i,r}\hat{u}_{P_{k-1},j,r}\right]\right| \\
&\leqslant \frac{\left(\frac{\lambda}{\rho_1-\lambda}\right)^2}{m-k}\sum_{r=1}^{k-1}\mathrm{E}\left[\sum_{i\neq j=k}^{m}\left|\hat{u}_{P_{k-1},i,1}\hat{u}_{P_{k-1},j,1}\hat{u}_{P_{k-1},i,r}\hat{u}_{P_{k-1},j,r}\right|\right] \\
&\leqslant O\left(\frac{1}{\theta_1^2 m}\right)\sum_{r=1}^{k-1}\mathrm{E}_p\left[\left(\sum_{i=k}^{m}\left|\hat{u}_{P_{k-1},i,1}\hat{u}_{P_{k-1},i,r}\right|\right)^2\right] \\
&\leqslant O\left(\frac{1}{\theta_1^2 m}\right)\sum_{r=1}^{k-1}\mathrm{E}_p\left[\left(\sum_{i=k}^{m}\hat{u}^2_{P_{k-1},i,1}\right)\left(\sum_{i=k}^{m}\hat{u}^2_{P_{k-1},i,r}\right)\right] \\
&= O\left(\frac{1}{\theta_1^3 m}\right).
\end{aligned}$$

Thus

$$\mathrm{Var}\left(\sum_{i=k}^{m}\frac{\hat{\lambda}_{P_{k-1},i}}{\rho_1 - \hat{\lambda}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,1}\hat{u}_{P_{k-1},i,k}\right) = O\left(\frac{1}{\theta_1^3 m}\right).$$

Therefore, because the expectation is 0 by invariance under rotation,

$$\sum_{i=k}^{m}\frac{\hat{\lambda}_{P_{k-1},i}}{\hat{\theta}_{P_k,1} - \hat{\lambda}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,1}\hat{u}_{P_{k-1},i,k} = O_p\left(\frac{1}{\theta_1^{3/2}\sqrt{m}}\right).$$

(b) We study

$$\frac{\hat{\theta}_{P_{k-1},1}}{\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1}}\hat{u}_{P_{k-1},1,1}\hat{u}_{P_{k-1},1,k}.$$

By Theorem 5.5.1 and Theorem 5.8.1 for $k-1$, we obtain respectively

$$\text{If } \theta_1 > D, \text{ for } D > 0 \text{ fixed }, \exists\, d(D) \text{ such that, } 1 > |\hat{u}_{P_{k-1},1,1}| > d(D) > 0,$$
$$\hat{u}_{P_{k-1},1,k} \sim \frac{1}{\sqrt{\theta_1 m}}.$$

We see thanks to Theorem 5.4.1 for k that

$$\frac{\hat{\theta}_{P_{k-1},1}}{\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1}} \sim \frac{\theta_1 m}{\min(\theta_1, \theta_k)}.$$

The result is straightforward.

(c) We study

$$\sum_{i=2}^{k-1} \frac{\hat{\theta}_{P_{k-1},i}}{\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,1} \hat{u}_{P_{k-1},i,k}.$$

By Theorem 5.8.1,

$$\begin{aligned} \hat{u}_{P_{k-1},i,1} &= O_p\left(\frac{\sqrt{\theta_1 \theta_i}}{(\theta_1 - \theta_i)\sqrt{m}}\right), \\ \hat{u}_{P_{k-1},i,k} &= O_p\left(\frac{1}{\sqrt{\theta_i}\sqrt{m}}\right). \end{aligned}$$

Therefore,

$$\frac{\hat{\theta}_{P_{k-1},i}}{\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,1} \hat{u}_{P_{k-1},i,k} = O_p\left(\frac{\sqrt{\theta_1}\theta_i}{(\theta_1 - \theta_i)^2 \sqrt{m}}\right).$$

Studying the different possibilities for θ_i and θ_1 leads to the result.

(d) We study

$$\sum_{i=k}^{m} \frac{\hat{\lambda}^2_{P_{k-1},i}}{(\hat{\theta}_{P_k,1} - \hat{\lambda}_{P_{k-1},i})^2} \hat{u}^2_{P_{k-1},i,k}.$$

A straightforward computation leads to

$$\sum_{i=k}^{m} \frac{\hat{\lambda}^2_{P_{k-1},i}}{(\hat{\theta}_{P_k,1} - \hat{\lambda}_{P_{k-1},i})^2} \hat{u}^2_{P_{k-1},i,k} \leqslant O_p\left(\frac{1}{\theta_1^2}\right)\left(1 - \sum_{i=1}^{k-1} \hat{u}^2_{P_{k-1},i,k}\right) = O_p\left(\frac{1}{\theta_1^2}\right).$$

(e) We study

$$\frac{\hat{\theta}^2_{P_{k-1},1}}{(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1})^2} \hat{u}^2_{P_{k-1},1,k}.$$

By Theorems 5.4.1 and 5.8.1,

$$\frac{\hat{\theta}^2_{P_{k-1},1}}{(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1})^2} \sim \frac{\theta_1^2 m^2}{\min(\theta_1, \theta_k)^2},$$
$$\hat{u}^2_{P_{k-1},1,k} \sim \frac{1}{m\theta_1}.$$

The result is straightforward.

(f) We study

$$\sum_{i=2}^{k-1} \frac{\hat{\theta}^2_{P_{k-1},i}}{(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},i})^2} \hat{u}^2_{P_{k-1},i,k}.$$

By Theorem 5.8.1 and 5.4.1,

$$\hat{u}^2_{P_{k-1},i,k} = O_p\left(\frac{1}{m\theta_i}\right).$$

Then,

$$\frac{\hat{\theta}^2_{P_{k-1},i}}{(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},i})^2} \hat{u}^2_{P_{k-1},i,k} = O_p\left(\frac{\theta_i}{(\theta_1 - \theta_i)^2 m}\right).$$

Studying the different possibilities for θ_i and θ_1 leads to the result.

(g) The result is obtained directly from Theorem 5.2.1.

(h) The same proof of (a) leads to the result.

(i) We study

$$\frac{\hat{\theta}_{P_{k-1},1}}{\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1}} \hat{u}_{P_{k-1},1,s} \hat{u}_{P_{k-1},1,k},$$

for $s = 2, ..., k-1$.

By Theorems 5.4.1 and 5.8.1 ,

$$\hat{u}_{P_{k-1},1,s} \hat{u}_{P_{k-1},1,k} \sim \frac{\min(\theta_1, \theta_s)}{\theta_1 \sqrt{\theta_s} m},$$
$$\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1} \sim \frac{\min(\theta_1, \theta_k)}{m}.$$

The result follows directly.

(j) We study

$$\sum_{i=2,\neq s}^{k-1} \frac{\hat{\theta}_{P_{k-1},i}}{\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,s} \hat{u}_{P_{k-1},i,k}.$$

By the Theorems 5.4.1 and 5.8.1 the result is straightforward as for part c.

(k) We study

$$\frac{\hat{\theta}_{P_{k-1},s}}{\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},s}} \hat{u}_{P_{k-1},s,s} \hat{u}_{P_{k-1},s,k}.$$

By Theorem 5.5.1 and Lemma 5.8.1,

$$\text{If } \theta_s > D, \text{ for } D > 0 \text{ fixed }, \exists\, d(D) \text{ such that, } 1 > |\hat{u}_{P_{k-1},s,s}| > d(D) > 0,$$
$$\hat{u}_{P_{k-1},s,k} \sim \frac{1}{\sqrt{\theta_s m}}.$$

The result is straightforward.

The link between $\hat{u}_{P_k,1,1}$ and $\tilde{u}_{P_k,1,1}$ is just obtained by basic notions of linear algebra and similar estimations of the norm.
We now prove the first point of the remark.

1. First we study $\operatorname{sign}(\hat{u}_{P_k,1,1})$ by investigating $\tilde{u}_{P_k,1,1}$ which was defined in the statement of the theorem. Then, by construction, the results hold for $\hat{u}_{P_k,1,1}$ because we just rescale $\tilde{u}_{P_k,1}$ to obtain $\hat{u}_{P_k,1}$. The theorem says

$$\begin{aligned}&\operatorname{sign}\left(\tilde{u}_{P_k,1,1}\right)\\ &= \operatorname{sign}\left(\frac{\overbrace{\sum_{i=k}^{m}\frac{\hat{\lambda}_{P_{k-1},i}}{\hat{\theta}_{P_k,1}-\hat{\lambda}_{P_{k-1},i}}\hat{u}_{P_{k-1},i,1}\hat{u}_{P_{k-1},i,k}}^{(a)O_p\left(\frac{1}{\theta_1^{3/2}\sqrt{m}}\right)} + \overbrace{\frac{\hat{\theta}_{P_{k-1},1}}{\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},1}}\hat{u}_{P_{k-1},1,1}\hat{u}_{P_{k-1},1,k}}^{(b)\sim\ \frac{\sqrt{\theta_1 m}}{\min(\theta_1,\theta_k)}} + \overbrace{\sum_{i=2}^{k-1}\frac{\hat{\theta}_{P_{k-1},i}}{\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},i}}\hat{u}_{P_{k-1},i,1}\hat{u}_{P_{k-1},i,k}}^{(c)O_p\left(\frac{1}{\theta_1^{1/2}m}\right)}}{\sqrt{\underbrace{\sum_{i=k}^{m}\frac{\hat{\lambda}^2_{P_{k-1},i}}{(\hat{\theta}_{P_k,1}-\hat{\lambda}_{P_{k-1},i})^2}\hat{u}^2_{P_{k-1},i,k}}_{(d)O_p\left(\frac{1}{\theta_1^2}\right)} + \underbrace{\frac{\hat{\theta}^2_{P_{k-1},1}}{(\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},1})^2}\hat{u}^2_{P_{k-1},1,k}}_{(e)\sim\ \frac{\theta_1 m}{\min(\theta_1,\theta_k)^2}} + \underbrace{\sum_{i=2}^{k-1}\frac{\hat{\theta}^2_{P_{k-1},i}}{(\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},i})^2}\hat{u}^2_{P_{k-1},i,k}}_{(f)O_p\left(\frac{1}{\theta_1 m}\right)}}}\right).\end{aligned}$$

The first convergence is directly obtained from

$$\operatorname{sign}\left(\tilde{u}_{P_k,1,1}\right) = \operatorname{sign}\left(\frac{\hat{\theta}_{P_{k-1},1}}{\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},1}}\hat{u}_{P_{k-1},1,1}\hat{u}_{P_{k-1},1,k} + O_p\left(\frac{1}{\theta_1^{1/2}m}\right) + O_p\left(\frac{1}{\theta_1^{3/2}\sqrt{m}}\right)\right).$$

Using Theorem 5.4.1 and assuming m and θ_1 sufficiently large lead to

$$\begin{aligned}\operatorname{sign}\left(\tilde{u}_{P_k,1,1}\right) &= \operatorname{sign}\left(\left(\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},1}\right)\hat{u}_{P_{k-1},1,k}\right)\\ &= \operatorname{sign}\left(\left(\theta_1-\theta_k\right)\hat{u}_{P_{k-1},1,k}\right).\end{aligned}$$

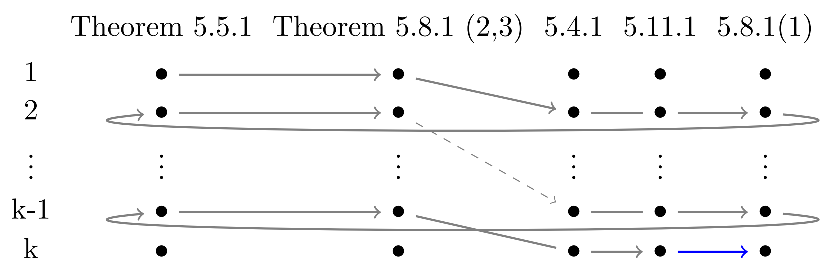
2. The second remark supposes a perturbation of order $k = 2$. Then, we already know how to investigate the behaviour of the first eigenvector. In order to obtain the second vector, we need to replace $\hat{\theta}_{P_k,1}$ by $\hat{\theta}_{P_k,2}$ in the formula and the order size changes. Similar arguments as above lead to the result.

□

For the last part of the proof of the blue part in the Figure 7.1, we study the first point of the component Theorem 5.8.1.

Proof. **Theorem 5.8.1**

To prove this result we can assume the grey results in the following picture as proven.



This proof computes $\hat{u}_{P_k,1,k}$, but the method can be used to study any components $\hat{u}_{P_k,s,t}$ where $s \neq t \in \{1, 2, ..., k\}$. In order to extend it we should use the notation defined in Notation 7.1.1.

First we assume the convention of Theorem 5.11.1, $\hat{u}_{P_k,1,k} > 0$.

$$
\begin{aligned}
&\langle \tilde{u}_{P_k,1}, e_k \rangle \\
&= \frac{\sum_{i=k}^{m} \frac{\hat{\lambda}_{P_{k-1},i}}{\hat{\theta}_{P_k,1}-\hat{\lambda}_{P_{k-1},i}} \hat{u}^2_{P_{k-1},i,k} + \frac{\hat{\theta}_{P_{k-1},1}}{\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},1}} \hat{u}^2_{P_{k-1},1,k} + \sum_{i=2}^{k-1} \frac{\hat{\theta}_{P_{k-1},i}}{\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},i}} \hat{u}^2_{P_{k-1},i,k}}{\sqrt{\underbrace{\sum_{i=k}^{m} \frac{\hat{\lambda}^2_{P_{k-1},i}}{(\hat{\theta}_{P_k,1}-\hat{\lambda}_{P_{k-1},i})^2} \hat{u}^2_{P_{k-1},i,k}}_{O_p\left(\frac{1}{\theta_1^2}\right)} + \underbrace{\frac{\hat{\theta}^2_{P_{k-1},1}}{(\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},1})^2} \hat{u}^2_{P_{k-1},1,k}}_{\sim\frac{\theta_1 m}{\min(\theta_1,\theta_k)^2}} + \underbrace{\sum_{i=2}^{k-1} \frac{\hat{\theta}^2_{P_{k-1},i}}{(\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},i})^2} \hat{u}^2_{P_{k-1},i,k}}_{O_p\left(\frac{1}{\theta_1 m}\right)}}} \\
&= \frac{1}{\theta_k - 1} \frac{1}{\sqrt{\frac{\hat{\theta}^2_{P_{k-1},1}}{(\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},1})^2} \hat{u}^2_{P_{k-1},1,k}}} + O_p\left(\frac{\min(\theta_1,\theta_k)^3}{\theta_k \theta_1^{7/2} m^{3/2}}\right) + O_p\left(\frac{\min(\theta_1,\theta_k)^3}{\theta_k \theta_1^{5/2} m^{5/2}}\right) \\
&= \frac{1}{\theta_k - 1} \frac{|\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1}|}{|\hat{\theta}_{P_{k-1},1}||\hat{u}_{P_{k-1},1,k}|} + O_p\left(\frac{\min(\theta_1,\theta_k)^3}{\theta_k \theta_1^{7/2} m^{3/2}}\right) + O_p\left(\frac{\min(\theta_1,\theta_k)^3}{\theta_k \theta_1^{5/2} m^{5/2}}\right).
\end{aligned}
$$

Then, $\hat{u}_{P_k,1} = P_k^{1/2} \tilde{u}_{P_k,1}/N_1$, and

$$
\begin{aligned}
N_1^2 &= \sum_{i=1}^{k-1} \tilde{u}^2_{P_k,i} + \theta_k \tilde{u}^2_{P_k,k} + \sum_{i=k+1}^{m} \tilde{u}^2_{P_k,i} \\
&= 1 + (\theta - 1)\tilde{u}^2_{P_k,k}.
\end{aligned}
$$

We also know by Theorem 5.4.1 that

$$
\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1} = -\frac{\hat{\theta}_{P_{k-1},1}\hat{\theta}_{P_k,1}(\theta_k - 1)}{\theta_k - 1 - \hat{\theta}_{P_k,1}} \hat{u}^2_{P_{k-1},1,k} + O_p\left(\frac{1}{m}\right) + O_p\left(\frac{\theta_1}{m^{3/2}}\right).
$$

and

$$
\hat{\theta}_{P_{k-1},1} - \hat{\theta}_{P_k,1} = O_p\left(\frac{\min(\theta_1,\theta_k)}{m}\right).
$$

Therefore, Theorem 5.11.1 and 5.4.1 for k leads to

$$
\begin{aligned}
\hat{u}_{P_k,1,k} &= \frac{\langle \tilde{u}_{P_k,1}, e_k \rangle}{\text{Norm}} \sqrt{\theta_k} \\
&= \left(\underbrace{\frac{1}{\theta_k - 1} \frac{|\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1}|}{|\hat{\theta}_{P_{k-1},1}||\hat{u}_{P_{k-1},1,k}|}}_{\sim\frac{\min(\theta_1,\theta_k)}{\theta_k \theta_1^{1/2} m^{1/2}}} + O_p\left(\frac{\min(\theta_1,\theta_k)^3}{\theta_k \theta_1^{7/2} m^{3/2}}\right) + O_p\left(\frac{\min(\theta_1,\theta_k)^3}{\theta_k \theta_1^{5/2} m^{5/2}}\right) \right) \frac{\sqrt{\theta_k}}{1 + O_p\left(\frac{1}{m}\right)} \\
&= \frac{1}{\sqrt{\theta_k}} \frac{|\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1}|}{|\hat{\theta}_{P_{k-1},1}||\hat{u}_{P_{k-1},1,k}|} + O_p\left(\frac{\min(\theta_1,\theta_k)}{\theta_k^{1/2}\theta_1^{1/2} m^{3/2}}\right) \\
&= \frac{\left| -\frac{\hat{\theta}_{P_{k-1},1}\hat{\theta}_{P_k,1}(\theta_k-1)}{\theta_k-1-\hat{\theta}_{P_k,1}} \hat{u}^2_{P_{k-1},1,k} + O_p\left(\frac{1}{m}\right) + O_p\left(\frac{\min(\theta_1,\theta_k)}{m^{3/2}}\right) \right|}{\sqrt{\theta_k}|\hat{\theta}_{P_{k-1},1}||\hat{u}_{P_{k-1},1,k}|} + O_p\left(\frac{\min(\theta_1,\theta_k)}{\theta_k^{1/2}\theta_1^{1/2} m^{3/2}}\right) \\
&= \frac{\sqrt{\theta_k}\hat{\theta}_{P_k,1}}{|\theta_k - \hat{\theta}_{P_k,1}|} |\hat{u}_{P_{k-1},1,k}| + O_p\left(\frac{\min(\theta_1,\theta_k)}{\theta_1^{1/2}\theta_k^{1/2} m}\right) + O_p\left(\frac{1}{\theta_1^{1/2}\theta_k^{1/2} m^{1/2}}\right).
\end{aligned}
$$

Note that the sign is always positive! We can use the Remark of Theorem 5.11.1 and set $\hat{u}_{P_s,i,i} > 0$ for $s = 1, 2, ..., k$ and $i = 1, 2, ..., s$. Then, the previous result becomes more convenient:

Under the sign condition for the eigenvector,

$$\hat{u}_{P_k,1,k} = \frac{\sqrt{\theta_k}\theta_1}{\theta_1-\theta_k}\hat{u}_{P_{k-1},1,k} + O_p\left(\frac{\min(\theta_1,\theta_k)}{\theta_1^{1/2}\theta_k^{1/2}m}\right) + O_p\left(\frac{1}{\theta_1^{1/2}\theta_k^{1/2}m^{1/2}}\right).$$

Therefore, we directly obtain the distribution when $\theta_1, \theta_k \to \infty$. Using the notation

$$\begin{aligned}\hat{\alpha}^2_{P_{k-1},1} &= \sum_{i=1}^{k-1}\left\langle \hat{u}_{P_{k-1},1}, \epsilon_i\right\rangle^2, \\ \hat{\alpha}^2_{P_{k-1},1} &= \alpha_1^2 + O_p\left(\frac{1}{\theta_1 m}\right) = 1 - \frac{M_2-1}{\theta_1} + O_p\left(\frac{1}{\theta_1^2}\right) + O_p\left(\frac{1}{\theta_1\sqrt{m}}\right)\end{aligned}$$

and the second part of this Theorem 5.8.1 for $k-1$,

$$\hat{u}_{P_{k-1},1,k}|\hat{\alpha}^2_{P_{k-1},1} \overset{Asy}{\sim} \mathbf{N}\left(0, \frac{1-\hat{\alpha}^2_{P_{k-1},1}}{m}\right),$$

then

$$\hat{u}_{P_k,1,k}|\hat{\alpha}^2_{P_k,1} \sim \mathbf{N}\left(0, \frac{\theta_k\theta_1^2}{(\theta_k-\theta_1)^2}\frac{\hat{\alpha}^2_{P_{k-1},1}-1}{m}\right) + O_p\left(\frac{\min(\theta_1,\theta_k)}{\theta_1^{1/2}\theta_k^{1/2}m}\right) + O_p\left(\frac{1}{\theta_1^{1/2}\theta_k^{1/2}m^{1/2}}\right)$$

and

$$\hat{u}_{P_k,1,k} \sim \mathbf{N}\left(0, \frac{\theta_k\theta_1}{(\theta_k-\theta_1)^2}\frac{M_2-1}{m}\right) + O_p\left(\frac{\min(\theta_1,\theta_k)}{\theta_1^{1/2}\theta_k^{1/2}m}\right) + O_p\left(\frac{1}{\theta_1^{1/2}\theta_k^{1/2}m^{1/2}}\right).$$

Finally, we extend this result to small eigenvalues,

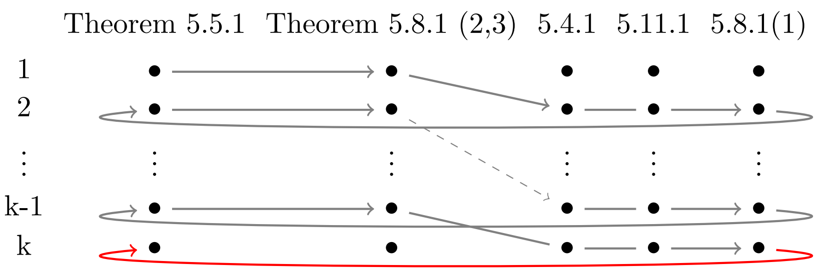
$$\text{If } \theta_1 \to \infty \text{ and } \theta_k \text{ is finite, then } \hat{u}_{P_k,1,k} = O_p\left(\frac{1}{\sqrt{\theta_1 m}}\right),$$

$$\text{If } \theta_1 \text{ and } \theta_k \text{ are finite, then } \hat{u}_{P_k,1,k} = O_p\left(\frac{1}{\sqrt{m}}\right).$$

□

Red

Using induction we show the part of the Invariant Theorem 5.5.1 shown in red in the picture. We assume the grey theorems as true.



Proof. **Theorem 5.5.1**

We assume the induction hypotheses and prove the result for k. The idea of the proof is to use Theorem 5.11.1 to simplify the k first entries of the eigenvector $\tilde{u}_{P_k,1}$. Then, we show that

$$\tilde{F}_{P_k}^2 = \sum_{i=k+1}^{m} \tilde{u}_{P_k,1,i}^2 = \sum_{i=k}^{m} \hat{u}_{P_{k-1},1,i}^2 + O_p\left(\frac{1}{m\theta_1}\right) = \hat{F}_{P_{k-1}}^2 + O_p\left(\frac{1}{m\theta_1}\right).$$

Finally, we easily prove

$$\hat{F}_{P_k}^2 = \hat{F}_{P_{k-1}}^2 + O_p\left(\frac{1}{m\theta_1}\right).$$

Remark 7.1.1.
The following proof studies $\sum_{i=k+1}^{m} \tilde{u}_{P_k,1,i}^2$ with $\theta_1 > \theta_2 > ... > \theta_k$. However, the proof is easily extended to $\sum_{i=k+1}^{m} \tilde{u}_{P_k,s,i}^2$ for $s = 1, 2, ..., k$ and $\theta_s > \theta_k$. Finally, the proof is also valid for $\theta_s > \theta_k$ with more difficult notation as in 7.1.1. In order to simplify the two expansions for the reader, we will not reduce values such as $\min(\theta_1, \theta_i)$.

A: First, we study

$$\tilde{\Sigma}_{P_k} = \hat{\Sigma}_{P_{k-1}} P_k$$

using Theorem 5.11.1 and 5.2.1. The eigenvectors of $\tilde{\Sigma}_{P_k}$ are

$$\tilde{u}_{P_k,i} = \frac{(\hat{\theta}_{P_k,i}\mathrm{I}_m - \hat{\Sigma}_{P_{k-1}})^{-1}\hat{\Sigma}_{P_{k-1}}\epsilon_k}{\sqrt{e_k^t\hat{\Sigma}_{P_{k-1}}(\hat{\theta}_{P_k,i}\mathrm{I} - \hat{\Sigma}_{\mathrm{P_{k-1}}})^{-2}\hat{\Sigma}_{\mathrm{P_{k-1}}}\epsilon_{\mathrm{k}}}}.$$

Then,

$$\begin{aligned}
&\left\langle \tilde{u}_{P_k,1}, e_s \right\rangle^2 \\
&= \frac{\left(\sum_{i=k}^{m} \frac{\hat{\lambda}_{P_{k-1},i}}{\hat{\theta}_{P_k,1}-\hat{\lambda}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,s}\hat{u}_{P_{k-1},i,k} + \frac{\hat{\theta}_{P_{k-1},1}}{\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},1}} \hat{u}_{P_{k-1},1,s}\hat{u}_{P_{k-1},1,k} + \sum_{i=2}^{k-1} \frac{\hat{\theta}_{P_{k-1},i}}{\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,s}\hat{u}_{P_{k-1},i,k}\right)^2}{\sum_{i=k}^{m} \frac{\hat{\lambda}_{P_{k-1},i}^2}{(\hat{\theta}_{P_k,1}-\hat{\lambda}_{P_{k-1},i})^2} \hat{u}_{P_{k-1},i,k}^2 + \frac{\hat{\theta}_{P_{k-1},1}^2}{(\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},1})^2} \hat{u}_{P_{k-1},1,k}^2 + \sum_{i=2}^{k-1} \frac{\hat{\theta}_{P_{k-1},i}^2}{(\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},i})^2} \hat{u}_{P_{k-1},i,k}^2} \\
&= \frac{(A_{1,s,k:m} + A_{1,s,1} + A_{1,s,2:k-1})^2}{D_{1,k:m} + D_{1,1} + D_{1,2:k-1}} \\
&= \frac{A_{1,s}^2}{D_1}.
\end{aligned}$$

The size of each element of the equation can be estimated by Theorem 5.11.1.

B: We investigate the norm of the noisy part of the eigenvector. We set

$$\tilde{F}_{P_k}^2 = \sum_{i=k+1}^{m} \tilde{u}_{P_k,1,i}^2 = 1 - \sum_{i=1}^{k} \tilde{u}_{P_k,1,i}^2 = 1 - \frac{\sum_{i=s}^{k} A_{1,s}^2}{D_1}.$$

We want to show that $\tilde{F}_{P_k}^2 \approx \hat{F}_{P_{k-1}}^2$ using Theorem 5.11.1.
First, we make an approximation of $A_{1,s}$, $A_{1,s}^2$ and D_1:

$$\begin{aligned}
A_{1,1} = &\overbrace{\frac{\hat{\theta}_{P_{k-1},1}}{\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1}} \hat{u}_{P_{k-1},1,1}\hat{u}_{P_{k-1},1,k}}^{O_p\left(\frac{\theta_1^{1/2} m^{1/2}}{\min(\theta_1,\theta_k)}\right)} + \overbrace{\sum_{i=2}^{k-1} \frac{\hat{\theta}_{P_{k-1},i}}{\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,1}\hat{u}_{P_{k-1},i,k}}^{O_p\left(\frac{1}{\theta_1^{1/2} m}\right)} \\
&+ O_p\left(\frac{1}{\theta_1^{3/2} m^{1/2}}\right),
\end{aligned}$$

$$A_{1,s} = \overbrace{\frac{\hat{\theta}_{P_{k-1},1}}{\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1}} \hat{u}_{P_{k-1},1,s} \hat{u}_{P_{k-1},1,k}}^{O_p\left(\frac{\min(\theta_1,\theta_s)}{\theta_s^{1/2}\min(\theta_1,\theta_k)}\right)} + \overbrace{\sum_{i=2,\neq s}^{k-1} \frac{\hat{\theta}_{P_{k-1},i}}{\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,s} \hat{u}_{P_{k-1},i,k}}^{O_p\left(\max_{i\neq 1,s,k}\left(\frac{\min(\theta_1,\theta_i)\min(\theta_s,\theta_i)}{\theta_s^{1/2}\theta_1\theta_i m^{1/2}}\right)\right)}$$
$$+ \overbrace{\frac{\hat{\theta}_{P_{k-1},s}}{\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},s}} \hat{u}_{P_{k-1},s,s} \hat{u}_{P_{k-1},s,k}}^{O_p\left(\frac{\min(\theta_1,\theta_s)}{\theta_s^{1/2}\theta_1 m^{1/2}}\right)} + O_p\left(\frac{1}{\theta_s^{1/2}\theta_1 m^{1/2}}\right),$$

$$A_{1,k} = \frac{1}{\theta_k - 1},$$

$$D_1 = \overbrace{\frac{\hat{\theta}^2_{P_{k-1},1}}{(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1})^2} \hat{u}^2_{P_{k-1},1,k}}^{O_p\left(\frac{\theta_1 m}{\min(\theta_1,\theta_k)^2}\right)} + \overbrace{\sum_{i=2}^{k-1} \frac{\hat{\theta}^2_{P_{k-1},i}}{(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},i})^2} \hat{u}^2_{P_{k-1},i,k}}^{O_p\left(\frac{1}{\theta_1 m}\right)} + O_p\left(\frac{1}{\theta_1^2}\right)$$
$$= O_p\left(\frac{\theta_1 m}{\min(\theta_1,\theta_k)^2}\right),$$

$$A^2_{1,k} = \frac{1}{(\theta_k - 1)^2},$$

$$A^2_{1,s} = \overbrace{\sum_{i=1}^{k-1} \frac{\hat{\theta}^2_{P_{k-1},i}}{(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},i})^2} \hat{u}^2_{P_{k-1},i,s} \hat{u}^2_{P_{k-1},i,k}}^{A_{1,s,1}}$$
$$+ \underbrace{2\sum_{i=1}^{k-1}\sum_{j>i}^{k-1} \frac{\hat{\theta}_{P_{k-1},i}\hat{\theta}_{P_{k-1},j}}{(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},i})(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},j})} \hat{u}_{P_{k-1},i,s} \hat{u}_{P_{k-1},i,k} \hat{u}_{P_{k-1},j,s} \hat{u}_{P_{k-1},j,k}}_{A_{1,s,2}}$$
$$+ O_p\left(\frac{1}{\theta_1 \min(\theta_1,\theta_k)}\right).$$

More investigations allow the estimation of $\sum_{s=2}^{k-1} A^2_{1,s}$:

$$\sum_{s=1}^{k-1} A^2_{1,s} = \sum_{s=1}^{k-1} A_{1,s,1} + \sum_{s=2}^{k-1} A_{1,s,2} + O_p\left(\frac{1}{\theta_1 \min(\theta_1,\theta_k)}\right),$$

$$\begin{aligned}
\sum_{s=1}^{k-1} A_{1,s,1} &= \sum_{s=1}^{k-1}\sum_{i=1}^{k-1} \frac{\hat{\theta}^2_{P_{k-1},i}}{(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},i})^2} \hat{u}^2_{P_{k-1},i,s} \hat{u}^2_{P_{k-1},i,k} \\
&= \sum_{i=1}^{k-1} \left(\sum_{s=1}^{k-1} \hat{u}^2_{P_{k-1},i,s}\right) \frac{\hat{\theta}^2_{P_{k-1},i}}{(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},i})^2} \hat{u}^2_{P_{k-1},i,k} \\
&= \left(\sum_{s=1}^{k-1} \hat{u}^2_{P_{k-1},1,s}\right) D_1 + \underbrace{\sum_{i=1}^{k-1} \underbrace{\left(\sum_{s=1}^{k-1} \hat{u}^2_{P_{k-1},i,s} - \sum_{s=1}^{k-1} \hat{u}^2_{P_{k-1},1,s}\right)}_{O_p\left(\frac{1}{\min(\theta_1,\theta_i)}\right) \text{ by induction}} \underbrace{\frac{\hat{\theta}^2_{P_{k-1},i}}{(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},i})^2} \hat{u}^2_{P_{k-1},i,k}}_{O_p\left(\frac{\min(\theta_1,\theta_i)}{\theta_1(\theta_1-\theta_i)m}\right)}}_{O_p\left(\max_{i=2,\dots,k-1} \frac{1}{\theta_1(\theta_1-\theta_i)m}\right)},
\end{aligned}$$

$$
\begin{aligned}
&\sum_{s=1}^{k-1} A_{1,s,2} \\
&= 2\sum_{s=1}^{k-1}\sum_{i=1}^{k-1}\sum_{j>i}^{k-1} \frac{\hat{\theta}_{P_{k-1},i}\hat{\theta}_{P_{k-1},j}}{(\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},i})(\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},j})}\hat{u}_{P_{k-1},i,s}\hat{u}_{P_{k-1},i,k}\hat{u}_{P_{k-1},j,s}\hat{u}_{P_{k-1},j,k} \\
&= 2\sum_{i=1}^{k-1}\sum_{j>i}^{k-1} \frac{\hat{\theta}_{P_{k-1},i}\hat{\theta}_{P_{k-1},j}}{(\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},i})(\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},j})}\hat{u}_{P_{k-1},i,k}\hat{u}_{P_{k-1},j,k}\left(\sum_{s=1}^{k-1}\hat{u}_{P_{k-1},i,s}\hat{u}_{P_{k-1},j,s}\right) \\
&= 2\sum_{i=1}^{k-1}\sum_{j>i}^{k-1} \underbrace{\frac{\hat{\theta}_{P_{k-1},i}\hat{\theta}_{P_{k-1},j}}{(\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},i})(\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},j})}\hat{u}_{P_{k-1},i,k}\hat{u}_{P_{k-1},j,k}}_{\text{If } i=1,\ O_p\left(\frac{\min(\theta_1,\theta_j)}{\min(\theta_1,\theta_k)\theta_1^{1/2}\theta_j^{1/2}}\right)\text{ and if } i>1,\ O_p\left(\frac{\min(\theta_i,\theta_1)\min(\theta_j,\theta_1)}{\theta_1^2\theta_i^{1/2}\theta_j^{1/2}m}\right)} \underbrace{\left(-\sum_{s=k}^{m}\hat{u}_{P_{k-1},i,s}\hat{u}_{P_{k-1},j,s}\right)}_{O_p\left(\frac{1}{\theta_i^{1/2}\theta_j^{1/2}}\right)} \\
&= O_p\left(\max_{j=2,\ldots,k-1}\frac{1}{\min(\theta_1,\theta_k)\max(\theta_1,\theta_j)}\right).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\tilde{F}^2_{P_k} &= 1-\frac{1}{D_1(\theta_k-1)^2}-\frac{\sum_{i=s}^{k-1}A^2_{1,s}}{D_1} \\
&= 1-O_p\left(\frac{1}{\theta_1 m}\right)-\sum_{s=1}^{k-1}\hat{u}^2_{P_{k-1},1,s}+O_p\left(\frac{\theta_1}{\max(\theta_1,\theta_k)^2 m}\right) \\
&= 1-\sum_{s=1}^{k-1}\hat{u}^2_{P_{k-1},1,s}+O_p\left(\frac{1}{\theta_1 m}\right) \\
&= \hat{F}_{P_{k-1}}+O_p\left(\frac{1}{\theta_1 m}\right).
\end{aligned}
$$

C: The result is proven for the eigenvector of

$$\tilde{\Sigma}_{P_k}=\hat{\Sigma}_{P_{k-1}}P_k.$$

Now, we need to study

$$\hat{\Sigma}_{P_k}=P_k^{1/2}\hat{\Sigma}_{P_{k-1}}P_k^{1/2}.$$

The link between the eigenvectors is

$$
\begin{aligned}
\hat{u}_{P_{k-1},1} &= \frac{P_k^{1/2}\tilde{u}_{P_k,1}}{\sqrt{\text{Norm}^2}} \\
\text{Norm}^2 &= \sum_{i=1}^{k-1}\tilde{u}^2_{P_k,1,i}+\theta_k\tilde{u}^2_{P_k,1,k}+\sum_{i=k+1}^{m}\tilde{u}^2_{P_k,1,i} \\
&= 1+\underbrace{(\theta_k-1)\overbrace{\tilde{u}^2_{P_k,1,k}}^{O_p\left(\frac{\theta_1}{\max(\theta_1,\theta_k)^2 m}\right)}}_{O_p\left(\frac{1}{m}\right)}
\end{aligned}
$$

Using the induction hypothesis, the result is true for $k-1$; therefore, by Theorem 5.3.1,

$$\hat{F}_{P_{k-1}}=\text{RV}\left(O_p\left(\frac{1}{\theta_1}\right),O_p\left(\frac{1}{\theta_1^2 m}\right)\right).$$

Then,

$$
\begin{aligned}
\hat{F}^2_{P_k} &= \sum_{i=k+1}^{m} \hat{u}^2_{P_{k-1},1,i} \\
&= \frac{1}{\text{Norm}^2} \sum_{i=k+1}^{m} \tilde{u}^2_{P_{k-1},1,i} \\
&= \frac{1}{1+O_p\left(\frac{1}{m}\right)} \tilde{F}^2_{P_k} \\
&= \frac{1}{1+O_p\left(\frac{1}{m}\right)} \left(\hat{F}_{P_{k-1}} + O_p\left(\frac{1}{\theta_1 m}\right)\right) \\
&= \hat{F}_{P_{k-1}} + O_p\left(\frac{1}{\theta_1 m}\right).
\end{aligned}
$$

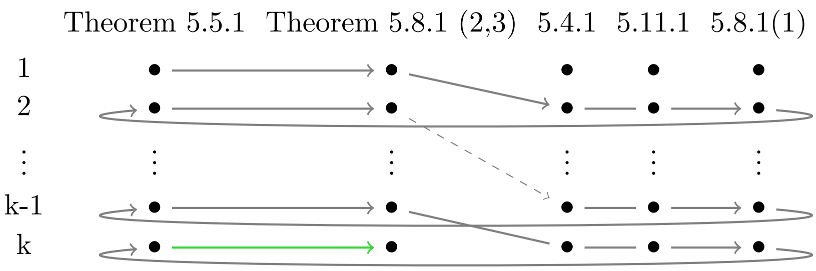
This last equation concludes the proof by induction

$$\sum_{i=1}^{k} \hat{u}^2_{P_k,1,i} = \hat{u}^2_{P_1,1,1} + O_p\left(\frac{1}{\theta_1 m}\right).$$

$\square$

Green

In this section we want to prove the green part in the following picture. In order to prove Theorem 5.8.1 (2, 3) for k, we only assume as true the grey results in the picture.



Proof. **Theorem 5.8.1 (2,3)**

To prove this theorem for k we use the same procedure as for $k = 1$.
Let

$$U = \begin{pmatrix} \hat{u}^t_{P_k,1} \\ \hat{u}^t_{P_k,2} \\ \vdots \\ \hat{u}^t_{P_k,m} \end{pmatrix} = \begin{pmatrix} \hat{u}_{P_k,1:k,1:k} & \hat{u}_{P_k,1:k,k+1:m} \\ \hat{u}_{P_k,k+1:m,1:k} & \hat{u}_{P_k,k+1:m,k+1:m}. \end{pmatrix}$$

and

$$O_k = \begin{pmatrix} \mathrm{I_k} & 0 \\ 0 & O_{m-k} \end{pmatrix},$$

where O_{m-k} is Haar invariant.

2. When P_k is canonical, we know that $\hat{\Sigma} \sim P_k^{1/2} W P_k^{1/2}$ and $O_k \hat{\Sigma} O_k^t$ follow the same distribution under Assumption 2.2.1. Therefore, $\hat{u}_{i,k+1:m}$ is rotationally invariant and $\mathrm{Corr}\left(\hat{u}_{i,j_1}, \hat{u}_{i,j_2}\right) = \delta_{j_1}(j_2)$. Knowing $\hat{u}_{P_k,1:m,1:k}$, we can show that $\hat{u}_{P_k,i,k+1:m}/||\hat{u}_{P_k,i,k+1:m}||$ is uniform for $i = 1, 2, ..., m$. Therefore, these statistics are independent (not jointly) of $\hat{u}_{P_k,1:m,1:k}$. Uniformity of $\hat{u}_{P_k,r,k+1:m}$ implies that, for $s = k+1, ..., m$ and $r = 1, 2, ..., k$,

$$\sqrt{m}\frac{\hat{u}_{P_k,r,s}}{||\hat{u}_{P_k,r,(k+1):m}||} = \sqrt{m}\frac{\hat{u}_{P_k,r,s}}{\sqrt{1-\hat{\alpha}^2_{P_k,r}}} \sim \mathbf{N}\left(0,1\right) + o_p\left(1\right),$$

where

$$\hat{\alpha}^2_{P_k,r} = \sum_{i=1}^{k} \left\langle \hat{u}_{P_k,r}, \epsilon_i \right\rangle^2.$$

By Slutsky's Theorem and the Invariant Angle Theorem 5.5.1 for k,

$$\hat{u}_{P_k,r,s} \sim \mathbf{N}\left(0, \frac{1-\alpha_r^2}{m}\right) + o_p\left(\frac{1}{\sqrt{m}}\right),$$

where $\alpha_r^2 = \lim_{m\to\infty} \hat{\alpha}^2_{P_k,r} = 1 - \frac{M_2-1}{\theta_r} + O_p\left(\frac{1}{\theta^2}\right) < 1$.

3. Then, we estimate the order of $\sum \hat{u}^2_{k+1:m,1}$.
Without loss of generality we assume that the perturbation

$$P_k = \mathrm{I}_m + \sum_{i=1}^{k} (\theta_i - 1)\epsilon_i \epsilon_i^t$$

respecting Assumption 2.2.2(A4) is such that

$$\begin{aligned} &\theta_1, \theta_2, ..., \theta_{k_1} \text{ are proportional,}\\ &\theta_{k_1+1}, \theta_{k_1+2}, ..., \theta_k \text{ are proportional.}\end{aligned}$$

Then by Theorem 5.5.1 and 5.8.1 Part 1 for perturbations of order k,

$$\begin{aligned}
\sum \hat{u}^2_{k+1:m,1:k_1} &= \sum \hat{u}^2_{k_1+1:m,1:k_1} - \sum \hat{u}^2_{k_1+1:k,1:k_1}\\
&= \sum \hat{u}^2_{1:k_1,k_1+1:m} + O_p\left(\frac{\min(\theta_1,\theta_k)}{\max(\theta_1,\theta_k)m}\right)\\
&= \sum \hat{u}^2_{1:k_1,k+1:m} + O_p\left(\frac{\min(\theta_1,\theta_k)}{\max(\theta_1,\theta_k)m}\right)\\
&\sim \mathrm{RV}\left(O\left(\frac{1}{\theta_1}\right), O\left(\frac{1}{\theta_1^2 m}\right)\right) + O_p\left(\frac{\min(\theta_1,\theta_k)}{\max(\theta_1,\theta_k)m}\right).
\end{aligned}$$

The result is straightforward.

□

7.1.5 Dot Product Theorem and its Invariant

In this section we prove the results concerning the partial dot product between two estimated eigenvectors. First, we show a useful small Lemma. Then, we investigate its distribution when $k = 2$. Finally, we prove the invariance to increasing k.

Prerequisite

Lemma 6.1.1.
Assuming W and $\hat{\Sigma}_{P_1}$ as in Theorem 5.6.1, then by construction of the eigenvectors using Theorem 5.11.1,

$$\begin{aligned}
\hat{u}_{P_1,1,2} &= \frac{W_{1,2}}{\sqrt{\theta_1}W_{1,1}} - \frac{W_{1,2}}{\theta_1^{3/2}}\left(-1/2+3/2M_2\right) + \frac{\left(W^2\right)_{1,2}}{\theta_1^{3/2}} + O_p\left(\frac{1}{\theta_1^{3/2}m}\right) + O_p\left(\frac{1}{\theta_1^{5/2}m^{1/2}}\right) \\
&= \frac{W_{1,2}}{\sqrt{\theta_1}} + O_p\left(\frac{1}{\theta_1^{1/2}m}\right) + O_p\left(\frac{1}{\theta_1^{3/2}m^{1/2}}\right), \\
\textstyle\sum_{i=2}^m \hat{\lambda}_{P_1,i}^2\hat{u}_{P_1,i,2}^2 &= W_{2,2} + O_p\left(\frac{1}{m}\right), \\
\textstyle\sum_{i=2}^m \hat{\lambda}_{P_1,i}\hat{u}_{P_1,i,1}\hat{u}_{P_1,i,2} &= W_{1,2}\frac{M_2}{\sqrt{\theta_1}} - \left(W^2\right)_{1,2}\frac{1}{\sqrt{\theta_1}} + O_p\left(\frac{1}{\theta_1^{1/2}m}\right) + O_p\left(\frac{1}{\theta_1^{3/2}m^{1/2}}\right).
\end{aligned}$$

Remark 6.1.1.1.
Because the perturbation is of order 1, the two sign conventions defined in 5.1.1 are the same.

(Page 91)

Proof. **Lemma 6.1.1**

The proofs of the three results use Theorem 5.11.1.
First, we recall that

$$\begin{aligned}
\sum_{i=2}^m \hat{\lambda}_{P_1,i}^2\hat{u}_{P_1,i,2}^2 &= \hat{\Sigma}_{P_1,2,2}^2 - \hat{\theta}_{P_1,1}^2\hat{u}_{P_1,1,2}^2, \\
\sum_{i=2}^m \hat{\lambda}_{P_1,i}\hat{u}_{P_1,i,1}\hat{u}_{P_1,i,2} &= \hat{\Sigma}_{P_1,1,2} - \hat{\theta}_{P_1,1}\hat{u}_{P_1,1,1}\hat{u}_{P_1,1,2}.
\end{aligned}$$

Moreover, if $\tilde{P}_1 = \left(\sqrt{\theta_1}-1\right)e_1e_1^t$, then

$$\begin{aligned}
\hat{\Sigma}_{P_1} &= W + W\tilde{P}_1 + \tilde{P}_1W + \tilde{P}_1W\tilde{P}_1, \\
\hat{\Sigma}_{P_1,1,2} &= W_{1,2}\sqrt{\theta_1}, \\
\left(\hat{\Sigma}_{P_1}^2\right)_{2,2} &= \left(W + W\tilde{P}_1 + \tilde{P}_1W + \tilde{P}_1W\tilde{P}_1\right)^2[2,2] \\
&= \left(W^2\right)_{2,2} + (\theta-1)\left(W_{1,2}\right)^2,
\end{aligned}$$

where $A[2,2]$ is the entry $A_{2,2}$ of the matrix A.
In order to prove the formulas, we need some prerequisite estimations of

$$\hat{u}_{P_1,1,1}^2,\ \hat{\theta}_{P_1,1}^2 \text{ and } \frac{\hat{u}_{P_1,1,2}}{\sqrt{1-\hat{u}_{P_1,1,1}^2}}.$$

A more precise estimation of $\hat{u}^2_{P_1,1,1}$ leads to

$$\begin{aligned}
\hat{u}^2_{P_1,1,1} &= 1 - \frac{\left(W^2\right)_{1,1} - \left(W_{1,1}\right)^2}{\theta_1 \left(W_{1,1}\right)^2} + \frac{1 + \frac{3\left(\left(W^2\right)_{1,1}\right)^2}{\left(W_{1,1}\right)^4} - \frac{2\left(W^2\right)_{1,1}}{\left(W_{1,1}\right)^2} - \frac{2\left(W^3\right)_{1,1}}{\left(W_{1,1}\right)^3}}{\theta_1^2} + O_p\left(\frac{1}{\theta_1^3}\right), \\
\hat{u}_{P_1,1,1} &= 1 - \frac{\left(W^2\right)_{1,1} - \left(W_{1,1}\right)^2}{2\theta_1 \left(W_{1,1}\right)^2} + \frac{1 + \frac{3\left(\left(W^2\right)_{1,1}\right)^2}{\left(W_{1,1}\right)^4} - \frac{2\left(W^2\right)_{1,1}}{\left(W_{1,1}\right)^2} - \frac{2\left(W^3\right)_{1,1}}{\left(W_{1,1}\right)^3}}{2\theta_1^2} + O_p\left(\frac{1}{\theta_1^3}\right), \\
\sqrt{1 - \hat{u}^2_{P_1,1,1}} &= \frac{1}{\sqrt{\theta_1}} \left(\frac{\sqrt{\left(W^2\right)_{1,1} - \left(W_{1,1}\right)^2}}{W_{1,1}} - \frac{W_{1,1}\left(1 + \frac{3\left(\left(W^2\right)_{1,1}\right)^2}{\left(W_{1,1}\right)^4} - \frac{2\left(W^2\right)_{1,1}}{\left(W_{1,1}\right)^2} - \frac{2\left(W^3\right)_{1,1}}{\left(W_{1,1}\right)^3}\right)}{2\theta\sqrt{\left(W^2\right)_{1,1} - \left(W_{1,1}\right)^2}} \right) \\
&\quad + O_p\left(\frac{1}{\theta_1^{5/2}}\right).
\end{aligned}$$

Then, we estimate $\hat{\theta}^2_{P_1,1}$,

$$\begin{aligned}
\frac{1}{\theta_1 - 1} &= \sum_{i=1}^{m} \frac{\hat{\lambda}_{W,i}}{\hat{\theta}_{P_1,1} - \hat{\lambda}_{W,i}} \hat{u}^2_{W,i,1} \\
&= \frac{W_{1,1}}{\hat{\theta}_{P_1,1}} + \frac{\left(W^2\right)_{1,1}}{\hat{\theta}^2_{P_1,1}} + O_p\left(\frac{1}{\theta_1^3}\right) \\
\Rightarrow \hat{\theta}_{P_1,1} &= \theta_1 W_{1,1} + \frac{\left(W^2\right)_{1,1} - \left(W_{1,1}\right)^2}{W_{1,1}} + O_p\left(\frac{1}{\theta_1}\right) \\
\Rightarrow \hat{\theta}^2_{P_1,1} &= \theta_1^2 W_{1,1}^2 + 2\theta_1 \left(\left(W^2\right)_{1,1} - \left(W_{1,1}\right)^2\right) + O_p(1).
\end{aligned}$$

Finally, we estimate the rescaled component,

$$\begin{aligned}
\frac{\hat{u}_{P_1,1,2}}{\sqrt{1 - \hat{u}^2_{P_1,1,1}}} &= \frac{\sum_{i=1}^{m} \frac{\hat{\lambda}_{W,i}}{\hat{\theta}_{P_1,1} - \hat{\lambda}_{W,i}} \hat{u}_{W,i,1} \hat{u}_{W,i,2}}{\sqrt{\sum_{s=2}^{m} \left(\sum_{i=1}^{m} \frac{\hat{\lambda}_{W,i}}{\hat{\theta}_{P_1,1} - \hat{\lambda}_{W,i}} \hat{u}_{W,i,1} \hat{u}_{W,i,s}\right)^2}} \\
&= \frac{\frac{1}{\hat{\theta}_{P_1,1}} W_{1,2} + \frac{1}{\hat{\theta}^2_{P_1,1}} \left(W^2\right)_{1,2} + O_p\left(\frac{1}{\theta^3\sqrt{m}}\right)}{\sqrt{\sum_{s=2}^{m} \left(\frac{1}{\hat{\theta}_{P_1,1}} W_{1,s} + \frac{1}{\hat{\theta}^2_{P_1,1}} \left(W^2\right)_{1,s} + O_p\left(\frac{1}{\theta_1^3\sqrt{m}}\right)\right)^2}} \\
&= \frac{W_{1,2} + \frac{1}{\hat{\theta}_{P_1,1}} \left(W^2\right)_{1,2} + O_p\left(\frac{1}{\theta_1^2\sqrt{m}}\right)}{\sqrt{\sum_{s=2}^{m} \left(\left(W_{1,s}\right)^2 + 2\frac{1}{\hat{\theta}_{P_1,1}} W_{1,s} \left(W^2\right)_{1,s} + O_p\left(\frac{1}{\theta_1^2 m}\right)\right)}} \\
&= \frac{W_{1,2} + \frac{1}{\hat{\theta}_{P_1,1}} \left(W^2\right)_{1,2} + O_p\left(\frac{1}{\theta_1^2\sqrt{m}}\right)}{\sqrt{\left(W^2\right)_{1,1} - \left(W_{1,1}\right)^2 + \frac{1}{\hat{\theta}_{P_1,1}} \left[\left(W^3\right)_{1,1} - W_{1,1}\left(W^2\right)_{1,1}\right] + O_p\left(\frac{1}{\theta_1^2 m}\right)}} \\
&= W_{1,2} \left(\frac{1}{\sqrt{\left(W^2\right)_{1,1} - \left(W_{1,1}\right)^2}} - \frac{\left(W^3\right)_{1,1} - W_{1,1}\left(W^2\right)_{1,1}}{\left(\left(W^2\right)_{1,1} - \left(W_{1,1}\right)^2\right)^{3/2} \hat{\theta}_{P_1,1}} \right) \\
&\quad + \frac{\left(W^2\right)_{1,2}}{\sqrt{\left(W^2\right)_{1,1} - \left(W_{1,1}\right)^2} \hat{\theta}_{P_1,1}} + O_p\left(\frac{1}{\theta_1^2\sqrt{m}}\right).
\end{aligned}$$

Using this estimation, the three formulas are easily proven.

We start with the first formula:

$$\begin{aligned}
\hat{u}_{P_1,1,2} &= \frac{\hat{u}_{P_1,1,2}}{\sqrt{1-\hat{u}^2_{P_1,1,1}}}\sqrt{1-\hat{u}^2_{P_1,1,1}} \\
&= \frac{W_{1,2}}{\sqrt{\theta_1}W_{1,1}} - \frac{W_{1,2}}{\theta_1^{3/2}}\left(-1/2+3/2M_2\right) + \frac{\left(W^2\right)_{1,2}}{\theta_1^{3/2}} + O_p\left(\frac{1}{\theta_1^{3/2}m}\right) + O_p\left(\frac{1}{\theta_1^{5/2}m^{1/2}}\right).
\end{aligned}$$

Then, the second formula:

$$\begin{aligned}
\sum_{i=2}^{m}\hat{\lambda}^2_{P_1,i}\hat{u}^2_{P_1,i,2} - \left(W^2\right)_{2,2} &= \left(\theta_1-1\right)\left(W_{1,2}\right)^2 - \hat{\theta}^2_{P_1,1}\hat{u}^2_{P_1,1,2} \\
&= \left(\theta_1-1\right)\left(W_{1,2}\right)^2 - \left(\theta_1^2\left(W_{1,1}\right)^2 + 2\theta_1\left(\left(W^2\right)_{1,1} - \left(W_{1,1}\right)^2\right) + O_p\left(1\right)\right) \\
&\qquad \left(\frac{\left(W_{1,2}\right)^2}{\theta_1\left(W_{1,1}\right)^2} + O_p\left(\frac{1}{\theta^2 m}\right)\right) \\
&= -\left(W_{1,2}\right)^2 - 2\frac{\left(W_{1,2}\right)^2\left(\left(W^2\right)_{1,1} - \left(W_{1,1}\right)^2\right)}{\left(W_{1,1}\right)^2} + O_p\left(\frac{1}{m}\right) \\
&= O_p\left(\frac{1}{m}\right).
\end{aligned}$$

Finally, some computations lead to the last formula,

$$\begin{aligned}
&\hat{\theta}_{P_1,1}\hat{u}_{P_1,1,1}\hat{u}_{P_1,1,2} \\
&\quad = \left(\theta_1 W_{1,1} + \frac{\left(W^2\right)_{1,1} - \left(W_{1,1}\right)^2}{W_{1,1}}\right)\left(1 - \frac{\left(W^2\right)_{1,1} - \left(W_{1,1}\right)^2}{2\theta_1\left(W_{1,1}\right)^2}\right) \\
&\qquad \left(\frac{W_{1,2}}{\sqrt{\theta_1}W_{1,1}} - \frac{W_{1,2}}{\theta_1^{3/2}}\left(-1/2+3/2M_2\right) + \frac{\left(W^2\right)_{1,2}}{\theta_1^{3/2}}\right) + O_p\left(\frac{1}{\theta_1^{1/2}m}\right) + O_p\left(\frac{1}{\theta_1^{3/2}m^{1/2}}\right) \\
&\quad = W_{1,2}\left(\sqrt{\theta_1} - \frac{M_2}{\sqrt{\theta_1}}\right) + \left(W^2\right)_{1,2}\frac{1}{\sqrt{\theta_1}} + O_p\left(\frac{1}{\theta_1^{1/2}m}\right) + O_p\left(\frac{1}{\theta_1^{3/2}m^{1/2}}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sum_{i=2}^{m}\hat{\lambda}_{P_1,i}\hat{u}_{P_1,i,1}\hat{u}_{P_1,i,2} &= \sqrt{\theta_1}W_{1,2} - \hat{\theta}_{P_1,1}\hat{u}_{P_1,1,1}\hat{u}_{P_1,1,2} \\
&= W_{1,2}\frac{M_2}{\sqrt{\theta_1}} - \left(W^2\right)_{1,2}\frac{1}{\sqrt{\theta_1}} + O_p\left(\frac{1}{\theta_1^{1/2}m}\right) + O_p\left(\frac{1}{\theta_1^{3/2}m^{1/2}}\right).
\end{aligned}$$

$\square$

Distribution

Theorem 5.6.1. *(Dot Product Theorem)*
Suppose that W respects Assumption 2.2.1 and $P_2 = \mathrm{I}_m + \sum_{i=1}^{2}(\theta_i - 1)e_i e_i^t$ with $\theta_1 > \theta_2$. We define

$$\hat{\Sigma}_{P_2} = P_2^{1/2}WP_2^{1/2} \textit{ and } \hat{\Sigma}_{P_1} = P_1^{1/2}WP_1^{1/2}.$$

Moreover, for $s,k=1,2$ and $s\leqslant k$, we define

$$\hat{u}_{P_k,s},\hat{\theta}_{P_k,s} \quad s.t. \quad \hat{\Sigma}_{P_k}\hat{u}_{P_k,s}=\hat{\theta}_{P_k,s}\hat{u}_{P_k,s},$$

where $\hat{\theta}_{P_k,s}=\hat{\lambda}_{\hat{\Sigma}_{P_k},s}$. Finally the present theorem uses the convention:

$$\textit{For } s=1,2,...,k \textit{ and } i=1,2,...,s,\ \hat{u}_{P_s,i,i}>0.$$

1. *Assuming that Assumptions 2.2.2(A2) and (A3) ($\theta_i=p_i\theta\to\infty$) hold,*

$$\begin{aligned}\sum_{s=3}^{m}\hat{u}_{P_2,1,s}\hat{u}_{P_2,2,s} &= \hat{u}_{P_2,1,2}\left(\frac{1}{\theta_1}-\frac{1}{\theta_2}\right)-\frac{1}{\theta_2^{1/2}}\sum_{j>1}^{m}\hat{\lambda}_{P_1,j}\hat{u}_{P_1,j,1}\hat{u}_{P_1,j,2}\\ &\quad +O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{1/2}m}\right)+O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{3/2}m^{1/2}}\right)\\ &= \frac{-\left(1+M_2\right)W_{1,2}+\left(W^2\right)_{1,2}}{\sqrt{\theta_1\theta_2}}+O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{1/2}m}\right)+O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{3/2}m^{1/2}}\right).\end{aligned}$$

Thus, we can estimate the distribution conditioned on the spectrum of W,

$$\begin{aligned}\sum_{s=3}^{m}\hat{u}_{P_2,1,s}\hat{u}_{P_k,2,s} &\sim \mathbf{N}\left(0,\frac{\left(1+M_2\right)^2\left(M_2-1\right)+\left(M_4-(M_2)^2\right)-2\left(1+M_2\right)\left(M_3-M_2\right)}{\theta_1\theta_2 m}\right)\\ &\quad +O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{1/2}m}\right)+O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{3/2}m^{1/2}}\right).\end{aligned}$$

2. *If $\theta_1\to\infty$ and θ_2 is finite, then*

$$\sum_{s=3}^{m}\hat{u}_{P_2,1,s}\hat{u}_{P_2,2,s} = O_p\left(\frac{1}{\sqrt{\theta_1 m}}\right).$$

3. *If θ_1 and θ_2 are finite, then*

$$\sum_{s=3}^{m}\hat{u}_{P_k,1,s}\hat{u}_{P_k,2,s}=O_p\left(\frac{1}{\sqrt{m}}\right).$$

Remark 5.6.1.1.

1. We can easily show

$$\begin{aligned}&\hat{u}_{P_2,1,2}\left(\frac{1}{\theta_1}-\frac{1}{\theta_2}\right)\delta+\sum_{s=3}^{m}\hat{u}_{P_2,1,s}\hat{u}_{P_2,2,s}\\ &=\frac{-\left(\delta+M_2\right)W_{1,2}+\left(W^2\right)_{1,2}}{\sqrt{\theta_1\theta_2}}+O_p\left(\frac{1}{\theta m}\right)+O_p\left(\frac{1}{\theta^2 m^{1/2}}\right)\\ &\sim\mathbf{N}\left(0,\frac{\left(\delta+M_2\right)^2\left(M_2-1\right)+\left(M_4-(M_2)^2\right)-2\left(\delta+M_2\right)\left(M_3-M_2\right)}{\theta_1\theta_2 m}\right)\\ &\quad +O_p\left(\frac{1}{\theta m}\right)+O_p\left(\frac{1}{\theta^2 m^{1/2}}\right).\end{aligned}$$

2. If W is a standard Wishart random matrix, then Assumptions 2.2.2(A2) and (A3) lead

to a Marcenko-Pastur spectrum and

$$\sum_{s=3}^{m} \hat{u}_{P_2,1,s}\hat{u}_{P_2,2,s} \sim \mathbf{N}\left(0, \frac{(1-\alpha_1^2)(1-\alpha_2^2)}{m}\right) + o_p\left(\frac{1}{\theta\sqrt{m}}\right),$$

where $\alpha_s^2 = \lim_{m\to\infty} \sum_{i=1}^{2} \langle \hat{u}_{P_2,s}, u_i \rangle^2$.

(Page 79)

Proof. **Theorem 5.6.1**

We begin this proof with a remark about the sign convention. This Theorem assumes $\hat{u}_{P_s,i,i} > 0$ for $s = 1, 2, ..., k$ and $i = 1, 2, ..., s$. The Theorem 5.11.1 builds the eigenvectors of random matrices with another sign convention,

$$\hat{u}_{P_s,i,s} > 0, \text{ for } s = 1, 2, ..., k \text{ and } i = 1, 2, ..., s.$$

We will use the same notation for both conventions and only precise the convention. The parts **A** and **B** use the convention of Theorem 5.11.1. The convention changes in the end of part **B**. Finally part **C** uses the convention of this theorem.
In order to prove the theorem, we divide the proof into three parts. The first part, **A**, tries to express the components of an eigenvector using Theorem 5.11.1. The second part, **B**, expresses the dot product of $\hat{\Sigma}_{P_2}$ with the eigenstructure of $\hat{\Sigma}_{P_1}$. Finally, with the previous part leading to a nice formula, we investigate in **C** the distribution of this statistic.
We will often replace $\hat{\theta}_{P_1,1}$ by $\hat{\lambda}_{P_1,1}$ to simplify computations.

A: For $t = 1, 2$, we study the expression:

$$\tilde{u}_{P_2,t,s} = \frac{\sum_{i=1}^{m} \frac{\hat{\lambda}_{P_1,i}}{\hat{\theta}_{P_2,t}-\hat{\lambda}_{P_1,i}} \hat{u}_{P_1,i,s}\hat{u}_{P_1,i,2}}{\sqrt{\sum_{i=1}^{m} \frac{\hat{\lambda}^2_{P_1,i}}{\left(\hat{\theta}_{P_2,t}-\hat{\lambda}_{P_1,i}\right)^2} \hat{u}^2_{P_1,i,2}}} = \frac{\sum_{i=1}^{m} \frac{\hat{\lambda}_{P_1,i}}{\hat{\theta}_{P_2,t}-\hat{\lambda}_{P_1,i}} \hat{u}_{P_1,i,s}\hat{u}_{P_1,i,2}}{\sqrt{D_t}},$$

where by Theorem 5.11.1 and assuming $\theta_1 > \theta_2$,

$$\begin{aligned}
D_1 &= \sum_{i=2}^{m} \frac{\hat{\lambda}^2_{P_1,i}}{(\hat{\theta}_{P_2,1} - \hat{\lambda}_{P_1,i})^2} \hat{u}^2_{P_1,i,2} + \frac{\hat{\theta}^2_{P_1,1}}{(\hat{\theta}_{P_2,1} - \hat{\theta}_{P_1,1})^2} \hat{u}^2_{P_1,1,2} \\
&= \underbrace{\frac{\hat{\theta}^2_{P_1,1}}{(\hat{\theta}_{P_2,1} - \hat{\theta}_{P_1,1})^2} \hat{u}^2_{P_1,1,2}}_{\sim \frac{\theta_1 m}{\theta_2^2}} + O_p\left(\frac{1}{\theta_1^2}\right), \\
D_2 &= \sum_{i=2}^{m} \frac{\hat{\lambda}^2_{P_1,i}}{(\hat{\theta}_{P_2,2} - \hat{\lambda}_{P_1,i})^2} \hat{u}^2_{P_1,i,2} + \frac{\hat{\theta}^2_{P_1,1}}{(\hat{\theta}_{P_2,2} - \hat{\theta}_{P_1,1})^2} \hat{u}^2_{P_1,1,2} \\
&= O_p\left(\frac{1}{\theta_2^2}\right) + O_p\left(\frac{1}{\theta_1 m}\right).
\end{aligned}$$

By Theorem 5.11.1, $\hat{u}_{P_2,t,s} = \frac{P_2^{1/2}\tilde{u}_{P_2,t}}{N_t}$, where

$$\begin{aligned}
N_t^2 &= \tilde{u}^2_{P_2,t,1} + \sum_{i=3}^{m} \tilde{u}^2_{P_2,t,i} + \tilde{u}^2_{P_2,t,2}\theta_2 \\
&= 1 + (\theta_2 - 1)\tilde{u}^2_{P_2,t,2} \\
&= 1 + \frac{1}{(\theta_2 - 1)D_t}.
\end{aligned}$$

Then,

$$\begin{aligned}
N_t^2 D_t &= D_t + \frac{1}{(\theta_2 - 1)}, \\
N_1^2 D_1 &= \frac{\hat{\theta}_{P_1,1}^2}{(\hat{\theta}_{P_2,1} - \hat{\theta}_{P_1,1})^2} \hat{u}_{P_1,1,2}^2 + O_p\left(\frac{1}{\theta_2}\right), \\
N_2^2 D_2 &= \frac{1}{(\theta_2 - 1)} + O_p\left(\frac{1}{\theta_2^2}\right) + O_p\left(\frac{1}{\theta_1 m}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{1}{N_1\sqrt{D_1}} &= \frac{|\hat{\theta}_{P_2,1} - \hat{\theta}_{P_1,1}|}{\hat{\theta}_{P_1,1}|\hat{u}_{P_1,1,2}|} + O_p\left(\frac{\theta_2^2}{\theta_1^{3/2} m^{3/2}}\right) \\
&= O_p\left(\frac{\theta_2}{\theta_1^{1/2} m^{1/2}}\right), \\
\frac{1}{N_2\sqrt{D_2}} &= \sqrt{\theta_2 - 1} + O_p\left(\frac{1}{\theta_2^{1/2}}\right) + O_p\left(\frac{\theta_2^{3/2}}{\theta_1 m}\right).
\end{aligned}$$

B: We are now able to study the quantity:

$$\sum_{s=3}^{m} \hat{u}_{P_k,1,s} \hat{u}_{P_k,2,s}$$

First,

$$\hat{u}_{P_2,1,s} \hat{u}_{P_2,2,s} = \frac{\sum_{i,j=1}^{m} \frac{\hat{\lambda}_{P_1,i}\hat{\lambda}_{P_1,j}}{(\hat{\theta}_{P_2,1} - \hat{\lambda}_{P_1,i})(\hat{\theta}_{P_2,2} - \hat{\lambda}_{P_1,i})} \hat{u}_{P_1,i,s}\hat{u}_{P_1,j,s}\hat{u}_{P_1,i,2}\hat{u}_{P_1,j,2}}{\sqrt{D_1 D_2} N_1 N_2}.$$

Then,

$$\begin{aligned}
&\sum_{s=3}^{m} \hat{u}_{P_2,1,s} \hat{u}_{P_2,2,s} \\
&= \frac{\sum_{s=3}^{m} \sum_{i,j=1}^{m} \frac{\hat{\lambda}_{P_1,i}\hat{\lambda}_{P_1,j}}{(\hat{\theta}_{P_2,1} - \hat{\lambda}_{P_1,i})(\hat{\theta}_{P_2,2} - \hat{\lambda}_{P_1,j})} \hat{u}_{P_1,i,s}\hat{u}_{P_1,j,s}\hat{u}_{P_1,i,2}\hat{u}_{P_1,j,2}}{\sqrt{D_1 D_2} N_1 N_2} \\
&= \frac{1}{\sqrt{D_1 D_2} N_1 N_2} \left(\sum_{i,j=1, i\neq j}^{m} \frac{\hat{\lambda}_{P_1,i}\hat{\lambda}_{P_1,j}}{(\hat{\theta}_{P_2,1} - \hat{\lambda}_{P_1,i})(\hat{\theta}_{P_2,2} - \hat{\lambda}_{P_1,j})} \hat{u}_{P_1,i,2}\hat{u}_{P_1,j,2} \left(\sum_{s=3}^{m} \hat{u}_{P_1,i,s}\hat{u}_{P_1,j,s}\right) \right. \\
&\quad \left. + \sum_{i=1}^{m} \frac{\hat{\lambda}_{P_1,i}^2}{(\hat{\theta}_{P_2,1} - \hat{\lambda}_{P_1,i})(\hat{\theta}_{P_2,2} - \hat{\lambda}_{P_1,i})} \hat{u}_{P_1,i,2}^2 \left(\sum_{s=3}^{m} \hat{u}_{P_1,i,s}^2\right)\right) \\
&= \frac{1}{\sqrt{D_1 D_2} N_1 N_2} \\
&\quad \left(\underbrace{- \sum_{i,j=1, i\neq j}^{m} \frac{\hat{\lambda}_{P_1,i}\hat{\lambda}_{P_1,j}}{(\hat{\theta}_{P_2,1} - \hat{\lambda}_{P_1,i})(\hat{\theta}_{P_2,2} - \hat{\lambda}_{P_1,j})} \left(\hat{u}_{P_1,i,2}^2 \hat{u}_{P_1,j,2}^2 + \hat{u}_{P_1,i,1}\hat{u}_{P_1,j,1}\hat{u}_{P_1,i,2}\hat{u}_{P_1,j,2}\right)}_{\text{Part 2}} \right. \\
&\quad \left. + \underbrace{\sum_{i=1}^{m} \frac{\hat{\lambda}_{P_1,i}^2}{(\hat{\theta}_{P_2,1} - \hat{\lambda}_{P_1,i})(\hat{\theta}_{P_2,2} - \hat{\lambda}_{P_1,i})} \hat{u}_{P_1,i,2}^2 \left(1 - \hat{u}_{P_1,i,1}^2 - \hat{u}_{P_1,i,2}^2\right)}_{\text{Part 1}} \right).
\end{aligned}$$

Using the part **A**,

$$\frac{1}{\sqrt{D_1 D_2} N_1 N_2} = O_p\left(\frac{\theta_2^{3/2}}{\theta_1^{1/2} m^{1/2}}\right).$$

Then, we can study Part 1 and Part 2 and neglect terms smaller than $O_p\left(\frac{1}{\theta_2^2}\right)$. (If at least one term is of order $\frac{1}{\theta_2^2}$.)

Part 1: We decompose the sum of Part 1 into $i = 1$ and $i > 1$. Then, using Theorems 5.11.1, 5.8.1 and 5.4.1, each term can be estimated.

1.1) i=1 :

$$\frac{\hat{\lambda}_{P_1,1}^2}{(\hat{\theta}_{P_2,1} - \hat{\lambda}_{P_1,1})(\hat{\theta}_{P_2,2} - \hat{\lambda}_{P_1,1})}\hat{u}_{P_1,1,2}^2\left(1 - \hat{u}_{P_1,1,1}^2 - \hat{u}_{P_1,1,2}^2\right)$$
$$= \underbrace{\frac{\boldsymbol{\hat{\theta}^2_{P_1,1}}}{(\boldsymbol{\hat{\theta}_{P_2,1}} - \boldsymbol{\hat{\theta}_{P_1,1}})(\boldsymbol{\hat{\theta}_{P_2,2}} - \boldsymbol{\hat{\theta}_{P_1,1}})}\boldsymbol{\hat{u}^2_{P_1,1,2}}\left(\mathbf{1} - \boldsymbol{\hat{\alpha}^2_{P_1,1}}\right)}_{O_p\left(\frac{1}{\theta_1\theta_2}\right)} - \underbrace{\frac{\hat{\theta}_{P_1,1}^2}{(\hat{\theta}_{P_2,1} - \hat{\theta}_{P_1,1})(\hat{\theta}_{P_2,2} - \hat{\theta}_{P_1,1})}\hat{u}_{P_1,1,2}^4}_{O_p\left(\frac{1}{\theta_1\theta_2 m}\right)}.$$

1.2) i>1 :

∗ First, we show a small non-optimal result

$$\sum_{i=2}^{m} \hat{u}_{P_1,i,2}^2 \hat{u}_{P_1,i,1}^2 = O_p\left(\frac{1}{\theta_1 m^{1/2}}\right).$$

We easily obtain this result by using some inequalities on the sums,

$$\begin{aligned}\sum_{i=2}^{m} \hat{u}_{P_1,i,2}^2 \hat{u}_{P_1,i,1}^2 &\leqslant \left(\sum_{i=2}^{m} \hat{u}_{P_1,i,2}^4\right)^{1/2}\left(\sum_{i=2}^{m} \hat{u}_{P_1,i,1}^4\right)^{1/2} \\ &= O_p\left(\frac{1}{\theta_1 m^{1/2}}\right).\end{aligned}$$

By Theorem 5.8.1 Part 3, $\sum_{i=2}^{m} \hat{u}_{P_1,i,1}^4 = O_p\left(\frac{1}{\theta_1^2}\right)$, and the estimation $\sum_{i=2}^{m} \hat{u}_{P_1,i,2}^4 = O_p\left(1/\sqrt{m}\right)$ is obtained by the spherical property. Indeed, because $\hat{u}_{P_1,i,2:m}$ is invariant by rotation, then $\hat{u}_{P_1,i,2:m}/||\hat{u}_{P_1,i,2:m}||$ is uniform. Therefore,

$$\mathrm{E}\left[\frac{\hat{u}_{P_1,i,2}^4}{||\hat{u}_{P_1,i,2:m}||^4}\right] = O_p\left(\frac{1}{m^2}\right) \text{ and } \mathrm{E}\left[\frac{\hat{u}_{P_1,i,2}^8}{||\hat{u}_{P_1,i,2:m}||^8}\right] = O_p\left(\frac{1}{m^4}\right).$$

We see that $\hat{u}_{P_1,i,2}^4 \sim \mathrm{RV}\left(O\left(\frac{1}{m^2}\right), O\left(\frac{1}{m^4}\right)\right)$. Finally, summing the random variables leads to

$$\begin{aligned}\mathrm{E}\left[\sum_{i=2}^{m} \hat{u}_{P_1,i,2}^4\right] &= O_p\left(\frac{1}{m}\right), \\ \mathrm{Var}\left(\sum_{i=2}^{m} \hat{u}_{P_1,i,2}^4\right) &= O_p\left(\frac{1}{m^2}\right).\end{aligned}$$

$*$ Now, we can estimate the sum of interest:

$$\begin{aligned}
&\sum_{i=2}^{m}\frac{\hat\lambda^2_{P_1,i}}{(\hat\theta_{P_2,1}-\hat\lambda_{P_1,i})(\hat\theta_{P_2,2}-\hat\lambda_{P_1,i})}\hat u^2_{P_1,i,2}\left(1-\hat u^2_{P_1,i,1}-\hat u^2_{P_1,i,2}\right)\\
&=\frac{1}{\hat\theta_{P_2,1}\hat\theta_{P_2,2}}\sum_{i=2}^{m}\hat\lambda^2_{P_1,i}\hat u^2_{P_1,i,2}\left(1-\hat u^2_{P_1,i,1}-\hat u^2_{P_1,i,2}\right)+O_p\left(\frac{1}{\theta_1\theta_2^2}\right)\\
&=\frac{1}{\hat\theta_{P_2,1}\hat\theta_{P_2,2}}\left(\sum_{i=2}^{m}\hat\lambda^2_{P_1,i}\hat u^2_{P_1,i,2}+O_p(1)\sum_{i=2}^{m}\hat u^2_{P_1,i,2}\hat u^2_{P_1,i,1}+O_p(1)\sum_{i=2}^{m}\hat u^4_{P_1,i,2}\right)+O_p\left(\frac{1}{\theta_1\theta_2^2}\right)\\
&=\frac{1}{\hat\theta_{P_2,1}\hat\theta_{P_2,2}}\sum_{i=2}^{m}\hat\lambda^2_{P_1,i}\hat u^2_{P_1,i,2}+O_p\left(\frac{1}{\theta_1\theta_2}\right)\sum_{i=2}^{m}\hat u^2_{P_1,i,2}\hat u^2_{P_1,i,1}\\
&\qquad+O_p\left(\frac{1}{\theta_1\theta_2}\right)\sum_{i=2}^{m}\hat u^4_{P_1,i,2}+O_p\left(\frac{1}{\theta_1\theta_2^2}\right)\\
&=\underbrace{\frac{\mathbf{1}}{\hat\theta_{\mathbf{P_2,1}}\hat\theta_{\mathbf{P_2,2}}}\sum_{\mathbf{i=2}}^{\mathbf{m}}\hat\lambda^{\mathbf{2}}_{\mathbf{P_1,i}}\hat{\mathbf{u}}^{\mathbf{2}}_{\mathbf{P_1,i,2}}}_{O_p\left(\frac{1}{\theta_1\theta_2}\right)}+O_p\left(\frac{1}{\theta_1^2\theta_2 m^{1/2}}\right)+O_p\left(\frac{1}{\theta_1\theta_2 m}\right)+O_p\left(\frac{1}{\theta_1\theta_2^2}\right).
\end{aligned}$$

Part 2: As for the previous part, we divide this term.

2.1) $\sum_{i,j=1,i\neq j}^{m}\frac{\hat\lambda_{P_1,i}\hat\lambda_{P_1,j}}{(\hat\theta_{P_2,1}-\hat\lambda_{P_1,i})(\hat\theta_{P_2,2}-\hat\lambda_{P_1,j})}\hat u_{P_1,i,1}\hat u_{P_1,j,1}\hat u_{P_1,i,2}\hat u_{P_1,j,2}$.

2.1.1) i=1,j>1 : We want to prove

$$\frac{\hat\theta_{\mathbf{P_1,1}}}{(\hat\theta_{\mathbf{P_2,1}}-\hat\theta_{\mathbf{P_1,1}})}\hat{\mathbf{u}}_{\mathbf{P_1,1,1}}\hat{\mathbf{u}}_{\mathbf{P_1,1,2}}\sum_{\mathbf{j>1}}^{\mathbf{m}}\frac{\hat\lambda_{\mathbf{P_1,j}}}{(\hat\theta_{\mathbf{P_2,2}}-\hat\lambda_{\mathbf{P_1,j}})}\hat{\mathbf{u}}_{\mathbf{P_1,j,1}}\hat{\mathbf{u}}_{\mathbf{P_1,j,2}}=O_p\left(\frac{1}{\theta_2^2}\right).$$

The order size follows from Theorems 5.11.1, 5.8.1 and 5.4.1,

$$\underbrace{\frac{\hat\theta_{P_1,1}}{(\hat\theta_{P_2,1}-\hat\theta_{P_1,1})}\hat u_{P_1,1,1}\hat u_{P_1,1,2}}_{O_p\left(\frac{\theta_1^{1/2}m^{1/2}}{\theta_2}\right)}\underbrace{\sum_{j>1}^{m}\frac{\hat\lambda_{P_1,j}}{(\hat\theta_{P_2,2}-\hat\lambda_{P_1,j})}\hat u_{P_1,j,1}\hat u_{P_1,j,2}}_{O_p\left(\frac{1}{\theta_1^{1/2}\theta_2 m^{1/2}}\right)}=O_p\left(\frac{1}{\theta_2^2}\right).$$

Remark 7.1.2.
The Theorem 5.11.1 estimates the order size of the second term for $\hat\theta_{P_2,1}$. However, the same proof is still valid in this case.

2.1.2) i>1,j=1 : Using the fact that $\hat\lambda_{P_1,i}$ is bounded for $i>1$, then

$$\underbrace{\frac{\hat\theta_{P_1,1}}{(\hat\theta_{P_2,2}-\hat\theta_{P_1,1})}}_{O_p(1)}\underbrace{\hat u_{P_1,1,1}}_{O_p(1)}\underbrace{\hat u_{P_1,1,2}}_{O_p\left(\frac{1}{m^{1/2}\theta_1^{1/2}}\right)}\underbrace{\sum_{i>1}^{m}\frac{\hat\lambda_{P_1,i}}{(\hat\theta_{P_2,1}-\hat\lambda_{P_1,i})}\hat u_{P_1,i,1}\hat u_{P_1,i,2}}_{O_p\left(\frac{1}{\theta_1^{3/2}m^{1/2}}\right)}=O_p\left(\frac{1}{\theta_1^2 m}\right).$$

2.1.3) i>1,j>1,i≠j :

$$\left|\sum_{i,j>1,i\neq j}^{m}\frac{\hat{\lambda}_{P_1,i}\hat{\lambda}_{P_1,j}}{(\hat{\theta}_{P_2,1}-\hat{\lambda}_{P_1,i})(\hat{\theta}_{P_2,2}-\hat{\lambda}_{P_1,j})}\hat{u}_{P_1,i,1}\hat{u}_{P_1,j,1}\hat{u}_{P_1,i,2}\hat{u}_{P_1,j,2}\right|$$

$$\leqslant\left(\frac{1}{\hat{\theta}_{P_2,1}\hat{\theta}_{P_2,2}}+O_p\left(\frac{1}{\theta_1\theta_2^2}\right)\right)\left(\sum_{i>1}^{m}\hat{\lambda}_{P_1,i}|\hat{u}_{P_1,i,1}||\hat{u}_{P_1,i,2}|\right)\left(\sum_{j>1}^{m}\hat{\lambda}_{P_1,j}|\hat{u}_{P_1,j,1}||\hat{u}_{P_1,j,2}|\right)$$

$$\leqslant\left(\frac{1}{\hat{\theta}_{P_2,1}\hat{\theta}_{P_2,2}}+O_p\left(\frac{1}{\theta_1\theta_2^2}\right)\right)\hat{\lambda}_{\max}^2\left(\sum_{i>1}^{m}\hat{u}_{P_1,i,1}^2\right)\left(\sum_{i>1}^{m}\hat{u}_{P_1,i,2}^2\right)$$

$$\leqslant\left(\frac{1}{\hat{\theta}_{P_2,1}\hat{\theta}_{P_2,2}}+O_p\left(\frac{1}{\theta_1\theta_2^2}\right)\right)\hat{\lambda}_{\max}^2\left(1-\hat{\alpha}_{P_1,1}^2\right)\left(1-\hat{u}_{P_1,1,2}^2\right)$$

$$=O_p\left(\frac{1}{\theta_1^2\theta_2}\right).$$

2.2)

$$\sum_{i,j=1,i\neq j}^{m}\frac{\hat{\lambda}_{P_1,i}\hat{\lambda}_{P_1,j}}{(\hat{\theta}_{P_2,1}-\hat{\lambda}_{P_1,i})(\hat{\theta}_{P_2,2}-\hat{\lambda}_{P_1,j})}\hat{u}_{P_1,i,2}^2\hat{u}_{P_1,j,2}^2$$

$$=\sum_{i,j=1}^{m}\frac{\hat{\lambda}_{P_1,i}\hat{\lambda}_{P_1,j}}{(\hat{\theta}_{P_2,1}-\hat{\lambda}_{P_1,i})(\hat{\theta}_{P_2,2}-\hat{\lambda}_{P_1,j})}\hat{u}_{P_1,i,2}^2\hat{u}_{P_1,j,2}^2-\sum_{i=1}^{m}\frac{\hat{\lambda}_{P_1,i}^2}{(\hat{\theta}_{P_2,1}-\hat{\lambda}_{P_1,i})(\hat{\theta}_{P_2,2}-\hat{\lambda}_{P_1,i})}\hat{u}_{P_1,i,2}^4$$

$$=\frac{\mathbf{1}}{(\boldsymbol{\theta_2}-\mathbf{1})^{\mathbf{2}}}-\underbrace{\frac{\hat{\theta}_{P_1,1}^2}{(\hat{\theta}_{P_2,1}-\hat{\theta}_{P_1,1})(\hat{\theta}_{P_2,2}-\hat{\theta}_{P_1,1})}\hat{u}_{P_1,1,2}^4}_{O_p\left(\frac{1}{\theta_1\theta_2 m}\right)}-\underbrace{\sum_{i=2}^{m}\frac{\hat{\lambda}_{P_1,i}^2}{(\hat{\theta}_{P_2,1}-\hat{\lambda}_{P_1,i})(\hat{\theta}_{P_2,2}-\hat{\lambda}_{P_1,i})}\hat{u}_{P_1,i,2}^4}_{O_p\left(\frac{1}{\theta_1\theta_2 m}\right)}.$$

Combining Part 1 and Part 2 leads to

$$\sum_{s=3}^{m}\hat{u}_{P_2,1,s}\hat{u}_{P_2,2,s}$$

$$=\frac{1}{\sqrt{D_1D_2}N_1N_2}\left(\frac{\hat{\theta}_{P_1,1}^2}{(\hat{\theta}_{P_2,1}-\hat{\theta}_{P_1,1})(\hat{\theta}_{P_2,2}-\hat{\theta}_{P_1,1})}\hat{u}_{P_1,1,2}^2\left(1-\hat{\alpha}_{P_1,1}^2\right)+\frac{1}{\hat{\theta}_{P_2,1}\hat{\theta}_{P_2,2}}\sum_{i=2}^{m}\hat{\lambda}_{P_1,i}^2\hat{u}_{P_1,i,2}^2\right.$$
$$\left.-\frac{\hat{\theta}_{P_1,1}}{(\hat{\theta}_{P_2,1}-\hat{\theta}_{P_1,1})}\hat{u}_{P_1,1,1}\hat{u}_{P_1,1,2}\frac{1}{\hat{\theta}_{P_2,2}}\sum_{j>1}^{m}\hat{\lambda}_{P_1,j}\hat{u}_{P_1,j,1}\hat{u}_{P_1,j,2}-\frac{1}{(\theta_2-1)^2}\right)$$
$$+O_p\left(\frac{\theta_2^{1/2}}{\theta_1^{3/2}m^{3/2}}\right)+O_p\left(\frac{1}{\theta_1^{3/2}\theta_2^{1/2}m^{1/2}}\right)$$

$$=\frac{|\hat{\theta}_{P_2,1}-\hat{\theta}_{P_1,1}|}{\hat{\theta}_{P_1,1}|\hat{u}_{P_1,1,2}|}\sqrt{\theta_2-1}\left(\frac{\hat{\theta}_{P_1,1}^2}{(\hat{\theta}_{P_2,1}-\hat{\theta}_{P_1,1})(\hat{\theta}_{P_2,2}-\hat{\theta}_{P_1,1})}\hat{u}_{P_1,1,2}^2\left(1-\hat{\alpha}_{P_1,1}^2\right)\right.$$
$$+\frac{1}{\hat{\theta}_{P_2,1}\hat{\theta}_{P_2,2}}\sum_{i=2}^{m}\hat{\lambda}_{P_1,i}^2\hat{u}_{P_1,i,2}^2$$
$$\left.-\frac{\hat{\theta}_{P_1,1}}{(\hat{\theta}_{P_2,1}-\hat{\theta}_{P_1,1})}\hat{u}_{P_1,1,1}\hat{u}_{P_1,1,2}\frac{1}{\hat{\theta}_{P_2,2}}\sum_{j>1}^{m}\hat{\lambda}_{P_1,j}\hat{u}_{P_1,j,1}\hat{u}_{P_1,j,2}-\frac{1}{(\theta_2-1)^2}\right)$$
$$+O_p\left(\frac{\theta_2^{1/2}}{\theta_1^{3/2}m^{3/2}}\right)+O_p\left(\frac{1}{\theta_1^{3/2}\theta_2^{1/2}m^{1/2}}\right)+O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{5/2}m^{1/2}}\right).$$

In this second part we simplify the terms using Theorems 5.8.1 and 5.4.1:

$$\hat{u}_{P_k,1,k} = \frac{\sqrt{\theta_k}\theta_1}{|\theta_k - \theta_1|}|\hat{u}_{P_{k-1},1,k}| + O_p\left(\frac{\min(\theta_1,\theta_k)}{\theta_1^{1/2}\theta_k^{1/2}m}\right) + O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{1/2}m^{1/2}}\right)$$

and

$$\hat{\theta}_{P_2,1} - \hat{\theta}_{P_1,1} = -\frac{\hat{\theta}_{P_1,1}\hat{\theta}_{P_2,1}(\theta_2-1)}{\theta_2 - 1 - \hat{\theta}_{P_2,1}}\hat{u}^2_{P_1,1,2} + O_p\left(\frac{\theta_2}{m^{3/2}}\right) + O_p\left(\frac{1}{m}\right).$$

We remember that without the convention $\hat{u}_{P_k,1,1} > 0$, then, by construction, we have $\hat{u}_{P_k,1,k} > 0$. Because $\theta_1 > \theta_2$,

- $P1$:

$$\begin{aligned}
P_1 &= \frac{|\hat{\theta}_{P_2,1} - \hat{\theta}_{P_1,1}|}{\hat{\theta}_{P_1,1}|\hat{u}_{P_1,1,2}|}\sqrt{\theta_2 - 1}\frac{\hat{\theta}^2_{P_1,1}}{(\hat{\theta}_{P_2,1} - \hat{\theta}_{P_1,1})(\hat{\theta}_{P_2,2} - \hat{\theta}_{P_1,1})}\hat{u}^2_{P_1,1,2}\left(1 - \hat{\alpha}^2_{P_1,1}\right) \\
&\overset{Asy}{=} \sqrt{\theta_2 - 1}\frac{\hat{\theta}_{P_1,1}}{\hat{\theta}_{P_2,2} - \hat{\theta}_{P_1,1}}|\hat{u}_{P_1,1,2}|\left(1 - \hat{\alpha}^2_{P_1,1}\right) \\
&\overset{Asy}{=} -\hat{u}_{P_2,1,2}\left(1 - \hat{\alpha}^2_{P_1,1}\right) + O_p\left(\frac{\theta_2^{1/2}}{\theta_1^{3/2}m}\right) + O_p\left(\frac{1}{\theta_1^{3/2}\theta_2^{1/2}m^{1/2}}\right).
\end{aligned}$$

  We use the notation $\overset{Asy}{=}$ because the probability that the sign is wrong tends to 0 in $1/m$ when θ_1 tends to infinity. Moreover, when θ_1 is finite, the order size is $1/\sqrt{m}$.

- $P2$:

$$\begin{aligned}
P_2 &= \frac{|\hat{\theta}_{P_2,1} - \hat{\theta}_{P_1,1}|}{\hat{\theta}_{P_1,1}|\hat{u}_{P_1,1,2}|}\sqrt{\theta_2 - 1}\frac{1}{\hat{\theta}_{P_2,1}\hat{\theta}_{P_2,2}}\sum_{i=2}^{m}\hat{\lambda}^2_{P_1,i}\hat{u}^2_{P_1,i,2} \\
&= \frac{\hat{\theta}_{P_1,1}\hat{\theta}_{P_2,1}\theta_2}{|\theta_2 - \hat{\theta}_{P_2,1}|}\hat{u}^2_{P_1,1,2}\frac{1}{\hat{\theta}_{P_1,1}|\hat{u}_{P_1,1,2}|}\sqrt{\theta_2 - 1}\frac{1}{\hat{\theta}_{P_2,1}\hat{\theta}_{P_2,2}}\sum_{i=2}^{m}\hat{\lambda}^2_{P_1,i}\hat{u}^2_{P_1,i,2} \\
&\quad + O_p\left(\frac{\theta_2^{1/2}}{\theta_1^{3/2}m}\right) + O_p\left(\frac{1}{\theta_1^{3/2}\theta_2^{1/2}m^{1/2}}\right) \\
&= \frac{|\hat{u}_{P_1,1,2}|\sqrt{\theta_2 - 1}}{|\theta_2 - \hat{\theta}_{P_2,1}|}\sum_{i=2}^{m}\hat{\lambda}^2_{P_1,i}\hat{u}^2_{P_1,i,2} + O_p\left(\frac{\theta_2^{1/2}}{\theta_1^{3/2}m}\right) + O_p\left(\frac{1}{\theta_1^{3/2}\theta_2^{1/2}m^{1/2}}\right) \\
&= \frac{|\hat{u}_{P_2,1,2}|}{\theta_1}\sum_{i=2}^{m}\hat{\lambda}^2_{P_1,i}\hat{u}^2_{P_1,i,2} + O_p\left(\frac{\theta_2^{1/2}}{\theta_1^{3/2}m}\right) + O_p\left(\frac{1}{\theta_1^{3/2}\theta_2^{1/2}m^{1/2}}\right).
\end{aligned}$$

- $P3$: Using Lemma 6.1.1,

$$\begin{aligned}
P_3 &= \frac{|\hat{\theta}_{P_2,1} - \hat{\theta}_{P_1,1}|}{\hat{\theta}_{P_1,1}|\hat{u}_{P_1,1,2}|}\sqrt{\theta_2 - 1}\frac{\hat{\theta}_{P_1,1}}{(\hat{\theta}_{P_2,1} - \hat{\theta}_{P_1,1})}\hat{u}_{P_1,1,1}\hat{u}_{P_1,1,2}\frac{1}{\hat{\theta}_{P_2,2}}\sum_{j>1}^{m}\hat{\lambda}_{P_1,j}\hat{u}_{P_1,j,1}\hat{u}_{P_1,j,2} \\
&\overset{Asy}{=} \operatorname{sign}\left(\hat{u}_{P_2,1,1}\right)\frac{1}{\theta_2^{1/2}}\sum_{j>1}^{m}\hat{\lambda}_{P_1,j}\hat{u}_{P_1,j,1}\hat{u}_{P_1,j,2} + O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{3/2}m^{1/2}}\right) + O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{1/2}m}\right),
\end{aligned}$$

  where the sign equality is obtained by the remark of Theorem 5.11.1 and tends to be correct in $1/m$.

- $P4$:

$$\begin{aligned} P_4 &= \frac{1}{\sqrt{D_1}}\sqrt{\theta_2-1}\frac{1}{(\theta_2-1)^2} \\ &= \frac{\sqrt{\theta_2-1}}{\theta_2-1}\tilde{u}_{P_2,1,2} \\ &= \frac{1}{\theta_2-1}\hat{u}_{P_2,1,2} \\ &= \frac{1}{\theta_2}\hat{u}_{P_2,1,2}+O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{3/2}m^{1/2}}\right) \end{aligned}$$

By construction we know that $\hat{u}_{P_2,1,2}>0$, but this is not the case for $\hat{u}_{P_2,1,1}$. We will correct this convention later. First, we can combine $P_1+P_2-P_4$ to obtain

$$\begin{aligned} P_1+P_2-P_4 &\overset{Asy}{=} \hat{u}_{P_2,1,2}\left(-\left(1-\hat{\alpha}^2_{P_1,1}\right)+\frac{\sum_{i=2}^m\hat{\lambda}^2_{P_1,i}\hat{u}^2_{P_1,i,2}}{\theta_1}-\frac{1}{\theta_2}\right) \\ &\quad +O_p\left(\frac{\theta_2^{1/2}}{\theta_1^{3/2}m}\right)+O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{3/2}m^{1/2}}\right) \\ &\overset{Asy}{=} \hat{u}_{P_2,1,2}\left(-\frac{\sum_{i=1}^m\hat{\lambda}^2_{W,i}\hat{u}^2_{W,i,1}-1}{\theta_1}+\frac{\sum_{i=2}^m\hat{\lambda}^2_{P_1,i}\hat{u}^2_{P_1,i,2}}{\theta_1}-\frac{1}{\theta_2}\right) \\ &\quad +O_p\left(\frac{\theta_2^{1/2}}{\theta_1^{3/2}m}\right)+O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{3/2}m^{1/2}}\right) \\ &\overset{Asy}{=} \hat{u}_{P_2,1,2}\left(\frac{1}{\theta_1}-\frac{1}{\theta_2}\right)+O_p\left(\frac{\theta_2^{1/2}}{\theta_1^{3/2}m}\right)+O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{3/2}m^{1/2}}\right). \end{aligned}$$

Indeed Lemma 6.1.1 shows that

$$\begin{aligned} \sum_{i=1}^m\hat{\lambda}^2_{W,i}\hat{u}^2_{W,i,1} &= \left(W^2\right)_{1,1}, \\ \sum_{i=2}^m\hat{\lambda}^2_{P_1,i}\hat{u}^2_{P_1,i,2} &= \left(W^2\right)_{2,2}+O_p\left(\frac{1}{m}\right). \end{aligned}$$

The result follows by invariance of W^2 under rotation.

Finally, we combine the different parts

$$\begin{aligned} P_1+P_2-P_3-P_4 &\overset{Asy}{=} \hat{u}_{P_2,1,2}\left(\frac{1}{\theta_1}-\frac{1}{\theta_2}\right)-\operatorname{sign}\left(\hat{u}_{P_2,1,1}\right)\frac{1}{\theta_2^{1/2}}\sum_{j>1}^m\hat{\lambda}_{P_1,j}\hat{u}_{P_1,j,1}\hat{u}_{P_1,j,2} \\ &\quad +O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{1/2}m}\right)+O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{3/2}m^{1/2}}\right), \end{aligned}$$

where the asymptotic equality is discussed in the remark 7.1.3.
We change the convention of the sign such that $\hat{u}_{P_2,i,i}>0$, $i=1,2$. Therefore, we multiply by $\operatorname{sign}\left(\hat{u}_{P_2,1,1}\right)$. With this convention $\hat{u}_{P_2,1,2}$ is not strictly positive anymore. Nevertheless, we keep using the same notation.

$$\begin{aligned} P_1+P_2-P_3-P_4 &= \hat{u}_{P_2,1,2}\left(\frac{1}{\theta_1}-\frac{1}{\theta_2}\right)-\frac{1}{\theta_2^{1/2}}\sum_{j>1}^m\hat{\lambda}_{P_1,j}\hat{u}_{P_1,j,1}\hat{u}_{P_1,j,2} \\ &\quad +O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{1/2}m}\right)+O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{3/2}m^{1/2}}\right). \end{aligned}$$

Remark 7.1.3.
First we remember that the O errors are in probability and take care of this possible fluctuation with probability tending to 0.
The simplification of $P_1 + P_2 - P_3 - P_4$ is possible thanks to the remark of Theorem 5.11.1 showing that the signs are correct with probability tending to 1 in $1/m$ when θ_2 is large.
In particular, there is a probability of order $1/m$ to have an error of size $O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{1/2}m^{1/2}}\right)$.
Luckily this rare error will not affect the moment estimation of the statistic.
Then, when θ_2 is finite, the formula just provides order size.

This estimation concludes part **B**. Nevertheless, the part **C** uses this formula to provide a distribution.

C: In this section we express

$$\hat{u}_{P_2,1,2}\left(\frac{1}{\theta_1} - \frac{1}{\theta_2}\right)\delta - \frac{1}{\theta_2^{1/2}}\sum_{j>1}^{m}\hat{\lambda}_{P_1,j}\hat{u}_{P_1,j,1}\hat{u}_{P_1,j,2} + O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{1/2}m}\right) + O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{3/2}m^{1/2}}\right)$$

as a function of the unit statistic defined in Theorem 5.11.2. First, using Theorem 5.8.1 and Lemma 6.1.1 leads to the following estimations,

$$\begin{aligned}
\hat{u}_{P_2,1,2} &= \frac{\sqrt{\theta_2}\theta_1}{|\theta_2 - \theta_1|}\hat{u}_{P_1,1,2} + O_p\left(\frac{\theta_2^{1/2}}{\theta_1^{1/2}m}\right) + O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{1/2}m^{1/2}}\right),\\
\hat{u}_{P_1,1,2} &= \frac{W_{1,2}}{\sqrt{\theta_1}} + O_p\left(\frac{1}{\theta_1^{3/2}m^{1/2}}\right) + O_p\left(\frac{1}{\theta_1^{1/2}m}\right),\\
\textstyle\sum_{i=2}^{m}\hat{\lambda}_{P_1,i}\hat{u}_{P_1,i,1}\hat{u}_{P_1,i,2} &= W_{1,2}\frac{M_2}{\sqrt{\theta_1}} - \left(W^2\right)_{1,2}\frac{1}{\sqrt{\theta_1}} + O_p\left(\frac{1}{\theta_1^{1/2}m}\right) + O_p\left(\frac{1}{\theta_1^{3/2}m^{1/2}}\right).
\end{aligned}$$

Therefore, we can show that

$$\begin{aligned}
&\hat{u}_{P_2,1,2}\left(\frac{1}{\theta_1} - \frac{1}{\theta_2}\right)\delta - \frac{1}{\theta_2^{1/2}}\sum_{j>1}^{m}\hat{\lambda}_{P_1,j}\hat{u}_{P_1,j,1}\hat{u}_{P_1,j,2}\\
&= \frac{-\left(\delta + M_2\right)W_{1,2} + \left(W^2\right)_{1,2}}{\sqrt{\theta_1\theta_2}} + O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{1/2}m}\right) + O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{3/2}m^{1/2}}\right).
\end{aligned}$$

The result is straightforward using a delta method and Theorem 5.11.2.

□

Invariant

Theorem 5.7.1. *(Invariant Dot Product Theorem)*
Suppose that W satisfies Assumption 2.2.1 and

$$P_{s,r} = \mathrm{I}_m + \sum_{i=s,r}^{2}(\theta_i - 1)e_ie_i^t$$

$$P_k = \mathrm{I}_m + \sum_{i=1}^{k}(\theta_i - 1)e_ie_i^t \textit{ respects 2.2.2 (A4)},$$

where $\theta_1 > \theta_2 > ... > \theta_k$. We define

$$\hat{\Sigma}_{P_{s,r}} = P_{s,r}^{1/2} W P_{s,r}^{1/2},$$
$$\hat{\Sigma}_{P_k} = P_k^{1/2} W P_k^{1/2}.$$

Moreover, for $s, r = 1, 2, ..., k$ with $s \neq r$, we define

$$\hat{u}_{P_{s,r},1}, \hat{\theta}_{P_{s,r},1} \quad s.t. \quad \hat{\Sigma}_{P_{s,r}} \hat{u}_{P_{s,r},1} = \hat{\theta}_{P_{s,r},1} \hat{u}_{P_{s,r},1},$$
$$\hat{u}_{P_k,s}, \hat{\theta}_{P_k,s} \quad s.t. \quad \hat{\Sigma}_{P_k} \hat{u}_{P_k,s} = \hat{\theta}_{P_k,s} \hat{u}_{P_k,s},$$

where $\hat{\theta}_{P_{s,r},1} = \hat{\lambda}_{\hat{\Sigma}_{P_{s,r},1}}$ and $\hat{\theta}_{P_k,s} = \hat{\lambda}_{\hat{\Sigma}_{P_k},s}$.
Assuming the convention:

$$\text{For } s = 1, 2, ..., k \text{ and } i = 1, 2, ..., s,\ \hat{u}_{P_s,i,i} > 0.$$

leads to

$$\boxed{\sum_{\substack{i=1 \\ i \neq s,r}}^{m} \hat{u}_{P_{s,r},1,i} \hat{u}_{P_{s,r},2,i} = \sum_{i=k+1}^{m} \hat{u}_{P_k,s,i} \hat{u}_{P_k,r,i} + O_p\left(\frac{1}{\sqrt{\theta_s \theta_r m}}\right).}$$

(Page 81)

Proof. **Theorem 5.7.1**

We begin this proof with two important remarks.

- This proof will assume the sign convention of Theorem 5.11.1. Nevertheless we correct this convention in the end of the proof.
- We use the notation 7.1.1 to prove the result only for $\theta_1 > \theta_2$ and relaxing the order of the other eigenvalues. This notation permutes the estimated eigenvalues and their eigenvector but the reader can also read this proof as if $\theta_1 > \theta_2 > ... > \theta_k$ and realize that the notation allows a generalisation. Moreover, we add the notation $\hat{\lambda}_{P_r,i} = \hat{\theta}_{P_r,i}$ for $i = 1, 2, ..., r$ in order to simplify formulas.

Thus, Theorem 5.11.1 leads to

$$\hat{u}_{P_k,1,s} \hat{u}_{P_k,2,s} = \frac{1}{\sqrt{D_1 D_2} N_1 N_2} \sum_{i,j} \frac{\hat{\lambda}_{P_{k-1},i} \hat{\lambda}_{P_{k-1},j}}{\left(\hat{\theta}_{P_k,1} - \hat{\lambda}_{P_{k-1},i}\right)\left(\hat{\theta}_{P_k,2} - \hat{\lambda}_{P_{k-1},j}\right)} \hat{u}_{P_{k-1},i,k} \hat{u}_{P_{k-1},j,k} \hat{u}_{P_{k-1},i,s} \hat{u}_{P_{k-1},j,s},$$

where N_1 and N_2 are scalars such that the vectors are unit. Therefore

$$\sum_{s=k+1}^{m} \hat{u}_{P_k,1,s} \hat{u}_{P_k,2,s} =$$
$$\frac{1}{\sqrt{D_1 D_2} N_1 N_2} \left(\underbrace{\sum_{i \neq j} \frac{\hat{\lambda}_{P_{k-1},i} \hat{\lambda}_{P_{k-1},j}}{\left(\hat{\theta}_{P_k,1} - \hat{\lambda}_{P_{k-1},i}\right)\left(\hat{\theta}_{P_k,2} - \hat{\lambda}_{P_{k-1},j}\right)} \hat{u}_{P_{k-1},i,k} \hat{u}_{P_{k-1},j,k} \left(-\sum_{r=1}^{k} \hat{u}_{P_{k-1},i,r} \hat{u}_{P_{k-1},j,r}\right)}_{\text{Part 2}} \right.$$
$$\left. \underbrace{\sum_{i=1}^{m} \frac{\hat{\lambda}_{P_{k-1},i}^2}{\left(\hat{\theta}_{P_k,1} - \hat{\lambda}_{P_{k-1},i}\right)\left(\hat{\theta}_{P_k,2} - \hat{\lambda}_{P_{k-1},i}\right)} \hat{u}_{P_{k-1},i,k}^2 \left(1 - \sum_{r=1}^{k} \hat{u}_{P_{k-1},i,r}^2\right)}_{\text{Part 1}} \right).$$

First we will study Part 1 and Part 2 in **A**. Then in **B**, we will show

$$\frac{1}{\sqrt{D_1 D_2} N_1 N_2} = O_p\left(\frac{\min(\theta_1,\theta_k)\min(\theta_2,\theta_k)}{\theta_1^{1/2}\theta_2^{1/2} m}\right).$$

Finally, in part **C**, we combine **A** and **B** to conclude the proof.

A: Assuming the previous estimation, we can neglect all the terms of order $o_p\left(\frac{\sqrt{m}}{\min(\theta_1,\theta_k)\min(\theta_2,\theta_k)}\right)$ in Part 1 and 2. The order size of the elements are obtained using Theorems 5.8.1, 5.3.1, 5.4.1, 5.11.1, the Invariant Angle Theorem 5.5.1, the Dot Product Theorem 5.6.1 and its Invariant Theorem 5.7.1.

Part 1 : We will show that we can neglect this part.

1.1) $i = 1$: Assuming without loss of generality that $\theta_1 < \theta_2$ leads to

$$\begin{aligned}
&\frac{\hat{\theta}^2_{P_{k-1},1}}{\left(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1}\right)\left(\hat{\theta}_{P_k,2} - \hat{\theta}_{P_{k-1},1}\right)} \hat{u}^2_{P_{k-1},1,k}\left(1 - \sum_{r=1}^{k} \hat{u}^2_{P_{k-1},1,r}\right) \\
&= \underbrace{\frac{\hat{\theta}^2_{P_{k-1},1}}{\left(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1}\right)\left(\hat{\theta}_{P_k,2} - \hat{\theta}_{P_{k-1},1}\right)} \hat{u}^2_{P_{k-1},1,k}}_{O_p\left(\frac{1}{\min(\theta_1,\theta_k)}\right)} \left(\underbrace{1 - \sum_{r=1}^{k-1} \hat{u}^2_{P_{k-1},1,r}}_{=(1-\hat{\alpha}^2_{P_{k-1},1})=O_p\left(\frac{1}{\theta_1}\right)} \quad \underbrace{- \hat{u}^2_{P_{k-1},1,k}}_{O_p\left(\frac{1}{\theta_1 m}\right)}\right) \\
&= O_p\left(\frac{1}{\theta_1 \min(\theta_1,\theta_k)}\right).
\end{aligned}$$

1.2) $i = 2$:

$$\frac{\hat{\theta}^2_{P_{k-1},2}}{\left(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},2}\right)\left(\hat{\theta}_{P_k,2} - \hat{\theta}_{P_{k-1},2}\right)} \hat{u}^2_{P_{k-1},2,k}\left(1 - \sum_{r=1}^{k} \hat{u}^2_{P_{k-1},2,r}\right) = O_p\left(\frac{1}{\theta_1 \min(\theta_2,\theta_k)}\right).$$

1.3) $i = 3, ..., k-1$:

$$\frac{\hat{\theta}^2_{P_{k-1},i}}{\left(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},i}\right)\left(\hat{\theta}_{P_k,2} - \hat{\theta}_{P_{k-1},i}\right)} \hat{u}^2_{P_{k-1},i,k}\left(1 - \sum_{r=1}^{k} \hat{u}^2_{P_{k-1},i,r}\right) = O_p\left(\frac{1}{\max(\theta_1,\theta_i)\max(\theta_2,\theta_i) m}\right).$$

1.4) $i \geqslant k$:

$$\begin{aligned}
&\frac{\hat{\lambda}^2_{P_{k-1},i}}{\left(\hat{\theta}_{P_k,1} - \hat{\lambda}_{P_{k-1},i}\right)\left(\hat{\theta}_{P_k,2} - \hat{\lambda}_{P_{k-1},i}\right)} \hat{u}^2_{P_{k-1},i,k}\left(1 - \sum_{r=1}^{k} \hat{u}^2_{P_{k-1},i,r}\right) = O_p\left(\frac{1}{\theta_1\theta_2 m}\right) \\
&\Rightarrow \sum_{i=k}^{m} \frac{\hat{\lambda}^2_{P_{k-1},i}}{\left(\hat{\theta}_{P_k,1} - \hat{\lambda}_{P_{k-1},i}\right)\left(\hat{\theta}_{P_k,2} - \hat{\lambda}_{P_{k-1},i}\right)} \hat{u}^2_{P_{k-1},i,k}\left(1 - \sum_{r=1}^{k} \hat{u}^2_{P_{k-1},i,r}\right) = O_p\left(\frac{1}{\theta_1\theta_2}\right).
\end{aligned}$$

Part 2 : The second part is trickier but many elements can be neglected.

2.1) $i \neq j \geqslant k$: By the previous part, if $i = j \geqslant k$, then the sum is $O_p\left(\frac{1}{\theta_1\theta_2}\right)$.

$$\begin{aligned}
&\left|\sum_{i\neq j\geqslant k} \frac{\hat{\lambda}_{P_{k-1},i}\hat{\lambda}_{P_{k-1},j}}{\left(\hat{\theta}_{P_k,1}-\hat{\lambda}_{P_{k-1},i}\right)\left(\hat{\theta}_{P_k,2}-\hat{\lambda}_{P_{k-1},j}\right)}\hat{u}_{P_{k-1},i,k}\hat{u}_{P_{k-1},j,k}\left(-\sum_{r=1}^{k}\hat{u}_{P_{k-1},i,r}\hat{u}_{P_{k-1},j,r}\right)+O_p\left(\frac{1}{\theta_1\theta_2}\right)\right| \\
&= \left|\sum_{i,j\geqslant k} \frac{\hat{\lambda}_{P_{k-1},i}\hat{\lambda}_{P_{k-1},j}}{\left(\hat{\theta}_{P_k,1}-\hat{\lambda}_{P_{k-1},i}\right)\left(\hat{\theta}_{P_k,2}-\hat{\lambda}_{P_{k-1},j}\right)}\hat{u}_{P_{k-1},i,k}\hat{u}_{P_{k-1},j,k}\left(-\sum_{r=1}^{k}\hat{u}_{P_{k-1},i,r}\hat{u}_{P_{k-1},j,r}\right)\right| \\
&= \left|\sum_{r=1}^{k}\sum_{i,j\geqslant k} \frac{\hat{\lambda}_{P_{k-1},i}\hat{\lambda}_{P_{k-1},j}}{\left(\hat{\theta}_{P_k,1}-\hat{\lambda}_{P_{k-1},i}\right)\left(\hat{\theta}_{P_k,2}-\hat{\lambda}_{P_{k-1},j}\right)}\hat{u}_{P_{k-1},i,k}\hat{u}_{P_{k-1},j,k}\left(-\hat{u}_{P_{k-1},i,r}\hat{u}_{P_{k-1},j,r}\right)\right| \\
&\leqslant O_p(1)\times\sum_{r=1}^{k}\frac{1}{\hat{\theta}_{P_k,2}\hat{\theta}_{P_k,1}}\left(\sum_{i\geqslant k}\left|\hat{\lambda}_{P_{k-1},i}\hat{u}_{P_{k-1},i,k}\hat{u}_{P_{k-1},i,r}\right|\right)^2 \\
&\leqslant O_p(1)\times\sum_{r=1}^{k}\frac{1}{\hat{\theta}_{P_k,2}\hat{\theta}_{P_k,1}}\left(\sum_{i\geqslant k}\hat{\lambda}^2_{P_{k-1},i}\hat{u}^2_{P_{k-1},i,k}\right)\left(\sum_{i\geqslant k}\hat{u}^2_{P_{k-1},i,r}\right) \\
&\leqslant O_p(1)\times\sum_{r=1}^{k}\frac{\lambda^2_{\max}}{\hat{\theta}_{P_k,2}\hat{\theta}_{P_k,1}}\left(\sum_{i\geqslant k}\hat{u}^2_{P_{k-1},i,k}\right)\left(\sum_{i\geqslant k}\hat{u}^2_{P_{k-1},i,r}\right) \\
&= O_p\left(\frac{1}{\theta_1\theta_2}\right).
\end{aligned}$$

2.2) $i = 2, ..., k-1,\ j \geqslant k$:

2.2.1) $r = 1, ..., k-1$:

$$\begin{aligned}
&\left|\sum_{j\geqslant k} \frac{\hat{\lambda}_{P_{k-1},i}\hat{\lambda}_{P_{k-1},j}}{\left(\hat{\theta}_{P_k,1}-\hat{\lambda}_{P_{k-1},i}\right)\left(\hat{\theta}_{P_k,2}-\hat{\lambda}_{P_{k-1},j}\right)}\hat{u}_{P_{k-1},i,k}\hat{u}_{P_{k-1},j,k}\left(-\sum_{r=2}^{k-1}\hat{u}_{P_{k-1},i,r}\hat{u}_{P_{k-1},j,r}\right)\right| \\
&\leqslant \sum_{j\geqslant k} \frac{\hat{\lambda}_{P_{k-1},i}\hat{\lambda}_{P_{k-1},j}}{\left(\hat{\theta}_{P_k,1}-\hat{\lambda}_{P_{k-1},i}\right)\left(\hat{\theta}_{P_k,2}-\hat{\lambda}_{P_{k-1},j}\right)}\left|\hat{u}_{P_{k-1},i,k}\right|\left|\hat{u}_{P_{k-1},j,k}\right|\left(\sum_{r=2}^{k-1}\left|\hat{u}_{P_{k-1},i,r}\right|\left|\hat{u}_{P_{k-1},j,r}\right|\right) \\
&\leqslant O_p\left(\frac{1}{\theta_1\theta_2}\right)\sum_{r=2}^{k-1}\hat{\lambda}_{P_{k-1},i}\underbrace{\left|\hat{u}_{P_{k-1},i,k}\hat{u}_{P_{k-1},i,r}\right|}_{O_p\left(\frac{1}{m^{1/2}}\right)}\underbrace{\sum_{j=k}^{m}\hat{\lambda}_{P_{k-1},j}\left|\hat{u}_{P_{k-1},j,k}\hat{u}_{P_{k-1},j,r}\right|}_{O_p\left(m^{1/2}\right)} \\
&= O_p\left(\frac{1}{\theta_1\theta_2}\right).
\end{aligned}$$

The size could be improved; however, this estimation is enough to justify neglecting the term.

2.2.2) $r = k$:

$$\begin{aligned}
&\left|\sum_{j\geqslant k} \frac{\hat{\lambda}_{P_{k-1},i}\hat{\lambda}_{P_{k-1},j}}{\left(\hat{\theta}_{P_k,1}-\hat{\lambda}_{P_{k-1},i}\right)\left(\hat{\theta}_{P_k,2}-\hat{\lambda}_{P_{k-1},j}\right)}\hat{u}_{P_{k-1},i,k}\hat{u}_{P_{k-1},j,k}\left(-\hat{u}_{P_{k-1},i,k}\hat{u}_{P_{k-1},j,k}\right)\right| \\
&\leqslant O_p\left(\frac{1}{\theta_1\theta_2}\right)\hat{\lambda}_{P_{k-1},i}\underbrace{\hat{u}^2_{P_{k-1},i,k}}_{O_p\left(\frac{1}{\theta_i m}\right)}\underbrace{\sum_{j=k}^{m}\hat{\lambda}_{P_{k-1},j}\hat{u}^2_{P_{k-1},j,k}}_{O_p(1)} \\
&= O_p\left(\frac{1}{\theta_1\theta_2 m}\right).
\end{aligned}$$

2.3) $i = 1,\ j \geqslant k$:

2.3.1) $r = 2, 3, 4, ..., k-1$:

$$\left|\sum_{j\geqslant k} \frac{\hat{\theta}_{P_{k-1},1}\hat{\lambda}_{P_{k-1},j}}{\left(\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},1}\right)\left(\hat{\theta}_{P_k,2}-\hat{\lambda}_{P_{k-1},j}\right)}\hat{u}_{P_{k-1},1,k}\hat{u}_{P_{k-1},j,k}\left(-\sum_{r=2}^{k-1}\hat{u}_{P_{k-1},1,r}\hat{u}_{P_{k-1},j,r}\right)\right|$$

$$= O_p\left(\frac{\theta_1 m}{\theta_2\min(\theta_1,\theta_k)}\right)\sum_{r=2}^{k-1}\underbrace{\left|\hat{u}_{P_{k-1},1,k}\hat{u}_{P_{k-1},1,r}\right|}_{O_p\left(\frac{\min(\theta_1,\theta_r)^{1/2}}{\theta_1^{1/2}\max(\theta_1,\theta_r)^{1/2}m}\right)}\sum_{j=k}^{m}\hat{\lambda}_{P_{k-1},j}\left|\hat{u}_{P_{k-1},j,k}\hat{u}_{P_{k-1},j,r}\right|$$

$$\leqslant \max_{r=2,\ldots,k-1}\left(O_p\left(\frac{\theta_1^{1/2}\min(\theta_1,\theta_r)^{1/2}}{\theta_2\max(\theta_1,\theta_r)^{1/2}\min(\theta_1,\theta_k)}\right)\left(\sum_{j=k}^{m}\hat{\lambda}^2_{P_{k-1},j}\hat{u}^2_{P_{k-1},j,k}\right)^{1/2}\left(\sum_{j=k}^{m}\hat{u}^2_{P_{k-1},j,r}\right)^{1/2}\right)$$

$$\leqslant O_p\left(\frac{1}{\theta_1\theta_2}\right).$$

2.3.2) $r = k$:

$$\left|\sum_{j\geqslant k} \frac{\hat{\theta}_{P_{k-1},1}\hat{\lambda}_{P_{k-1},j}}{\left(\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},1}\right)\left(\hat{\theta}_{P_k,2}-\hat{\lambda}_{P_{k-1},j}\right)}\hat{u}_{P_{k-1},1,k}\hat{u}_{P_{k-1},j,k}\left(-\hat{u}_{P_{k-1},1,k}\hat{u}_{P_{k-1},j,k}\right)\right|$$

$$= O_p\left(\frac{1}{\theta_2\min(\theta_1,\theta_k)}\right).$$

2.3.3) $r = 1$: We use the Theorem 5.11.1 for eigenvectors part (b) and (h).

$$\left|\sum_{j\geqslant k} \frac{\hat{\theta}_{P_{k-1},1}\hat{\lambda}_{P_{k-1},j}}{\left(\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},1}\right)\left(\hat{\theta}_{P_k,2}-\hat{\lambda}_{P_{k-1},j}\right)}\hat{u}_{P_{k-1},1,k}\hat{u}_{P_{k-1},j,k}\left(-\hat{u}_{P_{k-1},1,1}\hat{u}_{P_{k-1},j,1}\right)\right|$$

$$= \underbrace{\left|\frac{\hat{\theta}_{P_{k-1},1}}{\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},1}}\hat{u}_{P_{k-1},1,k}\hat{u}_{P_{k-1},1,1}\right|}_{O_p\left(\frac{\theta_1^{1/2}m^{1/2}}{\min(\theta_1,\theta_k)}\right)}\underbrace{\left|\sum_{j\geqslant k}\frac{\hat{\lambda}_{P_{k-1},j}}{\hat{\theta}_{P_k,2}-\hat{\lambda}_{P_{k-1},j}}\hat{u}_{P_{k-1},j,k}\hat{u}_{P_{k-1},j,1}\right|}_{O_p\left(\frac{1}{\theta_2\theta_1^{1/2}m^{1/2}}\right)}$$

$$= O_p\left(\frac{1}{\theta_2\min(\theta_1,\theta_k)}\right).$$

2.4) $j < k,\ i \geqslant k$: As in 2.2 and 2.3, we can show that this part is $O_p\left(\frac{1}{\theta_1\min(\theta_2,\theta_k)}\right)$.

2.5) $i, j < k$

2.5.1) $i, j < k,\ i \neq 1,\ j \neq 2$:

$$\underbrace{\frac{\hat{\theta}_{P_{k-1},i}\hat{\theta}_{P_{k-1},j}}{\left(\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},i}\right)\left(\hat{\theta}_{P_k,2}-\hat{\theta}_{P_{k-1},j}\right)}\hat{u}_{P_{k-1},i,k}\hat{u}_{P_{k-1},j,k}}_{O_p\left(\frac{\theta_i^{1/2}\theta_j^{1/2}}{\max(\theta_1,\theta_i)\max(\theta_2,\theta_j)m}\right)}\quad\underbrace{\left(-\sum_{r=1}^{k}\hat{u}_{P_{k-1},i,r}\hat{u}_{P_{k-1},j,r}\right)}_{O_p\left(\frac{1}{\theta_i^{1/2}\theta_j^{1/2}m^{1/2}}\right)\text{ (by induction on } k-1)}$$

$$= O_p\left(\frac{1}{\theta_1\theta_2 m^{3/2}}\right).$$

2.5.2) $i = 1,\ j = 3, 4, ..., k-1$:

$$\underbrace{\frac{\hat{\theta}_{P_{k-1},1}\hat{\theta}_{P_{k-1},j}}{\left(\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},1}\right)\left(\hat{\theta}_{P_k,2}-\hat{\theta}_{P_{k-1},j}\right)}\hat{u}_{P_{k-1},1,k}\hat{u}_{P_{k-1},j,k}}_{O_p\left(\frac{\theta_1^{1/2}\theta_j^{1/2}}{\min(\theta_1,\theta_k)\max(\theta_2,\theta_j)}\right)}\quad\underbrace{\left(-\sum_{r=1}^{k}\hat{u}_{P_{k-1},1,r}\hat{u}_{P_{k-1},j,r}\right)}_{O_p\left(\frac{1}{\theta_1^{1/2}\theta_j^{1/2}m^{1/2}}\right)\text{ (by induction on } k-1)}$$

$$= O_p\left(\frac{1}{\theta_2\min(\theta_1,\theta_k)m^{1/2}}\right).$$

2.5.3) $j = 2$, $i = 3, 4, ..., k-1$: By similar simplifications as 2.5.2,

$$\frac{\hat{\theta}_{P_{k-1},i}\hat{\theta}_{P_{k-1},2}}{\left(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},i}\right)\left(\hat{\theta}_{P_k,2} - \hat{\theta}_{P_{k-1},2}\right)}\hat{u}_{P_{k-1},i,k}\hat{u}_{P_{k-1},2,k}\left(-\sum_{r=1}^{k}\hat{u}_{P_{k-1},i,r}\hat{u}_{P_{k-1},2,r}\right)$$
$$= O_p\left(\frac{1}{\theta_1\min(\theta_2,\theta_k)m^{1/2}}\right).$$

2.5.4) $i = 1$, $j = 2$:

$$\underbrace{\frac{\hat{\theta}_{P_{k-1},1}\hat{\theta}_{P_{k-1},2}}{\left(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1}\right)\left(\hat{\theta}_{P_k,2} - \hat{\theta}_{P_{k-1},2}\right)}\hat{u}_{P_{k-1},1,k}\hat{u}_{P_{k-1},2,k}}_{\sim\left(\frac{\theta_1^{1/2}\theta_2^{1/2}m}{\min(\theta_1,\theta_k)\min(\theta_2,\theta_k)}\right)}\quad\underbrace{\left(-\sum_{r=1}^{k}\hat{u}_{P_{k-1},1,r}\hat{u}_{P_{k-1},2,r}\right)}_{O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{1/2}m^{1/2}}\right)\text{ (by induction on } k-1)}$$
$$= \mathbf{O_p}\left(\frac{\mathbf{m^{1/2}}}{\min(\boldsymbol{\theta_1},\boldsymbol{\theta_k})\min(\boldsymbol{\theta_2},\boldsymbol{\theta_k})}\right).$$

This term is non-negligible and its estimation is presented in **C**.

Therefore,

$$\sum_{s=k+1}^{m}\hat{u}_{P_k,1,s}\hat{u}_{P_k,2,s} =$$
$$\frac{1}{\sqrt{D_1D_2}N_1N_2}\left(\frac{\hat{\theta}_{P_{k-1},1}\hat{\theta}_{P_{k-1},2}}{\left(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1}\right)\left(\hat{\theta}_{P_k,2} - \hat{\theta}_{P_{k-1},2}\right)}\hat{u}_{P_{k-1},1,k}\hat{u}_{P_{k-1},2,k}\left(-\sum_{r=1}^{k}\hat{u}_{P_{k-1},1,r}\hat{u}_{P_{k-1},2,r}\right)\right)$$
$$+O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{1/2}m}\right).$$

B: In this paragraph we study $\frac{1}{\sqrt{D_1D_2}N_1N_2}$.

$$\begin{aligned}
D_1 &= \underbrace{\sum_{i=k}^{m}\frac{\hat{\lambda}^2_{P_{k-1},i}}{(\hat{\theta}_{P_k,1} - \hat{\lambda}_{P_{k-1},i})^2}\hat{u}^2_{P_{k-1},i,k}}_{O_p\left(\frac{1}{\theta_1^2}\right)} + \underbrace{\frac{\hat{\theta}^2_{P_{k-1},1}}{(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1})^2}\hat{u}^2_{P_{k-1},1,k}}_{\sim\frac{\theta_1 m}{\min(\theta_1,\theta_k)^2}} + \underbrace{\sum_{i=2}^{k-1}\frac{\hat{\theta}^2_{P_{k-1},i}}{(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},i})^2}\hat{u}^2_{P_{k-1},i,k}}_{O_p\left(\frac{1}{\theta_1 m}\right)}\\
&= \frac{\hat{\theta}^2_{P_{k-1},1}}{(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1})^2}\hat{u}^2_{P_{k-1},1,k} + O_p\left(\frac{1}{\theta_1 m}\right) + O_p\left(\frac{1}{\theta_1^2}\right)\\
&= O_p\left(\frac{\theta_1 m}{\min(\theta_1,\theta_k)^2}\right).
\end{aligned}$$

Because $\hat{u}_{P_k,t,s} = \frac{P_k^{1/2}\tilde{u}_{P_k,t}}{N_t}$, then

$$\begin{aligned}
N_1^2 &= \sum_{i\neq k}^{m}\tilde{u}^2_{P_k,1,i} + \tilde{u}^2_{P_k,1,k}\theta_k\\
&= 1 + (\theta_k - 1)\tilde{u}^2_{P_2,1,2}\\
&= 1 + \frac{1}{(\theta_k - 1)D_1}\\
&= 1 + O_p\left(\frac{\min(\theta_1,\theta_k)}{\max(\theta_1,\theta_k)m}\right).
\end{aligned}$$

We easily obtain

$$\begin{aligned}\frac{1}{N_1\sqrt{D_1}} &= \frac{|\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},1}|}{\hat{\theta}_{P_{k-1},1}|\hat{u}_{P_{k-1},1,k}|} + O_p\left(\frac{\min(\theta_1,\theta_k)}{\theta_1^{1/2}m^{3/2}}\right)\\ &= O_p\left(\frac{\min(\theta_1,\theta_k)}{\theta_1^{1/2}m^{1/2}}\right)\end{aligned}$$

and

$$\begin{aligned}\frac{1}{N_1N_2\sqrt{D_1D_2}} &= \frac{|\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},1}|}{\hat{\theta}_{P_{k-1},1}|\hat{u}_{P_{k-1},1,k}|}\frac{|\hat{\theta}_{P_k,2}-\hat{\theta}_{P_{k-1},2}|}{\hat{\theta}_{P_{k-1},2}|\hat{u}_{P_{k-1},2,k}|} + O_p\left(\frac{\min(\theta_1,\theta_k)}{\theta_1^{1/2}\theta_2^{1/2}m^{2}}\right)\\ &= O_p\left(\frac{\min(\theta_1,\theta_k)\min(\theta_2,\theta_k)}{\theta_1^{1/2}\theta_2^{1/2}m}\right).\end{aligned}$$

C: From **A** and **B**, we conclude using Theorem 5.4.1,

$$\begin{aligned}&\sum_{s=k+1}^{m}\hat{u}_{P_k,1,s}\hat{u}_{P_k,2,s}\\ &= \frac{\frac{\hat{\theta}_{P_{k-1},1}\hat{\theta}_{P_{k-1},2}}{\left(\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},1}\right)\left(\hat{\theta}_{P_k,2}-\hat{\theta}_{P_{k-1},2}\right)}\hat{u}_{P_{k-1},1,k}\hat{u}_{P_{k-1},2,k}\left(-\sum_{r=1}^{k}\hat{u}_{P_{k-1},1,r}\hat{u}_{P_{k-1},2,r}\right)}{\sqrt{D_1D_2}N_1N_2} + O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{1/2}m}\right)\\ &= \mathrm{sign}\left(\hat{u}_{P_{k-1},1,1}\hat{u}_{P_{k-1},2,2}\hat{u}_{P_k,1,1}\hat{u}_{P_k,2,2}\right)\left(-\sum_{r=1}^{k}\hat{u}_{P_{k-1},1,r}\hat{u}_{P_{k-1},2,r}\right) + O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{1/2}m}\right)\\ &= \mathrm{sign}\left(\hat{u}_{P_{k-1},1,1}\hat{u}_{P_{k-1},2,2}\hat{u}_{P_k,1,1}\hat{u}_{P_k,2,2}\right)\sum_{s=k+1}^{m}\hat{u}_{P_{k-1},1,s}\hat{u}_{P_{k-1},2,s} + O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{1/2}m}\right).\end{aligned}$$

Using the remark of Theorem 5.11.1, the sign of the third line is correct with a probability tending to 1 in $1/m$. Therefore, using the convention $\hat{u}_{P_s,i,i} > 0$ for $i = 1,2,...,s$ and $s = 1,2,...,k$ leads to

$$\begin{aligned}\sum_{s=k+1}^{m}\hat{u}_{P_k,1,s}\hat{u}_{P_k,2,s} &= \sum_{s=k+1}^{m}\hat{u}_{P_{k-1},1,s}\hat{u}_{P_{k-1},2,s} + O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{1/2}m}\right)\\ &= \sum_{s=2}^{m}\hat{u}_{P_2,1,s}\hat{u}_{P_2,2,s} + O_p\left(\frac{1}{\theta_1^{1/2}\theta_2^{1/2}m}\right),\end{aligned}$$

where we remember that the error O is in probability.

□

7.1.6 Invariant Double Angle Theorem

Corollary 5.10.1.
Suppose W_X and W_Y satisfies Assumption 2.2.1 and

$$\tilde{P}_s = \mathrm{I}_m + (\theta_s - 1)e_se_s^t,\ \textit{for } s = 1,2,...,k,$$
$$P_k = \mathrm{I}_m + \sum_{i=1}^{k}(\theta_i - 1)e_ie_i^t \textit{ respects 2.2.2 (A4)},$$

where $\theta_1 > \theta_2 > ... > \theta_k$. We define

$$\hat{\Sigma}_{X,\tilde{P}_s} = \tilde{P}_s^{1/2}W_X\tilde{P}_s^{1/2} \textit{ and } \hat{\Sigma}_{X,\tilde{P}_s} = \tilde{P}_s^{1/2}W_Y\tilde{P}_s^{1/2},$$
$$\hat{\Sigma}_{X,P_k} = P_k^{1/2}W_XP_k^{1/2} \textit{ and } \hat{\Sigma}_{Y,P_k} = P_k^{1/2}W_YP_k^{1/2}.$$

Moreover, for $s = 1, ..., k$, we define

$$\begin{array}{lll} \hat{u}_{\hat{\Sigma}_{X,\tilde{P}_s},1}, \hat{\theta}_{\hat{\Sigma}_{X,\tilde{P}_s},1} & s.t. & \hat{\Sigma}_{X,\tilde{P}_s} \hat{u}_{\hat{\Sigma}_{X,\tilde{P}_s},1} = \hat{\theta}_{\hat{\Sigma}_{X,\tilde{P}_s},1} \hat{u}_{\hat{\Sigma}_{X,\tilde{P}_s},1}, \\ \hat{u}_{\hat{\Sigma}_{X,P_k},s}, \hat{\theta}_{\hat{\Sigma}_{X,P_k},s} & s.t. & \hat{\Sigma}_{X,P_k} \hat{u}_{\hat{\Sigma}_{X,P_k},s} = \hat{\theta}_{\hat{\Sigma}_{X,P_k},s} \hat{u}_{\hat{\Sigma}_{X,P_k},s}, \end{array}$$

where $\hat{\theta}_{\hat{\Sigma}_{X,\tilde{P}_s},1} = \hat{\lambda}_{\hat{\Sigma}_{X,\tilde{P}_s,1}}$ and $\hat{\theta}_{\hat{\Sigma}_{X,P_k},s} = \hat{\lambda}_{\hat{\Sigma}_{X,P_k},s}$. The statistics of the group Y are defined in analogous manner.
Then,

$$\begin{aligned} \boxed{\left\langle \hat{u}_{\hat{\Sigma}_{X,\tilde{P}_s},1}, \hat{u}_{\hat{\Sigma}_{Y,\tilde{P}_s},1} \right\rangle^2 = \sum_{i=1}^{k} \left\langle \hat{u}_{\hat{\Sigma}_{X,P_k},s}, \hat{u}_{\hat{\Sigma}_{Y,P_k},s} \right\rangle^2 + O_p\left(\frac{1}{\theta_s m}\right)} \\ = \sum_{i=1}^{k+\epsilon} \left\langle \hat{u}_{\hat{\Sigma}_{X,P_k},s}, \hat{u}_{\hat{\Sigma}_{Y,P_k},i} \right\rangle^2 + O_p\left(\frac{1}{\theta_s m}\right), \end{aligned}$$

where ϵ is a small integer.

Remark 7.1.0.1.

1. The procedure of the proof shows an interesting invariant:
   Assuming the sign convention $\hat{u}_{P_s,i,i} > 0$ for $s = 1, 2, ..., k$ and $i = 1, 2, ..., s$,
   $$\sum_{i=k+1}^{m} \hat{u}_{P_k,1,i} \hat{\hat{u}}_{P_k,1,i} = \sum_{i=k}^{m} \hat{u}_{P_{k-1},1,i} \hat{\hat{u}}_{P_{k-1},1,i} + O_p\left(\frac{1}{\theta_1 m}\right).$$
2. The distribution of $\left\langle \hat{u}_{\hat{\Sigma}_{X,P_1},1}, \hat{u}_{\hat{\Sigma}_{Y,P_1},1} \right\rangle^2$ is computed in Theorem 5.3.1.
3. An error of ϵ principal components does not affect the asymptotic distribution of the general double angle. This property allows us to construct a robust test.

(Page 83)

Proof. **Corollary 5.10.1**

We change the notation in order to obtain shorter equations

$$\begin{array}{lll} \hat{\theta}_{P_s,t} = \hat{\theta}_{\hat{\Sigma}_{X,P_s},t}, & \hat{u}_{P_s,t} = \hat{u}_{\hat{\Sigma}_{X,P_s},t}, & \hat{\lambda}_{P_s,t} = \hat{\lambda}_{\hat{\Sigma}_{X,P_s},t} \\ \hat{\hat{\theta}}_{P_s,t} = \hat{\theta}_{\hat{\Sigma}_{Y,P_s},t}, & \hat{\hat{u}}_{P_s,t} = \hat{u}_{\hat{\Sigma}_{Y,P_s},t}, & \hat{\hat{\lambda}}_{P_s,t} = \hat{\lambda}_{\hat{\Sigma}_{Y,P_s},t}. \end{array}$$

Moreover, we use a specific notation for this proof,

$$\begin{array}{rcl} u^{c_s} & = & \frac{u_{1:s}}{||u_{1:s}||}, \text{ where } u \text{ is a vector of size } m, \\ \hat{\alpha}^2_{P_s,i} & = & ||\hat{u}_{P_s,i,1:s}||^2, \\ \hat{\hat{\alpha}}^2_{P_s,i} & = & ||\hat{\hat{u}}_{P_s,i,1:s}||^2. \end{array}$$

Finally, using the notation 7.1.1 and relaxing $\theta_1 > \theta_2 > ... > \theta_k$ allow us to study only $\hat{u}_{P_k,1}$ and $\hat{\hat{u}}_{P_k,1}$ without loss of generality.
The proof is essentially based on Theorem 5.3.1,5.6.1, 5.7.1 and 5.5.1.

1. First we study $\left\langle \hat{u}_{P_1,1}, \hat{\hat{u}}_{P_1,1} \right\rangle^2$:

$$\begin{aligned}\left\langle \hat{u}_{P_1,1}, \hat{\hat{u}}_{P_1,1} \right\rangle^2 &= \hat{u}^2_{P_1,1,1}\hat{\hat{u}}^2_{P_1,1,1} + 2\hat{u}_{P_1,1,1}\hat{\hat{u}}_{P_1,1,1}\sum_{i=2}^{m}\hat{u}_{P_1,1,i}\hat{\hat{u}}_{P_1,1,i} + \left(\sum_{i=1}^{m}\hat{u}_{P_1,1,i}\hat{\hat{u}}_{P_1,1,i}\right)^2 \\ &= \underbrace{\hat{u}^2_{P_1,1,1}\hat{\hat{u}}^2_{P_1,1,1}}_{\mathrm{RV}\left(O_p(1), O_p\left(\frac{1}{\theta_1^2 m}\right)\right)} + \underbrace{C_{P_1}}_{\mathrm{RV}\left(0, O_p\left(\frac{1}{\theta_1^2 m}\right)\right)} + O_p\left(\frac{1}{\theta_1^2 m}\right).\end{aligned}$$

2. Now we want to prove

$$\left\langle \hat{u}_{P_1,1}, \hat{\hat{u}}_{P_1,1} \right\rangle^2 = \sum_{i=1}^{k}\left\langle \hat{u}_{P_k,1}, \hat{\hat{u}}_{P_k,i} \right\rangle^2 + O_p\left(\frac{1}{\theta_1 m}\right).$$

Using Theorem 5.8.1 and 5.3.1,

$$\left\langle \hat{u}_{P_k,1}, \hat{\hat{u}}_{P_k,1} \right\rangle^2 = \left\langle \hat{u}_{P_k,1,1:k}, \hat{\hat{u}}_{P_k,1,1:k} \right\rangle^2 + \underbrace{2\hat{u}_{P_k,1,1}\hat{\hat{u}}_{P_k,1,1}\sum_{i=k+1}^{m}\hat{u}_{P_k,1,i}\hat{\hat{u}}_{P_k,1,i}}_{C_{P_k}} + O_p\left(\frac{1}{\theta_1 m}\right),$$

$$\left\langle \hat{u}_{P_k,1}, \hat{\hat{u}}_{P_k,s} \right\rangle^2 = \left\langle \hat{u}_{P_k,1,1:k}, \hat{\hat{u}}_{P_k,s,1:k} \right\rangle^2 + O_p\left(\frac{1}{\max(\theta_1,\theta_s)m}\right).$$

In this theorem we assume Assumption 2.2.2 (A4) and without loss of generality, we can assume $\theta_1,...,\theta_{k_1}$ are of same order and $\theta_{k_1+1},...,\theta_k$ are also of same order but different from the first group. Moreover, no assumptions are made between the groups except Assumption 2.2.2 (A4). This means that either all the eigenvalues are proportional or one group has finite eigenvalues. Therefore,

$$\sum_{i=1}^{k}\left\langle \hat{u}_{P_k,1}, \hat{\hat{u}}_{P_k,i} \right\rangle^2 = \sum_{i=1}^{k_1}\left\langle \hat{u}_{P_k,1}, \hat{\hat{u}}_{P_k,i} \right\rangle^2 + O_p\left(\frac{1}{\theta_1 m}\right).$$

Moreover, we easily see that for $i = 1, 2, ..., k_1$,

$$\begin{aligned}\hat{\hat{\alpha}}^2_{P_k,i} &= ||\hat{\hat{u}}_{P_s,i,1:k}||^2 \\ &= ||\hat{\hat{u}}_{P_s,i,1:k_1}||^2 + O_p\left(\frac{1}{\theta_1 m}\right).\end{aligned}$$

Thus

$$\begin{aligned}\sum_{i=1}^{k_1}\left\langle \hat{u}_{P_k,1}, \hat{\hat{u}}_{P_k,i} \right\rangle^2 &= \sum_{i=1}^{k_1}\left\langle \hat{u}_{P_k,1,1:k}, \hat{\hat{u}}_{P_k,i,1:k} \right\rangle^2 + C_k + O_p\left(\frac{1}{\theta_1 m}\right) \\ &= \sum_{i=1}^{k_1}\left\langle \hat{u}_{P_k,1,1:k_1}, \hat{\hat{u}}_{P_k,i,1:k_1} \right\rangle^2 + C_k + O_p\left(\frac{1}{\theta_1 m}\right) \\ &= \sum_{i=1}^{k_1}\hat{\alpha}^2_{P_k,1}\hat{\hat{\alpha}}^2_{P_k,i}\left\langle \hat{u}^{c_{k_1}}_{P_k,1}, \hat{\hat{u}}^{c_{k_1}}_{P_k,i} \right\rangle^2 + C_k + O_p\left(\frac{1}{\theta_1 m}\right) \\ &= \hat{\alpha}^2_{P_k,1}\hat{\hat{\alpha}}^2_{P_k,1}\sum_{i=1}^{k_1}\left\langle \hat{u}^{c_{k_1}}_{P_k,1}, \hat{\hat{u}}^{c_{k_1}}_{P_k,i} \right\rangle^2 + C_k + O_p\left(\frac{1}{\theta_1 m}\right) \\ &\quad + \hat{\alpha}^2_{P_k,1}\sum_{i=2}^{k_1}\left(\hat{\hat{\alpha}}^2_{P_k,i} - \hat{\hat{\alpha}}^2_{P_k,1}\right)\left\langle \hat{u}^{c_{k_1}}_{P_k,1}, \hat{\hat{u}}^{c_{k_1}}_{P_k,i} \right\rangle^2 \\ &= \hat{\alpha}^2_{P_1,1}\hat{\hat{\alpha}}^2_{P_1,1}\underbrace{\sum_{i=1}^{k_1}\left\langle \hat{u}^{c_{k_1}}_{P_k,1}, \hat{\hat{u}}^{c_{k_1}}_{P_k,i} \right\rangle^2}_{\text{Part 1}} + \underbrace{C_{P_k}}_{\text{Part 2}} + O_p\left(\frac{1}{\theta_1 m}\right),\end{aligned}$$

Where the last equality is obtained because for $i = 1, 2, ..., k_1$, $\hat{\hat{\alpha}}^2_{P_k,i} - \hat{\hat{\alpha}}^2_{P_k,1} = O_p\left(1/\theta_1\right)$. Therefore, we just need to show that

$$\sum_{i=1}^{k_1} \left\langle \hat{u}^{c_{k_1}}_{P_k,1}, \hat{\hat{u}}^{c_{k_1}}_{P_k,i} \right\rangle^2 = 1 + O_p\left(\frac{1}{\theta_1 m}\right),$$
$$C_{P_k} = C_{P_1} + O_p\left(\frac{1}{\theta_1 m}\right).$$

Part 1 : First we prove that

$$\sum_{i=1}^{k_1} \left\langle \hat{u}^{c_{k_1}}_{P_k,i}, \hat{\hat{u}}^{c_{k_1}}_{P_k,i} \right\rangle^2 = 1 + O_p\left(\frac{1}{\theta_1 m}\right).$$

We apply Gramm-Schmidt to $\hat{\hat{u}}^{c_{k_1}}_{P_k,1}, \hat{\hat{u}}^{c_{k_1}}_{P_k,2}, ..., \hat{\hat{u}}^{c_{k_1}}_{P_k,k_1}$,

$$\begin{aligned}
\hat{\hat{w}}_{P_k,1} &= \hat{\hat{u}}^{c_{k_1}}_{P_k,1}. \\
\hat{\hat{w}}_{P_k,2} &= \left(\hat{\hat{u}}^{c_{k_1}}_{P_k,2} - \left\langle \hat{\hat{u}}^{c_{k_1}}_{P_k,2}, \hat{\hat{w}}_{P_k,1} \right\rangle \hat{\hat{w}}_{P_k,1}\right)\left(1 + O_p\left(\frac{1}{\theta_1^2 m}\right)\right).
\end{aligned}$$

Indeed by Theorems 5.6.1 and 5.7.1,

$$||\hat{\hat{u}}^{c_{k_1}}_{P_k,2} - \left\langle \hat{\hat{u}}^{c_{k_1}}_{P_k,2}, \hat{\hat{w}}_{P_k,1} \right\rangle \hat{\hat{w}}_{P_k,1}|| = 1 - \underbrace{\left\langle \hat{\hat{u}}^{c_{k_1}}_{P_k,1}, \hat{\hat{u}}^{c_{k_1}}_{P_k,2} \right\rangle^2}_{-\alpha_{P_k,1}\alpha_{P_k,2}\sum_{i=k+1}^{m}\hat{u}_{P_k,1,i}\hat{u}_{P_k,2,i}} = 1 + O_p\left(\frac{1}{\theta_1^2 m}\right).$$

$$\hat{\hat{w}}_{P_k,p} = \left(\hat{\hat{u}}^{c_{k_1}}_{P_k,p} - \sum_{i=1}^{p-1} \left\langle \hat{\hat{u}}^{c_{k_1}}_{P_k,p}, \hat{\hat{w}}_{P_k,i} \right\rangle \hat{\hat{w}}_{P_k,i}\right)\left(1 + O_p\left(\frac{1}{\theta_1^2 m}\right)\right).$$

However, the norm is more difficult to estimate for $p = 3, 4, ..., k_1$:

$$\begin{aligned}
||\hat{\hat{u}}^{c_{k_1}}_{P_k,p} - \sum_{i=1}^{p-1} \left\langle \hat{\hat{u}}^{c_{k_1}}_{P_k,p}, \hat{\hat{w}}_{P_k,i} \right\rangle \hat{\hat{w}}_{P_k,i}|| &= 1 - \sum_{i=1}^{p-1} \left\langle \hat{\hat{u}}^{c_{k_1}}_{P_k,p}, \hat{\hat{w}}_{P_k,i} \right\rangle^2 \\
&= 1 - \sum_{i=1}^{p-1} \left\langle \hat{\hat{u}}^{c_{k_1}}_{P_k,p}, \sum_{j=1}^{i} a_j \hat{\hat{u}}^{c_{k_1}}_{P_k,j} \right\rangle^2, \text{ for some } |a_i| < 1, \\
&= 1 - \sum_{i=1}^{p-1} \sum_{j_1,j_2=1}^{i} \left\langle \hat{\hat{u}}^{c_{k_1}}_{P_k,p}, a_{j_1}\hat{\hat{u}}^{c_{k_1}}_{P_k,j_1} \right\rangle \left\langle \hat{\hat{u}}^{c_{k_1}}_{P_k,p}, a_{j_2}\hat{\hat{u}}^{c_{k_1}}_{P_k,j_2} \right\rangle \\
&= 1 + O_p\left(\frac{1}{\theta_1^2 m}\right).
\end{aligned}$$

Therefore, we can express the truncated eigenvectors in a orthonormal basis,

$$\Rightarrow \quad \hat{\hat{u}}^{c_k}_{P_k,1} = \hat{\hat{w}}_{P_k,1},$$

For $p = 2, ..., k_1$,

$$\hat{\hat{u}}^{c_{k_1}}_{P_k,p} = \left(\hat{\hat{w}}_{P_k,p} + \sum_{i=1}^{p-1} \left\langle \hat{\hat{u}}^{c_{k_1}}_{P_k,p}, \hat{\hat{w}}_{P_k,i} \right\rangle \hat{\hat{w}}_{P_k,i}\right)\left(1 + O_p\left(\frac{1}{\theta_1^2 m}\right)\right).$$

Thus

$$
\begin{aligned}
&\sum_{p=1}^{k_1}\left\langle \hat{u}^{c_{k_1}}_{P_k,1}, \hat{\hat{u}}^{c_{k_1}}_{P_k,p}\right\rangle^2 \\
&= \left\langle \hat{u}^{c_k}_{P_k,1}, \hat{\hat{w}}_{P_k,1}\right\rangle^2 + \sum_{p=2}^{k_1}\left\langle \hat{u}^{c_{k_1}}_{P_k,1}, \hat{\hat{w}}_{P_k,p} + \sum_{j=1}^{p-1}\left\langle \hat{\hat{u}}^{c_{k_1}}_{P_k,p}, \hat{\hat{w}}_{P_k,j}\right\rangle \hat{\hat{w}}_{P_k,j}\right\rangle^2 + O_p\left(\frac{1}{\theta_1^2 m}\right) \\
&= \sum_{p=1}^{k_1}\left\langle \hat{u}^{c_{k_1}}_{P_k,1}, \hat{\hat{w}}_{P_k,p}\right\rangle^2 + \left(\sum_{p=2}^{k_1}\left\langle \hat{u}^{c_{k_1}}_{P_k,1}, \sum_{j=1}^{p-1}\left\langle \hat{\hat{u}}^{c_{k_1}}_{P_k,p}, \hat{\hat{w}}_{P_k,j}\right\rangle \hat{\hat{w}}_{P_k,j}\right\rangle^2\right. \\
&\qquad \left. +2\sum_{p=2}^{k_1}\left\langle \hat{u}^{c_{k_1}}_{P_k,1}, \hat{\hat{w}}_{P_k,p}\right\rangle\left\langle \hat{u}^{c_{k_1}}_{P_k,1}, \sum_{j=1}^{p-1}\left\langle \hat{\hat{u}}^{c_{k_1}}_{P_k,p}, \hat{\hat{w}}_{P_k,j}\right\rangle \hat{\hat{w}}_{P_k,j}\right\rangle\right) + O_p\left(\frac{1}{\theta_1^2 m}\right) \\
&= 1 + \sum_{p=2}^{k_1}\left(\sum_{j=1}^{p-1}\left\langle \hat{\hat{u}}^{c_{k_1}}_{P_k,p}, \hat{\hat{w}}_{P_k,j}\right\rangle\left\langle \hat{u}^{c_{k_1}}_{P_k,1}, \hat{\hat{w}}_{P_k,j}\right\rangle\right)^2 \\
&\qquad +2\sum_{p=2}^{k_1}\sum_{j=1}^{p-1}\left\langle \hat{\hat{u}}^{c_{k_1}}_{P_k,p}, \hat{\hat{w}}_{P_k,j}\right\rangle\left\langle \hat{u}^{c_{k_1}}_{P_k,1}, \hat{\hat{w}}_{P_k,p}\right\rangle\left\langle \hat{u}^{c_{k_1}}_{P_k,1}, \hat{\hat{w}}_{P_k,j}\right\rangle + O_p\left(\frac{1}{\theta_1^2 m}\right) \\
&= 1 + A + B + O_p\left(\frac{1}{\theta_1^2 m}\right).
\end{aligned}
$$

Next we prove separately that A and B are negligible.

A : By Theorem 5.6.1, 5.7.1,

$\left\langle \hat{\hat{u}}^{c_{k_1}}_{P_k,p}, \hat{\hat{w}}_{P_k,j}\right\rangle,\ k_1 \geqslant p > j:$

$$
j = 1: \left\langle \hat{\hat{u}}^{c_{k_1}}_{P_k,p}, \hat{\hat{w}}_{P_k,1}\right\rangle = \left\langle \hat{\hat{u}}^{c_{k_1}}_{P_k,p}, \hat{\hat{u}}^{c_{k_1}}_{P_k,1}\right\rangle = O_p\left(\frac{1}{\theta_1\sqrt{m}}\right),
$$

$$
\begin{aligned}
j \neq 1: \left\langle \hat{\hat{u}}^{c_{k_1}}_{P_k,p}, \hat{\hat{w}}_{P_k,j}\right\rangle\left(1 + O_p\left(\frac{1}{\theta_1^2 m}\right)\right) &= \left\langle \hat{\hat{u}}^{c_{k_1}}_{P_k,p}, \hat{\hat{u}}^{c_{k_1}}_{P_k,j}\right\rangle - \sum_{i=1}^{j-1}\underbrace{\left\langle \hat{\hat{u}}^{c_{k_1}}_{P_k,j}, \hat{\hat{w}}_{P_k,i}\right\rangle}_{O_p\left(\frac{1}{\theta_1\sqrt{m}}\right)}\underbrace{\left\langle \hat{\hat{u}}^{c_{k_1}}_{P_k,p}, \hat{\hat{w}}_{P_k,i}\right\rangle}_{O_p\left(\frac{1}{\sqrt{m}}\right)} \\
&= O_p\left(\frac{1}{\theta_1\sqrt{m}}\right).
\end{aligned}
$$

$\left\langle \hat{u}^{c_{k_1}}_{P_k,1}, \hat{\hat{w}}_{P_k,j}\right\rangle, k_1 \geqslant j:$

$$
\begin{aligned}
j \neq 1: \left\langle \hat{u}^{c_{k_1}}_{P_k,1}, \hat{\hat{w}}_{P_k,j}\right\rangle\left(1 + O_p\left(\frac{1}{\theta_1^2 m}\right)\right) &= \left\langle \hat{u}^{c_{k_1}}_{P_k,1}, \hat{\hat{u}}^{c_{k_1}}_{P_k,j}\right\rangle - \sum_{i=1}^{j-1}\underbrace{\left\langle \hat{\hat{w}}_{P_k,i}, \hat{\hat{u}}^{c_{k_1}}_{P_k,j}\right\rangle}_{O_p\left(\frac{1}{\theta_1\sqrt{m}}\right)}\underbrace{\left\langle \hat{u}^{c_{k_1}}_{P_k,1}, \hat{\hat{w}}_{P_k,i}\right\rangle}_{=O_p(1)} \\
&= O_p\left(\frac{1}{\sqrt{m}}\right),
\end{aligned}
$$

$$
j = 1: \left\langle \hat{u}^{c_{k_1}}_{P_k,1}, \hat{\hat{w}}_{P_k,1}\right\rangle = O_p(1).
$$

Consequently,

$$
\left(\sum_{j=1}^{p-1}\left\langle \hat{\hat{u}}^{c_{k_1}}_{P_k,p}, \hat{\hat{w}}_{P_k,j}\right\rangle\left\langle \hat{u}^{c_{k_1}}_{P_k,1}, \hat{\hat{w}}_{P_k,j}\right\rangle\right)^2 = O_p\left(\frac{1}{\theta_1^2 m}\right)
$$

Therefore, $A = O_p\left(\frac{1}{\theta_1^2 m}\right)$.

B : The same estimations as previously lead to

$$B = 2\sum_{p=2}^{k_1}\sum_{j=1}^{p-1} \underbrace{\left\langle \hat{\hat{u}}_{P_k,p}^{c_{k_1}}, \hat{\hat{w}}_{P_k,j}\right\rangle}_{O_p\left(\frac{1}{\theta_1\sqrt{m}}\right)} \underbrace{\left\langle \hat{u}_{P_k,1}^{c_{k_1}}, \hat{\hat{w}}_{P_k,p}\right\rangle}_{O_p\left(\frac{1}{\sqrt{m}}\right)} \underbrace{\left\langle \hat{u}_{P_k,1}^{c_{k_1}}, \hat{\hat{w}}_{P_k,j}\right\rangle}_{O_p(1)} = O_p\left(\frac{1}{\theta_1 m}\right).$$

Therefore,

$$\sum_{i=1}^{k_1}\left\langle \hat{u}_{P_k,i}^{c_{k_1}}, \hat{\hat{u}}_{P_k,i}^{c_{k_1}}\right\rangle^2 = 1 + O_p\left(\frac{1}{\theta_1 m}\right).$$

Part 2 : In this part we prove the invariance of C_{P_1}. We need to show:

$$\begin{aligned} C_{P_k} &= 2\hat{u}_{P_k,1,1}\hat{\hat{u}}_{P_k,1,1}\sum_{i=k+1}^{m}\hat{u}_{P_k,1,i}\hat{\hat{u}}_{P_k,1,i} \\ &= 2\hat{u}_{P_1,1,1}\hat{\hat{u}}_{P_1,1,1}\sum_{i=2}^{m}\hat{u}_{P_1,1,i}\hat{\hat{u}}_{P_1,1,i} + O_p\left(\frac{1}{\theta_1 m}\right) \\ &= C_{P_1} + O_p\left(\frac{1}{\theta_1 m}\right). \end{aligned}$$

In order to prove this result we show $C_{P_k} = C_{P_{k-1}} + O_p\left(\frac{1}{\theta_1 m}\right)$ and more precisely,

$$2\hat{u}_{P_k,1,1}\hat{\hat{u}}_{P_k,1,1}\sum_{i=k+1}^{m}\hat{u}_{P_k,1,i}\hat{\hat{u}}_{P_k,1,i} = 2\hat{u}_{P_{k-1},1,1}\hat{\hat{u}}_{P_{k-1},1,1}\sum_{i=k}^{m}\hat{u}_{P_{k-1},1,i}\hat{\hat{u}}_{P_{k-1},1,i} + O_p\left(\frac{1}{\theta_1 m}\right).$$

The proof is similar to the proofs of invariant eigenvectors structure 5.7.1 and 5.5.1. We use Theorem 5.11.1 in order to estimate each term of the sum. Assuming P_{k-1} respects 2.2.2(A4) the last added eigenvalue can be either proportional to θ_1 or to the other group.
In this proof we do not use the convention of the sign : $\hat{u}_{P_k,i,i} > 0$ for $i = 1, 2, ..., k$.

We start by studying $\hat{u}_{P_k,1}$. As in Theorem 5.11.1, for $s > k$,

$$\begin{aligned} \hat{u}_{P_k,1,s} = \frac{1}{\sqrt{\hat{D}_1}\hat{N}_1}&\left(\sum_{i=k}^{m}\frac{\hat{\lambda}_{P_{k-1},i}}{\hat{\theta}_{P_k,1}-\hat{\lambda}_{P_{k-1},i}}\hat{u}_{P_{k-1},i,s}\hat{u}_{P_{k-1},i,k} + \frac{\hat{\theta}_{P_{k-1},1}}{\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},1}}\hat{u}_{P_{k-1},1,s}\hat{u}_{P_{k-1},1,k}\right. \\ &\left. + \sum_{i=2}^{k_1}\frac{\hat{\theta}_{P_{k-1},i}}{\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},i}}\hat{u}_{P_{k-1},i,s}\hat{u}_{P_{k-1},i,k} + \sum_{i=k_1+1}^{k-1}\frac{\hat{\theta}_{P_{k-1},i}}{\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},i}}\hat{u}_{P_{k-1},i,s}\hat{u}_{P_{k-1},i,k}\right). \end{aligned}$$

By a similar proof as part (a), (b) and (c) of Theorem 5.11.1,

$$\begin{aligned} \hat{A}_s &= \sum_{i=k}^{m}\frac{\hat{\lambda}_{P_{k-1},i}}{\hat{\theta}_{P_k,1}-\hat{\lambda}_{P_{k-1},i}}\hat{u}_{P_{k-1},i,s}\hat{u}_{P_{k-1},i,k} = O_p\left(\frac{1}{\sqrt{m}\theta_1}\right), \\ \hat{B}_s &= \frac{\hat{\theta}_{P_{k-1},1}}{\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},1}}\hat{u}_{P_{k-1},1,s}\hat{u}_{P_{k-1},1,k} \sim \frac{1}{\min(\theta_1,\theta_k)}, \\ \hat{C}_s &= \sum_{i=2}^{k_1}\frac{\hat{\theta}_{P_{k-1},i}}{\hat{\theta}_{P_k,1}-\hat{\theta}_{P_{k-1},i}}\hat{u}_{P_{k-1},i,s}\hat{u}_{P_{k-1},i,k} = O_p\left(\frac{1}{m\theta_1}\right), \end{aligned}$$

$$\hat{C}_s^G = \sum_{i=k_1+1}^{k-1} \frac{\hat{\theta}_{P_{k-1},i}}{\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,s} \hat{u}_{P_{k-1},i,k} = O_p\left(\frac{1}{m\theta_1}\right),$$

$$\hat{D}_1 = \frac{\hat{\theta}^2_{P_{k-1},1}}{\left(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1}\right)^2} \hat{u}^2_{P_{k-1},1,k} + O_p\left(\frac{1}{\theta_1^2}\right) + O_p\left(\frac{1}{\theta_1^2 m}\right),$$

$$\hat{N}_1 = 1 + O_p\left(\frac{\min(\theta_1, \theta_k)}{\max(\theta_1, \theta_k) m}\right).$$

Thus,

$$\hat{u}_{P_k,1,s} \quad = \quad \frac{1}{\sqrt{\hat{D}_1}\hat{N}_1}\left(\hat{A}_s + \hat{B}_s + \hat{C}_s + \hat{C}_s^G\right).$$

Therefore,

$$\sum_{s=k+1}^{m} \hat{u}_{P_k,1,i}\hat{\hat{u}}_{P_k,1,i} \quad = \quad \frac{\sum_{s=k+1}^m \left(\hat{A}_s + \hat{B}_s + \hat{C}_s + \hat{C}_s^G\right)\left(\hat{\hat{A}}_s + \hat{\hat{B}}_s + \hat{\hat{C}}_s + \hat{\hat{C}}_s^G\right)}{\sqrt{\hat{D}_1}\hat{N}_1\sqrt{\hat{\hat{D}}_1}\hat{\hat{N}}_1}.$$

Many terms are negligible,

$$\sum_{s=k+1}^{m} \hat{A}_s\hat{\hat{A}}_s = O_p\left(\frac{1}{\theta_1^2}\right), \quad \sum_{s=k+1}^{m} \hat{A}_s\hat{\hat{C}}_s = O_p\left(\frac{1}{\sqrt{m}\theta_1^2}\right),$$

$$\sum_{s=k+1}^{m} \hat{B}_s\hat{\hat{C}}_s = O_p\left(\frac{1}{\theta_1 \min(\theta_1, \theta_k)}\right), \quad \sum_{s=k+1}^{m} \hat{C}_s\hat{\hat{C}}_s = O_p\left(\frac{1}{m\theta_1^2}\right).$$

Moreover, because $\hat{\hat{u}}_{P_{k-1},1,s}$ is invariant by rotation, then

$$\begin{aligned}
\sum_{s=k+1}^{m} \hat{A}_s\hat{\hat{B}}_s \quad &= \quad \frac{\hat{\hat{\theta}}_{P_{k-1},1}}{\hat{\hat{\theta}}_{P_k,1} - \hat{\hat{\theta}}_{P_{k-1},1}} \hat{\hat{u}}_{P_{k-1},1,k} \underbrace{\sum_{s=k+1}^{m}\left(\hat{\hat{u}}_{P_{k-1},1,s} \sum_{i=k}^{m} \frac{\hat{\lambda}_{P_{k-1},i}}{\hat{\theta}_{P_k,1} - \hat{\lambda}_{P_{k-1},i}} \hat{u}_{P_{k-1},i,s}\hat{u}_{P_{k-1},i,k}\right)}_{O_p\left(\frac{1}{m\theta_1^2}\right)} \\
&= \quad O_p\left(\frac{1}{\theta_1 \min(\theta_1, \theta_k)}\right).
\end{aligned}$$

Using the remark of Theorem 5.11.1, the last term leads to

$$\begin{aligned}
&\sum_{s=k+1}^{m} \hat{u}_{P_k,1,i}\hat{\hat{u}}_{P_k,1,i} \\
&= \frac{1}{\sqrt{\hat{D}_1}\hat{N}_1\sqrt{\hat{\hat{D}}_1}\hat{\hat{N}}_1} \sum_{s=k+1}^{m} \hat{B}_s\hat{\hat{B}}_s + O_p\left(\frac{1}{\theta_1 m}\right) \\
&= \frac{\frac{\hat{\theta}_{P_{k-1},1}}{\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1}} \hat{u}_{P_{k-1},1,k} \frac{\hat{\hat{\theta}}_{P_{k-1},1}}{\hat{\hat{\theta}}_{P_k,1} - \hat{\hat{\theta}}_{P_{k-1},1}} \hat{\hat{u}}_{P_{k-1},1,k} \sum_{s=k+1}^m \hat{u}_{P_{k-1},1,s}\hat{\hat{u}}_{P_{k-1},1,s}}{\frac{\hat{\theta}_{P_{k-1},1}}{\left|\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1}\right|}\left|\hat{u}_{P_{k-1},1,k}\right| \frac{\hat{\hat{\theta}}_{P_{k-1},1}}{\left|\hat{\hat{\theta}}_{P_k,1} - \hat{\hat{\theta}}_{P_{k-1},1}\right|}\left|\hat{\hat{u}}_{P_{k-1},1,k}\right|} + O_p\left(\frac{1}{\theta_1 m}\right) \\
&= \operatorname{sign}\left(\left(\hat{\theta}_{P_k,1} - \hat{\theta}_{P_{k-1},1}\right)\hat{u}_{P_{k-1},1,k}\left(\hat{\hat{\theta}}_{P_k,1} - \hat{\hat{\theta}}_{P_{k-1},1}\right)\hat{\hat{u}}_{P_{k-1},1,k}\right) \sum_{s=k+1}^{m} \hat{u}_{P_{k-1},1,s}\hat{\hat{u}}_{P_{k-1},1,s} \\
&= \ \operatorname{sign}\left(\hat{u}_{P_k,1,1}\right)\operatorname{sign}\left(\hat{\hat{u}}_{P_k,1,1}\right)\ \operatorname{sign}\left(\hat{\hat{u}}_{P_{k-1},1,1}\right)\ \operatorname{sign}\left(\hat{u}_{P_{k-1},1,1}\right) \sum_{s=k+1}^{m} \hat{u}_{P_{k-1},1,s}\hat{\hat{u}}_{P_{k-1},1,s}.
\end{aligned}$$

Therefore,

$$2\hat{u}_{P_k,1,1}\hat{\hat{u}}_{P_k,1,1}\sum_{i=k+1}^{m}\hat{u}_{P_k,1,i}\hat{\hat{u}}_{P_k,1,i} = 2\hat{u}_{P_{k-1},1,1}\hat{\hat{u}}_{P_{k-1},1,1}\sum_{i=k}^{m}\hat{u}_{P_{k-1},1,i}\hat{\hat{u}}_{P_{k-1},1,i} + O_p\left(\frac{1}{\theta_1 m}\right)$$

and the remark is straightforward assuming the sign convention.

$\square$

7.1.7 Double dot product

Theorem 5.11.3.
Suppose W_X and W_Y satisfies Assumption 2.2.1 and $P_k = \mathrm{I}_m + \sum_{i=1}^{k}(\theta_i - 1)e_i e_i^t$ satisfies 2.2.2 (A4), where $\theta_1 > \theta_2 > ... > \theta_k$. We set

$$\hat{\Sigma}_X = \hat{\Sigma}_{X,P_k} = P_k^{1/2} W_X P_k^{1/2} \text{ and } \hat{\Sigma}_{Y,P_k} = P_k^{1/2} W_Y P_k^{1/2}.$$

and for $s = 1, ..., k$,

$$\begin{aligned} &\hat{u}_{\hat{\Sigma}_X,s}, \hat{\theta}_{\hat{\Sigma}_X,s} \quad s.t. \quad \hat{\Sigma}_X \hat{u}_{\hat{\Sigma}_X,s} = \hat{\theta}_{\hat{\Sigma}_X,s}\hat{u}_{\hat{\Sigma}_X,s}, \\ &\hat{u}_{\hat{\Sigma}_Y,s}, \hat{\theta}_{\hat{\Sigma}_Y,s} \quad s.t. \quad \hat{\Sigma}_Y \hat{u}_{\hat{\Sigma}_Y,s} = \hat{\theta}_{\hat{\Sigma}_Y,s}\hat{u}_{\hat{\Sigma}_Y,s}, \end{aligned}$$

where $\hat{\theta}_{\hat{\Sigma}_Y,s} = \hat{\lambda}_{\hat{\Sigma}_Y,s}$ and $\hat{\theta}_{\hat{\Sigma}_X,s} = \hat{\lambda}_{\hat{\Sigma}_X,s}$. To simplify the result we assume the sign convention:

For $s = 1, 2, ..., k$ and $i = 1, 2, ..., s$, $\hat{u}_{\hat{\Sigma}_X,i,i} > 0$, $\hat{u}_{\hat{\Sigma}_Y,i,i} > 0$.

Finally, we define

$$\tilde{u}_s = \hat{U}_X^t \hat{\hat{u}}_{\hat{\Sigma}_Y,s},$$

where,

$$\hat{U}_X = (v_1, v_2, \cdots, v_m) = \left(\hat{u}_{\hat{\Sigma}_X,1}, \hat{u}_{\hat{\Sigma}_X,2}, \cdots \hat{u}_{\hat{\Sigma}_X,k}, v_{k+1}, v_{k+2}, \cdots, v_m\right),$$

where the vectors $v_{k+1}, ..., v_m$ are chosen such that the matrix $\hat{U}_X$ is orthonormal. Then,

- *If $\theta_j, \theta_t \to \infty$:*

$$\begin{aligned} \sum_{i=k+1}^{m}\tilde{u}_{j,i}\tilde{u}_{t,i} = {} & \sum_{i=k+1}^{m}\hat{u}_{\hat{\Sigma}_Y,j,i}\hat{u}_{\hat{\Sigma}_Y,t,i} + \sum_{i=k+1}^{m}\hat{u}_{\hat{\Sigma}_X,j,i}\hat{u}_{\hat{\Sigma}_X,t,i} - \sum_{i=k+1}^{m}\hat{u}_{\hat{\Sigma}_X,j,i}\hat{u}_{\hat{\Sigma}_Y,t,i} \\ & - \sum_{i=k+1}^{m}\hat{u}_{\hat{\Sigma}_Y,j,i}\hat{u}_{\hat{\Sigma}_X,t,i} - \left(\hat{u}_{\hat{\Sigma}_X,t,j} + \hat{u}_{\hat{\Sigma}_Y,j,t}\right)\left(\hat{\alpha}^2_{\hat{\Sigma}_X,j} - \hat{\alpha}^2_{\hat{\Sigma}_X,t}\right) \\ & + O_p\left(\frac{1}{\theta_1 m}\right) + O_p\left(\frac{1}{\theta_1^2\sqrt{m}}\right), \end{aligned}$$

  where $\hat{\alpha}^2_{\hat{\Sigma}_X,t} = \sum_{i=1}^{k}\hat{u}^2_{\hat{\Sigma}_X,t,i}$.

- *If $\theta_j \to \infty$ and θ_t is finite:*

$$\sum_{i=k+1}^{m}\tilde{u}_{j,i}\tilde{u}_{t,i} = O_p\left(\frac{1}{\sqrt{m}\sqrt{\theta_1}}\right).$$

- *If θ_j and θ_t are finite:*

$$\sum_{i=k+1}^{m} \tilde{u}_{j,i}\tilde{u}_{t,i} = O_p\left(\frac{1}{\sqrt{m}}\right).$$

Moreover, for $s = 1, ..., k$, $t = 2, ..., k$ and $j = k+1, ...m$,

$$\begin{aligned}
\sum_{i=1}^{k} \tilde{u}_{s,i}^2 &= \sum_{i=1}^{k} \left\langle \hat{u}_{\hat{\Sigma}_X,i}, \hat{u}_{\hat{\Sigma}_Y,s} \right\rangle^2, \\
\tilde{u}_{s,s} &= \hat{u}_{\hat{\Sigma}_X,s,s}\hat{u}_{\hat{\Sigma}_Y,s,s} + O_p\left(\frac{1}{m}\right) + O_p\left(\frac{1}{\theta_s^{1/2} m^{1/2}}\right), \\
\tilde{u}_{s,t} &= \hat{u}_{\hat{\Sigma}_X,t,s} + \hat{u}_{\hat{\Sigma}_X,s,t} + O_p\left(\frac{\sqrt{\min(\theta_s,\theta_t)}}{m\sqrt{\max(\theta_s,\theta_t)}}\right) + O_p\left(\frac{1}{\theta_t m^{1/2}}\right), \\
\tilde{u}_{t,s} &= O_p\left(\frac{\sqrt{\min(\theta_s,\theta_t)}}{m\sqrt{\max(\theta_s,\theta_t)}}\right) + O_p\left(\frac{1}{\theta_s m^{1/2}}\right), \\
\tilde{u}_{s,j} &= \hat{u}_{\hat{\Sigma}_Y,s,j} - \hat{u}_{\hat{\Sigma}_X,s,j}\left\langle \hat{u}_{\hat{\Sigma}_Y,j}, \hat{u}_{\hat{\Sigma}_X,j} \right\rangle + O_p\left(\frac{1}{\theta_s^{1/2} m}\right).
\end{aligned}$$

(Page 88)

Proof. **Theorem 5.11.3**

The proof of this theorem is divided into two main parts. First, we suppose Assumption 2.2.2(A3). Thanks to this assumption, all the eigenvalues are of same order. This property will be helpful in the computation of the dot product. In the second part we suppose Assumption 2.2.2(A4).
As in the proof of Theorem 5.10.1, we change the notation in order to obtain shorter equations :

$$\begin{aligned}
\hat{\theta}_t = \hat{\theta}_{\hat{\Sigma}_{X,P_k},t}, \quad \hat{u}_t = \hat{u}_{\hat{\Sigma}_{X,P_k},t}, \\
\hat{\hat{\theta}}_t = \hat{\theta}_{\hat{\Sigma}_{Y,P_k},t}, \quad \hat{\hat{u}}_t = \hat{u}_{\hat{\Sigma}_{Y,P_k},t}.
\end{aligned}$$

Under Assumption 2.2.2(A3) : The proof is divided into two parts. First, we build the basis v_1, v_2,..., v_m using Gramm-Schmidt. Then, we directly compute the dot product.
Construction of $v_1, ..., v_m$
The construction of the terms $v_k + 1$,...,v_m is free. Therefore, we use Gramm-Schmidt and can neglect some terms thanks to Theorems 5.3.1, 5.5.1, 5.6.1,5.7.1 and 5.8.1.

1.
$$v_1 = \hat{u}_1, ..., v_k = \hat{u}_k.$$

2.
$$\begin{aligned}
\tilde{v}_{k+1} &= e_{k+1} - \sum_{i=1}^{k} \hat{u}_{i,k+1}\hat{u}_i, \\
||\tilde{v}_{k+1}||^2 &= 1 + \sum_{i=1}^{k} \hat{u}_{i,k+1}^2 - 2\sum_{i=1}^{k} \hat{u}_{i,k+1}^2 \\
&= 1 - \sum_{i=1}^{k} \hat{u}_{i,k+1}^2, \\
\Rightarrow v_{k+1} &= \frac{\tilde{v}_{k+1}}{||\tilde{v}_{k+1}||}.
\end{aligned}$$

3.

$$\begin{aligned}
\tilde{v}_{k+2} &= e_{k+2} - \sum_{i=1}^{k} \hat{u}_{i,k+2}\hat{u}_i - \langle e_{k+2}, \tilde{v}_{k+1}\rangle \frac{\tilde{v}_{k+1}}{||\tilde{v}_{k+1}||^2} \\
&= e_{k+2} - \sum_{i=1}^{k} \hat{u}_{i,k+2}\hat{u}_i + \left(\sum_{i=1}^{k} \hat{u}_{i,k+1}\hat{u}_{i,k+2}\right)\left(e_{k+1} - \sum_{i=1}^{k} \hat{u}_{i,k+1}\hat{u}_i\right) \underbrace{\frac{1}{1-\sum_{i=1}^{k}\hat{u}_{i,k+1}^2}}_{1+O_p\left(\frac{1}{\theta m}\right)} \\
&= e_{k+2} - \sum_{i=1}^{k} \hat{u}_{i,k+2}\hat{u}_i + \left(\underbrace{\sum_{i=1}^{k} \hat{u}_{i,k+1}\hat{u}_{i,k+2}}_{O_p\left(\frac{1}{\theta m}\right)} + O_p\left(\frac{1}{\theta^2 m^2}\right)\right) e_{k+1} - \sum_{i=1}^{k} O_p\left(\frac{1}{\theta^{3/2}m^{3/2}}\right)\hat{u}_i, \\
||\tilde{v}_{k+2}||^2 &= 1 - \sum_{i=1}^{k} \hat{u}_{i,k+2}^2 + O_p\left(\frac{1}{\theta^2 m}\right).
\end{aligned}$$

4.

$$\begin{aligned}
\tilde{v}_p &= e_p - \sum_{i=1}^{k} \hat{u}_{i,p}\hat{u}_i - \sum_{s=k+1}^{p-1} \langle e_p, \tilde{v}_s\rangle \frac{\tilde{v}_s}{||\tilde{v}_s||^2} \\
&= e_p - \sum_{i=1}^{k} \hat{u}_{i,p}\hat{u}_i + \sum_{s=k+1}^{p-1}\left(\left(\sum_{j=1}^{k}\hat{u}_{j,s}\hat{u}_{j,p}\right)\left(e_s - \sum_{i=1}^{k}\hat{u}_{i,s}\hat{u}_i\right)\underbrace{\frac{1}{1-\sum_{i=1}^{k}\hat{u}_{i,s}^2}}_{1+O_p\left(\frac{1}{\theta m}\right)}\right) \\
&= e_p - \sum_{i=1}^{k} \hat{u}_{i,p}\hat{u}_i + \left(\sum_{s=k+1}^{p-1}\left(\sum_{j=1}^{k}\hat{u}_{j,s}\hat{u}_{j,p}\right)e_s - \sum_{i,j=1}^{k}\underbrace{\left(\sum_{s=k+1}^{p-1}\hat{u}_{j,s}\hat{u}_{i,s}\right)}_{\substack{\text{by dot product} \\ \text{theorem }: O_p\left(\frac{1}{\theta\sqrt{m}}\right)}}\hat{u}_{j,p}\hat{u}_i\right)\left(1+O_p\left(\frac{1}{\theta m}\right)\right) \\
&= e_p - \sum_{i=1}^{k} \hat{u}_{i,p}\hat{u}_i + \sum_{s=k+1}^{p-1}\left(\underbrace{\sum_{i=1}^{k}\hat{u}_{i,s}\hat{u}_{i,p}}_{O_p\left(\frac{1}{\theta m}\right)} + O_p\left(\frac{1}{\theta^2 m^2}\right)\right) e_s - \sum_{i=1}^{k} O_p\left(\frac{1}{\theta^{3/2} m}\right)\hat{u}_i, \\
||\tilde{v}_p||^2 &= 1 - \sum_{i=1}^{k}\hat{u}_{i,p}^2 + O_p\left(\frac{1}{\theta^2 m}\right).
\end{aligned}$$

This concludes the construction of the orthonormal rotation

$$\hat{U} = \left(\hat{u}_1, \hat{u}_2, ..., \hat{u}_k, v_{k+1}, v_{k+2}, ..., v_m\right).$$

Simplification of the partial scalar product
First, we express $\tilde{u}_{j,p}$ as a function of $\hat{\hat{u}}_i$ and $\hat{u}_i$ for $i = 1, 2, ..., k$. Then, we estimate $\tilde{u}_{j,p}\tilde{u}_{t,p}$. Finally, we propose a formula for $\sum_{p=k+1}^{m} \tilde{u}_{j,p}\tilde{u}_{t,p}$.

For $p = k+1, k+2, ..., m$

$$\begin{aligned}
\tilde{u}_{j,p} &= \left\langle v_p, \hat{\hat{u}}_j \right\rangle \\
&= \frac{\left\langle e_p - \sum_{i=1}^{k} \hat{u}_{i,p}\hat{u}_i + \sum_{s=k+1}^{p-1} \left(\sum_{i=1}^{k} \hat{u}_{i,s}\hat{u}_{i,p} + O_p\left(\frac{1}{\theta^2 m^2}\right)\right) e_s - \sum_{i=1}^{k} O_p\left(\frac{1}{\theta^{3/2} m}\right) \hat{u}_i, \hat{\hat{u}}_j \right\rangle}{1 + O_p\left(\frac{1}{\theta m}\right)} \\
&= \frac{\hat{\hat{u}}_{j,p} - \hat{u}_{j,p}\left\langle \hat{u}_j, \hat{\hat{u}}_j \right\rangle - \sum_{i=1, i\neq j}^{k} \hat{u}_{i,p} \overbrace{\left\langle \hat{u}_i, \hat{\hat{u}}_j \right\rangle}^{\hat{u}_{i,j} + \hat{\hat{u}}_{j,i} + O_p\left(\frac{1}{m}\right) + O_p\left(\frac{1}{\theta m^{1/2}}\right)} \overbrace{+ \sum_{i=1}^{k} \hat{u}_{i,p} \sum_{s=k+1}^{p-1} \hat{\hat{u}}_{j,s}\hat{u}_{i,s}}^{O_p\left(\frac{1}{\theta^{3/2} m}\right)} + O_p\left(\frac{1}{\theta^{3/2} m}\right)}{1 + O_p\left(\frac{1}{\theta m}\right)} \\
&= \hat{\hat{u}}_{j,p} - \hat{u}_{j,p}\left\langle \hat{u}_j, \hat{\hat{u}}_j \right\rangle - \sum_{i=1, i\neq j}^{k} \hat{u}_{i,p}\left(\hat{u}_{i,j} + \hat{\hat{u}}_{j,i}\right) + O_p\left(\frac{1}{\theta^{1/2} m^{3/2}}\right) + O_p\left(\frac{1}{\theta^{3/2} m}\right).
\end{aligned}$$

$$\begin{aligned}
\tilde{u}_{j,p}\tilde{u}_{t,p} &= \left(\hat{\hat{u}}_{j,p} - \hat{u}_{j,p}\left\langle \hat{u}_j, \hat{\hat{u}}_j \right\rangle\right)\left(\hat{\hat{u}}_{t,p} - \hat{u}_{t,p}\left\langle \hat{u}_t, \hat{\hat{u}}_t \right\rangle\right) - \left(\hat{\hat{u}}_{t,p} - \hat{u}_{t,p}\left\langle \hat{u}_t, \hat{\hat{u}}_t \right\rangle\right) \sum_{i=1, i\neq j}^{k} \hat{u}_{i,p}\left(\hat{u}_{i,j} + \hat{\hat{u}}_{j,i}\right) \\
&\quad - \left(\hat{\hat{u}}_{j,p} - \hat{u}_{j,p}\left\langle \hat{u}_j, \hat{\hat{u}}_j \right\rangle\right) \sum_{i=1, i\neq t}^{k} \hat{u}_{i,p}\left(\hat{u}_{i,t} + \hat{\hat{u}}_{t,i}\right) + O_p\left(\frac{1}{\theta m^2}\right) + O_p\left(\frac{1}{\theta^2 m^{3/2}}\right) \\
&= \hat{\hat{u}}_{j,p}\hat{\hat{u}}_{t,p} - \hat{\hat{u}}_{j,p}\hat{u}_{t,p}\left\langle \hat{u}_t, \hat{\hat{u}}_t \right\rangle - \hat{\hat{u}}_{t,p}\hat{u}_{j,p}\left\langle \hat{u}_j, \hat{\hat{u}}_j \right\rangle + \hat{u}_{j,p}\hat{u}_{t,p}\left\langle \hat{u}_t, \hat{\hat{u}}_t \right\rangle\left\langle \hat{u}_j, \hat{\hat{u}}_j \right\rangle \\
&\quad - \left(\hat{\hat{u}}_{t,p} - \hat{u}_{t,p}\left\langle \hat{u}_t, \hat{\hat{u}}_t \right\rangle\right)\hat{u}_{t,p}\left(\hat{u}_{t,j} + \hat{\hat{u}}_{j,t}\right) - \left(\hat{\hat{u}}_{t,p} - \hat{u}_{t,p}\left\langle \hat{u}_t, \hat{\hat{u}}_t \right\rangle\right) \sum_{i=1, i\neq j,t}^{k} \hat{u}_{i,p}\left(\hat{u}_{i,j} + \hat{\hat{u}}_{j,i}\right) \\
&\quad - \left(\hat{\hat{u}}_{j,p} - \hat{u}_{j,p}\left\langle \hat{u}_j, \hat{\hat{u}}_j \right\rangle\right)\hat{u}_{j,p}\left(\hat{u}_{j,t} + \hat{\hat{u}}_{t,j}\right) - \left(\hat{\hat{u}}_{j,p} - \hat{u}_{j,p}\left\langle \hat{u}_j, \hat{\hat{u}}_j \right\rangle\right) \sum_{i=1, i\neq j,t}^{k} \hat{u}_{i,p}\left(\hat{u}_{i,t} + \hat{\hat{u}}_{t,i}\right) \\
&\quad + O_p\left(\frac{1}{\theta m^2}\right) + O_p\left(\frac{1}{\theta^2 m^{3/2}}\right).
\end{aligned}$$

$$\begin{aligned}
\sum_{p=k+1}^{m} \tilde{u}_{j,p}\tilde{u}_{t,p} &= \sum_{p=k+1}^{m} \hat{\hat{u}}_{j,p}\hat{\hat{u}}_{t,p} - \sum_{p=k+1}^{m} \hat{\hat{u}}_{j,p}\hat{u}_{t,p} - \sum_{p=k+1}^{m} \hat{\hat{u}}_{t,p}\hat{u}_{j,p} + \sum_{p=k+1}^{m} \hat{u}_{j,p}\hat{u}_{t,p} + O_p\left(\frac{1}{\theta^2 m^{1/2}}\right) \\
&\quad - \underbrace{\left(\hat{u}_{t,j} + \hat{\hat{u}}_{j,t}\right)\left(\sum_{p=k+1}^{m} \hat{\hat{u}}_{t,p}\hat{u}_{t,p}\right)}_{O_p\left(\frac{1}{\theta m}\right)} - \left(\hat{u}_{t,j} + \hat{\hat{u}}_{j,t}\right)\left(\sum_{p=k+1}^{m} \hat{u}_{t,p}^2\right)\left\langle \hat{u}_t, \hat{\hat{u}}_t \right\rangle \\
&\quad - \underbrace{\sum_{i=1, i\neq j,t}^{k} \left(\sum_{p=k+1}^{m} \hat{u}_{i,p}\hat{\hat{u}}_{t,p} - \sum_{p=k+1}^{m} \hat{u}_{i,p}\hat{u}_{t,p}\left\langle \hat{u}_t, \hat{\hat{u}}_t \right\rangle\right)\left(\hat{u}_{i,j} + \hat{\hat{u}}_{j,i}\right)}_{O_p\left(\frac{1}{\theta m}\right)} \\
&\quad - \underbrace{\left(\hat{u}_{j,t} + \hat{\hat{u}}_{t,j}\right)\left(\sum_{p=k+1}^{m} \hat{\hat{u}}_{j,p}\hat{u}_{j,p}\right)}_{O_p\left(\frac{1}{\theta m}\right)} - \left(\hat{u}_{j,t} + \hat{\hat{u}}_{t,j}\right)\left(\sum_{p=k+1}^{m} \hat{u}_{j,p}^2\right)\left\langle \hat{u}_j, \hat{\hat{u}}_j \right\rangle \\
&\quad - \underbrace{\sum_{i=1, i\neq j,t}^{k} \left(\sum_{p=k+1}^{m} \hat{u}_{i,p}\hat{\hat{u}}_{j,p} - \sum_{p=k+1}^{m} \hat{u}_{i,p}\hat{u}_{j,p}\left\langle \hat{u}_j, \hat{\hat{u}}_j \right\rangle\right)\left(\hat{u}_{i,t} + \hat{\hat{u}}_{t,i}\right)}_{O_p\left(\frac{1}{\theta m}\right)} \\
&\quad + O_p\left(\frac{1}{\theta m}\right) + O_p\left(\frac{1}{\theta^2 m^{1/2}}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{p=k+1}^{m} \hat{\hat{u}}_{j,p}\hat{\hat{u}}_{t,p} - \sum_{p=k+1}^{m} \hat{\hat{u}}_{j,p}\hat{u}_{t,p} - \sum_{p=k+1}^{m} \hat{\hat{u}}_{t,p}\hat{u}_{j,p} + \sum_{p=k+1}^{m} \hat{u}_{j,p}\hat{u}_{t,p} \\
&\quad - \left(\hat{u}_{t,j} + \hat{\hat{u}}_{j,t}\right)(1-\hat{\alpha}_t^2) - \underbrace{\left(\hat{u}_{j,t} + \hat{\hat{u}}_{t,j}\right)}_{-\left(\hat{u}_{t,j}+\hat{\hat{u}}_{j,t}+O_p\left(\frac{1}{m}\right)\right)} (1-\hat{\alpha}_j^2) + O_p\left(\frac{1}{\theta m}\right) + O_p\left(\frac{1}{\theta^2 m^{1/2}}\right) \\
&= \sum_{p=k+1}^{m} \hat{\hat{u}}_{j,p}\hat{\hat{u}}_{t,p} - \sum_{p=k+1}^{m} \hat{\hat{u}}_{j,p}\hat{u}_{t,p} - \sum_{p=k+1}^{m} \hat{\hat{u}}_{t,p}\hat{u}_{j,p} + \sum_{p=k+1}^{m} \hat{u}_{j,p}\hat{u}_{t,p} - \left(\hat{u}_{t,j} + \hat{\hat{u}}_{j,t}\right)(\hat{\alpha}_j^2 - \hat{\alpha}_t^2) \\
&\quad + O_p\left(\frac{1}{\theta m}\right) + O_p\left(\frac{1}{\theta^2 m^{1/2}}\right).
\end{aligned}
$$

Under Assumption 2.2.2(A4) : Under this assumption, the eigenvalues are separated into two groups: $\theta_1, \theta_2, ..., \theta_{k_1}$ tending to infinity and $\theta_{k_1+1}, \theta_{k_1+2}, ..., \theta_k$ of finite size.

If $\mathbf{k_1 < j, t \leqslant k}$: The same proof as shown previously is still valid. However, because some eigenvalues are finite, the formula leads to

$$
\sum_{i=k+1}^{m} \tilde{u}_{j,i}\tilde{u}_{t,i} = O_p\left(\frac{1}{m^{1/2}}\right).
$$

If $\mathbf{j \leqslant k_1}$ and $\mathbf{k_1 < t \leqslant k}$: Assuming $p \leqslant k_1 < r \leqslant k$, then by Theorem 5.8.1,

$$
\begin{aligned}
\tilde{u}_{p,r} &= \left\langle \hat{\hat{u}}_p, \hat{u}_r \right\rangle \\
&= \sum_{s=1}^{k_1} \underbrace{\hat{u}_{r,s}}_{O_p\left(\frac{1}{\theta_1^{1/2} m^{1/2}}\right)} \hat{\hat{u}}_{p,s} + \sum_{s=k_1+1}^{k} \hat{u}_{r,s} \underbrace{\hat{\hat{u}}_{p,s}}_{O_p\left(\frac{1}{\theta_1^{1/2} m^{1/2}}\right)} + \underbrace{\sum_{s=k+1}^{m} \hat{u}_{r,s}\hat{\hat{u}}_{p,s}}_{O_p\left(\frac{1}{\theta_1^{1/2} m^{1/2}}\right)} = O_p\left(\frac{1}{\theta_1^{1/2} m^{1/2}}\right).
\end{aligned}
$$

Then,

$$
\begin{aligned}
\sum_{i=k+1}^{m} \tilde{u}_{j,i}\tilde{u}_{t,i} &= -\sum_{i=1}^{k} \tilde{u}_{j,i}\tilde{u}_{t,i} \\
&= -\tilde{u}_{j,j}\tilde{u}_{t,j} - \tilde{u}_{j,t}\tilde{u}_{t,t} - \sum_{\substack{i=1 \\ i \neq j}}^{k_1} \tilde{u}_{j,i}\tilde{u}_{t,i} - \sum_{\substack{i=k_1+1 \\ i \neq t}}^{k} \tilde{u}_{j,i}\tilde{u}_{t,i} \\
&= O_p\left(\frac{1}{\theta_1^{1/2} m^{1/2}}\right).
\end{aligned}
$$

If $\mathbf{j, t \leqslant k_1}$: In this case we will apply two rotations to the eigenvectors in order to get the result. We define three rotation matrices:

$$
\begin{aligned}
\hat{U} &= (\hat{u}_1, \hat{u}_2, \cdots \hat{u}_k, v_{k+1}, v_{k+2}, \cdots, v_m), \\
\hat{U}_\infty &= (\hat{u}_1, \hat{u}_2, \cdots \hat{u}_{k_1}, v_{\infty,k+1}, v_{\infty,k+2}, \cdots, v_{\infty,m}), \\
\hat{U}_{\not\infty} &= \hat{U}_\infty^t \hat{U}.
\end{aligned}
$$

Therefore

$$
\hat{U}_{\not\infty}^t \hat{U}_\infty^t \hat{\hat{u}}_1 = \hat{U}^t \hat{\hat{u}}_1 = \tilde{u}_1.
$$

We define $\overset{\infty}{\tilde{u}}_1$ such that

$$\overset{\infty}{\tilde{u}}_1 = \hat{U}^t_{\infty} \hat{u}_1.$$

The proof is divided into two parts.

1. First, we estimate

$$\sum_{i=k_1+1}^{k} \tilde{u}_{j,i} \tilde{u}_{t,i} \quad = \quad O_p\left(\frac{1}{\theta_1 m}\right).$$

2. Then, we show

$$\sum_{i=1}^{k_1} \overset{\infty}{\tilde{u}}_{j,i} \overset{\infty}{\tilde{u}}_{t,i} \quad = \quad \sum_{i=1}^{k_1} \tilde{u}_{j,i} \tilde{u}_{t,i}.$$

1. Because

$$\hat{U}_{\not\infty} = \left(e_1, e_2, ..., e_{k_1}, \sum_{i=k_1+1}^{m} \left\langle \overset{\infty}{v}_i, \hat{u}_{k_1+1} \right\rangle e_i, \cdots, \sum_{i=k_1+1}^{m} \left\langle \overset{\infty}{v}_i, \hat{u}_m \right\rangle e_i \right),$$

does not affect the first k_1 first components,

$$\sum_{i=k_1+1}^{m} \overset{\infty}{\tilde{u}}_{j,i} \overset{\infty}{\tilde{u}}_{t,i} \quad = \quad \sum_{i=k_1+1}^{m} \tilde{u}_{j,i} \tilde{u}_{t,i}.$$

2. We compute

$$\begin{aligned}
\tilde{u}_1 &= \hat{U}^t_{\not\infty} \overset{\infty}{\tilde{u}}_1 \\
&= \left(\overset{\infty}{\tilde{u}}_{1,1}, \overset{\infty}{\tilde{u}}_{1,2}, \cdots, \overset{\infty}{\tilde{u}}_{1,k_1}, \left\langle \sum_{i=k_1+1}^{m} \left\langle \overset{\infty}{v}_i, \hat{u}_{k_1+1} \right\rangle e_i, \overset{\infty}{\tilde{u}}_1 \right\rangle, \cdots, , \left\langle \sum_{i=k_1+1}^{m} \left\langle \overset{\infty}{v}_i, \hat{u}_k \right\rangle e_i, \overset{\infty}{\tilde{u}}_1 \right\rangle, \cdots \right).
\end{aligned}$$

We now show that $\tilde{u}_{1,i}$ for $i = k_1+1, ..., k$ are $O_p\left(\frac{1}{\theta_1^{1/2} m^{1/2}}\right)$.

$$\begin{aligned}
\tilde{u}_{1,k_1+1} &= \left\langle \sum_{p=k_1+1}^{m} \left\langle \overset{\infty}{v}_p, \hat{u}_{k_1+1} \right\rangle e_p, \overset{\infty}{\tilde{u}}_1 \right\rangle \\
&= \sum_{p=k_1+1}^{m} \left\langle \overset{\infty}{v}_p, \hat{u}_{k_1+1} \right\rangle \left\langle e_p, \overset{\infty}{\tilde{u}}_1 \right\rangle.
\end{aligned}$$

We need to estimate this sum using the first part of the proof and Theorem 5.8.1.
For $p = k_1+1, k_1+2, ..., m$,

$$\begin{aligned}
\overset{\infty}{v}_p &= \frac{e_p - \sum_{i=1}^{k_1} \hat{u}_{i,p} \hat{u}_i + \sum_{s=k_1+1}^{p-1} \left(\underbrace{\sum_{i=1}^{k_1} \hat{u}_{i,s} \hat{u}_{i,p}}_{O_p\left(\frac{1}{\theta_1 m}\right)} + O_p\left(\frac{1}{\theta_1^2 m^2}\right) \right) e_s - \sum_{i=1}^{k_1} O_p\left(\frac{1}{m \theta_1^{3/2}}\right) \hat{u}_i}{1 - \sum_{i=1}^{k_1} \hat{u}_{i,p}^2 + O_p\left(\frac{1}{\theta_1^2 m}\right)}, \\
\overset{\infty}{\tilde{u}}_{1,p} &= \hat{\hat{u}}_{1,p} - \hat{u}_{1,p} \left\langle \hat{u}_1, \hat{\hat{u}}_1 \right\rangle - \sum_{i=1, i\neq 1}^{k_1} \hat{u}_{i,p} \left(\hat{u}_{i,1} + \hat{\hat{u}}_{1,i} \right) + O_p\left(\frac{1}{\theta_1^{1/2} m^{3/2}}\right) + O_p\left(\frac{1}{\theta_1^{3/2} m}\right) \\
&= \hat{\hat{u}}_{1,p} - \hat{u}_{1,p} \left\langle \hat{u}_1, \hat{\hat{u}}_1 \right\rangle + O_p\left(\frac{1}{\theta_1^{1/2} m}\right)
\end{aligned}$$

and

$$
\begin{aligned}
\left\langle \overset{\infty}{v}_p, \hat{u}_{k_1+1} \right\rangle &= \left(\hat{u}_{k_1+1,p} + \sum_{s=k_1+1}^{p-1} \left(\sum_{i=1}^{k_1} \hat{u}_{i,s}\hat{u}_{i,p} + O_p\left(\frac{1}{\theta_1^2 m^2}\right) \right) \hat{u}_{k_1+1,s} \right) \left(1 + O_p\left(\frac{1}{\theta_1 m}\right)\right) \\
&= \hat{u}_{k_1+1,p} + \underbrace{\sum_{s=k_1+1}^{p-1} \sum_{i=1}^{k_1} \hat{u}_{i,s}\hat{u}_{i,p}\hat{u}_{k_1+1,s}}_{O_p\left(\frac{1}{\theta_1 m}\right) \text{ by Theorem 5.6.1}} + O_p\left(\frac{1}{\theta_1 m^{3/2}}\right) \\
&= \hat{u}_{k_1+1,p} + O_p\left(\frac{1}{\theta_1 m}\right).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\tilde{u}_{1,k_1+1} &= \sum_{p=k_1+1}^{m} \left\langle \overset{\infty}{v}_p, \hat{u}_{k_1+1} \right\rangle \overset{\infty}{\tilde{u}}_{1,p} \\
&= \sum_{p=k_1+1}^{m} \left(\hat{u}_{k_1+1,p}\hat{\hat{u}}_{1,p} - \hat{u}_{k_1+1,p}\hat{u}_{1,p} \left\langle \hat{u}_1, \hat{\hat{u}}_1 \right\rangle \right) + O_p\left(\frac{1}{\theta_1^{1/2} m^{1/2}}\right) \\
&= O_p\left(\frac{1}{\theta_1^{1/2} m^{1/2}}\right).
\end{aligned}
$$

Where the last line is obtained by Theorems 5.6.1 and 5.7.1.

Therefore, because of a similar proof as in paragraph "Simplification of the partial scalar product" shows

$$
\begin{aligned}
\sum_{i=k_1+1}^{m} \overset{\infty}{\tilde{u}}_{j,i} \overset{\infty}{\tilde{u}}_{t,i} &= \sum_{i=k_1+1}^{m} \hat{\hat{u}}_{j,i}\hat{\hat{u}}_{t,i} + \sum_{i=k_1+1}^{m} \hat{u}_{j,i}\hat{u}_{t,i} - \sum_{i=k_1+1}^{m} \hat{u}_{j,i}\hat{\hat{u}}_{t,i} - \sum_{i=k_1+1}^{m} \hat{\hat{u}}_{j,i}\hat{u}_{t,i} \\
&\quad - \left(\hat{u}_{t,j} + \hat{\hat{u}}_{j,t} \right) \left(\hat{\alpha}_j^2 - \hat{\alpha}_t^2 \right) + O_p\left(\frac{1}{\theta_1 m}\right) + O_p\left(\frac{1}{\theta_1^2 m^{1/2}}\right).
\end{aligned}
$$

The proof of the formula is straightforward

$$
\begin{aligned}
\sum_{i=1}^{k_1} \tilde{u}_{j,i}\tilde{u}_{t,i} &= O_p\left(\frac{1}{\theta_1 m}\right), \\
\sum_{i=1}^{k_1} \overset{\infty}{\tilde{u}}_{j,i} \overset{\infty}{\tilde{u}}_{t,i} &= \sum_{i=1}^{k_1} \tilde{u}_{j,i}\tilde{u}_{t,i}.
\end{aligned}
$$

Finally, the small results are all proven in the proof of the formula and are left to the reader.

□

7.2 Proof of the Main Theorem

Results necessary to prove the Main Theorem 3.1.1 are separated into subsection.
First, we present a sketch of the proof without detail to explain the main idea.
Then, we prove some small useful lemmas of linear algebra.
Finally, we detail each part of the sketch of the proof.

Theorem 3.1.1.
Suppose $W_X, W_Y \in \mathbb{R}^{m\times m}$ respect Assumptions 2.2.1 and 3.1.1 and for $i = 1, 2, ..., k$, θ_i respects Assumptions 2.2.2 (A3) and (A1).

1. *Let* $\tilde{P}_i = \mathrm{I}_m + (\theta_i - 1)e_i e_i^t \in \mathbb{R}^{m\times m}$ *and define*

$$\hat{\Sigma}_{X,\tilde{P}_i} = \tilde{P}_i^{1/2} W_X \tilde{P}_i^{1/2} \text{ and } \hat{\Sigma}_{Y,\tilde{P}_i} = \tilde{P}_i^{1/2} W_Y \tilde{P}_i^{1/2}.$$

The induced filtered estimators become

$$\hat{\hat{\Sigma}}_{X,\tilde{P}_i} = \mathrm{I}_m + (\hat{\hat{\theta}}_{\hat{\Sigma}_{X,\tilde{P}_i}} - 1)\hat{u}_{\hat{\Sigma}_{X,\tilde{P}_i}} \hat{u}^t_{\hat{\Sigma}_{X,\tilde{P}_i}} \text{ and } \hat{\hat{\Sigma}}_{Y,\tilde{P}_i} = \mathrm{I}_m + (\hat{\hat{\theta}}_{\hat{\Sigma}_{Y,\tilde{P}_i}} - 1)\hat{u}_{\hat{\Sigma}_{Y,\tilde{P}_i}} \hat{u}^t_{\hat{\Sigma}_{Y,\tilde{P}_i}},$$

where $\hat{\hat{\theta}}_{\hat{\Sigma}_{X,\tilde{P}_i}}$ *and* $\hat{\hat{\theta}}_{\hat{\Sigma}_{Y,\tilde{P}_i}}$ *are the unbiased estimators of the largest eigenvalues of* $\hat{\Sigma}_{X,\tilde{P}_i}$ *and* $\hat{\Sigma}_{Y,\tilde{P}_i}$ *respectively, as defined in Definition 2.2.1.*

Then, conditioning on the spectra of W_X *and* W_Y*,*

$$\sqrt{m}\frac{\left(\lambda_{\max}\left(\hat{\hat{\Sigma}}_{X,\tilde{P}_i}^{-1/2}\hat{\hat{\Sigma}}_{Y,\tilde{P}_i}\hat{\hat{\Sigma}}_{X,\tilde{P}_i}^{-1/2}\right) - \lambda^+\right)}{\sigma^+} \sim \mathbf{N}(0,1) + o_{p;m}(1),$$

where

$$\lambda^+ = \sqrt{M_2^2 - 1} + M_2,$$

$$\begin{aligned}
\sigma^{+2} = & \frac{1}{(M_{2,X} + M_{2,Y} - 2)(M_{2,X} + M_{2,Y} + 2)} \\
& \Big(9M_{2,X}^4 M_{2,Y} + 4M_{2,X}^3 M_{2,Y}^2 + 4M_{2,X}^3 M_{2,Y} + 2M_{2,X}^3 M_{3,Y} - 2M_{2,X}^2 M_{2,Y}^3 \\
& +4M_{2,X}^2 M_{2,Y}^2 - 11M_{2,X}^2 M_{2,Y} - 8M_{3,X} M_{2,X}^2 M_{2,Y} + 2M_{2,X}^2 M_{2,Y} M_{3,Y} \\
& -2M_{2,X}^2 M_{3,Y} + M_{2,X}^2 M_{4,Y} + 4M_{2,X} M_{2,Y}^3 + M_{2,X} M_{2,Y}^2 + 4M_{2,X} M_{2,Y} \\
& -4M_{3,X} M_{2,X} M_{2,Y}^2 - 4M_{3,X} M_{2,X} M_{2,Y} - 2M_{2,X} M_{2,Y}^2 M_{3,Y} - 4M_{2,X} M_{2,Y} M_{3,Y} \\
& -6M_{2,X} M_{3,Y} + 2M_{4,X} M_{2,X} M_{2,Y} + 2M_{2,X} M_{2,Y} M_{4,Y} - 2M_{3,X} M_{2,Y}^2 \\
& +2M_{3,X} M_{2,Y} + M_{4,X} M_{2,Y}^2 + 4M_{2,X}^5 + 2M_{2,X}^4 - 4M_{3,X} M_{2,X}^3 - 13M_{2,X}^3 \\
& -2M_{3,X} M_{2,X}^2 + M_{4,X} M_{2,X}^2 - 2M_{2,X}^2 + 10M_{3,X} M_{2,X} + 4M_{2,X} + 4M_{3,X} \\
& -2M_{4,X} + M_{2,Y}^5 + 2M_{2,Y}^4 - M_{2,Y}^3 - 2M_{2,Y}^2 + 4M_{2,Y} - 2M_{2,Y}^3 M_{3,Y} \\
& -2M_{2,Y}^2 M_{3,Y} + 2M_{2,Y} M_{3,Y} + 4M_{3,Y} + M_{2,Y}^2 M_{4,Y} - 2M_{4,Y} - 4\Big) \\
& + \frac{1}{\sqrt{(M_{2,X} + M_{2,Y} - 2)(M_{2,X} + M_{2,Y} + 2)}} \\
& \Big(5M_{2,X}^3 M_{2,Y} - M_{2,X}^2 M_{2,Y}^2 + 2M_{2,X}^2 M_{2,Y} + 2M_{2,X}^2 M_{3,Y} - M_{2,X} M_{2,Y}^3 \\
& +2M_{2,X} M_{2,Y}^2 - 4M_{2,X} M_{2,Y} - 4M_{3,X} M_{2,X} M_{2,Y} - 2M_{2,X} M_{3,Y} + M_{2,X} M_{4,Y} \\
& -2M_{3,X} M_{2,Y} + M_{4,X} M_{2,Y} + 4M_{2,X}^4 + 2M_{2,X}^3 - 4M_{3,X} M_{2,X}^2 - 5M_{2,X}^2 \\
& -2M_{3,X} M_{2,X} + M_{4,X} M_{2,X} + 2M_{2,X} + 2M_{3,X} + M_{2,Y}^4 + 2M_{2,Y}^3 + M_{2,Y}^2 \\
& +2M_{2,Y} - 2M_{2,Y}^2 M_{3,Y} - 2M_{2,Y} M_{3,Y} - 2M_{3,Y} + M_{2,Y} M_{4,Y}\Big),
\end{aligned}$$

$$M_{s,X} = \frac{1}{m}\sum_{i=1}^{m}\hat{\lambda}^s_{W_X,i},$$
$$M_{s,Y} = \frac{1}{m}\sum_{i=1}^{m}\hat{\lambda}^s_{W_Y,i},$$
$$M_s = \frac{M_{s,X}+M_{s,Y}}{2}.$$

Moreover,

$$\sqrt{m}\frac{\left(\lambda_{\min}\left(\hat{\hat{\Sigma}}^{-1/2}_{X,\tilde{P}_i}\hat{\hat{\Sigma}}_{Y,\tilde{P}_i}\hat{\hat{\Sigma}}^{-1/2}_{X,\tilde{P}_i}\right)-\lambda^-\right)}{\sigma^-} \sim \mathbf{N}(0,1)+o_{p;m}(1),$$

where

$$\lambda^- = -\sqrt{M_2^2-1}+M_2,$$
$$\sigma^{-2} = \left(\lambda^-\right)^4\sigma^{+2}.$$

2. *Let* $P_k = \mathrm{I}_m + \sum_{i=1}^k(\theta_i-1)e_ie_i^t\mathbb{R}^{m\times m}$ *and define*

$$\hat{\Sigma}_{X,P_k} = P_k^{1/2}W_XP_k^{1/2} \text{ and } \hat{\Sigma}_{Y,P_k} = P_k^{1/2}W_YP_k^{1/2}.$$

The induced filtered estimators become

$$\hat{\hat{\Sigma}}_{X,P_k} = \mathrm{I}_m + \sum_{i=1}^k(\hat{\hat{\theta}}_{\hat{\Sigma}_{X,P_k},i}-1)\hat{u}_{\hat{\Sigma}_{X,P_k},i}\hat{u}^t_{\hat{\Sigma}_{X,P_k},i} \text{ and } \hat{\hat{\Sigma}}_{Y,P_k} = \mathrm{I}_m + \sum_{i=1}^k(\hat{\hat{\theta}}_{\hat{\Sigma}_{Y,P_k},i}-1)\hat{u}_{\hat{\Sigma}_{Y,P_k},i}\hat{u}^t_{\hat{\Sigma}_{Y,P_k},i},$$

where $\hat{\hat{\theta}}_{\hat{\Sigma}_{X,P_k},i}$ *and* $\hat{\hat{\theta}}_{\hat{\Sigma}_{Y,P_k},i}$ *are the unbiased estimators of the* $i^{\,th}$ *largest eigenvalue of* $\hat{\Sigma}_{X,P_k}$ *and* $\hat{\Sigma}_{Y,P_k}$ *respectively.*
Then, conditioning on the spectra of W_X *and* W_Y*,*

$$\lambda_{\max}\left(\hat{\hat{\Sigma}}^{-1/2}_{X,P_k}\hat{\hat{\Sigma}}_{Y,P_k}\hat{\hat{\Sigma}}^{-1/2}_{X,P_k}\right) = \lambda_{\max}\left(H^+\right)+1+O_p\left(\frac{1}{m}\right)+O_p\left(\frac{1}{\theta\sqrt{m}}\right),$$
$$\lambda_{\min}\left(\hat{\hat{\Sigma}}^{-1/2}_{X,P_k}\hat{\hat{\Sigma}}_{Y,P_k}\hat{\hat{\Sigma}}^{-1/2}_{X,P_k}\right) = \lambda_{\max}\left(H^-\right)+1+O_p\left(\frac{1}{m}\right)+O_p\left(\frac{1}{\theta\sqrt{m}}\right),$$

where

$$H^{\pm} = \zeta^{\pm}_{\infty}\begin{pmatrix} \hat{\zeta}^{\pm}_1/\zeta^{\pm}_{\infty} & w^{\pm}_{1,2} & w^{\pm}_{1,3} & \cdots & w^{\pm}_{1,k} \\ w^{\pm}_{2,1} & \hat{\zeta}^{\pm}_2/\zeta^{\pm}_{\infty} & w^{\pm}_{2,3} & \cdots & w^{\pm}_{2,k} \\ w^{\pm}_{3,1} & w^{\pm}_{3,2} & \hat{\zeta}^{\pm}_3/\zeta^{\pm}_{\infty} & \cdots & w^{\pm}_{3,k} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ w^{\pm}_{k,1} & w^{\pm}_{k,2} & w^{\pm}_{k,3} & \cdots & \hat{\zeta}^{\pm}_k/\zeta^{\pm}_{\infty} \end{pmatrix},$$

and

$$\hat{\zeta}^+_i = \lambda_{\max}\left(\hat{\hat{\Sigma}}^{1/2}_{X,\tilde{P}_i}\hat{\hat{\Sigma}}_{Y,\tilde{P}_i}\hat{\hat{\Sigma}}^{1/2}_{X,\tilde{P}_i}\right)-1,$$
$$\hat{\zeta}^-_i = \lambda_{\min}\left(\hat{\hat{\Sigma}}^{1/2}_{X,\tilde{P}_i}\hat{\hat{\Sigma}}_{Y,\tilde{P}_i}\hat{\hat{\Sigma}}^{1/2}_{X,\tilde{P}_i}\right)-1,$$
$$\zeta^{\pm}_{\infty} = \lim_{m\to\infty}\hat{\zeta}^{\pm}_i = \lambda^{\pm}-1,$$
$$w^{\pm}_{i,j} \sim \mathbf{N}\left(0,\frac{1}{m}\frac{2(M_{2,X}-1)(M_{2,Y}-1)+B^{\pm}_X+B^{\pm}_Y}{\left((\zeta^{\pm}_{\infty}-2M_2+1)^2+2(M_2-1)\right)^2}\right)+o_p\left(\frac{1}{\sqrt{m}}\right),$$

$$B_X^+ = \left(1 - M_2 + 2M_{2,X} + \sqrt{M_2^2 - 1}\right)^2 (M_{2,X} - 1)$$
$$+2\left(-1 + M_2 - 2M_{2,x} - \sqrt{M_2^2 - 1}\right)(M_{3,X} - M_{2,X}) + (M_{4,X} - M_{2,X}^2),$$
$$B_Y^+ = \left(1 + M_2 + M_{2,Y} - M_{2,X} - \sqrt{M_2^2 - 1}\right)^2 (M_{2,Y} - 1)$$
$$+2\left(-1 - M_2 - M_{2,Y} - M_{2,X} - \sqrt{M_2^2 - 1}\right)(M_{3,Y} - M_{2,Y}) + (M_{4,Y} - M_{2,Y}^2),$$
$$B_X^- = \left(1 - M_2 + 2M_{2,X} - \sqrt{M_2^2 - 1}\right)^2 (M_{2,X} - 1)$$
$$+2\left(-1 + M_2 - 2M_{2,x} + \sqrt{M_2^2 - 1}\right)(M_{3,X} - M_{2,X}) + (M_{4,X} - M_{2,X}^2),$$
$$B_Y^- = \left(1 + M_2 + M_{2,Y} - M_{2,X} + \sqrt{M_2^2 - 1}\right)^2 (M_{2,Y} - 1)$$
$$+2\left(-1 - M_2 - M_{2,Y} + M_{2,X} - \sqrt{M_2^2 - 1}\right)(M_{3,Y} - M_{2,Y}) + (M_{4,Y} - M_{2,Y}^2).$$

The matrices H^+ and H^- are strongly correlated. However, within a matrix, all the entries are jointly independent.

Remark 7.2.1.

1. Without Assumption 3.1.1, the red parts of the theorem are weaker. The entries $w_{i,j}$ have the same two first moments but not necessarily the asymptotic normal distribution. Then, the entries of $H^\pm$ are just uncorrelated instead of being independent.

2. If the spectra follow Marcenko-Pastur distributions, then

$$c = \frac{c_X + c_Y}{2},$$
$$\lambda^+ = c + \sqrt{c(c+2)} + 1,$$
$$\sigma^{+2} = c_X^3 + c_X^2 c_Y + 3c_X^2 + 4c_X c_Y - c_X + c_Y^2 + c_Y$$
$$+\frac{(8c_X + 2c_X^2 + \left(c_X^3 + 5c_X^2 + c_X^2 c_Y + 4c_X c_Y + 5c_X + 3c_Y + c_Y^2\right)\sqrt{c(c+2)}}{c+2},$$
$$w_{i,j}^+ \sim \mathbf{N}\left(0, \frac{\sigma_w^2}{m}\right),$$
$$\sigma_w^2 = \frac{2c_X\left(\sqrt{c(c+2)} + 2\right) + 2c_Y\left(-\sqrt{c(c+2)} + 2\right) + c_X^2 + c_Y^2}{4c\left(-\sqrt{c\,(c+2)} + c + 2\right)^2}.$$

3. If c_X tends to 0, then

$$\sigma^{+2} = \Big(M_{2,Y}^5 + 2M_{2,Y}^4 - 2M_{3,Y}M_{2,Y}^3 + M_{2,Y}^3 - 4M_{3,Y}M_{2,Y}^2 + M_{4,Y}M_{2,Y}^2 + 2M_{2,Y}^2$$
$$+2M_{4,Y}M_{2,Y} + 2M_{2,Y} - 2M_{3,Y} - M_{4,Y} - 2\Big)\Big/\Big((M_{2,Y} - 1)(M_{2,Y} + 3)\Big)$$
$$+\Big(M_{2,Y}^4 + M_{2,Y}^3 - 2M_{3,Y}M_{2,Y}^2 + 2M_{2,Y}^2 - 2M_{3,Y}M_{2,Y} + M_{4,Y}M_{2,Y}$$
$$-2M_{3,Y} + M_{4,Y}\Big)\Big/\sqrt{(M_{2,Y} - 1)(M_{2,Y} + 3)}.$$

4. When m is not large enough, the normality assumption of $\hat{\lambda}_i$ is not respected. In this case, and in particular if k is large, it could be profitable to estimate the order 1 residual spike ($\tilde{P}_i$) with the following algorithm.

   (a) Let $\hat{\lambda}_{W_X,i}$, $\hat{\lambda}_{W_Y,i}$ be the eigenvalues of W_X and W_Y respectively.

   (b) Generate u_x and u_y, two independent uniform unit vectors of size m.

   (c) Generate Z, a standard normal independent of u_x and u_y.

   (d) We define

$$W_{x,1,1} = \sum_{i=1}^{m} \hat{\lambda}_{W_X,i} u_{x,i}^2 \text{ and } W_{x,1,1}^2 = \sum_{i=1}^{m} \hat{\lambda}_{W_X,i} u_{x,i}^2,$$
$$W_{y,2,2} = \sum_{i=1}^{m} \hat{\lambda}_{W_Y,i} u_{y,i}^2 \text{ and } W_{y,2,2}^2 = \sum_{i=1}^{m} \hat{\lambda}_{W_Y,i} u_{y,i}^2.$$

   Assuming 2.2.1, the statistics $W_{x,1,1}$ and $W_{x,1,1}^2$ follow the distribution of the first entry of W_X and ${W_X}^2$ respectively.

   (e) Construct

$$\theta_x = \theta\ W_{x,1,1} \text{ and } \alpha_x^2 = 1 + 1/\theta - \theta/\theta_x^2\ W_{x,1,1}^2,$$
$$\theta_y = \theta\ W_{y,2,2} \text{ and } \alpha_y^2 = 1 + 1/\theta - \theta/\theta_y^2\ W_{y,2,2}^2,$$
$$\alpha^2 = \alpha_x^2\alpha_y^2 + 2\sqrt{\alpha_x^2\alpha_y^2} Z \frac{\sqrt{(1-\alpha_x^2)(1-\alpha_y^2)}}{m}.$$

   (f) Finally, because θ is large enough

$$\lambda_{\max}\left(\hat{\hat{\Sigma}}_{X,\tilde{P}_i}^{1/2} \hat{\hat{\Sigma}}_{Y,\tilde{P}_i} \hat{\hat{\Sigma}}_{X,\tilde{P}_i}^{1/2}\right) \sim \frac{\theta_y\left(1-\alpha^2\right) + 1 + \frac{\theta_y}{\theta_x} + \sqrt{-4\frac{\theta_y}{\theta_x} + \left(\theta_y\left(1-\alpha^2\right) + \frac{\theta_y}{\theta_x} + 1\right)^2}}{2},$$
$$\lambda_{\min}\left(\hat{\hat{\Sigma}}_{X,\tilde{P}_i}^{1/2} \hat{\hat{\Sigma}}_{Y,\tilde{P}_i} \hat{\hat{\Sigma}}_{X,\tilde{P}_i}^{1/2}\right) \sim \frac{\theta_y\left(1-\alpha^2\right) + 1 + \frac{\theta_y}{\theta_x} - \sqrt{-4\frac{\theta_y}{\theta_x} + \left(\theta_y\left(1-\alpha^2\right) + \frac{\theta_y}{\theta_x} + 1\right)^2}}{2}$$

   In practice the spectra of W_X and W_Y are not observed and will be replaced by the $m-k$ smallest eigenvalues of $\hat{\Sigma}_{X,P_k}$ and $\hat{\Sigma}_{Y,P_k}$.

5. Assuming that we would like to use Monte Carlo methods to estimate the distribution, we should first estimate the eigenvalues of the covariance matrices.
   Without the theorem, the loops of the simulation generating the residual spikes generate $O\left(m^2\right)$ elements.
   Using the theorem, the loops generate k^2 elements.
   Finally, using the previous algorithm, the loops generate $O(m)$ elements.

(Page 26)

7.2.1 Sketch of the proof

In this section we show the sketch of the proof. The details are presented in the next sections.

Residual spike for perturbations of order 1 **:**
Using Lemma 6.2.1 and Theorem 5.3.1, a delta method proves the first part of the Main Theorem

3.1.1 for perturbations of order 1.
The details are presented in Section 7.2.3.

Decomposition of the matrix :
Generalisation of the previous result to perturbations of order k is not straightforward. We recall that we want to study the largest eigenvalue of

$$\hat{\hat{\Sigma}}_{P_k,X}^{-1/2}\hat{\hat{\Sigma}}_{P_k,Y}\hat{\hat{\Sigma}}_{P_k,X}^{-1/2},$$

where $\hat{\hat{\Sigma}}_{P_k,X}$ is the filtered estimator of the random covariance matrix $\hat{\Sigma}_{P_k,X} = P_k^{1/2}W_X P_k^{1/2}$ defined in 2.2.1. First, we define a rotation matrix $\hat{U}_{P_k,X}$ such that $\hat{U}_{P_k,X}^t \hat{u}_{P_k,X,i} = e_i$ for $i = 1, 2, ..., k$ and $\hat{U}_{P_k,X}^t \hat{u}_{P_k,Y,i} = \tilde{u}_{P_k}$ as in Theorem 5.11.3. Then, we propose the matrices with same eigenvalues,

$$\Sigma_{P_k,X}^{-1/2}\tilde{\Sigma}_{P_k,Y}\Sigma_{P_k,X}^{-1/2} \quad = \quad \mathrm{I}_m + \sum_{i=1}^{k}\left[\Sigma_{P_k,X}^{-1/2}\left(\hat{\hat{\theta}}_{P_k,Y,i}-1\right)\tilde{u}_{P_k,i}\tilde{u}_{P_k,i}^t\Sigma_{P_k,X}^{-1/2} + \left(\frac{1}{\hat{\hat{\theta}}_{P_k,X,i}}-1\right)e_ie_i^t\right],$$

where $\Sigma_{P_k,X} = \mathrm{I}_m + \sum_{i=1}^k \left(\hat{\hat{\theta}}_{P_k,X,i}-1\right)e_ie_i^t$ and $\tilde{\Sigma}_{P_k,Y} = \mathrm{I}_m + \sum_{i=1}^k \left(\hat{\hat{\theta}}_{P_k,Y,i}-1\right)\tilde{u}_i\tilde{u}_i^t$.
The details are presented in Section 7.2.4.

Pseudo Invariance of the residual spike :
The residual spike is not invariant when k grows. We define for $i = 1, 2, ..., k$, $\hat{\Sigma}_{\tilde{P}_i,X} = \tilde{P}_i^{1/2}W_X\tilde{P}_i^{1/2}$, where $\tilde{P}_i = \mathrm{I}_m + (\theta_i - 1)e_ie_i^t$ is a perturbation of order 1.
Therefore, an invariance is proven,

$$\lambda\left(\Sigma_{P_k,X}^{-1/2}\left(\hat{\hat{\theta}}_{P_k,Y,i}-1\right)\tilde{u}_{P_k,i}\tilde{u}_{P_k,i}^t\Sigma_{P_k,X}^{-1/2} + \left(\frac{1}{\hat{\hat{\theta}}_{P_k,X,i}}-1\right)e_ie_i^t\right) = \lambda\left(\hat{\hat{\Sigma}}_{\tilde{P}_i,X}^{-1/2}\hat{\hat{\Sigma}}_{\tilde{P}_i,Y}\hat{\hat{\Sigma}}_{\tilde{P}_i,X}^{-1/2}\right) - 1 + O_p\left(\frac{1}{m}\right),$$

where $\lambda()$ provides the non null eigenvalues. This result is proven assuming either that θ_1 is large or that the two non-trivial residual spikes of the perturbation of order 1 are distinct. The second condition is clear because the eigenvectors are biased. However, when $n_X, n_Y >> m$, this could create some imprecision.
The details are presented in Section 7.2.5.

Pseudo residual eigenvectors :
The previous part demonstrates the pseudo invariance of the residual spike. The next step studies the pseudo residual eigenvectors.
For $s = 1, 2, ..., k$, we set

$$\hat{\zeta}_s^{\pm} = \lambda\left(\Sigma_{P_k,X}^{-1/2}\left(\hat{\hat{\theta}}_{P_k,Y,s}-1\right)\tilde{u}_{P_k,s}\tilde{u}_{P_k,s}^t\Sigma_{P_k,X}^{-1/2} + \left(\frac{1}{\hat{\hat{\theta}}_{P_k,X,s}}-1\right)e_se_s^t\right)$$

and

$$w_s^{\pm} \quad = \quad u\left(\Sigma_{P_k,X}^{-1/2}\left((\hat{\theta}_{P_k,Y,s}-1)\tilde{u}_{P_k,s}\tilde{u}_{P_k,Y,s}^t\right)\Sigma_{P_k,X}^{-1/2} + \left(\frac{1}{\hat{\theta}_{P_k,X,s}}-1\right)e_se_s^t\right),$$

its corresponding eigenvector. The notation $\pm$ allows $\hat{\zeta}_s^-$ and $\hat{\zeta}_s^+$ to be distinguished. We define for $s = 1, 2, ..., k$,

$$\zeta_\infty^{\pm}(\theta_s) = \lim_{m\to\infty}\hat{\zeta}_s^{\pm},$$
$$\zeta_\infty^{\pm} = \lim_{m,\theta_s\to\infty}\hat{\zeta}_s^{\pm}.$$

Then,

$$
\begin{aligned}
w^{\pm}_{s,s} &= \frac{\sqrt{\hat{\theta}_{P_k,X,s}}}{\mathrm{Norm}^{\pm}_s\sqrt{\hat{\theta}_{P_k,Y,s}-1}\tilde{u}_{P_k,s,s}}\left(\hat{\zeta}^{\pm}_s-\left(\hat{\theta}_{P_k,Y,s}-1\right)\left(1-\hat{\alpha}^2_{P_k,s}\right)\right)+O_p\left(\frac{1}{m}\right)\\
&= \frac{\sqrt{\theta_s}}{\mathrm{Norm}^{\pm}_s\sqrt{\theta_s-1}\alpha_s}\left(\zeta^{\pm}_{\infty}(\theta_s)-(\theta_s-1)\left(1-\alpha_s^2\right)\right)+O_p\left(\frac{1}{\sqrt{m}}\right)\\
&= \frac{\left(\zeta^{\pm}_{\infty}-2\left(M_2-1\right)\right)}{\sqrt{\left(\zeta^{\pm}_{\infty}-2\left(M_2-1\right)\right)^2+2\left(M_2-1\right)}}+O_p\left(\frac{1}{\sqrt{m}}\right)+o_{p;\theta}\left(1\right),
\end{aligned}
$$

$$
\begin{aligned}
w^{\pm}_{s,2:m\setminus s} &= \frac{\sqrt{\hat{\theta}_{P_k,Y,s}-1}\left(\frac{\tilde{u}_{P_k,s,1}}{\sqrt{\hat{\theta}_{P_k,X,1}}},...,\frac{\tilde{u}_{P_k,s,s-1}}{\sqrt{\hat{\theta}_{P_k,X,s-1}}}\frac{\tilde{u}_{P_k,s,s+1}}{\sqrt{\hat{\theta}_{P_k,X,s+1}}},...,\frac{\tilde{u}_{P_k,s,k}}{\sqrt{\hat{\theta}_{P_k,X,k}}},\tilde{u}_{P_k,s,k+1},...,\tilde{u}_{P_k,s,m}\right)}{\mathrm{Norm}^{\pm}_s}\\
&= \frac{\sqrt{\hat{\theta}_{P_k,Y,1}-1}\left(\frac{\tilde{u}_{P_k,s,1}}{\sqrt{\hat{\theta}_{P_k,X,1}}},...,\frac{\tilde{u}_{P_k,s,s-1}}{\sqrt{\hat{\theta}_{P_k,X,s-1}}},\frac{\tilde{u}_{P_k,s,s+1}}{\sqrt{\hat{\theta}_{P_k,X,s+1}}},...,\frac{\tilde{u}_{P_k,s,k}}{\sqrt{\hat{\theta}_{P_k,X,k}}},\tilde{u}_{P_k,s,k+1},...,\tilde{u}_{P_k,s,m}\right)}{\sqrt{\left(\zeta^{\pm}_{\infty}-2\left(M_2-1\right)\right)^2+2\left(M_2-1\right)+O_p\left(\frac{1}{\sqrt{m}}\right)}+o_{p;\theta}(1)},
\end{aligned}
$$

$$
\begin{aligned}
\left(\mathrm{Norm}^{\pm}_s\right)^2 &= \frac{\hat{\theta}_{P_k,X,s}\left(\hat{\zeta}^{\pm}_s-\left(\hat{\theta}_{P_k,Y,s}-1\right)\left(1-\hat{\alpha}^2_{P_k,s}\right)\right)^2}{\left(\hat{\theta}_{P_k,Y,s}-1\right)\tilde{u}^2_{P_k,s,s}}+\left(\hat{\theta}_{P_k,Y,s}-1\right)\left(1-\hat{\alpha}^2_{P_k,s}\right)+O_p\left(\frac{1}{m}\right)\\
&= \frac{\theta_s}{\left(\theta_s-1\right)\alpha_s^2}\left(\zeta^{\pm}_{\infty}\left(\theta_s\right)-\left(\theta_s-1\right)\left(1-\alpha_s^2\right)\right)^2+\left(\theta_s-1\right)\left(1-\alpha_s^2\right)+O_p\left(\frac{1}{\sqrt{m}}\right).
\end{aligned}
$$

The details of the proof are presented in Section 7.2.6.

Remark 7.2.2.
These results concerning pseudo residual structure are proven for all θ. This extension is not used in the Main Theorem, but it could be interesting in the robust part 4.2 to argue that the largest residual spike occurs when θ is large.

Dimension reduction :
The three previous parts showed that

$$
\lambda\left(\hat{\hat{\Sigma}}^{-1/2}_{P_k,X}\hat{\Sigma}_{P_k,Y}\hat{\hat{\Sigma}}^{-1/2}_{P_k,X}\right) = \lambda\left(\mathrm{I}_m+\sum_{i=1}^{k}\left[\Sigma^{-1/2}_{P_k,X}\left(\hat{\hat{\theta}}_{P_k,Y,i}-1\right)\tilde{u}_{P_k,i}\tilde{u}^t_{P_k,i}\Sigma^{-1/2}_{P_k,X}+\left(\frac{1}{\hat{\hat{\theta}}_{P_k,X,i}}-1\right)e_ie_i^t\right]\right)
$$

and for $i=1,2,...,k$, we showed

$$
\hat{\zeta}^+_i w^+_i w^{+\,t}_i+\hat{\zeta}^-_i w^-_i w^{-\,t}_i=\left[\Sigma^{-1/2}_{P_k,X}\left(\hat{\hat{\theta}}_{P_k,Y,i}-1\right)\tilde{u}_{P_k,i}\tilde{u}^t_{P_k,i}\Sigma^{-1/2}_{P_k,X}+\left(\frac{1}{\hat{\hat{\theta}}_{P_k,X,i}}-1\right)e_ie_i^t\right],
$$

where $\pm$ allows the distinction of the pseudo eigenvalues and eigenvectors such that $+$ is the largest and $-$ is the smallest. Some easy arguments of linear algebra in Lemmas 6.2.4 and 6.2.2 lead to

$$
\begin{aligned}
\lambda_{\max}\left(\hat{\hat{\Sigma}}^{-1/2}_{P_k,X}\hat{\hat{\Sigma}}_{P_k,Y}\hat{\hat{\Sigma}}^{-1/2}_{P_k,X}\right) &= \lambda_{\max}\left(H^+\right)+1+O_p\left(\frac{1}{m}\right),\\
\lambda_{\min}\left(\hat{\hat{\Sigma}}^{-1/2}_{P_k,X}\hat{\hat{\Sigma}}_{P_k,Y}\hat{\hat{\Sigma}}^{-1/2}_{P_k,X}\right) &= \lambda_{\min}\left(H^-\right)+1+O_p\left(\frac{1}{m}\right),
\end{aligned}
$$

where $H^{\pm}$ are matrices of dimension k such that

$$H^{\pm} = \begin{pmatrix} \hat{\zeta}_1^{\pm} & \sqrt{\hat{\zeta}_1^{\pm}\hat{\zeta}_2^{\pm}}\left\langle w_1^{\pm}, w_2^{\pm}\right\rangle & \sqrt{\hat{\zeta}_1^{\pm}\hat{\zeta}_3^{\pm}}\left\langle w_1^{\pm}, w_3^{\pm}\right\rangle & \cdots & \sqrt{\hat{\zeta}_k^{\pm}\hat{\zeta}_2^{\pm}}\left\langle w_1^{\pm}, w_k^{\pm}\right\rangle \\ \sqrt{\hat{\zeta}_2^{\pm}\hat{\zeta}_1^{\pm}}\left\langle w_2^{\pm}, w_1^{\pm}\right\rangle & \hat{\zeta}_2^{\pm} & \sqrt{\hat{\zeta}_2^{\pm}\hat{\zeta}_3^{\pm}}\left\langle w_2^{\pm}, w_3^{\pm}\right\rangle & \cdots & \sqrt{\hat{\zeta}_2^{\pm}\hat{\zeta}_k^{\pm}}\left\langle w_2^{\pm}, w_k^{\pm}\right\rangle \\ \sqrt{\hat{\zeta}_3^{\pm}\hat{\zeta}_1^{\pm}}\left\langle w_3^{\pm}, w_1^{\pm}\right\rangle & \sqrt{\hat{\zeta}_3^{\pm}\hat{\zeta}_2^{\pm}}\left\langle w_3^{\pm}, w_2^{\pm}\right\rangle & \hat{\zeta}_3^{\pm} & \cdots & \sqrt{\hat{\zeta}_3^{\pm}\hat{\zeta}_k^{\pm}}\left\langle w_3^{\pm}, w_k^{\pm}\right\rangle \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \sqrt{\hat{\zeta}_k^{\pm}\hat{\zeta}_1^{\pm}}\left\langle w_k^{\pm}, w_1^{\pm}\right\rangle & \sqrt{\hat{\zeta}_k^{\pm}\hat{\zeta}_2^{\pm}}\left\langle w_k^{\pm}, w_2^{\pm}\right\rangle & \sqrt{\hat{\zeta}_k^{\pm}\hat{\zeta}_3^{\pm}}\left\langle w_k^{\pm}, w_3^{\pm}\right\rangle & \cdots & \hat{\zeta}_k^{\pm} \end{pmatrix}.$$

The details are explained in Section 7.2.7.

Elements of H
In this part we assume Assumptions 2.2.2(A1) and (A3). The matrices $H^{\pm}$ are functions of $\hat{\zeta}_i^{\pm}$ and $\left\langle w_i^{\pm}, w_j^{\pm}\right\rangle$ for $i, j = 1, 2, ..., k$. By the pseudo invariance of the residual spike, we know that $\hat{\zeta}_i^{\pm}$ behaves like residual spikes of perturbation of order 1. However, the behaviour of $\left\langle w_i^{\pm}, w_j^{\pm}\right\rangle$ is unknown. Using Theorem 5.11.3 and the investigation of $w_i^{\pm}$ of the previous paragraph, $\left\langle w_i^{\pm}, w_j^{\pm}\right\rangle$ can be expressed in function of well-known statistics. We directly see that $\langle w_s^{\pm}, w_s^{\pm}\rangle = 1$. Moreover, for $s \neq t$,

$$\begin{aligned}\left\langle w_s^{\pm}, w_t^{\pm}\right\rangle &= \frac{\sqrt{\theta_s\theta_t}}{\left(\zeta_\infty^{\pm} - 2M_2 + 1\right)^2 + 2\left(M_2 - 1\right)} \times \\ &\quad -\Bigg(\sum_{p=k+1}^{m} \hat{u}_{P_k,Y,s,p}\hat{u}_{P_k,Y,t,p} + \sum_{p=k+1}^{m} \hat{u}_{P_k,X,s,p}\hat{u}_{P_k,X,t,p} \\ &\qquad - \sum_{p=k+1}^{m} \hat{u}_{P_k,Y,s,p}\hat{u}_{P_k,X,t,p} - \sum_{p=k+1}^{m} \hat{u}_{P_k,Y,t,p}\hat{u}_{P_k,X,s,p} \\ &\qquad - \left(\hat{u}_{P_k,X,t,s} + \hat{u}_{P_k,Y,s,t}\right)\left(\tilde{\alpha}_s^2 - \tilde{\alpha}_t^2 - \left(\zeta_\infty^{\pm} - 2\left(M_2 - 1\right)\right)\left(\frac{1}{\theta_t} - \frac{1}{\theta_s}\right)\right)\Bigg) \\ &\quad + O_p\left(\frac{1}{m}\right) + O_p\left(\frac{1}{\theta\sqrt{m}}\right).\end{aligned}$$

By using the distribution Theorems 5.6.1, 5.3.1 and the Invariant Theorems, we can compute the asymptotic moment of $\left\langle w_i^{\pm}, w_j^{\pm}\right\rangle$.

$$\begin{aligned}&\left\langle w_i^{\pm}, w_j^{\pm}\right\rangle \sim \mathrm{RV}\left(0, \frac{1}{m}\frac{2(M_{2,X}-1)(M_{2,Y}-1) + B_X^{\pm} + B_Y^{\pm}}{\left((\zeta_\infty^{\pm} - 2M_2 + 1)^2 + 2(M_2 - 1)\right)^2}\right) \\ &B_X^{+} = \left(1 - M_2 + 2M_{2,X} + \sqrt{M_2^2 - 1}\right)^2 (M_{2,X} - 1) \\ &\qquad + 2\left(-1 + M_2 - 2M_{2,x} - \sqrt{M_2^2 - 1}\right)(M_{3,X} - M_{2,X}) + (M_{4,X} - M_{2,X}^2) \\ &B_Y^{+} = \left(1 + M_2 + M_{2,Y} - M_{2,X} - \sqrt{M_2^2 - 1}\right)^2 (M_{2,Y} - 1) \\ &\qquad + 2\left(-1 - M_2 - M_{2,Y} - M_{2,X} - \sqrt{M_2^2 - 1}\right)(M_{3,Y} - M_{2,Y}) + (M_{4,Y} - M_{2,Y}^2)\end{aligned}$$

$$B_X^- = \left(1 - M_2 + 2M_{2,X} - \sqrt{M_2^2 - 1}\right)^2 (M_{2,X} - 1)$$
$$+2\left(-1 + M_2 - 2M_{2,x} + \sqrt{M_2^2 - 1}\right)(M_{3,X} - M_{2,X}) + (M_{4,X} - M_{2,X}^2)$$
$$B_Y^- = \left(1 + M_2 + M_{2,Y} - M_{2,X} + \sqrt{M_2^2 - 1}\right)^2 (M_{2,Y} - 1)$$
$$+2\left(-1 - M_2 - M_{2,Y} + M_{2,X} - \sqrt{M_2^2 - 1}\right)(M_{3,Y} - M_{2,Y}) + (M_{4,Y} - M_{2,Y}^2).$$

The details of the computation are presented in Section 7.2.8.

Normality discussion
When the perturbation is of order 1, we can easily show the asymptotic normality using 5.3.1.
When the perturbation is of order k and $n_X >> n_Y$, then the rotation of Theorem 5.11.3 is simpler and the joint normality is straightforward consequence of Theorem 5.11.2 .
Nevertheless, in the general case we can only express the entries of $H^{\pm}$ as a function of marginally Normal statistics. The Assumption 3.1.1 in particular says that these statistics are asymptotically jointly Normal. This leads to the result.
The details of the computations are presented in Section 7.2.9.

7.2.2 Prerequisite Lemmas

In order to prove the Main Theorem 3.1.1 we introduce small lemmas.

Lemma 6.2.1.
Suppose

$$D = \left(\mathrm{I}_m + (\theta - 1)u_X u_X^t\right)^{-1/2} \left(\mathrm{I}_m + (\theta - 1)u_Y u_Y^t\right) \left(\mathrm{I}_m + (\theta - 1)u_X u_X^t\right)^{-1/2}.$$

The eigenvalues of D are 1 *and*

$$\lambda(D) = -\frac{1}{2\theta}\left(-1 + \alpha^2 - 2\alpha^2\theta - \theta^2(1 - \alpha^2) \pm \sqrt{-4\theta^2 + [1 + \theta^2 - (-1 + \theta)^2\alpha^2]^2}\right),$$

where $\alpha^2 = \langle u_X, u_Y\rangle^2$.
Moreover, if

$$D_2 = \left(\mathrm{I}_m + (\theta_X - 1)u_X u_X^t\right)^{-1/2} \left(\mathrm{I}_m + (\theta_Y - 1)u_Y u_Y^t\right) \left(\mathrm{I}_m + (\theta_X - 1)u_X u_X^t\right)^{-1/2}.$$

The eigenvalues of D_2 are 1 *and*

$$\lambda(D_2) = \frac{1}{2}\left(\theta_Y + \alpha^2 - \theta_Y\alpha^2 + \frac{1 + (\theta_Y - 1)\alpha^2 \pm \sqrt{-4\theta_Y\theta_X + (1 + \theta_Y\theta_X - (\theta_Y - 1)(\theta_X - 1)\alpha^2)^2}}{\theta_X}\right),$$

where $\alpha^2 = \langle u_X, u_Y\rangle^2$.

(Page 91)

Proof.

First, we see that for any $m-2$ orthogonal vectors, orthogonal to u_X and u_Y, the eigenvalues are 1. Consequently, we just need to compute the two last eigenvalues, λ_1 and λ_2. We will consider the Trace of D_2 and D_2^2.

$$\begin{aligned}
\operatorname{Trace}(D_2) &= \operatorname{Trace}\left(\left(\mathrm{I}_m + (\theta_Y - 1)u_Y u_Y^t\right)\left(\mathrm{I}_m + \left(\frac{1}{\theta_X} - 1\right)u_X u_X^t\right)\right) \\
&= m + (\theta_Y - 1) + \left(\frac{1}{\theta_X} - 1\right) + (\theta_Y - 1)\left(\frac{1}{\theta_X} - 1\right)\langle u_X, u_Y\rangle^2 \\
&= m - 2 + \langle u_X, u_Y\rangle^2 + \theta_Y\left(1 - \langle u_X, u_Y\rangle^2\right) + \frac{1 + \langle u_Y, u_Y\rangle^2(\theta_Y - 1)}{\theta_X},
\end{aligned}$$

$$\begin{aligned}
\operatorname{Trace}\left(D_2^2\right) &= \operatorname{Trace}\left(\left(\left(\mathrm{I}_m + (\theta_Y - 1)u_Y u_Y^t\right)\left(\mathrm{I}_m + \left(\frac{1}{\theta_X} - 1\right)u_X u_X^t\right)\right)^2\right) \\
&= m - 2 + \langle u_X, u_Y\rangle^4 + \theta_Y^2\left(\langle u_X, u_Y\rangle^2 - 1\right)^2 - 2\theta_Y\langle u_X, u_Y\rangle^2\left(\langle u_X, u_Y\rangle^2 - 1\right) \\
&\quad + \frac{\left(1 + (\theta_Y - 1)\langle u_X, u_Y\rangle^2\right)^2}{\theta_X^2} - \frac{2(\theta_X - 1)^2\langle u_X, u_Y\rangle^2\left(\langle u_X, u_Y\rangle^2 - 1\right)}{\theta_X}.
\end{aligned}$$

Moreover,

$$\begin{aligned}
\operatorname{Trace}(D_2) &= m - 2 + \lambda_1 + \lambda_2, \\
\operatorname{Trace}\left(D_2^2\right) &= m - 2 + \lambda_1^2 + \lambda_2^2.
\end{aligned}$$

Therefore, the result is obtained by solving an equation of second order and is left to the reader.

□

Lemma 6.2.2.
Suppose $w_1, ..., w_k \in \mathbb{R}^m$ and $\lambda_1, ..., \lambda_k \in \mathbb{R}^*$, then if the function $\lambda()$ provides non-trivial eigenvalues,

$$\lambda\left(\sum_{i=1}^{k} \lambda_i w_i w_i^t\right) = \lambda\left(H\right),$$

where

$$H = \begin{pmatrix}
\lambda_1 & \sqrt{\lambda_1\lambda_2}\langle w_1, w_2\rangle & \sqrt{\lambda_1\lambda_3}\langle w_1, w_3\rangle & \cdots & \sqrt{\lambda_k\lambda_2}\langle w_1, w_k\rangle \\
\sqrt{\lambda_2\lambda_1}\langle w_2, w_1\rangle & \lambda_2 & \sqrt{\lambda_2\lambda_3}\langle w_2, w_3\rangle & \cdots & \sqrt{\lambda_2\lambda_k}\langle w_2, w_k\rangle \\
\sqrt{\lambda_3\lambda_1}\langle w_3, w_1\rangle & \sqrt{\lambda_3\lambda_2}\langle w_3, w_2\rangle & \lambda_3 & \cdots & \sqrt{\lambda_3\lambda_k}\langle w_3, w_k\rangle \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
\sqrt{\lambda_k\lambda_1}\langle w_k, w_1\rangle & \sqrt{\lambda_k\lambda_2}\langle w_k, w_2\rangle & \sqrt{\lambda_k\lambda_3}\langle w_k, w_3\rangle & \cdots & \lambda_k
\end{pmatrix}.$$

(Page 92)

Proof. **Lemma 6.2.2**

We define

$$\Lambda = \begin{pmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \lambda_2 & & 0 \\
\vdots & & \ddots & \vdots \\
0 & 0 & \cdots & \lambda_k
\end{pmatrix}, \; W = (w_1, w_2, ..., w_k) \in \mathbb{R}^{m \times k},$$

then

$$\begin{aligned}
\lambda\left(\sum_{i=1}^{k}\lambda_i w_i w_i^t\right) &= \lambda\left(W\Lambda W^t\right) = \lambda\left(\left(W\Lambda^{1/2}\right)\left(W\Lambda^{1/2}\right)^t\right)\\
&= \lambda\left(\left(W\Lambda^{1/2}\right)^t\left(W\Lambda^{1/2}\right)\right) \text{ (for nonzero eigenvalues!)}\\
&= \lambda\left(\Lambda^{1/2}W^tW\Lambda^{1/2}\right) = \lambda\left(H\right).
\end{aligned}$$

□

Lemma 6.2.3.
Suppose $e_1, w \in \mathbf{R}^m$ and $a, b \in \mathbf{R}$. Then, if $||w|| = 1$, the two ($\pm$) non-trivial eigenvalues and eigenvectors are

$$\begin{aligned}
\lambda^{\pm}\left(ae_1e_1^t + bww^t\right) &= \frac{1}{2}\left(a + b \pm \sqrt{4abw_1^2 + (a-b)^2}\right),\\
u^{\pm}\left(ae_1e_1^t + bww^t\right) &= \frac{1}{\text{Norm}^{\pm}}\left(\frac{\lambda^{\pm}\left(ae_1e_1^t + bww^t\right) + b\left(w1^2 - 1\right)}{bw_1}, w_2, w_3, w_4, ..., w_m\right),\\
\left(\text{Norm}^{\pm}\right)^2 &= \frac{\left(\lambda^{\pm}\left(ae_1e_1^t + bww^t\right) + b\left(w1^2 - 1\right)\right)^2}{b^2w_1^2} + 1 - w_1^2.
\end{aligned}$$

If $||w|| \neq 1$,

$$\begin{aligned}
\lambda^{\pm}\left(ae_1e_1^t + ww^t\right) &= \frac{1}{2}\left(\pm\sqrt{(a + ||w||^2)^2 - 4a\left(||w||^2 - w_1^2\right)} + a + ||w||^2\right),\\
u^{\pm}\left(ae_1e_1^t + ww^t\right) &= \frac{1}{\text{Norm}^{\pm}}\left(\frac{\lambda^{\pm}\left(ae_1e_1^t + ww^t\right) - ||w||^2 + w_1^2}{w_1}, w_2, w_3, w_4, ..., w_m\right),\\
\left(\text{Norm}^{\pm}\right)^2 &= \frac{\left(\lambda^{\pm}\left(ae_1e_1^t + ww^t\right) - ||w||^2 + w_1^2\right)^2}{w_1^2} + ||w||^2 - w_1^2.
\end{aligned}$$

(Page 92)

Proof. **Lemma 6.2.3**

These results were computed with Wolfram Mathematica 11.1.1. and the reader can check them by computing

$$\left(ae_1e_1^t + bww^t\right)u^{\pm} = \lambda^{\pm}u^{\pm}.$$

□

Lemma 6.2.4.
Suppose $u_1, ..., u_k \in \mathbb{R}^m$ are orthonormal and $\lambda_1 > ... > \lambda_k \in \mathbb{R}^+$ where k is finite. Suppose $v \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^+$ such that $\langle u_i, v\rangle = O_p\left(1/\sqrt{m}\right)$ and $\mu - \lambda_1 < d < 0$ for a fixed d, then

$$\lambda_{\max}\left(\sum_{i=1}^k \lambda_i u_i u_i^t + \mu v v^t\right) = \lambda_1 + O_p\left(\frac{1}{m}\right).$$

Moreover, if $\mu - \lambda_k > d_2 > 0$ for a fixed d_2,

$$\lambda_{\min}\left(\sum_{i=1}^k \lambda_i u_i u_i^t + \mu v v^t\right) = \lambda_k + O_p\left(\frac{1}{m}\right).$$

(Page 93)

Proof. **Lemma 6.2.4**

Suppose w is the maximum unit eigenvector of $\sum_{i=1}^k \lambda_i u_i u_i^t + \mu v v^t$. Then,

$$w = \sum_{i=1}^k \alpha_i u_i + \beta v,$$

where

$$\sum_{i=1}^k \alpha_i^2 + \beta^2 + 2\sum_{i=1}^k \alpha_i \beta \langle u_i, v\rangle = 1.$$

If $\beta = O_p(1/\sqrt{m})$, then

$$\begin{aligned}
w^t\left(\sum_{i=1}^k \lambda_i u_i u_i^t + \mu v v^t\right) w &= \sum_{i=1}^k \lambda_i \left(\alpha_i + \beta\langle v, u_i\rangle\right)^2 + \mu\left(\sum_{i=1}^k \alpha_i \langle u_i, v\rangle + \beta\right)^2 \\
&= \sum_{i=1}^k \lambda_i \alpha_i^2 + O_p\left(\frac{1}{m}\right) \\
&\leqslant \lambda_1 + O_p\left(\frac{1}{m}\right).
\end{aligned}$$

If β is larger than $O_p(1/\sqrt{m})$,

$$\begin{aligned}
w^t\left(\sum_{i=1}^k \lambda_i u_i u_i^t + \mu v v^t\right) w &= \sum_{i=1}^k \lambda_i \left(\alpha_i + \beta\langle v, u_i\rangle\right)^2 + \mu\left(\sum_{i=1}^k \alpha_i \langle u_i, v\rangle + \beta\right)^2 \\
&\leqslant \lambda_1\left(\sum_{i=1}^k \left(\alpha_i + \beta\langle v, u_i\rangle\right)^2\right) + \mu\left(\sum_{i=1}^k \alpha_i \langle u_i, v\rangle + \beta\right)^2 \\
&= \lambda_1\left(\sum_{i=1}^k \alpha_i^2 + 2\sum_{i=1}^k \alpha_i \beta\langle v, u_i\rangle\right) + \mu\left(2\beta\sum_{i=1}^k \alpha_i \langle u_i, v\rangle + \beta^2\right) + O_p\left(\frac{1}{m}\right) \\
&= \lambda_1 + (\mu - \lambda_1)\beta^2 + 2\mu\beta\sum_{i=1}^k \alpha_i \langle u_i, v\rangle + O_p\left(\frac{1}{m}\right) \\
&\leqslant \lambda_1 + O_p\left(\frac{1}{m}\right),
\end{aligned}$$

where the two last lines are obtained using $\sum_{i=1}^k \alpha_i^2 + 2\sum_{i=1}^k \alpha_i\beta\langle u_i, v\rangle = 1 - \beta^2$ and because

$$P\left\{(\mu - \lambda_1)\beta^2 + 2\mu\beta\sum_{i=1}^k \alpha_i \langle u_i, v\rangle < 0\right\} \underset{m\to\infty}{\to} 1.$$

On the other hand,

$$\lambda_{\max}\left(\sum_{i=1}^{k}\lambda_i u_i u_i^t+\mu v v^t\right)\geqslant u_1^t\left(\sum_{i=1}^{k}\lambda_i u_i u_i^t+\mu v v^t\right)u_1=\lambda_1+O_p\left(\frac{1}{m}\right).$$

This concludes the proof.

□

7.2.3 Residual spike for perturbations of order 1

First we assume a perturbation of order 1, $P_1=\mathrm{I}_m+(\theta_1-1)e_1e_1^t$. The Lemma 6.2.1 gives a formula for the residual spike as a function of $\hat{\hat{\theta}}_X$, $\hat{\hat{\theta}}_Y$ and $\hat{\alpha}^2=\langle\hat{u}_{X,1},\hat{u}_{Y,1}\rangle^2$. We call this formula, $\lambda^{\pm}\left(\hat{\hat{\theta}}_X,\hat{\hat{\theta}}_Y,\hat{\alpha}^2\right)$.
For all θ, the asymptotic joint distribution of the three parameters,$\hat{\hat{\theta}}_X$, $\hat{\hat{\theta}}_Y$ and $\hat{\alpha}^2=\langle\hat{u}_{X,1},\hat{u}_{Y,1}\rangle^2$, is known by Theorem 5.3.1. Therefore, applying Slutsky's Theorem shows the asymptotic normal distribution of the residual spike.
The computation of the asymptotic variance is left to the reader. It only requires the first derivative of the residual formula of Lemma 6.2.1.
This proves the first part of the Main Theorem 3.1.1.

7.2.4 Decomposition of the difference matrix

As proposed in the Section 7.2.1, we decompose the matrix $\hat{\hat{\Sigma}}_{P_k,X}^{-1/2}\hat{\hat{\Sigma}}_{P_k,Y}\hat{\hat{\Sigma}}_{P_k,X}^{-1/2}$.

$$\hat{\hat{\Sigma}}_{P_k,X}^{-1/2}\hat{\hat{\Sigma}}_{P_k,Y}\hat{\hat{\Sigma}}_{P_k,X}^{-1/2}=\mathrm{I}_m+\sum_{i=1}^{k}\left[\hat{\hat{\Sigma}}_{P_k,X}^{-1/2}\left(\hat{\hat{\theta}}_{P_k,Y,i}-1\right)\hat{u}_{P_k,Y,i}\hat{u}_{P_k,Y,i}^t\hat{\hat{\Sigma}}_{P_k,X}^{-1/2}+\left(\frac{1}{\hat{\hat{\theta}}_{P_k,X,i}}-1\right)\hat{u}_{P_k,X,i}\hat{u}_{P_k,X,i}^t\right].$$

Then, we define a rotation matrix $\hat{U}_{P_k,X}$ such that $\hat{U}_{P_k,X}^t\hat{u}_{P_k,X,i}=e_i$ and $\hat{U}_{P_k,X}^t\hat{u}_{P_k,Y,i}=\tilde{u}_{P_k}$ as in Theorem 5.11.3. Because this rotation does not affect the eigenvalues,

$$\begin{aligned}\lambda\left(\hat{\hat{\Sigma}}_{P_k,X}^{-1/2}\hat{\hat{\Sigma}}_{P_k,Y}\hat{\hat{\Sigma}}_{P_k,X}^{-1/2}\right)&=\lambda\left(\hat{U}_{P_k,X}\hat{\hat{\Sigma}}_{P_k,X}^{-1/2}\hat{U}_{P_k,X}^t\hat{U}_{P_k,X}\hat{\hat{\Sigma}}_{P_k,Y}\hat{U}_{P_k,X}^t\hat{U}_{P_k,X}\hat{\hat{\Sigma}}_{P_k,X}^{-1/2}\hat{U}_{P_k,X}^t\right)\\&=\lambda\left(\mathrm{I}_m+\sum_{i=1}^{k}\left[\Sigma_{P_k,X}^{-1/2}\left(\hat{\hat{\theta}}_{P_k,Y,i}-1\right)\tilde{u}_{P_k,i}\tilde{u}_{P_k,i}^t\Sigma_{P_k,X}^{-1/2}+\left(\frac{1}{\hat{\hat{\theta}}_{P_k,X,i}}-1\right)e_ie_i^t\right]\right),\end{aligned}$$

where $\Sigma_{P_k,X}=\hat{U}_{P_k,X}^t\hat{\hat{\Sigma}}_{P_k,X}\hat{U}_{P_k,X}=\mathrm{I}_m+\sum_{i=1}^{k}\left(\hat{\hat{\theta}}_{P_k,X,i}-1\right)e_ie_i^t$ and $\lambda()$ provides the eigenvalues of the matrices.

7.2.5 Pseudo invariant residual spike

Assuming that $\hat{\Sigma}_{P_k,X}=P_k^{1/2}W_XP_k^{1/2}$ and $\hat{\Sigma}_{P_k,Y}=P_k^{1/2}W_YP_k^{1/2}$, we define $\hat{\Sigma}_{\tilde{P}_i,X}=\tilde{P}_i^{1/2}W_X\tilde{P}_i^{1/2}$, where $\tilde{P}_i=\mathrm{I}_m+(\theta_i-1)e_ie_i^t$. We show that if θ_1 is large or if we assume that the residual spikes of the perturbations of order 1 are distinct,

$$\lambda\left(\Sigma_{P_k,X}^{-1/2}\left(\hat{\hat{\theta}}_{P_k,Y,i}-1\right)\tilde{u}_{P_k,i}\tilde{u}_{P_k,i}^t\Sigma_{P_k,X}^{-1/2}+\left(\frac{1}{\hat{\hat{\theta}}_{P_k,X,i}}-1\right)e_ie_i^t\right)=\lambda\left(\hat{\hat{\Sigma}}_{\tilde{P}_i,X}^{-1/2}\hat{\hat{\Sigma}}_{\tilde{P}_i,Y}\hat{\hat{\Sigma}}_{\tilde{P}_i,X}^{-1/2}\right)-1+O_p\left(\frac{1}{m}\right).$$

The proof of this equality is computed in two steps.

1. First we compute the non trivial eigenvalues and eigenvectors of

$$\Sigma_{P_k,X}^{-1/2}\left(\hat{\hat{\theta}}_{P_k,Y,i}-1\right)\tilde{u}_{P_k,i}\tilde{u}_{P_k,i}^t\Sigma_{P_k,X}^{-1/2}.$$

2. Then, using the Lemma 6.2.3, we establish the equality.

1. We define

$$\begin{aligned}\tilde{\Sigma}_X &= \mathrm{I} + \sum_{i=1}^{k} (\hat{\theta}_{X,i} - 1)\tilde{\tilde{u}}_i \tilde{\tilde{u}}_i^t, \\ \tilde{\tilde{u}}_{i,1} &= \tilde{u}_{1,i}.\end{aligned}$$

The vector $\tilde{\tilde{u}}_i$ is just $\hat{U}_Y^t \hat{u}_{X,i}$ or $\tilde{U}e_i$. Using the fact that for a matrix M, the non trivial eigenvector of $e_1 e_1^t M$ is e_1,

$$\begin{aligned}\lambda\left(\Sigma_X^{-1/2}\left((\hat{\theta}_{Y,1}-1)\tilde{u}_1\tilde{u}_1^t\right)\Sigma_X^{-1/2}\right) &= \lambda\left(\tilde{\Sigma}_X^{-1/2}\left((\hat{\theta}_{Y,1}-1)e_1e_1^t\right)\tilde{\Sigma}_X^{-1/2}\right) \\ &= \lambda\left(\left((\hat{\theta}_{Y,1}-1)e_1e_1^t\right)\tilde{\Sigma}_X^{-1}\right) \\ &= e_1^t\left((\hat{\theta}_{Y,1}-1)e_1e_1^t\right)\tilde{\Sigma}_X^{-1}e_1 \\ &= e_1^t(\hat{\theta}_{Y,1}-1)\left(\left(e_1e_1^t\right)+\sum_{i=1}^{k}\left(\frac{1}{\hat{\theta}_{X,i}}-1\right)\tilde{\tilde{u}}_{i,1}e_1\tilde{\tilde{u}}_i^t\right)e_1 \\ &= (\hat{\theta}_{Y,1}-1)\left(1+\sum_{i=1}^{k}\left(\frac{1}{\hat{\theta}_{X,i}}-1\right)\tilde{\tilde{u}}_{i,1}^2\right) \\ &= (\hat{\theta}_{Y,1}-1)\left(1+\sum_{i=1}^{k}\left(\frac{1}{\hat{\theta}_{X,i}}-1\right)\tilde{u}_{1,i}^2\right).\end{aligned}$$

The computation of the eigenvector leads to

$$\begin{aligned}u\left(\Sigma_X^{-1/2}\left((\hat{\theta}_{Y,1}-1)\tilde{u}_1\tilde{u}_1^t\right)\Sigma_X^{-1/2}\right) &\propto \Sigma_X^{-1/2}u\left(\left((\hat{\theta}_{Y,1}-1)\tilde{u}_1\tilde{u}_1^t\right)\Sigma_X^{-1}\right) \\ &\propto \Sigma_X^{-1/2}\tilde{u}_1 \\ &\propto \left(\frac{\tilde{u}_{1,1}}{\sqrt{\hat{\theta}_{X,1}}}, \frac{\tilde{u}_{1,2}}{\sqrt{\hat{\theta}_{X,2}}}, ..., \frac{\tilde{u}_{1,k}}{\sqrt{\hat{\theta}_{X,k}}}, \tilde{u}_{1,k+1}, ..., \tilde{u}_{1,m}\right).\end{aligned}$$

Because the previous eigenvector is not standardised, we compute its norm,

$$\begin{aligned}\left\|\left(\frac{\tilde{u}_{1,1}}{\sqrt{\hat{\theta}_{X,1}}}, \frac{\tilde{u}_{1,2}}{\sqrt{\hat{\theta}_{X,2}}}, ..., \frac{\tilde{u}_{1,k}}{\sqrt{\hat{\theta}_{X,k}}}, \tilde{u}_{1,k+1}, ..., \tilde{u}_{1,m}\right)\right\|^2 &= \frac{\tilde{u}_{1,1}^2}{\hat{\theta}_{X,1}} + \frac{\tilde{u}_{1,2}^2}{\hat{\theta}_{X,2}} + ... + \frac{\tilde{u}_{1,k}^2}{\hat{\theta}_{X,k}} + \tilde{u}_{1,k+1}^2 + ... + \tilde{u}_{1,m}^2 \\ &= 1 + \sum_{i=1}^{k}\left(\frac{1}{\hat{\theta}_{X,i}}-1\right)\tilde{u}_{1,i}^2 \\ &= \frac{\lambda\left(\hat{\hat{\Sigma}}_X^{-1/2}\left((\hat{\theta}_{Y,1}-1)\hat{u}_{Y,1}\hat{u}_{Y,1}^t\right)\hat{\hat{\Sigma}}_X^{-1/2}\right)}{\hat{\theta}_{Y,1}-1}.\end{aligned}$$

We conclude the first part with the two formulas:

$$\lambda\left(\Sigma_X^{-1/2}\left((\hat{\theta}_{Y,1}-1)\tilde{u}_1\tilde{u}_1^t\right)\Sigma_X^{-1/2}\right) = \left(\hat{\theta}_{Y,1}-1\right)\left(\sum_{i=1}^{k}\left(\frac{1}{\hat{\theta}_{X,i}}-1\right)\tilde{u}_{1,i}^2+1\right)$$

and

$$u\left(\Sigma_X^{-1/2}\left((\hat{\theta}_{Y,1}-1)\tilde{u}_1\tilde{u}_1^t\right)\Sigma_X^{-1/2}\right)$$
$$=\frac{\sqrt{\hat{\theta}_{Y,1}-1}}{\sqrt{\lambda\left(\hat{\hat{\Sigma}}_X^{-1/2}\left((\hat{\theta}_{Y,1}-1)\hat{u}_{Y,1}\hat{u}_{Y,1}^t\right)\hat{\hat{\Sigma}}_X^{-1/2}\right)}}\left(\frac{\tilde{u}_{1,1}}{\sqrt{\hat{\theta}_{X,1}}},\frac{\tilde{u}_{1,2}}{\sqrt{\hat{\theta}_{X,2}}},...,\frac{\tilde{u}_{1,k}}{\sqrt{\hat{\theta}_{X,k}}},\tilde{u}_{1,k+1},...,\tilde{u}_{1,m}\right).$$

2. The second part uses the Lemma 6.2.3 to establish the following relation

$$\Sigma_{P_k,X}^{-1/2}\left(\hat{\hat{\theta}}_{P_k,Y,1}-1\right)\tilde{u}_{P_k,1}\tilde{u}_{P_k,1}^t\Sigma_{P_k,X}^{-1/2}+\left(\frac{1}{\hat{\hat{\theta}}_{P_k,X,1}}-1\right)e_1e_1^t$$
$$=\lambda\left(\hat{\hat{\Sigma}}_{P_1,X}^{-1/2}\hat{\hat{\Sigma}}_{P_1,X}\hat{\hat{\Sigma}}_{P_1,X}^{-1/2}\right)-1+O_p\left(\frac{1}{m}\right).$$

(a) We start with the order 1,

$$\begin{aligned}\lambda\left(\hat{\hat{\Sigma}}_{P_1,X}^{-1/2}\hat{\hat{\Sigma}}_{P_1,X}\hat{\hat{\Sigma}}_{P_1,X}^{-1/2}\right) &= \lambda\left(\Sigma_{P_1,X}^{-1/2}\left((\hat{\theta}_{P_1,Y,1}-1)\tilde{u}_{P_1,1}\tilde{u}_{P_1,1}^t\right)\Sigma_{P_1,X}^{-1/2}+\left(\frac{1}{\hat{\theta}_{P_1,X,1}}-1\right)e_1e_1^t+\mathrm{I}\right)\\ &= \lambda\left(\hat{\eta}_{P_1}u_{\hat{\eta}_{P_1}}u_{\hat{\eta}_{P_1}}^t+\left(\frac{1}{\hat{\theta}_{P_1,X,1}}-1\right)e_1e_1^t\right)+1,\end{aligned}$$

where,

$$\begin{aligned}\hat{\eta}_{P_1} &= \lambda\left(\Sigma_X^{-1/2}\left((\hat{\theta}_{Y,1}-1)\tilde{u}_1\tilde{u}_1^t\right)\Sigma_X^{-1/2}\right)=\frac{\hat{\theta}_{P_1,Y,1}-1}{\hat{\theta}_{P_1,X,1}}\tilde{u}_{P_1,1,1}^2+\left(\hat{\theta}_{P_1,Y,1}-1\right)\left(1-\tilde{u}_{P_1,1,1}^2\right),\\ u_{\hat{\eta}_{P_1}} &= u\left(\Sigma_X^{-1/2}\left((\hat{\theta}_{Y,1}-1)\tilde{u}_1\tilde{u}_1^t\right)\Sigma_X^{-1/2}\right)=\frac{\sqrt{\hat{\theta}_{P_1,Y,1}-1}\left(\frac{\tilde{u}_{P_1,1,1}}{\sqrt{\hat{\theta}_{P_1,X,1}}},\tilde{u}_{P_1,1,2},...,\tilde{u}_{P_1,1,m}\right)}{\sqrt{\lambda\left(\Sigma_X^{-1/2}\left((\hat{\theta}_{Y,1}-1)\tilde{u}_1\tilde{u}_1^t\right)\Sigma_X^{-1/2}\right)}}.\end{aligned}$$

By the Lemma 6.2.3, the non trivial eigenvalues are functions of different parameters. Using a similar notation to the Lemma we set

$$\begin{aligned}a_{P_1} &= \frac{1}{\hat{\theta}_{P_1,X,1}}-1,\\ b_{P_1} &= \frac{\hat{\theta}_{P_1,Y,1}-1}{\hat{\theta}_{P_1,X,1}}\tilde{u}_{P_1,1,1}^2+\left(\hat{\theta}_{P_1,Y,1}-1\right)\left(1-\tilde{u}_{P_1,1,1}^2\right),\\ b_{P_1}w_{P_1}^2 &= \left(\hat{\theta}_{P_1,Y,1}-1\right)\frac{\tilde{u}_{P_1,1,1}^2}{\hat{\theta}_{P_1,X,1}}.\end{aligned}$$

The Lemma 6.2.3 provides a function g such that

$$\lambda\left(\hat{\eta}_{P_1}u_{\hat{\eta}_{P_1}}u_{\hat{\eta}_{P_1}}^t+\left(\frac{1}{\hat{\theta}_{P_1,X,1}}-1\right)e_1e_1^t\right)=g^{\pm}(a_{P_1},b_{P_1},b_{P_1}w_{P_1}^2).$$

Therefore

$$\lambda\left(\hat{\hat{\Sigma}}_{P_1,X}^{-1/2}\hat{\hat{\Sigma}}_{P_1,X}\hat{\hat{\Sigma}}_{P_1,X}^{-1/2}\right) = g^{\pm}(a_{P_1},b_{P_1},b_{P_1}w_{P_1}^2).$$

(b) For perturbations of order k,

$$\lambda\left(\hat{\hat{\Sigma}}_{P_k,X}^{-1/2}\left((\hat{\theta}_{P_k,Y,1}-1)\hat{u}_{P_k,Y,1}\hat{u}_{P_k,Y,1}^t\right)\hat{\hat{\Sigma}}_{P_k,X}^{-1/2}+\left(\frac{1}{\hat{\theta}_{P_k,X,1}}-1\right)\hat{u}_{P_k,X,1}\hat{u}_{P_k,X,1}^t\right)$$
$$=\lambda\left(\Sigma_{P_k,X}^{-1/2}\left((\hat{\theta}_{P_k,Y,1}-1)\tilde{u}_{P_k,1}\tilde{u}_{P_k,1}^t\right)\Sigma_{P_k,X}^{-1/2}+\left(\frac{1}{\hat{\theta}_{P_k,X,1}}-1\right)e_1e_1^t\right)$$
$$=\lambda\left(\hat{\eta}_{P_k}u_{\hat{\eta}_{P_k}}u_{\hat{\eta}_{P_k}}^t+\left(\frac{1}{\hat{\theta}_{P_k,X,1}}-1\right)e_1e_1^t\right),$$

where

$$\begin{aligned}\hat{\eta}_{P_k,1} &= \left(\hat{\theta}_{P_k,Y,1}-1\right)\left(\sum_{i=1}^{k}\left(\frac{1}{\hat{\theta}_{P_k,X,i}}-1\right)\tilde{u}_{P_k,1,i}^2+1\right)\\ &= \frac{\hat{\theta}_{P_k,Y,1}-1}{\hat{\theta}_{P_k,X,1}}\left(\sum_{i=1}^{k}\tilde{u}_{P_k,1,i}^2\right)+\left(\hat{\theta}_{P_k,Y,1}-1\right)\left(1-\sum_{i=1}^{k}\tilde{u}_{P_k,1,i}^2\right)+O_p\left(\frac{1}{m}\right),\\ u_{\hat{\eta}_{P_k,1}} &= \frac{\sqrt{\hat{\theta}_{P_k,Y,1}-1}\left(\frac{\tilde{u}_{P_k,1,1}}{\sqrt{\hat{\theta}_{P_k,X,1}}},\frac{\tilde{u}_{P_k,1,2}}{\sqrt{\hat{\theta}_{P_k,X,2}}},...,\frac{\tilde{u}_{P_k,1,k}}{\sqrt{\hat{\theta}_{P_k,X,k}}},\tilde{u}_{P_k,1,k+1},...,\tilde{u}_{P_k,1,m}\right)}{\sqrt{\lambda\left(\hat{\hat{\Sigma}}_{P_k,X}^{-1/2}\left((\hat{\theta}_{P_k,Y,1}-1)\hat{u}_{P_k,Y,1}\hat{u}_{P_k,Y,1}^t\right)\hat{\hat{\Sigma}}_{P_k,X}^{-1/2}\right)}}.\end{aligned}$$

As previously by the Lemma 6.2.3, the non-trivial eigenvalues are functions of different parameters. Using a similar notation to the Lemma we set

$$\begin{aligned}a_{P_k} &= \frac{1}{\hat{\theta}_{P_k,X,1}}-1,\\ b_{P_k} &= \frac{\hat{\theta}_{P_k,Y,1}-1}{\hat{\theta}_{P_k,X,1}}\left(\sum_{i=1}^{k}\tilde{u}_{P_k,1,i}^2\right)+\left(\hat{\theta}_{P_k,Y,1}-1\right)\left(1-\sum_{i=1}^{k}\tilde{u}_{P_k,1,i}^2\right)+O_p\left(\frac{1}{m}\right),\\ b_{P_k}w_{P_k,1}^2 &= \left(\hat{\theta}_{P_k,Y,1}-1\right)\frac{\tilde{u}_{P_k,1,1}^2}{\hat{\theta}_{P_k,X,1}}=\left(\hat{\theta}_{P_k,Y,1}-1\right)\frac{\sum_{i=1}^{k}\tilde{u}_{P_k,1,i}^2}{\hat{\theta}_{P_k,X,1}}+O_p\left(\frac{1}{m}\right).\end{aligned}$$

The Lemma 6.2.3 provides the function g such that

$$\lambda\left(\hat{\eta}_{P_k}u_{\hat{\eta}_{P_k}}u_{\hat{\eta}_{P_k}}^t+\left(\frac{1}{\hat{\theta}_{P_k,X,1}}-1\right)e_1e_1^t\right) = g^{\pm}(a_{P_k},b_{P_k},b_{P_k}w_{P_k,1}^2).$$

(c) Finally we show that

$$g^{\pm}(a_{P_k},b_{P_k},b_{P_k}w_{P_k,1}^2)=g^{\pm}(a_{P_1},b_{P_1},b_{P_1}w_{P_1,1}^2)+O_p\left(\frac{1}{m}\right).$$

By the Invariant Theorems,

$$\begin{aligned}&a_{P_k}=a_{P_1}+O_p\left(\frac{1}{\theta m}\right),\\ &b_{P_k}=b_{P_1}+O_p\left(\frac{1}{m}\right),\\ &b_{P_k}w_{P_k,1}^2=b_{P_1}w_{P_1}^2+O_p\left(\frac{1}{m}\right).\end{aligned}$$

Moreover, the three parameters do not converge to 0.
Because we know from Lemma 6.2.3 that g is continuous,

$$g^{\pm}(x,y,z) = \frac{1}{2}\left(x+y\pm\sqrt{4xz+(x-y)^2}\right).$$

This function is Lipschitz if $4xz+(x-y)^2$ is not closed to 0. The reader can show that the perturbation creates two residual spikes different from 1 when θ_1 is detectable. (In other cases the covariance matrices are the same.) In particular we can show that when θ_1 is large, the pseudo residual spike is distinct from 1.
Therefore, using this property we conclude,

$$\begin{aligned}
&\left|\left(\lambda\left(\hat{\hat{\Sigma}}_{P_1,X}^{-1/2}\hat{\hat{\Sigma}}_{P_1,X}\hat{\hat{\Sigma}}_{P_1,X}^{-1/2}\right)-1\right)-\left(\lambda\left(\hat{\eta}_{P_k,1}u_{\hat{\eta}_{P_k,1}}u^t_{\hat{\eta}_{P_k,1}}+\left(\frac{1}{\hat{\theta}_{P_k,X,1}}-1\right)e_1e_1^t\right)\right)\right| \\
&\qquad = \left|g^{\pm}(a_{P_1},b_{P_1},b_{P_1}w^2_{P_1,1})-g^{\pm}(a_{P_k},b_{P_k},b_{P_k}w^2_{P_k,1})\right| \\
&\qquad = O_p\left(\frac{1}{m}\right)
\end{aligned}$$

Remark 7.2.3.
The hypothesis assuming that $\lambda^+-\lambda^- \not\to 0$ is clear except when $n_X, n_Y >> m$. Nevertheless, the Main Theorem 3.1.1 assumes proportional values and so avoids this critical case.

7.2.6 Pseudo residual eigenvectors

Knowing the pseudo residual spike, it is not difficult to find the pseudo residual eigenvector. For $s=1,2,...,k$, suppose

$$w_s^{\pm} = u\left(\Sigma_{P_k,X}^{-1/2}\left((\hat{\theta}_{P_k,Y,s}-1)\tilde{u}_{P_k,s}\tilde{u}^t_{P_k,Y,s}\right)\Sigma_{P_k,X}^{-1/2}+\left(\frac{1}{\hat{\theta}_{P_k,X,s}}-1\right)e_se_s^t\right)$$

are the pseudo residual spikes corresponding to the eigenvalues

$$\hat{\zeta}_s^{\pm} = \lambda\left(\Sigma_{P_k,X}^{-1/2}\left((\hat{\theta}_{P_k,Y,s}-1)\tilde{u}_{P_k,s}\tilde{u}^t_{P_k,Y,s}\right)\Sigma_{P_k,X}^{-1/2}+\left(\frac{1}{\hat{\theta}_{P_k,X,s}}-1\right)e_se_s^t\right).$$

We define

$$\begin{aligned}
\zeta_\infty^{\pm}(\theta_s) &= \lim_{m\to\infty}\hat{\zeta}_s^{\pm}, \\
\zeta_\infty^{\pm} &= \lim_{\theta_s,m\to\infty}\hat{\zeta}_s^{\pm}.
\end{aligned}$$

Then,

$$\begin{aligned}
w_{s,s}^{\pm} &= \frac{\sqrt{\hat{\theta}_{P_k,X,s}}}{\mathrm{Norm}_s^{\pm}\sqrt{\hat{\theta}_{P_k,Y,s}-1}\tilde{u}_{P_k,s,s}}\left(\hat{\zeta}_s^{\pm}-\left(\hat{\theta}_{P_k,Y,s}-1\right)\left(1-\hat{\alpha}^2_{P_k,s}\right)\right)+O_p\left(\frac{1}{m}\right) \\
&= \frac{\sqrt{\theta_s}}{\mathrm{Norm}_s^{\pm}\sqrt{\theta_s-1}\alpha_s}\left(\zeta_\infty^{\pm}(\theta_s)-(\theta_s-1)\left(1-\alpha_s^2\right)\right)+O_p\left(\frac{1}{\sqrt{m}}\right) \\
&= \frac{\left(\zeta_\infty^{\pm}-2\left(M_2-1\right)\right)}{\sqrt{\left(\zeta_\infty^{\pm}-2\left(M_2-1\right)\right)^2+2\left(M_2-1\right)}}+O_p\left(\frac{1}{\sqrt{m}}\right)+o_{p;\theta_s}(1),
\end{aligned}$$

$$
\begin{aligned}
w^{\pm}_{s,2:m\setminus s} &= \frac{\sqrt{\hat{\theta}_{P_k,Y,s}-1}\left(\frac{\tilde{u}_{P_k,s,1}}{\sqrt{\hat{\theta}_{P_k,X,1}}},...,\frac{\tilde{u}_{P_k,s,s-1}}{\sqrt{\hat{\theta}_{P_k,X,s-1}}}\frac{\tilde{u}_{P_k,s,s+1}}{\sqrt{\hat{\theta}_{P_k,X,s+1}}},...,\frac{\tilde{u}_{P_k,s,k}}{\sqrt{\hat{\theta}_{P_k,X,k}}},\tilde{u}_{P_k,s,k+1},...,\tilde{u}_{P_k,s,m}\right)}{\mathrm{Norm}^{\pm}_s}, \\
&= \frac{\sqrt{\hat{\theta}_{P_k,Y,1}-1}\left(\frac{\tilde{u}_{P_k,s,1}}{\sqrt{\hat{\theta}_{P_k,X,1}}},...,\frac{\tilde{u}_{P_k,s,s-1}}{\sqrt{\hat{\theta}_{P_k,X,s-1}}},\frac{\tilde{u}_{P_k,s,s+1}}{\sqrt{\hat{\theta}_{P_k,X,s+1}}},...,\frac{\tilde{u}_{P_k,s,k}}{\sqrt{\hat{\theta}_{P_k,X,k}}},\tilde{u}_{P_k,s,k+1},...,\tilde{u}_{P_k,s,m}\right)}{\sqrt{\left(\zeta^{\pm}_{\infty}-2\left(M_2-1\right)\right)^2+2\left(M_2-1\right)+O_p\left(\frac{1}{\sqrt{m}}\right)}+o_{p;\theta_s}(1)},
\end{aligned}
$$

$$
\begin{aligned}
\left(\mathrm{Norm}^{\pm}_s\right)^2 &= \frac{\hat{\theta}_{P_k,X,s}\left(\hat{\zeta}^{\pm}_s-\left(\hat{\theta}_{P_k,Y,s}-1\right)\left(1-\hat{\alpha}^2_{P_k,s}\right)\right)^2}{\left(\hat{\theta}_{P_k,Y,s}-1\right)\tilde{u}^2_{P_k,s,s}}+\left(\hat{\theta}_{P_k,Y,s}-1\right)\left(1-\hat{\alpha}^2_{P_k,s}\right)+O_p\left(\frac{1}{m}\right) \\
&= \frac{\theta_s}{\left(\theta_s-1\right)\alpha^2_s}\left(\zeta^{\pm}_{\infty}\left(\theta_s\right)-\left(\theta_s-1\right)\left(1-\alpha^2_s\right)\right)^2+\left(\theta_s-1\right)\left(1-\alpha^2_s\right)+O_p\left(\frac{1}{\sqrt{m}}\right).
\end{aligned}
$$

We used the fact that the rate convergence of $\hat{\theta}_{P_k,X,s}$, $\hat{\theta}_{P_k,Y,s}$ and $\hat{\alpha}^2_{P_k,s}$ is in $1/\sqrt{m}$. Moreover, when θ is large, $\left(1-\alpha^2_s\right)=\frac{2(M_2-1)}{\theta_s}+O_p\left(1/\theta^2_s\right)$.

7.2.7 Dimension reduction

The previous parts showed that the non-trivial eigenvalues of

$$\hat{\hat{\Sigma}}^{-1/2}_{P_k,X}\hat{\hat{\Sigma}}_{P_k,Y}\hat{\hat{\Sigma}}^{-1/2}_{P_k,X}-\mathrm{I}_m$$

are the same eigenvalues of

$$\sum_{i=1}^{k}\hat{\zeta}^{+}_i w^{+}_i {w^{+}_i}^t+\sum_{i=1}^{k}\hat{\zeta}^{-}_i w^{-}_i {w^{-}_i}^t\in\mathbf{R}^m\times\mathbf{R}^m.$$

We can use 6.2.2 to show that for all non null eigenvalues,

$$\lambda_i\left(\sum_{i=1}^{k}\hat{\zeta}^{+}_i w^{+}_i {w^{+}_i}^t+\sum_{i=1}^{k}\hat{\zeta}^{-}_i w^{-}_i {w^{-}_i}^t\right)=\lambda_i\left(H\right),$$

where

$$
\begin{aligned}
H &= \begin{pmatrix} H^{+} & H^{b} \\ H^{bt} & H^{-}\end{pmatrix}, \\
H^{\pm} &= \begin{pmatrix}
\hat{\zeta}^{\pm}_1 & \sqrt{\hat{\zeta}^{\pm}_1\hat{\zeta}^{\pm}_2}\left\langle w^{\pm}_1,w^{\pm}_2\right\rangle & \sqrt{\hat{\zeta}^{\pm}_1\hat{\zeta}^{\pm}_3}\left\langle w^{\pm}_1,w^{\pm}_3\right\rangle & \cdots & \sqrt{\hat{\zeta}^{\pm}_k\hat{\zeta}^{\pm}_2}\left\langle w^{\pm}_1,w^{\pm}_k\right\rangle \\
\sqrt{\hat{\zeta}^{\pm}_2\hat{\zeta}^{\pm}_1}\left\langle w^{\pm}_2,w^{\pm}_1\right\rangle & \hat{\zeta}^{\pm}_2 & \sqrt{\hat{\zeta}^{\pm}_2\hat{\zeta}^{\pm}_3}\left\langle w^{\pm}_2,w^{\pm}_3\right\rangle & \cdots & \sqrt{\hat{\zeta}^{\pm}_2\hat{\zeta}^{\pm}_k}\left\langle w^{\pm}_2,w^{\pm}_k\right\rangle \\
\sqrt{\hat{\zeta}^{\pm}_3\hat{\zeta}^{\pm}_1}\left\langle w^{\pm}_3,w^{\pm}_1\right\rangle & \sqrt{\hat{\zeta}^{\pm}_3\hat{\zeta}^{\pm}_2}\left\langle w^{\pm}_3,w^{\pm}_2\right\rangle & \hat{\zeta}^{\pm}_3 & \cdots & \sqrt{\hat{\zeta}^{\pm}_3\hat{\zeta}^{\pm}_k}\left\langle w^{\pm}_3,w^{\pm}_k\right\rangle \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
\sqrt{\hat{\zeta}^{\pm}_k\hat{\zeta}^{\pm}_1}\left\langle w^{\pm}_k,w^{\pm}_1\right\rangle & \sqrt{\hat{\zeta}^{\pm}_k\hat{\zeta}^{\pm}_2}\left\langle w^{\pm}_k,w^{\pm}_2\right\rangle & \sqrt{\hat{\zeta}^{\pm}_k\hat{\zeta}^{\pm}_3}\left\langle w^{\pm}_k,w^{\pm}_3\right\rangle & \cdots & \hat{\zeta}^{\pm}_k
\end{pmatrix}, \\
H^{b} &= \begin{pmatrix}
0 & \sqrt{\hat{\zeta}^{+}_1\hat{\zeta}^{-}_2}\left\langle w^{+}_1,w^{-}_2\right\rangle & \sqrt{\hat{\zeta}^{+}_1\hat{\zeta}^{-}_3}\left\langle w^{+}_1,w^{-}_3\right\rangle & \cdots & \sqrt{\hat{\zeta}^{+}_k\hat{\zeta}^{-}_2}\left\langle w^{+}_1,w^{-}_k\right\rangle \\
\sqrt{\hat{\zeta}^{+}_2\hat{\zeta}^{-}_1}\left\langle w^{+}_2,w^{-}_1\right\rangle & 0 & \sqrt{\hat{\zeta}^{+}_2\hat{\zeta}^{-}_3}\left\langle w^{+}_2,w^{-}_3\right\rangle & \cdots & \sqrt{\hat{\zeta}^{+}_2\hat{\zeta}^{-}_k}\left\langle w^{+}_2,w^{-}_k\right\rangle \\
\sqrt{\hat{\zeta}^{+}_3\hat{\zeta}^{-}_1}\left\langle w^{+}_3,w^{-}_1\right\rangle & \sqrt{\hat{\zeta}^{+}_3\hat{\zeta}^{-}_2}\left\langle w^{+}_3,w^{-}_2\right\rangle & 0 & \cdots & \sqrt{\hat{\zeta}^{+}_3\hat{\zeta}^{-}_k}\left\langle w^{+}_3,w^{-}_k\right\rangle \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
\sqrt{\hat{\zeta}^{+}_k\hat{\zeta}^{-}_1}\left\langle w^{+}_k,w^{-}_1\right\rangle & \sqrt{\hat{\zeta}^{+}_k\hat{\zeta}^{-}_2}\left\langle w^{+}_k,w^{-}_2\right\rangle & \sqrt{\hat{\zeta}^{+}_k\hat{\zeta}^{-}_3}\left\langle w^{+}_k,w^{-}_3\right\rangle & \cdots & 0
\end{pmatrix}.
\end{aligned}
$$

Then, we can use Lemma 6.2.4 to argue that

$$\begin{aligned}\lambda_{\max}(H) &= \lambda_{\max}\left(H^{+}\right) + O_p\left(\frac{1}{m}\right),\\ \lambda_{\min}(H) &= \lambda_{\max}\left(H^{-}\right) + O_p\left(\frac{1}{m}\right).\end{aligned}$$

We will see that the covariance between all the entries of H^{+} is null. However, the entries of H^{+} and H^{-} are correlated. Therefore, this step is very useful to avoid the need to study this correlation.

7.2.8 Elements of H

Computation of the distribution of the entries of H requires the distributions of $\hat{\zeta}_i^{\pm}$ and $\left\langle w_i^{\pm}, w_j^{\pm}\right\rangle$, for $i, j = 1, 2, ..., k$ with $j \neq i$. By the pseudo Invariance Section 7.2.5,

$$\hat{\zeta}_i^{\pm} = \lambda^{\pm}\left(\hat{\hat{\Sigma}}_{\tilde{P}_i,X}^{-1/2}\hat{\hat{\Sigma}}_{\tilde{P}_i,Y}\hat{\hat{\Sigma}}_{\tilde{P}_i,X}^{-1/2}\right) - 1 + O_p\left(\frac{1}{m}\right).$$

Therefore, using the Section 7.2.3, we obtain the two first moments of the diagonal elements.
The off-diagonal terms are more difficult to estimate and we assume that Assumptions 2.2.2(A2) and (A3) hold.
First, we will express $\left\langle w_i^{\pm}, w_j^{\pm}\right\rangle$ as a function of the usual statistics when all the eigenvalues are of order θ. Then, we will compute its two first moments. Finally, a small argument similar to Lemma 6.2.4 leads to a result for all perturbations.

Remark 7.2.4.
Using Theorem 5.11.3, the reader can prove that the off diagonal terms are of order $O_p(1/\sqrt{m})$ when at least one eigenvalue θ_i or θ_j is finite (Assumption 2.2.2(A4)).

Formula Suppose $k > 1$ and

$$w_i^{\pm} = u\left(\Sigma_{P_k,X}^{-1/2}\left((\hat{\hat{\theta}}_{P_k,Y,i} - 1)\tilde{u}_{P_k,i}\tilde{u}_{P_k,Y,i}^{t}\right)\Sigma_{P_k,X}^{-1/2} + \left(\frac{1}{\hat{\hat{\theta}}_{P_k,X,i}} - 1\right)e_i e_i^t\right)$$

We want to prove the following formula :

$$\begin{aligned}\left\langle w_s^{\pm}, w_t^{\pm}\right\rangle &= \frac{\sqrt{\theta_s\theta_t}}{\left(\zeta_\infty^{\pm} - 2M_2 + 1\right)^2 + 2\left(M_2 - 1\right)} \times\\ &\quad\left(\sum_{p=k+1}^{m}\hat{u}_{P_k,Y,s,p}\hat{u}_{P_k,Y,t,p} + \sum_{p=k+1}^{m}\hat{u}_{P_k,X,s,p}\hat{u}_{P_k,X,t,p}\right.\\ &\qquad - \sum_{p=k+1}^{m}\hat{u}_{P_k,Y,s,p}\hat{u}_{P_k,X,t,p} - \sum_{p=k+1}^{m}\hat{u}_{P_k,Y,t,p}\hat{u}_{P_k,X,s,p}\\ &\qquad\left. - \left(\hat{u}_{P_k,X,t,s} + \hat{u}_{P_k,Y,s,t}\right)\left(\tilde{\alpha}_s^2 - \tilde{\alpha}_t^2 - \left(\zeta_\infty^{\pm} - 2\left(M_2 - 1\right)\right)\left(\frac{1}{\theta_t} - \frac{1}{\theta_s}\right)\right)\right)\\ &\quad + o_{p;m,\theta}\left(\frac{1}{m^{1/2}}\right),\end{aligned}$$

where $\lim_{m,\theta\to\infty}\frac{o_{p;m,\theta}\left(\frac{1}{m^{1/2}}\right)}{\frac{1}{m^{1/2}}} = 0$ with probability tending to 1.

- In section Pseudo residual eigenvector, 7.2.6 we proved that assuming $\epsilon = O_p\left(\frac{1}{m^{1/2}}\right) + o_{p;\theta}(1)$,

$$
\begin{aligned}
(w_{s,s})^{\pm} &= \frac{\left(\zeta_{\infty}^{\pm} - 2(M_2-1)\right)}{\sqrt{\left(\zeta_{\infty}^{\pm} - 2(M_2-1)\right)^2 + 2(M_2-1)}} + \epsilon, \\
w_{s,k+1:m}^{\pm} &= \frac{\sqrt{\hat{\theta}_{P_k,Y,s} - 1}\tilde{u}_{P_k,s,k+1:m}}{\sqrt{\left(\zeta_{\infty}^{\pm} - 2(M_2-1)\right)^2 + 2(M_2-1)} + \epsilon}, \\
w_{s,1:k\backslash s}^{\pm} &= \frac{\sqrt{\hat{\theta}_{P_k,Y,s} - 1}\left(\frac{\tilde{u}_{P_k,s,1}}{\sqrt{\hat{\theta}_{P_k,X,1}}}, ..., \frac{\tilde{u}_{P_k,s,s-1}}{\sqrt{\hat{\theta}_{P_k,X,s-1}}}, \frac{\tilde{u}_{P_k,s,s+1}}{\sqrt{\hat{\theta}_{P_k,X,s+1}}}, ..., \frac{\tilde{u}_{P_k,s,k}}{\sqrt{\hat{\theta}_{P_k,X,k}}}\right)}{\sqrt{\left(\zeta_{\infty}^{\pm} - 2(M_2-1)\right)^2 + 2(M_2-1)} + \epsilon}.
\end{aligned}
$$

- Then, by Theorem 5.11.3,

$$
\tilde{u}_{P_k,s,t} = \hat{u}_{P_k,X,t,s} + \hat{u}_{P_k,Y,s,t} + O_p\left(\frac{1}{m}\right) + O_p\left(\frac{1}{\theta m^{1/2}}\right).
$$

First, we set $b = +$ or $b = -$ and separate the scalar product in three parts.

$$
\left\langle w_s^b, w_t^b \right\rangle = \underbrace{\sum_{i=s,t} w_{s,i}^b w_{t,i}^b}_{3)} + \underbrace{\sum_{i \neq s,t}^{k} w_{s,i}^b w_{t,i}^b}_{1)} + \underbrace{\sum_{i=k+1}^{m} w_{s,i}^b w_{t,i}^b}_{2)}.
$$

1) If $k = 2$, the second term does not exist. However, if $k > 2$, then asymptotically for $i = 1, ..., k$, $i \neq s, t$,

$$
\begin{aligned}
w_{s,i}^b w_{t,i}^b &= \frac{\sqrt{(\hat{\theta}_{P_k,Y,s} - 1)(\hat{\theta}_{P_k,Y,t} - 1)}}{\left(\zeta_{\infty}^{b} - 2(M_2-1)\right)^2 + 2(M_2-1) + \epsilon} \frac{\tilde{u}_{P_k,s,i}}{\sqrt{\hat{\theta}_{P_k,X,t}}} \frac{\tilde{u}_{P_k,t,i}}{\sqrt{\hat{\theta}_{P_k,X,s}}} \\
&= O_p\left(\frac{1}{m}\right).
\end{aligned}
$$

2) By Theorem 5.11.3,

$$
\begin{aligned}
\sum_{i=k+1}^{m} \tilde{u}_{j,i}\tilde{u}_{t,i} &= \sum_{i=k+1}^{m} \hat{u}_{\hat{\Sigma}_Y,j,i}\hat{u}_{\hat{\Sigma}_Y,t,i} + \sum_{i=k+1}^{m} \hat{u}_{\hat{\Sigma}_X,j,i}\hat{u}_{\hat{\Sigma}_X,t,i} - \sum_{i=k+1}^{m} \hat{u}_{\hat{\Sigma}_X,j,i}\hat{u}_{\hat{\Sigma}_Y,t,i} \\
&\quad - \sum_{i=k+1}^{m} \hat{u}_{\hat{\Sigma}_Y,j,i}\hat{u}_{\hat{\Sigma}_X,t,i} - \left(\hat{u}_{\hat{\Sigma}_X,t,j} + \hat{u}_{\hat{\Sigma}_Y,j,t}\right)\left(\hat{\alpha}_{\hat{\Sigma}_X,j}^2 - \hat{\alpha}_{\hat{\Sigma}_X,t}^2\right) \\
&\quad + O_p\left(\frac{1}{\theta^{1/2} m}\right) + O_p\left(\frac{1}{\theta m^{1/2}}\right).
\end{aligned}
$$

Therefore,

$$
\sum_{i=k+1}^{m} w_{s,i}^b w_{t,i}^b = \frac{\sqrt{\hat{\theta}_{P_k,Y,s} - 1}\sqrt{\hat{\theta}_{P_k,Y,t} - 1}}{\sqrt{\left(\zeta_{\infty}^{b} - 2(M_2-1)\right)^2 + 2(M_2-1)} + \epsilon} \sum_{p=k+1}^{m} \tilde{u}_{P_k,s,p}\tilde{u}_{P_k,t,p}.
$$

3) Asymptotically,

$$\begin{aligned}
w^b_{s,s}w^b_{t,s} + w^b_{s,t}w^b_{t,t} &= \frac{\zeta^b_\infty - 2(M_2-1)+\epsilon}{\left(\zeta^b_\infty - 2(M_2-1)\right)^2 + 2(M_2-1)}\left(w_{t,s}+w_{s,t}\right) \\
&= \frac{\zeta^b_\infty - 2(M_2-1)}{\left(\zeta^b_\infty - 2(M_2-1)\right)^2 + 2(M_2-1)}\left(\frac{\sqrt{\theta_t}}{\sqrt{\theta_s}}\tilde{u}_{P_k,t,s} + \frac{\sqrt{\theta_s}}{\sqrt{\theta_t}}\tilde{u}_{P_k,s,t}\right) + O_p\left(\frac{1}{m^{1/2}}\right)\epsilon \\
&= \frac{\left(\zeta^b_\infty - 2(M_2-1)\right)\sqrt{\theta_s\theta_t}}{\left(\zeta^b_\infty - 2(M_2-1)\right)^2 + 2(M_2-1)} \\
&\quad \left(\frac{1}{\theta_s}\left(\hat{u}_{P_k,X,s,t}+\hat{u}_{P_k,Y,t,s}\right) + \frac{1}{\theta_t}\left(\hat{u}_{P_k,X,t,s}+\hat{u}_{P_k,Y,s,t}\right)\right) + O_p\left(\frac{1}{m^{1/2}}\right)\epsilon \\
&= \frac{\left(\zeta^b_\infty - 2(M_2-1)\right)\sqrt{\theta_s\theta_t}}{\left(\zeta^b_\infty - 2(M_2-1)\right)^2 + 2(M_2-1)} \\
&\quad \left(\frac{1}{\theta_t} - \frac{1}{\theta_s}\right)\left(\hat{u}_{P_k,X,t,s}+\hat{u}_{P_k,Y,s,t}\right) + o_{p;\theta,m}\left(\frac{1}{m^{1/2}}\right).
\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}
\left\langle w^b_s, w^b_t\right\rangle &= \frac{\sqrt{\theta_s\theta_t}}{\left(\zeta^b_\infty - 2M_2+1\right)^2 + 2(M_2-1)} \times \\
&\quad \left(\sum_{p=k+1}^m \hat{u}_{P_k,Y,s,p}\hat{u}_{P_k,Y,t,p} + \sum_{p=k+1}^m \hat{u}_{P_k,X,s,p}\hat{u}_{P_k,X,t,p}\right. \\
&\quad - \sum_{p=k+1}^m \hat{u}_{P_k,Y,s,p}\hat{u}_{P_k,X,t,p} - \sum_{p=k+1}^m \hat{u}_{P_k,Y,t,p}\hat{u}_{P_k,X,s,p} \\
&\quad \left. - \left(\hat{u}_{P_k,X,t,s}+\hat{u}_{P_k,Y,s,t}\right)\left(\tilde{\alpha}^2_t - \tilde{\alpha}^2_s - \left(\zeta^b_\infty - 2(M_2-1)\right)\left(\frac{1}{\theta_t}-\frac{1}{\theta_s}\right)\right)\right) \\
&\quad + o_{p;\theta,m}\left(\frac{1}{m^{1/2}}\right).
\end{aligned}$$

Moment We separate the formula into three parts

1. $\sum_{p=k+1}^m \hat{u}_{P_k,Y,s,p}\hat{u}_{P_k,Y,t,p} - \hat{u}_{P_k,Y,s,t}\left(\tilde{\alpha}^2_t - \tilde{\alpha}^2_s - \left(\zeta^b_\infty - 2(M_2-1)\right)\left(\frac{1}{\theta_t}-\frac{1}{\theta_s}\right)\right)$

2. $\sum_{p=k+1}^m \hat{u}_{P_k,X,s,p}\hat{u}_{P_k,X,t,p} + \hat{u}_{P_k,X,s,t}\left(\tilde{\alpha}^2_t - \tilde{\alpha}^2_s - \left(\zeta^b_\infty - 2(M_2-1)\right)\left(\frac{1}{\theta_t}-\frac{1}{\theta_s}\right)\right)$

3. $\sum_{p=k+1}^m \hat{u}_{P_k,Y,s,p}\hat{u}_{P_k,X,t,p} + \sum_{p=k+1}^m \hat{u}_{P_k,X,s,p}\hat{u}_{P_k,Y,t,p}$

Without loss of generality, we present the proof for $s=1$ and $t=2$.
In order to compute the moments of the first and second parts, we use the remark of Theorem 5.6.1,

$$\begin{aligned}
&\hat{u}_{P_2,1,2}\left(\frac{1}{\theta_1}-\frac{1}{\theta_2}\right)\delta + \sum_{i=3}^m \hat{u}_{P_2,1,i}\hat{u}_{P_2,2,i} \\
&\quad \sim \mathbf{N}\left(0, \frac{(1+M_2+\delta)^2(M_2-1) + \left(M_4-(M_2)^2\right) - 2(1+M_2+\delta)(M_3-M_2)}{\theta_1\theta_2 m}\right) \\
&\qquad + O_p\left(\frac{1}{\theta m}\right) + O_p\left(\frac{1}{\theta^2 m^{1/2}}\right).
\end{aligned}$$

1.
$$\begin{aligned}&\sum_{p=k+1}^{m} \hat{u}_{P_k,Y,1,p}\hat{u}_{P_k,Y,2,p} - \hat{u}_{P_k,Y,1,2}\left(\tilde{\alpha}_2^2 - \tilde{\alpha}_1^2 - \left(\zeta_\infty^b - 2\left(M_2-1\right)\right)\left(\frac{1}{\theta_2}-\frac{1}{\theta_1}\right)\right)\\ &= \sum_{p=k+1}^{m} \hat{u}_{P_k,Y,1,p}\hat{u}_{P_k,Y,2,p} + \hat{u}_{P_k,Y,1,2}\left(-M_{2,X} - \zeta_\infty^b + 2M_2 - 1\right)\left(\frac{1}{\theta_1}-\frac{1}{\theta_2}\right) + O_p\left(\frac{1}{\theta^2\sqrt{m}}\right).\end{aligned}$$

Using the remark, we set $\delta = -M_{2,X} + 2M_2 - 1 - \lambda^b$,

$$\begin{aligned}&\sum_{p=k+1}^{m} \hat{u}_{P_k,Y,1,p}\hat{u}_{P_k,Y,2,p} - \hat{u}_{P_k,Y,1,2}\left(\tilde{\alpha}_2^2 - \tilde{\alpha}_1^2 - \left(\zeta_\infty^b - 2\left(M_2-1\right)\right)\left(\frac{1}{\theta_2}-\frac{1}{\theta_1}\right)\right) + O_p\left(\frac{1}{\theta m}\right) + O_p\left(\frac{1}{\theta^2 m^{1/2}}\right)\\ &= \mathbf{N}\left(0, \frac{\left(1+M_2+M_{2,Y}-M_{2,X}\mp\sqrt{M_2^2-1}\right)^2\left(M_2-1\right) + \left(M_4-(M_2)^2\right) - 2\left(1+M_2+M_{2,Y}-M_{2,X}\mp\sqrt{M_2^2-1}\right)\left(M_3-M_2\right)}{\theta_1\theta_2 m}\right)\end{aligned}$$

2. Similar computation leads to

$$\begin{aligned}&\sum_{p=k+1}^{m} \hat{u}_{P_k,X,1,p}\hat{u}_{P_k,X,2,p} + \hat{u}_{P_k,X,1,2}\left(\tilde{\alpha}_2^2 - \tilde{\alpha}_1^2 - \left(\zeta_\infty^b - 2\left(M_2-1\right)\right)\left(\frac{1}{\theta_2}-\frac{1}{\theta_1}\right)\right) + O_p\left(\frac{1}{\theta m}\right) + O_p\left(\frac{1}{\theta^2 m^{1/2}}\right)\\ &= \mathbf{N}\left(0, \frac{\left(1-M_2+2M_{2,X}\pm\sqrt{M_2^2-1}\right)^2\left(M_2-1\right) + \left(M_4-(M_2)^2\right) - 2\left(1-M_2+2M_{2,X}\pm\sqrt{M_2^2-1}\right)\left(M_3-M_2\right)}{\theta_1\theta_2 m}\right).\end{aligned}$$

3. We easily show that

$$\begin{aligned}\sum_{p=k+1}^{m} \hat{u}_{P_k,Y,1,p}\hat{u}_{P_k,X,2,p} + \sum_{p=k+1}^{m} \hat{u}_{P_k,X,1,p}\hat{u}_{P_k,Y,2,p} &= \mathrm{RV}\left(0, \frac{2\left(M_{2,X}-1\right)\left(M_{2,Y}-1\right)}{\theta_1\theta_1 m}\right)\\ &\quad + O_p\left(\frac{1}{\theta m}\right) + O_p\left(\frac{1}{\theta^2 m^{1/2}}\right)\end{aligned}$$

Indeed the covariance between the first and the second term is negligible. This can be shown using the independence between X and Y, and Theorem 5.6.1.

$$\begin{aligned}&\mathrm{Cov}\left(\sum_{p=k+1}^{m} \hat{u}_{P_k,Y,1,p}\hat{u}_{P_k,X,2,p}, \sum_{p=k+1}^{m} \hat{u}_{P_k,X,1,p}\hat{u}_{P_k,Y,2,p}\right)\\ &\quad = \sum_{p=k+1}^{m} \mathrm{E}\left[\hat{u}_{P_k,Y,1,p}\hat{u}_{P_k,Y,2,p}\right] \sum_{p=k+1}^{m} \mathrm{E}\left[\hat{u}_{P_k,X,1,p}\hat{u}_{P_k,X,2,p}\right] = o_p\left(\frac{1}{\theta^2 m}\right).\end{aligned}$$

Because the null covariance between the three parts is easily proven by the independence between X and Y, we conclude:

$$\begin{aligned}\left\langle w_i^b, w_j^b\right\rangle &\sim \mathrm{RV}\left(0, \frac{1}{m}\frac{2(M_{2,X}-1)(M_{2,Y}-1) + B_X^b + B_Y^b}{\left((\zeta_\infty^\pm - 2M_2+1)^2 + 2(M_2-1)\right)^2}\right) + O_p\left(\frac{1}{\theta m}\right) + O_p\left(\frac{1}{\theta^2 m^{1/2}}\right),\\ B_X^+ &= \left(1 - M_2 + 2M_{2,X} + \sqrt{M_2^2-1}\right)^2\left(M_{2,X}-1\right)\\ &\qquad +2\left(-1+M_2-2M_{2,x}-\sqrt{M_2^2-1}\right)\left(M_{3,X}-M_{2,X}\right) + \left(M_{4,X}-M_{2,X}^2\right),\\ B_Y^+ &= \left(1+M_2+M_{2,Y}-M_{2,X}-\sqrt{M_2^2-1}\right)^2\left(M_{2,Y}-1\right)\\ &\qquad +2\left(-1-M_2-M_{2,Y}-M_{2,X}-\sqrt{M_2^2-1}\right)\left(M_{3,Y}-M_{2,Y}\right) + \left(M_{4,Y}-M_{2,Y}^2\right),\end{aligned}$$

$$B_X^- = \left(1 - M_2 + 2M_{2,X} - \sqrt{M_2^2 - 1}\right)^2 (M_{2,X} - 1)$$
$$+2\left(-1 + M_2 - 2M_{2,x} + \sqrt{M_2^2 - 1}\right)(M_{3,X} - M_{2,X}) + (M_{4,X} - M_{2,X}^2),$$
$$B_Y^- = \left(1 + M_2 + M_{2,Y} - M_{2,X} + \sqrt{M_2^2 - 1}\right)^2 (M_{2,Y} - 1)$$
$$+2\left(-1 - M_2 - M_{2,Y} + M_{2,X} - \sqrt{M_2^2 - 1}\right)(M_{3,Y} - M_{2,Y}) + (M_{4,Y} - M_{2,Y}^2),$$

where $b = +$ or $b = -$.

7.2.9 Normality discussion

Assuming $n_X >> n_Y$, the normality is straightforward to prove using 5.11.2 and 5.11.3. Nevertheless, when $n_X \sim n_Y$, new marginally normal statistics enter in the formula. These statistics are

$$\sum_{p=k+1}^{m} \hat{u}_{P_k,Y,s,p}\hat{u}_{P_k,X,t,p} + \sum_{p=k+1}^{m} \hat{u}_{P_k,X,s,p}\hat{u}_{P_k,Y,t,p}.$$

Despite this difficulty, the reader can check that assuming large θ, asymptotic joint normality of the entries of H is equivalent to asymptotic joint normality of

$$W_{X,s,t}, W_{Y,s,t}, \left({W_X}^2\right)_{s,t}, \left({W_Y}^2\right)_{s,t} \text{ and } \frac{1}{\sqrt{m}} \sum_{i=k+1}^{m} W_{X,s,i}W_{Y,t,i}$$

for $s, t = 1, 2, ..., k$. Note that joint normality holds for the first four elements by Theorem 5.11.2. The part left to the reader is nearly done. Theorem 5.6.1 and Lemma 6.1.1 already showed that nearly all the statistics that composed H are functions of the first four elements. Then, a similar proof to Theorem 5.10.1 proves that for $s, t = 1, 2, ..., k$,

$$\sum_{p=k+1}^{m} \hat{u}_{P_k,Y,s,p}\hat{u}_{P_k,X,t,p}$$

can be expressed in function of the statistics.

7.3 Proof of the robust theorems

In this part we use the moment in probability E_p and Var_p defined in 4.2.1.

7.3.1 Preliminary proofs

Before proving the theorems concerning the robustness, we need to prove the following small Lemma that characterises the residual spike when θ is large.

Lemma 6.3.1.
Suppose $\mathbf{X} \in \mathbb{R}^{m \times n_X}$ and $\mathbf{Y} \in \mathbb{R}^{m \times n_Y}$ are members of the same class, $\mathcal{C}_A(\Sigma)$, $\mathcal{C}_B$, $\mathcal{C}_C$ or $\mathcal{C}_D(u)$. Moreover, define $W_X = \frac{1}{n_X}\mathbf{X}\mathbf{X}^t$ and $\hat{\Sigma}_{\mathbf{X}} = \frac{1}{n_X}\mathbf{X}\mathbf{X}^t$, where $\mathbf{X} = P^{1/2}\mathbf{X}$, and P is a finite perturbation of order 1. (Without loss of generality we can assume a canonical perturbation, but $\hat{u}_{\mathbf{X},1:m,1}$ is not necessarily uniform. In particular the following formulas are only true for $\mathcal{C}_D(u)$ after a rotation to change u in e_1.)
The largest residual spike obtained using $\mathbf{X}$ and $\mathbf{Y}$ is

$$\lambda_{m,\mathcal{C}} = \frac{1}{2}\left(\hat{\hat{\theta}}_Y(1-\hat{\alpha}^2) + 1 + \frac{\hat{\hat{\theta}}_Y}{\hat{\hat{\theta}}_X} + \sqrt{-4\frac{\hat{\hat{\theta}}_Y}{\hat{\hat{\theta}}_X} + \left(\hat{\hat{\theta}}_Y(1-\hat{\alpha}^2) + \frac{\hat{\hat{\theta}}_Y}{\hat{\hat{\theta}}_X} + 1\right)^2}\right) + O_p\left(\frac{1}{\theta}\right),$$

where

$$\frac{1}{\theta-1} = \sum_{i=1}^{m} \frac{\hat{\lambda}_{W_X,i}}{\hat{\theta}_X - \hat{\lambda}_{W_X,i}} \hat{u}^2_{W_X,i,1},$$
$$\frac{1}{\hat{\hat{\theta}}_X - 1} = \frac{1}{m}\sum_{i=1}^{m} \frac{\hat{\lambda}_{W_X,i}}{\hat{\theta}_X - \hat{\lambda}_{W_X,i}},$$
$$\hat{\hat{\theta}}_X = \theta \sum_{i=1}^{m} \hat{\lambda}_{W_X,i}\hat{u}^2_{W_X,1,i} + O_p(1),$$
$$\hat{\alpha}^2_X = \frac{\theta}{(\theta-1)}\left(1 - \frac{\theta-1}{\hat{\theta}^2_X}\sum_{i=1}^{m}\hat{\lambda}^2_{W_X,i}\hat{u}^2_{W_X,i,1}\right) + O_p\left(\frac{1}{\theta^2}\right),$$
$$\hat{\alpha} = \hat{\alpha}_X\hat{\alpha}_Y + \sum_{i=2}^{m}\hat{u}_{\hat{\Sigma}_{\mathbf{X}},i}\hat{u}_{\hat{\Sigma}_{\mathbf{Y}},i} + O_p\left(\frac{1}{\theta^2}\right),$$

$\hat{u}_{W_X,1,i}$ *is the first entry of the* i^{th} *eigenvector of* W_X *and* $\hat{\theta}_X = \hat{\theta}_{\hat{\Sigma}_{\mathbf{X}},1}$ *is the largest eigenvalue of* $\hat{\Sigma}_{\mathbf{X}}$*. Moreover for class* $\mathcal{C}_A(\Sigma)$*,* $\mathcal{C}_B$ *and* $\mathcal{C}_C$*,*

$$\begin{aligned}
\frac{\hat{\hat{\theta}}_Y}{\hat{\hat{\theta}}_X} &= 1 + O_p\left(\frac{1}{\theta}\right) + \left(\frac{1}{m^{1/2}}\right), \\
\hat{\hat{\theta}}_Y(1-\hat{\alpha}^2) &= \frac{1}{\Sigma_{1,1}}\left(\left({W_X}^2\right)_{1,1} + \left({W_Y}^2\right)_{1,1}\right) - \frac{2}{\Sigma_{1,1}}\left(\Sigma^2\right)_{1,1} + O_p\left(\frac{1}{\theta}\right) + O_p\left(\frac{1}{m^{1/2}}\right).
\end{aligned}$$

(Page 93)

Proof. **Lemma 6.3.1**

First, we study the limit of the expectation of the residual spike when θ is large. Using Lemma 6.2.1,

$$\begin{aligned}
\lambda_{m,\mathcal{C}} &= \frac{1}{2}\left(\hat{\hat{\theta}}_Y + \hat{\alpha}^2 - \hat{\hat{\theta}}_Y\hat{\alpha}^2 + \frac{1 + (\hat{\hat{\theta}}_Y - 1)\hat{\alpha}^2 + \sqrt{-4\hat{\hat{\theta}}_Y\hat{\hat{\theta}}_X + \left(1 + \hat{\hat{\theta}}_Y\hat{\hat{\theta}}_X - (\hat{\hat{\theta}}_Y - 1)(\hat{\hat{\theta}}_X - 1)\hat{\alpha}^2\right)^2}}{\hat{\hat{\theta}}_X}\right) \\
&= \frac{1}{2}\left(\hat{\hat{\theta}}_Y(1-\hat{\alpha}^2) + 1 + \frac{\hat{\hat{\theta}}_Y}{\hat{\hat{\theta}}_X} + \sqrt{-4\frac{\hat{\hat{\theta}}_Y}{\hat{\hat{\theta}}_X} + \left(\hat{\hat{\theta}}_Y(1-\hat{\alpha}^2) + \frac{\hat{\hat{\theta}}_Y}{\hat{\hat{\theta}}_X} + 1\right)^2}\right) + O_p\left(\frac{1}{\theta}\right).
\end{aligned}$$

The idea of the proof consists of looking at the expectation of $\frac{\hat{\hat{\theta}}_Y}{\hat{\hat{\theta}}_X}$ and $\hat{\hat{\theta}}_Y(1-\hat{\alpha}^2)$.

1: By Theorem 5.3.1,

$$\frac{\hat{\hat{\theta}}_Y}{\hat{\hat{\theta}}_X} = \frac{\theta\sum_{i=1}^{m}\hat{\lambda}_{W_Y,i}\hat{u}^2_{W_Y,1,i} + O_p(1)}{\theta\sum_{i=1}^{m}\hat{\lambda}_{W_X,i}\hat{u}^2_{W_X,1,i} + O_p(1)} = \frac{W_{Y,1,1}}{W_{X,1,1}} + O_p(1/\theta) = 1 + O_p\left(\frac{1}{\theta}\right) + O_p\left(\frac{1}{m^{1/2}}\right).$$

2: By Theorem 5.3.1,

$$\begin{aligned}
\hat{\alpha}_X^2\hat{\alpha}_Y^2 &= \frac{\theta^2}{(\theta-1)^2}\left(1-\frac{\theta-1}{\theta^2\left(W_{X,1,1}\right)^2}\left({W_X}^2\right)_{,1,1}\right)\left(1-\frac{\theta-1}{\theta^2\left(W_{Y,1,1}\right)^2}\left({W_Y}^2\right)_{1,1}\right)+O_p\left(\frac{1}{\theta^2}\right)\\
&= \left(1+\frac{2}{\theta}-\frac{1}{(\theta-1)\left(W_{X,1,1}\right)^2}\left({W_X}^2\right)_{,1,1}-\frac{1}{(\theta-1)\left(W_{Y,1,1}\right)^2}\left({W_Y}^2\right)_{1,1}\right)+O_p\left(\frac{1}{\theta^2}\right)\\
&= \left(1+\frac{2}{\theta}-\frac{1}{(\theta-1)\left(\Sigma_{1,1}\right)^2}\left(\left({W_X}^2\right)_{,1,1}+\left({W_Y}^2\right)_{1,1}\right)\right)+O_p\left(\frac{1}{\theta^2}\right)+O_p\left(\frac{1}{\theta m^{1/2}}\right),\\
\hat{\alpha}_X\hat{\alpha}_Y &= \left(1+\frac{1}{\theta}-\frac{1}{(\theta-1)\left(\Sigma_{1,1}\right)^2}\frac{\left({W_X}^2\right)_{,1,1}+\left({W_Y}^2\right)_{1,1}}{2}\right)+O_p\left(\frac{1}{\theta^2}\right)+O_p\left(\frac{1}{\theta m^{1/2}}\right).
\end{aligned}$$

3: We can estimate $\sum_{i=2}^m \hat{u}_{\hat{\Sigma}_{\mathbf{X}},i}\hat{u}_{\hat{\Sigma}_{\mathbf{Y}},i}$ as in the proof of Lemma 6.1.1. Nevertheless, this previous proof does not directly solve the problem. Indeed, the term $\left(W^2\right)_{1,2}$ is not of order $1/\sqrt{m}$ anymore. The reader can show that the following extended result remains true. For $s = 2,3,...,m$, there exist B_1 and B_2 two finite constants such that

$$\left|\hat{u}_{\hat{\Sigma}_{\mathbf{X}},i,s}-\frac{W_{X,1,s}}{\theta^{1/2}W_{X,1,1}}\right|<\frac{B_1W_{X,1,s}}{\theta^{3/2}}+\frac{B_2\left({W_X}^2\right)_{1,2}}{\theta^{3/2}}.$$

Thus,

$$\begin{aligned}
\sum_{i=2}^m \hat{u}_{\hat{\Sigma}_{\mathbf{X}},i}\hat{u}_{\hat{\Sigma}_{\mathbf{Y}},i} &= \frac{1}{\theta(W_{X,1,1}W_{Y,1,1})}\sum_{i=2}^m\left(\Sigma_{1,s}\right)^2+O_p\left(\frac{1}{\theta^2}\right)\\
&= \frac{1}{\theta(\Sigma_{1,1})^2}\sum_{i=2}^m\left(\Sigma_{1,s}\right)^2+O_p\left(\frac{1}{\theta^2}\right).
\end{aligned}$$

This last result follows from bounded spectra of W_X and W_Y, which imply the following bounded quantities,

$$\begin{aligned}
\left(\sum_{s=2}^m W_{X,1,s}\left({W_Y}^2\right)_{1,s}\right)^2 &\leqslant \left({W_X}^2\right)_{1,1}\left({W_Y}^4\right)_{1,1},\\
\left(\sum_{s=2}^m \left({W_X}^2\right)_{1,s}W_{Y,1,s}\right)^2 &\leqslant \left({W_X}^4\right)_{1,1}\left({W_Y}^2\right)_{1,1},\\
\left(\sum_{s=2}^m \left({W_X}^2\right)_{1,s}\left({W_Y}^2\right)_{1,s}\right)^2 &\leqslant \left({W_X}^4\right)_{1,1}\left({W_Y}^4\right)_{1,1}.
\end{aligned}$$

Despite the fact that the error could be large for some Σ, when θ tends to infinity, it is small.

4:

$$\begin{aligned}
\hat{\alpha} &= \left(1+\frac{1}{\theta}-\frac{1}{\theta\left(\Sigma_{1,1}\right)^2}\left(\frac{\left({W_X}^2\right)_{,1,1}+\left({W_Y}^2\right)_{1,1}}{2}+\sum_{i=2}^m\left(\Sigma_{1,s}\right)^2\right)\right)+O_p\left(\frac{1}{\theta^2}\right)+O_p\left(\frac{1}{\theta m^{1/2}}\right),\\
\hat{\alpha}^2 &= \left(1+\frac{2}{\theta}-\frac{2}{\theta\left(\Sigma_{1,1}\right)^2}\left(\frac{\left({W_X}^2\right)_{,1,1}+\left({W_Y}^2\right)_{1,1}}{2}+\sum_{i=2}^m\left(\Sigma_{1,s}\right)^2\right)\right)+O_p\left(\frac{1}{\theta^2}\right)+O_p\left(\frac{1}{\theta m^{1/2}}\right).
\end{aligned}$$

Finally, because $\hat{\hat{\theta}}_Y=\theta W_{Y,1,1}+O_p(1)=\theta\Sigma_{1,1}+O_p(1)+O_p\left(m^{-1/2}\right)$,

$$\begin{aligned}
\hat{\hat{\theta}}_Y(1-\hat{\alpha}^2) &= -2\Sigma_{1,1}+\frac{1}{\Sigma_{1,1}}\left(\left({W_X}^2\right)_{,1,1}+\left({W_Y}^2\right)_{1,1}\right)-\frac{2}{\Sigma_{1,1}}\sum_{i=2}^m\left(\Sigma_{1,s}\right)^2+O_p\left(\frac{1}{\theta}\right)+O_p\left(\frac{1}{m^{1/2}}\right)\\
&= \frac{1}{\Sigma_{1,1}}\left(\left({W_X}^2\right)_{,1,1}+\left({W_Y}^2\right)_{1,1}\right)-\frac{2}{\Sigma_{1,1}}\left(\Sigma^2\right)_{1,1}+O_p\left(\frac{1}{\theta}\right)+O_p\left(\frac{1}{m^{1/2}}\right).
\end{aligned}$$

□

Then, we prove four lemmas necessary to prove 4.2.1.

Lemma 6.3.2.
In the class $\mathcal{C}_A(\Sigma)$ defined in Definition 4.2.3, the limit of the expectation of the residual spike is invariant of Σ. This means that:
Suppose $\mathbf{X}_\Sigma, \mathbf{Y}_\Sigma \in \mathcal{C}_A(\Sigma)$ and $\mathbf{X}_{\mathrm{I}_m}, \mathbf{Y}_{\mathrm{I}_m} \in \mathcal{C}_A(\mathrm{I}_m)$ such that

$$\mathbf{X}_\Sigma = P^{1/2}\Sigma^{1/2}\mathbf{X} \in \mathbb{R}^{m\times n_X} \text{ and } \mathbf{Y}_\Sigma = P^{1/2}\Sigma^{1/2}\mathbf{Y} \in \mathbb{R}^{m\times n_Y},$$
$$\mathbf{X}_{\mathrm{I}_m} = P^{1/2}\mathbf{X} \in \mathbb{R}^{m\times n_X} \text{ and } \mathbf{Y}_{\mathrm{I}_m} = P^{1/2}\mathbf{Y} \in \mathbb{R}^{m\times n_Y}.$$

Let $\lambda_{m,\Sigma}$ and λ_{m,I_m} be the resulting residual spikes using $\left(\mathbf{X}_\Sigma, \mathbf{Y}_\Sigma\right)$ and $\left(\mathbf{X}_{\mathrm{I}_m}, \mathbf{Y}_{\mathrm{I}_m}\right)$ respectively. We have

$$\mathrm{E}_{p(\min(m,\theta))}\left[\lambda_{m,\Sigma}\right] = \mathrm{E}_{p(\min(m,\theta))}\left[\lambda_{m,\mathrm{I}_m}\right].$$

(Page 94)

Proof. **Lemma 6.3.2**

Using 6.3.1, we just need to show

$$\mathrm{E}\left[T_{\Sigma,X}\right] = \mathrm{E}\left[T_{\mathrm{I}_m,X}\right] + o_{\min(m,\theta)}(1),$$

where

$$\begin{aligned}
T_{\Sigma,X} &= \frac{1}{\Sigma_{1,1}}\left({W_X}^2\right)_{1,1} - \frac{(\Sigma^2)_{1,1}}{\Sigma_{1,1}},\\
W_X &= \frac{1}{n_X}\Sigma^{1/2}\mathbf{X}\mathbf{X}^t\Sigma^{1/2},\\
T_{\mathrm{I}_m,X} &= \left({W_X}^2\right)_{1,1} - 1,\\
W_X &= \frac{1}{n_X}\mathbf{X}\mathbf{X}^t.
\end{aligned}$$

The proof is divided into two parts. First, we compute $\mathrm{E}\left[T_{\Sigma,X}\right]$ and, then, we compare it to $\mathrm{E}\left[T_{\mathrm{I}_m,X}\right]$.

1.

$$\begin{aligned}
\mathrm{E}\left[T_{\Sigma,X}\right] &= \frac{1}{\Sigma_{1,1}}\mathrm{E}\left[\left({W_X}^2\right)_{1,1}\right] - \frac{(\Sigma^2)_{1,1}}{\Sigma_{1,1}}\\
&= \frac{1}{\Sigma_{1,1}}\sum_{s=1}^m \mathrm{E}\left[(W_{X,1,s})^2\right] - \frac{(\Sigma^2)_{1,1}}{\Sigma_{1,1}}\\
&= \frac{1}{\Sigma_{1,1}}\sum_{s=1}^m \mathrm{Var}\left(W_{X,1,s}\right)\\
&= \frac{1}{\Sigma_{1,1}}\sum_{s=2}^m \mathrm{Var}\left(W_{X,1,s}\right) + O\left(\frac{1}{m}\right).
\end{aligned}$$

The last line is obtained using the condition $\mathrm{Var}\left(W_{X,1,s}\right) = O(1/m)$ of class $\mathcal{C}_A$.

We recall that $W_X = \Sigma^{1/2}\textcolor{orange}{W}_X\Sigma^{1/2}$. Then,

$$
\begin{aligned}
\sum_{s=2}^{m} \mathrm{Var}\left(W_{X,1,s}\right) &= \sum_{s=2}^{m} \mathrm{Var}\left(\left(\Sigma^{1/2}\textcolor{orange}{W}_X\Sigma^{1/2}\right)_{1,s}\right) \\
&= \underbrace{\sum_{s=2}^{m}\sum_{i,j}^{m} \left(\Sigma^{1/2}\right)^2_{1,i}\left(\Sigma^{1/2}\right)^2_{s,j} \mathrm{Var}\left(\textcolor{orange}{W}_{X,i,j}\right)}_{A} \\
&\quad \underbrace{+\sum_{s=2}^{m}\sum_{i>j}^{m} 2\left(\Sigma^{1/2}\right)_{1,i}\left(\Sigma^{1/2}\right)_{s,j}\left(\Sigma^{1/2}\right)_{1,j}\left(\Sigma^{1/2}\right)_{s,i} \mathrm{Var}\left(\textcolor{orange}{W}_{X,i,j}\right).}_{B}
\end{aligned}
$$

The last equality is obtained using the fact that $\textcolor{orange}{\mathbf{X}}_{\cdot,i}$ have independent entries. Indeed, assuming $p \leqslant q$ and $r \leqslant t$,

$$
\begin{aligned}
\mathrm{Cov}\left(\textcolor{orange}{W}_{X,p,q}, \textcolor{orange}{W}_{X,r,t}\right) &= \frac{1}{n_X^2}\sum_{i,j} \mathrm{Cov}\left(\textcolor{orange}{\mathbf{X}}_{p,i}\textcolor{orange}{\mathbf{X}}_{q,i}, \textcolor{orange}{\mathbf{X}}_{r,j}\textcolor{orange}{\mathbf{X}}_{t,j}\right) \\
&= \delta_{p=r,q=t} \mathrm{Var}\left(\textcolor{orange}{W}_{X,p,q}\right).
\end{aligned}
$$

The expression depends on $\mathrm{Var}\left(\textcolor{orange}{W}_{X,i,j}\right)$ and this terms can take four values.

$$
\begin{aligned}
&\text{If } i=j=1,\ \mathrm{Var}\left(\textcolor{orange}{W}_{X,i,j}\right) = \frac{V_{1,1}}{n_X}, \\
&\text{If } i=1,\ j>1\ \mathrm{Var}\left(\textcolor{orange}{W}_{X,i,j}\right) = \frac{V_{1,2}}{n_X}, \\
&\text{If } i=j>1,\ \mathrm{Var}\left(\textcolor{orange}{W}_{X,i,j}\right) = \frac{V_{2,2}}{n_X}, \\
&\text{If } i>1,\ j>1,\ i\neq j,\ \mathrm{Var}\left(\textcolor{orange}{W}_{X,i,j}\right) = \frac{V_{2,3}}{n_X}.
\end{aligned}
$$

Therefore,

A:

$$
\begin{aligned}
A = &\sum_{s=2}^{m}\sum_{i,j}^{m} \left(\Sigma^{1/2}\right)^2_{1,i}\left(\Sigma^{1/2}\right)^2_{s,j} \mathrm{Var}\left(\textcolor{orange}{W}_{X,i,j}\right) \\
= &\sum_{s=2}^{m}\left(\Sigma^{1/2}\right)^2_{1,1}\left(\Sigma^{1/2}\right)^2_{s,1}\frac{V_{1,1}}{n_X} + \sum_{s=2}^{m}\sum_{j>2}^{m}\left(\Sigma^{1/2}\right)^2_{1,1}\left(\Sigma^{1/2}\right)^2_{s,j}\frac{V_{1,2}}{n_X} + \sum_{s=2}^{m}\sum_{i>2}^{m}\left(\Sigma^{1/2}\right)^2_{1,i}\left(\Sigma^{1/2}\right)^2_{s,1}\frac{V_{1,2}}{n_X} \\
&+\sum_{s=2}^{m}\sum_{i>2}^{m}\left(\Sigma^{1/2}\right)^2_{1,i}\left(\Sigma^{1/2}\right)^2_{s,i}\frac{V_{2,2}}{n_X} + \sum_{s=2}^{m}\sum_{\substack{j>2\\ i>2\\ i\neq j}}^{m}\left(\Sigma^{1/2}\right)^2_{1,i}\left(\Sigma^{1/2}\right)^2_{s,j}\frac{V_{2,3}}{n_X} \\
= &\left(\Sigma^{1/2}\right)^2_{1,1}\left(\Sigma_{1,1} - \left(\Sigma^{1/2}\right)^2_{1,1}\right)\frac{V_{1,1}}{n_X} + \sum_{s=2}^{m}\left(\Sigma^{1/2}\right)^2_{1,1}\left(\Sigma_{s,s} - \left(\Sigma^{1/2}\right)^2_{s,1}\right)\frac{V_{1,2}}{n_X} \\
&+\left(\Sigma_{1,1} - \left(\Sigma^{1/2}\right)^2_{1,1}\right)^2\frac{V_{1,2}}{n_X} + \sum_{i>2}^{m}\left(\Sigma^{1/2}\right)^2_{1,i}\left(\Sigma_{i,i} - \left(\Sigma^{1/2}\right)^2_{1,i}\right)\frac{V_{2,2}}{n_X} \\
&+\sum_{s=2}^{m}\left(\Sigma_{1,1} - \left(\Sigma^{1/2}\right)^2_{1,1}\right)\left(\Sigma_{s,s} - \left(\Sigma^{1/2}\right)^2_{s,1}\right)\frac{V_{2,3}}{n_X} \\
= &\left(\Sigma^{1/2}\right)^2_{1,1} mV_{1,2}/n_x + \Sigma_{1,1}\max_{i=2,\ldots,m}\{\Sigma_{i,i}\}\frac{V_{2,2}}{n_X} + \left(\Sigma_{1,1} - \left(\Sigma^{1/2}\right)^2_{1,1}\right) m\frac{V_{2,3}}{n_X} + O\left(\frac{1}{n_X}\right) \\
= &\Sigma_{1,1}c_X V_{2,3} + \left(\Sigma^{1/2}\right)^2_{1,1} c_X\left(V_{1,2} - V_{2,3}\right) + O\left(\frac{1}{n_X}\right).
\end{aligned}
$$

Below, we show that $V_{1,2} - V_{2,3} = o_p(1)$. For $r < t$,

$$\begin{aligned}
\mathrm{Var}\left(W_{X,r,t}\right) &= \mathrm{Var}\left(\frac{1}{n_X}\sum_{i=1}^{n_X} \mathbf{X}_{r,i}\mathbf{X}_{t,i}\right) \\
&= \frac{1}{n_X^2}\sum_{i,j=1}^{n_X} \mathrm{Cov}\left(\mathbf{X}_{r,i}\mathbf{X}_{t,i}, \mathbf{X}_{r,j}\mathbf{X}_{t,j}\right) \\
&= \frac{1}{n_X^2}\sum_{i,j=1}^{n_X} \mathrm{E}\left[\mathbf{X}_{r,i}\mathbf{X}_{t,i}\mathbf{X}_{r,j}\mathbf{X}_{t,j}\right] \\
&= \frac{1}{n_X^2}\sum_{i,j=1}^{n_X} \mathrm{E}\left[\mathbf{X}_{r,i}\mathbf{X}_{r,j}\right]\mathrm{E}[\mathbf{X}_{t,i}\mathbf{X}_{t,j}] .
\end{aligned}$$

Therefore,

$$\begin{aligned}
V_{1,2} &= \frac{1}{n_X}\sum_{i,j=1}^{n_X} \mathrm{E}\left[\mathbf{X}_{1,i}\mathbf{X}_{1,j}\right]\mathrm{E}[\mathbf{X}_{2,i}\mathbf{X}_{2,j}] = \frac{1}{n_X}\sum_{i,j=1}^{n_X} \Psi_{1;i,j}\Psi_{i,j}, \\
V_{2,3} &= \frac{1}{n_X}\sum_{i,j=1}^{n_X} \Psi_{i,j}^2, \\
V_{1,2} - V_{2,3} &= \frac{1}{n_X}\sum_{i,j=1}^{n_X} \Psi_{i,j}\left(\Psi_{1;i,j} - \Psi_{i,j}\right) \\
&\leqslant \frac{1}{n_X}\sqrt{\sum_{i,j=1}^{n_X} \Psi_{i,j}^2}\sqrt{\sum_{i,j=1}^{n_X} \left(\Psi_{1;i,j} - \Psi_{i,j}\right)^2} \\
&= o_m(1),
\end{aligned}$$

where the last line is obtained by the condition on Ψ in the class $\mathcal{C}_A$ defined in Definition 4.2.3.

B: The fact that $\lambda_{\max}(\Sigma)$ is bounded implies $\Sigma_{i,i}$ is bounded for $i = 1, 2, ..., m$ and Hölder inequalities,

$$\begin{aligned}
B &= \sum_{s=2}^{m}\sum_{i>j}^{m} 2\left(\Sigma^{1/2}\right)_{1,i}\left(\Sigma^{1/2}\right)_{s,j}\left(\Sigma^{1/2}\right)_{1,j}\left(\Sigma^{1/2}\right)_{s,i}\mathrm{Var}\left(W_{X,i,j}\right) \\
&= \sum_{s=2}^{m}\sum_{i=2}^{m} 2\left(\Sigma^{1/2}\right)_{1,i}\left(\Sigma^{1/2}\right)_{s,1}\left(\Sigma^{1/2}\right)_{1,1}\left(\Sigma^{1/2}\right)_{s,i} V_{1,2}/n_X \\
&\quad + \sum_{s=2}^{m}\sum_{i>j>1}^{m} 2\left(\Sigma^{1/2}\right)_{1,i}\left(\Sigma^{1/2}\right)_{s,j}\left(\Sigma^{1/2}\right)_{1,j}\left(\Sigma^{1/2}\right)_{s,i} V_{2,3}/n_X \\
&= \sum_{s=2}^{m}\sum_{i>j}^{m} 2\left(\Sigma^{1/2}\right)_{1,i}\left(\Sigma^{1/2}\right)_{s,j}\left(\Sigma^{1/2}\right)_{1,j}\left(\Sigma^{1/2}\right)_{s,i} V_{2,3}/n_X + O\left(\frac{1}{n_X}\right) \\
&= \sum_{s=2}^{m}\sum_{i,j=1}^{m} \left(\Sigma^{1/2}\right)_{1,i}\left(\Sigma^{1/2}\right)_{s,j}\left(\Sigma^{1/2}\right)_{1,j}\left(\Sigma^{1/2}\right)_{s,i} V_{2,3}/n_X \\
&\quad - \sum_{s=2}^{m}\sum_{i}^{m} \left(\Sigma^{1/2}\right)_{1,i}^2\left(\Sigma^{1/2}\right)_{s,i}^2 V_{2,3}/n_X + O\left(\frac{1}{n_X}\right) \\
&= O\left(\frac{1}{n_X}\right).
\end{aligned}$$

Therefore, we conclude

$$\mathrm{E}\left[T_{\Sigma,X}\right] = c_X V_{1,2} + o_m(1).$$

2. On the other hand, when $\Sigma = \mathrm{I}_m$,

$$\begin{aligned}
\mathrm{E}\left[T_{\mathrm{I}_m,X}\right] &= \sum_{s=2}^{m} \mathrm{Var}\left(W_{X,1,s}\right) \\
&= \frac{1}{n_X} \sum V_{1,s} \\
&= c_X V_{1,2} + o_m(1).
\end{aligned}$$

Therefore, the expectation in probability of the residual spike is asymptotically independent of Σ in the class $\mathcal{C}_A$.

$\square$

Lemma 6.3.3.
We define in the class $\mathcal{C}_A(\mathrm{I}_m)$, $\mathbf{X}_{\mathrm{I}_m,\mathcal{L}_X} = P^{1/2}\mathbf{X}_{\mathcal{L}_X} \in \mathbb{R}^{m\times n_X}$ and $\mathbf{Y}_{\mathrm{I}_m,\mathcal{L}_Y} = P^{1/2}\mathbf{Y}_{\mathcal{L}_Y} \in \mathbb{R}^{m\times n_Y}$. Moreover, $\mathbf{X}_{\mathcal{L}_X}$ and $\mathbf{Y}_{\mathcal{L}_Y}$ have temporal structure such that

$$\begin{aligned}
\mathbf{X}_{\mathcal{L}_X,1,\cdot}, \mathbf{X}_{\mathcal{L}_X,2,\cdot}, \ldots, \mathbf{X}_{\mathcal{L}_X,m,\cdot} &\sim \mathcal{L}_X, \\
\mathbf{Y}_{\mathcal{L}_Y,1,\cdot}, \mathbf{Y}_{\mathcal{L}_Y,2,\cdot}, \ldots, \mathbf{Y}_{\mathcal{L}_Y,m,\cdot} &\sim \mathcal{L}_Y.
\end{aligned}$$

We define $W_{X,\mathcal{L}_X} = \frac{1}{n_X}\mathbf{X}_{\mathcal{L}_X}\mathbf{X}^t_{\mathcal{L}_X}$ and $W_{Y,\mathcal{L}_Y} = \frac{1}{n_Y}\mathbf{Y}_{\mathcal{L}_Y}\mathbf{Y}^t_{\mathcal{L}_Y}$.
If the spectra of $\left(W_{X,\mathcal{L}_1}, W_{Y,\mathcal{L}_3}\right)$ and $\left(W_{X,\mathcal{L}_2}, W_{Y,\mathcal{L}_4}\right)$ are rescaled and the second moments of the spectra are the same, i.e.

$$\begin{aligned}
&\mathrm{E}\left[\sum_{i=1}^m \hat\lambda_{W_{X,\mathcal{L}_1},i}\right] = \mathrm{E}\left[\sum_{i=1}^m \hat\lambda_{W_{X,\mathcal{L}_2},i}\right] = 1 \text{ and } \mathrm{E}\left[\sum_{i=1}^m \hat\lambda^2_{W_{X,\mathcal{L}_1},i}\right] = \mathrm{E}\left[\sum_{i=1}^m \hat\lambda^2_{W_{X,\mathcal{L}_2},i}\right], \\
&\mathrm{E}\left[\sum_{i=1}^m \hat\lambda_{W_{Y,\mathcal{L}_3},i}\right] = \mathrm{E}\left[\sum_{i=1}^m \hat\lambda_{W_{Y,\mathcal{L}_4},i}\right] = 1 \text{ and } \mathrm{E}\left[\sum_{i=1}^m \hat\lambda^2_{W_{Y,\mathcal{L}_3},i}\right] = \mathrm{E}\left[\sum_{i=1}^m \hat\lambda^2_{W_{Y,\mathcal{L}_4},i}\right],
\end{aligned}$$

then

$$\mathrm{E}_{p(\min(m,\theta))}\left[\lambda_{m,\mathcal{L}_1,\mathcal{L}_3}\right] = \mathrm{E}_{p(\min(m,\theta))}\left[\lambda_{m,\mathcal{L}_2,\mathcal{L}_4}\right],$$

where $\lambda_{m,\mathcal{L}_1,\mathcal{L}_3}$ and $\lambda_{m,\mathcal{L}_2,\mathcal{L}_4}$ are the resulting residual spikes using $\left(\mathbf{X}_{\mathrm{I}_m,\mathcal{L}_1}, \mathbf{Y}_{\mathrm{I}_m,\mathcal{L}_3}\right)$ and $\left(\mathbf{X}_{\mathrm{I}_m,\mathcal{L}_2}, \mathbf{Y}_{\mathrm{I}_m,\mathcal{L}_4}\right)$ respectively.

(Page 94)

Proof. **Lemma 6.3.3**

Lemma 6.3.1 implies

$$\hat{\hat\theta}_Y(1-\hat\alpha^2) = \frac{1}{\Sigma_{1,1}}\left(\left({W_X}^2\right)_{,1,1} + \left({W_Y}^2\right)_{1,1}\right) - \frac{2}{\Sigma_{1,1}}\left(\Sigma^2\right)_{1,1} + O_p\left(\frac{1}{\theta}\right) + O_p\left(\frac{1}{m^{1/2}}\right).$$

Thus, we only need to check the result for

$$T_{X,\mathcal{L}_X} = \left({W_{X,\mathcal{L}_X}}^2\right)_{1,1} - 1,$$

where, in this case, $\left(W_{X,\mathcal{L}_X}{}^2\right)_{1,1} = \left(W_{X,\mathcal{L}_X}{}^2\right)_{1,1}$. Therefore, we need to show that

$$\mathrm{E}\left[T_{X,\mathcal{L}_1}\right] = \mathrm{E}\left[T_{X,\mathcal{L}_2}\right].$$

Because each row of the matrix $\mathbf{X}_{\mathcal{L}_1}$ has the same distribution, we obtain

$$\begin{aligned}
\frac{1}{m}\mathrm{E}\left[\sum_{i=1}^{m}\hat{\lambda}^2_{W_{X,\mathcal{L}_1},i}\right] &= \frac{1}{m}\mathrm{E}\left[\sum_{i=1}^{m}\hat{\lambda}^2_{W_{X,\mathcal{L}_1},i}\sum_{j=1}^{m}\hat{u}^2_{W_{X,\mathcal{L}_1},i,j}\right] \\
&= \frac{1}{m}\sum_{j=1}^{m}\mathrm{E}\left[\sum_{i=1}^{m}\hat{\lambda}^2_{W_{X,\mathcal{L}_1},i}\hat{u}^2_{W_{X,\mathcal{L}_1},i,j}\right] \\
&= \frac{1}{m}\sum_{j=1}^{m}\mathrm{E}\left[\left(W_{X,\mathcal{L}_1}{}^2\right)_{j,j}\right] \\
&= \mathrm{E}\left[\left(W_{X,\mathcal{L}_1}{}^2\right)_{1,1}\right].
\end{aligned}$$

The last line is true because each row has same distribution and can be permuted. Therefore, $\mathrm{E}\left[\left(W_{X,\mathcal{L}_1}{}^2\right)_{1,1}\right] = \mathrm{E}\left[\left(W_{X,\mathcal{L}_1}{}^2\right)_{i,i}\right]$ for $i = 2, ..., m$. The same path leads to

$$\frac{1}{m}\mathrm{E}\left[\sum_{i=1}^{m}\hat{\lambda}^2_{W_{X,\mathcal{L}_2},i}\right] = \mathrm{E}\left[\left(W_{X,\mathcal{L}_2}{}^2\right)_{1,1}\right].$$

This concludes the proof.

□

Lemma 6.3.4.
Suppose that $\mathbf{X}_{\mathcal{L}_X}, \mathbf{X}_{\mathcal{L}_X^*} \in \mathbb{R}^{m\times n_X}$ *and* $\mathbf{Y}_{\mathcal{L}_Y}, \mathbf{Y}_{\mathcal{L}_Y^*} \in \mathbb{R}^{m\times n_Y}$ *are in* $\mathcal{C}_A\left(\mathrm{I}_m\right)$ *such that,*

$$\begin{aligned}
\mathbf{X}_{\mathcal{L}_X} &= P^{1/2}\mathbf{X}_{\mathcal{L}_X}, \quad \mathbf{Y}_{\mathcal{L}_Y} = P^{1/2}\mathbf{Y}_{\mathcal{L}_Y}, \\
\mathbf{X}_{\mathcal{L}_X^*} &= P^{1/2}\mathbf{X}_{\mathcal{L}_X^*}, \quad \mathbf{Y}_{\mathcal{L}_Y^*} = P^{1/2}\mathbf{Y}_{\mathcal{L}_Y^*},
\end{aligned}$$

where $P = \mathrm{I}_m + (\theta - 1)e_1e_1^t$.
Moreover, $\left(\mathbf{X}_{\mathcal{L}_X}, \mathbf{Y}_{\mathcal{L}_Y}\right)$ *and* $\left(\mathbf{X}_{\mathcal{L}_X^*}, \mathbf{Y}_{\mathcal{L}_Y^*}\right)$ *have temporal structure such that*

$$\begin{aligned}
&\mathbf{X}_{\mathcal{L}_X,1,\cdot}, \mathbf{X}_{\mathcal{L}_X,2,\cdot}, ..., \mathbf{X}_{\mathcal{L}_X,m,\cdot} \sim \mathcal{L}_X, \\
&\mathbf{Y}_{\mathcal{L}_Y,1,\cdot}, \mathbf{Y}_{\mathcal{L}_Y,2,\cdot}, ..., \mathbf{Y}_{\mathcal{L}_Y,m,\cdot} \sim \mathcal{L}_Y, \\
&\mathbf{X}_{\mathcal{L}_X^*,1,\cdot} \sim \mathcal{L}_{X,1}, \\
&\mathbf{X}_{\mathcal{L}_X^*,2,\cdot}, \mathbf{X}_{\mathcal{L}_X^*,3,\cdot}, ..., \mathbf{X}_{\mathcal{L}_X^*,m,\cdot} \sim \mathcal{L}_X, \\
&\mathbf{Y}_{\mathcal{L}_Y^*,1,\cdot} \sim \mathcal{L}_{Y,1}, \\
&\mathbf{Y}_{\mathcal{L}_Y^*,2,\cdot}, \mathbf{Y}_{\mathcal{L}_Y^*,3,\cdot}, ..., \mathbf{Y}_{\mathcal{L}_Y^*,m,\cdot} \sim \mathcal{L}_Y.
\end{aligned}$$

In particular, $\mathrm{Cov}\left(\mathbf{X}_{\mathcal{L}_X,1,\cdot}\right) = \Psi_{1,X} = \Psi_X$, $\mathrm{Cov}\left(\mathbf{Y}_{\mathcal{L}_Y,1,\cdot}\right) = \Psi_{1,Y} = \Psi_Y$, $\mathrm{Cov}\left(\mathbf{X}_{\mathcal{L}^*,1,\cdot}\right) = \Psi_{1,X}^*$, $\mathrm{Cov}\left(\mathbf{Y}_{\mathcal{L}^*,1,\cdot}\right) = \Psi_{1,Y}^*$, $\Delta_X = \Psi_X - \Psi_{1,X}^*$ *and* $\Delta_Y = \Psi_Y - \Psi_{1,Y}^*$ *respect the conditions of the class* $\mathcal{C}_A(\mathrm{I}_m)$.
Therefore we have

$$\mathrm{E}_{p(\min(m,\theta))}\left[\lambda_{m,\mathcal{L}_X,\mathcal{L}_Y}\right] = \mathrm{E}_{p(\min(m,\theta))}\left[\lambda_{m,\mathcal{L}_X^*,\mathcal{L}_Y^*}\right],$$

where $\lambda_{m,\mathcal{L}_X,\mathcal{L}_Y}$ *and* $\lambda_{m,\mathcal{L}_X^*,\mathcal{L}_Y^*}$ *are the resulting residual spikes using* $\left(\mathbf{X}_{\mathcal{L}_X}, \mathbf{Y}_{\mathcal{L}_Y}\right)$ *and* $\left(\mathbf{X}_{\mathcal{L}_X^*}, \mathbf{Y}_{\mathcal{L}_Y^*}\right)$ *respectively.*

(Page 95)

Proof. **Lemma 6.3.4**

Using Lemma 6.3.1, we just show that the expectations of $\left(W_{X,\mathcal{L}_X}{}^2\right)_{1,1}$ and $W^2{}_{X,\mathcal{L}^*_X,1,1}$ tend to the same value.

$$\begin{aligned}
\mathrm{E}\left[\left(W_{X,\mathcal{L}^*_X}{}^2\right)_{1,1}\right] &= \frac{1}{n_X^2}\sum_{r=1}^{m}\sum_{i,j=1}^{n_X}\mathrm{E}\left[\mathbf{X}_{\mathcal{L}^*_X,1,i}\mathbf{X}_{\mathcal{L}^*_X,1,j}\mathbf{X}_{\mathcal{L}^*_X,r,i}\mathbf{X}_{\mathcal{L}^*_X,r,j}\right]\\
&= \frac{1}{n_X^2}\sum_{i,j=1}^{n_X}\mathrm{E}\left[\mathbf{X}^2_{\mathcal{L}^*_X,1,i}\mathbf{X}^2_{\mathcal{L}^*_X,1,j}\right]+\frac{1}{n_X^2}\sum_{i,j=1}^{n_X}\mathrm{E}\left[\mathbf{X}_{\mathcal{L}^*_X,1,i}\mathbf{X}_{\mathcal{L}^*_X,1,j}\sum_{r=2}^{m}\mathbf{X}_{\mathcal{L}^*_X,r,i}\mathbf{X}_{\mathcal{L}^*_X,r,j}\right]\\
&= \frac{1}{n_X^2}\mathrm{E}\left[\left(\sum_{i=1}^{n_X}\mathbf{X}^2_{\mathcal{L}^*_X,1,i}\right)\right]+\frac{1}{n_X^2}\sum_{i,j=1}^{n_X}\mathrm{E}\left[\mathbf{X}_{\mathcal{L}^*_X,1,i}\mathbf{X}_{\mathcal{L}^*_X,1,j}\right]\mathrm{E}\left[\sum_{r=2}^{m}\mathbf{X}_{\mathcal{L}^*_X,r,i}\mathbf{X}_{\mathcal{L}^*_X,r,j}\right]\\
&= \mathrm{Var}\left(W_{X,\mathcal{L}^*_X,1,1}\right)+1+\frac{1}{n_X^2}\sum_{i,j=1}^{n_X}\mathrm{E}\left[\mathbf{X}_{\mathcal{L}^*_X,1,i}\mathbf{X}_{\mathcal{L}^*_X,1,j}\right]\mathrm{E}\left[\sum_{r=2}^{m}\mathbf{X}_{\mathcal{L}^*_X,r,i}\mathbf{X}_{\mathcal{L}^*_X,r,j}\right].
\end{aligned}$$

The same computation leads to

$$\mathrm{E}\left[\left(W_{X,\mathcal{L}_X}{}^2\right)_{1,1}\right] = \mathrm{Var}\left(W_{X,\mathcal{L}_X,1,1}\right)+1+\frac{1}{n_X^2}\sum_{i,j=1}^{n_X}\mathrm{E}\left[\mathbf{X}_{\mathcal{L}_X,1,i}\mathbf{X}_{\mathcal{L}_X,1,j}\right]\mathrm{E}\left[\sum_{r=2}^{m}\mathbf{X}_{\mathcal{L}_X,r,i}\mathbf{X}_{\mathcal{L}_X,r,j}\right].$$

Because in class $\mathcal{C}_A$, $\mathrm{Var}\left(W_{X,\mathcal{L}^*_X,1,1}\right)$ and $\mathrm{Var}\left(W_{X,\mathcal{L}_X,1,1}\right)$ are $O\left(1/m\right)$,

$$\begin{aligned}
&\left|\mathrm{E}_{\mathcal{L}^*_X}\left[\left(W_{X,\mathcal{L}^*_X}{}^2\right)_{1,1}\right]-\mathrm{E}\left[\left(W_{X,\mathcal{L}_X}{}^2\right)_{1,1}\right]\right|\\
&=\left|\frac{1}{n_X^2}\sum_{i,j=1}^{n_X}\left(\mathrm{E}\left[\mathbf{X}_{\mathcal{L}_{1,X},i,1}\mathbf{X}_{\mathcal{L}_{1,X},j,1}\right]-\mathrm{E}\left[\mathbf{X}_{\mathcal{L}_X,i,1}\mathbf{X}_{\mathcal{L}_X,j,1}\right]\right)\mathrm{E}\left[\sum_{r=2}^{m}\mathbf{X}_{\mathcal{L}_X,i,r}\mathbf{X}_{\mathcal{L}_X,j,r}\right]+O\left(\frac{1}{m}\right)\right|\\
&=\left|\frac{c_X}{n_X}\sum_{i,j=1}^{n_X}\left(\Psi^*_{1,X;i,j}-\Psi_{X;i,j}\right)\Psi_{X;i,j}+O\left(\frac{1}{m}\right)\right|\\
&\leqslant\frac{c_X}{n_X}\sqrt{\mathrm{Trace}\left(\Delta_X^2\right)\mathrm{Trace}\left(\Psi_X^2\right)}+O\left(\frac{1}{m}\right)\\
&=o_m(1).
\end{aligned}$$

This concludes the proof.

□

Lemma 6.3.5.
Suppose that $\mathbf{X}=P^{1/2}\mathbf{X}\in\mathbb{R}^{m\times n_X}$ *and* $\mathbf{Y}=P^{1/2}\mathbf{Y}\in\mathbb{R}^{m\times n_Y}$ *are two random matrices in class* $\mathcal{C}_A(\Sigma)$ *such that*

$$\begin{aligned}
&\frac{1}{m}\sum_{i=1}^{m}\hat\lambda_{W_X,i}=\frac{1}{m}\sum_{i=1}^{m}\hat\lambda_{W_Y,i}=1,\\
&\frac{1}{m}\sum_{i=1}^{m}\hat\lambda^2_{W_X,i}=M_{2,X}\ \textit{and}\ \frac{1}{m}\sum_{i=1}^{m}\hat\lambda^2_{W_Y,i}=M_{2,Y},\\
&M_2=\frac{M_{2,X}+M_{2,Y}}{2}.
\end{aligned}$$

We define $\lambda_{m,\mathcal{C}_A(\Sigma)}$ as the resulting residual spike between $\mathbf{X}$ *and* $\mathbf{Y}$.
The Main Theorem 3.1.1 says that knowing the spectra, the estimator of the expectation of residual spike assuming our model is

$$\lambda^{+}(M_2) = M_2 + \sqrt{M_2^2 - 1}.$$

This expectation of $\lambda_{m,\mathcal{C}_A(\Sigma)}$ is conservative in the sense of the second robustness defined in 4.2.2. This means that the following results are true when n_X, n_Y and θ tends to infinity.

1: If $\Sigma = \mathrm{I}_m$, then $\lambda^{+}(M_2) = M_2 + \sqrt{M_2^2 - 1} = \mathrm{E}_{p(\min(m,\theta))}\left[\lambda_{m,\mathcal{C}_A(\Sigma)}\right]$.

2: If $\Sigma \neq \mathrm{I}_m$, then $\lambda^{+}(M_2) = M_2 + \sqrt{M_2^2 - 1} > \mathrm{E}_{p(\min(m,\theta))}\left[\lambda_{m,\mathcal{C}_A(\Sigma)}\right]$.

(Page 95)

Proof. **Lemma 6.3.5**

Our simple model assumes $\Sigma = \mathrm{I}_m$ and elliptical data. Therefore, Lemma 6.3.1 is still true and we just need to investigate $\frac{\hat{\hat{\theta}}_Y}{\hat{\hat{\theta}}_X}$ and $\hat{\hat{\theta}}_Y(1-\hat{\alpha}^2)$. We call E_{simple} and $\mathrm{E}_{\mathcal{C}_A(\Sigma)}$, the expectations assuming our simple model and a model in the class $\mathcal{C}_A(\Sigma)$. By analogy we define $\mathrm{E}_{p;simple}$ and $\mathrm{E}_{p;\mathcal{C}_A(\Sigma)}$ as the expectation in probability.

$$\begin{aligned}
\mathrm{E}_{p(m);simple}\left[\frac{\hat{\hat{\theta}}_Y}{\hat{\hat{\theta}}_X}\right] &= 1 + O\left(\frac{1}{m}\right), \\
\mathrm{E}_{p(\min(m,\theta));simple}\left[\hat{\hat{\theta}}_Y(1-\hat{\alpha}^2)\middle| M_2\right] &= \mathrm{E}_{simple}\left[\left({W_X}^2\right)_{,1,1} + \left({W_Y}^2\right)_{1,1}\middle| M_2\right] - 2 \\
&= 2M_2 - 2.
\end{aligned}$$

In order to prove the robustness in expectation of the estimators obtained assuming this model, we just need to show that

$$\mathrm{E}_{p(\min(m,\theta));simple}\left[\hat{\hat{\theta}}_Y(1-\hat{\alpha}^2)\middle| M_2\right] \geqslant \mathrm{E}_{p(\min(m,\theta));\mathcal{C}_A(\Sigma)}\left[\hat{\hat{\theta}}_Y(1-\hat{\alpha}^2)\middle| M_2\right].$$

Using the Lemma 6.3.4, we can assume that $\Psi_1 = \Psi$ in the class $\mathcal{C}_A$.

1: If $\Sigma = \mathrm{I}_m$, then using the restriction, $\Psi_1 = \Psi$, leads to

$$\begin{aligned}
\mathrm{E}_{\mathcal{C}_A(\mathrm{I}_m)}\left[\left({W_X}^2\right)_{,1,1}\right] &= \frac{1}{m}\sum_{i=1}^{m} \mathrm{E}_{\mathcal{C}_A(\mathrm{I}_m)}\left[\left({W_X}^2\right)_{,i,i}\right] + o_m(1) \\
&= \frac{1}{m}\mathrm{E}_{\mathcal{C}_A(\mathrm{I}_m)}\left[\mathrm{Trace}\left(\left({W_X}^2\right)\right)\right] + o_m(1) \\
&= M_{2,X} + o_m(1).
\end{aligned}$$

Therefore,

$$\mathrm{E}_{p(\min(m,\theta);simple}\left[\hat{\hat{\theta}}_Y(1-\hat{\alpha}^2)\middle| M_2\right] = 2M_2 - 2 = \mathrm{E}_{p(\min(m,\theta);\mathcal{C}(\mathrm{I}_m)}\left[\hat{\hat{\theta}}_Y(1-\hat{\alpha}^2)\right].$$

2: If $\Sigma \neq \mathrm{I}_m$, then, by using the proof of Lemma 6.3.2 (and its estimations of A and B), we obtain

$$\begin{aligned}
\mathrm{E}_{p(\min(m,\theta));\mathcal{C}_A(\Sigma)}\left[\hat{\hat{\theta}}_Y(1-\hat{\alpha}^2)\right] &= \mathrm{E}_{\mathcal{C}_A(\Sigma)}\left[\frac{1}{\Sigma_{1,1}}\left(\left({W_X}^2\right)_{,1,1} + \left({W_Y}^2\right)_{1,1}\right) - \frac{2}{\Sigma_{1,1}}\left(\Sigma^2\right)_{1,1}\right] \\
&= \frac{\sum_{s=2}^{m}\mathrm{Var}_{\mathcal{C}_A(\Sigma)}\left(W_{X,1,s}\right) + \sum_{s=2}^{m}\mathrm{Var}_{\mathcal{C}_A(\Sigma)}\left(W_{Y,1,s}\right)}{\Sigma_{1,1}} \\
&= m\left(\mathrm{Var}_{\mathcal{C}_A(\Sigma)}\left(W_{X,1,2}\right) + \mathrm{Var}_{\mathcal{C}_A(\Sigma)}\left(W_{Y,1,2}\right)\right),
\end{aligned}$$

where $W_X = \frac{1}{n_X}\mathbf{X}\mathbf{X}^t$ is such that $W_X = \Sigma^{1/2} W_X \Sigma^{1/2}$.
We remember that assuming our model leads to

$$\mathrm{E}_{p(\min(m,\theta));simple}\left[\hat{\hat{\theta}}_Y(1-\hat{\alpha}^2)\middle| M_2\right] = 2M_2 - 2.$$

Therefore, using the same argument as above and the restriction $\Psi_1 = \Psi$,

$$\begin{aligned}
2M_2 - 2 &= \mathrm{E}_{\mathcal{C}_A(\Sigma)}\left[\frac{1}{m}\left(\mathrm{Trace}\left(\left(W_X{}^2\right)\right) + \mathrm{Trace}\left(W^2{}_Y\right)\right)\right] - 2 \\
&= \sum_{i=1}^{m} \mathrm{E}_{\mathcal{C}_A(\Sigma)}\left[\left(\left(W_X{}^2\right)_{,i,i} + \left(W_Y{}^2\right)_{,i,i}\right) - 2\left(\Sigma^2\right)_{i,i}\right] + \frac{2}{m}\mathrm{Trace}\left(\Sigma^2\right) - 2 \\
&= \frac{1}{m}\sum_{i=1}^{m} \Sigma_{i,i}\left(m\left(\mathrm{Var}_{\mathcal{C}_A(\Sigma)}\left(W_{X,1,2}\right) + \mathrm{Var}_{\mathcal{C}_A(\Sigma)}\left(W_{Y,1,2}\right)\right)\right) \\
&\qquad + \frac{2}{m}\mathrm{Trace}\left(\Sigma^2\right) - 2 \\
&= m\left(\mathrm{Var}_{\mathcal{C}_A(\Sigma)}\left(W_{X,1,2}\right) + \mathrm{Var}_{\mathcal{C}_A(\Sigma)}\left(W_{Y,1,2}\right)\right) + \frac{2}{m}\mathrm{Trace}\left(\Sigma^2\right) - 2.
\end{aligned}$$

Because $\left(\mathrm{Trace}\left(\Sigma \mathrm{I}_m\right)\right)^2 < m\mathrm{Trace}\left(\Sigma^2\right)$, then $\frac{2}{m}\mathrm{Trace}\left(\Sigma^2\right) - 2 > 0$. Therefore

$$2M_2 - 2 > \mathrm{E}_{p(\min(m,\theta));\mathcal{C}_A(\Sigma)}\left[\hat{\hat{\theta}}_Y(1-\hat{\alpha}^2)\right].$$

Finally, the last inequality follows from the formula of the residual spike.

$$\lambda^+(M_2) = M_2 + \sqrt{M_2^2 - 1} > \mathrm{E}_{p(\min(m,\theta))}\left[\lambda_{m,\mathcal{C}_A(\Sigma)}\right].$$

$\square$

7.3.2 Proof of Model A

Theorem 4.2.1.
Assume $\mathbf{X} \in \mathbb{R}^{m\times n_X}$ and $\mathbf{Y} \in \mathbb{R}^{m\times n_Y}$ are random matrices from class $\mathcal{C}_A(\Sigma)$ defined in 4.2.3. By definition of $\mathcal{C}_A$,

$$\mathbf{X} = P^{1/2}\Sigma^{1/2}\mathbf{X} \text{ and } \mathbf{Y} = P^{1/2}\Sigma^{1/2}\mathbf{Y}.$$

We define

$$\tilde{\mathbf{X}} = P^{1/2}\tilde{\mathbf{X}} \text{ and } \tilde{\mathbf{Y}} = P^{1/2}\tilde{\mathbf{Y}},$$

such that $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ are invariant by rotation with spectra equal to the spectra of $\mathbf{X}$ and $\mathbf{Y}$ respectively. Moreover, we define $\lambda_{A,m}$ and $\lambda_{0,m}$ as the largest residual spikes obtained using $\left(\mathbf{X},\mathbf{Y}\right)$ and $\left(\tilde{\mathbf{X}},\tilde{\mathbf{Y}}\right)$ respectively.
Then,

1. *The new data $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ satisfy the conditions of our model.*

2. *Our model is asymptotically weakly robust on the right,*

$$\mathrm{E}_{p(\min(m,\theta))}\left[\lambda_{A,m}\right] \leqslant \mathrm{E}_{p(\min(m,\theta))}\left[\lambda_{0,m}\right] \text{ with equality when } \Sigma = \mathrm{I}_m.$$

3. *If* $\text{Trace}(\Sigma) - 1 > d$, *where* d *is a positive constant, our model is asymptotically strongly robust on the right,*

$$\mathrm{E}_{p(\min(m,\theta))}[\lambda_{A,m}] + D(d) < \mathrm{E}_{p(\min(m,\theta))}[\lambda_{0,m}],$$
$$\textit{where } D(d) \textit{ is positive and depends on } d,$$
$$\lim_{m,\theta\to\infty} \mathrm{Var}_{p(\min(m,\theta))}(\lambda_{A,m}) = \lim_{m,\theta\to\infty} \mathrm{Var}_{p(\min(m,\theta))}(\lambda_{0,m}) = 0.$$

Remark 4.2.1.1.

1. The theorem shows that applying the procedure based on our base model to data generated by model A leads to conservative tests.

2. This theorem can be extended to the minimum residual spike to show robustness on the left.

(Page 44)

Proof. **Theorem 4.2.1**

The proof of Theorem 4.2.1 uses four lemmas.

1: By Lemma 6.3.2, the expectation of the residual spike depends only on the temporal structure when θ is large. Therefore, the result is invariant of Σ.

2: Using the first point, we can consider only models with $\Sigma = \mathrm{I}_m$. By Lemma 6.3.3, if $\Sigma = \mathrm{I}_m$, the expectation is the same for all the temporal structures with the same two first moments of the spectrum.

3: A small fluctuation of the first rows (time) does not affect the expectation by Lemma 6.3.4.

4: Lemma 6.3.5 proves that applying our model leads to a conservative test when the true model is A.

Moreover, in the class C_A, the variance of $W_{X,1,1}$ and $\left({W_X}^2\right)_{1,1}$ tends to 0. Combining this observation with Lemma 6.3.1 shows the limit of the residual spikes variances.
The combination of the four lemmas and the convergence of the variances prove Theorem 4.2.1.

□

7.3.3 Proof of Model B

Theorem 4.2.2.
Assume $\mathbf{X} \in \mathbb{R}^{m\times n_X}$ *and* $\mathbf{Y} \in \mathbb{R}^{m\times n_Y}$ *are random matrices from class* $\mathcal{C}_B$ *defined in 4.2.4. By definition of* $\mathcal{C}_B$*,*

$$\mathbf{X} = P^{1/2}\mathbf{X} \textit{ and } \mathbf{Y} = P^{1/2}\mathbf{Y},$$

where $P = \mathrm{I}_m + (\theta - 1)uu^t$. *We define*

$$\tilde{\mathbf{X}} = \tilde{P}^{1/2}\mathbf{X} \textit{ and } \tilde{\mathbf{Y}} = \tilde{P}^{1/2}\mathbf{Y},$$

such that $\tilde{P} = \mathrm{I}_m + (\tilde{\theta} - 1)uu^t$ *and* $\lim \frac{\tilde{\theta}}{\sqrt{m}} = \infty$. *Moreover, we define* $\lambda_{B,m}$ *and* $\lambda_{0,m}$ *as the largest residual spikes obtained using* $(\mathbf{X}, \mathbf{Y})$ *and* $(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}})$ *respectively.*
If the criterion defined in 4.2.5 is such that

$$\forall\theta > 0,\ \mu_\lambda(\theta, S_{\mathbf{X}}, S_{\mathbf{Y}}) < \lim_{\theta\to\infty} \mu_\lambda(\theta, S_{\mathbf{X}}, S_{\mathbf{Y}}) + o(1/\sqrt{m}),$$

then

1. *If the variances in probability of $\hat{\theta}_X$, $\hat{\theta}_Y$, $\langle \hat{u}_X, \hat{u}_Y \rangle$ are $O(1/m)$, our model is asymptotically strongly robust on the right,*

$$\mathrm{E}_{p(m)}[\lambda_{B,m}] \leqslant \mathrm{E}_{p(m)}[\lambda_{0,m}].$$

The equality occurs when

$$\lim_{m\to\infty} \frac{\mathrm{Var}_{p(m)}(\lambda_{B,m})}{\mathrm{Var}_{p(m)}(\lambda_{0,m})} = 1.$$

2. *If the variances in probability of $\hat{\theta}_X$, $\hat{\theta}_Y$, $\langle \hat{u}_X, \hat{u}_Y \rangle$ are $o_m(1)$, our model is asymptotically weakly robust on the right,*

$$\mathrm{E}_{p(m)}[\lambda_{B,m}] \leqslant \mathrm{E}_{p(m)}[\lambda_{0,m}].$$

Remark 4.2.2.1.

1. If the criterion is satisfied, then applying the Main Theorem to data generated by model B leads to conservative tests.

2. This theorem can be extended to the minimum residual spike to show robustness on the left.

(Page 45)

Proof. **Theorem 4.2.2**

We recall the formula for the residual spike :

$$\hat{\lambda}_{residual} = \frac{1}{2}\left(\hat{\hat{\theta}}_Y + \hat{\alpha}^2 - \hat{\hat{\theta}}_Y\hat{\alpha}^2 + \frac{1 + (\hat{\hat{\theta}}_Y - 1)\hat{\alpha}^2 + \sqrt{-4\hat{\hat{\theta}}_Y\hat{\hat{\theta}}_X + \left(1 + \hat{\hat{\theta}}_Y\hat{\hat{\theta}}_X - (\hat{\hat{\theta}}_Y - 1)(\hat{\hat{\theta}}_X - 1)\hat{\alpha}^2\right)^2}}{\hat{\hat{\theta}}_X}\right)$$

1 : First, we note that if the variance of $\hat{\theta}_X$, $\hat{\theta}_Y$, $\hat{\alpha}^2$ is $O_p(1/m^\beta)$, the expectation of the residual spike tends to the function of the criterion 4.2.5, $\mu_\lambda(\theta, S_X, S_Y)$, with S_X and S_Y the spectra of the random matrices before applying the perturbation P. The error is of size $o_p(1/m^{\beta/2})$. (If the variances are just $o(1)$, the error is $o_p(1)$.)
However, we need to check that asymptotically $\mathrm{E}[\hat{\alpha}] = \alpha = \alpha_X\alpha_Y$. Indeed we already know by construction that $\hat{\hat{\theta}}_X$ is asymptotically an unbiased estimator of θ_X.

2 : Because the matrices $\mathbf{X}$ and $\mathbf{Y}$ are in the class $\mathcal{C}_B$ defined in 4.2.4, we can suppose without loss of generality that $u = e_1$. Therefore,

$$\mathrm{E}[\hat{\alpha}] = \mathrm{E}\left[\hat{u}_{\hat{\Sigma}_\mathbf{X},1}\hat{u}_{\hat{\Sigma}_\mathbf{Y},1}\right] + \mathrm{E}\left[\sum_{i=2}^m \hat{u}_{\hat{\Sigma}_\mathbf{X},i}\hat{u}_{\hat{\Sigma}_\mathbf{Y},i}\right] = \alpha_X\alpha_Y + \sum_{i=2}^m \mathrm{E}\left[\hat{u}_{\hat{\Sigma}_\mathbf{X},i}\right]\mathrm{E}\left[\hat{u}_{\hat{\Sigma}_\mathbf{Y},i}\right] = \alpha_X\alpha_Y,$$

where $\hat{u}_{\hat{\Sigma}_\mathbf{X}}$ and $\hat{u}_{\hat{\Sigma}_\mathbf{Y}}$ are the largest eigenvectors of $\hat{\Sigma}_\mathbf{X} = \frac{\mathbf{X}\mathbf{X}^t}{n_X}$ and $\hat{\Sigma}_\mathbf{Y} = \frac{\mathbf{Y}\mathbf{Y}^t}{n_Y}$ respectively. The last equality is obtained because the invariance by rotation implies $\mathrm{E}\left[\hat{u}_{\hat{\Sigma}_\mathbf{Y},i}\right] = \mathrm{E}\left[-\hat{u}_{\hat{\Sigma}_\mathbf{Y},i}\right] = 0$ for $i > 1$.

3: Therefore, the criterion 4.2.5 is the expectation of the residual spike of model B with a small error of size $o_p(1/m^{\beta/2})$ or $o_p(1)$ depending on the variance of the estimators.

Therefore, assuming that the criterion $\mu_\lambda(\theta)$ increases as a function of θ, the largest expectation appears for large θ and this scenario corresponds to our model.

□

7.3.4 Proof of Model C

Theorem 4.2.3.
Assume $\mathbf{X} \in \mathbb{R}^{m\times n_X}$ *and* $\mathbf{Y} \in \mathbb{R}^{m\times n_Y}$ *are random matrices from class* $\mathcal{C}_C$ *defined in 4.2.6. By definition of* $\mathcal{C}_B$*,*

$$\mathbf{X} = P^{1/2}\mathbf{X} \text{ and } \mathbf{Y} = P^{1/2}\mathbf{Y},$$

where $P = \mathrm{I}_m + (\theta - 1)uu^t$*. We define*

$$\tilde{\mathbf{X}} = \tilde{P}^{1/2}\tilde{\mathbf{X}} \text{ and } \tilde{\mathbf{Y}} = \tilde{P}^{1/2}\tilde{\mathbf{Y}},$$

such that $\tilde{P} = \mathrm{I}_m + (\tilde{\theta} - 1)uu^t$, $\lim \frac{\tilde{\theta}}{\sqrt{m}} = \infty$ *and* $\tilde{\mathbf{X}}$ *(respectively* $\tilde{\mathbf{Y}}$*) is invariant by rotation with spectrum equal to the spectrum of* $\mathbf{X}$ *(respectively* $\mathbf{Y}$*). Moreover, we define* $\lambda_{C,m}$ *and* $\lambda_{0,m}$ *as the largest residual spikes obtained using* $(\mathbf{X}, \mathbf{Y})$ *and* $(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}})$ *respectively.*
If the criterion defined in 4.2.5 is such that

$$\forall\theta > 0,\ \mu_\lambda(\theta, S_{\mathbf{X}}, S_{\mathbf{Y}}) < \lim_{\theta\to\infty} \mu_\lambda(\theta, S_{\mathbf{X}}, S_{\mathbf{Y}}) + o(1/\sqrt{m})$$

and if the variance in probability of $\hat{\theta}_X$*,* $\hat{\theta}_Y$*,* $\langle \hat{u}_X, \hat{u}_Y\rangle$ *are* $o_m(1)$*, our model is asymptotically weakly robust on the right,*

$$\mathrm{E}_{p(m)}[\lambda_{C,m}] \leqslant \mathrm{E}_{p(m)}[\lambda_{0,m}].$$

Remark 4.2.3.1.

1. We can show that assuming data of the classes $\mathcal{C}_B$ and $\mathcal{C}_C$ with same spectrum leads to

$$\mathrm{E}_{p(m)}[\lambda_{C,m}(\theta)] \leqslant \mathrm{E}_{p(m)}[\lambda_{B,m}(\theta)].$$

2. If θ is large, $\mathrm{E}_{p(m)}[\lambda_{C,m}] = \mathrm{E}_{p(m)}[\lambda_{0,m}]$ but we have no information about the variance.
3. If the criterion is satisfied, then applying the Main Theorem to data generated by model C leads to conservative tests.
4. This theorem can be extended to the minimum residual spike to show robustness on the left.

(Page 47)

Proof. **Theorem 4.2.3**

The proof of 4.2.3 is the same as the proof of Lemma 4.2.2. Because the data are symmetric and identical, we can apply rotations that permute elements of the matrices or change the sign of rows. Therefore, the expectation of the residual spike is exactly the same as for model B, page 202.

□

7.3.5 Proof of Model D

Theorem 4.2.4.
Assume $\mathbf{X} \in \mathbb{R}^{m\times n_X}$ *and* $\mathbf{Y} \in \mathbb{R}^{m\times n_Y}$ *are random matrices from class* $\mathcal{C}_D(u)$ *defined in 4.2.7 for* $u \in \mathbb{R}^m$*. By definition of* $\mathcal{C}_D$*,*

$$\mathbf{X} = P^{1/2}\mathbf{X} = P^{1/2}\Sigma^{1/2}\mathbf{X} \text{ and } \mathbf{Y} = P^{1/2}\mathbf{Y} = P^{1/2}\Sigma^{1/2}\mathbf{Y},$$

where $P = \mathrm{I}_m + (\theta - 1)uu^t$. *We define*

$$\tilde{\mathbf{X}} = \tilde{P}^{1/2}\tilde{\mathbf{X}} \text{ and } \tilde{\mathbf{Y}} = \tilde{P}^{1/2}\tilde{\mathbf{Y}},$$

such that $\tilde{P} = \mathrm{I}_m + (\tilde{\theta} - 1)uu^t$, $\lim \frac{\tilde{\theta}}{\sqrt{m}} = \infty$ *and* $\tilde{\mathbf{X}}$ *(respectively* $\tilde{\mathbf{Y}}$*) is invariant by rotation with spectrum equal to the spectrum of* $\mathbf{X}$ *(respectively* $\mathbf{Y}$*). Moreover, we define* $\lambda_{D,m}$ *and* $\lambda_{0,m}$ *as the largest residual spikes obtained using* $\left(\mathbf{X}, \mathbf{Y}\right)$ *and* $\left(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}\right)$ *respectively.*
If the criterion defined in 4.2.5 is such that

$$\forall \theta > 0,\ \mu_\lambda\left(\theta, S_{\mathbf{X}}, S_{\mathbf{Y}}, \hat{U}_{\mathbf{X}}, \hat{U}_{\mathbf{Y}}\right) < \lim_{\theta\to\infty} \mu_\lambda\left(\theta, S_{\mathbf{X}}, S_{\mathbf{Y}}, \hat{U}_{\mathbf{X}}, \hat{U}_{\mathbf{Y}}\right) + o(1/\sqrt{m})$$

and the variance of $\langle \hat{u}_X, \hat{u}_Y\rangle$ *is* $o_m(1)$, *then*

1. *If the criterion 4.2.8 is such that* $T = 0$, *then our model is asymptotically weakly robust on the right,*

$$\mathrm{E}_{p(m)}\left[\lambda_{D,m}\right] \leqslant \mathrm{E}_{p(m)}\left[\lambda_{0,m}\right].$$

2. *If the criterion of 4.2.8 is such that* $T > d$ *for* d *a positive constant, then our model is asymptotically strongly robust on the right,*

$$\mathrm{E}_{p(m)}\left[\lambda_{D,m}\right] + D(d) < \mathrm{E}_{p(m)}\left[\lambda_{0,m}\right] + o_m(1), \text{ where } D(d) \text{ is a positive function of } d,$$
$$\lim_{m\to\infty} \mathrm{Var}_{p(m)}\left(\lambda_{D,m}\right) = \lim_{m\to\infty} \mathrm{Var}_{p(m)}\left(\lambda_{0,m}\right) = 0.$$

Remark 4.2.4.1.

1. If $\mathrm{E}_{p(m)}\left[\lambda_{D,m}\right] = \mathrm{E}_{p(m)}\left[\lambda_{0,m}\right]$ when θ tends to infinity, the asymptotic variances are not necessarily the same. We recall that models letting θ tend to infinity are in class $\mathcal{C}_A$. Therefore, assuming $\Sigma \neq \mathrm{I}_m$ leads to conservative tests.

2. Simulations seem to show that for all θ, when $\Sigma \neq \mathrm{I}_m$, we usually have

$$\mathrm{E}_{p(m)}\left[\lambda_{D,m}\right] < \mathrm{E}_{p(m)}\left[\lambda_{0,m}\right].$$

3. If the criterion is respected and the variance of the residual spike tends to 0, then applying the Main Theorem to data generated by model D leads to conservative tests.

4. This theorem can be extended to the minimum residual spike to show robustness on the left.

(Page 49)

Proof. **Theorem 4.2.4**

To prove Theorem 4.2.4, we need to understand how the T criterion 4.2.8 is built. Then, the proof uses simple results of linear algebra.
In this theorem we use the following notation:

$$\hat{\Sigma}_X = \frac{\mathbf{X}\mathbf{X}^t}{n_X},\quad \hat{u}_X = \hat{u}_{\hat{\Sigma}_X,1},\quad \hat{\theta}_X = \hat{\lambda}_{\hat{\Sigma}_X,1}$$
$$\hat{\Sigma}_Y = \frac{\mathbf{Y}\mathbf{Y}^t}{n_X},\quad \hat{u}_Y = \hat{u}_{\hat{\Sigma}_Y,1},\quad \hat{\theta}_Y = \hat{\lambda}_{\hat{\Sigma}_Y,1}$$

1 : First we can without loss of generality apply the same invariant by rotation matrix to $\mathbf{X}$ and $\mathbf{Y}$ in order to change u by e_1. Then, the matrices without perturbation, $\mathbf{X}$ and $\mathbf{Y}$ are rotationally invariant but not independent anymore.

2 : We realize that only the expectation of $\hat{\alpha}$ changes compared to model B. (Indeed the expectations of $\hat{\hat{\theta}}_X$ and $\hat{\hat{\theta}}_X$ are asymptotically θ.)

$$\mathrm{E}[\hat{\alpha}] = \alpha_X \alpha_Y + \sum_{i=2}^{m} \mathrm{E}\left[\hat{u}_{Y,i}\hat{u}_{X,i}\right] = \alpha_X\alpha_Y + m\,\mathrm{E}\left[\hat{u}_{Y,2}\hat{u}_{X,2}\right].$$

The characterisation 5.11.1 of the perturbed eigenvector shows that the component i of the eigenvector is

$$\frac{\hat{u}_{X,s}}{\sqrt{1-\hat{u}_{X,1}^2}} = \frac{Q_{\mathbf{X},1,s}}{\sqrt{\sum_{t\neq 1}(Q_{\mathbf{X},1,s})^2}},$$
$$Q_{\mathbf{X},1,s} = \left(W_X\left(\theta \mathrm{I}_m - W_X\right)^{-1}\right)_{1,s},$$

where $W_X = \frac{1}{n_X}\mathbf{X}\mathbf{X}^t$.
Therefore, we have

$$\mathrm{E}\left[\frac{\hat{u}_{Y,2}\hat{u}_{X,2}}{\sqrt{1-\hat{u}_{Y,1}^2}\sqrt{1-\hat{u}_{X,1}^2}}\right] = \mathrm{E}\left[\frac{Q_{\mathbf{X},1,2}}{\sqrt{\sum_{t\neq 1}(Q_{\mathbf{X},1,t})^2}}\frac{Q_{\mathbf{Y},1,2}}{\sqrt{\sum_{t\neq 1}(Q_{\mathbf{Y},1,t})^2}}\right].$$

This last equation explains the T criterion. The scalar product is computed from $Q_{\mathbf{X}}$ and $Q_{\mathbf{Y}}$. Therefore, the criterion replacing $Q_{\mathbf{X}}$ and $Q_{\mathbf{Y}}$ by $Q_{\mathbf{X}}$ and $Q_{\mathbf{Y}}$ seems intuitive. In the next part, we will show that this replacement is valid.

3: The previous part express $\mathrm{E}\left[\hat{u}_{Y,2}\hat{u}_{X,2}\right]$ as a function of $Q_{\mathbf{X},1,2}$; however, we can only look at $Q_{\mathbf{X},2,3}$. To prove that replacing $\mathbf{X}$ by $\mathbf{X}$ has no asymptotic impact, we need to show

$$Q_{\mathbf{X},2,3} \sim Q_{\mathbf{X},2,3},$$

where

$$\begin{aligned} Q_{\mathbf{X},2,3} &= \left(W_X\left(\theta \mathrm{I}_m - W_X\right)^{-1}\right)_{2,3}, \\ Q_{\mathbf{X},2,3} &= \left(\left(\hat{\Sigma}_X - \hat{\theta}_X\hat{u}_X\hat{u}_X^t\right)\left(\theta\mathrm{I}_m - \left(\hat{\Sigma}_X - \hat{\theta}_X\hat{u}_X\hat{u}_X^t\right)\right)^{-1}\right)_{2,3}. \end{aligned}$$

To simplify equation, we use the following notation:

$$Q\left(M,\theta\right) = \left(M\left(\theta\mathrm{I}_m - M\right)^{-1}\right).$$

The equivalence is proven in three steps.

(a) $Q\left(W_X,\theta\right)_{2,3} = Q\left(\hat{\Sigma}_{X,-1,-1},\theta\right)_{1,2} + O_p\left(\frac{1}{\theta^2 m}\right)$, where $\hat{\Sigma}_{X,-1,-1}$ is the matrix $\hat{\Sigma}_X$ without the first row and the first column.

(b) $Q\left(\hat{\Sigma}_{X,-1,-1},\theta\right)_{1,2} = Q\left(\hat{\Sigma}_X,\theta\right)_{2,3} + O_p\left(\frac{1}{\theta m}\right)$.

(c) $Q\left(\hat{\Sigma}_X,\theta\right)_{2,3} = Q\left(\hat{\Sigma}_X - \hat{\theta}_X\hat{u}_X\hat{u}_X^t,\theta\right)_{2,3} + O_p\left(\frac{1}{\theta m}\right)$.

(a) $Q\left(W_X,\theta\right)_{2,3} = Q\left(\hat{\Sigma}_{X,-1,-1},\theta\right)_{1,2} + O_p\left(\frac{1}{\theta^2 m}\right)$.

Because $\hat{\Sigma}_X = P^{1/2}W_X P^{1/2}$ and P is a canonical perturbation of order 1, then

$$\Sigma_{X,-1,-1} = W_{X,-1,-1}.$$

Moreover,

$$\begin{aligned}
Q\left(W_X,\theta\right)_{2,3} &= \left(W_X\left(\theta \mathrm{I}_m - W_X\right)^{-1}\right)_{2,3} \\
&= \left(\begin{pmatrix} W_{X,1,1} & W_{X,1,-1} \\ W_{X,-1,1} & W_{X,-1,-1}\end{pmatrix}\left(\theta \mathrm{I}_m - \begin{pmatrix} W_{X,1,1} & W_{X,1,-1} \\ W_{X,-1,1} & W_{X,-1,-1}\end{pmatrix}\right)^{-1}\right)_{2,3} \\
&= \left(\begin{pmatrix} W_{X,1,1} & W_{X,1,-1} \\ W_{X,-1,1} & W_{X,-1,-1}\end{pmatrix}\begin{pmatrix} \theta - W_{X,1,1} & -W_{X,1,-1} \\ -W_{X,-1,1} & \theta \mathrm{I}_{m-1} - W_{X,-1,-1}\end{pmatrix}^{-1}\right)_{2,3}.
\end{aligned}$$

If $\kappa = (\theta - W_{X,1,1}) - W_{X,1,-1}(\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1}W_{X,-1,1}$, then $\kappa \sim \theta$ and

$$\begin{aligned}
&\begin{pmatrix} \theta - W_{X,1,1} & -W_{X,1,-1} \\ -W_{X,-1,1} & \theta \mathrm{I}_{m-1} - W_{X,-1,-1}\end{pmatrix}^{-1} \\
&= \begin{pmatrix} 1/\kappa & 1/\kappa(\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1}W_{X,1,-1} \\ 1/\kappa\left((\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1}W_{X,1,-1}\right)^t & (\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1} + 1/\kappa(\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1}W_{X,1,-1}W_{X,-1,1}(\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1}\end{pmatrix}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&Q\left(W_X,\theta\right)_{2,3} \\
&= \begin{pmatrix} W_{X,2,1} & W_{X,2,-1}\end{pmatrix}\begin{pmatrix} \frac{1}{\kappa}(\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1}W_{X,1,-1} \\ (\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1} + \frac{1}{\kappa}(\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1}W_{X,1,-1}W_{X,-1,1}(\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1}\end{pmatrix}_{\cdot,2} \\
&= W_{X,2,1}\frac{1}{\kappa}\left((\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1}W_{X,1,-1}\right)_2 + \left(W_{X,-1,-1}(\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1}\right)_{1,2} \\
&\quad + \left(W_{X,-1,-1}\frac{1}{\kappa}(\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1}W_{X,1,-1}W_{X,-1,1}(\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1}\right)_{1,2} \\
&= \left(W_{X,-1,-1}(\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1}\right)_{1,2} + O_p\left(\frac{1}{\theta^2 m}\right).
\end{aligned}$$

The last line is obtained in two steps.

- Using the invariance by rotation of $\vec{B} = \left((\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1}W_{X,1,-1}\right) \in \mathbb{R}^{m-1}$ implies that, for $s = 3, 4, ..., m$,

  $$B_2 = \left((\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1}W_{X,1,-1}\right)_1 \sim \left((\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1}W_{X,1,-1}\right)_{s-1} = B_s.$$

  Then,

  $$\begin{aligned}
  \left(\sum_{s=2}^{m}|B_s|\right)^2 &\leqslant m\left(\sum_{s=2}^{m}(B_s)^2\right) \\
  &= m||\vec{B}||^2 \\
  &= m(W_{X,1,-1})^t(\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-2}W_{X,1,-1} \\
  &\leqslant m\lambda_{\max}\left((\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1}\right)^2 ||W_{X,1,-1}||^2.
  \end{aligned}$$

  Because the spectrum of W_X is bounded, there exists C a constant such that $||W_{X,1,-1}||^2 < C^2$. Therefore,

  $$\mathrm{E}\left[|B_2|\right] \leqslant C\frac{\lambda_{\max}\left((\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1}\right)}{\sqrt{m}} + O_p\left(\frac{1}{\theta m}\right).$$

  Because $\kappa \sim \theta$ and $W_{X,2,1} = O_p\left(1/\sqrt{m}\right)$, then

  $$W_{X,2,1}1/\kappa\left((\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1}W_{X,1,-1}\right)_2 = O_p\left(\frac{1}{\theta^2 m}\right).$$

- By invariance under rotation, for $1 < i \neq j \leqslant m$,

$$\begin{aligned} & B_{2,3} \\ & \quad = \left(W_{X,-1,-1}(\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1} W_{X,1,-1} W_{X,-1,1} (\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1}\right)_{1,2} \\ & \sim \left(W_{X,-1,-1}(\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1} W_{X,1,-1} W_{X,-1,1} (\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1}\right)_{i-1,j-1} \\ & \quad = B_{i,j}. \end{aligned}$$

$$\begin{aligned} \left(\sum_{\substack{i,j=2 \\ i \neq j}}^{m} |B_{i,j}| \right)^2 & \leqslant \left(\sum_{i,j=2}^{m} (B_{i,j})^2 \right) m^2 \\ & = \mathrm{Trace}\left(B^2\right) m^2 \\ & \leqslant C \lambda_{\max} \left((\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1} W_{X,-1,-1} (\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1} \right)^2 m^2. \end{aligned}$$

Therefore

$$\mathrm{E}\left[|B_{2,3}|\right] \leqslant C \frac{\lambda_{\max}\left((\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1} W_{X,-1,-1} (\theta \mathrm{I}_{m-1} - W_{X,-1,-1})^{-1}\right)}{m} + o_p\left(\frac{1}{\theta^2 m}\right).$$

Because $\kappa \sim \theta$, the last term is of order $O_p\left(\frac{1}{\theta^3 m}\right)$.

This concludes the first step.

(b) $Q\left(\hat{\Sigma}_{X,-1,-1}, \theta\right)_{1,2} = Q\left(\hat{\Sigma}_X, \theta\right)_{2,3} + O_p\left(\frac{1}{\theta m}\right)$.

This equality is proven as in the previous step by replacing $W_{X,1,1}$ and $W_{X,-1,1}$ by $\hat{\Sigma}_{X,1,1}$ and $\hat{\Sigma}_{X,-1,1}$ respectively. In this case, the constant C change,

$$C = ||\hat{\Sigma}_{X,-1,1}|| = O_p(\theta).$$

However, we still can prove the result with a weaker power of $1/\theta$.

(c) $Q\left(\hat{\Sigma}_X, \theta\right)_{2,3} = Q\left(\hat{\Sigma}_X - \hat{\theta}_X \hat{u}_X \hat{u}_X^t, \theta\right)_{2,3} + O_p\left(\frac{1}{\theta m}\right)$.

We directly compute

$$\begin{aligned} Q\left(\hat{\Sigma}_X - \hat{\theta}_X \hat{u}_X \hat{u}_X^t, \theta\right)_{2,3} & = \left(\left(\hat{\Sigma}_X - \hat{\theta}_X \hat{u}_X \hat{u}_X^t\right) \left(\theta \mathrm{I}_m - \left(\hat{\Sigma}_X - \hat{\theta}_X \hat{u}_X \hat{u}_X^t\right)\right)^{-1} \right)_{2,3} \\ & = \left(\sum_{i=2}^{m} \frac{\hat{\lambda}_{\hat{\Sigma}_X,i}}{\theta - \hat{\lambda}_{\hat{\Sigma}_X,i}} \hat{u}_{\hat{\Sigma}_X,i} \hat{u}_{\hat{\Sigma}_X,i}^t \right)_{2,3} \\ & = \left(\hat{\Sigma}_X \left(\theta \mathrm{I}_m - \hat{\Sigma}_X\right)^{-1} \right)_{2,3} - \frac{\hat{\theta}}{\theta - \hat{\theta}} \hat{u}_2 \hat{u}_3 \\ & = \left(\hat{\Sigma}_X \left(\theta \mathrm{I}_m - \hat{\Sigma}_X\right)^{-1} \right)_{2,3} + O_p\left(\frac{1}{\theta m}\right). \end{aligned}$$

This concludes the third step.

3: Finally if T is the criterion,

$$\begin{aligned} \mathrm{E}\left[\hat{\alpha}\right] & = \alpha_X \alpha_Y + m\,\mathrm{E}\left[\frac{\hat{u}_{Y,2} \hat{u}_{X,2}}{\sqrt{1 - \hat{u}_{Y,1}^2} \sqrt{1 - \hat{u}_{X,1}^2}} \right] \sqrt{1 - \alpha_Y^2} \sqrt{1 - \alpha_X^2} + o_m(1) \\ & = \alpha_X \alpha_Y + m\,\mathrm{E}\left[\frac{Q_{\mathbf{X},1,2}}{\sqrt{\sum_{t \neq 1} (Q_{\mathbf{X},1,t})^2}} \frac{Q_{\mathbf{Y},1,2}}{\sqrt{\sum_{t \neq 1} (Q_{\mathbf{Y},1,t})^2}} \right] \sqrt{1 - \alpha_Y^2} \sqrt{1 - \alpha_X^2} + o_m(1) \\ & = \alpha_X \alpha_Y + T \sqrt{1 - \alpha_Y^2} \sqrt{1 - \alpha_X^2} + o_m(1). \end{aligned}$$

The theorem assumes the criterion defined in 4.2.5 μ_λ as strictly increasing with θ.
First, this condition proves that if $T = 0$, then $\mu_{\lambda_0} \leqslant \mu_{\lambda_D} + o_m(1)$.
Then, the expectation of the residual spike is strictly decreasing as a function of T. Therefore, the estimation μ_{λ_0}, neglecting strictly positive T, is asymptotically strictly larger than μ_{λ_D}.
Moreover, assuming that the variance of the angle tends to 0 proves the second point of the theorem.

□

7.3.6 Proof of Power

Theorem 4.3.1.
Suppose $\mathbf{X} = P^{1/2}\mathbf{X} \in \mathbb{R}^{m\times n_X}$ *and* $\mathbf{Y} = P^{1/2}\mathbf{Y} \in \mathbb{R}^{m\times n_Y}$ *are data such that for* $\mathbf{X} \in \mathbb{R}^{m\times n_X}$ *and* $\mathbf{Y} \in \mathbb{R}^{m\times n_Y}$ *2.2.1 and furthermore* $P \in \mathbb{R}^{m\times m}$ *satisfies Assumption 2.2.2(A4). We test*

$$\begin{aligned} &\mathrm{H}_0 : P_X = P_Y, \\ &\mathrm{H}_1 : P_Y = P_X + (\lambda - 1)vv^t, \end{aligned}$$

where P_X *is a finite perturbation such that*

$$P_X \quad = \quad \mathrm{I}_m + \sum_{i=1}^{k} (\theta_i - 1) u_i u_i^t.$$

We define the ***Power*** *as*

$$\beta_m = P_m \left\{ \mathrm{H}_0 \textit{ rejected} \right\}.$$

1. *Assuming the continuity properties defined in 4.3.1, then the* ***Power*** *tends to* 1 *in the following two cases:*
   (a) $\langle v, u_i\rangle = 0$ *for* $i = 1, 2, ..., k$,
   (b) $\sum_{i=1}^k \langle v, u_i\rangle^2 \neq 1$,
2. *Assuming the classical multivariate assumption,* $n_X, n_Y \to \infty$ *with* $\frac{m}{n_X}, \frac{m}{n_Y} \to 0$ *and a fixed dependence between the columns (temporal structure), then the* ***Power*** *tends to* 1 *in the two cases defined above.*
3. *Assuming* $m, n_X, n_Y \to \infty$ *with* $c_X = \frac{m}{n_X}$, $c_Y = \frac{m}{n_Y}$, *then*
   (a) *If* $\langle v, u_i\rangle = 0$ *for* $i = 1, 2, ..., k$, *and*

$$\lambda > M_2 + \sqrt{M_2^2 - 1},$$

where

$$M_2 = \frac{M_{2,X} + M_{2,Y}}{2}, \; M_{2,X} = \frac{1}{m}\sum_{i=1}^{m} \hat{\lambda}_{\mathbf{X},i}^2 \textit{ and } M_{2,Y} = \frac{1}{m}\sum_{i=1}^{m} \hat{\lambda}_{\mathbf{Y},i}^2,$$

the ***Power*** *tends to* 1.

(Page 52)

Proof. **Theorem 4.3.1**

We prove each part of Theorem 4.3.1 in a different order.

$\langle v, u_i \rangle = 0$:

- $m/n_X = c_X$ and $m/n_Y = c_Y$:
  Under H_0, the asymptotic residual spike tends to $M_2 + \sqrt{M_2^2 - 1}$. We just need to show that the asymptotic residual spike tends to be larger than λ under H_1.
  Under H_1, we build $\hat{\hat{\lambda}}$, the unbiased estimator of λ used in $\hat{\hat{\Sigma}}_Y$ with its corresponding eigenvector $\hat{v}$, then

$$\begin{aligned}
\lambda_{\max}\left(\hat{\hat{\Sigma}}_X^{-1/2}\hat{\hat{\Sigma}}_Y\hat{\hat{\Sigma}}_X^{-1/2}\right) &\geqslant \hat{v}^t\hat{\hat{\Sigma}}_X^{-1}\hat{\hat{\Sigma}}_Y\hat{v} \\
&\geqslant \hat{\hat{\lambda}}\hat{v}^t\hat{\hat{\Sigma}}_X^{-1}\hat{v} \\
&\geqslant \hat{\hat{\lambda}} + \hat{\hat{\lambda}}\sum_{i=1}^{k}(1/\hat{\hat{\theta}}_{X,i} - 1)\left\langle \hat{v}, \hat{u}_{X,i}\right\rangle^2 \\
&\geqslant \hat{\hat{\lambda}} + O_p\left(1/m\right).
\end{aligned}$$

  The first part of the result follows because $\hat{\hat{\lambda}}$ tends to λ. Therefore, when $\lambda > M_2 + \sqrt{M_2^2 - 1}$, the power tends to 1.
- c_X and c_Y tend to 0:
  When n_X and n_Y tend to infinity, we are in the classical theory. Under H_0, the residual spike tends to 1 and because the previous computation still holds under H_1, the power tends to 1.
- Continuity property:
  Assuming the continuity property, λ tends to infinity and the power tends to 1.

$\sum_{i=1}^k \langle v, u_i\rangle^2 \neq 1$:
In this case the difference is not orthogonal. We use the same computation as previously,

$$\begin{aligned}
\lambda_{\max}\left(\hat{\hat{\Sigma}}_X^{-1/2}\hat{\hat{\Sigma}}_Y\hat{\hat{\Sigma}}_X^{-1/2}\right) &\geqslant \hat{\hat{\lambda}} + \hat{\hat{\lambda}}\sum_{i=1}^{k}(1/\hat{\hat{\theta}}_{X,i} - 1)\left\langle \hat{v}, \hat{u}_{X,i}\right\rangle^2 \\
&\geqslant \hat{\hat{\lambda}} - \hat{\hat{\lambda}}\sum_{i=1}^{k}\left\langle \hat{v}, \hat{u}_{X,i}\right\rangle^2 \\
&\geqslant \hat{\hat{\lambda}}\left(1 - \sum_{i=1}^{k}\left\langle \hat{v}, \hat{u}_{X,i}\right\rangle^2\right).
\end{aligned}$$

- c_X and c_Y tends to 0:
  Under H_0, the residual spike tends to 1, but, under H_1, $\hat{\hat{\lambda}}$ does not tend to 1. Therefore the power tends to 1.
- Continuity property:
  Because $\sum_{i=1}^k \left\langle \hat{v}, \hat{u}_{X,i}\right\rangle^2 < 1$ and λ tends to infinity, then $\lambda_{\max}\left(\hat{\hat{\Sigma}}_X^{-1/2}\hat{\hat{\Sigma}}_Y\hat{\hat{\Sigma}}_X^{-1/2}\right)$ necessarily tends to infinity. Therefore the power tends to 1.

□

7.4 Application proofs

7.4.1 Spectrum estimation

Theorem 4.4.1.
Using the notation of Section 4.4.1 and assuming ϕ is an increasing function on the support of the spectrum, the following holds.

1. *If the values $\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}$ for $i = k_X, k_X + 1, ..., k$ are finite ($k \geqslant k_X$),*

$$0 < \frac{1}{m}\sum_{i=1}^{m} \phi\left(\hat{\lambda}_{W_X,i}\right) - \frac{1}{m-k_X}\sum_{i=k_X+1}^{m} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right) \leqslant O_p\left(\frac{1}{m}\right).$$

In particular if $\rho \in \mathbb{R}$ is outside the spectrum support, for $s_1, s_2 \geqslant 0$,

$$M_{s_1,s_2,W_X}(\rho) = \frac{1}{m}\sum_{i=1}^{m} \frac{\left(\hat{\lambda}_{W_X,i}\right)^{s_1}}{\left(\rho - \hat{\lambda}_{W_X,i}\right)^{s_2}} + O_p(1/m) = M_{s_1,s_2,\hat{\Sigma}_{\mathbf{X}}}(\rho) + O_p(1/m).$$

More precisely

$$\begin{aligned}&\frac{k_X \sum_{i=k_X+1}^{m} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right)}{m(m-k_X)}\\ &\qquad\leqslant \frac{1}{m}\sum_{i=1}^{m} \phi\left(\hat{\lambda}_{W_X,i}\right) - \frac{1}{m-k_X}\sum_{i=k_X+1}^{m} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right)\\ &\qquad\qquad\leqslant \frac{2k\phi\left(\hat{\lambda}_{W_X,1}\right) - \sum_{i=k_X+1}^{k} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right)}{m} + \frac{k_X \sum_{i=k_X+1}^{m} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right)}{m(m-k_X)}.\end{aligned}$$

2. *The estimator can be improved when $\phi(x) = x^2$ using another estimator. We assume that the variance of the entries of the data is σ^2 and $\sigma^2 - M_{1,\hat{\Sigma}_{\mathbf{X}}} = O_p\left(\frac{1}{mn_X}\right)$, then*

$$\begin{aligned}&\frac{1}{\sigma^2}\frac{1}{m}\sum_{i=1}^{m}\left(\hat{\lambda}_{W_X,i}\right)^2 - \frac{\sum_{i=k_X+1}^{m}\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}(k_X)\right)^2 + 2k_X\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}\right)^2}{(m-k_X)M_{1,\hat{\Sigma}_{\mathbf{X}}} + 2k_X\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}}\\ &\quad\leqslant \frac{2k\left(\hat{\lambda}_{W_X,1}\right)^2 - \sum_{i=k_X+1}^{k}\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right)^2 - 2k_X\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}\right)^2}{\sigma^2 m}\\ &\quad+\frac{\left(\sum_{i=k_X+1}^{m}\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right)^2\right)}{m\sigma^2 M_{1,\hat{\Sigma}_{\mathbf{X}}}}\\ &\qquad\left(M_{1,\hat{\Sigma}_{\mathbf{X}}} - \sigma^2 + \frac{-\sum_{i=k_X+1}^{k}\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i} + 2k_X\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1} - (k+1)M_{1,\hat{\Sigma}_{\mathbf{X}}}}{m}\right) + O_p\left(1/m^2\right).\end{aligned}$$

Remark 4.4.1.1.

1. The first inequality shows that the error of the intuitive estimator is of order $1/m$. Moreover, we always underestimate the true value and this underestimation can lead to a non-conservative test. Nevertheless, asymptotically, this estimation is enough and simulations show very good performance with reasonably large m.

2. We propose the second estimator to improve the conservative properties of the first one.

The first estimator underestimates $M_{2,X}$ with an error of size

$$\frac{2k\phi\left(\hat{\lambda}_{W_X,1}\right)-\sum_{i=k_X+1}^{k}\phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right)}{m}=\frac{2k\left(\hat{\lambda}_{W_X,1}\right)^2-\sum_{i=k_X+1}^{k}\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right)^2}{m}.$$

This estimator has also an error of order $1/m$ and tends to bound $M_{2,X}$. However, assuming $\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}$ close to $\hat{\lambda}_{W_X,1}$, the terms that contribute to the underestimation are

$$\frac{2k_X\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}}{m} \text{ and } \frac{(k-k_X)\left(\hat{\lambda}_{W_X,1}\right)^2}{\sigma^2 m}.$$

The error is of order $1/m$, but the numerator is smaller than $k\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}\right)^2$.

3. We do not treat the case $k_X > k$ because,
   - If $k_X \leqslant k$, the estimators are $O_p(1/m)$ when the missing perturbations are small. Moreover, underestimation of k leads to overestimation of M_2
   - If $k_X > k$, we tend to underestimate M_2 and the procedure loses the property of conservativeness.

   Before concluding that choosing $k_X < k$ is better, we should remember that this scenario creates errors on the statistics used for the Main Theorem 3.1.1. A better answer could be:

   Choose $\mathbf{k_X < k}$ to estimate the spectrum and $\mathbf{k_X > k}$ for the rest of the procedure.

   Of course all this discussion can be neglected when m is large enough.

(Page 54)

Proof. **Theorem 4.4.1**

We look at the eigenvalue of W_X and $P_k^{1/2}W_XP_k^{1/2}$. Without loss of generality we assume P_k is canonical.
We define $\tilde{P}_i = \mathrm{I}_m + (\theta_i - 1)e_ie_i^t$. Therefore,

$$P_k^{1/2}W_XP_k^{1/2} = \tilde{P}_k^{1/2}\tilde{P}_{k-1}^{1/2}\cdots\tilde{P}_1^{1/2}W_X\tilde{P}_1^{1/2}\cdots\tilde{P}_{k-1}^{1/2}\tilde{P}_k^{1/2}.$$

1. We first compute the error of the intuitive estimator. We start by assuming $k_X = k = 1$, then $k = k_X \geqslant 1$ and finally $k \geqslant k_X$.
   (a) First we show the result for a perturbation of order $k = 1$.

   $$\hat{\Sigma}_{P_1,X} = P_1^{1/2}W_XP_1^{1/2} = \begin{pmatrix}\hat{\Sigma}_{P_1,X,1,1} & \hat{\Sigma}_{P1,X,1,2:m}\\ \hat{\Sigma}_{P_1,X,2:m,1} & W_{X,2:m,2:m}\end{pmatrix}$$

   We use the notation $W_{X,-1,-1} = W_{X,2:m,2:m}$ to simplify equations. By the Cauchy interlacing law for symmetric matrices we have

   $\hat{\lambda}_{W_X,1} \geqslant \hat{\lambda}_{W_{X,-1,-1},1} \geqslant \hat{\lambda}_{W_X,2} \geqslant \hat{\lambda}_{W_{X,-1,-1},2} \geqslant \cdots \geqslant \hat{\lambda}_{W_X,m-1} \geqslant \hat{\lambda}_{W_{X,-1,-1},m-1} \geqslant \hat{\lambda}_{W_X,m}$.

   Therefore,

   $$0 \leqslant \sum_{i=1}^{m}\phi\left(\hat{\lambda}_{W_X,i}\right) - \sum_{i=1}^{m-1}\phi\left(\hat{\lambda}_{W_{X,-1,-1},i}\right) \leqslant \phi\left(\hat{\lambda}_{W_X,1}\right).$$

Using the same argument,

$$0 \leqslant \sum_{i=1}^{m-1} \phi\left(\hat{\lambda}_{W_X[-1,-1],i}\right) - \sum_{i=2}^{m} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{1,X},i}\right) \quad \leqslant \quad \phi\left(\hat{\lambda}_{W_{X,-1,-1},1}\right).$$

Therefore,

$$0 \quad \leqslant \quad \sum_{i=1}^{m} \phi\left(\hat{\lambda}_{W_X,i}\right) - \sum_{i=2}^{m} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{1,X},i}\right) \leqslant \phi\left(\hat{\lambda}_{W_X[-1,-1],1}\right) + \phi\left(\hat{\lambda}_{W_X,1}\right) \leqslant 2\phi\left(\hat{\lambda}_{W_X,1}\right).$$

Finally,

$$\begin{aligned} -\frac{\sum_{i=2}^{m} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{1,X},i}\right)}{m(m-1)} & \\ &\leqslant \frac{1}{m}\sum_{i=1}^{m} \phi\left(\hat{\lambda}_{W_X,i}\right) - \frac{1}{m-1}\sum_{i=2}^{m} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{1,X},i}\right) \\ &\qquad \leqslant \frac{2\phi\left(\hat{\lambda}_{W_X,1}\right)}{m} - \frac{\sum_{i=2}^{m} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{1,X},i}\right)}{m(m-1)}. \end{aligned}$$

(b) We can apply the same idea to perturbations of order k and assuming $k_X = k$.

$$\hat{\Sigma}_{\mathbf{X}} = P_k^{1/2} W_X P_k^{1/2} = \begin{pmatrix} \hat{\Sigma}_{\mathbf{X},1:k,1:k} & \hat{\Sigma}_{\mathbf{X},1,(k+1):m} \\ \hat{\Sigma}_{\mathbf{X},(k+1):m,1} & W_{X,(k+1):m,(k+1):m} \end{pmatrix}.$$

By successive applications of the Cauchy interlacing law,

$$0 \leqslant \sum_{i=1}^{m-s} \phi\left(\hat{\lambda}_{W_{X,-s,-s},i}\right) - \sum_{i=1}^{m-s-1} \phi\left(\hat{\lambda}_{W_{X,-s-1,-s-1},i}\right) \quad \leqslant \quad \phi\left(\hat{\lambda}_{W_{X,-s,-s},1}\right).$$

Therefore,

$$0 \leqslant \sum_{i=1}^{m} \phi\left(\hat{\lambda}_{W_X,i}\right) - \sum_{i=1}^{m-k} \phi\left(\hat{\lambda}_{W_{X,-k,-k},i}\right) \quad \leqslant \quad \sum_{s=0}^{k-1} \phi\left(\hat{\lambda}_{W_{X,-s,-s},1}\right).$$

Moreover, by successive Cauchy interlacing law,

$$0 \leqslant \sum_{i=k-s}^{m-s-1} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X},-s-1,-s-1},i}\right) - \sum_{i=k+1-s}^{m-s} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X},-s,-s},i}\right) \quad \leqslant \quad \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X},-s-1,-s-1},k-s}\right).$$

Then, because $\phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X},-s-1,-s-1},k-s}\right) \leqslant \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X},-k,-k},1}\right)$ and $\hat{\Sigma}_{\mathbf{X},-k,-k} = W_{X,-k,-k}$,

$$0 \leqslant \sum_{i=1}^{m-k} \phi\left(\hat{\lambda}_{W_{X,-k,-k},i}\right) - \sum_{i=k+1}^{m} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right) \quad \leqslant \quad k\phi\left(\hat{\lambda}_{W_{X,-k,-k},1}\right).$$

Finally,

$$\begin{aligned} 0 \leqslant \sum_{i=1}^{m} \phi\left(\hat{\lambda}_{W_X,i}\right) - \sum_{i=k+1}^{m} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right) \quad &\leqslant \quad \sum_{s=0}^{k-1} \phi\left(\hat{\lambda}_{W_{X,-s,-s},1}\right) + k\phi\left(\hat{\lambda}_{W_{X,-k,-k},1}\right) \\ &\leqslant \quad 2k\phi\left(\hat{\lambda}_{W_X,1}\right). \end{aligned}$$

(c) Suppose that we choose $k_X \leqslant k$ principal components. Then, the previous equation leads to

$$0 \leqslant \sum_{i=1}^{m} \phi\left(\hat{\lambda}_{W_X,i}\right) - \sum_{i=k_X+1}^{m} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right) \leqslant 2k\phi\left(\hat{\lambda}_{W_X,1}\right) - \sum_{i=k_X+1}^{k} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right).$$

Therefore

$$\begin{aligned} &\frac{k_X \sum_{i=k_X+1}^{m} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right)}{m(m-k_X)} \\ &\qquad \leqslant \frac{1}{m}\sum_{i=1}^{m} \phi\left(\hat{\lambda}_{W_X,i}\right) - \frac{1}{m-k_X}\sum_{i=k_X+1}^{m} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right) \\ &\qquad\qquad \leqslant \frac{2k\phi\left(\hat{\lambda}_{W_X,1}\right) - \sum_{i=k_X+1}^{k} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right)}{m} + \frac{k_X \sum_{i=k_X+1}^{m} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right)}{m(m-k_X)}. \end{aligned}$$

In particular, if the eigenvalues $\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}$ for $i = k_X, k_X+1, ..., k$ are finite,

$$0 < \frac{1}{m}\sum_{i=1}^{m} \phi\left(\hat{\lambda}_{W_X,i}\right) - \frac{1}{m-k_X}\sum_{i=k_X+1}^{m} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right) \leqslant O_p\left(\frac{1}{m}\right).$$

2. The error can be improved with the estimator

$$\frac{\sum_{i=k_X+1}^{m} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right) + 2k_X\phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}\right)}{(m-k_X)M_{1,\hat{\Sigma}_{\mathbf{X}}}(k_X) + 2k_X\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}}.$$

Suppose $\sigma = \frac{1}{n_X}\,\mathrm{E}\left[Tr\left(W_X\right)\right]$, then we compute the following error,

$$\begin{aligned} \epsilon &= \frac{1}{\sigma^2}\frac{1}{m}\sum_{i=1}^{m} \phi\left(\hat{\lambda}_{W_X,i}\right) - \frac{\sum_{i=k_X+1}^{m} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right) + 2k_X\phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}\right)}{(m-k_X)M_{1,\hat{\Sigma}_{\mathbf{X}}}(k_X) + 2k_X\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}} \\ &= \frac{\sum_{i=1}^{m} \phi\left(\hat{\lambda}_{W_X,i}\right) - \sum_{i=k_X+1}^{m} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right) - 2k_X\phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}\right)}{\sigma^2 m} \\ &\qquad + \left(\sum_{i=k_X+1}^{m} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right) + 2k_X\phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}\right)\right) \\ &\qquad\qquad \left(\frac{1}{\sigma^2 m} - \frac{1}{(m-k_X)M_{1,\hat{\Sigma}_{\mathbf{X}}}(k_X) + 2k_X\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}}\right). \end{aligned}$$

By the previous computations, we see that

$$\begin{aligned} &\sum_{i=1}^{m} \phi\left(\hat{\lambda}_{W_X,i}\right) - \sum_{i=k_X+1}^{m} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right) - 2k_X\phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}\right) \\ &\qquad \leqslant 2k\phi\left(\hat{\lambda}_{W_X,1}\right) - \sum_{i=k_X+1}^{k} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right) - 2k_X\phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}\right). \end{aligned}$$

and using $(m-k_X)M_{1,\hat{\Sigma}_{\mathbf{X}}}(k_X) = (m-k)M_{1,\hat{\Sigma}_{\mathbf{X}}}(k) - \sum_{i=k_X+1}^{k} \hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}$,

$$
\begin{aligned}
&\frac{1}{\sigma^2 m} - \frac{1}{(m-k_X)M_{1,\hat{\Sigma}_{\mathbf{X}}}(k_X) + 2k_X\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}} \\
&\quad = \frac{1}{\sigma^2 m} - \frac{1}{(m-k)M_{1,\hat{\Sigma}_{\mathbf{X}}}(k) - \sum_{i=k_X+1}^{k} \hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i} + 2k_X\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}} \\
&\quad = \frac{(m-k)\left(M_{1,\hat{\Sigma}_{\mathbf{X}}}(k) - \sigma^2\right) - \sum_{i=k_X+1}^{k} \hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i} + 2k_X\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1} - (k+1)\sigma^2}{m(m-k)\sigma^2 M_{1,\hat{\Sigma}_{\mathbf{X}}}(k) + m\sigma^2\left(2k_X\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1} - \sum_{i=k_X+1}^{k} \hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right)}.
\end{aligned}
$$

We assumed that $\sigma^2 - M_{1,\hat{\Sigma}_{\mathbf{X}}} = O_p\left(\frac{1}{m}\right)$; therefore,

$$
\begin{aligned}
&= \frac{(m-k)\left(M_{1,\hat{\Sigma}_{\mathbf{X}}}(k) - \sigma^2\right) - \sum_{i=k_X+1}^{k} \hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i} + 2k_X\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1} - (k+1)\sigma^2}{m^2\sigma^2 M_{1,\hat{\Sigma}_{\mathbf{X}}}} + O_p\left(\frac{1}{m^3}\right) \\
&\leqslant \frac{M_{1,\hat{\Sigma}_{\mathbf{X}}}(k) - \sigma^2}{m\sigma^2 M_{1,\hat{\Sigma}_{\mathbf{X}}}} + \frac{-\sum_{i=k_X+1}^{k} \hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i} + 2k_X\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1} - (k+1)M_{1,\hat{\Sigma}_{\mathbf{X}}}(k)}{m^2\sigma^2 M_{1,\hat{\Sigma}_{\mathbf{X}}}(k)} + O_p\left(\frac{1}{m^3}\right).
\end{aligned}
$$

Finally,

$$
\begin{aligned}
\epsilon \quad \leqslant \quad & \frac{2k\phi\left(\hat{\lambda}_{W_X,1}\right) - \sum_{i=k_X+1}^{k} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right) - 2k_X\phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1}\right)}{\sigma^2 m} \\
& + \frac{\left(\sum_{i=k_X+1}^{m} \phi\left(\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i}\right)\right)}{m\sigma^2 M_{1,\hat{\Sigma}_{\mathbf{X}}}} \\
& \quad \left(M_{1,\hat{\Sigma}_{\mathbf{X}}} - \sigma^2 + \frac{-\sum_{i=k_X+1}^{k} \hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},i} + 2k_X\hat{\lambda}_{\hat{\Sigma}_{\mathbf{X}},k_X+1} - (k+1)M_{1,\hat{\Sigma}_{\mathbf{X}}}(k)}{m}\right) + O_p\left(1/m^2\right).
\end{aligned}
$$

Replacing $\phi(x)$ by x^2 concludes the proof.

$\square$

Chapter 8

Simulations

In this section we present some simulation results in order to convince the reader that the most important theorems announced in the thesis are confirmed. In addition some conjectures, not proven in the thesis, are supported by the simulations. In order to highlight some weaknesses of the procedure developed in this thesis, some values in the tables will be in red.
In section 8.6.1 we give an example of analysis. All the useful routines are proposed in our R package "RMTResidualSpike".

8.1 Main Theorem

In this first section, we extend the simulations of Chapter 3. We test our Main Theorem 3.1.1 under different hypotheses on $\mathbf{X} \in \mathbf{R}^{m \times n_X}$ and $\mathbf{Y} \in \mathbf{R}^{m \times n_Y}$ (recall that $W_X = \frac{1}{n_X}\mathbf{X}\mathbf{X}^t$ and $W_Y = \frac{1}{n_Y}\mathbf{Y}\mathbf{Y}^t$):

1. The matrices $\mathbf{X}$ and $\mathbf{Y}$ contain independent standard normal entries.
2. The columns of the matrices $\mathbf{X}$ and $\mathbf{Y}$ are i.i.d. Multivariate Student with 8 degrees of freedom. For $i = 1, 2, ..., n_X$ and $j = 1, 2..., n_Y$,
$$X_{\cdot,i} \overset{i.i.d.}{\sim} \frac{\mathbf{N}\left(\vec{0}, \mathrm{I}_m\right)}{\sqrt{\frac{\chi_8^2}{8}}} \text{ and } Y_{\cdot,j} \overset{i.i.d.}{\sim} \frac{\mathbf{N}\left(\vec{0}, \mathrm{I}_m\right)}{\sqrt{\frac{\chi_8^2}{8}}}$$
3. The rows of $\mathbf{X}$ and $\mathbf{Y}$ are i.i.d. ARMA entries of parameters AR $= (0.6, 0.2)$ and MA $= (0.5, 0.2)$. Moreover, the traces of the matrices are standardised by the estimated variance.

The two Tables 8.1 and 8.2 compare the estimations of the mean and the standard error of the residual spikes $(\hat{\mu}, \hat{\sigma})$ to their empirical values (μ, σ).
The simulations are computed for the three scenarios described above. The perturbations $P_k = \mathrm{I}_m + \sum_{i=1}^k (\theta_i - 1) u_i u_i$ are without loss of generality assumed canonical and the eigenvalues θ_i are all fixed to 5000.
The numbers in red highlight bad results. Indeed, the number of data is large but the procedure is not conservative enough. We explain this red values by the strong temporal correlation of the data.

Details of the simulations of Tables 8.1 and 8.2

We apply our procedure to elliptical data after adding a perturbation $P_k = \mathrm{I}_m + \sum_{i=1}^k (\theta_i - 1) u_i u_i^t$ of order either $k = 1$, 4 or 10 with eigenvalues $\theta_i = 5000$.
This leads to observed residual spikes and repeating the experience 1000 times allows the estimation of their moments.
On the other hand, the theoretical moment of the extreme residual spikes are obtained by applying Theorem 3.1.1 using the spectra of $\mathbf{X}$ and $\mathbf{Y}$.

$$\lambda_{\max}\left(\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\hat{\hat{\Sigma}}_{Y,P_k}\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\right)$$

1. Normal entries.

n_Y \ n_X		100				500				1000				2000			
	k \ m	100		1000		100		1000		100		1000		100		1000	
		$(\hat{\mu}, \hat{\sigma})$	(μ, σ)	$(\hat{\mu}, \hat{\sigma})$	(μ, σ)	$(\hat{\mu}, \hat{\sigma})$	(μ, σ)	$(\hat{\mu}, \hat{\sigma})$	(μ, σ)	$(\hat{\mu}, \hat{\sigma})$	(μ, σ)	$(\hat{\mu}, \hat{\sigma})$	(μ, σ)	$(\hat{\mu}, \hat{\sigma})$	(μ, σ)	$(\hat{\mu}, \hat{\sigma})$	(μ, σ)
100	1	(3.69, 0.47)	(3.77, 0.48)	(21.91, 2.38)	(22.32, 2.36)	(2.88, 0.25)	(2.94, 0.25)	(13.93, 0.71)	(14.15, 0.69)	(2.76, 0.21)	(2.78, 0.21)	(13.01, 0.60)	(13.14, 0.57)	(2.73, 0.22)	(2.75, 0.23)	(12.39, 0.52)	(12.59, 0.50)
	4	(4.67, 0.41)	(4.85, 0.49)	(26.66, 1.92)	(28.21, 2.84)	(3.37, 0.20)	(3.39, 0.22)	(15.32, 0.56)	(15.50, 0.60)	(3.18, 0.18)	(3.19, 0.18)	(14.12, 0.45)	(14.30, 0.44)	(3.15, 0.19)	(3.17, 0.20)	(13.40, 0.38)	(13.52, 0.41)
	10	(5.64, 0.40)	(6.29, 0.71)	(31.28, 1.77)	(35.98, 3.38)	(3.87, 0.18)	(3.88, 0.21)	(16.66, 0.47)	(16.95, 0.55)	(3.60, 0.16)	(3.56, 0.15)	(15.24, 0.40)	(15.39, 0.45)	(3.59, 0.16)	(3.56, 0.16)	(14.37, 0.34)	(14.56, 0.34)
500	1					(1.87, 0.12)	(1.87, 0.12)	(5.85, 0.33)	(5.86, 0.34)	(1.72, 0.09)	(1.73, 0.09)	(4.79, 0.20)	(4.81, 0.21)	(1.63, 0.07)	(1.63, 0.08)	(4.28, 0.15)	(4.29, 0.16)
	4					(2.10, 0.09)	(2.11, 0.10)	(6.45, 0.24)	(6.58, 0.30)	(1.90, 0.07)	(1.90, 0.08)	(5.19, 0.16)	(5.22, 0.17)	(1.78, 0.06)	(1.77, 0.06)	(4.56, 0.11)	(4.56, 0.12)
	10					(2.32, 0.09)	(2.33, 0.10)	(7.06, 0.21)	(7.21, 0.29)	(2.07, 0.07)	(2.06, 0.07)	(5.57, 0.14)	(5.63, 0.17)	(1.92, 0.05)	(1.90, 0.05)	(4.85, 0.10)	(4.87, 0.11)
1000	1									(1.57, 0.07)	(1.57, 0.07)	(3.73, 0.15)	(3.74, 0.15)	(1.48, 0.06)	(1.47, 0.06)	(3.19, 0.10)	(3.19, 0.11)
	4									(1.71, 0.06)	(1.70, 0.06)	(4.02, 0.11)	(4.04, 0.13)	(1.58, 0.04)	(1.58, 0.04)	(3.39, 0.08)	(3.39, 0.08)
	10									(1.84, 0.05)	(1.84, 0.05)	(4.31, 0.10)	(4.36, 0.12)	(1.69, 0.04)	(1.68, 0.04)	(3.58, 0.07)	(3.59, 0.08)
2000	1													(1.37, 0.04)	(1.37, 0.04)	(2.62, 0.08)	(2.63, 0.08)
	4													(1.46, 0.03)	(1.45, 0.03)	(2.77, 0.06)	(2.77, 0.06)
	10													(1.54, 0.03)	(1.53, 0.03)	(2.92, 0.05)	(2.93, 0.06)

2. Multivariate Student.

n_Y \ n_X		100				500				1000				2000			
	k \ m	100		1000		100		1000		100		1000		100		1000	
		$(\hat{\mu}, \hat{\sigma})$	(μ, σ)	$(\hat{\mu}, \hat{\sigma})$	(μ, σ)	$(\hat{\mu}, \hat{\sigma})$	(μ, σ)	$(\hat{\mu}, \hat{\sigma})$	(μ, σ)	$(\hat{\mu}, \hat{\sigma})$	(μ, σ)	$(\hat{\mu}, \hat{\sigma})$	(μ, σ)	$(\hat{\mu}, \hat{\sigma})$	(μ, σ)	$(\hat{\mu}, \hat{\sigma})$	(μ, σ)
100	1	(5.00, 1.23)	(5.37, 1.24)	(31.66, 7.70)	(33.68, 7.67)	(3.27, 0.37)	(3.35, 0.36)	(18.45, 2.87)	(19.15, 2.96)	(3.24, 0.39)	(3.33, 0.40)	(16.33, 1.70)	(16.69, 1.52)	(3.16, 0.42)	(3.27, 0.40)	(18.30, 3.61)	(18.72, 3.24)
	4	(7.38, 1.03)	(7.56, 1.32)	(47.52, 6.91)	(49.82, 9.18)	(4.00, 0.31)	(4.07, 0.35)	(24.02, 2.34)	(24.75, 3.40)	(4.04, 0.33)	(4.13, 0.38)	(19.53, 1.30)	(19.89, 1.65)	(3.95, 0.33)	(3.98, 0.41)	(25.24, 3.02)	(25.23, 3.76)
	10	(9.76, 1.03)	(10.40, 1.71)	(62.93, 6.28)	(69.96, 9.63)	(4.75, 0.29)	(4.76, 0.33)	(29.61, 2.13)	(31.50, 3.54)	(4.86, 0.32)	(4.86, 0.36)	(22.77, 1.18)	(23.40, 1.70)	(4.78, 0.32)	(4.83, 0.41)	(32.12, 2.80)	(32.94, 3.51)
500	1					(2.16, 0.17)	(2.15, 0.16)	(7.56, 0.62)	(7.62, 0.57)	(1.92, 0.13)	(1.91, 0.13)	(6.12, 0.42)	(6.22, 0.43)	(1.81, 0.11)	(1.84, 0.11)	(6.15, 0.90)	(6.32, 1.02)
	4					(2.49, 0.14)	(2.47, 0.16)	(8.76, 0.50)	(8.87, 0.61)	(2.16, 0.10)	(2.14, 0.10)	(6.96, 0.33)	(7.00, 0.40)	(2.03, 0.09)	(2.03, 0.10)	(8.03, 0.81)	(8.12, 1.14)
	10					(2.83, 0.12)	(2.79, 0.15)	(9.99, 0.45)	(10.17, 0.63)	(2.40, 0.09)	(2.36, 0.10)	(7.77, 0.30)	(7.99, 0.43)	(2.24, 0.08)	(2.24, 0.10)	(9.99, 0.72)	(10.69, 1.41)
1000	1									(1.72, 0.10)	(1.71, 0.09)	(5.28, 0.87)	(5.34, 0.84)	(1.70, 0.17)	(1.70, 0.16)	(3.94, 0.21)	(3.92, 0.21)
	4									(1.91, 0.08)	(1.89, 0.08)	(7.05, 0.78)	(7.22, 1.50)	(2.03, 0.15)	(2.01, 0.21)	(4.36, 0.17)	(4.41, 0.24)
	10									(2.09, 0.07)	(2.07, 0.08)	(8.80, 0.71)	(9.92, 1.83)	(2.37, 0.14)	(2.42, 0.24)	(4.77, 0.15)	(4.91, 0.29)
2000	1													(1.46, 0.06)	(1.46, 0.05)	(3.17, 0.13)	(3.18, 0.12)
	4													(1.57, 0.04)	(1.56, 0.05)	(3.43, 0.10)	(3.43, 0.10)
	10													(1.68, 0.04)	(1.66, 0.05)	(3.66, 0.09)	(3.67, 0.10)

3. ARMA $((0.6, 0.2), (0.5, 0.2))$.

n_Y \ n_X		100				500				1000				2000			
	k \ m	100		1000		100		1000		100		1000		100		1000	
		$(\hat{\mu}, \hat{\sigma})$	(μ, σ)	$(\hat{\mu}, \hat{\sigma})$	(μ, σ)	$(\hat{\mu}, \hat{\sigma})$	(μ, σ)	$(\hat{\mu}, \hat{\sigma})$	(μ, σ)	$(\hat{\mu}, \hat{\sigma})$	(μ, σ)	$(\hat{\mu}, \hat{\sigma})$	(μ, σ)	$(\hat{\mu}, \hat{\sigma})$	(μ, σ)	$(\hat{\mu}, \hat{\sigma})$	(μ, σ)
100	1	(28.96, 12.69)	(36.79, 16.64)	(251.42, 105.05)	(333.04, 156.12)	(19.48, 7.79)	(25.23, 9.14)	(154.16, 50.52)	(187.38, 61.43)	(15.59, 5.62)	(18.96, 5.99)	(139.57, 46.00)	(170.16, 49.33)	(15.10, 5.05)	(18.11, 5.88)	(132.97, 47.80)	(163.83, 51.00)
	4	(54.47, 13.21)	(65.76, 25.01)	(466.97, 110.66)	(570.47, 218.52)	(35.83, 7.83)	(41.70, 8.71)	(254.86, 48.94)	(290.63, 51.26)	(27.23, 5.62)	(30.07, 5.66)	(231.95, 44.43)	(261.05, 38.51)	(25.11, 4.72)	(28.40, 4.30)	(227.38, 43.68)	(251.50, 37.94)
	10	(84.26, 13.36)	(92.84, 28.21)	(703.57, 108.72)	(791.92, 255.82)	(51.87, 7.68)	(55.35, 8.47)	(369.98, 50.60)	(388.93, 55.27)	(39.91, 6.00)	(40.14, 4.46)	(332.81, 44.09)	(331.52, 33.62)	(36.21, 4.74)	(34.65, 2.84)	(325.40, 43.65)	(313.54, 24.40)
500	1					(7.24, 1.55)	(7.70, 1.67)	(55.71, 11.52)	(58.73, 12.43)	(5.86, 1.06)	(6.15, 1.21)	(42.51, 7.07)	(45.19, 8.13)	(5.29, 0.96)	(5.57, 1.01)	(36.00, 5.82)	(36.68, 5.54)
	4					(10.36, 1.37)	(10.66, 1.68)	(78.46, 9.99)	(82.43, 13.21)	(8.09, 0.97)	(8.38, 1.13)	(57.47, 6.27)	(59.21, 7.58)	(7.28, 0.85)	(7.40, 0.96)	(46.98, 4.82)	(48.43, 4.76)
	10					(13.38, 1.31)	(14.76, 1.97)	(100.86, 8.87)	(112.61, 16.04)	(10.30, 0.93)	(10.47, 1.20)	(71.75, 5.73)	(74.17, 6.73)	(9.25, 0.82)	(9.26, 0.90)	(58.59, 4.59)	(58.84, 4.30)
1000	1									(4.46, 0.66)	(4.52, 0.73)	(28.97, 3.98)	(29.43, 4.26)	(3.83, 0.48)	(3.93, 0.54)	(22.44, 2.76)	(23.08, 2.87)
	4									(5.77, 0.57)	(5.82, 0.61)	(36.95, 3.37)	(37.93, 3.71)	(4.79, 0.42)	(4.83, 0.47)	(28.01, 2.35)	(28.29, 2.38)
	10									(7.10, 0.54)	(7.34, 0.73)	(44.83, 3.18)	(47.80, 4.46)	(5.81, 0.39)	(5.84, 0.46)	(33.31, 2.08)	(33.98, 2.39)
2000	1													(3.09, 0.37)	(3.17, 0.34)	(15.35, 1.61)	(15.58, 1.60)
	4													(3.78, 0.29)	(3.82, 0.34)	(18.45, 1.25)	(18.69, 1.43)
	10													(4.51, 0.26)	(4.62, 0.36)	(21.45, 1.13)	(22.16, 1.49)

Table 8.1 – Simulations of the maximum residual spike. The values $(\hat{\mu}, \hat{\sigma})$ and (μ, σ) are respectively the estimations of the mean and the standard deviation of the residual spikes obtained by respectively the Main Theorem 3.1.1 and empirical methods. In red we see that some results are less accurate despite the large values of n_X and n_Y. This is probably due to the temporal correlation of the data that reduces the number of independent columns. (Number of replicates: 500, $\theta_i = 5000$)

$$\lambda_{\min}\left(\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\hat{\hat{\Sigma}}_{Y,P_k}\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\right)$$

1. Normal entries.

n_Y \ n_X		100				500				1000				2000			
	k \ m	100		1000		100		1000		100		1000		100		1000	
		$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)
100	1	(0.269, 0.035)	(0.278, 0.035)	(0.045, 0.005)	(0.045, 0.005)	(0.347, 0.031)	(0.348, 0.042)	(0.072, 0.003)	(0.070, 0.009)	(0.363, 0.028)	(0.353, 0.043)	(0.077, 0.003)	(0.076, 0.010)	(0.368, 0.031)	(0.362, 0.042)	(0.080, 0.003)	(0.079, 0.011)
	4	(0.201, 0.026)	(0.222, 0.022)	(0.036, 0.004)	(0.036, 0.003)	(0.271, 0.027)	(0.272, 0.028)	(0.056, 0.005)	(0.055, 0.005)	(0.286, 0.027)	(0.276, 0.029)	(0.060, 0.005)	(0.058, 0.006)	(0.289, 0.029)	(0.283, 0.030)	(0.063, 0.006)	(0.060, 0.006)
	10	(0.138, 0.022)	(0.170, 0.016)	(0.027, 0.003)	(0.028, 0.003)	(0.192, 0.026)	(0.207, 0.021)	(0.039, 0.005)	(0.041, 0.004)	(0.204, 0.025)	(0.215, 0.021)	(0.042, 0.005)	(0.042, 0.004)	(0.204, 0.027)	(0.214, 0.021)	(0.043, 0.006)	(0.044, 0.005)
500	1					(0.535, 0.034)	(0.544, 0.035)	(0.171, 0.009)	(0.171, 0.010)	(0.579, 0.030)	(0.579, 0.033)	(0.208, 0.009)	(0.208, 0.011)	(0.612, 0.028)	(0.619, 0.030)	(0.234, 0.008)	(0.233, 0.013)
	4					(0.471, 0.026)	(0.484, 0.021)	(0.154, 0.007)	(0.154, 0.007)	(0.519, 0.025)	(0.519, 0.023)	(0.188, 0.008)	(0.187, 0.007)	(0.554, 0.023)	(0.566, 0.023)	(0.212, 0.007)	(0.210, 0.009)
	10					(0.409, 0.023)	(0.436, 0.018)	(0.136, 0.006)	(0.138, 0.005)	(0.456, 0.022)	(0.469, 0.020)	(0.166, 0.007)	(0.168, 0.007)	(0.495, 0.021)	(0.517, 0.018)	(0.187, 0.007)	(0.187, 0.007)
1000	1									(0.637, 0.028)	(0.641, 0.028)	(0.269, 0.011)	(0.268, 0.010)	(0.676, 0.025)	(0.682, 0.026)	(0.314, 0.010)	(0.313, 0.012)
	4									(0.582, 0.022)	(0.587, 0.019)	(0.247, 0.008)	(0.248, 0.008)	(0.625, 0.020)	(0.635, 0.021)	(0.291, 0.008)	(0.289, 0.009)
	10									(0.528, 0.020)	(0.547, 0.017)	(0.227, 0.007)	(0.230, 0.006)	(0.573, 0.018)	(0.594, 0.017)	(0.268, 0.007)	(0.269, 0.007)
2000	1													(0.730, 0.023)	(0.729, 0.021)	(0.382, 0.011)	(0.383, 0.011)
	4													(0.684, 0.018)	(0.689, 0.016)	(0.360, 0.009)	(0.362, 0.008)
	10													(0.641, 0.016)	(0.655, 0.013)	(0.339, 0.007)	(0.342, 0.007)

2. Multivariate Student.

n_Y \ n_X		100				500				1000				2000			
	k \ m	100		1000		100		1000		100		1000		100		1000	
		$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)
100	1	(0.203, 0.047)	(0.217, 0.034)	(0.031, 0.007)	(0.035, 0.005)	(0.309, 0.035)	(0.311, 0.042)	(0.054, 0.008)	(0.058, 0.009)	(0.307, 0.038)	(0.308, 0.042)	(0.062, 0.006)	(0.056, 0.008)	(0.318, 0.040)	(0.299, 0.040)	(0.055, 0.011)	(0.054, 0.009)
	4	(0.120, 0.034)	(0.162, 0.021)	(0.019, 0.005)	(0.026, 0.004)	(0.229, 0.029)	(0.242, 0.026)	(0.038, 0.006)	(0.043, 0.005)	(0.226, 0.032)	(0.235, 0.026)	(0.046, 0.005)	(0.043, 0.005)	(0.232, 0.031)	(0.233, 0.025)	(0.036, 0.007)	(0.039, 0.005)
	10	(0.049, 0.028)	(0.126, 0.014)	(0.008, 0.004)	(0.020, 0.002)	(0.152, 0.026)	(0.185, 0.021)	(0.022, 0.005)	(0.033, 0.004)	(0.148, 0.027)	(0.181, 0.019)	(0.030, 0.005)	(0.032, 0.003)	(0.151, 0.029)	(0.177, 0.019)	(0.019, 0.006)	(0.029, 0.004)
500	1					(0.465, 0.036)	(0.440, 0.039)	(0.132, 0.011)	(0.132, 0.011)	(0.522, 0.034)	(0.535, 0.039)	(0.163, 0.012)	(0.147, 0.011)	(0.550, 0.032)	(0.515, 0.032)	(0.161, 0.025)	(0.160, 0.017)
	4					(0.386, 0.030)	(0.375, 0.025)	(0.110, 0.008)	(0.111, 0.009)	(0.446, 0.029)	(0.464, 0.032)	(0.141, 0.009)	(0.129, 0.007)	(0.484, 0.026)	(0.458, 0.024)	(0.119, 0.018)	(0.131, 0.014)
	10					(0.312, 0.027)	(0.322, 0.021)	(0.088, 0.007)	(0.093, 0.008)	(0.368, 0.026)	(0.404, 0.029)	(0.120, 0.007)	(0.114, 0.005)	(0.419, 0.023)	(0.409, 0.020)	(0.082, 0.015)	(0.108, 0.009)
1000	1									(0.583, 0.033)	(0.562, 0.034)	(0.190, 0.031)	(0.196, 0.022)	(0.587, 0.059)	(0.628, 0.036)	(0.255, 0.014)	(0.249, 0.013)
	4									(0.517, 0.026)	(0.504, 0.026)	(0.134, 0.022)	(0.159, 0.021)	(0.494, 0.045)	(0.560, 0.032)	(0.228, 0.011)	(0.225, 0.009)
	10									(0.450, 0.023)	(0.452, 0.023)	(0.085, 0.018)	(0.126, 0.015)	(0.409, 0.037)	(0.515, 0.023)	(0.203, 0.008)	(0.204, 0.009)
2000	1													(0.684, 0.026)	(0.676, 0.027)	(0.315, 0.013)	(0.310, 0.013)
	4													(0.631, 0.021)	(0.631, 0.020)	(0.289, 0.010)	(0.285, 0.010)
	10													(0.580, 0.019)	(0.589, 0.019)	(0.264, 0.008)	(0.262, 0.009)

3. ARMA ((0.6, 0.2), (0.5, 0.2)).

n_Y \ n_X		100				500				1000				2000			
	k \ m	100		1000		100		1000		100		1000		100		1000	
		$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)
100	1	(0.035, 0.015)	(0.033, 0.017)	(0.004, 0.002)	(0.004, 0.002)	(0.052, 0.020)	(0.041, 0.017)	(0.007, 0.002)	(0.005, 0.002)	(0.065, 0.025)	(0.054, 0.020)	(0.007, 0.002)	(0.006, 0.002)	(0.067, 0.022)	(0.057, 0.021)	(0.007, 0.003)	(0.006, 0.002)
	4	(0.005, 0.012)	(0.016, 0.005)	(0.001, 0.001)	(0.002, 0.001)	(0.012, 0.015)	(0.023, 0.007)	(0.002, 0.002)	(0.003, 0.001)	(0.017, 0.019)	(0.030, 0.008)	(0.002, 0.002)	(0.003, 0.001)	(0.021, 0.018)	(0.030, 0.008)	(0.002, 0.002)	(0.003, 0.001)
	10	(−0.025, 0.010)	(0.011, 0.003)	(−0.002, 0.001)	(0.001, 0.000)	(−0.026, 0.012)	(0.016, 0.004)	(−0.002, 0.001)	(0.002, 0.001)	(−0.027, 0.016)	(0.020, 0.005)	(−0.003, 0.002)	(0.002, 0.001)	(−0.024, 0.014)	(0.021, 0.005)	(−0.003, 0.002)	(0.003, 0.001)
500	1					(0.137, 0.030)	(0.136, 0.027)	(0.018, 0.004)	(0.018, 0.004)	(0.170, 0.032)	(0.164, 0.032)	(0.023, 0.004)	(0.023, 0.004)	(0.189, 0.035)	(0.184, 0.034)	(0.028, 0.005)	(0.028, 0.005)
	4					(0.082, 0.022)	(0.095, 0.014)	(0.011, 0.003)	(0.013, 0.002)	(0.106, 0.025)	(0.119, 0.018)	(0.016, 0.003)	(0.017, 0.002)	(0.121, 0.026)	(0.131, 0.017)	(0.019, 0.003)	(0.020, 0.003)
	10					(0.027, 0.019)	(0.070, 0.009)	(0.004, 0.002)	(0.009, 0.001)	(0.047, 0.021)	(0.089, 0.012)	(0.008, 0.003)	(0.012, 0.002)	(0.057, 0.023)	(0.098, 0.012)	(0.010, 0.003)	(0.015, 0.002)
1000	1									(0.225, 0.034)	(0.223, 0.032)	(0.035, 0.005)	(0.035, 0.005)	(0.265, 0.035)	(0.258, 0.033)	(0.045, 0.006)	(0.044, 0.006)
	4									(0.158, 0.026)	(0.173, 0.019)	(0.025, 0.004)	(0.027, 0.003)	(0.194, 0.026)	(0.203, 0.022)	(0.033, 0.004)	(0.035, 0.003)
	10									(0.093, 0.023)	(0.135, 0.013)	(0.016, 0.003)	(0.021, 0.002)	(0.123, 0.023)	(0.162, 0.014)	(0.023, 0.004)	(0.028, 0.002)
2000	1													(0.322, 0.036)	(0.325, 0.037)	(0.065, 0.007)	(0.065, 0.007)
	4													(0.250, 0.028)	(0.263, 0.023)	(0.052, 0.005)	(0.053, 0.004)
	10													(0.180, 0.025)	(0.220, 0.017)	(0.040, 0.004)	(0.046, 0.003)

Table 8.2 – Simulations of the minimum residual spike. The values $(\hat{\mu},\hat{\sigma})$ and (μ,σ) are respectively the estimations of the mean and the standard deviation of the residual spikes obtained by respectively the Main Theorem 3.1.1 and empirical methods. (Number of replicates: 500, $\theta_i = 5000$)

8.1.1 Blue and orange estimations of the residual spike

Suppose $n_X > n_Y$. We easily see that

$$\lambda_{\max}\left(\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\hat{\hat{\Sigma}}_{Y,P_k}\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\right) = \frac{1}{\lambda_{\min}\left(\hat{\hat{\Sigma}}_{Y,P_k}^{-1/2}\hat{\hat{\Sigma}}_{X,P_k}\hat{\hat{\Sigma}}_{Y,P_k}^{-1/2}\right)}, \tag{8.1}$$

$$\lambda_{\min}\left(\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\hat{\hat{\Sigma}}_{Y,P_k}\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\right) = \frac{1}{\lambda_{\max}\left(\hat{\hat{\Sigma}}_{Y,P_k}^{-1/2}\hat{\hat{\Sigma}}_{X,P_k}\hat{\hat{\Sigma}}_{Y,P_k}^{-1/2}\right)}. \tag{8.2}$$

Using the Main Theorem, estimation of the blue or the orange values leads to an error of order $O_p(1/m)$ in the equalities. We already observed this difference in Section 3.1.1. The Table 8.3 compares the estimations of the blue and the orange values assuming that n_X is larger than n_Y.
Recall that we try to estimate the distributions of $\lambda_{\min}\left(\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\hat{\hat{\Sigma}}_{Y,P_k}\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\right)$ and $\lambda_{\max}\left(\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\hat{\hat{\Sigma}}_{Y,P_k}\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\right)$. The simulations results show that conservative properties can be maintained if we observe the following rules.

- Estimate the maximum by applying the Main Theorem to

$$\lambda_{\min}\left(\hat{\hat{\Sigma}}_{Y,P_k}^{-1/2}\hat{\hat{\Sigma}}_{X,P_k}\hat{\hat{\Sigma}}_{Y,P_k}^{-1/2}\right).$$

  Then invert the distribution.

- Estimate the minimum by applying the Main Theorem to

$$\lambda_{\min}\left(\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\hat{\hat{\Sigma}}_{Y,P_k}\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\right).$$

On the other hand, when both n_X and n_Y are sufficiently large, it is seems reasonable to use only the blue estimators for the minimum and the maximum.
Although inversion of the estimates leads to conservative properties, it usually also leads to worse estimates.

Details of the simulations of Table 8.3

We apply our procedure to normal data after adding perturbations $P_k = \mathrm{I}_m + \sum_{i=1}^{k}(\theta_i - 1)u_i u_i^t$ of order $k = 1$, 4 or 10 with eigenvalues $\theta_i = 5000$.
Repeating the experience 1000 times, we can estimate the moments of the observed extreme residual spikes.
The observed residual spikes are obtained using $\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\hat{\hat{\Sigma}}_{Y,P_k}\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}$. On the other hand, the theoretical residual spikes are obtained by two different ways. Either we apply our procedure to $\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\hat{\hat{\Sigma}}_{Y,P_k}\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}$, or to $\hat{\hat{\Sigma}}_{Y,P_k}^{-1/2}\hat{\hat{\Sigma}}_{X,P_k}\hat{\hat{\Sigma}}_{Y,P_k}^{-1/2}$. In this last orange case, we invert the estimations.

8.2 Robustness

In this section we present simulation results to check the validity of our procedure by simulating some scenarios assuming that the columns of $\mathbf{X}$ and $\mathbf{Y}$ have i.i.d. entries with a particular distribution and the data have the form:

1. $\mathbf{X} = P^{1/2}\mathbf{X}$ and $\mathbf{Y} = P^{1/2}\mathbf{X}$.
2. $\mathbf{X} = P^{1/2}\Sigma^{1/2}\mathbf{X}\Psi_X^{1/2}$ and $\mathbf{Y} = P^{1/2}\Sigma^{1/2}\mathbf{X}\Psi_Y^{1/2}$, where $\Psi_X \in \mathbb{R}^{n_X \times n_X}$ and $\Psi_Y \in \mathbb{R}^{n_Y \times n_Y}$ are temporal covariances that do not create spikes and $\Sigma \in \mathbb{R}^{m \times m}$ is a spatial covariance matrix that does not create spikes.

$$\lambda_{\max}\left(\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\hat{\hat{\Sigma}}_{Y,P_k}\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\right)$$

n_Y \ n_X		100				400				800			
	m	100		400		100		400		100		400	
	k	4	10	4	10	4	10	4	10	4	10	4	10
100	(μ,σ)	(4.86, 0.50)	(6.24, 0.64)	(12.87, 1.28)	(16.54, 1.69)	(3.38, 0.23)	(3.86, 0.21)	(7.90, 0.41)	(8.86, 0.41)	(3.25, 0.20)	(3.67, 0.18)	(7.22, 0.34)	(7.95, 0.29)
	$(\hat{\mu},\hat{\sigma})$	(4.69, 0.40)	(5.63, 0.38)	(12.21, 0.92)	(14.46, 0.84)	(3.35, 0.21)	(3.84, 0.19)	(7.77, 0.37)	(8.65, 0.32)	(3.23, 0.19)	(3.69, 0.17)	(7.12, 0.30)	(7.87, 0.27)
	$(\hat{\mu},\hat{\sigma})$	(5.07, 0.73)	(7.60, 1.47)	(13.00, 1.53)	(18.01, 2.51)	(3.58, 0.39)	(4.53, 0.53)	(8.41, 0.94)	(9.99, 0.98)	(3.47, 0.39)	(4.34, 0.49)	(7.75, 0.86)	(9.09, 0.88)
400	(μ,σ)					(2.31, 0.13)	(2.58, 0.12)	(4.22, 0.21)	(4.76, 0.21)	(2.04, 0.08)	(2.23, 0.08)	(3.54, 0.14)	(3.87, 0.13)
	$(\hat{\mu},\hat{\sigma})$					(2.30, 0.12)	(2.57, 0.10)	(4.19, 0.19)	(4.65, 0.17)	(2.04, 0.08)	(2.24, 0.08)	(3.53, 0.13)	(3.84, 0.11)
	$(\hat{\mu},\hat{\sigma})$					(2.35, 0.15)	(2.77, 0.19)	(4.26, 0.24)	(4.93, 0.27)	(2.08, 0.11)	(2.37, 0.13)	(3.60, 0.18)	(4.04, 0.19)
800	(μ,σ)									(1.80, 0.07)	(1.95, 0.07)	(2.87, 0.11)	(3.13, 0.10)
	$(\hat{\mu},\hat{\sigma})$									(1.80, 0.07)	(1.96, 0.06)	(2.86, 0.10)	(3.10, 0.09)
	$(\hat{\mu},\hat{\sigma})$									(1.82, 0.08)	(2.03, 0.09)	(2.89, 0.12)	(3.20, 0.12)

$$\lambda_{\min}\left(\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\hat{\hat{\Sigma}}_{Y,P_k}\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\right)$$

n_Y \ n_X		100				400				800			
	m	100		400		100		400		100		400	
	k	4	10	4	10	4	10	4	10	4	10	4	10
100	(μ,σ)	(0.208, 0.022)	(0.160, 0.016)	(0.079, 0.008)	(0.062, 0.006)	(0.273, 0.028)	(0.211, 0.021)	(0.113, 0.012)	(0.085, 0.009)	(0.283, 0.029)	(0.217, 0.021)	(0.119, 0.013)	(0.089, 0.009)
	$(\hat{\mu},\hat{\sigma})$	(0.201, 0.026)	(0.136, 0.023)	(0.078, 0.009)	(0.056, 0.007)	(0.274, 0.028)	(0.195, 0.026)	(0.115, 0.010)	(0.081, 0.010)	(0.284, 0.028)	(0.201, 0.026)	(0.124, 0.011)	(0.086, 0.011)
	$(\hat{\mu},\hat{\sigma})$	(0.215, 0.018)	(0.178, 0.012)	(0.082, 0.006)	(0.069, 0.004)	(0.284, 0.023)	(0.237, 0.016)	(0.118, 0.010)	(0.098, 0.006)	(0.293, 0.024)	(0.244, 0.016)	(0.126, 0.011)	(0.105, 0.007)
400	(μ,σ)					(0.430, 0.024)	(0.380, 0.018)	(0.235, 0.012)	(0.209, 0.009)	(0.478, 0.025)	(0.426, 0.019)	(0.272, 0.013)	(0.242, 0.010)
	$(\hat{\mu},\hat{\sigma})$					(0.427, 0.026)	(0.363, 0.023)	(0.234, 0.013)	(0.203, 0.011)	(0.480, 0.025)	(0.413, 0.023)	(0.276, 0.013)	(0.240, 0.012)
	$(\hat{\mu},\hat{\sigma})$					(0.436, 0.022)	(0.389, 0.016)	(0.239, 0.011)	(0.215, 0.008)	(0.485, 0.022)	(0.437, 0.017)	(0.279, 0.012)	(0.252, 0.009)
800	(μ,σ)									(0.555, 0.021)	(0.511, 0.017)	(0.350, 0.013)	(0.322, 0.010)
	$(\hat{\mu},\hat{\sigma})$									(0.550, 0.023)	(0.493, 0.020)	(0.346, 0.013)	(0.313, 0.012)
	$(\hat{\mu},\hat{\sigma})$									(0.555, 0.020)	(0.511, 0.015)	(0.350, 0.012)	(0.323, 0.009)

Table 8.3 – Estimation of the extreme residual spike first moments when $n_X > n_Y$ by applying the Main Theorem 3.1.1 to $\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\hat{\hat{\Sigma}}_{Y,P_k}\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}$ or $\hat{\hat{\Sigma}}_{Y,P_k}^{-1/2}\hat{\hat{\Sigma}}_{X,P_k}\hat{\hat{\Sigma}}_{Y,P_k}^{-1/2}$. The orange eigenvalues estimations are inverted to be compare to the blue one as in equations 8.1 and 8.2. (Number of replicates: 1000, $\theta_i = 5000$)

In these cases, we present the number of false rejections in the table 8.4 using our procedure with a level of 0.05. Note in particular that the second table contains only zeroes. This is, however, to be expected from the theory, because when the trace of Σ^2 is larger than m, we tend to overestimate the location with probability 1.
Moreover, the red results show that our model does not work very well with the binomial distribution with success probability 0.9. We believe that the reason of this failure is the large fourth moment. The same underestimation could appear with a strongly asymmetric gamma distribution.
Despite the weakness of the estimation of some quantiles, our procedure estimates the location of the spikes very well. The results show that wrongly assuming that $\Sigma \neq \mathrm{I}_m$ leads to asymptotic conservative tests for any distribution.
Indeed, assuming $\Sigma = \mathrm{I}_m$ when $\Sigma \neq \mathrm{I}_m$ leads to an overestimation of order $O_p(1)$. On the other hand, the location underestimation or overestimation due to non-spherical distribution is small compared to the previous overestimation due to assuming $\Sigma = \mathrm{I}_m$.

Details of the simulations of Table 8.4

We apply perturbations $P_k = \mathrm{I}_m + \sum_{i=1}^{k}(\theta_i - 1)u_i u_i^t$ of order $k = 1$, 4 and 10 with $\theta_i = 5000$ to matrices containing i.i.d. data. Then we use the procedure with level 0.05 assuming that the spectra of $\mathbf{X}$ and $\mathbf{Y}$ is known. (Estimation are proposed in the next section.)
Finally by repeating the operation $n = 1000$ times, we estimate the percentage of false rejections of our generated data respecting $\mathrm{H}_0 : \Sigma_X = \Sigma_Y$.

8.2.1 Small perturbations

Our procedure implicitly assumes that the largest and the smallest residual spikes occur for large θ. This assumption can be proven for particular spectral distributions such as Marcenko-Pastur. Nevertheless, the universality of this assumption is wrong.
The second part of this section assumes perturbations of order 1 and in Figure 8.1 exhibits the curve of the residual spike as a function of θ. As expected, all the curves are monotonic and diverge from 1. Nevertheless, we also create an artificial example where the curve decreases!
To understand why this counterexample is different from the other simulations we need to explain how we generate the curve.

- The increasing curves are built from data $\mathbf{X} = P^{1/2}\Sigma^{1/2}\mathbf{X}\Psi \in \mathbb{R}^{m\times n_X}$ and $\mathbf{Y} = P^{1/2}\Sigma^{1/2}\mathbf{Y}\Psi \in \mathbb{R}^{m\times n_Y}$ defined as previously where $P = \mathrm{I}_m + (\theta - 1)uu^t$. Thus the covariance matrices become $\hat{\Sigma}_X = \frac{1}{n_X}\mathbf{X}\mathbf{X}^t$ and $\hat{\Sigma}_Y = \frac{1}{n_Y}\mathbf{Y}\mathbf{Y}^t$ both in $\mathbb{R}^{m\times m}$ and we can construct the residual spike as a function of θ.

- The decreasing curve is built directly from $W_X \in \mathbb{R}^{m\times m}$ and $W_Y \in \mathbb{R}^{m\times m}$ such that $\hat{\Sigma}_X = P^{1/2}\Sigma^{1/2}W_X\Sigma^{1/2}P^{1/2}$ and $\hat{\Sigma}_Y = P^{1/2}\Sigma^{1/2}W_Y\Sigma^{1/2}P^{1/2}$. In addition, $W_X = O_X\Lambda_X O_X^t$ and $W_Y = O_Y\Lambda_Y O_Y^t$, where $O_X, O_Y \in \mathbb{R}^{m\times m}$ are uniformly distributed and $\Lambda_X, \Lambda_Y \in \mathbb{R}^{m\times m}$ are diagonal matrices with entries generated by independent exponential random variables of parameter 1.

Therefore, we argue that this second scenario is less plausible in practice. However, even if it occurs, the criteria 4.2.5 allows us to detect it.

Remark 8.2.1.
The last curve is noisy before becoming monotone. This behaviour is due to largest eigenvalues being created by the spectra instead of the perturbations. Indeed in some cases, $\lambda_{\max}(\mathbf{X})$ is such that the perturbation of order 1 is not detectable as the largest eigenvalue of $\mathbf{X} = P^{1/2}\mathbf{X}$.

1. $\Sigma = \mathrm{I}_m$, $\Psi_X = \mathrm{I}_{n_X}$, $\Psi_Y = \mathrm{I}_{n_Y}$

$\lambda_{\min}\left(\hat{\Sigma}_{X,P_k}^{-1/2}\hat{\Sigma}_{Y,P_k}\hat{\Sigma}_{X,P_k}^{-1/2}\right)$

n_Y	k	Gamma(6,2)-3						Binomial(0.5)-0.5						Binomial(0.9)-0.9					
	n_X	100		200		400		100		200		400		100		200		400	
	m	200	400	200	400	200	400	200	400	200	400	200	400	200	400	200	400	200	400
100	1	0.06	0.06	0.12	0.15	0.18	0.22	0.00	0.00	0.00	0.00	0.01	0.01	0.14	0.14	0.20	0.25	0.24	0.30
	4	0.02	0.03	0.05	0.07	0.08	0.14	0.00	0.00	0.00	0.01	0.01	0.02	0.15	0.17	0.25	0.27	0.29	0.34
	10	0.00	0.00	0.00	0.01	0.01	0.01	0.00	0.00	0.00		0.00	0.00	0.04	0.07	0.09	0.14	0.11	0.16
200	1			0.06	0.06	0.09	0.12			0.00	0.00	0.01	0.00			0.14	0.13	0.17	0.21
	4			0.01	0.03	0.06	0.09			0.00	0.00	0.01	0.01			0.11	0.15	0.18	0.27
	10			0.00	0.01	0.01	0.01			0.00	0.00	0.00	0.00			0.03	0.07	0.06	0.13
400	1					0.05	0.05					0.00	0.00					0.12	0.13
	4					0.02	0.04					0.00	0.00					0.09	0.14
	10					0.00	0.01					0.00	0.00					0.02	0.07

$\lambda_{\max}\left(\hat{\Sigma}_{X,P_k}^{-1/2}\hat{\Sigma}_{Y,P_k}\hat{\Sigma}_{X,P_k}^{-1/2}\right)$

n_Y	k	Gamma(6,2)-3						Binomial(0.5)-0.5						Binomial(0.9)-0.9					
	n_X	100		200		400		100		200		400		100		200		400	
	m	200	400	200	400	200	400	200	400	200	400	200	400	200	400	200	400	200	400
100	1	0.09	0.10	0.07	0.09	0.06	0.06	0.00	0.00	0.02	0.01	0.01	0.02	0.16	0.21	0.15	0.16	0.11	0.12
	4	0.19	0.20	0.10	0.14	0.08	0.07	0.04	0.05	0.02	0.02	0.02	0.02	0.36	0.40	0.23	0.29	0.14	0.15
	10	0.49	0.61	0.20	0.30	0.10	0.14	0.32	0.37	0.10	0.14	0.05	0.05	0.71	0.79	0.39	0.52	0.14	0.25
200	1			0.08	0.08	0.06	0.08			0.01	0.00	0.02	0.01			0.14	0.16	0.13	0.14
	4			0.13	0.17	0.07	0.10			0.02	0.03	0.03	0.02			0.29	0.34	0.17	0.23
	10			0.24	0.38	0.12	0.16			0.13	0.16	0.05	0.07			0.47	0.60	0.25	0.32
400	1					0.08	0.09					0.01	0.01					0.13	0.15
	4					0.09	0.13					0.03	0.02					0.23	0.27
	10					0.17	0.21					0.06	0.07					0.31	0.40

2. $\mathbf{N}(0,1)$, $W_i \sim \mathrm{Wishart}(\sigma_i^2 \mathrm{I}_{m/3}, 5m)$ are fixed

$\lambda_{\min}\left(\hat{\Sigma}_{X,P_k}^{-1/2}\hat{\Sigma}_{Y,P_k}\hat{\Sigma}_{X,P_k}^{-1/2}\right)$

n_Y	k	$\Sigma = \mathrm{Wishart}(\mathrm{I}_m, 5m)$, $\Psi = \mathrm{I}$						$\Sigma = \mathrm{Diag}(W_1, W_2, W_3)$, $\Psi = \mathrm{I}$						$\Sigma = \mathrm{Diag}(W_1, W_2, W_3)$, $\Psi = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots \\ \rho & 1 & \rho & \cdots \\ \rho^2 & \rho & 1 & \cdots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix}$, $\rho = 0.6$					
	n_X	100		200		400		100		200		400		100		200		400	
	m	200	400	200	400	200	400	200	400	200	400	200	400	200	400	200	400	200	400
100	1	0.00	0.01	0.04	0.05	0.04	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
	4	0.00	0.00	0.00	0.01	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	10	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
200	1			0.00	0.02	0.00	0.01			0.00	0.00	0.00	0.00			0.00	0.00	0.00	0.00
	4			0.00	0.00	0.00	0.00			0.00	0.00	0.00	0.00			0.00	0.00	0.00	0.00
	10			0.00	0.00	0.00	0.00			0.00	0.00	0.00	0.00			0.00	0.00	0.00	0.00
400	1					0.00	0.00					0.00	0.00					0.00	0.00
	4					0.00	0.00					0.00	0.00					0.00	0.00
	10					0.00	0.00					0.00	0.00					0.00	0.00

$\lambda_{\max}\left(\hat{\Sigma}_{X,P_k}^{-1/2}\hat{\Sigma}_{Y,P_k}\hat{\Sigma}_{X,P_k}^{-1/2}\right)$

n_Y	k	$\Sigma = \mathrm{Wishart}(\mathrm{I}_m, 5m)$, $\Psi = \mathrm{I}$						$\Sigma = \mathrm{Diag}(W_1, W_2, W_3)$, $\Psi = \mathrm{I}$						$\Sigma = \mathrm{Diag}(W_1, W_2, W_3)$, $\Psi = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots \\ \rho & 1 & \rho & \cdots \\ \rho^2 & \rho & 1 & \cdots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix}$, $\rho = 0.6$					
	n_X	100		200		400		100		200		400		100		200		400	
	m	200	400	200	400	200	400	200	400	200	400	200	400	200	400	200	400	200	400
100	1	0.03	0.03	0.00	0.01	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	4	0.05	0.02	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
	10	0.25	0.35	0.00	0.09	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.05	0.00	0.00	0.00	0.00
200	1			0.00	0.01	0.00	0.01			0.00	0.00	0.00	0.00			0.00	0.00	0.00	0.00
	4			0.00	0.00	0.00	0.00			0.00	0.00	0.00	0.00			0.00	0.00	0.00	0.00
	10			0.01	0.04	0.00	0.00			0.00	0.00	0.00	0.00			0.00	0.00	0.00	0.00
400	1					0.00	0.00					0.00	0.00					0.00	0.00
	4					0.00	0.00					0.00	0.00					0.00	0.00
	10					0.00	0.00					0.00	0.00					0.00	0.00

Table 8.4 – Simulations of the percentage of false rejections when the α-level of the procedure is 0.05. In the second table, the values σ_i^2 are $0.5, 1, 4$ for $i = 1, 2, 3$. The red values are not accurate enough. However they show that the location of the residual spike is well estimated. (Number of replicates: 1000, $\theta_i = 5000$)

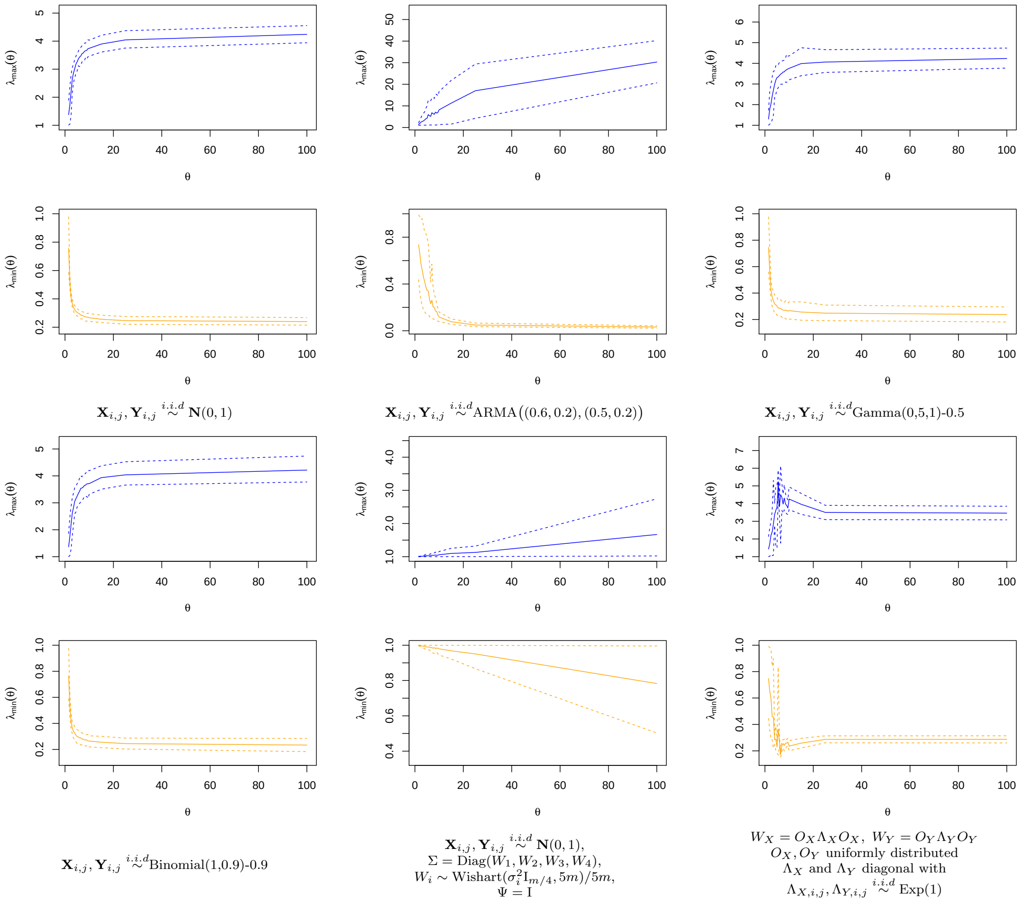


Figure 8.1 – Estimation of residual spikes (with quantiles 0.05 and 0.95) as a function of θ, the eigenvalue of P, a perturbation of order 1. The simulations always assume $m = 400$, $n_X = 800$ and $n_Y = 200$.

8.3 Estimation of the spectrum

Usually we do not know the spectra of the random matrices $\mathbf{X} \in \mathbf{R}^{m \times n_X}$ and $\mathbf{Y} \in \mathbf{R}^{m \times n_Y}$, but we estimate them in the Main Theorem 3.1.1 by using the spectra of $\mathbf{X} = P^{1/2}\mathbf{X}$ and $\mathbf{Y} = P^{1/2}\mathbf{Y}$ as in Theorem 4.4.1.

The Table 8.5 presents the false rejection rates of our procedure assuming a level of 0.05. Two ways of estimating the spectra are tested, the usual and the robust method defined in Theorem 4.4.1.

As expected from the theorem, the usual estimator tends to underestimate the extreme residual spikes and this can be corrected by the more conservative robust estimator.

When n_X or m is small, an interesting idea could be to combine this robust estimation of the spectra and the orange estimator of the residual spike introduced in Table 8.3.

Despite the poorer results of the procedure when k is large, the table confirms a good asymptotic performance for reasonably large k.

Details of the simulations of Table 8.5

We apply perturbations of order $k = 1, 4, 10$ with $\theta_i = 5000$ to normal data.

Then, we use the procedure with level 0.05. In particular, we use two different methods to estimate the moments of the spectra, the usual and the robust one defined in 4.4.1.

Finally by repeating the operation $n = 1000$ times, we estimate the percentage of false rejections of our generated data respecting $\mathrm{H}_0 : \Sigma_X = \Sigma_Y$.

$$\lambda_{\max}\left(\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\hat{\hat{\Sigma}}_{Y,P_k}\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\right)$$

n_Y \ n_X		100						500						1000					
	k \ m	50		100		500		50		100		500		50		100		500	
		Usual	Robust	Usual	Robust	Usual	Robust	Usual	Robust	Usual	Robust	Usual	Robust	Usual	Robust	Usual	Robust	Usual	Robust
100	1	0.07	0.02	0.08	0.04	0.07	0.05	0.07	0.01	0.06	0.01	0.06	0.03	0.06	0.01	0.07	0.02	0.06	0.02
	4	0.16	0.00	0.15	0.00	0.17	0.05	0.14	0.00	0.11	0.00	0.11	0.00	0.11	0.00	0.11	0.00	0.09	0.00
	10	0.50	0.00	0.52	0.00	0.56	0.06	0.34	0.00	0.32	0.00	0.29	0.00	0.31	0.00	0.29	0.00	0.26	0.00
500	1							0.08	0.01	0.04	0.01	0.06	0.04	0.06	0.01	0.06	0.00	0.06	0.03
	4							0.09	0.00	0.08	0.00	0.10	0.01	0.07	0.00	0.07	0.00	0.07	0.00
	10							0.19	0.00	0.22	0.00	0.20	0.00	0.19	0.00	0.16	0.00	0.18	0.00
1000	1													0.05	0.00	0.05	0.01	0.07	0.04
	4													0.08	0.00	0.08	0.00	0.07	0.00
	10													0.14	0.00	0.16	0.00	0.15	0.00

$$\lambda_{\min}\left(\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\hat{\hat{\Sigma}}_{Y,P_k}\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\right)$$

n_Y \ n_X		100						500						1000					
	k \ m	50		100		500		50		100		500		50		100		500	
		Usual	Robust	Usual	Robust	Usual	Robust	Usual	Robust	Usual	Robust	Usual	Robust	Usual	Robust	Usual	Robust	Usual	Robust
100	1	0.04	0.00	0.04	0.02	0.04	0.02	0.09	0.02	0.10	0.04	0.22	0.16	0.10	0.02	0.13	0.06	0.22	0.17
	4	0.00	0.00	0.00	0.00	0.02	0.00	0.01	0.00	0.04	0.00	0.07	0.00	0.04	0.00	0.03	0.00	0.07	0.00
	10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
500	1							0.06	0.00	0.04	0.00	0.04	0.02	0.06	0.00	0.07	0.02	0.07	0.04
	4							0.03	0.00	0.03	0.00	0.02	0.00	0.04	0.00	0.03	0.00	0.06	0.00
	10							0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00
1000	1													0.05	0.00	0.06	0.00	0.05	0.02
	4													0.03	0.00	0.03	0.00	0.03	0.00
	10													0.01	0.00	0.01	0.00	0.02	0.00

Table 8.5 – Simulations of the procentage of false rehection at level $\alpha = 0.05$ when the spectra are estimated by its usual estimator or its robust estimator proposed in Theorem 4.4.1. The red values are not accurate enough, but only occur for small m, n_X n_Y or for large k. However, the location of the residual spike is well estimated. (Number of replicates: 1000, $\Theta_i = 5000$)

8.4 Estimation of k

In this section we show in Table 8.6 that a small underestimation or overestimation of k does not affect the estimation of residual spikes.
Recall that in the simulations of Section 3.1, we assumed the spectra of $\mathbf{X} = P^{-1/2}\mathbf{X}$ and $\mathbf{Y} = P^{-1/2}\mathbf{Y}$ are known. In this scenario, neglecting overly large perturbations led to loose conservatism. The simulations of this section estimate the spectra using the observed spectra of $\mathbf{X} = P^{1/2}\mathbf{X}$ and $\mathbf{Y} = P^{1/2}\mathbf{Y}$. In this case it seems that,

Errors on k lead to conservative procedures in all cases. However, underestimation of k can lead to a large loss of power!

Details of the simulations of Table 8.6

We apply a perturbation of order $k = 4$ to normal data.
Then we use the procedure with different $k_{est} = 1, 2, 3, 4, 5, 6$.
The moments are estimated using the usual estimators of the spectra assuming $k = k_{est}$ and $n = 1000$ replicates of the experiment.
The correct perturbation, P_4, that is applied to the data has eigenvalues, $\theta_1 = 1000$, $\theta_2 = 200$, $\theta_3 = 16$, $\theta_4 = 2.1$.

8.5 Criterion

In Section 4.2, we investigate the criterion T defined in 4.2.8. In this section we use simulation to argue that assuming class $\mathcal{C}_D$ as defined in 4.2.7, T is never negative. In particular it becomes strictly positive when $\Sigma \neq \mathrm{I}_m$. The simulations of Table 8.7 compute the criterion for different scenarios.
The simulations always give positive values for the criterion. Therefore, our procedure leads to conservative tests in class $\mathcal{C}_D$.

8.6 An application

In this section, we apply our procedure to data $\mathbf{X}$ and $\mathbf{Y}$. First, each step is briefly explained. Then, an analysis is presented on simulated data together with the mathematical work and the important plots.
This procedure is not unique and other solutions better adapted to the problem could be implemented. For example, the choice of k and the number of perturbations, could certainly be improved. The goal of this section is to provide a procedure as conservative as possible with reasonably good asymptotic power.

1. First, we center the data with regard to the rows and columns.
2. Then, we need to estimate k and rescale the variance. However, in order to rescale the matrix we need to know k! There are two ways to select k:
   (a) The user chooses k by looking at the spectra and keeping in mind that overestimation is preferable to underestimation of the actual value.
   (b) i. By looking at the spectra, we can select different interesting values for k_X and k_Y to test.
      ii. Assuming some k_X and k_Y, we rescale the matrices by assuming $\hat{\theta}_{X,1}, ..., \hat{\theta}_{X,k_X}$ perturbations in the matrix $\mathbf{X}$ and $\hat{\theta}_{Y,1}, ..., \hat{\theta}_{Y,k_Y}$ perturbations in the matrix $\mathbf{Y}$.
      iii. If the perturbations $\hat{\theta}$ added to one matrix respect the condition proposed in Section 4.4.4 depending on k_X and k_Y, then the couple (k_X, k_Y) is relevant.

$\lambda_{\max}\left(\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\hat{\hat{\Sigma}}_{Y,P_k}\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\right)$

n_Y \ n_X		200				400				800			
	k_{est} \ m	200		400		200		400		200		400	
		$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)
200	1	(88.78, 55.48)	(3.86, 1.92)	(85.85, 57.82)	(6.23, 2.84)	(87.34, 55.85)	(3.44, 1.54)	(106.76, 67.23)	(5.04, 1.98)	(94.28, 58.81)	(3.17, 1.40)	(91.32, 58.26)	(4.51, 1.83)
	2	(7.86, 2.12)	(4.11, 0.42)	(8.15, 1.23)	(6.40, 0.57)	(7.30, 2.14)	(3.43, 0.30)	(6.60, 0.93)	(5.16, 0.37)	(6.63, 1.88)	(3.13, 0.26)	(6.48, 1.07)	(4.52, 0.27)
	3	(4.27, 0.30)	(4.23, 0.31)	(6.62, 0.43)	(6.46, 0.48)	(3.55, 0.19)	(3.53, 0.22)	(5.29, 0.26)	(5.19, 0.29)	(3.20, 0.15)	(3.18, 0.16)	(4.65, 0.19)	(4.58, 0.21)
	4	(4.42, 0.28)	(4.29, 0.33)	(6.90, 0.40)	(6.59, 0.47)	(3.63, 0.18)	(3.57, 0.22)	(5.44, 0.24)	(5.27, 0.31)	(3.25, 0.14)	(3.21, 0.16)	(4.77, 0.19)	(4.64, 0.22)
	5	(4.53, 0.28)	(4.33, 0.32)	(7.05, 0.39)	(6.55, 0.49)	(3.71, 0.17)	(3.59, 0.21)	(5.55, 0.23)	(5.30, 0.30)	(3.29, 0.14)	(3.21, 0.15)	(4.85, 0.18)	(4.69, 0.22)
	6	(4.63, 0.27)	(4.34, 0.33)	(7.25, 0.39)	(6.70, 0.49)	(3.77, 0.16)	(3.64, 0.22)	(5.66, 0.24)	(5.36, 0.29)	(3.38, 0.14)	(3.32, 0.17)	(4.90, 0.18)	(4.71, 0.22)
400	1					(89.97, 58.12)	(2.74, 1.04)	(93.15, 58.71)	(3.91, 1.40)	(102.49, 64.57)	(2.42, 0.84)	(85.13, 55.71)	(3.39, 1.11)
	2					(6.71, 2.18)	(2.82, 0.22)	(5.82, 1.05)	(4.00, 0.27)	(6.14, 1.96)	(2.46, 0.16)	(5.19, 0.98)	(3.35, 0.18)
	3					(2.89, 0.15)	(2.86, 0.16)	(4.10, 0.20)	(4.06, 0.22)	(2.54, 0.11)	(2.50, 0.11)	(3.44, 0.13)	(3.41, 0.15)
	4					(2.97, 0.14)	(2.90, 0.17)	(4.21, 0.19)	(4.06, 0.21)	(2.55, 0.10)	(2.51, 0.11)	(3.50, 0.12)	(3.41, 0.14)
	5					(3.02, 0.14)	(2.92, 0.16)	(4.29, 0.18)	(4.09, 0.23)	(2.61, 0.10)	(2.54, 0.11)	(3.56, 0.12)	(3.43, 0.14)
	6					(3.06, 0.13)	(2.92, 0.16)	(4.37, 0.18)	(4.12, 0.22)	(2.62, 0.09)	(2.57, 0.11)	(3.59, 0.12)	(3.45, 0.14)
800	1									(91.91, 56.71)	(2.06, 0.53)	(90.85, 59.77)	(2.69, 0.74)
	2									(5.52, 1.85)	(2.10, 0.13)	(5.00, 1.22)	(2.73, 0.14)
	3									(2.17, 0.08)	(2.14, 0.08)	(2.81, 0.10)	(2.79, 0.11)
	4									(2.19, 0.08)	(2.17, 0.08)	(2.85, 0.10)	(2.79, 0.11)
	5									(2.22, 0.08)	(2.17, 0.08)	(2.90, 0.10)	(2.80, 0.11)
	6									(2.24, 0.07)	(2.18, 0.09)	(2.92, 0.09)	(2.80, 0.11)

$\lambda_{\min}\left(\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\hat{\hat{\Sigma}}_{Y,P_k}\hat{\hat{\Sigma}}_{X,P_k}^{-1/2}\right)$

n_Y \ n_X		200				400				800			
	k_{est} \ m	200		400		200		400		200		400	
		$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)
200	1	(0.012, 0.008)	(0.297, 0.093)	(0.011, 0.008)	(0.181, 0.053)	(0.012, 0.008)	(0.333, 0.102)	(0.010, 0.007)	(0.219, 0.060)	(0.011, 0.007)	(0.360, 0.105)	(0.011, 0.008)	(0.246, 0.065)
	2	(0.113, 0.057)	(0.245, 0.024)	(0.119, 0.024)	(0.159, 0.014)	(0.120, 0.066)	(0.290, 0.027)	(0.146, 0.028)	(0.193, 0.016)	(0.133, 0.070)	(0.318, 0.030)	(0.149, 0.036)	(0.217, 0.018)
	3	(0.231, 0.020)	(0.239, 0.019)	(0.149, 0.011)	(0.156, 0.011)	(0.274, 0.019)	(0.281, 0.021)	(0.183, 0.012)	(0.189, 0.014)	(0.302, 0.019)	(0.306, 0.022)	(0.206, 0.012)	(0.210, 0.015)
	4	(0.220, 0.018)	(0.235, 0.017)	(0.141, 0.011)	(0.153, 0.011)	(0.262, 0.019)	(0.279, 0.021)	(0.174, 0.011)	(0.188, 0.013)	(0.291, 0.019)	(0.305, 0.021)	(0.197, 0.013)	(0.210, 0.016)
	5	(0.211, 0.018)	(0.233, 0.017)	(0.137, 0.010)	(0.153, 0.011)	(0.254, 0.018)	(0.276, 0.021)	(0.168, 0.011)	(0.187, 0.014)	(0.281, 0.019)	(0.306, 0.022)	(0.190, 0.012)	(0.210, 0.016)
	6	(0.205, 0.017)	(0.230, 0.018)	(0.131, 0.010)	(0.151, 0.011)	(0.244, 0.018)	(0.275, 0.020)	(0.162, 0.011)	(0.186, 0.014)	(0.271, 0.018)	(0.305, 0.023)	(0.183, 0.013)	(0.210, 0.015)
400	1					(0.011, 0.008)	(0.402, 0.105)	(0.011, 0.008)	(0.277, 0.066)	(0.010, 0.007)	(0.448, 0.108)	(0.013, 0.009)	(0.317, 0.073)
	2					(0.130, 0.083)	(0.357, 0.028)	(0.162, 0.043)	(0.251, 0.016)	(0.131, 0.085)	(0.407, 0.028)	(0.183, 0.053)	(0.296, 0.018)
	3					(0.342, 0.021)	(0.349, 0.020)	(0.241, 0.013)	(0.247, 0.013)	(0.389, 0.021)	(0.397, 0.021)	(0.286, 0.014)	(0.291, 0.015)
	4					(0.332, 0.020)	(0.347, 0.018)	(0.235, 0.013)	(0.246, 0.013)	(0.382, 0.019)	(0.395, 0.020)	(0.278, 0.013)	(0.290, 0.015)
	5					(0.323, 0.019)	(0.344, 0.019)	(0.228, 0.012)	(0.245, 0.013)	(0.372, 0.019)	(0.393, 0.019)	(0.271, 0.013)	(0.289, 0.015)
	6					(0.316, 0.019)	(0.343, 0.018)	(0.223, 0.012)	(0.244, 0.013)	(0.365, 0.018)	(0.392, 0.021)	(0.265, 0.012)	(0.289, 0.014)
800	1									(0.011, 0.008)	(0.509, 0.103)	(0.012, 0.008)	(0.391, 0.074)
	2									(0.148, 0.103)	(0.477, 0.027)	(0.184, 0.075)	(0.367, 0.017)
	3									(0.459, 0.020)	(0.467, 0.018)	(0.355, 0.014)	(0.361, 0.014)
	4									(0.453, 0.018)	(0.463, 0.018)	(0.349, 0.013)	(0.359, 0.013)
	5									(0.447, 0.018)	(0.465, 0.018)	(0.340, 0.013)	(0.356, 0.013)
	6									(0.439, 0.017)	(0.460, 0.018)	(0.337, 0.013)	(0.358, 0.014)

Table 8.6 – Residual spike moment for normal entries, $k = 4$ and $\theta_1 = 1000$, $\theta_2 = 200$, $\theta_3 = 16$, $\theta_4 = 2.1$. Then, the value k is estimated by k_{est}. (Number of replicates: 1000)

		$\Sigma = \mathrm{I}_m$, $\Psi = \mathrm{I}$						$\Sigma = \Lambda$, $\Lambda \sim \mathrm{MP}(c = 0.5)$, $\Psi = \mathrm{I}$						$\Sigma = \mathrm{Diag}(5, 5, ..., 1, 1, ..., 0.5, 0.5, ...)$, $\Psi = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots \\ \rho & 1 & \rho & \cdots \\ \rho^2 & \rho & 1 & \cdots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix}$, $\rho = 0.6$					
n_Y \ n_X		100		200		400		100		200		400		100		200		400	
	k \ m	200	400	200	400	200	400	200	400	200	400	200	400	200	400	200	400	200	400
100	1	0.009	0.001	-0.005	-0.002	0.015	0.004	0.185	0.108	0.272	0.136	0.283	0.176	0.158	0.086	0.238	0.128	0.283	0.165
	4	0.002	0.004	0.015	0.009	0.004	-0.000	0.189	0.108	0.266	0.127	0.270	0.181	0.168	0.089	0.239	0.125	0.267	0.159
	10	0.022	0.008	0.036	0.004	0.045	0.016	0.176	0.104	0.262	0.135	0.268	0.182	0.181	0.092	0.241	0.118	0.283	0.160
200	1			0.002	0.004	0.014	0.002			0.315	0.196	0.436	0.261			0.304	0.175	0.358	0.217
	4			0.012	0.008	0.019	0.005			0.320	0.197	0.425	0.265			0.279	0.165	0.377	0.214
	10			0.051	0.006	0.067	0.020			0.309	0.197	0.444	0.253			0.295	0.176	0.372	0.217
400	1					0.011	0.002					0.485	0.326					0.434	0.279
	4					0.052	0.018					0.484	0.322					0.462	0.295
	10					0.075	0.025					0.479	0.323					0.430	0.292

Table 8.7 – Results of the simulations of the criterion 4.2.8 for data with normal entries. The order size of T under H_0 is $O(1/m^{1/2}) \approx 0.07$ and 0.05. The values in the table are computed once from generated data and may fluctuate.

iv. Finally, the user wisely chooses $k = \max_{\mathcal{P}}(k_X, k_Y)$, where $\mathcal{P}$ is the group of the relevant pairs. In the case of doubt, the user can also set $k_{X,rob} \leqslant k$ and $k_{Y,rob} \leqslant k$ two probable underestimations of the true values.

Using appropriate $k_{X,rob}$ and $k_{Y,rob}$, we can rescale the matrices $\mathbf{X}$ and $\mathbf{Y}$ to create $\mathbf{X}$ and $\mathbf{Y}$.

3. Next, we apply the procedure to $\mathbf{X}$ and $\mathbf{Y}$ by assuming the relevant k. This leads to two observed extreme residual spikes.

4. We compute the distribution of the residual spike. As in the previous part, we assume k perturbations. Nevertheless, to ensure conservative properties, the estimation of M_2 uses $(k_{X,rob}, k_{Y,rob})$. In this step we can use either the usual estimator of M_2 or the bounded estimator defined in Theorem 4.4.1.

5. Finally, we can compare the extreme values with their distribution under H_0 for testing purposes or we can compute the expectation and the variance of the theoretical residual spike and use Chebyshev's inequality to discuss the likelihood of H_0.

Remark 8.6.1.

1. By simulations of Section 8.4, the choice of k does not affect the conservative properties. However strong underestimation greatly reduces the power. This explains the advice to overestimate k.

2. We could also rescale the matrices by groups as in Section 4.4.2. Moreover we can rescaled each columns by its variance. As consequence of this strong transformation, we will loose some power.

8.6.1 Analysis

We observe data $\mathbf{X} \in \mathbf{R}^{m \times n_X}$ and $\mathbf{Y} \in \mathbf{R}^{m \times n_Y}$ that we suppose already centred in rows and columns using Section 4.4.2. Knowing the size of the perturbation, k, we can estimate the variances,

$$\hat{\sigma}_X^2 = \frac{1}{m-k} \sum_{i=k+1}^{m} \lambda_i \left(\frac{1}{n_X} \mathbf{X}\mathbf{X}^t \right),$$

$$\hat{\sigma}_Y^2 = \frac{1}{m-k} \sum_{i=k+1}^{m} \lambda_i \left(\frac{1}{n_Y} \mathbf{Y}\mathbf{Y}^t \right).$$

In this example we have:

$$m = 1000,\ n_X = 2000,\ n_Y = 500.$$

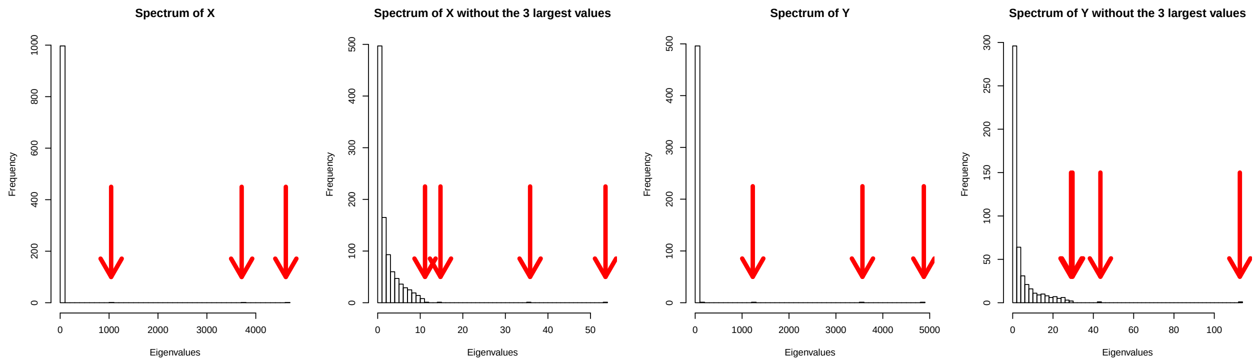


Figure 8.2 – Spectra of **X** and **Y** with largest isolated eigenvalues indicated by arrows.

Method 1

The first method overestimates k based on Figure 8.2. The spectrum of **X** seems to have 6 isolated eigenvalues, but we could argue that two other eigenvalues are perturbations. The spectrum **Y** clearly shows 5 isolated eigenvalues and at most 2 additional ones.
We thus set $k = 8$ knowing that we probably overestimate the true value.
Knowing k, we can rescale the matrices **X** and **Y** by $\hat{\sigma}_X$ and $\hat{\sigma}_Y$ respectively to create the covariance matrices

$$\hat{\Sigma}_X = \frac{1}{n_X \hat{\sigma}_X^2} \mathbf{X}\mathbf{X}^t \text{ and } \hat{\Sigma}_Y = \frac{1}{n_Y \hat{\sigma}_Y^2} \mathbf{Y}\mathbf{Y}^t.$$

Next, we filter the matrices as in definition 2.2.1.

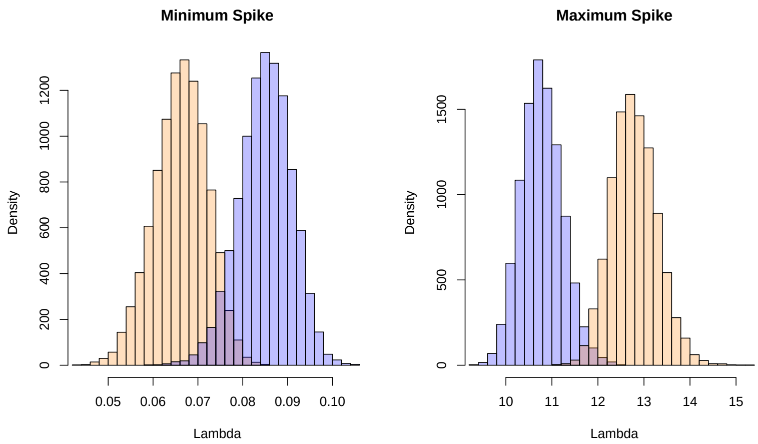
$$\hat{\hat{\Sigma}}_X = \mathrm{I}_m + \sum_{i=1}^{k} \left(\hat{\hat{\theta}}_{X,i} - 1 \right) \hat{u}_{\hat{\Sigma}_X,i} \hat{u}_{\hat{\Sigma}_X,i}^t,$$
$$\hat{\hat{\theta}}_{X,i} = 1 + \frac{1}{\frac{1}{m-k} \sum_{j=k+1}^{m} \frac{\hat{\lambda}_{\hat{\Sigma}_X,j}}{\hat{\lambda}_{\hat{\Sigma}_X,i} - \hat{\lambda}_{\hat{\Sigma}_X,j}}}.$$

The computed residual spikes of $\hat{\hat{\Sigma}}_X^{-1} \hat{\hat{\Sigma}}_Y$ are shown in Table 8.8.

$\lambda_{\max}$	56.03	10.25	9.88	8.96	8.29	7.27	5.71	5.10
$\lambda_{\min}$	0.04	0.10	0.13	0.13	0.18	0.21	0.34	0.36

Table 8.8 – Observed residual spikes.

Using Figure 8.3 a, these values are compared to the theoretical distributions of the extreme residual spikes assuming equality of the perturbations of order k. The distribution in blue uses the usual estimator of the spectra and the orange uses the conservative estimator introduced in Theorem 4.4.1. The moments of the spectra are summarize in Figure 8.3 b.



a

	$\lambda_{\min}$	$\lambda_{\max}$
Usual	(0.085, 0.006)	(10.75, 0.45)
Robust	(0.067, 0.006)	(12.67, 0.47)

b

Figure 8.3 – a: Distribution of the extreme residual spike assuming equality of the covariance, $k = 8$ and θ_i large. (Robust estimation of the spectra in orange.) b: Estimated residual spikes moments, (μ, σ) using usual or robust estimators of the moments spectra.

We finally clearly detect two residual spikes. Figure 8.4 presents the residual eigenvectors of the residual eigenvalues.

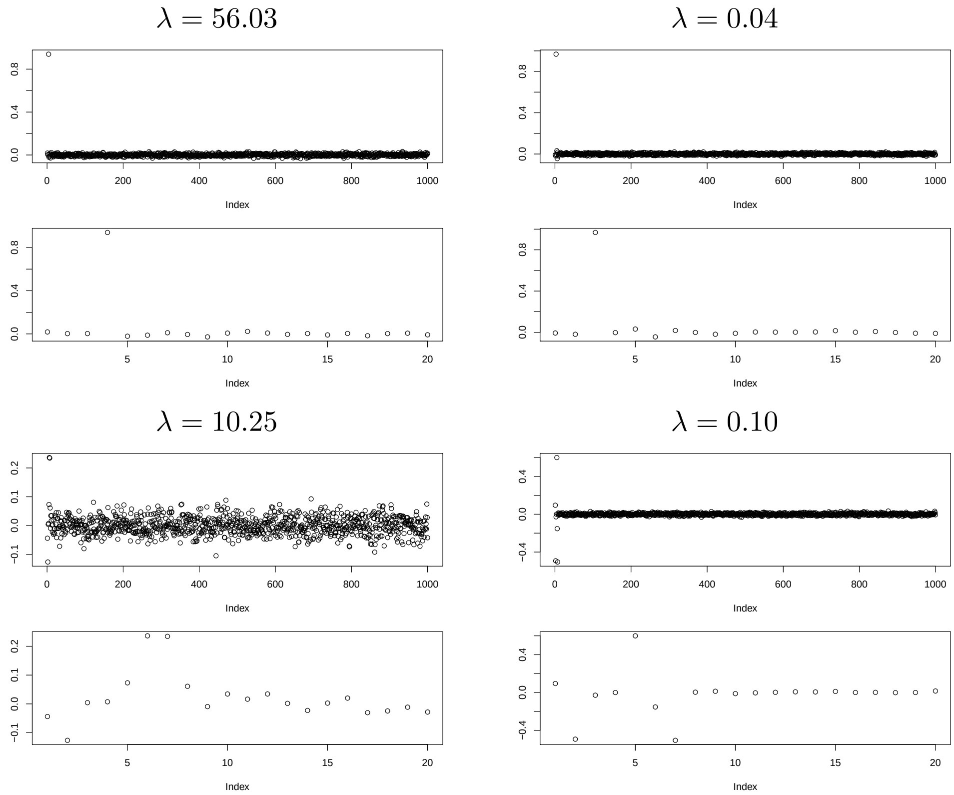


Figure 8.4 – Representation of the entiere residual eigenvectors and only the 20 first entries.

We conclude that the differences are in direction e_3 and e_4. As we see in the figure, two other eigenvectors also exhibit a structure. Without our test, we could have concluded that they also represent significant differences, but this residual structure is merely due to the biased estimation of the eigenvectors.

A structure in a residual eigenvector does not imply a real difference!

8.6.2 Method 2

An alternative way to choose k could be to define the set of the possible pairs k_X and k_Y by looking at the spectra of $\mathbf{X}$ and $\mathbf{Y}$ on the Figure 8.2. In our example, we set $\mathcal{P} = \{\{4,5,6,7\} \times \{4,5,6,7\}\}$. Next we estimate σ^2 with $\hat{\sigma}^2_{X,\mathcal{P}_i}$ and $\hat{\sigma}^2_{Y,\mathcal{P}_i}$ for a pair $\mathcal{P}_i \in \mathcal{P}$.
The second method tries to be less conservative by estimating k wisely. In case of overestimation of k, we are in the same case as previously. On the other hand, neglecting some values could lead to a mistake.
Therefore, an algorithm checks if the neglected eigenvalues could interfere with the maximum residual spike. This algorithm is presented in Section 4.4.4. Assuming that we choose the quantile 0.05, we obtain the following table:

k_Y \ k_X	4	5	6	7
4	1	1	0	0
5	1	1	0	0
6	0	0	0	0
7	0	0	0	0

When the table displays 1, the pairs (k_X, k_Y) contain only perturbations that can interfere with the maximum residual spike. When the value is 0, this means that with high probability the smallest estimated eigenvalues will not affect the residual spike.
In this example we would choose $k = 5$ and the rest of the analysis is similar to the first method. In this case it provides the same results, but in some scenarios it could increase the power.

8.7 Temporal algorithm

In this section, we apply the algorithm presented in section 4.4.3 to simulated data

1. without temporal perturbations,
2. with small temporal perturbations of order $\sqrt{m}$,
3. with large temporal perturbations of order m^2.

The data are of the form:

$$\mathbf{X} = P^{1/2}\mathbf{X}P_{X,T}^{1/2} \in \mathbf{R}^{m\times n_X} \text{ and } \mathbf{Y} = P^{1/2}\mathbf{Y}P_{Y,T}^{1/2} \in \mathbf{R}^{m\times n_Y},$$

where $P \in \mathbb{R}^{m\times m}$ is the same spatial perturbation of order k for both groups and $P_{X,T} \in \mathbb{R}^{n_X\times n_X}$, $P_{Y,T} \in \mathbb{R}^{n_Y\times n_Y}$ are temporal perturbations of order k_T. We apply the algorithm to matrices with normal entries and present the results in Table 8.9.
The main conclusion of the simulations is that the algorithm seems to improve the non-asymptotic conservative properties of the procedure.

$\lambda_{\max}\left(\hat{\hat{\Sigma}}^{-1/2}_{X,P_k}\hat{\hat{\Sigma}}_{Y,P_k}\hat{\hat{\Sigma}}^{-1/2}_{X,P_k}\right)$

	$P_{X,T}=I_{n_X}$, $P_{Y,T}=I_{n_Y}$		Small perturbations $P_{X,T}$ and $P_{Y,T}$		Large perturbations $P_{X,T}$ and $P_{Y,T}$	
k_T	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)
1	(4.32,0.27)	(4.23,0.26)	(4.34,0.29)	(4.31,0.31)	(4.35,0.27)	(4.18,0.26)
3			(4.30,0.24)	(4.38,0.27)	(4.37,0.25)	(4.19,0.24)

$\lambda_{\min}\left(\hat{\hat{\Sigma}}^{-1/2}_{X,P_k}\hat{\hat{\Sigma}}_{Y,P_k}\hat{\hat{\Sigma}}^{-1/2}_{X,P_k}\right)$

	$P_{X,T}=I_{n_X}$, $P_{Y,T}=I_{n_Y}$		Small perturbations $P_{X,T}$ and $P_{Y,T}$		Large perturbations $P_{X,T}$ and $P_{Y,T}$	
k_T	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)	$(\hat{\mu},\hat{\sigma})$	(μ,σ)
1	(0.230,0.015)	(0.229,0.017)	(0.230,0.015)	(0.230,0.018)	(0.227,0.013)	(0.234,0.019)
3			(0.227,0.017)	(0.226,0.018)	(0.230,0.017)	(0.233,0.018)

Table 8.9 – Moment of the extreme residual spikes when $m = 300$, $n_X = 400$, $n_Y = 200$ and $k = 2$. Small temporal perturbation are of order $\sqrt{m}$ and large pertrubation are of order m^2.

8.8 Confirmation by simulation of some important theorems

In this section, we demonstrate by simulation the central theorems of this thesis. We assume three scenarios,

1. $n_X = 400$, $n_Y = 400$, $m = 400$, $\rho = 0.5$,
2. $n_X = 400$, $n_Y = 200$, $m = 300$, $\rho = 0.5$,
3. $n_X = 200$, $n_Y = 200$, $m = 400$, $\rho = 0.5$.

First, we generate $W_X = \frac{1}{n_X}\mathbf{X}\mathbf{X}^t \in \mathbf{R}^{m \times n_X}$ and $W_Y = \frac{1}{n_Y}\mathbf{Y}\mathbf{Y}^t \in \mathbf{R}^{m \times n_Y}$, where

$$\mathbf{X}_{\cdot,i} \sim \mathbf{N}_m\left(\vec{0}, \mathrm{I}_m\right) \text{ and } \mathbf{X}_{\cdot,i+1} = \rho \mathbf{X}_{\cdot,i} + \sqrt{1-\rho^2}\ \epsilon_{i+1}, \text{ with } \epsilon_{i+1} \overset{i.i.d}{\sim} \mathbf{N}_m\left(\vec{0}, \mathrm{I}_m\right),$$
$$\mathbf{Y}_{\cdot,i} \sim \mathbf{N}_m\left(\vec{0}, \mathrm{I}_m\right) \text{ and } \mathbf{Y}_{\cdot,i+1} = \rho \mathbf{Y}_{\cdot,i} + \sqrt{1-\rho^2}\ \epsilon_{i+1}, \text{ with } \epsilon_{i+1} \overset{i.i.d}{\sim} \mathbf{N}_m\left(\vec{0}, \mathrm{I}_m\right).$$

Then, we apply canonical perturbation $P_k = \mathrm{I}_m + \sum_{i=1} k(\theta_i - 1)e_i e_i^t \in \mathbf{R}^{m \times m}$ to the matrices W_X and W_Y,

$$\hat{\Sigma}_{X,P_k} = P_k^{1/2} W_X P_k^{1/2},$$
$$\hat{\Sigma}_{Y,P_k} = P_k^{1/2} W_Y P_k^{1/2}.$$

8.8.1 Invariant Double Angle Theorem

In this part we compute the generalized double angle of Theorem 5.10.1,

$$\sum_{i=1}^{k} \left\langle \hat{u}_{\hat{\Sigma}_{X,P_k},k}, \hat{u}_{\hat{\Sigma}_{Y,P_k},i} \right\rangle^2,$$

where we assume $\theta_1 < \theta_i$ for $i > 1$.

Figure 8.5 shows that in the first scenario, the value of the double angle is invariant of k. Indeed for a fixed color pink, we generate W_X and W_Y. We successively apply P_1, P_2,...,P_8 to the pink matrices and compute the pink statistics. Then we generate new matrices W_X and W_Y with same spectra as W_X and W_Y. Applying successively P_1, P_2,...,P_8 to the matrices leads to the cyan lines.

The fluctuations within the colour lines are small compared to the fluctuations between the color lines. The Table 8.10 shows that the moments of 1000 replicates of $\mathbf{X}$ and $\mathbf{Y}$ coincide with the Theorem 5.3.1 when the perturbation is large.

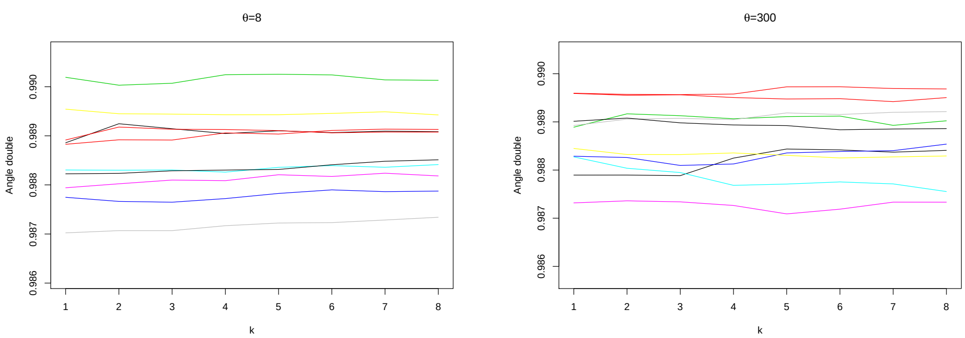


Figure 8.5 – Generalized double angle corresponding to θ_1 of $\left(P_k^{1/2} W_X P_k^{1/2}, P_k^{1/2} W_Y P_k^{1/2}\right)$, $\left(P_k^{1/2} W_X P_k^{1/2}, P_k^{1/2} W_Y P_k^{1/2}\right)$, $\left(P_k^{1/2} W_X P_k^{1/2}, P_k^{1/2} W_Y P_k^{1/2}\right)$,... in function of k. (The spectrum of all the W_X (respectively W_Y) are the same.) The perturbation is $P_k = \mathrm{I}_m + \sum_{i=1}^{k}(\theta_i - 1)e_i e_i^t$ with $\theta_1 = 8$ in the first figure and $\theta_1 = 300$ in the second one. The fluctuation between the colors lines represents the variance of the statistic, $O_p(1/\sqrt{m})$, and the stability inside each line shows the invariance, $O_p(1/m)$.

k		Theory	1	2	3	8
scenario 1	$1-\mu$	0.0113	0.0114	0.0113	0.0113	0.0113
	σ	0.0010	0.0011	0.0011	0.0011	0.0011
scenario 2	$1-\mu$	0.0128	0.0124	0.0123	0.0123	0.0123
	σ	0.0014	0.0015	0.0015	0.0015	0.0015
scenario 3	$1-\mu$	0.0219	0.0222	0.0222	0.0222	0.0222
	σ	0.0023	0.0025	0.0026	0.0026	0.0027

Table 8.10 – Moments of the double angle when the perturbation is large, $\theta_1 = 300$ and $\theta_i > \theta_1$ for $i > 1$. (Number of replicates: 1000)

8.8.2 Invariant Dot Product Theorem

In this part, we compute the dot product for $k > 1$ of Theorem 5.7.1,

$$\sum_{i=k+1}^{m} \hat{u}_{\hat{\Sigma}_{X,P_k},i,k}, \hat{u}_{\hat{\Sigma}_{X,P_k},i,2},$$

where we assume $\theta_1, \theta_2 < \theta_i$ for $i > 2$ and the convention $\hat{u}_{\hat{\Sigma}_{X,P_k},i,i} > 0$.
Figure 8.6 shows that in the three scenarios, the value of the dot product is invariant of k. This figure is explained in the previous section. The Table 8.11 shows that the moments of 1000 replicates of $\mathbf{X}$ and $\mathbf{Y}$ coincide with the Theorem 5.3.1 when the perturbation is large.

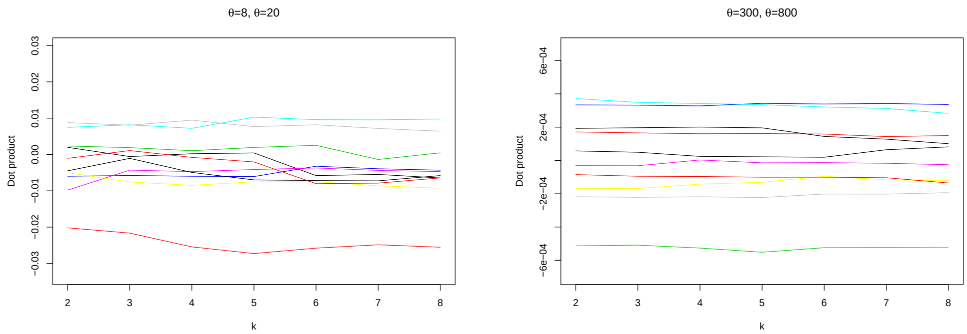


Figure 8.6 – Dot product corresponding to θ_1 and θ_2 of $P_k^{1/2} W_X P_k^{1/2}$, $P_k^{1/2} W_X P_k^{1/2}$, $P_k^{1/2} W_X P_k^{1/2}$,... in function of k. (The spectrum of all the W_X are the same.) The perturbation is $P_k = \mathrm{I}_m + \sum_{i=1}^{k} (\theta_i - 1) e_i e_i^t$ with $\theta_1 = 8$, $\theta_2 = 20$ in the first figure and $\theta_1 = 300$, $\theta_2 = 800$ in the second one. The fluctuation between the colors lines represents the variance of the statistic, $O_p(1/\sqrt{m})$, and the stability inside each line shows the invariance, $O_p(1/m)$.

k		Theory	1	2	3	8
scenario 1	μ	0	0.0000	0.0000	0.0000	0.0000
	σ	0.000208	0.000214	0.000214	0.000213	0.000212
scenario 2	μ	0	0.0000	0.0000	0.0000	0.0000
	σ	0.000182	0.000176	0.000176	0.000175	0.000177
scenario 3	μ	0	0.0000	0.0000	0.0000	0.0000
	σ	0.000483	0.000493	0.00049	0.000492	0.000498

Table 8.11 – Moments of the dot product when the perturbation is large, $\theta_1 = 300$, $\theta_2 = 800$ and $\theta_i > \theta_2$ for $i > 2$.

Chapter 9

Conclusion

In this thesis, we defined the new notion of the residual spike in Chapter 2. Using it, we built a statistical test for comparing two populations in Chapter 3. The proposed test is based on some hypotheses. Therefore, we investigated

1. Its good robust properties in Section 4.2.
2. The consequences of the preprocessing in Section 4.4.
3. The impact of correlated data on the results in Section 4.4.3.

On the other hand, in Chapter 5, we proved many theorems of Random matrix theory such as the central theorem of this thesis, the Invariant Angle Theorem 5.5.1.

This thesis also addressed difficulties in relaxing some assumptions, and we now list some interesting possibilities for future work.

Extension to any eigenvalues: The Main Theorem 3.1.1 assumes large eigenvalues. An interesting extension could show the distribution of the residual spike for any perturbations.

Large eigenvalues: We assume that real data will always create a larger residual spike when the eigenvalues are large. An interesting question concerns the necessary conditions to satisfy this assumption. Recall that we already proposed some criteria.

General distribution: The simulations of Chapter 8 seem to show that the variance of the residual spike depends on the fourth moment of the random matrices entries. It would be of interest to find a bound on this variance as a function of the fourth moment.

Faster procedure: The computation of the spectra is the most computationally intensive part of the procedure. Finding a bound on the extreme residual spikes as a function of "worst spectra" computed only with a few extreme eigenvalues of the spectra would be of interest. Because we only look at the first four moments of the spectra, another interesting work could compare the speed of the estimation of the spectra with the traces of the first four powers of the covariance matrices.

Outlier and temporal perturbation: We already proposed an heuristic algorithm to cancel temporal perturbations in Section 4.4.3. An interesting alternative could be to use robust estimators of the covariance matrices. This approach is proposed by Couillet et al. [2014], Couillet et al. [2015] or Kammoun and Alouini [2017].

Complex and quaternion: Many existing results of random matrices concern random matrices with complex or quaternion numbers. Some examples using quaternions could be Mays and Ponsaing [2017], Mays [2013] or Wang [2009]. The extension of our procedure beyond $\mathbb{R}$ could find applications in shape analysis as in Dryden and Mardia [2016].

Bibliography

G. W. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to Random Matrices*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2009. ISBN 9780521194525. URL https://doi.org/10.1017/CBO9780511801334.

T. W. Anderson. *An introduction to Multivariate Statistical Analysis*. Wiley publications in statistics. Wiley, 1958.

T. W. Anderson. *An introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Statistics. Wiley, 2003. ISBN 9780471360919.

Z. Bai and J. W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*. Springer, 2010. URL https://doi.org/10.1007/978-1-4419-0661-8.

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, 2009. URL https://doi.org/10.1214/09-AOS694.

Z. D. Bai and J. W. Silverstein. CLT for linear spectral statistics of large-dimensional sample covariance matrices. *The Annals of Probability*, 32(1A):553–605, 2004. URL https://doi.org/10.1214/aop/1078415845.

J. Baik and J. W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97:1382–1408, 2005. URL https://doi.org/10.1016/j.jmva.2005.08.003.

J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for non-null complex covariance matrices. *The Annals of Probability*, 33(5):1643–1697, 2005. URL https://doi.org/10.1214/009117905000000233.

Z. Bao, G. Pan, and W. Zhou. Universality for the largest eigenvalue of sample covariance matrices with general population. *The Annals of Statistics*, 43(1):382–421, 2015. URL https://doi.org/10.1214/14-AOS1281.

F. Benaych-Georges and N. R. Rao. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227:494–521, 2009. URL https://doi.org/10.1016/j.aim.2011.02.007.

F. Benaych-Georges, A. Guionnet, and M. Maida. Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. *Electronic Journal of Probability*, 16:1621–1662, 2010. URL https://doi.org/10.1214/EJP.v16-929.

V. Bentkus. A Lyapunov type bound in $\mathbf{R}^d$. *Theory of Probability and Its Applications*, 49(2):311–323, 2005. URL https://doi.org/10.1137/S0040585X97981123.

P. Bianchi, M. Debbah, M. Maida, and J. Najim. Performance of statistical tests for single source detection using random matrix theory. *Information Theory, IEEE Transactions on*, 57:2400–2419, 2011. URL https://doi.org/10.1109/TIT.2011.2111710.

A. Bloemendal, A. Knowles, H.-T. Yau, and J. Yin. On the principal components of sample covariance matrices. *Probability Theory and Related Fields*, 164(1):459–552, 2016. ISSN 1432-2064. URL `https://doi.org/10.1007/s00440-015-0616-x`.

A. Bose. *Patterned Random matrices.* Chapman and Hall/CRC, 2018. ISBN 9781138591462.

K. Cawse-Nicholson, M. Sears, A. Robin, S. Damelin, K. Wessels, F. van den Bergh, and R. Mathieu. Using random matrix theory to determine the number of endmembers in a hyperspectral image. *2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, WHISPERS 2010 - Workshop Program*, pages 1–4, 2010. URL `https://doi.org/10.1109/WHISPERS.2010.5594854`.

R. Couillet, F. Pascal, and J. Silverstein. Robust estimates of covariance matrices in large dimensional regime. *Information Theory, IEEE Transactions on*, 60(11):7269–7278, 11 2014. URL `https://doi.org/10.1109/TIT.2014.2354045`.

R. Couillet, F. Pascal, and J. Silverstein. The random matrix regime of maronna's m-estimator with elliptically distributed samples. *Journal of Multivariate Analysis*, 139:56–78, 7 2015. URL `https://doi.org/10.1016/j.jmva.2015.02.020`.

I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis, with Applications in R. Second Edition.* John Wiley and Sons, Chichester, 2016. ISBN 978-0470699621.

N. El Karoui. Tracy-widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *The Annals of Probability*, 35(2):663–714, 2007. URL `https://doi.org/10.1214/009117906000000917`.

N. El Karoui. Spectrum estimation for large dimensional covariance matrices using random matrix theory. *The Annals of Statistics*, 36(6):2757–2790, 2008. URL `https://doi.org/10.1214/07-AOS581`.

I. M. Johnstone. On the distribution of the largest eigenvalue in principal component analysis. *The Annals of Statistics*, 29(2):295–327, 2001. URL `https://doi.org/10.1214/aos/1009210544`.

A. Kammoun and M.-S. Alouini. The random matrix regime of maronna's m-estimator for observations corrupted by elliptical noises. *Journal of Multivariate Analysis*, 162:51–70, 11 2017. URL `https://doi.org/10.1016/j.jmva.2017.08.002`.

T. Kato. *Perturbation Theory.* 132. Springer, The address, 1 edition, 1995. ISBN 978-3-662-12678-3. URL `https://doi.org/10.1007/978-3-662-12678-3`.

V. A. Marchenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Math. USSR*, 1:457–483, 1967. URL `https://doi.org/10.1070/SM1967v001n04ABEH001994`.

K. Mardia, J. Kent, and J. Bibby. *Multivariate Analysis.* Probability and mathematical statistics. Academic press, 1979. ISBN 9780124712508. URL `https://doi.org/10.1002/bimj.4710240520`.

A. Mays. A real quaternion spherical ensemble of random matrices. *Journal of Statistical Physics*, 153(1):48–69, 10 2013. URL `https://doi.org/10.1007/s10955-013-0808-7`.

A. Mays and A. Ponsaing. An induced real quaternion spherical ensemble of random matrices. *Random Matrices: Theory and Applications*, 6(1):51–70, 11 2017. URL `https://doi.org/10.1142/S2010326317500010`.

X. Mestre and M. Lagunas. Modified subspace algorithms for doa estimation with large arrays. *Signal Processing, IEEE Transactions on*, 56:598–614, 2008. URL `https://doi.org/10.1109/TSP.2007.907884`.

R. J. Muirhead. *Aspect of Multivariate Statistical Theory*. Wiley Series in Probability and Statistics. Wiley-Interscience, 2005. ISBN 9780471769859. URL `https://doi.org/10.1002/9780470316559`.

D. Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, 17, 2007.

N. S. Pillai and J. Yin. Edge universality of correlation matrices. *The Annals of Statistics*, 40(3): 1737–1763, 2012. URL `https://doi.org/10.1214/12-AOS1022`.

N. S. Pillai and J. Yin. Universality of covariance matrices. *The Annals of Applied Probability*, 24(3): 935–1001, 2014. URL `https://doi.org/10.1214/13-AAP939`.

K. Ray, W. Jonathan, G. Basu, and P. Panigrahi. Random matrix route to image denoising. *2012 International Conference on Systems and Informatics, ICSAI 2012*, pages 1975–1980, 2012. URL `https://doi.org/10.1109/ICSAI.2012.6223437`.

J. W. Silverstein. Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *Journal of Multivariate Analysis*, 55:331–339, 1995. URL `https://doi.org/10.1006/jmva.1995.1083`.

T. Tao. Topics in random matrix theory. `http://www.math.hkbu.edu.hk/~ttang/UsefulCollections/matrix-book-2011-08.pdf`, 2012.

P. Vallet, W. Hachem, P. Loubaton, X. Mestre, and J. Najim. An improved music algorithm based on low rank perturbation of large random matrices. In *2011 IEEE Statistical Signal Processing Workshop (SSP)*, pages 689–692, 2011. URL `https://doi.org/10.1109/SSP.2011.5967795`.

D. Wang. The largest sample eigenvalue distribution in the rank 1 quaternionic spiked model of wishart ensemble. *The Annals of Probability*, 37(4):1273–1328, 07 2009. URL `https://doi.org/10.1214/08-AOP432`.

K. Wang. Random covariance matrices: Universality of local statistics of eigenvalues up to the edge. *Random Matrices: Theory and Applications*, 1:24–45, 2012. URL `https://doi.org/10.1142/S2010326311500055`.

M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(2):387–392, 1985. URL `https://doi.org/10.1109/TASSP.1985.1164557`.

Rémy Mariétan
Statisticien

Chemin des Croix-Rouges 8
1007 Lausanne
Tél : 079 482 47 36
Email : remy.marietan@alumni.epfl.ch

Suisse
Célibataire
16.09.90 (28 ans)

FORMATION

2014 – 2019	Doctorat en Statistique à l'EPFL.
2009 – 2014	Bachelor et Master en mathématique appliquée à l'EPFL.
2004 – 2009	Maturité gymnasiale au Collège des Creusets à Sion, section mathématique.

EXPERIENCE PROFESSIONNELLE

2014 – 2018 Doctorant à l'EPFL :

- 75% : Recherche en théorie des matrices aléatoires et construction d'applications basées sur les récentes avancées du domaine.
- 25% : Travail d'assistant académique pour différents cours du Bachelor au Master dans des domaines statistiques.

2013 – 2014 Stagiaire chez Hauffmann-La Roche en département de **Biostatistique** :

Développement d'une procédure de bioéquivalence basée sur des statistiques multivariées et des **méthodes bayésiennes**.

2012-2013, 2017-2018 Enseignant en mathématique à l'ECAV à temps partiel.

CONNAISSANCES

Statistiques

Cours suivis	Modèle de **régression généralisée**, **séries temporelles**, **méthodes de Monte Carlo**, méthodes de Biostatistiques, statistiques spatiales.
Cours donnés	Théorie de la statistique, **statistiques multivariées, statistiques robustes et non-paramétriques**.

Informatique

R, Latex	Connaissance très approfondie pendant le doctorat et à travers divers projets.
Python	Bonne connaissance acquise durant le doctorat.
Matlab, C++	Langages appris et utilisés durant le Bachelor.

Langues

Français	Langue maternelle.
Anglais	Bonne connaissance (B2), séjours linguistiques en 2010 et 2012, enseignement et travail de recherche réalisé en anglais à l'EPFL.

Ce document a été imprimé au Centre d’impression EPFL,
imprimerie climatiquement neutre, certifiée myClimate.