arXiv: 1708.02107

Adaptive Estimation of Nonparametric Geometric Graphs

Yohann De Castro, Claire Lacour and Thanh Mai Pham Ngoc

Laboratoire de Mathématiques d'Orsay
Univ. Paris-Sud, CNRS, Université Paris-Saclay, 91405 Orsay, France

Abstract: This article studies the recovery of graphons when they are convolution kernels on compact (symmetric) metric spaces. This case is of particular interest since it covers the situation where the probability of an edge depends only on some unknown nonparametric function of the distance between latent points, referred to as Nonparametric Geometric Graphs (NGG).
In this setting, almost minimax adaptive estimation of NGG is possible using a spectral procedure combined with a Goldenshluger-Lepski adaptation method. The latent spaces covered by our framework encompasses (among others) compact symmetric spaces of rank one, namely real spheres and projective spaces. For these latter, explicit computations of the eigenbasis and of the model complexity can be achieved, leading to quantitative non-asymptotic results. The time complexity of our method scales cubicly in the size of the graph and exponentially in the regularity of the graphon. Hence, this paper offers an algorithmically and theoretically efficient procedure to estimate smooth NGG.
As a by product, this paper shows a non-asymptotic concentration result on the spectrum of integral operators defined by symmetric kernels (not necessarily positive).

MSC 2010 subject classifications: Primary 62G05; secondary 60C05, 60B15.

Keywords and phrases: Graphon, Random Networks, Spectral Estimation, Kernel Matrix, Convolution Operator, Sphere.

1. Introduction

Over the recent years, the study of networks has become prevailing in many fields. Through the advent of social networks, biological neural networks, food webs, protein interaction in genomics and World wide web for instance, large scale data have become available. Extracting information from those repositories of data is a true challenge. Random graphs prove to be particularly relevant to model real-world networks. They are capable to capture complex interactions between actors of a system. Vertices of a random graph usually represent entities of a system and the edges stand for the presence of a specified relation between those entities. An important statistical problem is seeking better and more informative representations of random graphs.

Following the seminal work of [13] various random graphs models have been suggested, see [5, 26, 20, 23] and references therein. Aside from classical random graphs, random geometric graphs, see [28], have emerged as an interesting alternative to model real networks having spatial content. Examples include the Internet (where the nodes are the routers) and other physical communication networks such as road networks or neural networks in the brain. Recall that a random geometric graph is an undirected graph in which each vertex is assigned a latent (unobservable) random label in some metric spaces $\mathbf{S}$. Two vertices are connected by an edge if the distance between them is smaller than some threshold. Assuming that the underlying metric is the unit sphere $\mathbb{S}^{\mathbf{d}-1}$ and latent variables drawn from the uniform distribution on $\mathbb{S}^{\mathbf{d}-1}$, the paper [8] considered the problem of testing if the observed graph is an Erdös-Rényi one (no geometric structure) or a geometric graph on the sphere where points are connected if their distance is smaller than some threshold.

More generally, random graphs with latent space can be characterized by the so-called graphon. In fact, graphons can be seen as kernel functions for latent position graphs. For more insight about the theory of graphon, we refer to the excellent monograph of [22]. In the case of graphons defining positive

imsart-generic ver. 2014/10/16 file: Graph17.tex date: February 1, 2018

definite kernels, the paper [33] proved that the eigen-decomposition of the adjacency matrix yields consistent estimator of the graphon feature maps involving the latent variables. Besides, nonparametric representations of graphons has gained attention. Statistical approaches on estimating graphons have been developed using Least-Squares estimation [19] or Maximum Likelihood estimation [37]. Dealing with estimation of (sparse) graphons from the observation of the adjacency matrix, the paper [19] derives sharp rates of convergence for the $\mathbf{L}^2$ loss for the Stochastic Block Model.

1.1. A Statistical Pledge for Structured Latent Spaces

The graphons are limiting objects that describe large dense graphs. The graphon model [22] is standardly and without loss of generality formulated choosing $[0,1]$ as latent space. In this model, given latent points $x_1,\dots,x_n \in [0,1]$, the probability to draw an edge between i and j is $W(x_i,x_j)$ where W is a function from $[0,1]^2$ onto $[0,1]$, referred to as a graphon. This model is general and well referenced in the literature—as mentioned earlier, the reader may consult the book [22] for further details.

However, this model may underneath intrinsic features of a random graph. For instance, recall the prefix attachment graph model [22, page 190] where the nodes are added one at a time and each new node connects to a random previous node and all its predecessors. In this model, the graph sequence converges in cut distance [22, Proposition 11.42] to the graphon W_{pref} defined as

$$\forall (x_1,y_1),(x_2,y_2) \in [0,1]^2, \quad W_{\mathrm{pref}}((x_1,y_1),(x_2,y_2)) = \mathbb{1}(x_1 < x_2 y_2) + \mathbb{1}(x_2 < x_1 y_1)\,, \tag{1}$$

up to a measure preserving homomorphism of the latent space $[0,1]^2$. From a statistical point of view, the estimation of the function $((x_1,y_1),(x_2,y_2)) \mapsto \mathbb{1}(x_1 < x_2 y_2)$ from sample points $((x_k,y_k))_k$ uniformly distributed on $[0,1]^2$ is a well understood standard task.

Yet one may also represent this graphon on the standard latent space $[0,1]$. And, in this case, one cannot represent this graphon using the indicator function of two symmetric convex sets with piecewise smooth border as done in (1). Actually, in this case, a fractal-like structure appear and the statistical estimation of this function seems more difficult than in (1). Our statement may be loose here but one may emphasize that there may exist better latent spaces than $[0,1]$ on which the graphon may presents a simple and better estimable formulation.

An other important statistical issue is that, by construction, graphons are defined on an equivalent class "*up to a measure preserving homomorphism*" and it can be challenging to have a simple description from an observation given by sampled graph—since one has to deal with all possible composition of a bivariate function by any measure preserving homomorphism. In this paper, we circumvent this disappointing statistical issue restraining our attention to graph models for which the probability of appearance of an edge depends as a nonparametric function of the distance between latent points.

1.2. Main results

In this paper, we focus on latent metric spaces for which the distance is invariant by translation (or conjugation) of pairs of points. This natural assumption leads to consider that the latent space $\mathbf{S}$ has some group structure, namely it is a compact Lie group or some compact symmetric space. Hence, consider graphons defined as functions $\mathbf{p}$ of (the cosine of) the distance γ (normalized so that the range of γ equals $[0,\pi]$) of some compact Lie group $\mathbf{S}$, or more generally of some compact symmetric space, see Section 4. In this case, the graphon is given by

$$\forall x,y \in \mathbf{S}, \quad W(x,y) = \mathbf{p}(\cos\gamma(x,y)) = \mathbf{p}(\cos\gamma(z,e)) \quad \text{and} \quad z = xy^{-1}\,,$$

where y^{-1} is the inverse of y, e denotes the identity element of $\boldsymbol{S}$ and $\mathbf{p}$ is a function from $[-1,1]$ onto $[0,1]$ referred to as the "*envelope*". In the case when $\boldsymbol{S}$ is the Euclidean sphere, we consider graphons that are a function $\mathbf{p}$ of cosine of the distance, namely $\cos\gamma(x,y) = \langle x, y\rangle$, between latent points $x, y \in \boldsymbol{S}$.

First, note that W, viewed as an integral operator on square-integrable functions, is a compact convolution (on the left) operator. The convolution (on the left) kernel is simply

$$z \in \boldsymbol{S} \mapsto \mathbf{p}(\cos\gamma(z, e)) \in [0,1]\,.$$

Then the main point is that there exists an $\mathbf{L}^2$-decomposition of the Hilbert space of square integrable functions such that the eigenfunctions basis of the convolution kernel (and the graphon W viewed as a integral operator) depends only on the latent space $\boldsymbol{S}$ and does not depend on the function $\mathbf{p}$. This basis is the irreducible characters in the (Lie) Group case and the zonal spherical functions in the non-Group case, see Cartan's Classification of "sscc" Lie Groups and "ssccss" in Section 4 for further details. This decomposition can be pushed on $[-1,1]$ and one gets an $\mathbf{L}^2$-decomposition of the envelope function $\mathbf{p}$ such that the orthonormal basis (Z_ℓ) depends only on the latent space $\boldsymbol{S}$ and does not depend on $\mathbf{p}$, see (2). Furthermore, the eigenvalues $\lambda^\star = (\lambda^\star_k)_{k\geq 0}$ of the kernel W are exactly (up to some known multiplicities and up to some known multiplicative constants) the coefficients of $\mathbf{p}$ onto the orthonormal basis $(Z_\ell)_{\ell\geq 0}$. Hence, the graphon W is entirely described by the univariate function $\mathbf{p}$ defined on $[-1,1]$. It follows that this subclass of graphons may be well suited for estimation since it reduces to estimate a simple univariate function on $[-1,1]$.

Now, consider the case when $\boldsymbol{S}$ is one of the compact symmetric space of rank one—namely real spheres or real/complex/quaternionic/octonionic projective spaces. In this case, one can explicitly give the decomposition of the envelope function $\mathbf{p}$. One can prove that the orthonormal polynomials $(Z_\ell)_{\ell\geq 0}$ are the orthonormal polynomials (more precisely, normalized Jacobi polynomials) of some Beta law with known shape parameters (α, β), see Table 1 for the explicit values. This decomposition is given by

$$\mathbf{p} = \sum_{\ell\geq 0} \sqrt{d_\ell}\,\mathbf{p}^\star_\ell Z_\ell \quad \text{and} \quad \mathbf{p}^\star_\ell = \frac{1}{\sqrt{d_\ell}} \langle \mathbf{p}, Z_\ell\rangle_{\mathbf{L}^2([-1,1],\mathbf{w})}\,, \tag{2}$$

in $\mathbf{L}^2([-1,1],\mathbf{w})$ where $\mathbf{w}$ denotes the density function of the Beta distribution. We further assume that there exists $s > 0$, a (Sobolev) regularity parameter, such that

$$\forall R \geq 1\,, \quad \sum_{\ell > R} d_\ell (\mathbf{p}^\star_\ell)^2 \leq C(\mathbf{p}, s, \boldsymbol{S}) R^{-2s}\,.$$

for some constant $C(\mathbf{p}, s, \boldsymbol{S}) > 0$ and for some known dimensions $(d_\ell)_{\ell\in\mathbb{N}}$ (given by the representation of the group/quotient $\boldsymbol{S}$) that depends only on $\boldsymbol{S}$, see Table 1. This assumption governs the regularity of the kernel W and it can be understood that the derivative of order s (in the Laplacian on $\boldsymbol{S}$ sense) of W is square-integrable. In this case, one can build an estimator $\widehat{\lambda}^{\widehat{R}}$ (from the spectrum of the adjacency matrix of the graph) of the spectrum $\lambda^\star$ of W (viewed as an integral operator) such that

$$\mathbb{E}\Big[\delta_2^2\big(\widehat{\lambda}^{\widehat{R}}, \lambda^\star\big)\Big] = \mathcal{O}\Big[\Big(\frac{n}{\log n}\Big)^{-\frac{2s}{2s+(\mathbf{d}-1)}}\Big],$$

where n is the size of the graph, $\mathbf{d}$ is the dimension of the latent space (actually, $\boldsymbol{S}$ is a $(\mathbf{d}-1)$-manifold) and δ_2 is the ℓ_2 distance between spectra, see (7) for a definition. We uncover the minimax rate of estimating a s-regular function on a space of (Riemannian) dimension $\mathbf{d}-1$ up to a multiplicative log factor. This result is stated in Theorem 6 without adaptation to the smoothness parameter, Theorem 7 and Corollary 8 with smoothness adaptation, and Theorem 9 and Corollary 10 for adaptive estimation of the envelope function $\mathbf{p}$ at rate $\mathcal{O}(\log n/n)$ when $\mathbf{p}$ is a polynomial. The general statement for compact symmetric spaces is given by Theorem 11.

Note that our results hold for general convolution kernels and not necessarily semidefinite positive kernels. Indeed, it is often assumed in the literature, see for instance [14, 29, 33, 32], that the graphon W is a semidefinite positive kernels. In this case, the adjacency matrix of the random graph is almost surely semidefinite positive, which is a strong requirement in Graph theory. To bypass this limitation, our approach does not use any RKHS representation but a new non-asymptotic concentration result on the integral operator, see Theorem 2 and Corollary 3. The rates uncovered by these results allow us to introduce a minimax adaptive estimation procedure of the spectrum of the graphon.

From a computational point of view, Theorem 5 enlightens on the time complexity of our estimator. Remarkably, the time complexity is $n^3 + (R_{\max} + 2)!$, that is cubic in the graph size n (as any spectral method) and exponential in the number of coefficients $\mathbf{p}_\ell^\star$ one has to estimate. The spatial complexity is quadratic in n as one has to store the adjacency matrix of the graph.

1.3. Outline

The convergence of the spectrum of the "*matrix of probabilities*" towards the spectrum of the integral operator in a non-asymptotic frame is given in Section 2. Then, we begin our study by a comprehensive example on the **d**-dimensional sphere in Section 3. Interestingly, we uncover that the spectrum of the graphon (viewed as a kernel operator) presents a structure: the eigenvalues have prescribed multiplicities and the eigenvectors are fixed—they are the spherical harmonics. Adaptive minimax estimation of the spectrum of the graphon W (viewed as an integral operator) is proved and computational complexities are discussed. Extensions to compact symmetric spaces is done in Section 4. Numerical experiments are presented in Section 5. The proofs are given in the appendix.

2. Spectral Convergence of the Sampled graphons

2.1. Estimating the Matrix of Probabilities

We denote $[n] := \{1, \dots, n\}$ for all $n \geq 1$. Consider a random undirected graph $\boldsymbol{G}$ with n nodes and assume that we observe its $n \times n$ adjacency matrix $\boldsymbol{A}$ given by entries $\boldsymbol{A}_{ij} \in \{0, 1\}$ where $\boldsymbol{A}_{ij} = 1$ if the nodes i and j are connected and $\boldsymbol{A}_{ij} = 0$ otherwise. We set $\boldsymbol{A}_{ii} = 0$ on its diagonal entries for all $i \in [n]$ and we assume that $\boldsymbol{A}_{ij}$ are independent Bernoulli random variables with $(\boldsymbol{\Theta}_0)_{ij} := \mathbb{P}\{\boldsymbol{A}_{ij} = 1\}$ for $1 \leq i < j \leq n$. We denote by $\boldsymbol{\Theta}_0$ the $n \times n$ symmetric matrix with entries $(\boldsymbol{\Theta}_0)_{ij}$ for $1 \leq i < j \leq n$ and zero diagonal entries. This is a matrix of probabilities associated to the random graph $\boldsymbol{G}$. Throughout this paper, we denote by

$$\widehat{\boldsymbol{T}}_n := (1/n)\boldsymbol{A} \quad \text{and} \quad \boldsymbol{T}_n := (1/n)\boldsymbol{\Theta}_0 \,. \tag{3}$$

Using the central limit theorem, it is elementary that, with high probability,

$$\|\widehat{\boldsymbol{T}}_n - \boldsymbol{T}_n\|_F^2 \in \left[\frac{1}{n^2} \sum_{i \neq j} (\boldsymbol{\Theta}_0)_{ij}(1 - (\boldsymbol{\Theta}_0)_{ij}) \pm \mathcal{O}\left(\frac{1}{n}\right) \right],$$

where $\|\cdot\|_F$ denotes the Frobenius norm. We witness that the Euclidean distance between $\widehat{\boldsymbol{T}}_n$ and $\boldsymbol{T}_n$ is of the order of the average value of the entries $(\boldsymbol{\Theta}_0)_{ij}$. The Frobenius norm seems an inappropriate prediction loss. Instead, our analysis leverages the operator norm $\|\cdot\|$ loss to account for the distance between the observation $\widehat{\boldsymbol{T}}_n$ and the target parameter $\boldsymbol{T}_n$. Furthermore, a near optimal error bound can be derived for the operator norm $\|\cdot\|$ loss as shown in [2].

Proposition 1 (Bandeira & van Handel [2]). *There exists a universal constant $C_0 > 0$ such that for all $\alpha \in (0, 1)$, it holds*

$$\mathbb{P}\left\{ \|\widehat{\boldsymbol{T}}_n - \boldsymbol{T}_n\| \geq 3\frac{\sqrt{2\boldsymbol{D}_0}}{n} + C_0 \frac{\sqrt{\log(n/\alpha)}}{n} \right\} \leq \alpha \tag{4}$$

where $\boldsymbol{D}_0 = \max_{i \in [n]} \Big[\sum_{j \in [n]} (\boldsymbol{\Theta}_0)_{ij}(1 - (\boldsymbol{\Theta}_0)_{ij}) \Big] \leq n/4.$

A proof is recalled in Appendix A.1. Proposition 1 is of particular interest giving an error bound on each eigenvalue $\lambda_k(\boldsymbol{T}_n)$ of $\boldsymbol{T}_n$, where $\lambda_k(M)$ denotes the k-th largest eigenvalue of the symmetric matrix M.

Indeed, it holds, with probability greater that $1 - n\exp(-n)$,

$$\forall k \in [n], \quad |\lambda_k(\widehat{\boldsymbol{T}}_n) - \lambda_k(\boldsymbol{T}_n)| \leq \|\widehat{\boldsymbol{T}}_n - \boldsymbol{T}_n\| = \mathcal{O}(1/\sqrt{n}), \tag{5}$$

by Weyl's perturbation Theorem, see [4, page 63] for instance.

2.2. Non-Asymptotic Error Bounds on the Kernel Spectrum in δ_2-metric

We understand that the spectrum of $\widehat{\boldsymbol{T}}_n$ can be a good approximation of the spectrum of $\boldsymbol{T}_n$ in the sense of (5). Assuming a graphon W model we can link the spectrum of $\boldsymbol{T}_n$ (sampled graphon onto the latent points $X_1, \ldots, X_n$ see below) to the spectrum of an integral operator $\mathbb{T}_W$ defined by the graphon W viewed as a symmetric kernel. More precisely, we consider $J := (\boldsymbol{S}, \mathscr{A}, \boldsymbol{\sigma})$ a probability space on $\boldsymbol{S}$ endowed with measure $\boldsymbol{\sigma}$ on the σ-algebra $\mathscr{A}$, and $W : \boldsymbol{S} \times \boldsymbol{S} \to [0,1]$ a symmetric $\boldsymbol{\sigma}$-measurable function. The couple (J, W) is referred to as a graphon, see for instance [22, Chapter 13]. We then define a probabilistic model on $\boldsymbol{\Theta}_0$ setting

$$(\boldsymbol{\Theta}_0)_{i,j} = W(X_i, X_j) \text{ for } i \neq j \text{ and } 0 \text{ otherwise}$$

where $X_1, \ldots, X_n$ are i.i.d. drawn w.r.t. $\boldsymbol{\sigma}$. Assume that the kernel satisifies $W \in \boldsymbol{L}^2(\boldsymbol{S} \times \boldsymbol{S}, \boldsymbol{\sigma} \otimes \boldsymbol{\sigma})$, so that

$$\forall x \in \boldsymbol{S},\ \forall g \in \mathbf{L}^2(\boldsymbol{S}, \boldsymbol{\sigma}), \quad (\mathbb{T}_W g)(x) = \int_S W(x,y) g(y) \mathrm{d}\boldsymbol{\sigma}(y),$$

defines a symmetric Hilbert-Schmidt operator $\mathbb{T}_W$ on $\mathbf{L}^2(\boldsymbol{S}, \boldsymbol{\sigma})$ and we can invoke the spectral theorem. Hence, it holds that, in the $\boldsymbol{L}^2(\boldsymbol{S} \times \boldsymbol{S}, \boldsymbol{\sigma} \otimes \boldsymbol{\sigma})$-sense,

$$\text{for almost every } x, y \in \boldsymbol{S}, \quad W(x,y) = \sum_{k \geq 1} \lambda_k^\star \phi_k(x) \phi_k(y), \tag{6}$$

for an $\boldsymbol{L}^2(\boldsymbol{S}, \boldsymbol{\sigma})$-orthonormal basis $(\phi_i)_{i \geq 1}$. This operator has a discrete spectrum, i.e. a countable multiset $\lambda^\star$ of nonzero (real) labeled eigenvalues $(\lambda_k^\star)_{k \geq 1}$ such that $\lambda_k^\star \to 0$. In particular, every nonzero eigenvalue has finite multiplicity. We are free to choose any labeling of the target eigenvalues $(\lambda_k^\star)_{k \geq 1}$ and observe that our results are valid for any choice of labeling. For instance, we can standardly label the eigenvalues in decreasing order with respect to their absolute values such that $|\lambda_1^\star| \geq |\lambda_2^\star| \geq \cdots$ and this gives results whose error rates (typically $\|W - W_R\|_2$ see below) are in terms of the best $\boldsymbol{L}^2$-approximation of rank R of the kernel W. An other choice may result in labeling the eigenvalues in increasing order of "frequencies". This labeling is natural for instance when we have a representation by spherical harmonics of the kernel as in Section 3. This gives results whose error rates are in terms of the best approximation by low frequency (i.e. the R first frequencies) kernels.

Given two sequences x and y of real numbers—completing finite sequences by zeros—such that it holds $\sum x_i^2 + y_i^2 < \infty$, we standardly define the ℓ_2-rearrangement distance $\delta_2(x,y)$ as

$$\delta_2(x,y) := \inf_{\pi \in \mathscr{P}} \Big[\sum (x_i - y_{\pi(i)})^2 \Big]^{\frac{1}{2}},$$

where the infimum is taken over $\mathscr{P}$ the set of permutations with finite support. Using Hardy-Littlewood rearrangement inequality [16, Theorem 368], it is standard to observe that

$$\delta_2(x,y) = \lim_{N \to \infty} \left[\sum_{k=-N}^{N} (x_k - y_k)^2 \right]^{\frac{1}{2}}, \tag{7}$$

with the convenient notation $x_{-1} \leq x_{-2} \leq \ldots \leq 0 \leq \ldots \leq x_2 \leq x_1 \leq x_0$ (respectively $y_{-1} \leq y_{-2} \leq \ldots \leq 0 \leq \ldots \leq y_2 \leq y_1 \leq y_0$) where we denote $x = (x_k)_{k \in \mathbb{Z}}$ (respectively $y = (y_k)_{k \in \mathbb{Z}}$) completing with zeros if necessary.

Using this metric we can compare the (finite) spectrum $\lambda(\boldsymbol{T}_n)$ of $\boldsymbol{T}_n$ to the (infinite) spectrum $\lambda^\star$ of $\mathbb{T}_W$. To the best of our knowledge, existing results on this issue assume that W is a *positive kernel* and use a RKHS representation and/or Mercer theorem. This assumption might seem meaningless for a graphon. Indeed, it implies that $\mathbb{T}_W$ is semi-definite and if $W = W_H$ is a "step-function" kernel representing a finite graph H, it implies that the adjacency matrix of H is semi-definite which might be seen as restrictive. In this article, we bypass this limitation with the next result based on the analysis developed in [21] and some recent development in random matrix concentration, see [34] for instance.

Theorem 2. *Let $W \in \boldsymbol{L}^2(\boldsymbol{S}\times\boldsymbol{S},\boldsymbol{\sigma}\otimes\boldsymbol{\sigma})$ be a symmetric kernel and let $(\phi_k)_{k\geq1}$ be an orthonormal eigenbasis as in (6). Let $R \geq 1$ and $\alpha \in (0,1/3)$. Set*

$$\rho(R) := \max\left[1, \left\|\sum_{r=1}^{R}\phi_r^2\right\|_\infty - 1\right] \quad \text{and} \quad \forall x,y\in\boldsymbol{S},\ W_R(x,y) := \sum_{i=1}^{R}\lambda_i^\star\phi_i(x)\phi_i(y).$$

Then, for all $n^3 \geq \rho(R)\log(2R/\alpha)$, it holds

$$\begin{aligned}\delta_2(\lambda(\boldsymbol{T}_n),\lambda^\star) \leq & 2\|W-W_R\|_2 + \|W-W_R\|_\infty\Big[\frac{2\log(2/\alpha)}{n}\Big]^{\frac14}\\ & + \|W_R\|_2\Big[\Big[\frac{\rho(R)\log(2R/\alpha)}{n}\Big]^{\frac12} + \Big[\frac{2\rho(R)}{n}\Big(1+\max_{1\leq r\leq R}\|\phi_r^2\|_\infty\sqrt{\frac{\log(R/\alpha)}{2n}}\Big)\Big]^{\frac12}\Big],\end{aligned}$$

with probability at least $1-3\alpha$.

A proof of Theorem 2 can be found in Appendix A.2. This result shows that for all $n \geq n_0(R)$, it holds that $\delta_2(\lambda(\boldsymbol{T}_n),\lambda^\star) \leq 2\|W-W_R\|_2 + C_0(R)\,n^{-\frac14}$ with probability at least $1-3\alpha$, where the constants $n_0(R)\geq 1$ and $C_0(R) > 0$ may depend on R, the orthogonal basis $(\phi_k)_{k\in[R]}$, α and the graphon W. We have the following improvement for canonical kernels, see [11, Definition 3.5.1] for a definition.

Corollary 3. *Assume further that the kernel $(W-W_R)^2(x,y) - \mathbb{E}\big((W-W_R)^2\big)$ is canonical, namely*

$$\text{For almost every } x\in\boldsymbol{S},\quad \mathbb{E}\big((W-W_R)^2(x,X_1)\big) = \mathbb{E}\big((W-W_R)^2(X_1,X_2)\big)\,,$$

then there exist universal constants $C_1, C_2 > 0$ such that for all $n^3 \geq \rho(R)\log(2R/\alpha)$, it holds

$$\begin{aligned}\delta_2(\lambda(\boldsymbol{T}_n),\lambda^\star) \leq & 2\|W-W_R\|_2 + \|W-W_R\|_\infty\Big[\frac{C_1\log(C_2/\alpha)}{n}\Big]^{\frac12}\\ & + \|W_R\|_2\Big[\Big[\frac{\rho(R)\log(2R/\alpha)}{n}\Big]^{\frac12} + \Big[\frac{2\rho(R)}{n}\Big(1+\max_{1\leq r\leq R}\|\phi_r^2\|_\infty\sqrt{\frac{\log(R/\alpha)}{2n}}\Big)\Big]^{\frac12}\Big],\end{aligned}$$

with probability at least $1-3\alpha$.

A proof of this corollary can be found in Appendix A.6.

3. The Sphere Example, Prelude of Symmetric Compact Spaces

From a general point of view, this article focuses on the case where the value $W(x,y)$ depends on a nonparametric function $\mathbf{p}$ of the distance between the points x and y of a latent space $\boldsymbol{S}$ assumed a compact Lie group or a compact symmetric space, see Section 4 for further details. Such assumptions on the graphon W allows to lead the spectral analysis a step further. In this section, we restrict our analysis to the pleasant case of $\boldsymbol{S} := \mathbb{S}^{\mathbf{d}-1}$ the unit sphere of $\mathbb{R}^{\mathbf{d}}$ equipped with the uniform probability measure $\boldsymbol{\sigma}$ and the usual scalar product $\langle\cdot,\cdot\rangle$. In the literature, a popular model is given by the Random Geometric Graph for which the value $W(x,y)$ depends on the distance between the points x and y of the latent space $\mathbb{S}^{\mathbf{d}-1}$ and $W(x,y)=\mathbb{1}_{\langle x,y\rangle\geq\tau}$ for some threshold $\tau\in(-1,1)$ as in [12, 8]. From now on, assume that W only depends on the distance between latent points, namely

$$\forall x,y\in\mathbb{S}^{\mathbf{d}-1},\quad W(x,y)=\mathbf{p}(\langle x,y\rangle)$$

where $\mathbf{p}:[-1,1]\to[0,1]$ is an unknown function that is to be estimated. We refer to $\mathbf{p}$ as the "*envelope*" function.

3.1. Harmonic Analysis on $\mathbb{S}^{\mathbf{d}-1}$

Let us start by providing a brief overview on Fourier analysis on $\mathbb{S}^{\mathbf{d}-1}$. As pointed out above, in this case the operator $\mathbb{T}_W$ is a convolution (on the left) operator. Its spectral decomposition (6) satisfies that the orthonormal basis $(\phi_k)_k$ does not depend on $\mathbf{p}$ and the spectrum $\lambda(\mathbb{T}_W)$ is exactly described by the Fourier coefficients $(\mathbf{p}^\star_\ell)_\ell$ of $\mathbf{p}$, see [10, Lemma 1.2.3]. This remark remains true when the latent space $\boldsymbol{S}$ is assumed a compact Lie group or a compact symmetric space, see Section 4 for further details.

In the spherical case, the orthonormal basis of eigenfunctions consists of the real spherical harmonics. The following material can be found in [10]. Let us denote $\mathscr{H}_\ell$ the space of real spherical harmonics of degree ℓ with orthonormal basis $(Y_{\ell j})_{j\in[d_\ell]}$ where

$$d_\ell:=\dim(\mathscr{H}_\ell)=\binom{\ell+\mathbf{d}-1}{\ell}-\binom{\ell+\mathbf{d}-3}{\ell-2} \tag{8}$$

for $\ell\geq2$ and $d_0=1$, $d_1=\mathbf{d}$. Note that d_ℓ is therefore of order $\ell^{\mathbf{d}-2}$. In the sequel we identify $(\phi_k)_{k\geq1}=(Y_{\ell j})_{\ell\geq0,\,j\in[d_\ell]}$ so that the spectral decomposition (6) reads

$$\forall x,y\in\mathbb{S}^{\mathbf{d}-1},\quad W(x,y)=\mathbf{p}(\langle x,y\rangle)=\sum_{\ell\geq0}\mathbf{p}^\star_\ell\Big[\underbrace{\sum_{j=1}^{d_\ell}Y_{\ell j}(x)Y_{\ell j}(y)}_{\text{Zonal Harmonic}}\Big], \tag{9}$$

where $\lambda^\star=\{\mathbf{p}^\star_0,\mathbf{p}^\star_1,\dots,\mathbf{p}^\star_1,\dots,\mathbf{p}^\star_\ell,\dots,\mathbf{p}^\star_\ell,\dots\}$ and $\sum_{j=1}^{d_\ell}Y_{\ell j}(x)Y_{\ell j}(y)$ is a zonal harmonic of degree ℓ, see [10, Chapter 2]. The eigenvalue $\mathbf{p}^\star_\ell$ has multiplicity d_ℓ if the eigenvalues are all distinct. Furthermore, it holds that

$$\mathbf{p}^\star_\ell:=\Big(\frac{c_\ell b_{\mathbf{d}}}{d_\ell}\Big)\int_{-1}^{1}\mathbf{p}(t)G^\beta_\ell(t)\mathbf{w}_\beta(t)dt,$$

where G^β_ℓ denotes the *Gegenbauer* polynomial of degree ℓ defined for

$$\beta=\frac{\mathbf{d}-1}{2},\quad \mathbf{w}_\beta(x):=(1-x^2)^{\beta-1},\quad c_\ell:=\frac{2\ell+\mathbf{d}-2}{\mathbf{d}-2}\quad\text{and}\quad b_{\mathbf{d}}:=\frac{\Gamma(\frac{\mathbf{d}}{2})}{\Gamma(\frac12)\Gamma(\frac{\mathbf{d}}{2}-\frac12)},$$

with Γ the Gamma function, see [10, Chapter 2]. We recall that the Gegenbauer polynomials are orthogonal polynomials on the interval $[-1,1]$ with respect to the weight function $\mathbf{w}_\beta$. Besides, one can recover $\mathbf{p}\in\boldsymbol{L}^2([-1,1],\mathbf{w}_\beta)$ thanks to the identity

$$\mathbf{p}=\sum_{\ell\geq0}\Big[\sqrt{d_\ell}\mathbf{p}^\star_\ell\Big]\Big[\underbrace{G^\beta_\ell/\|G^\beta_\ell\|_{L^2([-1,1],\mathbf{w}_\beta)}}_{Z_\ell}\Big]=\sum_{\ell\geq0}\mathbf{p}^\star_\ell c_\ell G^\beta_\ell\,. \tag{10}$$

Remark 1. *Note that $\mathbf{p}^\star_\ell$ is the eigenvalue of the operator $\mathbb{T}_W$ associated to the eigenspace $\mathscr{H}_\ell$, $(\sqrt{d_\ell}\mathbf{p}^\star_\ell)_{\ell\geq0}$ are the coordinates of $\mathbf{p}\in\boldsymbol{L}^2([-1,1],\mathbf{w}_\beta)$ in the orthonormal basis $(Z_\ell)_{\ell\geq0}$, where $Z_\ell:=G^\beta_\ell/\|G^\beta_\ell\|_{L^2([-1,1],\mathbf{w}_\beta)}$. Note that requiring $W\in\boldsymbol{L}^2(\mathbb{S}^{\mathbf{d}-1}\times\mathbb{S}^{\mathbf{d}-1},\boldsymbol{\sigma}\otimes\boldsymbol{\sigma})$ is equivalent to $\mathbf{p}\in\boldsymbol{L}^2([-1,1],\mathbf{w}_\beta)$.*

Let $R \geq 0$ and define

$$\widetilde{R} := \sum_{\ell=0}^{R} d_\ell = \binom{R+\mathbf{d}-1}{R} + \binom{R+\mathbf{d}-2}{R-1}, \tag{11}$$

where the last equality is obtained with the telescoping sum using (8). Furthermore, we get that

$$\widetilde{R} \leq \frac{2(R+\mathbf{d}-1)^{\mathbf{d}-1}}{(\mathbf{d}-1)!} = \mathcal{O}(R^{\mathbf{d}-1}),$$

and this quantity is the dimension of Spherical Harmonics of degree less than R.

3.2. A Glimpse into Weighted Sobolev Spaces

Some of our result concern "*smooth graphons*" for which a regularity assumption is required. Following [27], we can define our approximation space defining the "*Weighted Sobolev*" space with the eigenvalues of the Laplacian on the Sphere. More precisely, let $s > 0$ a regularity parameter and $f \in \boldsymbol{L}^2((-1,1),\mathbf{w}_\beta)$ such that $f = \sum_{\ell\geq 0} f_\ell^\star c_\ell G_\ell^\beta$ in $\boldsymbol{L}^2$, we define

$$\|f\|^*_{Z^s_{\mathbf{w}_\beta}((-1,1))} = \left[\sum_{\ell=0}^{\infty} d_\ell |f_\ell^\star|^2 (1+(\ell(\ell+2\beta+1))^s)\right]^{\frac{1}{2}}$$

and

$$Z^s_{\mathbf{w}_\beta}((-1,1)) = \left\{f \in \boldsymbol{L}^2((-1,1),\mathbf{w}_\beta) \;:\; \|f\|^*_{Z^s_{\mathbf{w}_\beta}((-1,1))} < \infty\right\}.$$

Then, if $\mathbf{p}$ belongs to the Weighted Sobolev $Z^s_{\mathbf{w}_\beta}((-1,1))$ with smoothness $s > 0$, it holds

$$\sum_{\ell>R} d_\ell (\mathbf{p}_\ell^\star)^2 = \sum_{\ell>R} d_\ell (\mathbf{p}_\ell^\star)^2 \frac{1+(\ell(\ell+2\beta+1))^s}{1+(\ell(\ell+2\beta+1))^s} \leq C(\mathbf{p},s,\mathbf{d}) R^{-2s}, \tag{12}$$

where $C(\mathbf{p},s,\mathbf{d}) > 0$ is a constant that may depend on $\mathbf{p}$, s or $\mathbf{d}$.

3.3. Spectrum Consistency of the Matrix of Probabilities

Under this framework, Corollary 3 can be written as follows.

Proposition 4. *There exists a universal constant $C > 0$ such that for all $\alpha \in (0,1/3)$ and for all $n^3 \geq \widetilde{R}\log(2\widetilde{R}/\alpha)$, it holds*

$$\delta_2(\lambda(\boldsymbol{T}_n),\lambda^\star) \leq 2\Big[\sum_{\ell>R} d_\ell (\mathbf{p}_\ell^\star)^2\Big]^{\frac{1}{2}} + C\sqrt{\widetilde{R}(1+\log(\widetilde{R}/\alpha))/n}$$

with probability at least $1-3\alpha$. Moreover, if $\mathbf{p}$ belongs to the Weighted Sobolev space $Z^s_{\mathbf{w}_\beta}((-1,1))$, then for n large enough

$$\mathbb{E}[\delta_2^2(\lambda(\boldsymbol{T}_n),\lambda^\star)] \leq C' \left[\frac{n}{\log n}\right]^{-\frac{2s}{2s+(\mathbf{d}-1)}}$$

where C' only depends on s, $\mathbf{d}$ and $\|\mathbf{p}\|^*_{Z^s_{\mathbf{w}_\beta}((-1,1))}$.

A proof can be found in Appendix A.7. These theoretical results show that the eigenvalues of $\boldsymbol{T}_n$ converges towards the unknown spectrum $\lambda^\star$.

3.4. Nonparametric Estimation of the Kernel Spectrum

Let us now define our estimation procedure. Recall that we observe a graph and then its $n \times n$ adjacency matrix $\boldsymbol{A}$, where $\boldsymbol{A}_{ij}$ are independent Bernoulli random variables. Our model is that

$$\mathbb{P}\{\boldsymbol{A}_{ij} = 1\} = (\boldsymbol{\Theta}_0)_{ij} = W(X_i, X_j) = \mathbf{p}(\langle X_i, X_j \rangle), \qquad 1 \le i < j \le n,$$

where $X_1, \ldots, X_n$ are i.i.d. uniform variables on $\mathbb{S}^{\mathbf{d}-1}$. Our aim is to recover the envelope function $\mathbf{p}$ using only observations $\boldsymbol{A}$, the variables X_i being unobserved. The idea is to estimate the coefficients $\mathbf{p}^\star_\ell$ of $\mathbf{p}$ in the Gegenbauer polynomial basis, using that

$$\lambda^\star := \{\mathbf{p}^\star_0, \mathbf{p}^\star_1, \ldots, \mathbf{p}^\star_1, \ldots, \mathbf{p}^\star_\ell, \ldots, \mathbf{p}^\star_\ell, \ldots\}$$

is close to $\lambda(\boldsymbol{T}_n)$ and this latter is close to the spectrum

$$\lambda := \lambda(\widehat{\boldsymbol{T}}_n)$$

of our observable $\widehat{\boldsymbol{T}}_n = (1/n)\boldsymbol{A}$. Let us fix $R \ge 0$ some resolution level, and denote

$$\lambda^{\star R} := \big(\underbrace{\mathbf{p}^\star_0}_{d_0}, \underbrace{\mathbf{p}^\star_1, \ldots, \mathbf{p}^\star_1}_{d_1}, \ldots, \underbrace{\mathbf{p}^\star_R, \ldots, \mathbf{p}^\star_R}_{d_R}\big)$$

the first coefficients of $\mathbf{p}$, *i.e.,* the first eigenvalues of $\mathbb{T}_W$—not necessarily the largest. In view of (9) and defining $\widetilde{R}$ as in (11), we understand that the $\widetilde{R}$ first eigenvalues of $\mathbb{T}_W$ belong to the convex set

$$\mathscr{M}_R := \Big\{(\underbrace{u^\star_0}_{d_0}, \underbrace{u^\star_1, \ldots, u^\star_1}_{d_1}, \ldots, \underbrace{u^\star_R, \ldots, u^\star_R}_{d_R}) \in \mathbb{R}^{\widetilde{R}}\Big\}. \tag{13}$$

Remark 2. *Consider the convex set $\mathscr{M}_R^{[0,1]}$ of coefficients $(u^\star_0, u^\star_1, \ldots, u^\star_1, \ldots, u^\star_R, \ldots, u^\star_R)$ corresponding to a function between 0 and 1, namely*

$$\mathscr{M}_R^{[0,1]} := \Big\{(u^\star_0, u^\star_1, \ldots, u^\star_1, \ldots, u^\star_R, \ldots, u^\star_R) \in \mathbb{R}^{\widetilde{R}} \text{ s.t. there exists an extension } (u^\star_\ell)_{\ell > R}$$
$$\text{s.t. for a.e. } t \in [-1,1], \quad 0 \le \sum_{\ell=0}^{\infty} u^\star_\ell c_\ell G^\beta_\ell(t) \le 1\Big\}.$$

Note that $\lambda^{\star R} \in \mathscr{M}_R^{[0,1]}$ and that for all $x \in \mathscr{M}_R$

$$\delta_2(\mathscr{P}_{\mathscr{M}_R^{[0,1]}}(x), \lambda^{\star R}) \le \delta_2(x, \lambda^{\star R})$$

where $\mathscr{P}_{\mathscr{M}_R^{[0,1]}}$ denotes the $\mathbf{L}^2$-projection onto $\mathscr{M}_R^{[0,1]}$. It follows that all the results presented applies if we substitute $\mathscr{M}_R$ by $\mathscr{M}_R^{[0,1]}$. But, since we do not use the fact that the coefficients $(u^\star_0, u^\star_1, \ldots, u^\star_1, \ldots, u^\star_R, \ldots, u^\star_R)$ correspond to a function between 0 and 1 in our proofs and our numerical study, we choose to alleviate presentation using $\mathscr{M}_R$ instead of $\mathscr{M}_R^{[0,1]}$.

We assume that $n \ge \widetilde{R}$ and we denote $\mathfrak{S}_n$ the set of all permutation of $[n]$. We define the estimator $\widehat{\lambda}^R$ as the closest sequence to λ which belongs to the set of "*admissible*" spectra $\mathscr{M}_R$ as follows:

$$\widehat{\lambda}^R \in \operatorname*{arg\,min}_{u \in \mathscr{M}_R} \min_{\sigma \in \mathfrak{S}_n} \Big\{\sum_{k=1}^{\widetilde{R}} (u_k - \lambda_{\sigma(k)})^2 + \sum_{k=\widetilde{R}+1}^{n} \lambda^2_{\sigma(k)}\Big\}. \tag{14}$$

where we recall that λ denotes the spectrum of $\widehat{\boldsymbol{T}}_n$. We denote $\widehat{\mathbf{p}}^R_\ell$ the stage values of $\widehat{\lambda}^R$, such that

$$\widehat{\lambda}^R = (\widehat{\lambda}^R_1, \ldots, \widehat{\lambda}^R_{\widetilde{R}}) = (\widehat{\mathbf{p}}^R_0, \widehat{\mathbf{p}}^R_1, \ldots, \widehat{\mathbf{p}}^R_1, \ldots, \widehat{\mathbf{p}}^R_R, \ldots, \widehat{\mathbf{p}}^R_R).$$

One can check that

$$\widehat{\mathbf{p}}_\ell^R = \frac{1}{d_\ell}\sum_{k=\widetilde{\ell-1}}^{\widetilde{\ell}} \lambda_{\sigma(k)}$$

where σ (that depends on R) is a permutation achieving the minimum in (14) and we use the notation (11) with the convention $\widetilde{-1}=1$. Furthermore, the true complexity of this estimator is not $n!$ which matches the complexity of $\mathfrak{S}_n$. The true computation complexity of our estimator is at most $(R+2)!$ as shown by the next theorem.

Theorem 5 (Computational Complexity)**.** *Let $R\geq 0$ such that $\widetilde{R}\leq n$. For any sequence of real numbers $(\lambda_k)_{k=1}^n$ such that $\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_n$ it holds that*

$$\exists \mathfrak{H}_R\subseteq\mathfrak{S}_n \text{ s.t. } \forall u\in\mathscr{M}_R,\ \min_{\sigma\in\mathfrak{S}_n}\Big\{\sum_{k=1}^{\widetilde{R}}(u_k-\lambda_{\sigma(k)})^2+\sum_{k=\widetilde{R}+1}^{n}\lambda_{\sigma(k)}^2\Big\}=\min_{\sigma\in\mathfrak{H}_R}\Big\{\sum_{k=1}^{\widetilde{R}}(u_k-\lambda_{\sigma(k)})^2+\sum_{k=\widetilde{R}+1}^{n}\lambda_{\sigma(k)}^2\Big\}$$

where the set $\mathfrak{H}_R$ depends only on R and has size at most $(R+2)!$.

A proof can be found in Appendix B.1. This proof is constructive and it gives the expression of $\mathfrak{H}_R$.

Remark 3. *Remark that the hypothesis $\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_n$ is not necessary and can be removed. Indeed, if $\tau\in\mathfrak{S}_n$ a permutation such that $\lambda_{\tau(1)}\geq\lambda_{\tau(2)}\geq\ldots\geq\lambda_{\tau(n)}$ then it holds that*

$$\exists \mathfrak{H}_R\subseteq\mathfrak{S}_n \text{ s.t. } \forall u\in\mathscr{M}_R,\ \min_{\sigma\in\mathfrak{S}_n}\Big\{\sum_{k=1}^{\widetilde{R}}(u_k-\lambda_{\sigma(k)})^2+\sum_{k=\widetilde{R}+1}^{n}\lambda_{\sigma(k)}^2\Big\}=\min_{\sigma\in\mathfrak{H}_R}\Big\{\sum_{k=1}^{\widetilde{R}}(u_k-\lambda_{\sigma\circ\tau(k)})^2+\sum_{k=\widetilde{R}+1}^{n}\lambda_{\sigma\circ\tau(k)}^2\Big\}$$

where the set $\mathfrak{H}_R$ depends only on R and has size at most $(R+2)!$.

Remark 4. *Interestingly the computational complexity of our estimateur depends linearly on the sample size n which is important when observing large networks. However, it depends as $R\exp R$ in the complexity R of the model. Hence, it is relevant for large networks and low degree R kernels. However, if the experimenter knows that the eigenvalues are monotone (when sorting the eigenvalues so that the corresponding eigenspaces have increasing dimensions) then the complexity is linear in R.*

Using Proposition 1 and Theorem 4 we can prove that $\widehat{\lambda}^R$ is a relevant estimator of the true first eigenvalues $\lambda^{\star R}$ as shown in the next theorem.

Theorem 6. *There exists a universal constant $\kappa_0>0$ such that the following holds. For all $\alpha\in(0,1)$, if $n^3\geq(2\widetilde{R})^3\vee\widetilde{R}\log(2\widetilde{R}/\alpha)$, with probability greater that $1-3\alpha$, it holds*

$$\delta_2(\widehat{\lambda}^R,\lambda^{\star R})\leq 4\delta_2(\lambda^{\star R},\lambda^\star)+\kappa_0\sqrt{\widetilde{R}\big(1+\log\big(\widetilde{R}/\alpha\big)\big)/n}.$$

Moreover, there exists a universal constant $\kappa_1>0$ such that, if $n\geq 2\widetilde{R}$ then

$$\mathbb{E}[\delta_2^2(\widehat{\lambda}^R,\lambda^{\star R})]\leq\kappa_1\Big\{\delta_2^2(\lambda^{\star R},\lambda^\star)+\frac{\widetilde{R}\log n}{n}\Big\}.$$

A proof can be found in Appendix A.8.

To go further we need to analyze the behavior of the bias term $\delta_2(\lambda^{\star R},\lambda^\star)$ as a function of R under some regularity conditions on the envelope $\mathbf{p}$. Indeed we can write

$$\delta_2(\lambda^{\star R},\lambda^\star)^2=\sum_{k>\widetilde{R}}|\lambda_k^\star|^2=\sum_{\ell>R}d_\ell(\mathbf{p}_\ell^\star)^2.$$

Assume that $\mathbf{p}$ belongs to the weighted Sobolev space $Z^s_{\mathbf{w}_\beta}((-1,1))$ of regularity $s > 0$ defined in Section 3.2. Thus, since $\widetilde{R} = \mathcal{O}(R^{\mathbf{d}-1})$, using (12) and setting $R_{\text{opt}} = \lfloor (n/\log n)^{\frac{1}{2s+\mathbf{d}-1}} \rfloor$, we get

$$\mathbb{E}\left[\delta_2^2(\widehat{\lambda}^{R_{\text{opt}}}, \lambda^\star)\right] \le 2\delta_2^2(\lambda^{\star R_{\text{opt}}}, \lambda^\star) + 2\,\mathbb{E}\,\delta_2^2(\widehat{\lambda}^{R_{\text{opt}}}, \lambda^{\star R_{\text{opt}}}) \lesssim R_{\text{opt}}^{-2s} + \frac{\widetilde{R}_{\text{opt}} \log n}{n} \lesssim \left[\frac{n}{\log n}\right]^{-\frac{2s}{2s+(\mathbf{d}-1)}}.$$

Thus we recover a classical nonparametric rate of convergence for the estimation of a function with smoothness s in a space of dimension $\mathbf{d}-1$, see [17] for instance. We also face a classical issue of nonparametric statistics: *how to choose R, given that the best theoretical choice R_{opt} depends on the unknown smoothness s?* This is the point of the next section.

3.5. *Adaptation to the Smoothness of* $\mathbf{p}$

Let us define $\mathcal{R} = \{1, 2, \ldots, R_{\max}\}$ the possible values for R, with $2\widetilde{R}_{\max} \le n$. Following the Goldenshluger-Lepski method [15], set

$$B(R) := \max_{R' \in \mathcal{R}} \left\{ \delta_2(\widehat{\lambda}^{R'}, \widehat{\lambda}^{R' \wedge R}) - \kappa \sqrt{\frac{\widetilde{R}' \log n}{n}} \right\}, \tag{15}$$

where $\kappa > 0$ is a constant to be specified later. This function can be seen as an estimation of the (unknown) bias $\delta_2(\lambda^{\star R}, \lambda^\star)$. Then we define our final resolution level $\widehat{R}$ as a minimizer of an approximation of the risk as

$$\widehat{R} \in \arg\min_{R \in \mathcal{R}} \left\{ B(R) + \kappa \sqrt{\frac{\widetilde{R} \log n}{n}} \right\}. \tag{16}$$

The estimator of $\lambda^\star$ is then $\widehat{\lambda}^{\widehat{R}}$, which depends on the choice of constant κ in (15) and (16). The following results shows that this estimator is as good as the best one of the collection $(\widehat{\lambda}^R)_{R \in \mathcal{R}}$, up to a constant C, provided that κ is large enough.

Theorem 7. *Let $\widehat{\lambda}^{\widehat{R}}$ the estimator defined by* (14), (15) *and* (16). *There exist numerical constants $C > 0$ and $\kappa_0 > 0$* (*as in Theorem* 6) *such that , if $\kappa \ge \kappa_0 \sqrt{11}$, with probability $1 - 3n^{-8}$*

$$\delta_2(\widehat{\lambda}^{\widehat{R}}, \lambda^\star) \le C \min_{R \in \mathcal{R}} \left\{ \delta_2(\lambda^{\star R}, \lambda^\star) + \kappa \sqrt{\frac{\widetilde{R} \log n}{n}} \right\}.$$

Moreover, for $\kappa \ge \kappa_0 \sqrt{5}$, there exists a numerical constant $C' > 0$ such that

$$\mathbb{E}[\delta_2^2(\widehat{\lambda}^{\widehat{R}}, \lambda^\star)] \le C' \min_{R \in \mathcal{R}} \left\{ \delta_2^2(\lambda^{\star R}, \lambda^\star) + \kappa^2 \frac{\widetilde{R} \log n}{n} \right\}.$$

A proof can be found in Appendix A.10. Thus we choose $\kappa \ge \kappa_0 \sqrt{5}$ in (15) and (16), the practical choice of the tuning constant κ will be tackled in Section 5. Note also that the interesting choice of $\mathcal{R}$ is such that $R_{opt} \in \mathcal{R}$ which is the case for $cn^{\frac{1}{2s+\mathbf{d}-1}} \le R_{\max}$ where $c > 0$ is a constant. A more simple choice of $R_{\max}$ may be $cn \le 2\widetilde{R}_{\max} \le n$ where $0 < c < 1$ is a constant. In these cases, we get the following rate of convergence.

Corollary 8. *Assume that $\mathbf{p}$ belongs to the Weighted Sobolev space $Z^s_{\mathbf{w}_\beta}((-1,1))$. Then there exists a constant $C > 0$ depending only on $\|\mathbf{p}\|^*_{Z^s_{\mathbf{w}_\beta}((-1,1))}$, s and $\mathbf{d}$ such that*

$$\mathbb{E}[\delta_2^2(\widehat{\lambda}^{\widehat{R}}, \lambda^\star)] \le C \left[\frac{n}{\log n}\right]^{-\frac{2s}{2s+(\mathbf{d}-1)}}.$$

This means that the algorithm automatically adapts $\widehat{R}$ to the unknown smoothness s of $\mathbf{p}$: it chooses a small resolution level for smooth functions and a greater $\widehat{R}$ for irregular functions, that provides the best result in each case. Assuring that this is the optimal rate of convergence is beyond the scope of this paper but we believe that it cannot converge faster than $n^{-\frac{2s}{2s+(\mathbf{d}-1)}}$, since it is the usual rate for estimating a function in a $(\mathbf{d}-1)$-dimensional context.

The final step is to define the following estimator of envelope $\mathbf{p}$ by

$$\forall t \in [-1,1], \quad \widehat{\mathbf{p}}^{\widehat{R}}(t) := \sum_{\ell=0}^{\widehat{R}} \widehat{\mathbf{p}}_{\ell}^{\widehat{R}} c_\ell G_\ell^{\beta}(t). \tag{17}$$

3.6. Estimating the envelope function

Inferring from the estimation of $\lambda^\star$ to the estimation of $\mathbf{p}$, we face an identifiability problem. Indeed, consider for instance the case $\mathbf{d}=3$, which implies $\beta=1$, $d_\ell = 2\ell+1$, $c_\ell = 2\ell+1$. For $\mu>0$, let

$$\begin{aligned}
\mathbf{p}_a &= \frac{1}{2}c_0 G_0^\beta + \mu c_1 G_1^\beta + 0\times c_2 G_2^\beta + 0\times c_3 G_3^\beta + \mu c_4 G_4^\beta\,,\\
\mathbf{p}_b &= \frac{1}{2}c_0 G_0^\beta + 0\times c_1 G_1^\beta + \mu c_2 G_2^\beta + \mu c_3 G_3^\beta + 0\times c_4 G_4^\beta
\end{aligned}$$

Then the associated spectrum are

$$\begin{aligned}
\lambda_a^\star &= (1/2,\ \underbrace{\mu,\mu,\mu}_{3},\ \underbrace{0,0,0,0,0}_{5},\ \underbrace{0,0,0,0,0,0,0}_{7},\ \underbrace{\mu,\mu,\mu,\mu,\mu,\mu,\mu,\mu,\mu}_{9})\\
\lambda_b^\star &= (1/2,\ \underbrace{0,0,0}_{3},\ \underbrace{\mu,\mu,\mu,\mu,\mu}_{5},\ \underbrace{\mu,\mu,\mu,\mu,\mu,\mu,\mu}_{7},\ \underbrace{0,0,0,0,0,0,0,0,0}_{9})
\end{aligned}$$

which are indistinguishable in δ_2 metric, although $\|\mathbf{p}_a-\mathbf{p}_b\|_2 = \mu\sqrt{24}$. Furthermore, note that, for $\mu \le 1/24$, these functions have values in $[0,1]$.

Remark 5. *A natural question is then:* Can we recover the right eigenvalues labels from the empirical eigenvectors?
Under stronger requirements (RKHS-type assumptions), convergence of the eigenvectors of $\boldsymbol{A}/n$ towards the eigenfunctions of the integral operator $\mathbb{T}_W$ may be proved as in [33]. Essentially, it is possible to prove that the orthogonal projections Π_ℓ onto eigenspaces of $\boldsymbol{A}/n$ are closed in operator norm to the $n\times n$ matrix with entries $\sum_{j=1}^{d_\ell} Y_{\ell j}(X_i)Y_{\ell j}(X_j)$ given by the Zonal Harmonics. Unfortunately, this statistics depends on the latent points and suffers from the "agnostic" error as explained in [19]. While possible theoretically, it seems difficult in practice to use the information of the observed eigenvectors to uncover the right labels of the eigenvalues.

Nevertheless we can state a result in the case of a finite spectrum of distinct eigenvalues.

Proposition 9. *Assume that the envelope function $\mathbf{p}$ is polynomial of degree D, i.e., $\mathbf{p}_\ell^\star = 0$ for any $\ell > D$ and $\mathbf{p}_D^\star \neq 0$. Assume also that all nonzeros $\mathbf{p}_\ell^\star$ for $\ell \in \{0,\dots,D\}$ are distinct. If $R \ge D$ and n is large enough then*

$$\|\widehat{\mathbf{p}}^R - \mathbf{p}\|_2^2 \le 11\kappa_0^2 \frac{\widetilde{R}\log n}{n}\,,$$

with probability greater that $1-3n^{-8}$ where $\kappa_0>0$ is the constant defined in Theorem 6. Furthermore, it holds

$$\mathbb{E}[\|\widehat{\mathbf{p}}^R - \mathbf{p}\|_2^2] \le (18+4\kappa_0^2)\frac{\widetilde{R}\log n}{n}\,,$$

for n large enough.

A proof can be found in Appendix A.11. Note that we uncover (up to a log factor) the parametric rate of estimation. Let us now state what the adaptive procedure defined by (15) and (16) can do in this polynomial case.

Corollary 10. *Assume that the envelope function* $\mathbf{p}$ *is polynomial of degree* D*, i.e.,* $\mathbf{p}_\ell^\star = 0$ *for any* $\ell > D$ *and* $\mathbf{p}_D^\star \neq 0$*. Assume also that all nonzeros* $\mathbf{p}_\ell^\star$ *for* $\ell \in \{0, \dots, D\}$ *are distinct. If* $R_{\max} \geq D$*, there exists a numerical constant* C *such that, if* n *large enough, then* $\widehat{R} \geq D$ *a.s. and*

$$\mathbb{E}[\|\widehat{\mathbf{p}}^{\widehat{R}} - \mathbf{p}\|_2^2] \leq C\widetilde{D}\left(\frac{\log n}{n}\right).$$

A proof can be found in Appendix A.12. Here again, the parametric rate of estimation is attained by the adaptive procedure.

4. Extensions to Compact Symmetric Spaces

The aim of this section is to extend the previous result on spheres to numerous spaces such as compact Lie groups and compact symmetric spaces. A useful reference might be the books [36, 9] or the nice survey written in [25, Chapter 3] (see also [24] for a presentation of compact symmetric spaces) which has been useful to polish this section.

4.1. Harmonic Analysis on Compact Symmetric Spaces

In this section, we consider that $(\boldsymbol{S}, \gamma)$ is a compact Lie group with an invariant Riemannian metric γ, or more generally a compact symmetric space. The definitions will be given below when describing *Cartan's Classification* and, to be specific, this section focuses on (semi)simple connected compact Lie groups (sscc in short) and simple simply connected compact symmetric spaces (ssccss in short). These structures encompass spheres, projective spaces, Grassmannians, and orthogonal or unitary groups; and one can handle explicit eigenvectors computations in this framework.

Consider again that the graphon $W(g,h)$ depends only on (the cosine of) the distance $\gamma(g,h)$ (normalized so that the range of γ equals $[0,\pi]$) between points $g,h \in \boldsymbol{S}$ such that

$$W(g,h) = \mathbf{p}(\cos\gamma(g,h)) = \mathbf{p}(\cos\gamma(gh^{-1}, e_{\boldsymbol{S}})) =: p(gh^{-1})$$

where $e_{\boldsymbol{S}}$ denotes the identity element and $p(g) = \mathbf{p}(\cos\gamma(g, e_{\boldsymbol{S}}))$. Also we assume that $0 \leq W \leq 1$ since W defines a probability matrix. In particular, W is square-integrable on the compact $\boldsymbol{S} \times \boldsymbol{S}$. Observe that estimating W reduces to estimate $\mathbf{p}$ that reduces to estimate p and *vice versa*. By definition of the distance, note that

- When $\boldsymbol{S}$ is a sscc Lie group, the function p is invariant by conjugation, namely $p(hgh^{-1}) = p(g)$ for any latent points $g,h \in \boldsymbol{S}$. We denote by $\mathbf{L}^2(\boldsymbol{S})^{\boldsymbol{S}}$ the space of square-integrable functions p on $\boldsymbol{S}$ that are invariant by conjugation.
- When $\boldsymbol{S} = G/K$ is a ssccss, the function p is bi-K-invariant, namely $p(k_1 g k_2) = p(g)$ for any $k_1, k_2 \in K$ and $g \in G$. We denote by $\mathbf{L}^2(K \backslash G/K)$ the space of square-integrable functions on G that are bi-K-invariants.

In particular, Peter–Weyl's decomposition (presented below) gives an $\mathbf{L}^2$-decomposition of p in these settings. The measure on $\boldsymbol{S}$ is the Haar measure (normalized to be a probability measure), denoted $\mathrm{d}g$, standardly defined for any compact topological group $\boldsymbol{S}$. The harmonic analysis on $\boldsymbol{S}$ is based on the

Fourier transform of the space $\mathbf{L}^2(\boldsymbol{S}, \mathrm{d}g)$ of square integrable (complex valued) functions on $\boldsymbol{S}$. This space $\mathbf{L}^2(\boldsymbol{S}, \mathrm{d}g)$ is a Hilbert space for the scalar product

$$\langle f_1, f_2 \rangle = \int_S \overline{f_1}(g) f_2(g) \mathrm{d}g \,.$$

We define also the convolution product

$$(f_1 * f_2)(g) = \int_S f_1(gh^{-1}) f_2(h) \mathrm{d}h \,.$$

Now, recall that W defines a symmetric Hilbert-Schmidt operator $\mathbb{T}_W$ on $\mathbf{L}^2(\boldsymbol{S}, \mathrm{d}g)$ and the spectral theorem (6) gives

$$W(g,h) = \sum_{k \geq 1} \lambda_k^\star \phi_k(g) \phi_k(h) \,,$$

for an $\boldsymbol{L}^2(\boldsymbol{S}, \mathrm{d}g)$-orthonormal basis $(\phi_i)_{i \geq 1}$. Remark also that

$$(\mathbb{T}_W(f))(g_1) = \int_S W(g_1, g_2) f(g_2) \mathrm{d}g_2 = \int_S W(g_1 g_2^{-1}, e_S) f(g_2) \mathrm{d}g_2 = \int_S p(g_1 h^{-1}) f(h) \mathrm{d}h = (p * f)(g_1)$$

for all $f \in \boldsymbol{L}^2(\boldsymbol{S}, \mathrm{d}g)$. We deduce that $\mathbb{T}_W$ is the convolution on the left by p. We continue with a short reminder on harmonic analysis on compact groups and compact quotients.

Representation of Compact Groups and Irreducible Characters The first ingredient is representations of any compact group $\boldsymbol{S}$. It is defined by a finite dimensional complex vector space V and by a continuous morphism of groups $\rho : \boldsymbol{S} \to \mathrm{GL}(V)$ where $\mathrm{GL}(V)$ denotes the group of isomorphisms of V. A linear representation (V, ρ) is *irreducible* if one cannot find a subspace W such that $0 \subsetneq W \subsetneq V$ and that is $\boldsymbol{S}$-stable, *i.e.*, for all $w \in W$ and all $g \in \boldsymbol{S}$, one has $\rho(g)(w) \in W$. If V is a linear representation then one can always split it into irreducible components

$$V = \bigoplus_{\mathbf{r} \in \widehat{S}} m_{\mathbf{r}} V^{\mathbf{r}}$$

where $\widehat{\boldsymbol{S}}$ is the countable set of isomorphism classes of irreducible representations $\mathbf{r} = (\rho^{\mathbf{r}}, V^{\mathbf{r}})$ of $\boldsymbol{S}$ and $m_{\mathbf{r}} \geq 1$. Furthermore, we denote by

$$\mathrm{ch}^{\mathbf{r}}(g) = \mathrm{tr}(\rho^{\mathbf{r}}(g)) \,,$$

the irreducible characters associated to the irreducible representation $\mathbf{r} = (\rho^{\mathbf{r}}, V^{\mathbf{r}})$ of $\boldsymbol{S}$ where tr denotes the trace operator on $\mathrm{End}_{\mathbb{C}}(V^{\mathbf{r}})$ the set of (complex) endomorphisms of $V^{\mathbf{r}}$. In particular, since $\rho^{\mathbf{r}}(g)$ is unitary, it holds

$$\forall g \in \boldsymbol{S}, \quad |\mathrm{ch}^{\mathbf{r}}(g)| \leq d_{\mathbf{r}} = \mathrm{ch}^{\mathbf{r}}(e_S) \,,$$

where $d_{\mathbf{r}}$ is the dimension of $V^{\mathbf{r}}$. Also, note that

$$\mathrm{ch}^{\mathbf{r}} * \mathrm{ch}^{\mathbf{s}} = \frac{\delta_{\mathbf{rs}}}{d_{\mathbf{r}}} \mathrm{ch}^{\mathbf{r}} \,,$$

where $\delta_{\mathbf{rs}}$ denotes the Kroneker delta.

Peter–Weyl's Decomposition The Peter–Weyl's Decomposition shows that $(\mathrm{ch}^{\mathbf{r}})_{\mathbf{r} \in \widehat{S}}$ is an orthonormal basis of $\mathbf{L}^2(\boldsymbol{S})^S$. It follows that

$$p = \sum_{\mathbf{r} \in \widehat{S}} \langle p, \mathrm{ch}^{\mathbf{r}} \rangle \mathrm{ch}^{\mathbf{r}}$$

in $\mathbf{L}^2(\boldsymbol{S})^S$. Using that $\mathbb{T}_W$ is a left convolution operator by p, we find that $(\mathrm{ch}^{\mathbf{r}})_{\mathbf{r} \in \widehat{S}}$ is an eigenfunction basis of $\mathbb{T}_W$ associated to the eigenvalues $(\lambda_{\mathbf{r}}^\star)_{\mathbf{r} \in \widehat{S}}$ given by

$$\lambda_{\mathbf{r}}^\star = \frac{\langle p, \mathrm{ch}^{\mathbf{r}} \rangle}{d_{\mathbf{r}}} \,,$$

with multiplicity $d_{\mathbf{r}}^2 = \dim(\mathrm{End}_{\mathbb{C}}(V^{\mathbf{r}}))$.

Compact Gelfand Pairs and Zonal Spherical Functions There is an extension of this decomposition to quotients $S = G/K$ of a compact topological group G by a closed subgroup K. The most convenient setting for this extension is the one of compact Gelfand pairs defined as follows.

Definition (Gelfand Pair). *We say that (G,K) is a Gelfand pair if for any irreducible representation $V^{\mathbf{r}}$ of G, the space of K-fixed vectors*

$$V^{\mathbf{r},K} = \left\{ v \in V^{\mathbf{r}} \ :\ \forall k \in K\,,\ \rho^{\mathbf{r}}(k)(v) = v \right\}$$

has dimension at most one.

An irreducible representation $V^{\mathbf{r}}$ is called *spherical* if $\dim_{\mathbb{C}}(V^{\mathbf{r},K}) = 1$. We denote by $\widehat{G}^K$ the set of spherical representations of the Gelfand pair (G,K). If $\mathbf{r} \in \widehat{G}^K$ then we denote by $e^{\mathbf{r}}$ a unit vector vector in $V^{\mathbf{r},K}$ which is unique up to a multiplicative complex constant of modulus one. The *zonal spherical functions*

$$\mathrm{zon}^{\mathbf{r}}(g) = \sqrt{d_{\mathbf{r}}} \left\langle e^{\mathbf{r}}, \rho^{\mathbf{r}}(g)(e^{\mathbf{r}}) \right\rangle_{V^{\mathbf{r}}}$$

where $d_{\mathbf{r}}$ is the dimension of $V^{\mathbf{r}}$. In particular, since $\rho^{\mathbf{r}}(g)$ is unitary and $e^{\mathbf{r}}$ normalized, it holds

$$\forall g \in G, \quad |\mathrm{zon}^{\mathbf{r}}(g)| \leq \sqrt{d_{\mathbf{r}}} = \mathrm{zon}^{\mathbf{r}}(e_G)\,, \tag{18}$$

where e_G is the identity element of G. Also, note that

$$\mathrm{zon}^{\mathbf{r}} * \mathrm{zon}^{\mathbf{s}} = \frac{\delta_{\mathbf{rs}}}{\sqrt{d_{\mathbf{r}}}} \mathrm{zon}^{\mathbf{r}}\,.$$

Cartan's Extension of Peter–Weyl's Decomposition In the case of bi-K-invariant functions on G, an extension of Peter–Weyl's decomposition theorem shows that $(\mathrm{zon}^{\mathbf{r}})_{\mathbf{r} \in \widehat{G}^K}$ is an orthonormal basis of $\mathbf{L}^2(K \backslash G/K)$. It follows that

$$p = \sum_{\mathbf{r} \in \widehat{G}^K} \langle p, \mathrm{zon}^{\mathbf{r}} \rangle \mathrm{zon}^{\mathbf{r}}$$

in $\mathbf{L}^2(K \backslash G/K)$. Using that $\mathbb{T}_W$ is a left convolution operator by p, we find that $(\mathrm{zon}^{\mathbf{r}}))_{\mathbf{r} \in \widehat{G}^K}$ is an eigenfunction of $\mathbb{T}_W$ associated to the eigenvalue $(\lambda^\star_{\mathbf{r}})_{\mathbf{r} \in \widehat{G}^K}$ given by

$$\lambda^\star_{\mathbf{r}} = \frac{\langle p, \mathrm{zon}^{\mathbf{r}} \rangle}{\sqrt{d_{\mathbf{r}}}}\,.$$

with multiplicity $d_{\mathbf{r}} = \dim(V^{\mathbf{r}})$. The reader may recognize here the case of the sphere studied in the previous section.

Cartan's Classification of sscc Lie Groups and ssccss Now, a crucial question is how explicit are these decompositions. We begin with the notion of sscc Lie Groups that is based on Cartan's criterion for semisimplicity. It implies that a simply connected compact Lie group can always be written as a direct product of simple simply connected compact Lie group (in short sscc Lie group). Here, by simple we mean a Lie group S whose Lie algebra is simple, that is nonabelian and without non-trivial ideal. Interestingly, Cartan's classification of sscc Lie groups shows that any sscc Lie group fall into one of the following infinite families:

Group type

- Special unitary group $\mathrm{SU}(n+1)$,
- Odd spin group $\mathrm{Spin}(2n+1)$,
- Compact symplectic group $\mathrm{USp}(n)$,
- Even spin group $\mathrm{Spin}(2n)$,

or, it is one of the five exceptional compact Lie groups.

The sscc Lie groups belong to a larger class of compact Riemannian manifolds called symmetric spaces. Moreover, any simply connected compact symmetric space is isometric to a product of simple simply connected compact symmetric spaces (in short ssccss), which cannot be split further. A classification of all the ssccss S has been proposed by Cartan which shows that S is either of Group type (see above) or one of the following objects

non-Group type In this case, $\boldsymbol{S}$ falls into one of the following infinite families:

- Real Grassmannians $\mathrm{SO}(p+q)/(\mathrm{SO}(p)\times\mathrm{SO}(q))$,
- Complex Grassmannians $\mathrm{SU}(p+q)/(\mathrm{SU}(p)\times\mathrm{SU}(q))$,
- Quaternionic Grassmannians $\mathrm{USp}(p+q)/(\mathrm{USp}(p)\times\mathrm{USp}(q))$,
- Space of real structures on a complex space $\mathrm{SU}(n)/\mathrm{SO}(n)$,
- Space of quaternionic structures on an even complex space $\mathrm{SU}(2n)/\mathrm{USp}(n)$,
- Space of complex structures on a quaternionic space $\mathrm{USp}(n)/\mathrm{U}(n)$,
- Space of complex structures on an even real space $\mathrm{SO}(2n)/\mathrm{SU}(n)$,

or, it is one of the twelve exceptional sscc symmetric spaces.

Remark that, for all the ssccss examples, the eigenfunctions of the spectral decomposition of $\mathbb{T}_W$ do not depend on $\mathbb{T}_W$ and they are irreducible characters in the group case and zonal spherical functions in the non-group case.

Weyl's Highest Weight theorem and Cartan–Helgason's Extension Given a ssccss, we can make explicit the set $\widehat{\boldsymbol{S}}$ in the group case, and the set $\widehat{G}^K$ in the non-group case thanks to the Weyl's highest weight theorem and Cartan–Helgason's extension, see [25, Chapter 3] for a short and well written introduction. The highest weight theorem is completed by a formula for the irreducible character $\mathrm{ch}^{\mathbf{r}}$ of the module $V^{\mathbf{r}}$ with highest weight $\mathbf{r}$, see for instance [9, Chapter 22] and Weyl's integration formula [9, Chapter 17].
The same analysis can be lead in the non-group case. The only additional difficulty is the manipulation of zonal spherical functions. This issue will be handled by considering compact symmetric spaces of rank 1 in the following.

Now, we are ready to extend the previous results on the sphere to other latent spaces $\boldsymbol{S}$, namely the compact symmetric spaces of rank 1.

4.2. Compact Symmetric Spaces of Rank One

We focus here on the interesting case of compact symmetric spaces of rank one for which the zonal spherical functions can be explicitly computed. Indeed, one has the following classification of the compact symmetric spaces of rank one and of the corresponding spherical representations, see [25, Chapter 3] and [35] for instance. A compact symmetric space of rank one is ssccss that is 2-point homogeneous, namely

- **[Compact Symmetric Spaces of Rank One]** *Given two pairs of points (x_1,x_2) and (y_1,y_2) such that $\gamma(x_1,x_2)=\gamma(y_1,y_2)$, there is an isometry of $\boldsymbol{S}$ that maps x_1 (resp. x_2) onto y_1 (resp. y_2).*

The compact symmetric spaces of rank one are

- the real spheres $\mathbb{S}^{\mathbf{d}-1}=\mathrm{SO}(\mathbf{d})/\mathrm{SO}(\mathbf{d}-1)$,
- the real projective spaces $\mathbb{RP}^{\mathbf{d}-1}=\mathrm{SO}(\mathbf{d})/\mathrm{O}(\mathbf{d}-1)$,
- the complex projective spaces $\mathbb{CP}^{\mathbf{d}-1}=\mathrm{SU}(\mathbf{d})/\mathrm{U}(\mathbf{d}-1)$,
- the quaternionic projective spaces $\mathbb{HP}^{\mathbf{d}-1}=\mathrm{USp}(\mathbf{d})/(\mathrm{USp}(\mathbf{d}-1)\times\mathrm{USp}(1))$,
- or the octonionic projective plane $\mathbb{OP}^2=F_4/\mathrm{Spin}(9)$.

In the case of compact symmetric spaces of rank 1, one can explicitly described the spherical representations $\widehat{G}^K$. The dimension $d_\ell:=\dim_{\mathbb{C}}(V^{\ell\omega_0})$ of the ℓ-th spherical representation $V^{\ell\omega_0}$ are given in Table 1. One can even describe the zonal spherical functions of these spaces, and thus compute the eigenvalues $\mathbf{p}^\star_\ell$ (recall that their multiplicities d_ℓ are given by Table 1).

For compact symmetric spaces of rank one, on can define

- a probability density function $\mathbf{w}(t)$ on $[-1,1]$ defined as the density of the pushforward measure of the Haar measure by the map $x \mapsto t = \cos(\gamma(x,e))$,
- the pushforwards Z_ℓ on $[-1,1]$ of the zonal spherical functions, normalized so that they are an orthonormal basis of $\mathbf{L}^2([-1,1],\mathbf{w})$, the space of square-integrable functions with respect to the weight function $\mathbf{w}$ on $[-1,1]$.

$\boldsymbol{S}$	d_ℓ	$\widetilde{R} = \sum_{0\leq\ell\leq R} d_\ell$	$t = \cos(\gamma(x,e))$	Density $\mathbf{w}$
$\mathbb{S}^{\mathbf{d}-1}$	$\binom{\ell+\mathbf{d}-1}{\ell} - \binom{\ell+\mathbf{d}-3}{\ell-2}$	$\frac{2R+\mathbf{d}-1}{R+\mathbf{d}-1}\binom{R+\mathbf{d}-1}{\mathbf{d}-1}$	$x_{\mathbf{d}}$	$\mathrm{Beta}(\frac{\mathbf{d}-1}{2}, \frac{\mathbf{d}-1}{2})$
$\mathbb{RP}^{\mathbf{d}-1}$	$\frac{(6+\mathbf{d}^2+8\ell(2\ell-3)+\mathbf{d}(8\ell-5))\Gamma(\mathbf{d}+2\ell-3)}{\Gamma(\mathbf{d}-1)\Gamma(1+2\ell)}$	$\frac{4R+\mathbf{d}-1}{2R+\mathbf{d}-1}\binom{2R+\mathbf{d}-1}{\mathbf{d}-1}$	$\frac{x_{\mathbf{d}}^2-x_1^2-\cdots-x_{\mathbf{d}-1}^2}{x_1^2+\cdots+x_{\mathbf{d}}^2}$	$\mathrm{Beta}(\frac{\mathbf{d}-1}{2}, \frac{1}{2})$
$\mathbb{CP}^{\mathbf{d}-1}$	$\frac{2\ell+\mathbf{d}}{\mathbf{d}}\binom{\ell+\mathbf{d}-1}{\mathbf{d}-1}^2 - \frac{2\ell+\mathbf{d}-2}{\mathbf{d}}\binom{\ell+\mathbf{d}-2}{\mathbf{d}-1}^2$	$\frac{2R+\mathbf{d}}{\mathbf{d}}\binom{R+\mathbf{d}-1}{\mathbf{d}-1}^2$	$\frac{\lvert x_{\mathbf{d}}\rvert^2-\lvert x_1\rvert^2-\cdots-\lvert x_{\mathbf{d}-1}\rvert^2}{\lvert x_1\rvert^2+\cdots+\lvert x_{\mathbf{d}}\rvert^2}$	$\mathrm{Beta}(\mathbf{d}-1, 1)$
$\mathbb{HP}^{\mathbf{d}-1}$	$\frac{2(\ell+1)(\mathbf{d}(4\mathbf{d}^2-1)+2\mathbf{d}(4\mathbf{d}-1)\ell+(4\mathbf{d}-1)\ell^2)\Gamma(2\mathbf{d}+\ell-1)\Gamma(2\mathbf{d}+\ell)}{\Gamma(2\mathbf{d})\Gamma(2+2\mathbf{d})\Gamma(2+\ell)^2}$	$\frac{2R+2\mathbf{d}+1}{(2\mathbf{d}+1)(R+1)}\binom{R+2\mathbf{d}}{2\mathbf{d}}\binom{R+2\mathbf{d}-1}{2\mathbf{d}-1}$	$\frac{\lvert x_{\mathbf{d}}\rvert^2-\lvert x_1\rvert^2-\cdots-\lvert x_{\mathbf{d}-1}\rvert^2}{\lvert x_1\rvert^2+\cdots+\lvert x_{\mathbf{d}}\rvert^2}$	$\mathrm{Beta}(2\mathbf{d}-2, 2)$
$\mathbb{OP}^2$	$\frac{(4+\ell)(5+\ell)^2(6+\ell)}{924}$	$\frac{2R+11}{385}\binom{R+7}{4}\binom{R+10}{10}$	$\frac{\lvert x_3\rvert^2-\lvert x_1\rvert^2-\lvert x_2\rvert^2}{\lvert x_1\rvert^2+\lvert x_2\rvert^2+\lvert x_3\rvert^2}$	$\mathrm{Beta}(8, 4)$

TABLE 1
Review of the dimensions d_ℓ of the spherical representations, the distance $\cos(\gamma(x,e))$ to the identity e, the weight function $\mathbf{w}(t)$ of the compact symmetric spaces $\boldsymbol{S}$ of rank 1. These latter are respectively the density $\mathbf{w}$ and the orthonormal polynomials of the beta law on $[-1,1]$ with shape parameters (α,β), see (19).

In Table 1, one hase the following standard parameterizations of latent space $\boldsymbol{S}$:

- the real sphere $\mathbb{S}^{\mathbf{d}-1}$ is endowed with the coordinates $x = (x_1, \ldots, x_{\mathbf{d}})$ such that $\|x\|_2 = 1$ and the "*north pole*" is given by $e = (0, \ldots, 0, 1)$. We denote the "*weight function*" by $\mathbf{w}(x)$, it is the density of the push forward measure of the Haar measure by the map $x \mapsto \cos(\gamma(x,e))$ where we recall that $\gamma(x,e) = \arccos x_{\mathbf{d}}$.
- the projective space $\mathbb{FP}^{\mathbf{d}-1}$ (where $\mathbb{F} = \mathbb{R}, \mathbb{C}, \mathbb{H}$ or $\mathbb{O}$) is endowed with projective coordinates $[x_1 : x_2 : \cdots : x_{\mathbf{d}}]$ with the x_i's in $\mathbb{F}$, and the "north pole" is given by $e = [0 : \cdots : 0 : 1]$. We denote the "weight function" by $\mathbf{w}(x)$, it is the density of the push forward measure of the Haar measure by the map $x \mapsto \cos(\gamma(x,e))$ where we recall that $\gamma(x,e) = 2\arccos(|x_{\mathbf{d}}|/\|x\|_2)$.

One can show that the Jacobi polynomials (resp. beta distributions on $[-1,1]$) are the pushforward zonal spherical functions Z_ℓ (resp. the Haar measure) with shape parameters (α,β) depending on the base field and the dimension, see Table 1. In the case of real spheres, these Jacobi polynomials are the Legendre/Gegenbauer polynomials seen in Section 3. We recall that for shape parameters (α,β) the beta density distribution $\mathbf{w}$ is given by

$$\mathbf{w}(t) = \frac{\Gamma(\alpha+\beta)}{2^{\alpha+\beta-1}\Gamma(\alpha)\Gamma(\beta)}(1-t)^{\alpha-1}(1+t)^{\beta-1}\mathbb{1}_{[-1,1]}(t)\,, \tag{19}$$

where Γ is the Gamma function. In particular, recall that one has

$$\mathbf{p} = \sum_\ell \sqrt{d_\ell}\,\mathbf{p}_\ell^\star Z_\ell \quad \text{and} \quad \mathbf{p}_\ell^\star = \frac{1}{\sqrt{d_\ell}}\langle \mathbf{p}, Z_\ell\rangle_{\mathbf{L}^2([-1,1],\mathbf{w})}\,,$$

in $\mathbf{L}^2([-1,1],\mathbf{w})$. We further assume that there exists $s > 0$, a (Sobolev) regularity parameter, such that

$$\forall R \geq 1\,, \quad \sum_{\ell > R} d_\ell (\mathbf{p}_\ell^\star)^2 \leq C(\mathbf{p}, s, \boldsymbol{S}) R^{-2s}\,.$$

for some constant $C(\mathbf{p}, s, \boldsymbol{S}) > 0$ and for dimensions $(d_\ell)_{\ell\geq 0}$ that depends only on $\boldsymbol{S}$.

Now, recall the definition of the set of models $\mathcal{M}_R$ in (13) (the dimensions $(d_\ell)_{\ell\geq 0}$ are given by Table 1), of the estimator $\widehat{\lambda}^R$ in (14), of the adaptation $\widehat{R}$ in (16), of $\widehat{\mathbf{p}}^{\widehat{R}}$ in (17), and of $\boldsymbol{T}_n$ in (3). Our estimation procedure is the same as in the sphere example the only difference is that the dimensions $(d_\ell)_{\ell\geq 0}$, $\widetilde{R}$ and the zonal spherical function Z_ℓ depend on the latent space under consideration, see Table 1.

Theorem 11. *Let $\boldsymbol{S}$ be a compact symmetric space of rank one with Riemanian dimension $\mathbf{d}-1$. There exist constants $C_0, C_1, C_2, \kappa_0, \kappa_1 > 0$ such that the following holds. Let $\alpha \in (0,1/3)$ and $n, R \geq 0$ such that $n \geq 2\widetilde{R}$ and $n^3 \geq \widetilde{R}\log(2\widetilde{R}/\alpha)$ where $\widetilde{R}$ is given in Table 1. Then it holds,*

- **[Convergence of the matrix of probabilities]**

$$\delta_2(\lambda(\boldsymbol{T}_n), \lambda^\star) \leq 2\Big[\sum_{\ell>R} d_\ell(\mathbf{p}_\ell^\star)^2\Big]^{\frac{1}{2}} + C_0\sqrt{\widetilde{R}(1+\log(\widetilde{R}/\alpha))/n}$$

  with probability at least $1-3\alpha$ and

$$\mathbb{E}[\delta_2^2(\lambda(\boldsymbol{T}_n), \lambda^\star)] = \mathcal{O}\left[\left(\frac{n}{\log n}\right)^{-\frac{2s}{2s+(\mathbf{d}-1)}}\right].$$

- **[Convergence of the matrix of finite rank approximation]**

$$\delta_2(\widehat{\lambda}^R, \lambda^{\star R}) \leq 4\delta_2(\lambda^{\star R}, \lambda^\star) + \kappa_0\sqrt{\widetilde{R}\left(1+\log\left(\widetilde{R}/\alpha\right)\right)/n}.$$

  with probability at least $1-3\alpha$ and

$$\mathbb{E}[\delta_2^2(\widehat{\lambda}^R, \lambda^{\star R})] \leq \kappa_1\left\{\delta_2^2(\lambda^{\star R}, \lambda^\star) + \frac{\widetilde{R}\log n}{n}\right\}.$$

- **[Convergence of the adaptation]** *For $\kappa \geq \kappa_0\sqrt{11}$, it holds that*

$$\delta_2(\widehat{\lambda}^{\widehat{R}}, \lambda^\star) \leq C_1 \min_{R\in\mathcal{R}}\left\{\delta_2(\lambda^{\star R}, \lambda^\star) + \kappa\sqrt{\frac{\widetilde{R}\log n}{n}}\right\},$$

  with probability $1-3n^{-8}$. Furthermore, for $\kappa \geq \kappa_0\sqrt{5}$, it holds that

$$\mathbb{E}[\delta_2^2(\widehat{\lambda}^{\widehat{R}}, \lambda^\star)] \leq C_2 \min_{R\in\mathcal{R}}\left\{\delta_2^2(\lambda^{\star R}, \lambda^\star) + \kappa^2\frac{\widetilde{R}\log n}{n}\right\}.$$

A proof can be found in Appendix A.13. Note that the same results as in Proposition 9 and Corollary 10 hold when $\boldsymbol{S}$ is a compact symmetric space of rank one. Namely, adaptive estimation of the envelope function $\mathbf{p}$ is possible when $\mathbf{p}$ is a polynomial.

5. Numerical Experiments

5.1. Simulations

In this section we shall assess the performances of our estimation procedure by estimating numerous envelope functions $\mathbf{p}$. We consider the example of $\boldsymbol{S} = \mathbb{S}^2$, the unit sphere in dimension $\mathbf{d} = 3$. The functions G_ℓ^β turn to be the Legendre polynomials and the dimension of the space of spherical harmonics of degree ℓ is $d_\ell = 2\ell + 1$.

First, we shall explain how our algorithm works in practice to compute the adaptive estimator $\widehat{\mathbf{p}}^{\widehat{R}}$ of $\mathbf{p}$, see (14) and (17). For sake of clarity, we deal with a simple example. Suppose we are given an adjacency matrix $\boldsymbol{A}$ of size 20×20 and we set $R_{\max} = 1$. Thus $n = 20$, $d_0 = 1$ and $d_1 = 3$.

Step 1 Compute the 20 eigenvalues of A and sort them in decreasing order $\lambda_{(1)} \geq \cdots \geq \lambda_{(20)}$, see Figure 1.

Step 2 Take $0 \leq R \leq R_{\max}$. Generate $\mathfrak{S}_{R+2}$, the set of all permutation of $\{\mathbf{0}, d_0, \ldots, d_R\}$, the set with $R+2$ elements. The factor $+1$ in $R+2 = (R+1)+1$ is due to the “zeros” (represented by the symbol $\mathbf{0}$) to be placed, see Step 3 for a proper definition. For instance, for $R = 1$, we have

$$\mathfrak{S}_{R+1} = \begin{cases} \sigma_1 = [d_1, \quad d_0, \quad \mathbf{0}] \\ \sigma_2 = [d_1, \quad \mathbf{0}, \quad d_0] \\ \sigma_3 = [d_0, \quad d_1, \quad \mathbf{0}] \\ \sigma_4 = [d_0, \quad \mathbf{0}, \quad d_1] \\ \sigma_5 = [\mathbf{0}, \quad d_0, \quad d_1] \\ \sigma_6 = [\mathbf{0}, \quad d_1, \quad d_0] \end{cases}$$

Step 3 For each permutation $\sigma_{i, i\in\{1,\ldots,6\}}$ of $\mathfrak{S}_{R+2}$, compute the following $(\widetilde{\mathbf{p}}_{\sigma_i,\ell})_{\ell\in\{0,1,2\}}$ which are the “stage means” of the $\lambda_{(i), i\in\{1,\ldots,20\}}$'s according to the order of appearance of the d_ℓ's in the permutation σ_i. For instance, for $\sigma_1 = [d_1, \quad d_0, \quad \mathbf{0}]$ (see Figure 1), we get

$$\widetilde{\mathbf{p}}_{\sigma_1,2} = \frac{1}{3}\sum_{\ell=1}^{3} \lambda_{(\ell)}, \quad \widetilde{\mathbf{p}}_{\sigma_1,1} = \lambda_{(4)}, \quad \widetilde{\mathbf{p}}_{\sigma_1,0} = 0,$$

and for $\sigma_4 = [d_0, \quad \mathbf{0}, \quad d_1]$ one gets

$$\widetilde{\mathbf{p}}_{\sigma_2,1} = \lambda_{(1)}, \quad \widetilde{\mathbf{p}}_{\sigma_2,0} = 0, \quad \widetilde{\mathbf{p}}_{\sigma_2,2} = \frac{1}{3}\sum_{i=18}^{20} \lambda_{(\ell)}.$$

In Step 2, we have called “zeros” the fact that we always set $\widetilde{\mathbf{p}}_{\sigma_i,0} = 0$.

Step 4 For each permutation σ_i, compute the corresponding vector $\widetilde{\lambda}_{\sigma_i}$ of size 20, containing the $\widetilde{\mathbf{p}}_{\sigma_i,\ell}$ with multiplicity d_ℓ. Then compute the risk $\mathrm{Score}(\sigma_i)$ for each σ_i. For example for $\sigma_1 = [d_1, \quad d_0, \quad \mathbf{0}]$ (see Figure 1), one gets

$$\widetilde{\lambda}_{\sigma_1} = (\underbrace{\widetilde{\mathbf{p}}_{\sigma_1,2}, \widetilde{\mathbf{p}}_{\sigma_1,2}, \widetilde{\mathbf{p}}_{\sigma_1,2}}_{d_1=3}, \underbrace{\widetilde{\mathbf{p}}_{\sigma_1,1}}_{d_0=1}, \underbrace{0, \ldots, 0}_{n-d_0-d_1=16})$$

and its risk is $\mathrm{Score}(\sigma_1) = \sum_{\ell=1}^{20} (\lambda_{(\ell)} - \widetilde{\mathbf{p}}_{\sigma_1,\ell})^2$.

Step 5 Select the permutation $\sigma_{\min}$ such that $\sigma_{\min} = \arg\min_{\sigma_i} \mathrm{Score}(\sigma_i)$.

Step 6 Get the estimate $\widehat{\lambda}^R$ defined by

$$\widehat{\lambda}^R = (\underbrace{\widetilde{\mathbf{p}}_{\sigma_{\min},1}}_{d_0=1}, \underbrace{\widetilde{\mathbf{p}}_{\sigma_{\min},2}, \widetilde{\mathbf{p}}_{\sigma_{\min},2}, \widetilde{\mathbf{p}}_{\sigma_{\min},2}}_{d_1=3}) = (\widehat{\mathbf{p}}_0^R, \widehat{\mathbf{p}}_1^R, \widehat{\mathbf{p}}_1^R, \widehat{\mathbf{p}}_1^R),$$

see (14).

Step 7 Iterate Steps 2 to 6 for $R = 0$ to $R_{\max}$. Compute the level $\widehat{R}$ according to (16) and the adaptive estimator $\widehat{\mathbf{p}}^{\widehat{R}}(t)$ according to (17).

Step 8 Troncate $\widehat{\mathbf{p}}^{\widehat{R}}(t)$ so as to it belongs to $[0, 1]$.

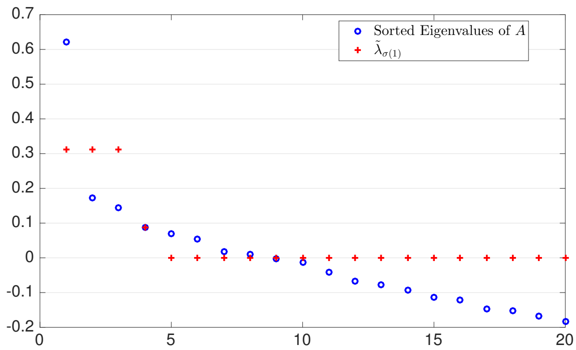

FIGURE 1. *Plot of the 20 sorted eigenvalues $\lambda_{(i)}$ of adjacency matrix $\mathbf{A}$ and the values of vector $\widetilde{\lambda}_{\sigma(1)}$.*

Of course, the choice of level R is crucial and the estimation is sensitive to R. That is why we use our selection method, as described in Section 3.5 (see Step 7 in the description of the algorithm above). As almost all estimators selection methods, this Goldenshluger-Lepski method uses an hyper-parameter κ. Our theoretical result ensures a good performance as soon as κ is large enough, but it is well known that a more precise choice is better in practice. Heuristics exist to calibrate κ, but they are all based on the behavior of the estimator for very large R (see for instance [3]). Hence these techniques are not possible here, due to the computational cost of the estimation when R is large (we can hardly consider larger than $R_{\max} = 7$ because of the complexity in $(R_{\max} + 2)!$). Fortunately, the stability of the estimation allows us to choose here a fixed κ, namely $\kappa = 0.25$, and this choice ensures good selection of $\hat{R}$ for a wide range of functions $\mathbf{p}$.

Now, let us deal with the estimation of the six following envelope functions $\mathbf{p}$

$$\begin{aligned}
\mathbf{p}_1(t) &= \Big(\frac{1+t}{2}\Big)^4, \\
\mathbf{p}_2(t) &= \mathbb{1}_{t>0.7}, \\
\mathbf{p}_3(t) &= e^{-(t-1)^2}, \\
\mathbf{p}_4(t) &= 0.5 + 0.5\sin\Big(\pi t/2\Big), \\
\mathbf{p}_5(t) &= \frac{1}{3} + \frac{1}{12}\Big(35t^4 - 30t^2 + 3\Big), \\
\mathbf{p}_6(t) &= t^{10}\mathbb{1}_{t>0}.
\end{aligned}$$

We consider graphs of size $n = 5000$. We set $R_{\max} = 4$ and $\kappa = 0.25$ for the adaptive selection rule of R, see (16).

Figure 2 presents our simulation results. For each envelope function $\mathbf{p}$, we represent on the top side, the estimated coefficients $\widehat{\mathbf{p}}_\ell^{\hat{R}}$ and the true coefficients $\mathbf{p}_\ell^\star$ with their multiplicity $2\ell + 1$. On the bottom side, we represent the estimated envelope function $\widehat{\mathbf{p}}$ and the true $\mathbf{p}$. Note that our procedure is not constrained by dealing with envelope functions $\mathbf{p}$ defining positive kernels W. Such an example is given by the step function $\mathbf{p}_2$ as its Fourier coefficients $\mathbf{p}_{2,\ell}^\star$'s can be negative, see Figure 2.

The estimation of all functions are good except for the step function $\mathbf{p}_2$ which is more demanding due to its discontinuity. Despite that function $\mathbf{p}_6$ is not easy to be estimated because of its flatness, our estimation is satisfying. Furthermore, it is interesting to remark that except for $\mathbf{p}_2$, the estimated coefficients are very close to the true ones.

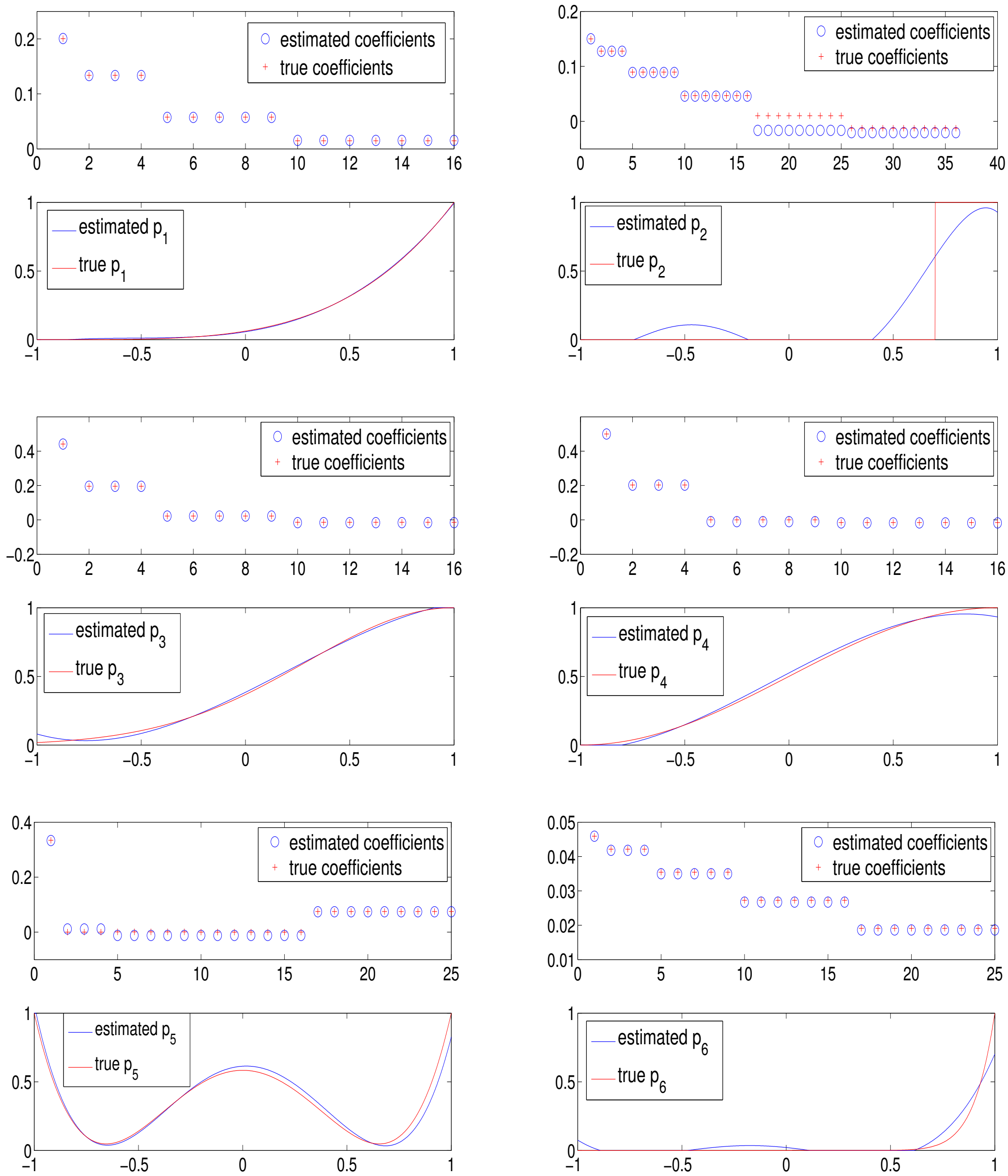

FIGURE 2. *Estimation of envelope functions* $\mathbf{p}_1, \ldots, \mathbf{p}_6$.

5.2. Real data: Grévy's zebras in Kenya

To depict our estimator on real datas, we use the dataset provided by the Koblenz Network Collection (http://konect.uni-koblenz.de/). This dataset is presented as follows: "*This undirected network contains interactions between 27 Grévy's zebras in Kenya. A node represents a zebra and an edge between two zebras shows that there was an interaction between them during the study*". For further information about the study, the reader may consult [30]. We display the undirected graph structure of this dataset on Figure 3.

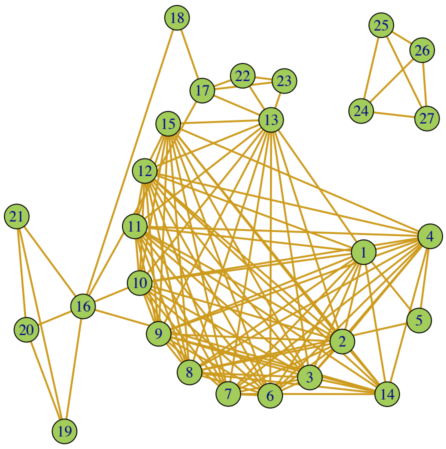

FIGURE 3. *Layout of the zebras' network*

We aim at estimating the envelop function $\mathbf{p}$. We compute the adjacency matrix A and its eigenvalues (see Figure 4). Considering that n is small, we may discard large $\mathbf{d}$ values (high dimensional spheres). Furthermore, the spectrum of A displays two clear stages of size 1 and 3, hence we set $\mathbf{d} = 3$ (since $d_1 = \mathbf{d}$ where d_1 is the multiplicity of the eigenvalue λ_1). Next we implement our adaptive estimation procedure from **Step 1** until **Step 8** described in Section 5.1 with $R_{\max} = 4$ and $\kappa = 0.25$. The procedure selects $\hat{R} = 2$.

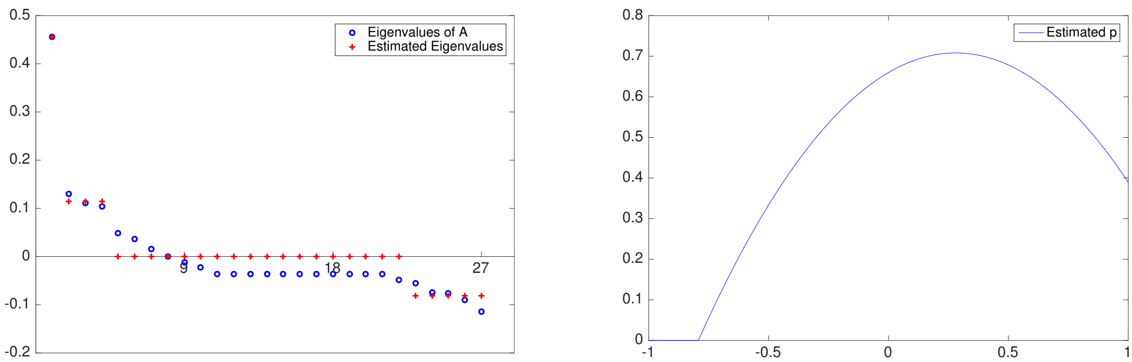

FIGURE 4. *Left panel: empirical (blue dots) and estimated (red crosses) spectra. Right panel: estimated envelop function.*

Figure 4 shows that the estimated spectrum fits in a satisfying way the empirical one. It gives also the estimated envelope function $\widehat{\mathbf{p}}^{\hat{R}}$. For this study, we may conclude that the zebras tend to interact with others which are moderately close to them but when they become too close, they less interact.

Acknowledgements: The authors would like to thank Pierre Loïc Méliot for many useful discussions on compact symmetric spaces.

Appendix A: Proofs

A.1. Proof of Proposition 1

This result is a consequence of [2, Corollary 3.12] and [2, Remark 3.19] with $X_{ij} = \boldsymbol{A}_{ij} - (\boldsymbol{\Theta}_0)_{ij}$ a centered but not symmetric random variable, $\varepsilon = 1/2$ say, $\widetilde{\sigma}^2 = \boldsymbol{D}_0$ by definition, and observing that $\widetilde{\sigma}_*^2 = \max_{ij}((\boldsymbol{\Theta}_0)_{ij} \vee (1-(\boldsymbol{\Theta}_0)_{ij})) \leq 1$. It gives

$$\forall t > 0, \quad \mathbb{P}\left\{\|\boldsymbol{A}-\boldsymbol{\Theta}_0\| \geq 3\sqrt{2\boldsymbol{D}_0} + Ct\right\} \leq n\exp(-t^2),$$

for some universal constant $C > 0$.

A.2. Proof of Theorem 2

Let $R \geq 1$ and define

$$\begin{aligned}
\Phi_i^n &:= (1/\sqrt{n})(\phi_i(X_1), \ldots, \phi_i(X_n)) \in \mathbb{R}^n, \\
K_R &:= \mathrm{Diag}(\lambda_1(\mathbb{T}_W) \ldots, \lambda_R(\mathbb{T}_W)) \in \mathbb{R}^{R\times R}, \\
E_{R,n} &:= (\langle \Phi_i^n, \Phi_j^n \rangle - \delta_{ij})_{i,j\in[R]} \in \mathbb{R}^{R\times R}, \\
X_{R,n} &:= \left[\Phi_1^n \cdots \Phi_R^n\right] \in \mathbb{R}^{n\times R}, \\
A_{R,n} &:= \left(X_{R,n}^\top X_{R,n}\right)^{\frac12} \in \mathbb{R}^{R\times R} \text{ and note that } A_{R,n}^2 = \mathrm{Id}_R + E_{R,n}, \\
\boldsymbol{T}_{R,n} &:= \sum_{r=1}^R \lambda_r(\mathbb{T}_W)\Phi_r^n(\Phi_r^n)^\top = X_{R,n}K_R X_{R,n}^\top \in \mathbb{R}^{n\times n}, \\
\widetilde{\boldsymbol{T}}_{R,n} &:= ((1-\delta_{ij})\boldsymbol{T}_{R,n})_{i,j\in[n]} \in \mathbb{R}^{n\times n}, \\
\boldsymbol{T}_{R,n}^\star &:= A_{R,n}K_R A_{R,n} \in \mathbb{R}^{R\times R}, \\
\text{and } W_R(x,y) &:= \sum_{i=1}^R \lambda_i(\mathbb{T}_W)\phi_i(x)\phi_i(y),
\end{aligned}$$

where the last identity holds point-wise. Observe that $A_{R,n}^2 = \mathrm{Id}_R + E_{R,n}$. It holds

$$\delta_2(\lambda(\mathbb{T}_W), \lambda(\mathbb{T}_{W_R})) = \Big(\sum_{r>R} \lambda_r^2(\mathbb{T}_W)\Big)^{\frac12}. \tag{20}$$

Note the equalities between spectra $\lambda(\mathbb{T}_{W_R}) = \lambda(K_R)$ and $\lambda(\boldsymbol{T}_{R,n}) = \lambda(\boldsymbol{T}_{R,n}^\star)$ where the last one follows by using a SVD of $X_{R,n}$. Hence, we deduce that

$$\delta_2(\lambda(\mathbb{T}_{W_R}), \lambda(\boldsymbol{T}_{R,n})) = \delta_2(\lambda(K_R), \lambda(\boldsymbol{T}_{R,n}^\star)) \leq \|\boldsymbol{T}_{R,n}^\star - K_R\|_F = \|A_{R,n}K_R A_{R,n} - K_R\|_F\,,$$

by Hoffman-Wielandt inequality, see [21, page 118] for instance. Equation (4.8) at [21, page 127] gives that

$$\delta_2(\lambda(\mathbb{T}_{W_R}), \lambda(\boldsymbol{T}_{R,n})) \leq \sqrt{2}\|K_R\|_F\|E_{R,n}\| = \sqrt{2}\|W_R\|_2\|E_{R,n}\|\,, \tag{21}$$

Actually, one can remove the constant $\sqrt{2}$ using Ostrowski's theorem, see [7, Theorem A.2] for instance. Also, by Hoffman-Wielandt inequality, we have

$$\delta_2(\lambda(\boldsymbol{T}_{R,n}), \lambda(\widetilde{\boldsymbol{T}}_{R,n})) \leq \|\widetilde{\boldsymbol{T}}_{R,n} - \boldsymbol{T}_{R,n}\|_F = \Big[\frac{1}{n^2}\sum_{i=1}^n W_R^2(X_i,X_i)\Big]^{\frac12}, \tag{22}$$

and

$$\delta_2(\lambda(\widetilde{\boldsymbol{T}}_{R,n}), \lambda(\boldsymbol{T}_n)) \leq \|\widetilde{\boldsymbol{T}}_{R,n} - \boldsymbol{T}_n\|_F = \Big[\frac{1}{n^2}\sum_{i\neq j}(W-W_R)^2(X_i,X_j)\Big]^{\frac12}. \tag{23}$$

Invoke Lemma 12 to bound (21), Lemma 13 to bound (22) and Lemma 14 to bound (23).

Lemma 12. *Let $R \geq 1$ and denote by $\rho(R) := \max(1, \|\sum_{r=1}^{R} \phi_r^2\|_\infty - 1)$ then it holds*

$$\forall t > 0, \quad \mathbb{P}\left\{\|E_{R,n}\| \geq t\right\} \leq 2R \exp\Big[-\frac{n}{2\rho(R)}\frac{t^2}{1+t/(3n)}\Big].$$

In particular, for all $\alpha \in (0,1)$ and for $n^3 \geq \rho(R)\log(2R/\alpha)$, it holds

$$\mathbb{P}\left\{\|E_{R,n}\| \geq \sqrt{\frac{\rho(R)\log(2R/\alpha)}{n}}\right\} \leq \alpha\,.$$

Lemma 13. *Let $R \geq 1$ and $\alpha \in (0,1)$ then, with probability at least $1-\alpha$, it holds*

$$\frac{1}{n^2}\sum_{i=1}^{n} W_R^2(X_i, X_i) \leq \left[1 + \max_{1\leq r\leq R}\|\phi_r^2\|_\infty \sqrt{\frac{\log(R/\alpha)}{2n}}\right]\frac{2\rho(R)\|W_R\|^2}{n}\,.$$

Lemma 14. *It holds, for all $\alpha \in (0,1)$,*

$$\mathbb{P}\left\{\frac{1}{n(n-1)}\sum_{i\neq j}(W-W_R)^2(X_i,X_j) \geq \sum_{r>R}\lambda_r^2(\mathbb{T}_W) + \|W-W_R\|_\infty^2\sqrt{\frac{\log(2/\alpha)}{n-1}}\right\} \leq \alpha\,.$$

These lemmas are proven in Appendix A.3, Appendix A.4 and Appendix A.5. Collecting (20), (21), (22) and (23), the triangular inequality gives the result.

A.3. *Proof of Lemma 12*

Observe that $nE_{R,n} = \sum_{i=1}^{n}(Z_R(X_i)Z_R^\top(X_i) - \mathrm{Id}_R)$ is a sum of independent centered symmetric matrices where we denote by $Z_R(x) := (\phi_1(x), \ldots, \phi_R(x))$. In particular, $Z_R(X_i)Z_R^\top(X_i)$ are rank one matrices so that it holds

$$\begin{aligned}
\|Z_R(X_i)Z_R^\top(X_i) - \mathrm{Id}_R\| &= 1 \vee \big(\|Z_R(X_i)\|_2^2 - 1\big) \\
&= 1 \vee \big((\sum_{r=1}^{R}\phi_r^2(X_i)) - 1\big) \\
&\leq 1 \vee \big(\|\sum_{r=1}^{R}\phi_r^2\|_\infty - 1\big) =: \rho(R)\,.
\end{aligned}$$

Moreover, one has

$$\begin{aligned}
\sigma_{R,n}^2 :&= n\|\mathbb{E}\big((Z_R(X_1)Z_R^\top(X_1) - \mathrm{Id}_R)^2\big)\| \\
&= n\|\mathbb{E}\big(\|Z_R(X_1)\|_2^2 Z_R(X_1)Z_R^\top(X_1) - 2Z_R(X_1)Z_R^\top(X_1) - \mathrm{Id}_R\big)\| \\
&= n\|\mathbb{E}\big(\|Z_R(X_1)\|_2^2 Z_R(X_1)Z_R^\top(X_1)\big) - \mathrm{Id}_R\| \\
&\leq n\Big\|\|\sum_{r=1}^{R}\phi_r^2\|_\infty \mathbb{E}\big(Z_R(X_1)Z_R^\top(X_1)\big) - \mathrm{Id}_R\Big\| \\
&= n\Big\|\|\sum_{r=1}^{R}\phi_r^2\|_\infty \mathrm{Id}_R - \mathrm{Id}_R\Big\| \\
&= n\Big|\|\sum_{r=1}^{R}\phi_r^2\|_\infty - 1\Big| \leq n\rho(R)
\end{aligned}$$

where we invoke that a.s. $\|Z_R(X_1)\|_2^2 Z_R(X_1)Z_R^\top(X_1) \preccurlyeq (\|\sum_{r=1}^R \phi_r^2\|_\infty) Z_R(X_1)Z_R^\top(X_1)$. It follows from the matrix Bernstein inequality given in [34, Theorem 6.1.1]

$$\forall t>0, \quad \mathbb{P}\left\{\|E_{R,n}\| \geq t\right\} \leq 2R\exp\Big[-\frac{n}{2\rho(R)}\frac{t^2}{1+t/(3n)}\Big].$$

Indeed, we have used [34, Theorem 6.1.1] with

$$\begin{aligned} X_k &\leftarrow Z_R(X_i)Z_R^\top(X_i) - \mathrm{Id}_R, \\ R &\leftarrow \rho(R), \\ Y &\leftarrow nE_{R,n}, \\ \sigma^2 \leq n\|\mathbb{E}\left(X_1^2\right)\| &\leftarrow \sigma_{R,n}^2, \\ t &\leftarrow nt, \end{aligned}$$

according to the notation of [34] on the left hand side and our notation on the right hand side. It proves the lemma.

A.4. Proof of Lemma 13

Observe that

$$\begin{aligned} \frac{1}{n^2}\sum_{i=1}^n W_R^2(X_i,X_i) &= \frac{1}{n^2}\sum_{i=1}^n \Big(\sum_{r=1}^R \lambda_r(\mathbb{T}_W)\phi_r^2(X_i)\Big)^2 \\ &= \frac{1}{n^2}\sum_{r,s\in[R]} \lambda_r(\mathbb{T}_W)\lambda_s(\mathbb{T}_W)\Big(\sum_{i=1}^n \phi_r^2(X_i)\phi_s^2(X_i)\Big) \\ &= x^\top \boldsymbol{A} x \leq \|\boldsymbol{A}\|\|x\|_2^2 \end{aligned}$$

with $x = (\lambda_1(\mathbb{T}_W)/\sqrt{n}, \ldots, \lambda_R(\mathbb{T}_W)/\sqrt{n})$ and $\boldsymbol{A} = ((1/n)\sum_{i=1}^n \phi_r^2(X_i)\phi_s^2(X_i))_{r,s}$. Note that $\boldsymbol{A}$ is an irreducible and aperiodic matrix since its coefficients are positive. It follows by Perron-Frobenius theorem that

$$\|\boldsymbol{A}\| \leq \frac{1}{n}\max_{1\leq r\leq R}\Big(\sum_{s=1}^R\sum_{i=1}^n \phi_r^2(X_i)\phi_s^2(X_i)\Big)$$

Now, this last quantity can be upper bounded as follows

$$\begin{aligned} \frac{1}{n}\sum_{s=1}^R\sum_{i=1}^n \phi_r^2(X_i)\phi_s^2(X_i) &= \frac{1}{n}\sum_{i=1}^n \phi_r^2(X_i)\Big(\sum_{s=1}^R \phi_s^2(X_i)\Big), \\ &\leq \Big(\frac{1}{n}\sum_{i=1}^n \phi_r^2(X_i)\Big)(1+\rho(R)). \end{aligned}$$

Using the bound

$$\phi_r^2(X_1) \leq \max_{1\leq r\leq R}\|\phi_r^2\|_\infty =: a_R\,,$$

and Hoeffding inequality [6, page 34], we deduce that

$$\forall t>0, \quad \mathbb{P}\left\{\frac{1}{n}\sum_{i=1}^n \phi_r^2(X_i) > \mathbb{E}\left(\phi_r^2(X_1)\right) + t\right\} \leq \exp\Big(-\frac{2nt^2}{a_R^2}\Big).$$

Observe that $\mathbb{E}\left(\phi_r^2(X_1)\right) = 1$. Let $\alpha \in (0,1)$, choosing $t^2 = a_R^2\log(R/\alpha)/(2n)$ and taking an union bound, it holds that

$$\mathbb{P}\left\{\forall r\in[R], \quad \frac{1}{n}\sum_{i=1}^n \phi_r^2(X_i) \leq 1 + \frac{a_R\log^{\frac{1}{2}}(R/\alpha)}{\sqrt{2n}}\right\} \geq 1-\alpha$$

It results in

$$\mathbb{P}\left\{\|\boldsymbol{A}\| \leq \Big(1+\frac{(1+\rho(R))\log^{\frac{1}{2}}(R/\alpha)}{\sqrt{2n}}\Big)(1+\rho(R))\right\} \geq 1-\alpha$$

On this event, we deduce that

$$\begin{aligned}\frac{1}{n^2}\sum_{i=1}^{n} W_R^2(X_i,X_i) &\leq \|\boldsymbol{A}\|\|x\|_2^2,\\ &\leq \Big(1+\frac{a_R\log^{\frac{1}{2}}(R/\alpha)}{\sqrt{2n}}\Big)\frac{(1+\rho(R))\|W_R\|^2}{n},\end{aligned}$$

which gives the result.

A.5. Proof of Lemma 14

By a standard inequality of Hoeffding [18], for a bounded kernel h, for all $\alpha \in (0,1)$,

$$\mathbb{P}\left\{\Big|\frac{1}{n(n-1)}\sum_{i\neq j} h(X_i,X_j) - \mathbb{E}\left(h(X_1,X_2)\right)\Big| > \|h\|_\infty\sqrt{\frac{\log(2/\alpha)}{n-1}}\right\} \leq \alpha$$

Applying this result for $h=(W-W_R)^2$ and noticing that

- $\mathbb{E}\left(h(X_1,X_2)\right) = \|W-W_R\|_2^2 = \sum_{r>R}\lambda_r^2(\mathbb{T}_W)$,
- $\|h\|_\infty = \|W-W_R\|_\infty^2$,

the result follows.

A.6. Proof of Corollary 3

The symmetric kernel $h := (W-W_R)^2 - \mathbb{E}\left((W-W_R)^2\right)$ is σ-canonical, see [11, Definition 3.5.1] for a definition. The following important improvement of Hoeffding's inequalities for canonical kernels was proved by [1], it holds that there exists two universal constants $C_1>0$ and $C_2>0$ such that for all $\alpha \in (0,1)$,

$$\mathbb{P}\left\{\Big|\frac{1}{n(n-1)}\sum_{i\neq j} h(X_i,X_j)\Big| > C_1\|h\|_\infty\frac{\log(C_2/\alpha)}{n}\right\} \leq \alpha\,.$$

We deduce that it holds, for all $\alpha \in (0,1)$,

$$\mathbb{P}\left\{\frac{1}{n(n-1)}\sum_{i\neq j}(W-W_R)^2(X_i,X_j) \geq \|W-W_R\|_2^2 + C_1\|W-W_R\|_\infty^2\frac{\log(C_2/\alpha)}{n}\right\} \leq \alpha\,,$$

which proves the corollary substituting Lemma 14 by the aforementioned inequality.

A.7. Proof of Proposition 4

Define

$$\forall t\in[-1,1],\quad \mathbf{p}^R(t) := \sum_{\ell=0}^{R}\mathbf{p}_\ell^\star c_\ell G_\ell^\beta(t)\,.$$

We apply Corollary 3 to the kernel as follows.

$$\forall x,y\in\mathbb{S}^{\mathbf{d}-1},\quad W_{\widetilde{R}}(x,y):=\sum_{\ell=0}^{R}\mathbf{p}_\ell^\star c_\ell G_\ell^\beta(\langle x,y\rangle)=\mathbf{p}^R(\langle x,y\rangle),$$

First, note that

$$\|W-W_{\widetilde{R}}\|_2=\|\mathbf{p}-\mathbf{p}^R\|_2=\Big[\sum_{\ell>R}d_\ell(\mathbf{p}_\ell^\star)^2\Big]^{\frac{1}{2}}. \tag{24}$$

Next, invoke [10, Corollary 1.2.7] to get that

$$\forall\ell\geq 0,\quad \sum_{j=1}^{d_\ell}Y_{\ell j}^2=d_\ell\,.$$

It follows that the quantity $\rho(\widetilde{R})$ of Theorem 2 simplifies to $\rho(\widetilde{R})\leq\widetilde{R}$. Furthermore, it holds

$$\forall x\in\mathbb{S}^{\mathbf{d}-1},\quad W_{\widetilde{R}}(x,x)=\sum_{\ell=0}^{R}\mathbf{p}_\ell^\star c_\ell G_\ell^\beta(1)=\sum_{\ell=0}^{R}d_\ell\mathbf{p}_\ell^\star\,, \tag{25}$$

since $G_\ell^\lambda(1)=d_\ell/c_\ell$. Then by Hoffman-Wielandt inequality, we have

$$\delta_2(\lambda(\boldsymbol{T}_{\widetilde{R},n}),\lambda(\widetilde{\boldsymbol{T}}_{\widetilde{R},n}))\leq\|\widetilde{\boldsymbol{T}}_{\widetilde{R},n}-\boldsymbol{T}_{\widetilde{R},n}\|_F=\Big[\frac{1}{n^2}\sum_{i=1}^{n}W_{\widetilde{R}}^2(X_i,X_i)\Big]^{\frac{1}{2}}=\frac{1}{\sqrt{n}}\Big|\sum_{\ell=0}^{R}d_\ell\mathbf{p}_\ell^\star\Big|,$$

almost surely. And we use this bound instead of the one of Lemma 13. The following result follows:

$$\begin{aligned}\delta_2(\lambda(\mathbb{T}_{W_{\widetilde{R}}}),\lambda(\boldsymbol{T}_n))\leq&\Big[\sum_{\ell=0}^{R}d_\ell(\mathbf{p}_\ell^\star)^2\Big]^{\frac{1}{2}}\Big[\frac{\widetilde{R}\log(2\widetilde{R}/\alpha)}{n}\Big]^{\frac{1}{2}}\\&+\frac{1}{\sqrt{n}}\Big|\sum_{\ell=0}^{R}d_\ell\mathbf{p}_\ell^\star\Big|+\|\mathbf{p}-\mathbf{p}^R\|_2+\|\mathbf{p}-\mathbf{p}^R\|_\infty\Big[\frac{C_1\log(C_2/\alpha)}{n}\Big]^{\frac{1}{2}}\end{aligned} \tag{26}$$

with probability at least $1-3\alpha$.

Let us study the various terms appearing in (26). First, by orthonormality

$$\sum_{\ell=0}^{R}d_\ell|\mathbf{p}_\ell^\star|^2=\|\mathbf{p}_R\|_2^2\leq\|\mathbf{p}\|_2^2\leq 2$$

since $\mathbf{p}_R$ is the orthogonal projection of $\mathbf{p}$, and $|\mathbf{p}|\leq 1$. Next, using Cauchy-Schwarz inequality

$$\Big|\sum_{\ell=0}^{R}d_\ell\mathbf{p}_\ell^\star\Big|\leq\Big(\sum_{\ell=0}^{R}d_\ell\Big)^{1/2}\Big(\sum_{\ell=0}^{R}d_\ell|\mathbf{p}_\ell^\star|^2\Big)^{1/2}\leq\sqrt{2\widetilde{R}}.$$

Now $\|\mathbf{p}-\mathbf{p}_R\|_\infty\leq 1+\|\mathbf{p}_R\|_\infty$, with $\|\mathbf{p}_R\|_\infty\leq\sum_{\ell=0}^{R}|\mathbf{p}_\ell^\star c_\ell|\|G_\ell^\beta\|_\infty$. But $\|G_\ell^\beta\|_\infty=G_\ell^\beta(1)$ by Formula (4.7.1) and Theorems 7.32.1 and 7.33.1 of [31] so

$$\|\mathbf{p}_R\|_\infty\leq\sum_{\ell=0}^{R}|\mathbf{p}_\ell^\star c_\ell|G_\ell^\lambda(1)=\sum_{\ell=0}^{R}|\mathbf{p}_\ell^\star|d_\ell\leq\sqrt{2\widetilde{R}}\,.$$

Finally, (26) becomes

$$\begin{aligned}\delta_2(\lambda(\mathbb{T}_{W_{\widetilde{R}}}),\lambda(\boldsymbol{T}_n))\leq&\sqrt{2}\Big[\frac{\widetilde{R}\log(2\widetilde{R}/\alpha)}{n}\Big]^{\frac{1}{2}}+\frac{\sqrt{2\widetilde{R}}}{\sqrt{n}}\\&+\Big[\sum_{\ell>R}d_\ell(\mathbf{p}_\ell^\star)^2\Big]^{\frac{1}{2}}+\Big(1+\sqrt{2\widetilde{R}}\Big)\Big[\frac{C_1\log(C_2/\alpha)}{n}\Big]^{\frac{1}{2}}\end{aligned}$$

Hence, since $\widetilde{R} \geq 1$ and $\log n \leq n$, there exists a numerical constant $C > 0$ such that, with probability at least $1-3\alpha$

$$\delta_2(\lambda(\mathbb{T}_{W_{\widetilde{R}}}), \lambda(\boldsymbol{T}_n)) \leq \Big[\sum_{\ell > R} d_\ell(\mathbf{p}_\ell^\star)^2\Big]^{\frac{1}{2}} + C\sqrt{\frac{\widetilde{R}(1+\log(\widetilde{R}/\alpha))}{n}}. \tag{27}$$

Adding (24) gives the first statement of Proposition 4.

Now let us denote by Ω the set with probability larger than $1-3\alpha$ such that the previous inequality is true. One has

$$\delta_2(\lambda(\boldsymbol{T}_n), \lambda^\star) = \delta_2(\lambda(\boldsymbol{T}_n), \lambda^\star)\mathbb{1}_\Omega + \delta_2(\lambda(\boldsymbol{T}_n), \lambda^\star)\mathbb{1}_{\Omega^c}.$$

Observe that each $|\lambda_k(\boldsymbol{T}_n)|$ is bounded by $\rho(\boldsymbol{T}_n)$ the spectral radius of $\boldsymbol{T}_n$. Since $\boldsymbol{T}_n := (1/n)\boldsymbol{\Theta}_0$, it holds that $\rho(\widehat{\boldsymbol{T}}_n) \leq \|\boldsymbol{\Theta}_0/n\|_F \leq 1$. Then

$$\delta_2(\lambda(\boldsymbol{T}_n), \lambda^\star) \leq \delta_2(\lambda(\boldsymbol{T}_n), 0) + \delta_2(0, \lambda^\star) \leq \sqrt{n} + \|\mathbf{p}\|_2 \tag{28}$$

which entails $\delta_2^2(\lambda(\boldsymbol{T}_n), \lambda^\star) \leq (1+\sqrt{2})^2 n$. Hence, using this bound and previous inequality,

$$\begin{aligned}
\mathbb{E}\,\delta_2^2(\lambda(\boldsymbol{T}_n), \lambda^\star) &= \mathbb{E}\left(\delta_2^2(\lambda(\boldsymbol{T}_n), \lambda^\star)\mathbb{1}_\Omega\right) + (1+\sqrt{2})^2 n\,\mathbb{P}(\Omega^c) \\
&\leq 8\Big[\sum_{\ell > R} d_\ell(\mathbf{p}_\ell^\star)^2\Big] + 2C^2\frac{\widetilde{R}(1+\log(\widetilde{R}/\alpha))}{n} + 3\alpha(1+\sqrt{2})^2 n
\end{aligned}$$

as soon as $n^3 \geq \widetilde{R}\log(2\widetilde{R}/\alpha)$. We choose $\alpha = n^{-2}$, and assume $n \geq 2\widetilde{R}$. Then

$$\widetilde{R}\log(2\widetilde{R}/\alpha) = \widetilde{R}\log(2\widetilde{R}n^2) \leq \frac{n}{2}\log(n^3) \leq n^3\,,$$

and

$$\begin{aligned}
\mathbb{E}\left(\delta_2^2(\lambda(\boldsymbol{T}_n), \lambda^\star)\right) &\leq 8\Big[\sum_{\ell > R} d_\ell(\mathbf{p}_\ell^\star)^2\Big] + 2C^2\frac{\widetilde{R}(1+\log(\widetilde{R}n^2))}{n} + 3(1+\sqrt{2})^2 n^{-1} \\
&\leq 8\Big[\sum_{\ell > R} d_\ell(\mathbf{p}_\ell^\star)^2\Big] + C'\frac{R^{\mathbf{d}-1}\log n}{n}
\end{aligned}$$

since $\widetilde{R} = \mathcal{O}(R^{\mathbf{d}-1})$. Now we assume that $\mathbf{p}$ belongs to the Weighted Sobolev $Z^s_{\mathbf{w}_\beta}((-1,1))$. Then, using (12), for all R such that $n \geq 2\widetilde{R}$, it holds

$$\mathbb{E}\left(\delta_2^2(\lambda(\boldsymbol{T}_n), \lambda^\star)\right) \leq 8C(\mathbf{p}, s, \mathbf{d})R^{-2s} + C'\frac{R^{\mathbf{d}-1}\log n}{n}$$

To conclude it is sufficient to choose $R = \lfloor (n/\log n)^{\frac{1}{2s+\mathbf{d}-1}} \rfloor$.

A.8. *Proof of Theorem 6*

We use the notation of the previous proofs and, in particular, the notation of Appendix A.7. The heart of the proof lies in the following proposition, proved in Appendix A.9.

Proposition 15. *Let $R \geq 0$ such that $2\widetilde{R} \leq n$. It holds $\delta_2(\widehat{\lambda}^R, \lambda^{\star R}) \leq 4\delta_2(\lambda(\mathbb{T}_{W_{\widetilde{R}}}), \lambda(\boldsymbol{T}_n)) + \sqrt{2\widetilde{R}}\|\widehat{\boldsymbol{T}}_n - \boldsymbol{T}_n\|$.*

Now, using inequality (4), we know that with probability at least $1-\alpha$ it holds

$$\|\widehat{\boldsymbol{T}}_n - \boldsymbol{T}_n\| \leq \frac{3}{\sqrt{2n}} + C_0 \frac{\sqrt{\log(n/\alpha)}}{n}.$$

Moreover, by (27) in proof of Proposition 4, for all $n^3 \geq \widetilde{R}\log(2\widetilde{R}/\alpha)$,

$$\delta_2(\lambda(\mathbb{T}_{W_{\widetilde{R}}}), \lambda(\boldsymbol{T}_n)) \leq \Big[\sum_{\ell > R} d_\ell (\mathbf{p}_\ell^\star)^2\Big]^{\frac{1}{2}} + C\sqrt{\frac{\widetilde{R}(1+\log(\widetilde{R}/\alpha))}{n}}.$$

Thus there exists a numerical constant $\kappa_0 > 0$ such that, with probability at least $1-3\alpha$

$$\delta_2(\widehat{\lambda}^R, \lambda^{\star R}) \leq 4\Big[\sum_{\ell > R} d_\ell (\mathbf{p}_\ell^\star)^2\Big]^{\frac{1}{2}} + \kappa_0\sqrt{\frac{\widetilde{R}(1+\log(\widetilde{R}/\alpha))}{n}}.$$

if $n^3 \geq (2\widetilde{R})^3 \vee \widetilde{R}\log(2\widetilde{R}/\alpha)$, that gives the first statement of Theorem 6.

Now let us denote by Ω the set with probability larger than $1-3\alpha$ such that the previous inequality is true. One has

$$\delta_2(\widehat{\lambda}^R, \lambda^{\star R}) = \delta_2(\widehat{\lambda}^R, \lambda^{\star R})\mathbb{1}_\Omega + \delta_2(\widehat{\lambda}^R, \lambda^{\star R})\mathbb{1}_{\Omega^c}$$

As for (28), we can prove the coarse bound

$$\delta_2(\widehat{\lambda}^R, \lambda^{\star R}) \leq \sqrt{\widetilde{R}} + \|\mathbf{p}\|_2 \leq (1+\sqrt{2})\sqrt{\widetilde{R}}.$$

Hence, using this bound and previous inequality,

$$\begin{aligned}
\mathbb{E}\left(\delta_2^2(\widehat{\lambda}^R, \lambda^{\star R})\right) &= \mathbb{E}\left(\delta_2^2(\widehat{\lambda}^R, \lambda^{\star R})\mathbb{1}_\Omega\right) + (1+\sqrt{2})^2\widetilde{R}\,\mathbb{P}(\Omega^c) \\
&\leq 32\left[\sum_{\ell > R} d_\ell (\mathbf{p}_\ell^\star)^2\right] + 2\kappa_0^2 \frac{\widetilde{R}\left(1+\log(\widetilde{R}/\alpha)\right)}{n} + 3\alpha(1+\sqrt{2})^2\widetilde{R},
\end{aligned}$$

as soon as $n^3 \geq (2\widetilde{R})^3 \vee \widetilde{R}\log(2\widetilde{R}/\alpha)$. We choose $\alpha = n^{-1}$, and assume $n \geq 2\widetilde{R}$. Then $\widetilde{R}\log(2\widetilde{R}/\alpha) \leq n^3$ and

$$\begin{aligned}
\mathbb{E}\left(\delta_2^2(\widehat{\lambda}^R, \lambda^{\star R})\right) &\leq 32\left[\sum_{\ell > R} d_\ell (\mathbf{p}_\ell^\star)^2\right] + 2\kappa_0^2 \frac{\widetilde{R}\left(1+\log(\widetilde{R}n)\right)}{n} + 3(1+\sqrt{2})^2\frac{\widetilde{R}}{n} \\
&\leq 32\left[\sum_{\ell > R} d_\ell (\mathbf{p}_\ell^\star)^2\right] + (6\kappa_0^2 + 18)\frac{\widetilde{R}\log n}{n}.
\end{aligned}$$

This completes the proof.

A.9. Proof of Proposition 15

- Define Δ_R as follows

$$\forall x, y \in \mathbb{R}^{2\widetilde{R}}, \quad \Delta_R^2(x,y) := \min_{\sigma \in \mathfrak{S}_{2\widetilde{R}}} \Big\{\sum_{k=1}^{2\widetilde{R}} (x_k - y_{\sigma(k)})^2\Big\},$$

where $\mathfrak{S}_{2\widetilde{R}}$ denotes the set of permutations on $[2\widetilde{R}]$.

Once again, using Hardy-Littlewood rearrangement inequality [16, Theorem 368], it holds that

$$\forall x,y\in\mathbb{R}^{2\widetilde{R}} \text{ s.t. } x_1\geq\ldots\geq x_{2\widetilde{R}} \text{ and } y_1\geq\ldots\geq y_{2\widetilde{R}},\quad \Delta_R^2(x,y):=\sum_{k=1}^{2\widetilde{R}}(x_k-y_k)^2\,.$$

Completing with $\widetilde{R}$ zeros, we denote also

$$\widehat{\Lambda}^R:=(\underbrace{\widehat{\mathbf{p}}_0}_{d_0},\underbrace{\widehat{\mathbf{p}}_1,\ldots,\widehat{\mathbf{p}}_1}_{d_1},\ldots,\underbrace{\widehat{\mathbf{p}}_R,\ldots,\widehat{\mathbf{p}}_R}_{d_R},\underbrace{0,\ldots,0}_{\widetilde{R}})\in\mathbb{R}^{2\widetilde{R}}\,,$$
$$\text{and}\quad \Lambda^{\star R}:=(\mathbf{p}_0^\star,\mathbf{p}_1^\star,\ldots,\mathbf{p}_1^\star,\ldots,\mathbf{p}_R^\star,\ldots,\mathbf{p}_R^\star,0,\ldots,0)\in\mathbb{R}^{2\widetilde{R}}\,.$$

Since R does not vary in this proof, we have denoted $\widehat{\mathbf{p}}_\ell:=\widehat{\mathbf{p}}_\ell^R$. Observe that $\delta_2(\widehat{\lambda}^R,\lambda^{\star R})=\Delta_R(\widehat{\Lambda}^R,\Lambda^{\star R})$ using the property described in (7) and Hardy-Littlewood rearrangement inequality [16, Theorem 368] again.

∘ Recall that it holds $\lambda(\mathbb{T}_{W_{\widetilde{R}}})=\{0,\mathbf{p}_0^\star,\mathbf{p}_1^\star,\ldots,\mathbf{p}_1^\star,\ldots,\mathbf{p}_R^\star,\ldots,\mathbf{p}_R^\star\}$ where zero is the only eigenvalue with infinite multiplicity. In particular, remark that the vector $(\mathbf{p}_0^\star,\mathbf{p}_1^\star,\ldots,\mathbf{p}_1^\star,\ldots,\mathbf{p}_R^\star,\ldots,\mathbf{p}_R^\star)$ belongs to $\mathscr{M}_R$. We begin by defining

$$(\overline{\mathbf{p}}_0,\ldots,\overline{\mathbf{p}}_R,\ldots,\overline{\mathbf{p}}_R)\in\arg\min_{u\in\mathscr{M}_R}\min_{\sigma\in\mathfrak{S}_n}\Big\{\sum_{k=1}^{\widetilde{R}}(u_k-\lambda_{\sigma(k)}(\boldsymbol{T}_n))^2+\sum_{k=\widetilde{R}+1}^{n}\lambda_{\sigma(k)}(\boldsymbol{T}_n)^2\Big\}\,,\tag{29}$$

where $\mathfrak{S}_n$ denotes the set of permutation on $[n]$. Also, define

$$\forall x,y\in\mathbb{S}^{\mathbf{d}-1},\quad \overline{W}_{\widetilde{R}}(x,y)=\sum_{\ell=0}^{R}\overline{\mathbf{p}}_\ell c_\ell G_\ell^\beta(\langle x,y\rangle)\,,$$

and observe that $\lambda(\mathbb{T}_{\overline{W}_{\widetilde{R}}})=\{0,\overline{\mathbf{p}}_0,\overline{\mathbf{p}}_1,\ldots,\overline{\mathbf{p}}_1,\ldots,\overline{\mathbf{p}}_R,\ldots,\overline{\mathbf{p}}_R\}$ where zero is the only eigenvalue with infinite multiplicity. Denote $\overline{\sigma}\in\mathfrak{S}_n$ the permutation that achieves the minimum in (29). We have the following intermediate result.

Lemma 16. *It holds*

$$\delta_2^2(\lambda(\mathbb{T}_{\overline{W}_{\widetilde{R}}}),\lambda(\boldsymbol{T}_n))=\sum_{k=1}^{\widetilde{R}}(\overline{\mathbf{p}}_k-\lambda_{\overline{\sigma}(k)}(\boldsymbol{T}_n))^2+\sum_{k=\widetilde{R}+1}^{n}\lambda_{\overline{\sigma}(k)}(\boldsymbol{T}_n)^2\leq\delta_2^2(\lambda(\mathbb{T}_{W_{\widetilde{R}}}),\lambda(\boldsymbol{T}_n))\,,$$

where $(\overline{\mathbf{p}}_\ell)_\ell$ is defined by (29).

Proof. Observe that $\lambda(\mathbb{T}_{\overline{W}_{\widetilde{R}}})$ has at most $\widetilde{R}$ nonzero eigenvalues. Using again Hardy-Littlewood rearrangement inequality [16, Theorem 368] and (7), one may deduce that $\delta_2^2(\lambda(\mathbb{T}_{\overline{W}_{\widetilde{R}}}),\lambda(\boldsymbol{T}_n))$ reads $\sum_{k=1}^{\widetilde{R}}(\overline{\mathbf{p}}_k-\lambda_{\sigma(k)}(\boldsymbol{T}_n))^2+\sum_{k=\widetilde{R}+1}^{n}\lambda_{\sigma(k)}(\boldsymbol{T}_n)^2$ for some permutation $\sigma\in\mathfrak{S}_n$. Taking the infimum leads to the left hand side equality.

Then, observe that $\lambda(\mathbb{T}_{W_{\widetilde{R}}})$ has at most $\widetilde{R}$ nonzero eigenvalues. Using again Hardy-Littlewood rearrangement inequality [16, Theorem 368] and (7), one may deduce again that $\delta_2^2(\lambda(\mathbb{T}_{\overline{W}_{\widetilde{R}}}),\lambda(\boldsymbol{T}_n))$ reads $\sum_{k=1}^{\widetilde{R}}(\mathbf{p}_k^\star-\lambda_{\sigma(k)}(\boldsymbol{T}_n))^2+\sum_{k=\widetilde{R}+1}^{n}\lambda_{\sigma(k)}(\boldsymbol{T}_n)^2$ for some permutation $\sigma\in\mathfrak{S}_n$. Furthermore, recall that $(\mathbf{p}_0^\star,\mathbf{p}_1^\star,\ldots,\mathbf{p}_1^\star,\ldots,\mathbf{p}_R^\star,\ldots,\mathbf{p}_R^\star)$ belongs to $\mathscr{M}_R$ and, hence, it is admissible to Program (29). In particular, the value of the objective at this point is always greater than the minimal value. This gives the right hand side inequality. □

∘ Similarly, denote $((\widehat{\mathbf{p}}_\ell),\widehat{\sigma})$ a point that achieves the minimum in (14). Now, consider

$$S:=\overline{\sigma}([\widetilde{R}])\cup\widehat{\sigma}([\widetilde{R}])\,,$$

and $S^c:=[n]\setminus S$, and define $s:=\#S\leq 2\widetilde{R}\leq n$.

On can check that

$$\widehat{\mathbf{p}}_\ell = \frac{1}{d_\ell} \sum_{k=\widetilde{\ell-1}}^{\widetilde{\ell}} \lambda_{\widehat{\sigma}(k)} \quad \text{and} \quad \overline{\mathbf{p}}_\ell = \frac{1}{d_\ell} \sum_{k=\widetilde{\ell-1}}^{\widetilde{\ell}} \lambda_{\overline{\sigma}(k)}$$

with the convention $\widetilde{-1} = 1$.

∘ Denote by $\mathfrak{S}_{S,n}$ the set of permutation $\sigma \in \mathfrak{S}_n$ such that $\sigma([s]) = S$, $\mathfrak{S}_S$ the set of bijections from $[s]$ onto S and $\mathfrak{S}_s$ the set of permutations of $[s]$. It is clear that $\mathfrak{S}_S \simeq \mathfrak{S}_s$. Observe that

$$\begin{aligned}
(\widehat{\mathbf{p}}_\ell) &= \arg\min_{u\in\mathscr{M}_R} \min_{\sigma\in\mathfrak{S}_n} \Big\{ \sum_{k=1}^{\widetilde{R}} (u_k - \lambda_{\sigma(k)})^2 + \sum_{k=\widetilde{R}+1}^{n} \lambda_{\sigma(k)}^2 \Big\} \\
&= \arg\min_{u\in\mathscr{M}_R} \min_{\sigma\in\mathfrak{S}_{S,n}} \Big\{ \sum_{k=1}^{\widetilde{R}} (u_k - \lambda_{\sigma(k)})^2 + \sum_{k=\widetilde{R}+1}^{n} \lambda_{\sigma(k)}^2 \Big\}
\end{aligned}$$

since one of the permutation $\sigma \in \mathfrak{S}_n$ that achieves the minimum in the first row satisfies $\sigma \in \mathfrak{S}_{S,n}$ and it follows that $(\widehat{\mathbf{p}}_\ell)$ is the arg minimum of the second program. Now, separating the terms $\lambda_{\sigma(k)}^2$ for $k > \widetilde{R}$, we obtain

$$\begin{aligned}
(\widehat{\mathbf{p}}_\ell) &= \arg\min_{u\in\mathscr{M}_R} \min_{\sigma\in\mathfrak{S}_{S,n}} \Big\{ \sum_{k=1}^{\widetilde{R}} (u_k - \lambda_{\sigma(k)})^2 + \sum_{k=\widetilde{R}+1}^{s} \lambda_{\sigma(k)}^2 + \sum_{t\in S^c} \lambda_t^2 \Big\} \\
&= \arg\min_{u\in\mathscr{M}_R} \min_{\sigma\in\mathfrak{S}_S} \Big\{ \sum_{k=1}^{\widetilde{R}} (u_k - \lambda_{\sigma(k)})^2 + \sum_{k=\widetilde{R}+1}^{s} \lambda_{\sigma(k)}^2 + \sum_{t\in S^c} \lambda_t^2 \Big\} \\
&= \arg\min_{u\in\mathscr{M}_R} \min_{\sigma\in\mathfrak{S}_S} \Big\{ \sum_{k=1}^{\widetilde{R}} (u_k - \lambda_{\sigma(k)})^2 + \sum_{k=\widetilde{R}+1}^{s} \lambda_{\sigma(k)}^2 \Big\}. \qquad (30)
\end{aligned}$$

Similarly, one can check that

$$(\overline{\mathbf{p}}_\ell) = \arg\min_{u\in\mathscr{M}_R} \min_{\sigma\in\mathfrak{S}_S} \Big\{ \sum_{k=1}^{\widetilde{R}} (u_k - \lambda_{\sigma(k)}(\boldsymbol{T}_n))^2 + \sum_{k=\widetilde{R}+1}^{s} \lambda_{\sigma(k)}(\boldsymbol{T}_n)^2 \Big\}.$$

∘ Consider the restriction $\dot{\Delta}_R$ of Δ_R to $\mathbb{R}^s$ defined as follows

$$\forall x, y \in \mathbb{R}^s, \quad \dot{\Delta}_R^2(x,y) := \min_{\sigma\in\mathfrak{S}_s} \Big\{ \sum_{k=1}^{s} (x_k - y_{\sigma(k)})^2 \Big\}.$$

Using (5) and Weyl's inequality [4, page 63] and by abuse of notation, note that

$$\dot{\Delta}_R((\lambda_k(\boldsymbol{T}_n))_{k\in S}, (\lambda_k)_{k\in S}) \le \Big[\sum_{k\in S} (\lambda_k - \lambda_k(\boldsymbol{T}_n))^2 \Big]^{\frac{1}{2}} \le \sqrt{s} \|\widehat{\boldsymbol{T}}_n - \boldsymbol{T}_n\|.$$

Moreover, using (30) and by abuse of notation, remark that

$$\begin{aligned}
\dot{\Delta}_R^2((\widehat{\mathbf{p}}_\ell), (\lambda_k)_{k\in S}) &= \min_{\sigma\in\mathfrak{S}_S} \Big\{ \min_{u\in\mathscr{M}_R} \sum_{k=1}^{\widetilde{R}} (u_k - \lambda_{\sigma(k)})^2 + \sum_{k=\widetilde{R}+1}^{s} \lambda_{\sigma(k)}^2 \Big\}. \\
&\le \min_{\sigma\in\mathfrak{S}_S} \Big\{ \sum_{k=1}^{\widetilde{R}} (\overline{\mathbf{p}}_k - \lambda_{\sigma(k)})^2 + \sum_{k=\widetilde{R}+1}^{s} \lambda_{\sigma(k)}^2 \Big\}. \\
&= \dot{\Delta}_R^2((\overline{\mathbf{p}}_\ell), (\lambda_k)_{k\in S})
\end{aligned}$$

where $(\widehat{\mathbf{p}}_\ell) = (\widehat{\mathbf{p}}_0, \widehat{\mathbf{p}}_1, \ldots, \widehat{\mathbf{p}}_1, \ldots, \widehat{\mathbf{p}}_R, \ldots, \widehat{\mathbf{p}}_R, 0, \ldots, 0) \in \mathbb{R}^s$ completing with $s - \widetilde{R}$ zeros.

- Using (29), Lemma 16 and by abuse of notation, observe that

$$
\begin{aligned}
\dot{\Delta}_R^2((\overline{\mathbf{p}}_\ell),(\lambda_k(\boldsymbol{T}_n))_{k\in S}) &= \min_{\sigma\in\mathfrak{S}_s}\Big\{\sum_{k=1}^{\widetilde{R}}(\overline{\mathbf{p}}_k-\lambda_{\sigma(k)}(\boldsymbol{T}_n))^2+\sum_{k=\widetilde{R}+1}^{s}\lambda_{\sigma(k)}(\boldsymbol{T}_n)^2\Big\}.\\
&\le \min_{\sigma\in\mathfrak{S}_s}\Big\{\sum_{k=1}^{\widetilde{R}}(\overline{\mathbf{p}}_k-\lambda_{\sigma(k)}(\boldsymbol{T}_n))^2+\sum_{k=\widetilde{R}+1}^{s}\lambda_{\sigma(k)}(\boldsymbol{T}_n)^2+\sum_{t\in S^c}\lambda_t(\boldsymbol{T}_n)^2\Big\}.\\
&= \min_{\sigma\in\mathfrak{S}_n}\Big\{\sum_{k=1}^{\widetilde{R}}(\overline{\mathbf{p}}_k-\lambda_{\sigma(k)}(\boldsymbol{T}_n))^2+\sum_{k=\widetilde{R}+1}^{n}\lambda_{\sigma(k)}(\boldsymbol{T}_n)^2\Big\},\\
&= \sum_{k=1}^{\widetilde{R}}(\overline{\mathbf{p}}_k-\lambda_{\overline{\sigma}(k)}(\boldsymbol{T}_n))^2+\sum_{k=\widetilde{R}+1}^{n}\lambda_{\overline{\sigma}(k)}(\boldsymbol{T}_n)^2\\
&= \delta_2^2(\lambda(\mathbb{T}_{\overline{W}_{\widetilde{R}}}),\lambda(\boldsymbol{T}_n))\\
&\le \delta_2^2(\lambda(\mathbb{T}_{W_{\widetilde{R}}}),\lambda(\boldsymbol{T}_n))
\end{aligned}
$$

where we denote by $(\overline{\mathbf{p}}_\ell)=(\overline{\mathbf{p}}_0,\overline{\mathbf{p}}_1,\dots,\overline{\mathbf{p}}_1,\dots,\overline{\mathbf{p}}_R,\dots,\overline{\mathbf{p}}_R,0,\dots,0)\in\mathbb{R}^s$ completing with $s-\widetilde{R}$ zeros.

- Using that $\dot{\Delta}_R$ is a semi-distance—in particular the triangular inequality holds, one deduces

$$
\begin{aligned}
\dot{\Delta}_R((\widehat{\mathbf{p}}_\ell),(\overline{\mathbf{p}}_\ell)) &\le \dot{\Delta}_R((\widehat{\mathbf{p}}_\ell),(\lambda_k)_{k\in S})+\dot{\Delta}_R((\lambda_k)_{k\in S},(\lambda_k(\boldsymbol{T}_n))_{k\in S})+\dot{\Delta}_R((\lambda_k(\boldsymbol{T}_n))_{k\in S},(\overline{\mathbf{p}}_\ell))\\
&\le 2\delta_2(\lambda(\mathbb{T}_{W_{\widetilde{R}}}),\lambda(\boldsymbol{T}_n))+\sqrt{s}\|\widehat{\boldsymbol{T}}_n-\boldsymbol{T}_n\|,
\end{aligned}
$$

combining the aforementioned inequalities.

Define $\overline{\Lambda}^R:=(\overline{\mathbf{p}}_0,\overline{\mathbf{p}}_1,\dots,\overline{\mathbf{p}}_1,\dots,\overline{\mathbf{p}}_R,\dots,\overline{\mathbf{p}}_R,0,\dots,0)\in\mathbb{R}^{2\widetilde{R}}$ completing with $\widetilde{R}$ zeros, and remark that

$$
\Delta_R(\widehat{\Lambda}^R,\overline{\Lambda}^R)\le\dot{\Delta}_R((\widehat{\mathbf{p}}_\ell),(\overline{\mathbf{p}}_\ell))\le 2\delta_2(\lambda(\mathbb{T}_{W_{\widetilde{R}}}),\lambda(\boldsymbol{T}_n))+\sqrt{2\widetilde{R}}\|\widehat{\boldsymbol{T}}_n-\boldsymbol{T}_n\|.
$$

- It remains to bound $\Delta_R(\Lambda^{\star R},\overline{\Lambda}^R)$. Note that $\Delta_R(\Lambda^{\star R},\overline{\Lambda}^R)=\delta_2(\lambda(\mathbb{T}_{W_{\widetilde{R}}}),\lambda(\mathbb{T}_{\overline{W}_{\widetilde{R}}}))$. Then, invoke Lemma 16 to get that

$$
\Delta_R(\Lambda^{\star R},\overline{\Lambda}^R)\le\delta_2(\lambda(\mathbb{T}_{W_{\widetilde{R}}}),\lambda(\boldsymbol{T}_n))+\delta_2(\lambda(\boldsymbol{T}_n),\lambda(\mathbb{T}_{\overline{W}_{\widetilde{R}}}))\le 2\delta_2(\lambda(\mathbb{T}_{W_{\widetilde{R}}}),\lambda(\boldsymbol{T}_n)).
$$

Finally we obtain the following bound:

$$
\delta_2(\widehat{\lambda}^R,\lambda^{\star R})\le 4\delta_2(\lambda(\mathbb{T}_{W_{\widetilde{R}}}),\lambda(\boldsymbol{T}_n))+\sqrt{2\widetilde{R}}\|\widehat{\boldsymbol{T}}_n-\boldsymbol{T}_n\| \tag{31}
$$

for all sample size $n\ge 2\widetilde{R}$.

A.10. Proof of Theorem 7

In this proof we denote $D(R)=\sqrt{\widetilde{R}\log n/n}$, so that $B(R)=\max_{R'\in\mathscr{R}}\left\{\delta_2(\widehat{\lambda}^{R'},\widehat{\lambda}^{R'\wedge R})-\kappa D(R')\right\}$. Fix some $R\in\mathscr{R}$. First decompose

$$
\delta_2(\widehat{\lambda}^{\widehat{R}},\lambda^\star)\le\delta_2(\widehat{\lambda}^{\widehat{R}},\widehat{\lambda}^{\widehat{R}\wedge R})+\delta_2(\widehat{\lambda}^{\widehat{R}\wedge R},\widehat{\lambda}^{R})+\delta_2(\widehat{\lambda}^{R},\lambda^\star).
$$

Using the definition of $B(R)$ and $B(\widehat{R})$ it holds that

$$
\delta_2(\widehat{\lambda}^{\widehat{R}},\lambda^\star)\le B(R)+\kappa D(\widehat{R})+B(\widehat{R})+\kappa D(R)+\delta_2(\widehat{\lambda}^{R},\lambda^\star).
$$

We now use the definition of $\widehat{R}$ to write $\delta_2(\widehat{\lambda}^{\widehat{R}}, \lambda^\star) \le 2B(R) + 2\kappa D(R) + \delta_2(\widehat{\lambda}^R, \lambda^\star)$. The last term can be split in $\delta_2(\widehat{\lambda}^R, \lambda^\star) \le \delta_2(\widehat{\lambda}^R, \lambda^{\star R}) + \delta_2(\lambda^{\star R}, \lambda^\star)$. Thus,

$$\delta_2(\widehat{\lambda}^{\widehat{R}}, \lambda^\star) \le 2B(R) + 2\kappa D(R) + \delta_2(\widehat{\lambda}^R, \lambda^{\star R}) + \delta_2(\lambda^{\star R}, \lambda^\star). \tag{32}$$

We shall now control the term $B(R)$. Denote $a_+ = \max(a, 0)$ the positive part of any real a. Let us write

$$\begin{aligned} B(R) &= \max_{R' \in \mathscr{R}} \{\delta_2(\widehat{\lambda}^{R'}, \widehat{\lambda}^{R' \wedge R}) - \kappa D(R')\} \\ &\le \max_{R' \in \mathscr{R}, R' \ge R} \{\delta_2(\widehat{\lambda}^{R'}, \widehat{\lambda}^{R}) - \kappa D(R')\}_+ \\ &\le \max_{R' \in \mathscr{R}, R' \ge R} \{\delta_2(\widehat{\lambda}^{R'}, \lambda^{\star R'}) + \delta_2(\lambda^{\star R'}, \lambda^{\star R}) + \delta_2(\lambda^{\star R}, \widehat{\lambda}^{R}) - \kappa D(R')\}_+ \end{aligned}$$

Now $\delta_2^2(\lambda^{\star R'}, \lambda^{\star R}) = \sum_{k=\widetilde{R}+1}^{\widetilde{R}'} |\lambda_k^\star|^2 \le \delta_2^2(\lambda^{\star R}, \lambda^\star)$. Then

$$B(R) \le \max_{R' \in \mathscr{R}, R' \ge R} \{\delta_2(\widehat{\lambda}^{R'}, \lambda^{\star R'}) - \kappa D(R')\}_+ + \delta_2(\lambda^{\star R}, \lambda^\star) + \delta_2(\lambda^{\star R}, \widehat{\lambda}^{R}) .$$

Finally, combining this with (32),

$$\begin{aligned} \delta_2(\widehat{\lambda}^{\widehat{R}}, \lambda^\star) &\le 2 \max_{R' \in \mathscr{R}, R' \ge R} \{\delta_2(\widehat{\lambda}^{R'}, \lambda^{\star R'}) - \kappa D(R')\}_+ + 2\kappa D(R) + 3\delta_2(\widehat{\lambda}^{R}, \lambda^{\star R}) + 3\delta_2(\lambda^{\star R}, \lambda^\star) , \\ &\le 5 \max_{R' \in \mathscr{R}, R' \ge R} \{\delta_2(\widehat{\lambda}^{R'}, \lambda^{\star R'}) - \kappa D(R')\}_+ + 3\delta_2(\lambda^{\star R}, \lambda^\star) + 5\kappa D(R) . \end{aligned}$$

Now, we invoke Theorem 6 and a union bound to insure that, if $n^3 \ge (2\widetilde{R}_{\max})^3 \vee \widetilde{R}_{\max} \log(2\widetilde{R}_{\max}/\alpha)$ then, with probability greater that $1 - 3|\mathscr{R}|\alpha$, it holds

$$\forall R' \in \mathscr{R} , \qquad \delta_2(\widehat{\lambda}^{R'}, \lambda^{\star R'}) \le 4\delta_2(\lambda^{\star R'}, \lambda^\star) + \kappa_0 \sqrt{\frac{\widetilde{R}'}{n}\left(1 + \log\left(\frac{\widetilde{R}'}{\alpha}\right)\right)}$$

We choose $\alpha = n^{-1-q}$, then $\widetilde{R}_{\max} \log(2\widetilde{R}_{\max}/\alpha) = \widetilde{R}_{\max} \log(2\widetilde{R}_{\max} n^{q+1}) \le n \log(n^{q+2})/2 < 0.1(q+2)n^3$ since $x^{-2} \log x \le 0.09$ and also note that $1 + \log\big(\widetilde{R}/\alpha\big) \le (q+3)\log n$. If $q + 2 \le 10$ then it holds that $n^3 > \widetilde{R}_{\max} \log(2\widetilde{R}_{\max}/\alpha)$ and with probability $1 - 3n^{-q}$

$$\forall R' \in \mathscr{R} , \qquad \delta_2(\widehat{\lambda}^{R'}, \lambda^{\star R'}) \le 4\delta_2(\lambda^{\star R'}, \lambda^\star) + \kappa_0 \sqrt{q+3} D(R')$$

Then, with probability $1 - 3n^{-q}$, provided that $\kappa \ge \kappa_0 \sqrt{q+3}$

$$\begin{aligned} \delta_2(\widehat{\lambda}^{\widehat{R}}, \lambda^\star) &\le 5 \max_{R' \in \mathscr{R}, R' \ge R} \{4\delta_2(\lambda^{\star R'}, \lambda^\star)\}_+ + 3\delta_2(\lambda^{\star R}, \lambda^\star) + 5\kappa D(R) \\ &\le 23\delta_2(\lambda^{\star R}, \lambda^\star) + 5\kappa D(R) \end{aligned}$$

Since it holds for any R, the first inequality of Theorem 7 is proved by choosing $q = 8$.

The second statement will follow by the same roadmap as in the end of proof A.8. Let us denote by Ω the set with probability larger than $1 - 3n^{-q}$ such that the previous inequality is true, and let us find a coarse bound on $\delta_2^2(\widehat{\lambda}^R, \lambda^\star)$. Remind that $\delta_2(\widehat{\lambda}^R, \lambda^{\star R}) \le (1 + \sqrt{2})\sqrt{\widetilde{R}}$ for all R, see (28). Furthermore

$$\delta_2(\lambda^{\star R}, \lambda^\star) = \Big[\sum_{\ell > R} d_\ell(\mathbf{p}_\ell^\star)^2\Big]^{\frac{1}{2}} \le \|\mathbf{p}\|_2 \le \sqrt{2} .$$

Hence, using this bound and previous inequality, for all $R \in \mathscr{R}$,

$$\begin{aligned} \mathbb{E}\left(\delta_2^2(\widehat{\lambda}^{\widehat{R}}, \lambda^\star)\right) &\le \mathbb{E}\left(\delta_2^2(\widehat{\lambda}^{\widehat{R}}, \lambda^\star)\mathbb{1}_\Omega\right) + (1 + 2\sqrt{2})^2 \widetilde{R}_{\max} \mathbb{P}(\Omega^c) \\ &\le 2(23)^2 \delta_2^2(\lambda^{\star R}, \lambda^\star) + 2(5)^2 \kappa^2 D^2(R) + (1 + 2\sqrt{2})^2 \widetilde{R}_{\max} 3n^{-q} \\ &\le 2(23)^2 \left(\delta_2^2(\lambda^{\star R}, \lambda^\star) + \kappa^2 D^2(R) + n^{1-q}\right) \end{aligned}$$

provided that $\kappa \ge \kappa_0 \sqrt{q+3}$. The conclusion follows, choosing for instance $q = 2$.

A.11. Proof of Proposition 9

Note that $\delta_2^2(\lambda^{\star R},\lambda^\star) = \sum_{k>\widetilde{R}} |\lambda_k^\star|^2 = \sum_{\ell>R} d_\ell |\mathbf{p}_\ell^\star|^2$ and this quantity vanishes when $R \geq D$. From Theorem 6 and assuming that $n^3 \geq (2\widetilde{R})^3 \vee \widetilde{R}\log(2\widetilde{R}/\alpha)$, we derive that, for $R \geq D$, it holds

$$\delta_2(\widehat{\lambda}^R,\lambda^{\star R}) \leq \kappa_0 \sqrt{\widetilde{R}\big(1+\log\big(\widetilde{R}/\alpha\big)\big)/n}$$

with probability at least $1-3\alpha$. Remark also that $\mathbf{p}^R = \mathbf{p}$ as soon as $R \geq D$, where $\mathbf{p}^R(t) := \sum_{\ell=0}^R \mathbf{p}_\ell^\star c_\ell G_\ell^\beta(t)$. We now work on the set with probability $1-3\alpha$ given by Theorem 6.

We denote

$$\delta := \min_{0\leq i\neq j\leq D;\ \mathbf{p}_i^\star\neq 0} |\mathbf{p}_i^\star - \mathbf{p}_j^\star| \wedge |\mathbf{p}_i^\star| > 0\,,$$

and note that, for n large enough, it holds $\delta_2(\widehat{\lambda}^R,\lambda^{\star R}) < \delta/2$. Then there exists a permutation $\sigma^\star \in \mathfrak{S}_n$ such that for all $k \in [n]$, $|\widehat{\lambda}^R_{\sigma^\star(k)} - \lambda_k^{\star R}| < \delta/2$. Now, observe that

$$\begin{aligned}
\widehat{\lambda}^R &= (\underbrace{\,\widehat{\mathbf{p}}_0^R\,}_{d_0}, \underbrace{\widehat{\mathbf{p}}_1^R,\dots,\widehat{\mathbf{p}}_1^R}_{d_1},\dots,\underbrace{\widehat{\mathbf{p}}_D^R,\dots,\widehat{\mathbf{p}}_D^R}_{d_D},\dots,\underbrace{\widehat{\mathbf{p}}_R^R,\dots,\widehat{\mathbf{p}}_R^R}_{d_R},0,\dots),\\
\lambda^{\star R} &= (\underbrace{\,\mathbf{p}_0^\star\,}_{d_0}, \underbrace{\mathbf{p}_1^\star,\dots,\mathbf{p}_1^\star}_{d_1},\dots,\underbrace{\mathbf{p}_D^\star,\dots,\mathbf{p}_D^\star}_{d_D},0,\dots).
\end{aligned}$$

We deduce that if $\delta_2(\widehat{\lambda}^R,\lambda^{\star R}) < \delta/2$ then for all h,i,j,k,ℓ such that $\mathbf{p}_h^\star \neq 0$ it holds

$$\begin{aligned}
&\text{If}\quad |\widehat{\mathbf{p}}_k^R - \mathbf{p}_h^\star| \vee |\widehat{\mathbf{p}}_\ell^R - \mathbf{p}_h^\star| \leq \delta/2 \ (\text{resp. } |\widehat{\mathbf{p}}_k^R - \mathbf{p}_i^\star| \vee |\widehat{\mathbf{p}}_k^R - \mathbf{p}_j^\star| \leq \delta/2)\\
&\text{then}\quad \widehat{\mathbf{p}}_k^R = \widehat{\mathbf{p}}_\ell^R \ (\text{resp. } \mathbf{p}_i^\star = \mathbf{p}_j^\star)\,.
\end{aligned}\tag{33}$$

Indeed, one $\widehat{\mathbf{p}}_\ell^R$ cannot be at the same time at a distance less than $\delta/2$ to some $\mathbf{p}_i^\star \neq 0$ and at a distance less that $\delta/2$ to some $\mathbf{p}_j^\star$ since these latter are both at a distance of δ. Necessarily the permutation $\sigma^\star$ is such that the group of eigenvalues $\mathbf{p}_i^\star \neq 0$ of multiplicity d_i matches with the group of eigenvalues $\widehat{\mathbf{p}}_i^R$ with the same multiplicity—recall that the multiplicities d_ℓ are pairwise different since the sequence d_ℓ is increasing. Thanks to (33) it holds

$$\delta_2^2(\widehat{\lambda}^R,\lambda^{\star R}) = \sum_{h;\ \mathbf{p}_h^\star\neq 0} d_h(\widehat{\mathbf{p}}_h^R - \mathbf{p}_h^\star)^2 + \sum_{\ell;\ \mathbf{p}_\ell^\star=0} d_\ell(\widehat{\mathbf{p}}_\ell^R)^2 = \|\widehat{\mathbf{p}}^R - \mathbf{p}\|_2^2\,,$$

noticing that $\|\widehat{\mathbf{p}}^R - \mathbf{p}^R\|_2^2 = \sum_{\ell=0}^R d_\ell(\widehat{\mathbf{p}}_\ell^R - \mathbf{p}_\ell^\star)^2$. It follows that if

$$n^3 \geq (2\widetilde{R})^3 \vee \widetilde{R}\log(2\widetilde{R}/\alpha) \quad \text{and} \quad 2\kappa_0\sqrt{\widetilde{R}\big(1+\log\big(\widetilde{R}/\alpha\big)\big)/n} < \min_{0\leq i\neq j\leq D;\ \mathbf{p}_i^\star\neq 0} |\mathbf{p}_i^\star - \mathbf{p}_j^\star| \wedge |\mathbf{p}_i^\star|$$

then $\|\widehat{\mathbf{p}}^R - \mathbf{p}\|_2 \leq \kappa_0\sqrt{\widetilde{R}\big(1+\log\big(\widetilde{R}/\alpha\big)\big)/n}$ with probability at least $1-3\alpha$. Again, we choose $\alpha = n^{-1-q}$, then $\widetilde{R}\log(2\widetilde{R}/\alpha) = \widetilde{R}\log(2\widetilde{R}n^{q+1}) \leq n\log(n^{q+2})/2 < 0.1(q+2)n^3$ since $x^{-2}\log x \leq 0.09$ and also note that $1+\log(\widetilde{R}/\alpha) \leq (q+3)\log n$. With $q=8$, it holds that $n^3 > \widetilde{R}\log(2\widetilde{R}/\alpha)$. Now, if

$$n \geq 2\widetilde{R} \quad \text{and} \quad 2\kappa_0\sqrt{11\widetilde{R}\log(n)/n} < \min_{0\leq i\neq j\leq D;\ \mathbf{p}_i^\star\neq 0} |\mathbf{p}_i^\star - \mathbf{p}_j^\star| \wedge |\mathbf{p}_i^\star|$$

then $\|\widehat{\mathbf{p}}^R - \mathbf{p}\|_2 \leq \kappa_0\sqrt{11\widetilde{R}\log(n)/n}$ with probability $1-3n^{-8}$, as claimed.

For the second statement, let us denote by Ω the set with probability larger than $1-3n^{-q}$ such that the previous inequality is true, and let us find a coarse bound on $\|\widehat{\mathbf{p}}^R - \mathbf{p}\|_2^2$, for instance $\|\widehat{\mathbf{p}}^R - \mathbf{p}\|_2 \leq (1+\sqrt{2})\sqrt{\widetilde{R}}$. Hence, using this bound and previous inequality, it holds

$$\mathbb{E}\big(\|\widehat{\mathbf{p}}^R - \mathbf{p}\|_2^2\big) \leq \mathbb{E}\big(\|\widehat{\mathbf{p}}^R - \mathbf{p}\|_2^2 \mathbb{1}_\Omega\big) + (1+\sqrt{2})^2\widetilde{R}\,\mathbb{P}(\Omega^c) \leq \frac{\kappa_0^2(q+3)\widetilde{R}\log n}{n} + 3(1+\sqrt{2})^2\widetilde{R}n^{-q}$$

recalling that $1+\log(\widetilde{R}/\alpha) \leq (q+3)\log n$. The conclusion follows, choosing $q=1$.

A.12. Proof of Corollary 10

From Theorem 7, with probability $1-3n^{-8}$

$$\begin{aligned}\delta_2(\widehat{\lambda}^{\widehat{R}},\lambda^\star) &\leq C\min\left(\min_{R<D}\left(\delta_2(\lambda^{\star R},\lambda^\star)+\kappa\sqrt{\frac{\widetilde{R}\log n}{n}}\right),\min_{R\geq D}\left(\kappa\sqrt{\frac{\widetilde{R}\log n}{n}}\right)\right)\\ &\leq C\min\left(\min_{R<D}\left(\delta_2(\lambda^{\star R},\lambda^\star)+\kappa\sqrt{\frac{\widetilde{R}\log n}{n}}\right),\kappa\sqrt{\frac{\widetilde{D}\log n}{n}}\right)\end{aligned}$$

Then, with probability $1-3n^{-8}$,

$$\delta_2(\widehat{\lambda}^{\widehat{R}},\lambda^\star)\leq C\kappa\sqrt{\frac{\widetilde{D}\log n}{n}}\underset{n\to\infty}{\longrightarrow}0.$$

Thus, reasonning as in proof A.11, it holds

$$\delta_2^2(\widehat{\lambda}^{\widehat{R}},\lambda^\star)=||\widehat{\mathbf{p}}^{\widehat{R}}-\mathbf{p}||_2^2=\sum_\ell d_\ell(\widehat{\mathbf{p}}_\ell^{\widehat{R}}-\mathbf{p}^\star)^2,$$

If (by contradiction) $\widehat{R}<D$, then $\delta_2(\lambda^{\star\widehat{R}},\lambda^\star)\geq d_D|\mathbf{p}_D^\star|^2>0$ and then $\delta_2(\widehat{\lambda}^{\widehat{R}},\lambda^\star)$ cannot tend to 0. Thus necessarily $\widehat{R}\geq D$.Moreover, since $\delta_2^2(\widehat{\lambda}^{\hat{R}},\lambda^\star)=||\widehat{\mathbf{p}}^{\hat{R}}-\mathbf{p}||_2^2$, with probability $1-3n^{-8}$

$$||\widehat{\mathbf{p}}^{\hat{R}}-\mathbf{p}||_2^2\leq C^2\kappa^2\frac{\widetilde{D}\log n}{n}.$$

Finally we can write

$$\begin{aligned}\mathbb{E}\left(||\widehat{\mathbf{p}}^{\widehat{R}}-\mathbf{p}||_2^2\right)&\leq\mathbb{E}\left(||\widehat{\mathbf{p}}^{\widehat{R}}-\mathbf{p}||_2^2\mathbb{1}_\Omega\right)+(1+\sqrt{2})^2\widetilde{R}_{\max}\,\mathbb{P}(\Omega^c)\\ &\leq C^2\kappa^2\frac{\widetilde{D}\log n}{n}+3(1+\sqrt{2})^2\widetilde{R}_{\max}n^{-8}\leq(C^2\kappa^2+9)\frac{\widetilde{D}\log n}{n}.\end{aligned}$$

A.13. Proof of Theorem 11

The proof follows the same guidelines as in the sphere example. The only difference is that we do not have Gegenbauer polynomials but normalized Jacobi polynomials Z_ℓ now. In particular, we have previously used the fact that Gegenbauer polynomials are bounded. Here, the same result holds in virtue of (18).

To be specific, when $\boldsymbol{S}$ is a compact symmetric space, note that

- $\sum_{r=1}^{R}\phi_r^2=\sum_{\mathbf{r}=0}^{R-1}\sqrt{d_{\mathbf{r}}}\,\mathrm{zon}^{\mathbf{r}}(e_S)=\sum_{\mathbf{r}=0}^{R-1}d_{\mathbf{r}}=\widetilde{R-1}$ and we get that $\rho(R)\leq\widetilde{R}$ when invoking Lemma 12 or Theorem 2;
- we define $\mathbf{p}^R(t):=\sum_{\ell=0}^{R}\sqrt{d_\ell}\mathbf{p}_\ell^\star Z_\ell(t)$ and we get that

$$\begin{aligned}W_{\widetilde{R}}(x,y)&=\mathbf{p}^R(\cos(\gamma(x,y)))\\ W_{\widetilde{R}}(x,x)&=\sum_{\ell=0}^{R}\sqrt{d_\ell}\mathbf{p}_\ell^\star Z_\ell(1)=\sum_{\ell=0}^{R}d_\ell\mathbf{p}_\ell^\star,\end{aligned}$$

  by (18). This identity can be used in place of (25).

Using these inequalities and following the same guidelines as in the sphere example, one can prove the result.

Appendix B: Computational Considerations

B.1. Proof of Theorem 5

Without loss of generality, assume that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. Similarly, let $u \in \mathscr{M}_R$ and remember that we can group the coordinates of u in groups of sizes d_ℓ for $\ell = 0, \ldots, R$. Reordering by decreasing order, there exists $\tau \in \mathfrak{S}_{R+1}$ such that

$$\begin{aligned} &\underbrace{u_{\widetilde{\tau(1)-1}+1} = \ldots = u_{\widetilde{\tau(1)}}}_{d_{\tau(1)}} \geq \ldots \geq \underbrace{u_{\widetilde{\tau(q)-1}+1} = \ldots = u_{\widetilde{\tau(q)}}}_{d_{\tau(q)}} \geq 0 > \\ &\qquad\qquad \underbrace{u_{\widetilde{\tau(q+1)-1}+1} = \ldots = u_{\widetilde{\tau(q+1)}}}_{d_{\tau(q+1)}} \geq \ldots \geq \underbrace{u_{\widetilde{\tau(R)-1}+1} = \ldots = u_{\widetilde{\tau(R+1)}}}_{d_{\tau(R+1)}}, \end{aligned}$$

for some $q \in \mathbb{N}$. We may consider that $q = 0$ and respectively $q = R+1$ in degenerate cases when all the coefficients are negative and respectively non negative. Remember that $u \in \mathbb{R}^{\widetilde{R}}$ and set $u_k = 0$ for $k > \widetilde{R}$ such that, completing with zeros, consider that $u \in \mathbb{R}^n$. One has

$$\begin{aligned} &\underbrace{u_{\widetilde{\tau(1)-1}+1} = \ldots = u_{\widetilde{\tau(1)}}}_{d_{\tau(1)}} \geq \ldots \geq \underbrace{u_{\widetilde{\tau(q)-1}+1} = \ldots = u_{\widetilde{\tau(q)}}}_{d_{\tau(q)}} \geq \underbrace{u_{\widetilde{R}+1} = \ldots = u_n}_{n-\widetilde{R}} = 0 > \\ &\qquad\qquad \underbrace{u_{\widetilde{\tau(q+1)-1}+1} = \ldots = u_{\widetilde{\tau(q+1)}}}_{d_{\tau(q+1)}} \geq \ldots \geq \underbrace{u_{\widetilde{\tau(R+1)-1}+1} = \ldots = u_{\widetilde{\tau(R+1)}}}_{d_{\tau(R+1)}}. \end{aligned}$$

Note that

$$\min_{\sigma \in \mathfrak{S}_n} \Big\{ \sum_{k=1}^{\widetilde{R}} (u_k - \lambda_{\sigma(k)})^2 + \sum_{k=\widetilde{R}+1}^{n} \lambda_{\sigma(k)}^2 \Big\} = \delta_2^2((\lambda_k)_{k=1}^n, (u_k)_{k=1}^n) = \min_{\sigma' \in \mathfrak{S}_n} \Big\{ \sum_{k=1}^{n} (u_{\sigma'(k)} - \lambda_k)^2 \Big\}, \tag{34}$$

taking $\sigma' = \sigma^{-1}$. Using Hardy-Littlewood rearrangement inequality [16, Theorem 368], it is standard to observe that

$$\begin{aligned} (34) = &\underbrace{(u_{\widetilde{\tau(1)-1}+1} - \lambda_1)^2 + \cdots + (u_{\widetilde{\tau(1)}} - \lambda_{d_{\tau(1)}})^2}_{d_{\tau(1)}} + \cdots \\ &+ \underbrace{(u_{\widetilde{\tau(q)-1}+1} - \lambda_{d_{\tau(1)}+\cdots+d_{\tau(q-1)}+1})^2 + \cdots + (u_{\widetilde{\tau(q)}} - \lambda_{d_{\tau(1)}+\cdots+d_{\tau(q)}})^2}_{d_{\tau(q)}} \\ &+ \underbrace{\lambda^2_{d_{\tau(1)}+\cdots+d_{\tau(q)}+1} + \cdots + \lambda^2_{d_{\tau(1)}+\cdots+d_{\tau(q)}+n-\widetilde{R}}}_{n-\widetilde{R}} \\ &+ \underbrace{(u_{\widetilde{\tau(q+1)-1}+1} - \lambda_{d_{\tau(1)}+\cdots+d_{\tau(q)}+n-\widetilde{R}+1})^2 + \cdots + (u_{\widetilde{\tau(q+1)}} - \lambda_{d_{\tau(1)}+\cdots+d_{\tau(q+1)}+n-\widetilde{R}})^2}_{d_{\tau(q+1)}} \\ &+ \cdots \\ &+ \underbrace{(u_{\widetilde{\tau(R+1)-1}+1} - \lambda_{d_{\tau(1)}+\cdots+d_{\tau(R)}+n-\widetilde{R}})^2 + \cdots + (u_{\widetilde{\tau(R+1)}} - \lambda_n)^2}_{d_{\tau(R+1)}}. \end{aligned}$$

Hence a permutation σ' achieving the minimum in (34) is given by

$$\sigma^{-1} = \sigma' = \begin{pmatrix} k & \sigma'(k) \\ 1 & \widetilde{\tau(1)-1}+1 \\ \vdots & \vdots \\ d_{\tau(1)} & \widetilde{\tau(1)} \\ \vdots & \vdots \\ d_{\tau(1)}+\cdots+d_{\tau(q-1)}+1 & \widetilde{\tau(q)-1}+1 \\ \vdots & \vdots \\ d_{\tau(1)}+\cdots+d_{\tau(q)} & \widetilde{\tau(q)} \\ d_{\tau(1)}+\cdots+d_{\tau(q)}+1 & \widetilde{R}+1 \\ \vdots & \vdots \\ d_{\tau(1)}+\cdots+d_{\tau(q)}+n-\widetilde{R} & n \\ \vdots & \vdots \\ d_{\tau(1)}+\cdots+d_{\tau(R)}+n-\widetilde{R} & \widetilde{\tau(R+1)-1}+1 \\ \vdots & \vdots \\ n & \widetilde{\tau(R+1)} \end{pmatrix}$$

Remark that this permutation can be explicitly written given $\tau \in \mathfrak{S}_{R+1}$ and $q \in [0, R]$. It follows that the set of permutations σ' achieving the minimum in the right hand side of (34) is in one to one correspondence with a subset of $\mathfrak{S}_{R+2}$. Since $\sigma = \sigma'^{-1}$ in (34) the same result holds true for the permutation σ achieving the minimum of the left hand side of (34), proving the result. We define $\mathfrak{H}_R$ has the set of permutation σ achieving the minimum of the left hand side of (34). The proof given here is constructive and it gives an explicit expression of $\mathfrak{H}_R$.

References

[1] M. A. Arcones and E. Giné. Limit theorems for U-processes. *The Annals of Probability*, pages 1494–1542, 1993.

[2] A. S. Bandeira, R. van Handel, et al. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *The Annals of Probability*, 44(4):2479–2506, 2016.

[3] J.-P. Baudry, C. Maugis, and B. Michel. Slope heuristics: overview and implementation. *Statistics and Computing*, 22(2):455–470, Mar 2012.

[4] R. Bhatia. *Matrix analysis*, volume 169. Springer Science & Business Media, 2013.

[5] B. Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.

[6] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities: A nonasymptotic theory of independence*. OUP Oxford, 2013.

[7] M. L. Braun. Accurate error bounds for the eigenvalues of the kernel matrix. *Journal of Machine Learning Research*, 7(Nov):2303–2328, 2006.

[8] S. Bubeck, J. Ding, R. Eldan, and M. Z. Rácz. Testing for high-dimensional geometry in random graphs. *Random Structures & Algorithms*, 2016.

[9] D. Bump. *Lie groups*, volume 225. Springer Science & Business Media, 2013.

[10] F. Dai and Y. Xu. *Approximation theory and harmonic analysis on spheres and balls*. Springer, 2013.

[11] V. De la Pena and E. Giné. *Decoupling: from dependence to independence*. Springer Science & Business Media, 2012.

[12] L. Devroye, A. György, G. Lugosi, F. Udina, et al. High-dimensional random geometric graphs and their clique number. *Electronic Journal of Probability*, 16:2481–2508, 2011.

[13] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.

[14] J. Ferreira, V. A. Menegatto, and A. Peron. Integral operators on the sphere generated by positive definite smooth kernels. *Journal of complexity*, 24(5):632–647, 2008.

[15] A. V. Goldenshluger and O. V. Lepski. General selection rule from a family of linear estimators. *Theory Probab. Appl.*, 57(2):209–226, 2013.

[16] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge university press, 1952.

[17] R. Hasminskii and I. Ibragimov. On density estimation in the view of Kolmogorov's ideas in approximation theory. *Ann. Statist.*, 18(3):999–1010, 09 1990.

[18] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

[19] O. Klopp, A. B. Tsybakov, N. Verzelen, et al. Oracle inequalities for network models and sparse graphon estimation. *The Annals of Statistics*, 45(1):316–354, 2017.

[20] E. D. Kolaczyk. *Statistical analysis of network data*. Springer Series in Statistics. Springer, New York, 2009. Methods and models.

[21] V. Koltchinskii and E. Giné. Random matrix approximation of spectra of integral operators. *Bernoulli*, pages 113–167, 2000.

[22] L. Lovász. *Large networks and graph limits*, volume 60. American Mathematical Soc., 2012.

[23] C. Matias and S. Robin. Modeling heterogeneity in random graphs through latent space models: a selective review. *ESAIM: Proceedings and Surveys*, 47:55–74, 2014.

[24] P.-L. Méliot. The cut-off phenomenon for brownian motions on compact symmetric spaces. *Potential Analysis*, 40(4):427–509, 2014.

[25] P.-L. Méliot. Techniques d'analyse harmonique et résultats asymptotiques en théorie des probabilités. Habilitation à Diriger des Recherches, 2017.

[26] M. E. J. Newman. The structure and function of complex networks. *SIAM Rev.*, 45(2):167–256, 2003.

[27] S. Nicaise. Jacobi polynomials, weighted sobolev spaces and approximation results of some singularities. *Mathematische Nachrichten*, pages 117–140, 2000.

[28] M. Penrose. *Random geometric graphs*, volume 5 of *Oxford Studies in Probability*. Oxford University Press, Oxford, 2003.

[29] L. Rosasco, M. Belkin, and E. D. Vito. On learning with integral operators. *Journal of Machine Learning Research*, 11(Feb):905–934, 2010.

[30] S. R. Sundaresan, I. R. Fischhoff, J. Dushoff, and D. I. Rubenstein. Network metrics reveal differences in social organization between two fission–fusion species, Grevy's zebra and onager. *Oecologia*, 151(1):140–149, Feb 2007.

[31] G. Szegő. *Orthogonal polynomials*. American Mathematical Society, Providence, R.I., fourth edition, 1975. American Mathematical Society, Colloquium Publications, Vol. XXIII.

[32] M. Tang, A. Athreya, D. L. Sussman, V. Lyzinski, C. E. Priebe, et al. A nonparametric two-sample hypothesis testing problem for random graphs. *Bernoulli*, 23(3):1599–1630, 2017.

[33] M. Tang, D. L. Sussman, C. E. Priebe, et al. Universally consistent vertex classification for latent positions graphs. *The Annals of Statistics*, 41(3):1406–1430, 2013.

[34] J. A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434, 2012.

[35] V. V. Volchkov and V. V. Volchkov. *Harmonic analysis of mean periodic functions on symmetric spaces and the Heisenberg group*. Springer Science & Business Media, 2009.

[36] J. A. Wolf. *Harmonic analysis on commutative spaces*. Number 142 in Mathematical Surveys and Monographs. American Mathematical Soc., 2007.

[37] P. J. Wolfe and S. C. Olhede. Nonparametric graphon estimation. *arXiv preprint arXiv:1309.5936*, 2013.