A NICE APPLICATION OF KERNEL METHODS FOR RANDOM GRAPHS

based on a thesis by Araya Valdivia (2020a)

SIMONE MARIA GIANCOLA[*] [†]

Last Modified on April 23, 2025

CONTENTS

INTRODUCTION In this short document we summarize the first part of a thesis about kernel methods in random graphs (Araya Valdivia 2020a). Emphasis is on intuition, proofs are sketched, computations are sometimes explicit, especially in tedious parts. References are to a minimum: for related work and context, we reroute the reader to the original papers (Araya Valdivia 2020b; Araya Valdivia and Yohann 2019) and the thesis (Araya Valdivia 2020a). Section 1 is an introduction to the tools needed. Section 2 presents the model and the algorithm. From subsection 2.II onwards, we give partial details about the proofs.

NOTATION Most of the symbols are standard. The only difference we make is between what is random and what is *not*, what is scalar, what is vectorial and what is matricial. For example, $a, b, c, x, y, z, \alpha, \beta, \gamma$ is a variable, while $\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{x}, \mathrm{y}, \mathrm{z}, \upalpha, \upbeta, \upgamma$ is a random variable. Similarly, $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ is a vector; $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{x}, \mathbf{y}, \mathbf{z}, \boldsymbol{\upalpha}, \boldsymbol{\upbeta}, \boldsymbol{\upgamma}$ is a random vector. Again, $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\Theta}$ is a matrix; $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{\Theta}$ is a random matrix. An operator is denoted as $\mathsf{A}, \mathsf{B}, \mathsf{C}$. An expectation such as $\mathbb{E}_{\mathrm{x}}[\mathrm{x}y\mathrm{z}] = \int xy\mathrm{z}\,\mathrm{d}\mathbb{P}[x]$ is such that y is deterministic, and we integrate out against x which is deterministic once it is expressed inside an integral, keeping z random throughout.

1 TOOLS

{sec:tools

In this section we briefly summarize the tools needed for our presentation.

We consider a probability space (Ω, μ). A kernel is a symmetric measurable function $K : \Omega \times \Omega \to \mathbb{R}$ which is in L^2 for the underlying measure considered. Such measure is often just $\mu \times \mu$. Given a kernel, we write its integral operator as:

$$\mathsf{T}_K(f)(x) : \mathrm{L}^2(\Omega, \mu) \to \mathrm{L}^2(\Omega, \mu) \tag{1.1}$$

$$f \mapsto \int_\Omega K(\cdot, y) f(y)\,\mathrm{d}\mu(y). \tag{1.2}$$

[*] while at Université Paris-Saclay, Orsay institute

[†] simonegiancola09@gmail.com

Let us recall some important notions adapted to our setting.

Definition 1.3 (Hilbert-Schmidt operator). *Let $(e_i)_{i\in\mathbb{I}}, (f_j)_{j\in\mathbb{J}}$ are bases of a separable Hilbert space $\mathscr{H}$ with associated norm $\|\cdot\|$. Suppose $\mathbf{A}$ is a linear bounded operator. If the equivalent sums:* {def:hilbe

$$\sum_{i\in\mathbb{I}} \|\mathbf{A}e_i\|^2 = \sum_{j\in\mathbb{J}} \|\mathbf{A}^* f_j\|^2 = \sum_{(i,j)\in\mathbb{I}\times\mathbb{J}} |\langle \mathbf{A}e_i, f_j\rangle|^2 \tag{1.4}$$

are convergent, then $\mathbf{A}$ is a Hilbert-Schmidt operator.

Definition 1.5 (Compact operator). *An operator $\mathbf{T} : \mathscr{X} \to \mathscr{Y}$ is compact if the image of the unit ball in $\mathscr{X}$ is relatively compact in $\mathscr{Y}$. Namely, the closure of the image of the unit ball is compact.*

Proposition 1.6. *Suppose $K \in \mathrm{L}^2(\Omega, \mu)$. The operator $\mathbf{T}_K$ is compact, self-adjoint and Hilbert-Schmidt.*

Proof. See (Hirsch and Lacombe 1999, pg. 216) or the comment in (Araya Valdivia 2020a, pg. 16). □

Definition 1.7 (Operator norm). *For a bounded linear operator $\mathbf{A}$ on a separable Hilbert space $\mathscr{H}$ we define the operator norm as $\|\mathbf{A}\|_{\mathrm{op}} := \sup_{h\in\mathscr{H}:\|h\|=1} \|\mathbf{A}h\|$.* {def:opera

Proposition 1.8 (Spectral theorem). *Let $\mathbf{A} \in \mathcal{K}(\mathscr{H})$ be compact and self-adjoint in a separable Hilbert space. Then, there exists an at most countable basis of orthonormal pairs $\{e_i, \lambda_i\}_{i\in\mathbb{I}} \in (\mathscr{H}, \mathbb{R})^{\times\mathbb{I}}$ satisfying $|\lambda_i| \overset{i\to\infty}{\to} 0$ if $\mathbb{I} = \mathbb{N}$ and such that:* {prop:spec

$$\mathbf{A} = \sum_{i\in\mathbb{I}} \lambda_i e_i \otimes e_i^*, \qquad \text{when acting on any } u \in \mathscr{H}. \tag{1.9}$$

Said sum is convergent in operator norm. Moreover, the spectrum of $\mathbf{A}$, denoted as $\lambda^{\mathbf{A}} := \{\lambda_i\}_{i\in\mathbb{I}} \cup \{0\} \subset \mathbb{R}$ has the properties that 0 is its only accumulation point, the eigenvalues can be ordered, and we may complete $\{e_i\}_{i\in\mathbb{I}}$ to a basis by adding the needed functions with associated $\lambda_\ell = 0$. From these, we may rewrite the equality above as:

$$\mathbf{A} = \sum_{\lambda\in\lambda^{\mathbf{A}}\setminus\{0\}} \lambda \mathbf{P}_\lambda, \tag{1.10}$$

where $\mathbf{P}_\lambda$ is the orthogonal projection onto the eigenspace associated to λ. Alternatively, we may use the decomposition:

$$\mathbf{A}u = \sum_{i\in\mathbb{I}} \lambda_i \langle u, e_i\rangle e_i. \tag{1.11}$$

Definition 1.12 (Finite-rank operator). *A self-adjoint operator such that the spectral decomposition holds for $\mathbb{I}$ being finite, so that there is a "finite-dimensional" representation of the kernel by truncating the eigenvalues.*

Remark 1.13. *The practical difference with the theory of symmetric matrices is that we add the zero eigenvalue to the spectrum, and we have eigenfunctions instead of eigenvectors.*

Remark 1.14. *From now onwards, we take $\mathbb{I} = \mathbb{N}$ and possibly complete the eigenfunctions to a basis with eigenvalues being zero, or know that the spectrum converges to zero when it is ordered. By convention, we order eigenvalues increasingly, so for a sequence $\{\lambda_i\}_{i\in\mathbb{N}}$ we take the ordering:*

$$\{\lambda_i\}_{i\in\mathbb{N}} \rightsquigarrow |\lambda_0| \geqslant |\lambda_1| \geqslant \cdots \geqslant 0. \tag{1.15}$$

Remark 1.16. *We take eigenfunctions to be unit norm, so a compact self-adjoint operator on a separable Hilbert space has the additional property that:*

$$\|K\|^2_{\mathrm{L}^2(\Omega\times\Omega,\mu\times\mu)} = \int_{\Omega\times\Omega} K^2(x,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) = \sum_{i\in\mathbb{N}} \lambda_i^2 < \infty, \tag{1.17}$$

where we used the assumption that K is in $\mathrm{L}^2(\Omega, \mu)$ and a simple Cauchy-Schwartz inequality. Then, by comparison with convergent series, we have necessarily that $|\lambda_i| < {}^{C}/\sqrt{i}$ for some constant C. In the following sections, we will improve this suboptimal bound with additional assumptions on K.

Since we want to do learning, we will study samples from a distribution and their statistics. Given a collection $\{\mathrm{x}_i\}_{i\leqslant n} \overset{\text{i.i.d.}}{\sim} \mu^{\otimes n}$ we are interested in the *kernel matrix*:

$$\mathbf{T}^{(n)} := \frac{1}{n}\mathbf{K}^{(n)} \in \mathbb{R}^{n\times n}, \qquad \mathrm{K}^{(n)}_{ij} := K(\mathrm{x}_i, \mathrm{x}_j)\ \forall i,j \in [n]. \tag{1.18}$$

If we embed the space of eigenvalue sequences with the distance:

$$\mathrm{d}_2(a,b) := \inf_{\pi\in\mathrm{Sym}(\mathbb{N}):|\mathrm{supp}(\pi)|<\infty} \sqrt{\sum_{i\geqslant 1}(a_i - b_{\pi(i)})^2}, \tag{1.19}$$

{{eqn:spec

then we have a law of large numbers-type result for L^2 kernels in the sense that:

$$\mathrm{d}_2\left(\boldsymbol{\lambda}^{\widetilde{\mathbf{T}}^{(n)}}, \boldsymbol{\lambda}^{(\mathbf{T}_K)}\right) \underset{a.s.}{\overset{n\to\infty}{\to}} 0, \qquad \widetilde{\mathbf{T}}^{(n)} := (1-\mathbb{1}_{i=j})\mathbf{T}^{(n)}_{ij}. \tag{1.20}$$

{{eqn:LLN

See (Koltchinskii and Giné 2000, thm. 3.1).
In particular, we know the asymptotic result and the right scalings to derive *non-asymptotic inequalities*.

1.I *Regularity in Sobolev spaces*

{subsec:so

It turns out that to obtain sharp concentration result we need good notions of eigenvalue decay (Araya Valdivia 2020a, sec. 2.4). In this subsection, we quickly present the main idea and the reduction to our special case.

Consider a measurable metric space $(\mathscr{X}, \mu, \mathrm{d})$. We equip the space $\mathrm{L}^2(\mathscr{X}, \mu)$ with an orthonormal basis $\{\phi_k\}_{k\in\mathbb{K}}$ for a countable set $\mathbb{K}$.

Definition 1.21 (Weighted Sobolev space, general)**.** *For a given set of special weights $\omega : \mathbb{K} \to \mathbb{R}_+$ the weighted Sobolev space is the space of functions that are L^2 integrable with respect to a re-weighted notion of norm. Formally, we use the orthonormal basis to decompose functions as $(\lambda_k, \phi_k)_{k\in\mathbb{K}}$ and construct the following space:* {def:weigh

$$\mathscr{S}_\omega(\mathscr{X}) = \left\{ f \mid f \overset{\mathrm{L}^2}{=} \sum_{k\in\mathbb{K}} \lambda_k \phi_k, \qquad \|f\|^2_\omega := \sum_{k\in\mathbb{K}} \frac{|\lambda_k|^2}{\omega(k)} < \infty \right\}, \qquad \textit{paired with } \|\cdot\|_\omega. \tag{1.22}$$

To give a formal example, for a measurable metric space $(\Omega, \mu, \mathrm{d})$, if we set $\mathscr{X} = \Omega \times \Omega$ and $\nu = \mu \times \mu$ we can build an orthonormal basis of the product space $\mathscr{X}$ by taking tensor products of the orthonormal basis $(e_k)_{k\in\mathbb{K}}$ in $\mathrm{L}^2(\Omega, \mu)$, namely $\phi_{k\ell} = e_k \otimes e_\ell$ for all $k, \ell \in \mathbb{K}$. By comparison with the harmonic series, a sufficient condition for convergence is that:

$$\frac{\lambda_k^2}{\omega(k)} = \frac{1}{k^{1+\eta}}, \qquad \text{for some } \eta > 0. \tag{1.23}$$

If we restrict to open subsets of the Euclidean space $\Omega \subset \mathbb{R}^d$, we can slightly simplify definition 1.21.

Definition 1.24 (Euclidean weighted Sobolev space)**.** *Let $\Omega \subset \mathbb{R}^d$ and $\varrho : \Omega \to \mathbb{R}_+$ be a locally integrable function.*[1] *Then, the (p, ϱ) weighted Sobolev space is the space of locally integrable functions that have a good L^2 norm with respect to $\mathrm{d}\varrho$ and also weak derivatives have a good L^2 norm with respect to ϱ.*[2] *The p number chooses how large weak derivatives must be in the sense that $p = |\boldsymbol{\alpha}|$ for $\boldsymbol{\alpha}$ a multi-index denoting the number of times we derive in a given direction. Mathematically:* {def:eucli

$$\mathscr{S}_{(2,p)}(\Omega, \varrho) := \left\{ f \mid \ \textit{locally integrable}, \ \|f\|_{(p,\varrho)} < \infty \right\}, \tag{1.25}$$

where for D the weak derivative operator we have:

$$\|f\|_{(p,\varrho)} := \sqrt{\int_\Omega |f(x)|^2 \,\mathrm{d}\varrho(x)} + \sqrt{\sum_{|\boldsymbol{\alpha}|=p} |\mathrm{D}^{\boldsymbol{\alpha}} f(x)|^2 \,\mathrm{d}\varrho(x)}. \tag{1.26}$$

1.II *Dot product kernels on the sphere*

Let $\Omega = \mathbb{S}^{d-1}$ be the unit sphere with $d \geqslant 3$, the function $\rho : \mathbb{S}^{d-1} \times \mathbb{S}^{d-1} \to \mathbb{R}_+$ be the geodesic distance and σ be the uniform measure on the sphere. While seemingly complicated, we have quick ways to build intuition on the last two objects. The geodesic distance depends only on the inner product, as $\rho(x,y) = \arccos(\langle x, y\rangle)$. {subsec:do

[1] In words, a function integrable on every compact that defines a Radon measure.
[2] Weak derivatives are as always defined with integration by parts.

The uniform measure is the measure of the random variables $\mathbf{x} = \mathbf{z}/\|\mathbf{z}\|_2$ where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}_d, \boldsymbol{I}_d)$. Since a kernel is a notion of alignment in functional space, i.e. some fancy inner product, we say that our kernel K is a dot product kernel if $K(x,y) = f(\cos\rho(x,y)) = f(\langle x,y\rangle)$ for some function $f : [-1,1] \to [0,1]$. There are some immediate consequences of this construction.

Fact 1.27. *The fact that $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is represented as a scalar function $f : \mathbb{R} \to \mathbb{R}$ means that we can check if it satisfies the Euclidean Sobolev condition in definition 1.24 for a given weight ϱ and p in its scalar representation.*
On the technical viewpoint Nicaise (2000) proves that weighted Sobolev spaces as in definitions 1.21 and 1.24 are equivalent in the following sense.[3] *The equivalence is between metric spaces. Explicitly, we have that the Sobolev space:*

$$\mathscr{S}_{(\omega)}(\mathscr{X}), \qquad \mathscr{X} = [-1,1], \qquad \omega(k) = \frac{1}{1+k(k+d-1)}, \tag{1.28}$$

and:

$$\mathscr{S}_{(2,p)}(\mathscr{X},\varrho), \qquad \mathscr{X} = [-1,1], \qquad \mathrm{d}\varrho(x) = (1-x^2)^{d-3/2}\,\mathrm{d}x, \tag{1.29}$$ {{eqn:sobo

are equivalent. Interestingly, the sequence $l_k = k(k+d-1)$ coincides with the eigenvalues of the Laplace-Beltrami operator of the d dimensional sphere,[4] *and $\mathrm{d}\rho$ is the measure giving weights for the orthogonality relation between Gegenbauer polynomials with parameter $\gamma = d-2/2$, meaning that two different Gegenbauer polynomials are orthogonal in $\mathrm{L}^2([-1,1],\varrho)$. In what follows, we refer to $\|\cdot\|_{(p,\varrho)}$ as the norm in the L^p space with underlying measure ϱ, which in our case depends implicitly on γ.*[5]

Definition 1.30 (Order δ-Sobolev regularity). *A function $f : \Omega \to [-1,1]$ is δ-regular if it belongs to the weighted Sobolev space $\mathscr{S}_{2,\delta}([-1,1], \mathrm{d}\varrho)$.* {def:order

Remark 1.31 (Another δ Sobolev regularity). *From (Castro, Lacour, and Ngoc 2020, eqn. 2). Following the previous observation, we can give an explicit decomposition of f in the basis of $\mathrm{L}^2(\Omega,\varrho)$. We have $f = \sum_{\ell\geqslant 0}\langle f, r_\ell\rangle_{\mathrm{L}^2([-1,1],\varrho)}$ for r_ℓ the orthonormal polynomials, we say f (or equivalently in this case the kernel) is $\delta > 0$ Sobolev regular if:* {rem:anoth

$$\textit{for all } R \geqslant 1 \qquad \sum_{\ell > R}\langle f, r_\ell\rangle^2_{\mathrm{L}^2([-1,1],\varrho)} \leqslant C_{(f,\delta,\mathbb{S}^{d-1})}R^{-2\delta}, \tag{1.32}$$

where $C_{(f,\delta,\mathbb{S}^{d-1})}$ is a constant independent of the cutoff R. In words, it is merely a decay condition in the coefficients of the decomposition of the kernel when seen as a map from the reals to the reals.

Remark 1.33. *In (Castro, Lacour, and Ngoc 2020) it is mentioned that the regularity condition amounts to requiring that the derivative of order δ in the Laplacian of the sphere $\mathbb{S}^{d-1}$ is square integrable.*

Fact 1.34. *Let K be a dot product kernel on the sphere. Then it is rotationally invariant and $\mathbf{T}_K$ is a convolution operator. The basis of eigenvectors in $\mathbb{S}^{d-1}$ for $\mathbf{T}_K$ is independent of K and composed of the spherical harmonics.*

Proof. See (Dai and Xu 2013, chap. 1). □

If we decompose along the spherical harmonics, it is a matter of niceness to keep the eigen-spaces $\mathcal{F}_\ell$ explicit. Each is of dimension d_ℓ and eigenvalue λ_ℓ^* without multiplicity. The dot-product kernel writes:

$$f(\langle x,y\rangle) = \sum_{\ell\geqslant 0}\lambda_\ell^*\sum_{j=0}^{d_\ell} f_{k\ell}(x)f_{k\ell}(y). \tag{1.35}$$ {{eqn:dot

The **order** is of the spherical harmonics, **not w.r.t absolute magnitude**. For these we have many nice explicit formulas.

Fact 1.36 (Summary of spherical harmonics). *The eigen-spaces have dimension d_ℓ with $d_0 = 1, d_1 = d$ and:* {fct:summa

$$d_\ell = \binom{\ell+d-1}{\ell} - \binom{\ell+d-3}{\ell-2} = O\left(\ell^{d-2}\right). \tag{1.37}$$

For any orthonormal basis $\{f_{k\ell}\}_{k=1}^{d_\ell}$ of the space $\mathcal{F}_\ell$ an addition theorem holds (Dai and Xu 2013, eqn. 1.2.8):

$$z_\ell(x,y) = \sum_{k=0}^{d_\ell} f_{k\ell}(x)f_{k\ell}(y), \tag{1.38}$$ {{eqn:addi

[3] The result is interesting for what we will see later, so we report the summary of (**sec. 2.6**)arayavaldiviaKernelSpectralLearning2020.

[4] The Laplace-Beltrami operator is a generalization of the Laplacian to any type of curved space. It is always the divergence of the gradient, this time the Gradient on the sphere.

[5] The generic notion of Sobolev weight is $(1-x^2)^{\gamma-1/2}$, but we will take γ fixed here, so we would rather stress this.

where we term the LHS "zonal polynomial". In particular, the expression does not depend on the basis $\{f_{k\ell}\}_{k=1}^{d_\ell}$. If we use the expression for $f(\langle x,y\rangle)$ in equation 1.35 and the zonal harmonics we have alternatively:

$$f(\langle x,y\rangle) = \sum_{\ell\geqslant 0} \lambda_\ell^* z_\ell(x,y). \tag{1.39}$$

{{eqn:zona

Interestingly enough, we can estimate the growth rate of the eigenvalues above by reconnecting the expression with rescaled Gegenbauer polynomials. Eventually, we know the growth of these and can conclude that it is sufficient to apply the inequalities that follow. Let us build up on this intuition by unrolling the result of (Dai and Xu 2013, thm. 1.2.6, cor. 1.2.7) and (Araya Valdivia 2020a, prop. 12, rem. 2). After this, we will be able to write a concentration result for eigenvalues via a series of growth rates.

To reconnect zonal polynomials and Gegenbauer polynomials it is just a matter of rescaling. Given $x, y \in \mathbb{S}^{d-1}$ points, $\ell \in \mathbb{N}$ a level of spherical polynomials and $d \geqslant 3$, we set $\gamma := {}^{d-2}/_{2}$ and find that:

$$z_\ell(x,y) = c_\ell g_\ell^\gamma(\langle x,y\rangle) = c_\ell \sqrt{d_\ell}\, \tilde{g}_\ell^\gamma(\langle x,y\rangle), \qquad c_\ell := \frac{\ell+\gamma}{\gamma}, \qquad \tilde{g}_\ell^\gamma := \frac{g_\ell^\gamma}{\|g_\ell^\gamma\|_{(2,\varrho)}}, \tag{1.40}$$

{{eqn:zona

where we remind that ϱ depends implicitly on γ. Such Gegenbauer polynomials are orthogonal in $L^2([-1,1],\varrho)$. The zonal polynomial encapsulates a notion of alignment and is naturally maximized at $x = y$. Having Orthogonal polynomials with respect to a reference measure ϱ we clearly have that:

$$\int_{\mathbb{S}^{d-1}} \tilde{g}_\ell^\gamma(x)\tilde{g}_r^\gamma(x)\varrho(x)\,\mathrm{d}x = \delta_{\ell,r}, \tag{1.41}$$

and using the decomposition of f in equation 1.35 together with its "zonal" representation in equation 1.39 and the zonal-Gegenbauer connection of equation 1.40 we find that:

$$\lambda_\ell^* = \frac{\Gamma({}^{d}/_{2})\ell!}{\sqrt{\pi}\Gamma({}^{d-1}/_{2})(2d-2)^{(\ell)}} \int_{-1}^{1} f(x) g_\ell^\gamma(x)\varrho(x)\,\mathrm{d}x, \tag{1.42}$$

{{eqn:grow

where $(a)^{(\ell)}$ is the rising factorial symbol. This is particularly interesting because equation 1.42 is amenable to extracting a growth rate of eigenvalues, and at the same time eigen-vectors are fixed (they are on the sphere). From such growth rate, we will derive concentration results.

1.III *Some useful inequalities*

{subsec:so

Our objective in this quick subsection is to summarize the structure of the statements needed for good concentration of the kernel matrix eigenvalues and the kernel eigenvalues. Araya Valdivia (2020a) builds a multi-step series of hypothesis that lead to such result, which is the basis for what follows. In a nutshell, we are interested in attaining a parametric rate of closeness. Let us begin from the spectral decomposition in the L^2 sense of a given kernel:

$$K \overset{L^2}{=} \sum_{k\in\mathbb{N}} \lambda_k \phi_k \otimes \phi_k, \tag{1.43}$$

which is yet another consequence of proposition 1.8. To establish good results, we need the following well-behavedness assumption.

Assumption 1.44 (Main). *The spectrum is summable in the sense that:* {ass:main}

$$\left\| \sum_{k\geqslant 1} |\lambda_k| \phi_k^2 \right\|_\infty < \infty. \tag{1.45}$$

From it, we can basically reduce ourselves to matrix concentration type inequalities. For these, we define a variance proxy:

Definition 1.46 (Variance proxy). *Given a kernel K, its i-order variance proxy is:* {def:varia

$$v(i) := \left\| \sum_{k=1}^{i} \phi_k^2 \right\|_\infty, \tag{1.47}$$

namely the sup norm of the eigenfunctions associated to the largest i eigenvalues.

From this notion, we seek concentration based on when we truncate the variance proxy. It will naturally depend on the residual error and will be finer as we take i to be large. Moreover, it will be asympotic, in the spirit of a quantified law of large numbers, with tunable probability. From this statement, we will derive depending on all terms a concentration with high probability at a given optimal truncation.

Theorem 1.48 ((Araya Valdivia 2020a), thm. 3). *Let $K : \Omega \times \Omega \to [0,1]$ satisfy assumption 1.44. For a cut level $i \in \mathbb{N}$ define the residual:* {thm:main

$$r(i) := \min\left\{ r \in \mathbb{N} \;\middle|\; |\lambda_i| > \max\left\{ \sum_{k>r} |\lambda_k|, \sqrt{r \sum_{k>r} \lambda_k^2} \right\} \right\}. \tag{1.49}$$

Then, there exists a critical n_c such that for all larger sample sizes and all $\alpha \in (0,1)$ probability levels we have that the kernel matrix $\mathbf{T}^{(n)}$ satisfies:

$$\left| \lambda_i^{(\mathbf{T}^{(n)})} - \lambda_i \right| \lesssim |\lambda_i| \sqrt{\frac{v(r_i) \log\left({}^{r_i}/_{\alpha}\right)}{n}}, \qquad \textit{with probability larger than } 1 - \alpha. \tag{1.50}$$

Remark 1.51. *Notice how the statement is at fixed $i \in \mathbb{N}$, so it is eigenvalue by eigenvalue.*

From this generic statement Araya Valdivia (2020a) finds finer results under additional assumptions. There are three of them, and they all roughly do the following:

- set a growth rate on the eigenvalues $|\lambda_i|$;
- set a growth rate on the sup norm of the eigenfunctions $\|\phi_i\|_\infty$;

Assumption 1.52 (Assumptions $\mathrm{H}_1, \mathrm{H}_2, \mathrm{H}_3$ in (Araya Valdivia 2020a)). *We have the following three settings.* {ass:main

(H_1) *Take a power law decay of eigenvalues $|\lambda_i| = i^{-\delta}$ for some $\delta > 0$ and a power law growth of eigenfunctions $\|\phi_i\|_\infty = i^s$ where to make assumption 1.44 hold we need $\delta > 2s + 1$.*

(H_2) *Take an exponential decay of eigenvalues $|\lambda_i| = \mathrm{e}^{-i\delta}$ for some $\delta > 0$ and a power law growth of eigenfunctions $\|\phi_i\|_\infty = i^s$ where to make assumption 1.44 hold we need $\delta > s$.*

(H_3) *Take an exponential decay of eigenvalues $|\lambda_i| = \mathrm{e}^{-i\delta}$ for some $\delta > 0$ and an exponential growth of eigenfunctions $\|\phi_i\|_\infty = \mathrm{e}^{is}$ where to make assumption 1.44 hold we need $\delta > 2s$.*

Decoupling the contributions into the condition of assumption 1.44, which involves eigenvalues and eigenfunctions, we warp theorem 1.48. We improve mainly because we remove the critical sample size requirement and obtain a *non-asymptotic* result. In exchange, we need the i index to vary with the sample size n as follows.

Theorem 1.53 ((Araya Valdivia 2020a), thm. 4). *Let K be a kernel satisfying either of the conditions in assumption 1.52. Then, with probability larger than $1 - \alpha$ we have a bound of the form:* {thm:main

$$\left| \lambda_i^{(\mathbf{T}^{(n)})} - \lambda_i \right| \lesssim b(i, n, \mathrm{hyp}) \log {}^{1}/_{\alpha}, \tag{1.54}$$

where $b(i, n, \mathrm{hyp})$ depends on the index, the sample size and the hypothesis chosen (table in (Araya Valdivia 2020a, pg. 20)).

Remark 1.55. *Again, this result is at fixed $i \in \mathbb{N}$ eigenvalue.*

In light of the two remarks below theorems 1.48-1.53, we report that the results adapt to a full spectrum concentration in terms of the spectral distance defined in equation 1.19. By analogy with matrix results, it is termed a Hoffman-Wielandt type inequality.

Corollary 1.56 (Hoffman-Wielandt type inequality). *Let K satisfy H_2 or H_3, or H_1 with the added condition that $\delta > 2s + 2$. Then, with probability larger than $1 - \alpha$:* {cor:hoffm

$$\mathrm{d}_2(\boldsymbol{\lambda}^{\mathbf{T}^{(n)}}, \boldsymbol{\lambda}^{\mathbf{T}_K}) \lesssim_\alpha \frac{1}{\sqrt{n}}, \tag{1.57}$$

where the asymptotic inequality depends on the probability level α.

In our application, we will show that dot product kernels enjoy this generic concentration nicely. This means that they satisfy either of the assumptions. A key point is that for spherically symmetric (dot-product) kernels like the ones we consider the eigenvalues have an analytic expression. Without this peculiar advantage, one can resort to the Sobolev regularity approach.

How do we make assumption 1.44 true? Using equation 1.40 it suffices to have:

$$\sum_{\ell \geqslant 0} |\lambda_\ell^*| d_\ell < \infty, \tag{1.58}$$

and from the scaling of d_ℓ in fct. 1.36 a sufficient condition on the growth rate of the eigenvalues is:[6]

$$|\lambda_\ell^*| \in O\left(\ell^{1-d-\epsilon}\right), \text{for any } \epsilon > 0. \tag{1.59}$$

In order to satisfy the more expressive hypotheses, we use the addition theorem (i.e. eqn. 1.38) to bound the variance proxy of definition 1.46. From the two, we find that $v(i) \in O(i)$ (Araya Valdivia 2020a, lem. 13), and so reconnecting with the triplet in assumption 1.52:

- since $s = 0$ in the eigenvector growth rate, H_1 is satisfied if $\lambda_i \in O\left(i^{-\delta}\right)$ for some $\delta > 1$;
- for the same reason, $s = 0$ and $\mathrm{H}_2, \mathrm{H}_3$ are satisfied if $\lambda_i \in O\left(\mathrm{e}^{-\delta i}\right)$ for some $\delta > 0$.

Thanks to the growth rate of the variance proxy, we can specify theorem 1.53 to become sharper.

Proposition 1.60 ((Araya Valdivia 2020a), lem. 14 and cor. 15)**.** *Consider a kernel K such that $v(i) \in O(i)$ for all $i \in \mathbb{N}$ (e.g. a dot product kernel). Then, theorem 1.53 is true, with the added fact that the constant $b(i, n, \mathrm{hyp})$, where* hyp *depended on s and i where i depended on s are true with $s = 0$. See the table in (Araya Valdivia 2020a, pg. 20) for the statements.* {prop: sh

Consequently, consider $K(x, y) = f(\langle x, y\rangle)$ a dot product kernel on the sphere where $f \in \mathscr{S}_{(2,p)}([-1, 1], \varrho)$ for ϱ as in equation 1.29. For any $\alpha \in (0, 1)$, there exists $\epsilon > 0$ such that with probability larger than $1 - \alpha$ for all $i \in [n]$ indices we have a good concentration of the kernel matrix eigenvalues:

$$\left|\lambda_i^{\mathbf{K}^{(n)}} - \lambda_i\right| \lesssim_\alpha i^{-\eta+1/2} n^{-1/2}, \qquad \eta := \frac{p+\epsilon}{d-1} + \frac{1}{2}. \tag{1.61}$$

2 Inferring distances from spherical data

{sec:algor

In this section we apply the tools to the problem of learning distances from a noisy sample of a graphon, i.e. morally a continuous graph. Let us begin by introducing the main objects.

Throughout, we place ourselves in $(\mathbb{S}^{d-1}, \sigma)$ where σ is the uniform measure on the sphere. A classical result is that $\mathbb{S}^{d-1}$ is separable as $\mathbb{R}^d$ is.[7] A graphon is a kernel function that takes two points and returns a "continuous" edge, namely $W : \mathbb{S}^{d-1} \times \mathbb{S}^{d-1} \to [0, 1]$. To pose a learning task, we sample n vectors uniformly at random, form the dataset $\mathbb{D} = \{x_i\}_{i=1}^n$ and construct the Gram matrix of distances:

$$\mathbf{G}^\star \in \mathbb{R}^{n \times n}, \qquad \mathrm{G}_{ij} := \left\langle \mathbf{x}_i, \mathbf{x}_j \right\rangle. \tag{2.1}$$

To build a model, let us consider the probability matrix:

$$\mathbf{T}^{(n)} := \frac{1}{n}\mathbf{\Theta} \in \mathbb{R}^{n \times n}, \qquad \Theta_{ij} = W_n(\mathbf{x}_i, \mathbf{x}_j) = \rho_n W(\mathbf{x}_i, \mathbf{x}_j), \tag{2.2}$$

where ρ_n is a scaling factor.

Assumption 2.3 (Relative sparsity regime)**.** *For technical reasons, we take $\rho_n \in \Omega\left(\log n/n\right)$.* {ass:relat

When the graphon function W depends only on the inner product, i.e. it is a dot product kernel, we say it is a *geometric graphon*. Then, $\mathbf{\Theta}$ is symmetric, and we may build an adjacency matrix by thresholding the probabilities:

$$\mathbf{A}^{(n)} \in \mathbb{R}^{n \times n}, \qquad n\mathrm{A}_{ij}^{(n)} \sim \mathrm{Ber}(\mathbf{\Theta}_{ij}). \tag{2.4}$$

We have three main objects:

[6] Just compare with the harmonic series.

[7] A metric space $\mathbb{R}^d$ is separable if and only if it is second countable (i.e. it has a countable basis), and second countability passes to subspaces. A sphere is a subspace of $\mathbb{R}^d$, which is second countable.

- The random probability matrix $\mathbf{T}^{(n)}$ which normalizes the W-similarity, with eigenvalues $\boldsymbol{\lambda}^{\mathbf{T}^{(n)}}$ ordered non-increasingly in absolute value;
- the random observed adjacency matrix $\mathbf{A}^{(n)}$ which thresholds $\boldsymbol{\Theta}$ with eigenvalues $\boldsymbol{\lambda}^{\mathbf{A}^{(n)}}$ ordered non-increasingly in absolute value;
- the deterministic integral operator T_W which is Hilbert-Schmidt on the sphere having eigenvalues λ^* indexed by the degree of spherical harmonics.

Our objective is to learn the latent distances in the graphon from an observation. The key argument will be that we can establish various concentration results. We summarize them below.
If W has Sobolev regularity δ then (Araya Valdivia and Yohann 2019, thm. 2):

$$\mathrm{d}_2\left(\boldsymbol{\lambda}^{1/\rho_n \mathbf{T}^{(n)}}, \boldsymbol{\lambda}^{\mathsf{T}_W}\right) \lesssim_\alpha \left(\frac{\log n}{n}\right)^{\delta/2\delta+d-1}, \qquad \text{with probability larger than } 1-\alpha. \tag{2.5}$$

{{eqn:de c

The observed adjancency matrix approaches the probability matrix in operator norm:[8]

$$\mathbb{E}\left[\left\|\mathbf{A}^{(n)} - \mathbf{T}^{(n)}\right\|_{\mathsf{op}}\right] \lesssim \max\left\{\frac{\rho_n}{\sqrt{n}}, \frac{\sqrt{\log n}}{n}\right\}, \tag{2.6}$$

which translates into a bound found in (Araya Valdivia 2020a, eqn. 3.3, thm. 32):

$$\frac{1}{\rho_n}\left\|\mathbf{A}^{(n)} - \mathbf{T}^{(n)}\right\|_{\mathsf{op}} \lesssim_{\alpha/4} C \max\left\{\frac{1}{\sqrt{\rho_n n}}, \frac{\sqrt{\log n}}{\rho_n n}\right\}. \tag{2.7}$$

Morally, we will want to connect our observed $\mathbf{A}^{(n)}$ with T_W using refinements of these two. The mere results are concentration inequalities for the spectrum and the eigenvectors-functions.
In particular, using the fact that the eigenfunctions of T_W are the Gegenbauer polynomials, if we only take the first (linear) eigenfunction and apply the addition theorem we will find that:

$$g_1^\gamma(\langle \mathbf{x}_i, \mathbf{x}_j \rangle) = \frac{1}{c_1}\sum_{k=1}^{d} \phi_k(\mathbf{x}_i)\phi_k(\mathbf{x}_j), \qquad \gamma := \frac{d-2}{2}, \qquad c_1 := \frac{d}{d-2} \tag{2.8}$$

or better:

$$g_1^{(\gamma)}(\langle \mathbf{x}_i, \mathbf{x}_j \rangle) = \frac{\gamma}{d}\sum_{k=1}^{d} \phi_k(\mathbf{x}_i)\phi_k(\mathbf{x}_j) = 2\gamma \langle \mathbf{x}_i, \mathbf{x}_j \rangle, \qquad \gamma := \frac{d-2}{2}, \tag{2.9}$$

since the first Gegenbauer polynomial is linear ((Dai and Xu 2013, app. B2)). From this we can deduce the following for the true Gram matrix:

$$\mathbf{G}^\star =:= \frac{1}{n}\left[\langle \mathbf{x}_i, \mathbf{x}_j \rangle\right]_{i,j \in [n]} = \frac{1}{d}\mathbf{V}^\star [\mathbf{V}^\star]^\top, \tag{2.10}$$

where we defined:

$$\mathbf{V}^\star = \left[\mathbf{v}^{\star;1} \quad \mathbf{v}^{\star;2}, \ldots, \mathbf{v}^{\star;d}\right], \qquad \mathbf{v}^{\star;j} := \begin{bmatrix} \phi_j(\mathbf{x}_1)/\sqrt{n} \\ \vdots \\ \phi_j(\mathbf{x}_n)/\sqrt{n}. \end{bmatrix} \tag{2.11}$$

Therefore, we want to understand well the first eigenvector in the spherical harmonics since its $d_1 = d$ associated eigenvectors make an exact expression of the Gram matrix.
Eventually, we will prove that for a sufficient number of observations we can take "good" eigenvectors of $\mathbf{A}^{(n)}$ to estimate well the population Gram matrix in Frobenius norm, min-max optimally!

Just like any statistical problem, we need to ensure *identifiability*. When speaking about matrices it is common to assume this by some quantified separation of eigenvalues. In this line of work, there are no

[8] One starts from the bound $\mathbb{E}\left[\left\|\mathbf{A}^{(n)} - \mathbf{T}^{(n)}\right\|_{\mathsf{op}}\right] \lesssim \sqrt{d_0/n} + \sqrt{d_0^* \log n/n}$ where $d_0 := \max_{i \in [n]} \sum_{j=1}^{n} \Theta_{ij}(1-\Theta_{ij})$ and $d_0^* := \max_{ij}\left\|\mathrm{T}_{ij}^{(n)} - \mathrm{A}_{ij}^{(n)}\right\|_\infty$. It first appeared in (Bandeira and Handel 2016, cor. 3.3). Our scaling gives a more explicit upper bound.

surprises: we will need to ensure that the eigenvalue of the first Gegenbauer polynomial is well-separated from the others. Let us then define the spectral gap as:

$$\operatorname{Gap}(W) := \min_{j \neq 1} |\lambda_1^* - \lambda_j^*|. \tag{2.12}$$

Then, we remove annoying cases: the only accumulation point of the spectrum is zero, and the spherical harmonics eigenvalues are counted with multiplicity. The only cases in which the gap is null is if $\lambda_1^* = 0$ or the multiplicity is larger than one (see (Araya Valdivia 2020a, prop. 23)). In both cases, we would have issues with identifiability (i.e. there are two eigen-spaces). Assuming this does not hold is for sanity, not for simplicity. Having identifiability, we need to ensure that the observations are strong enough. To go "above" the noise, we will define a "good event E" under which the signal of the observations is above (an expression of) the spectral gap. While complicated at first sight, we just require the problem to be feasible. To be precise, the event will be:

$$E_n := \left\{ \max \left\{ \mathrm{d}_2 \left(\boldsymbol{\lambda}^{1/\rho_n \mathbf{T}^{(n)}}, \boldsymbol{\lambda}^{\mathsf{T}_W} \right), \frac{2^{9/2}\sqrt{d} \left\| \mathbf{A}^{(n)} - \mathbf{T}^{(n)} \right\|_{\mathsf{op}}}{\rho_n \operatorname{Gap}(W)} \right\} \leqslant \frac{\operatorname{Gap}(W)}{4} \right\} \tag{2.13}$$

Without regard to the details, we are just saying that our two main quantities are, upon correct rescaling, both not too noisy with respect to the natural separation of the deterministic object we want to estimate. The desired properties of these event are that:

(i) it has high probability, or at least a quantifiable tunable probability;

(ii) under it, we can estimate the population Gram matrix from the observed Gram matrix well and algorithmically;

(iii) it has a good dependence on n to find a critical scaling of the dataset size to make inference;

For (i), we just need a lemma.

Lemma 2.14. *If* $\operatorname{Gap}(W) > 0$ *there exists a critical* $n_c \equiv n_c(W, \alpha)$ *such that for all* $n \geqslant n_c$ *and* $\alpha \in (0,1)$ *it holds that* $\mathbb{P}[E_n] \geqslant 1 - \alpha/2$. {lem:high

For (ii) we will need more work, for (iii) we remark that it is evident from the proof technique. Unlike other nice results, there is no closed-form formula for n in terms of the other parameters but rather a set of inequalities identifying a *region* in the parameter space. It is less aesthetic, but it is still a plug-in information: if we know the parameters, we can directly answer if inference is feasible or not. For the expression, see (Araya Valdivia 2020a, rem. 6).

2.1 *The algorithm and its properties*

{subsec:th

We now focus on (ii). The main idea behind the algorithm is in the following statements.

Proposition 2.15 ((Araya Valdivia 2020a) proposition 25)**.** {prop:isol
On the event E_n*, the spectrum of* $\mathbf{A}^{(n)}$ *has:*

(bulk) a unique set d eigenvalues with diameter smaller than $\rho_n \operatorname{Gap}(W)/2$*;*

(rest) the others are at distance higher than $\rho_n \operatorname{Gap}(W)/2$ *from such bulk.*

Theorem 2.16 ((Araya Valdivia 2020a), theorem 26)**.** *Let* W *be a graphon on the sphere* $\mathbb{S}^{d-1}$ *that is* δ*-regular and has positive gap* $\operatorname{Gap}(W) > 0$*. Then, there exists a set of* d *eigenvectors of* $\mathbf{A}^{(n)}$ *that estimates well the true Gram matrix in the following sense. Letting* $\{\mathbf{v}^{(j)}\}_{j=1}^{d}$ *be the columns of* $\hat{\mathbf{V}}$*, we construct the estimator:* {thm:algor

$$\widehat{\mathbf{G}} := \gamma \hat{\mathbf{V}} \hat{\mathbf{V}}^{\top}, \tag{2.17}$$

and can quantify the error as:

$$\left\| \mathbf{G}^{\star} - \widehat{\mathbf{G}} \right\|_{\mathrm{F}} = O\left(\frac{n^{-\delta/(2\delta+d-1)}}{\operatorname{Gap}(W)} \right). \tag{2.18}$$ {{eqn:erro

Remark 2.19. *The eigenvectors forming $\hat{\mathbf{V}}$ form the isolated bulk of proposition 2.15. So E_n is really the good event where above identifiability we can actually work out this specific algorithm.*

Remark 2.20. *The rate of equation 2.18 is min-max optimal in the function space of δ-regular functions in $d-1$ dimensions for non-parametric regression (Emery, Nemirovski, and Voiculescu 2000, chap. 2).*

Thanks to these two results, we can write down a procedure to recover these eigenvectors. This is the objective of algorithm 1. The idea is to reconstruct the eigen-space of λ_1^*. It is done by finding a subset of dimension $d_1 = d$ of the eigenvalues of $\mathbf{A}^{(n)}$ denoted as $\boldsymbol{\lambda}^{\mathbf{A}^{(n)}} \in \mathbb{R}^n$ where all of the eigenvalues are jointly close to λ_1^*. The fact that E_n holds with high probability allows us to be always in a "good case" for this task. We will then have d candidates, and compare them via the following quantity:

$$\mathrm{Gap}(\mathbf{A}^{(n)};\mathbb{I}) := \min_{i\notin\mathbb{I}}\max_{j\in\mathbb{I}} |\lambda_i^{\mathbf{A}^{(n)}} - \lambda_j^{\mathbf{A}^{(n)}}|, \qquad \text{where } |\mathbb{I}| = d. \tag{2.21}$$

By analogy with the graphon case, the spectral gap of the adjacency matrix will be the largest gap among any possible candidate:

$$\mathrm{Gap}(\mathbf{A}^{(n)};d) := \max_{\mathbb{I}\subset[1:n-1]\ |\mathbb{I}|=d} \mathrm{Gap}(\mathbf{A}^{(n)};\mathbb{I}). \tag{2.22}$$ {{eqn:gap

Remark 2.23. *By (Araya Valdivia 2020a, prop. 36), we may ignore the eigenvalue that gets close to zero, which is shown to be close to λ_0^*, hence far from λ_1^*.*

So far, we have to search over a large collection of subsets, namely all those that have dimension d, which has cardinality $\binom{n-2}{d}$. Fortunately, we can simplify the search space, as it turns out that sets of consecutive sets of indices dominate the optimization problem in equation 2.22.

Lemma 2.24 ((Araya Valdivia 2020a) lem. 27)**.** *The quantity* $\mathrm{Gap}(\mathbf{A}^{(n)};d)$ *is attained by a set $\mathbb{I}$ with $|\mathbb{I}| = d$ by construction and $\mathbb{I} = \{i_1, \ldots, i_d\}$ corresponding to d consecutive eigenvalues of $\mathbf{A}^{(n)}$ sorted in decreasing order.*

Algorithm 1 Harmonic eigen-cluster (HEiC) algorithm (Araya Valdivia 2020a)

{alg:HEICD

```
Require: $(A^{(n)}, d)$ adjacency matrix and dimension;
Ensure: gap quantification, eigenvalues $\{\lambda^{(j)}\}_{j=1}^d$ and associated eigenvectors $\{v^{(j)}\}_{j=1}^d$.
  $\Lambda_{\text{sort}} \leftarrow \{\lambda_1^A, \ldots, \lambda_{n-1}^A\}$ sorted decreasingly;
  $\Lambda_{\text{sol}} \leftarrow \Lambda_{\text{sort}}[1 : d+1]$                    ▷ biggest $d$ eigenvalues;
  $i \leftarrow 2$;
  gap $\leftarrow \mathrm{Gap}(A^{(n)}; [d])$;
  while $i \leqslant n-d$ do
      if $\mathrm{Gap}(A^{(n)}; [i : i+d]) >$ gap then
          $\Lambda_{\text{sol}} \leftarrow \Lambda_{\text{sort}}[i : i+d]$;
      end if
      $i \leftarrow i+1$;
  end while
```

Putting it all together, we found that algorithm 1 attains the min-max optimal rate of non-parametric inference in δ regular Sobolev spaces. We can recover latent distances in the best possible way with a rather simple procedure. The key is understanding that the linear Gegenbauer polynomial (i.e. a rescaled version of the latent distances in a dot product kernel), is represented by d (consecutive) eigenvalues of the kernel operator. Under our good event, such d eigenvalues are well approximated by the eigenvalues of the observed graph $\mathbf{A}^{(n)}$, via the chain $\mathbf{A}^{(n)} \rightsquigarrow \mathbf{T}^{(n)} \rightsquigarrow \mathbf{T}_W$, crucially using concentration inequalities.
The thesis of Araya Valdivia (2020a) proceeds further with more general algorithms and settings; we reroute the interested reader to the original work.

2.11 *Proof sketch of main theorem*

{subsec:pr

In this subsection, we aim to argue a proof for proposition 2.15 and theorem 2.16. Throughout, we use the hat-none-star notation to denote:

- $\hat{\mathbf{V}}$ the matrix of d eigenvectors from $\mathbf{A}^{(n)}$ which is effectively an estimator;
- $\mathbf{V}$ the matrix of eigenvectors of $\mathbf{T}^{(n)}$, which is non-observable;
- $\mathbf{V}^\star$ the "true" matrix of eigenvectors of $\mathbf{T}_W$, in this case associated to the first Gegenbauer polynomial.

In particular, we avoided placing a hat on $\mathbf{A}^{(n)}$ because it is not at all an estimator. Araya Valdivia (2020a) uses the notation $\hat{\mathbf{T}}^{(n)}$ for it.

Proof of high probability event, lemma 2.14. We use the result of Bandeira and Handel (2016), reported in (Araya Valdivia 2020a, thm. 32) on the centered matrix $\mathbf{M} = \mathbf{A}^{(n)} - \mathbf{T}^{(n)}$. With probability larger than $1 - \alpha$ the following inequality holds:

$$\left\| \mathbf{A}^{(n)} - \mathbf{T}^{(n)} \right\|_{\mathsf{op}} \lesssim_\alpha C \max \left\{ \sqrt{\frac{\rho_n}{n}}, \frac{\sqrt{\log n}}{n} \right\}. \tag{2.25}$$

In particular, we used that $\max_i \sum_{j=1}^n \Theta_{ij}(1 - \Theta_{ij}) \leqslant \max_i \sum_{j=1}^n \Theta_{ij} \in O\left(n\rho_n\right)$ as $\Theta_{ij} = \rho_n W(\mathbf{x}_i, \mathbf{x}_j) \geqslant 0$. Reordering the ρ_n scaling, we find:

$$\frac{1}{\rho_n} \left\| \mathbf{A}^{(n)} - \mathbf{T}^{(n)} \right\|_{\mathsf{op}} \lesssim_\alpha C \max \left\{ \frac{1}{\sqrt{\rho_n n}}, \frac{\sqrt{\log n}}{\rho_n n} \right\}. \tag{2.26}$$

Now we would like to upper bound further, but we do not know *a priori* if the second term in the max is big or not. This is where assumption 2.3 is needed. We take the relative sparse regime, and the second term is just an $o\left(1\right)$, so for n large enough, say larger than a critical n_{c}:

$$\frac{1}{\rho_n} \left\| \mathbf{A}^{(n)} - \mathbf{T}^{(n)} \right\|_{\mathsf{op}} \lesssim_\alpha \frac{[\mathrm{Gap}(W)]^2}{2^{17/2}\sqrt{d}}, \tag{2.27}$$

where the RHS is just a convenient number for later. With an application of the Davis-Kahan theorem (use the exact formulation of (Araya Valdivia 2020a, thm. 29)) and a perturbation result (Araya Valdivia 2020a, lem. 37) we have:

$$\left\| \hat{\mathbf{V}}\hat{\mathbf{V}}^\top - \mathbf{V}\mathbf{V}^\top \right\|_{\mathsf{F}} \leqslant \frac{2^{5/2}\sqrt{d}1/\rho_n \left\| \mathbf{A}^{(n)} - \mathbf{T}^{(n)} \right\|_{\mathsf{op}}}{1/\rho_n \Delta} \lesssim_\alpha \frac{\rho_n \left(\mathrm{Gap}(W)\right)^2}{\Delta}, \qquad \Delta := \mathrm{d}(\{\lambda_{i_j}^{\mathbf{T}^{(n)}}\}_{j=1}^d, \lambda^{\mathbf{T}^{(n)}} \backslash \{\lambda_{i_j}^{\mathbf{T}^{(n)}}\}_{j=1}^d), \tag{2.28}$$

{{eqn:ugly

where we used the fact that we are under the event E_n and the inequality on the operator norm is up to α factors controlling its tails. The graphon belongs to a δ regular Sobolev space, so applying (Araya Valdivia 2020a, thm. 33) which is a result of Castro, Lacour, and Ngoc (2020), we find back equation 2.5. Furthermore, under the event E_n we have:

$$C \left(\frac{\log n}{n} \right)^{\delta/2\delta + d - 1} \leqslant \frac{\mathrm{Gap}(W)}{8}. \tag{2.29}$$

The two statements are a consequence of the fact that we are in E_n. Therefore, for a fixed couple $(\alpha, \mathrm{Gap}(W))$, i.e. a graphon and a confidence threshold, there exists a critical size n_{c} such that once crossed, we have $\mathbb{P}\left[E_n\right] \geqslant \alpha/2$. It suffices to notice that upper bounds depend only on this tuple. □

Proof of isolated bulk, proposition 2.15. When $\mathrm{Gap}(W) > 0$ we saw that there is identifiability with no ambiguity of the eigenvalue λ_1^* of $\mathbf{T}_W$, being it the only one with multiplicity $d_1 = d$. As a consequence, there is a unique set of d eigenvalues of $\mathbf{T}^{(n)}$, which we remind is normalized to $1/\rho_n$ separated by at least $3/4\mathrm{Gap}(W)$ by the inequality on the d_2 distance we just found in the lemma (i.e. equation 2.5 combined with the further upper bound under E_n by $\mathrm{Gap}(W)/8$). By the triangular inequality, we have immediately that:

$$\frac{1}{\rho_n}\Delta \geqslant \frac{3}{4}\mathrm{Gap}(W), \tag{2.30}$$

where we defined Δ in equation 2.28. Using exactly this equation, we can link with a bound on the empirical eigenvalues, those from $\mathbf{A}^{(n)}$ in $\hat{\mathbf{V}}\hat{\mathbf{V}}^\top$, rescaled by $1/\rho_n$. There necessarily exist d eigenvalues $\boldsymbol{\lambda}_{\mathsf{special}}^{\mathbf{A}^{(n)}} := (\lambda_{i_j}^{\mathbf{A}^{(n)}})_{j=1}^d$ such that:

$$\left\| \hat{\mathbf{V}}\hat{\mathbf{V}}^\top - \mathbf{V}\mathbf{V} \right\|_{\mathsf{F}} \leqslant \frac{\mathrm{Gap}(W)}{8}, \qquad \hat{\mathbf{V}} \text{ associated to eigenvecs of } \boldsymbol{\lambda}_{\mathsf{special}}^{\mathbf{A}^{(n)}}, \text{ rescaled.} \tag{2.31}$$

It suffices to apply the classical Hoffman-Wielandt inequality now (see for example (Bhatia 1997, thm. VI.4.1)). It tells us that the squared distance between the spectrum is less than the Frobenius norm of the matrices, once we sort the spectrums in order. These are still upper bounded by $\mathrm{Gap}(W)/8$ then. With a further application of the triangle inequality we conclude that:[9]

$$\hat{\Delta} := \mathrm{d}(\lambda^{\mathbf{A}^{(n)}}_{\mathsf{special}}, \lambda^{\mathbf{A}^{(n)}} \backslash \lambda^{\mathbf{A}^{(n)}}_{\mathsf{special}}) \geqslant \frac{\rho_n \mathrm{Gap}(W)}{2}, \tag{2.32}$$

meaning that the special piece of the spectrum we chose is in its entirety separated from the rest by this factor. □

Proof sketch of algorithmic performance in theorem 2.16. Full details are in the original paper (Araya Valdivia and Yohann 2019). In the statement, we relate $\mathbf{G}^\star$ the true Gram matrix of distances with $\hat{\mathbf{G}}$ the distances estimated from the adjacency matrix $\mathbf{A}^{(n)}$. As we said, we will connect them passing by the same Gram matrix, now for $\mathbf{T}^{(n)}$. Morally, the first step is the simple triangle inequality:

$$\left\|\hat{\mathbf{G}} - \mathbf{G}^\star\right\|_{\mathsf{F}} \leqslant \left\|\hat{\mathbf{G}} - \mathbf{G}\right\|_{\mathsf{F}} + \left\|\mathbf{G} - \mathbf{G}^\star\right\|_{\mathsf{F}}. \tag{2.33}$$

The former is easy to control. We just need to reconnect with the eigendecomposition (just a rescaling) and use equation 2.28. Recall that $\gamma = d-2/d$ to find:

$$\left\|\hat{\mathbf{G}} - \mathbf{G}\right\|_{\mathsf{F}} = \frac{1}{c_1}\left\|\hat{\mathbf{V}}\hat{\mathbf{V}}^\top - \mathbf{V}\mathbf{V}^\top\right\|_{\mathsf{F}}, \qquad c_1 = \frac{d-2}{2} \tag{2.34}$$

and when n is large enough and the gap $\mathrm{Gap}(W)$ is positive:

$$\gamma\left\|\hat{\mathbf{V}}\hat{\mathbf{V}}^\top - \mathbf{V}\mathbf{V}^\top\right\|_{\mathsf{F}} \lesssim_\alpha C_W \frac{d-2}{d}\frac{\sqrt{d}}{\sqrt{n}} = C_W \frac{d-2}{\sqrt{dn}}. \tag{2.35}$$

Next, we bound the other term. The starting point is a decomposition of it into three further terms, which accounts for different contributions to the error. We do the following:

- build a projection matrix into the column span of $\mathbf{V}^\star$, i.e. $\mathbf{G}_{\mathsf{proj}} := \mathbf{V}^\star([\mathbf{V}^\star]^\top\mathbf{V}^\star)^{-1}[\mathbf{V}^\star]^\top$;
- build the approximation matrix at a radius R to choose later, i.e. $\mathbf{G}_R$ the Gram matrix for the eigenvectors of $\mathbf{T}^{(n)}_R := 1/n(W_R(\mathbf{x}_i, \mathbf{x}_j))_{i,j\in[n]}$.

We then start from a triangle inequality to highlight terms:

$$\left\|\mathbf{G} - \mathbf{G}^\star\right\|_{\mathsf{F}} \leqslant \left\|\mathbf{G} - \mathbf{G}_R\right\|_{\mathsf{F}} + \left\|\mathbf{G}_R - \mathbf{G}_{\mathsf{proj}}\right\|_{\mathsf{F}} + \left\|\mathbf{G}_{\mathsf{proj}} - \mathbf{G}^\star\right\|_{\mathsf{F}}. \tag{2.36}$$

The first term is easy, we truncate at $R = O\left((n/\log n)^{1/2\delta+d-1}\right)$ and apply Davis Kahan in the form of (Araya Valdivia 2020a, thm. 29). This gives:

$$\begin{aligned}\left\|\mathbf{G} - \mathbf{G}_R\right\|_{\mathsf{F}} &\leqslant C\frac{\left\|\mathbf{T}^{(n)} - \mathbf{T}^{(n)}_R\right\|_{\mathsf{F}}}{\Delta} && \text{Davis-Kahan} \quad (2.37)\\ &\leqslant \frac{C}{\Delta}\left(\frac{n}{\log n}\right)^{-\delta/2\delta+d-1}. && (2.38)\end{aligned}$$

For the third term, we use a representation of the Frobenius distance of an outer product matrix and its column rank projection (namely (Araya Valdivia 2020a, lem. 38)):

$$\left\|\mathbf{G}_{\mathsf{proj}} - \mathbf{G}^\star\right\|_{\mathsf{F}} = \left\|\boldsymbol{I}_d - [\mathbf{V}^\star]^\top\mathbf{V}^\star\right\|_{\mathsf{F}}, \tag{2.39}$$

and a concentration result for sub-gaussian outer products (Araya Valdivia 2020a, thm. 34) taken from (Vershynin 2010, prop. 2.1), to find:

$$\left\|\boldsymbol{I}_d - [\mathbf{V}^\star]^\top\mathbf{V}^\star\right\|_{\mathsf{F}} \lesssim_\alpha \frac{d}{\sqrt{n}}. \tag{2.40}$$

So for the third term we have:

$$\left\|\mathbf{G}_{\mathsf{proj}} - \mathbf{G}^\star\right\|_{\mathsf{F}} \lesssim_\alpha \frac{d}{\sqrt{n}}. \tag{2.41}$$

The middle term is the most intricate. We do not get into the details, but it requires:

[9] Here for simplicity we do not write that $\mathbf{A}^{(n)}$ is rescaled by $1/\rho_n$. One should take the normalized result and conclude with this last equation.

- a representation of Frobenius distance by Bhatia (1997, pg. 202);
- a perturbation result by Bhatia (1997, thm. VII.3.1) found in (Araya Valdivia 2020a, thm. 31);
- an application of Ostrowskii's inequality (see (Araya Valdivia 2020a, cor. 18)) combined with a further bound by Castro, Lacour, and Ngoc (2020, lem. 12).

The result is that:

$$\left\|\mathbf{G}_{\mathsf{proj}} - \mathbf{G}_R\right\|_{\mathsf{F}} \lesssim_\alpha \frac{1}{\mathrm{Gap}(W)} \left(\frac{n}{\log n}\right)^{-\delta/2\delta+d-1}. \tag{2.42}$$

Taking the four inequalities into account, the one most surviving asymptotically is the last one.

□

2.III *Ideas from proofs of the concentration results*

{subsec:el

The concentration results we used in the random geometric graph model have a common thread. Below, we will try to summarize it. The proofs of theorems 1.48 and 1.53 follow three steps:

1. approximation;
2. perturbation;
3. concentration.

APPROXIMATION Fix a truncation $R \in \mathbb{N}$. We decompose the kernel W into its best-in-L^2 rank R approximation and a residual. As for all nuclear norms, the Young-Eckart-Mirsky theorem tells us that the approximation is the truncation to the largest R singular values. When looking at $\mathbf{T}^{(n)}$, we will have then a residual matrix:

$$(\mathrm{E}_R^{(n)})_{ij} := \frac{1}{n} \sum_{k>R} \lambda_k \phi_k(\mathbf{x}_i)\phi_k(\mathbf{x}_j). \tag{2.43}$$

We want to decouple the expression for $\mathbf{T}^{(n)}$ into three objects:

- the rank R approximation;
- the projection of the first R eigenvectors of the residual;
- the projection of the other eigenvectors of the residual.

A good representation for this decomposition requires preliminary definitions. Let:

$$\begin{aligned}
\mathbf{\Phi}_R &:= \frac{1}{\sqrt{n}} \begin{bmatrix} \boldsymbol{\phi}_R(\mathbf{x}_1) \\ \vdots \\ \boldsymbol{\phi}_R(\mathbf{x}_n) \end{bmatrix} \in \mathbb{R}^{n\times R} && (2.44)\\
\boldsymbol{\phi}_R(\mathbf{x}_i) &:= [\phi_1(\mathbf{x}_i), \dots, \phi_R(\mathbf{x}_i)] \qquad \forall i \in [n]. && (2.45)\\
\mathbf{\Lambda}_R &:= \mathrm{diag}\{\lambda_1, \dots, \lambda_R\} && (2.46)\\
\mathbf{\Phi}_R^\perp &:= \{\text{orthonormal basis of orthogonal complement of } \mathbf{\Phi}_R\}. && (2.47)
\end{aligned}$$

Then we define projections onto the spaces spanned by vectors in $\mathbf{\Phi}_R$ and its orthogonal complement. These are:

$$\begin{aligned}
\mathbf{P}_R &:= \mathbf{\Phi}_R(\mathbf{\Phi}_R^\top \mathbf{\Phi}_R)^{-1}\mathbf{\Phi}_R^\perp && (2.48)\\
\mathbf{Q}_R &:= \mathbf{\Phi}_R^\perp([\mathbf{\Phi}_R^\perp]^\top \mathbf{\Phi}_R)^{-1}[\mathbf{\Phi}_R^\perp]^\top && (2.49)\\
&= \mathbf{\Phi}_R^\perp[\mathbf{\Phi}_R^\perp]^\top, && (2.50)
\end{aligned}$$

where we used that the perpendicular space is created with an orthonormal basis. A trivial decomposition of the residual matrix is according to the various projections onto these spaces, which is:

$$\mathbf{E}_R^{(n)} = \mathbf{Q}_R\mathbf{E}_R^{(n)}\mathbf{Q}_R + \mathbf{Q}_R\mathbf{E}_R^{(n)}\mathbf{P}_R + \mathbf{P}_R\mathbf{E}_R^{(n)}\mathbf{Q}_R + \mathbf{P}_R\mathbf{E}_R^{(n)}\mathbf{P}_R. \tag{2.51}$$

From this, we may start from the decomposition into truncation and residual to obtain a matrix form:

$$\begin{aligned}
\mathbf{T}^{(n)} &= \mathbf{\Phi}_R\mathbf{\Lambda}_R\mathbf{\Phi}_R^\top + \mathbf{E}_R^{(n)} && (2.52)\\
&= \mathbf{\Phi}_R\mathbf{\Lambda}_R\mathbf{\Phi}_R^\top + \mathbf{Q}_R\mathbf{E}_R^{(n)}\mathbf{Q}_R + \mathbf{Q}_R\mathbf{E}_R^{(n)}\mathbf{P}_R + \underbrace{\mathbf{P}_R\mathbf{E}_R^{(n)}\mathbf{Q}_R + \mathbf{P}_R\mathbf{E}_R^{(n)}\mathbf{P}_R}_{:=\mathbf{A}}. && (2.53)\\
&= \mathbf{\Phi}_R\mathbf{\Lambda}_R\mathbf{\Phi}_R^\top + \mathbf{\Phi}_R^\perp[\mathbf{\Phi}_R^\perp]^\top\mathbf{E}_R^{(n)}\mathbf{\Phi}_R^\perp[\mathbf{\Phi}_R^\perp]^\top + \mathbf{A} && (2.54)\\
&= \begin{bmatrix}\mathbf{\Phi}_R & \mathbf{\Phi}_R^\perp\end{bmatrix}\underbrace{\begin{bmatrix}\mathbf{\Lambda}_R & \mathbf{0}\\ \mathbf{0} & \underbrace{[\mathbf{\Phi}_R^\perp]^\top\mathbf{E}_R^{(n)}\mathbf{\Phi}_R^\perp}_{:=\mathbf{M}_{>R}}\end{bmatrix}}_{:=\mathbf{M}}\begin{bmatrix}\mathbf{\Phi}_R^\top\\ [\mathbf{\Phi}_R^\perp]^\top\end{bmatrix} + \mathbf{A} && (2.55)\\
&= \underbrace{\begin{bmatrix}\mathbf{\Phi}_R & \mathbf{\Phi}_R^\perp\end{bmatrix}\mathbf{M}\begin{bmatrix}\mathbf{\Phi}_R & \mathbf{\Phi}_R^\perp\end{bmatrix}^\top}_{\mathbf{M}_{\mathbf{\Phi}_R}} + \mathbf{A}, && (2.56)
\end{aligned}$$

which is just an algebraic manipulation.

PERTURBATION AND CONCENTRATION We apply Weyl's inequality seeing $\mathbf{A}$ as the perturbing matrix. This allows us to connect the eigenvalues of $\mathbf{M}_{\mathbf{\Phi}_R}$ and $\mathbf{T}^{(n)}$ with respect to the operator norm of $\mathbf{A}$. Now, we want to recover a result with $\mathbf{M}$ instead of $\mathbf{M}_{\Phi_R}$, so we can apply a multiplicative inequality such as Ostrowskii's.[10] In particular, we will bound the distance of eigenvalues of $\mathbf{M}_{\mathbf{\Phi}_R}$ and $\mathbf{M}$. Putting the two together, we obtain:

$$|\lambda_i^{\mathbf{T}^{(n)}} - \lambda_i^{\mathbf{M}}| \leqslant |\lambda_i(\mathbf{M})| \left\| \mathbf{\Phi}_R^\top\mathbf{\Phi}_R - \boldsymbol{I}_d \right\|_{\mathsf{op}} + \|\mathbf{A}\|_{\mathsf{op}}, \qquad \forall i \in [n] \qquad (2.57) \quad \texttt{\{\{eqn:firs}$$

Moreover, we know the "high modes" above R part of the spectrum of $\mathbf{M}$ obeys the following inequality (operator norm is larger than eigenvalues):

$$\lambda_i^{\mathbf{M}_{>R}} \leqslant \|\mathbf{M}_{>R}\|_{\mathsf{op}} = \left\| [\mathbf{\Phi}_R^\perp]^\top\mathbf{E}^{(n)}\mathbf{\Phi}_R^\perp \right\|_{\mathsf{op}}, \qquad \forall i \in [n-R]. \qquad (2.58) \quad \texttt{\{\{eqn:seco}$$

In the concentration step, Araya Valdivia (2020b) upper bounds further the terms in the RHS of equation 2.57 and extracts results for the high-modes spectrum in equation 2.58. While some terms are standard, others require some non-classical concentration results and techniques (see the comments in (Araya Valdivia 2020b, pg. 10)). We do not provide details for the sake of space.

REFERENCES

Araya Valdivia, Ernesto (2020a). "Kernel Spectral Learning and Inference in Random Geometric Graphs". PhD thesis (cit. on pp. 1–3, 5–13).

— (Oct. 2020b). *Relative Concentration Bounds for the Spectrum of Kernel Matrices*. DOI: `10.48550/arXiv.1812.02108`. arXiv: `1812.02108 [stat]` (cit. on pp. 1, 14).

Araya Valdivia, Ernesto and De Castro Yohann (2019). "Latent Distance Estimation for Random Geometric Graphs". In: *Advances in Neural Information Processing Systems*. Vol. 32. Curran Associates, Inc. (cit. on pp. 1, 8, 12).

Bandeira, Afonso S. and Ramon van Handel (July 2016). "Sharp Nonasymptotic Bounds on the Norm of Random Matrices with Independent Entries". In: *The Annals of Probability* 44.4. ISSN: 0091-1798. DOI: `10.1214/15-AOP1025`. arXiv: `1408.6185 [math]` (cit. on pp. 8, 11).

Bhatia, Rajendra (1997). *Matrix Analysis*. Vol. 169. Graduate Texts in Mathematics. New York, NY: Springer. ISBN: 978-1-4612-6857-4 978-1-4612-0653-8. DOI: `10.1007/978-1-4612-0653-8` (cit. on pp. 12, 13).

Castro, Yohann De, Claire Lacour, and Thanh Mai Pham Ngoc (Apr. 2020). *Adaptive Estimation of Nonparametric Geometric Graphs*. DOI: `10.48550/arXiv.1708.02107`. arXiv: `1708.02107 [math]` (cit. on pp. 4, 11, 13).

Dai, Feng and Yuan Xu (2013). *Approximation Theory and Harmonic Analysis on Spheres and Balls*. Springer Monographs in Mathematics. New York, NY: Springer. ISBN: 978-1-4614-6659-8 978-1-4614-6660-4. DOI: `10.1007/978-1-4614-6660-4` (cit. on pp. 4, 5, 8).

[10] By multiplicative we mean an analog of Weyl's but for multiplications of matrices.

Emery, Michel, Arkadi Nemirovski, and Dan Voiculescu (2000). *Lectures on Probability Theory and Statistics*. Ed. by Pierre Bernard. Vol. 1738. Lecture Notes in Mathematics. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 978-3-540-67736-9 978-3-540-45029-0. DOI: 10.1007/BFb0106703 (cit. on p. 10).

Hirsch, Francis and Gilles Lacombe (1999). *Elements of Functional Analysis*. Graduate Texts in Mathematics 192. New York, NY: Springer. ISBN: 978-1-4612-7146-8 978-1-4612-1444-1. DOI: 10.1007/978-1-4612-1444-1 (cit. on p. 2).

Koltchinskii, Vladimir and Evarist Giné (Feb. 2000). "Random Matrix Approximation of Spectra of Integral Operators". In: *Bernoulli* 6.1, pp. 113–167. ISSN: 1350-7265 (cit. on p. 3).

Nicaise, Serge (May 2000). "Jacobi Polynomials, Weighted Sobolev Spaces and Approximation Results of Some Singularities". In: *Mathematische Nachrichten* 213.1, pp. 117–140. ISSN: 0025-584X, 1522-2616. DOI: 10.1002/(SICI)1522-2616(200005)213:1<117::AID-MANA117>3.0.CO;2-A (cit. on p. 4).

Vershynin, Roman (Dec. 2010). *How Close Is the Sample Covariance Matrix to the Actual Covariance Matrix?* DOI: 10.48550/arXiv.1004.3484. arXiv: 1004.3484 [math] (cit. on p. 12).