Journal of Machine Learning Research () Submitted 12/04; Published

On Learning with Integral Operators

Lorenzo Rosasco, LROSASCO@MIT.EDU
Center for Biological and Computational Learning, MIT,
Cambridge, MA, USA
& Dipartimento di Informatica e Scienze dell'Informazione,
Università di Genova, Italy

Mikhail Belkin, MBELKIN@CSE.OHIO-STATE.EDU
Department of Computer Science and Engineering,
Ohio State University, U.S.A.

Ernesto De Vito DEVITO@DIMA.UNIGE.IT
Dipartimento di Scienze per l'Architettura,
Università di Genova, Italy
& INFN, Sezione di Genova, Italy

Editor:

Abstract

A large number of learning algorithms, for example, spectral clustering, kernel Principal Components Analysis and many manifold methods are based on estimating eigenvalues and eigenfunctions of operators defined by a similarity function or a kernel, given empirical data. Thus for the analysis of algorithms, it is an important problem to be able to assess the quality of such approximations. The contribution of our paper is two-fold:
1. We use a technique based on a concentration inequality for Hilbert spaces to provide new much simplified proofs for a number of results in spectral approximation.
2. Using these methods we provide several new results for estimating spectral properties of the graph Laplacian operator extending and strengthening results from von Luxburg et al. (2008).

Keywords: spectral convergence, empirical operators, learning integral operators, perturbation methods

1. Introduction

A broad variety of methods for machine learning and data analysis from Principal Components Analysis (PCA) to Kernel PCA, Laplacian-based spectral clustering and manifold methods, rely on estimating eigenvalues and eigenvectors of certain data-dependent matrices. In many cases these matrices can be interpreted as empirical versions of underlying integral operators or closely related objects, such as continuous Laplacian operators. Thus establishing connections between empirical operators and their continuous counterparts is essential to understanding these algorithms. In this paper, we propose a method for analyzing empirical operators based on concentration inequalities in Hilbert spaces. This

technique together with perturbation theory results allows us to derive a number of results on spectral convergence in an exceptionally simple way. We note that the approach using concentration inequalities in a Hilbert space has already been proved useful for analyzing supervised kernel algorithms, see De Vito et al. (2005b), Yao et al. (2007), Bauer et al. (2007), Smale and Zhou (2005). Here we develop on this approach to provide a detailed and comprehensive study of perturbation results for empirical estimates of integral operators as well as empirical graph Laplacians.

In recent years several works started considering these connections. The first study of this problem appeared in Koltchinskii and Giné (2000), Koltchinskii (1998), where the authors consider integral operators defined by a kernel. In Koltchinskii and Giné (2000) the authors study the relation between the spectrum of an integral operator with respect to a probability distribution and its (modified) empirical counterpart in the framework of U-statistics. In particular they prove that the ℓ_2 distance between the two (ordered) spectra goes to zero under the assumption that the kernel is symmetric and square integrable. Moreover, under some stronger conditions, rates of convergence and distributional limit theorems are obtained. The results are based on an inequality due to Lidskii and to Wielandt for finite dimensional matrices and the Marcinkiewicz law of large numbers. In Koltchinskii (1998) similar results were obtained for convergence of eigenfunctions and, using the triangle inequality, for spectral projections. These investigations were continued in Mendelson and Pajor (2005, 2006), where it was shown that, under the assumption that the kernel is of positive type, the problem of eigenvalue convergence reduces to the study of how the random operator $\frac{1}{n}\sum_{i=1}^{n} X_i \otimes X_i$ deviates from its average $\mathbb{E}[X \otimes X]$, with respect to the operator norm, where $X, X_1, \ldots, X_n$ are i.i.d ℓ_2 random vectors. The result is based on a symmetrization technique and on the control of a suitable Radamacher complexity.
The above studies are related to the problem of consistency of kernel PCA considered in Shawe-Taylor et al. (2002, 2004) and refined in Zwald et al. (2004), Zwald and Blanchard (2006). In particular, Shawe-Taylor et al. (2002, 2004) study the deviation of the sum of the all but the largest k eigenvalues of the empirical matrix to its mean using McDiarmid inequality. The above result is improved in Zwald et al. (2004) where fast rates are provided by means of a localized Rademacher complexities approach. The results in Zwald and Blanchard (2006) are a development of the results in Koltchinskii (1998). Using a new perturbation result the authors study directly the convergence of the whole subspace spanned by the first k eigenvectors and are able to show that only the gap between the k and $k+1$ eigenvalue affects the estimate. All the above results hold for symmetric, positive definite kernels.

A second related series of works considered convergence of the graph Laplacian in various settings , see for example Belkin (2003), Lafon (2004), Belkin and Niyogi (2005), Hein et al. (2005), Hein (2006), Singer (2006), Giné and Koltchinskii (2006). These papers discuss convergence of the graph Laplacian directly to the Laplace-Beltrami operator. Convergence of the normalized graph Laplacian applied to a fixed smooth function on the manifold is discussed in Hein et al. (2005), Singer (2006), Lafon (2004). Results showing uniform convergence over some suitable class of test functions are presented in Hein (2006), Giné and Koltchinskii (2006). Finally, convergence of eigenvalues and eigenfunctions for the case of the uniform distribution was shown in Belkin and Niyogi (2007).

Unlike these works, where the kernel function is chosen adaptively depending on the number of points, we will be primarily interested in convergence of the graph Laplacian to its continuous (population) counterpart for a *fixed* weight function. Von Luxburg et al. (2004) study the convergence of the second eigenvalue which is relevant in spectral clustering problems. These results are extended in von Luxburg et al. (2008), where operators are defined on the space of continuous functions. The analysis is performed in the context of perturbation theory in Banach spaces and bounds on individual eigenfunctions are derived. The problem of out-of-sample extension is considered via a Nyström approximation argument. By working in Banach spaces the authors have only mild requirements for the weight function defining the graph Laplacian, at the price of having to do a fairly complicated analysis.

Our contribution is twofold. In the first part of the paper, we assume that the kernel K is symmetric and positive definite. We start considering the problem of out-of-sample extension of the kernel matrix and discuss a singular value decomposition perspective on Nyström-like extensions. More precisely, we show that a finite rank (extension) operator acting on the Reproducing Kernel Hilbert space $\mathcal{H}$ defined by K can be naturally associated with the empirical kernel matrix: the two operators have same eigenvalues and related eigenvectors/eigenfunctions. The kernel matrix and its extension can be seen as compositions of suitable restriction and extension operators that are explicitly defined by the kernel. A similar result holds true for the asymptotic integral operator, whose restriction to $\mathcal{H}$ is a Hilbert-Schmidt operator. We can use concentration inequalities for operator valued random variables and perturbation results to derive concentration results for eigenvalues (taking into account the multiplicity), as well as for the sums of eigenvalues. Moreover, using a perturbation result for spectral projections, we derive finite sample bounds for the deviation between the spectral projection associated with the k largest eigenvalues. We recover several known results with simplified proofs, and derive new results.

In the second part of the paper, we study the convergence of the asymmetric normalized graph Laplacian to its continuous counterpart. To this aim we consider a fixed positive symmetric weight function satisfying some smoothness conditions. These assumptions allow us to introduce a suitable intermediate Reproducing Kernel Hilbert space $\mathcal{H}$, which is, in fact, a Sobolev space. We describe explicitly restriction and extension operators and introduce a finite rank operator with spectral properties related to those of the graph Laplacian. Again we consider the law of large numbers for operator-valued random variables to derive concentration results for empirical operators. We study behavior of eigenvalues as well as the deviation of the corresponding spectral projections with respect to the Hilbert-Schmidt norm. To obtain explicit estimates for spectral projections we generalize the perturbation result in Zwald and Blanchard (2006) to deal with non-self-adjoint operators. From a technical point the main difficulty in studying the asymmetric graph Laplacian is that we no longer assume the weight function to be positive definite so that there is no longer a natural Reproducing Kernel Hilbert space space associated with it. In this case we have to deal with non-self-adjoint operators and the functional analysis becomes more involved. Comparing to von Luxburg et al. (2008), we note that the RKHS $\mathcal{H}$ replaces the Banach space of continuous functions. Assuming some regularity assumption on the weight functions we can exploit the Hilbert space structure to obtain more explicit results. Among other things, we derive explicit convergence rates for a large class of weight functions. Finally we note

that for the case of positive definite weight functions results similar to those presented here have been independently derived by Smale and Zhou (2008).

The paper is organized as follows. We start by introducing the necessary mathematical objects in Section 2. We recall some facts about the properties of linear operators in Hilbert spaces, such as their spectral theory and some perturbation results, and discuss some concentration inequalities in Hilbert spaces. This technical summary section aims at making this paper self-contained and provide the reader with a (hopefully useful) overview of the needed tools and results. In Section 3, we study the spectral properties of kernel matrices generated from random data. We study concentration of operators obtained by an out-of-sample extension using the kernel function and obtain considerably simplified derivations of several existing results on eigenvalues and eigenfunctions. We expect that these techniques will be useful in analyzing algorithms requiring spectral convergence. In fact, in Section 4, we apply these methods to prove convergence of eigenvalues and eigenvectors of the asymmetric graph Laplacian defined by a fixed weight function. We refine the results in von Luxburg et al. (2008), which, to the best of our knowledge, is the only other paper considering this problem so far.

2. Notation and preliminaries.

In this section we will discuss various preliminary results necessary for the further development.

2.1 Operator theory

We first recall some basic notions in operator theory, see for example Lang (1993). In the following we let $A : \mathcal{H} \to \mathcal{H}$ be a (linear) bounded operator, where $\mathcal{H}$ is a complex (separable) Hilbert space with scalar product (norm) $\langle \cdot, \cdot \rangle$ ($\|\cdot\|$) and $(e_j)_{j\geq 1}$ a Hilbert basis in $\mathcal{H}$. We often use the notation $j \geq 1$ to denote a sequence or a sum from 1 to p where p can be infinite. The set of bounded operators on $\mathcal{H}$ is a Banach space with respect to the operator norm $\|A\|_{\mathcal{H},\mathcal{H}} = \|A\| = \sup_{\|f\|=1} \|Af\|$. If A is a bounded operator, we let A^* be its adjoint, which is a bounded operator with $\|A^*\| = \|A\|$.
A bounded operator A is Hilbert-Schmidt if $\sum_{j\geq 1} \|Ae_j\|^2 < \infty$ for some (any) Hilbert basis $(e_j)_{j\geq 1}$. The space of Hilbert-Schmidt operators is also a Hilbert space (a fact which will be a key in our development) endowed with the scalar product $\langle A, B \rangle_{HS} = \sum_j \langle Ae_j, Be_j \rangle$ and we denote by $\|\cdot\|_{HS}$ the corresponding norm. In particular, Hilbert-Schmidt operators are compact.

A closely related notion is that of a *trace class* operator. We say that a bounded operator A is trace class, if $\sum_{j\geq 1} \left\langle \sqrt{A^*A} e_j, e_j \right\rangle < \infty$ for some (any) Hilbert basis $(e_j)_{j\geq 1}$ (where $\sqrt{A^*A}$ is the square root of the positive operator A^*A defined by spectral theorem, see for example Lang (1993). In particular, $\mathrm{Tr}(A) = \sum_{j\geq 1} \langle Ae_j, e_j \rangle < \infty$ and $\mathrm{Tr}(A)$ is called the trace of A. The space of trace class operators is a Banach space endowed with the norm $\|A\|_{TC} = \mathrm{Tr}(\sqrt{A^*A})$. Trace class operators are also Hilbert Schmidt (hence compact). The following inequalities relate the different operator norms:

$$\|A\| \leq \|A\|_{HS} \leq \|A\|_{TC}.$$

It can also be shown that for any Hilbert-Schmidt operator A and bounded operator B we have

$$\begin{aligned} \|AB\|_{HS} &\leq \|A\|_{HS}\|B\|, \\ \|BA\|_{HS} &\leq \|B\|\|A\|_{HS}. \end{aligned} \tag{1}$$

Remark 1 *If the context is clear we will simply denote the norm and the scalar product by $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ respectively. However, we will add a subscript when comparing norms in different spaces. When A is a bounded operator, $\|A\|$ always denotes the operator norm.*

2.2 Spectral Theory for Compact Operators

Recall that the spectrum of a matrix K can be defined as the set of eigenvalues $\lambda \in \mathbb{C}$, s.t. $\det(K - \lambda I) = 0$, or, equivalently, such that $\lambda I - K$ does not have a (bounded) inverse. This definition can be generalized to operators. Let $A : \mathcal{H} \to \mathcal{H}$ be a bounded operator, we say that $\lambda \in \mathbb{C}$ belongs to the spectrum $\sigma(A)$, if $(A - \lambda I)$ does not have a bounded inverse. For any $\lambda \notin \sigma(A)$, $R(\lambda) = (A - \lambda I)^{-1}$ is the *resolvent operator*, which is by definition a bounded operator. If A is a compact operator, then $\sigma(A) \setminus \{0\}$ consists of a countable family of isolated points with finite multiplicity $|\lambda_1| \geq |\lambda_2| \geq \cdots$ and either $\sigma(A)$ is finite or $\lim_{n\to\infty} \lambda_n = 0$, see for example Lang (1993).

If the bounded operator A is self-adjoint ($A = A^*$, analogous to a symmetric matrix in the finite-dimensional case), the eigenvalues are real. Each eigenvalue λ has an associated *eigenspace* which is the closed subspace of all eigenvectors with eigenvalue λ. A key result, known as the *Spectral Theorem*, ensures that

$$A = \sum_{i=1}^{\infty} \lambda_i P_{\lambda_i},$$

where P_λ is the *orthogonal projection operator* onto the eigenspace associated with λ. Moreover, it can be shown that the projection P_λ can be written explicitly in terms of the resolvent operator. Specifically, we have the following remarkable equality:

$$P_\lambda = \frac{1}{2\pi i} \int_{\Gamma \subset \mathbb{C}} (\gamma I - A)^{-1} d\gamma,$$

where the integral can be taken over any closed simple rectifiable curve $\Gamma \subset \mathbb{C}$ (with positive direction) containing λ and no other eigenvalue. We note that while an integral of an operator-valued function may seem unfamiliar, it is defined along the same lines as an integral of an ordinary real-valued function. Despite the initial technicality, the above equation allows for far simpler analysis of eigenprojections than other seemingly more direct methods.

This analysis can be extended to operators, which are not self-adjoint, to obtain a decomposition parallel to the Jordan canonical form for matrices. To avoid overloading this section, we postpone the precise technical statements for that case to the Appendix B.

Remark 2 *Though in manifold and spectral learning we typically work with real valued functions, in this paper we will consider complex vector spaces to be able to use certain*

results from the spectral theory of (possibly non self-adjoint) operators. However, if the reproducing kernel and the weight function are both real valued, as usually is the case in machine learning, we will show that all functions of interest are real valued as well.

2.3 Reproducing Kernel Hilbert Space (RKHS)

Let X be a subset of $\mathbb{R}^d$. A *Reproducing Kernel Hilbert space* is a Hilbert space $\mathcal{H}$ of functions $f : X \to \mathbb{C}$, such that all the evaluation functionals are bounded, that is

$$f(x) \leq C_x \|f\| \qquad \text{for some constant } C_x.$$

It can be shown that there is a unique conjugate symmetric positive definite kernel function $K : X \times X \to \mathbb{C}$, called *reproducing kernel*, associated with $\mathcal{H}$ and the following reproducing property holds

$$f(x) = \langle f, K_x \rangle \,, \tag{2}$$

where $K_x := K(\cdot, x)$. It is also well known (Aronszajn, 1950) that any conjugate symmetric positive definite kernel K uniquely defines a reproducing kernel Hilbert space whose reproducing kernel is K. We will assume that the kernel is continuous and bounded, so that the elements of $\mathcal{H}$ are bounded continuous functions, the space $\mathcal{H}$ is separable and is compactly embedded in $\mathcal{C}(X)$ with the compact-open topology (Aronszajn, 1950).

Remark 3 *The set X can be taken to be any locally compact separable metric space and the assumption about continuity can be weakened. However, the above setting will simplify some technical considerations, in particular in Section 4.2 where Sobolev spaces are considered.*

2.4 Concentration Inequalities in Hilbert spaces

We recall that if $\xi_1, \ldots, \xi_n$ are independent (real-valued) random variables with zero mean and such that $|\xi_i| \leq C, \quad i = 1, \ldots, n$, then Hoeffding inequality ensures that $\forall \varepsilon > 0$,

$$\mathrm{P}\left[\left| \frac{1}{n} \sum_{i=1}^{n} \xi_i \right| \geq \varepsilon \right] \leq 2e^{-\frac{n\varepsilon^2}{2C^2}}.$$

If we set $\tau = \frac{n\varepsilon^2}{2C^2}$ then we can express the above inequality saying that with probability at least (with confidence) $1 - 2e^{-\tau}$,

$$\left| \frac{1}{n} \sum_{i=1}^{n} \xi_i \right| \leq \frac{C\sqrt{2\tau}}{\sqrt{n}}. \tag{3}$$

Similarly if $\xi_1, \ldots, \xi_n$ are zero mean independent random variables with values in a separable complex Hilbert space and such that $\|\xi_i\| \leq C, \quad i = 1, \ldots, n$, then the same inequality holds with the absolute value replaced by the norm in the Hilbert space, that is, the following bound

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \xi_i \right\| \leq \frac{C\sqrt{2\tau}}{\sqrt{n}} \tag{4}$$

holds true with probability at least $1 - 2e^{-\tau}$ (Pinelis, 1992).

Remark 4 *In the cited reference the concentration inequality* (4) *is stated for real Hilbert spaces. However, a complex Hilbert space $\mathcal{H}$can be viewed as a real vector space with the scalar product given by $\langle f, g\rangle_{\mathcal{H}_\mathbb{R}} = (\langle f, g\rangle_{\mathcal{H}} + \langle g, f\rangle_{\mathcal{H}})/2$, so that $\|f\|_{\mathcal{H}_\mathbb{R}} = \|f\|_{\mathcal{H}}$. This last equality implies that* (4) *holds also for complex Hilbert spaces.*

2.5 Perturbation theory

First we recall some results on perturbation of eigenvalues and eigenfunctions. About eigenvalues, we need to recall the notion of *extended enumeration* of discrete eigenvalues. We adapt the definition of Kato (1987), which is given for an arbitrary self-adjoint operator, to the compact operators. Let $A : \mathcal{H} \to \mathcal{H}$ be a compact operator, an extended enumeration is a sequence of real numbers where every nonzero eigenvalue of A appears as many times as its multiplicity and the other values (if any) are zero. An enumeration is an extended numeration where any element of the sequence is an isolated eigenvalue with finite multiplicity. If the sequence is infinite, this last condition is equivalent to the fact that any element is nonzero.

The following result due to Kato (1987) is an extension to infinite dimensional operators of an inequality due to Lidskii for finite rank operator.

Theorem 5 (Kato 1987) *Let $\mathcal{H}$ be a separable Hilbert space with A, B self-adjoint compact operators. Let $(\gamma_j)_{j\geq 1}$, be an enumeration of discrete eigenvalues of $B - A$, then there exist extended enumerations $(\beta_j)_{j\geq 1}$ and $(\alpha_j)_{j\geq 1}$ of discrete eigenvalues of B and A respectively such that,*

$$\sum_{j\geq 1} \phi(\beta_j - \alpha_j) \leq \sum_{j\geq 1} \phi(\gamma_j),$$

where ϕ is any nonnegative convex function with $\phi(0) = 0$.

By choosing $\phi(t) = |t|^p$, $p \geq 1$, the above inequality becomes

$$\sum_{j\geq 1} |\beta_j - \alpha_j|^p \leq \sum_{j\geq 1} |\gamma_j|^p.$$

Letting $p = 2$ and recalling that $\|B - A\|_{HS}^2 = \sum_{j\geq 1} |\gamma_j|^2$, it follows that

$$\sum_{j\geq 1} |\beta_j - \alpha_j|^2 \leq \|B - A\|_{HS}^2.$$

Moreover, since $\lim_{p\to\infty}(\sum_{j\geq 1} |\gamma_j|^p)^{1/p} = \sup_{j\geq 1} |\gamma_j| = \|B - A\|$, we have that

$$\sup_{j\geq 1} |\beta_j - \alpha_j| \leq \|B - A\|.$$

Given an integer N, let m_N be the sum of the multiplicities of the first N nonzero top eigenvalues of A, it is possible to prove that the sequences $(\alpha_j)_{j\geq 1}$ and $(\beta_j)_{j\geq 1}$ in the above proposition can be chosen in such a way that

$$\begin{aligned} &\alpha_1 \geq \sigma_2 \geq \ldots \geq \alpha_{m_N} > \alpha_j && j > m_N \\ &\beta_1 \geq \beta_2 \geq \ldots \geq \beta_{m_N} \geq \beta_j && j > m_N. \end{aligned}$$

However in general we need to consider extended enumerations, which are not necessarily decreasing sequence, in order to take into account the kernel spaces of A and B, which are potentially infinite dimensional vector spaces (also see the remark after Theorem II in Kato (1987)).

To control the spectral projections associated with one or more eigenvalues we need the following perturbation result due to Zwald and Blanchard (2006) (see also Theorem 20 in Section 4.3). Let A be a positive compact operator such that $\sigma(A)$ is infinite. Given $N \in \mathbb{N}$, let P_N^A be the orthogonal projection on the eigenvectors corresponding to the top N distinct eigenvalues $\alpha_1 > \ldots > \alpha_N$ and α_{N+1} the next one.

Proposition 6 *Let A be a compact positive operator. Given an integer N, if B is another compact positive operator such that $\|A - B\| \leq \frac{\alpha_N - \alpha_{N+1}}{4}$, then*

$$\|P_D^B - P_N^A\| \leq \frac{2}{\alpha_N - \alpha_{N+1}} \|A - B\|$$

where the integer D is such that the dimension of the range of P_D^B is equal to the dimension of the range of P_N^A. If A and B are Hilbert-Schmidt, in the above bound the operator norm can be replaced by the Hilbert-Schmidt norm.

We note that a bound on the projections associated with simple eigenvalues implies that the corresponding eigenvectors are close since, if u and v are taken to be normalized and such that $\langle u, v \rangle > 0$, then the following inequality holds

$$\|P_u - P_v\|_{HS}^2 = 2(1 - \langle u, v \rangle^2) \geq 2(1 - \langle u, v \rangle) = \|u - v\|_{\mathcal{H}}^2.$$

3. Integral Operators defined by a Reproducing Kernel

Let X be a subset of $\mathbb{R}^d$ and $K : X \times X \to \mathbb{C}$ be a reproducing kernel satisfying the assumptions stated in Section 2.3. Let ρ be a probability measure on X and denote by $L^2(X, \rho)$ the space of square integrable (complex) functions with norm $\|f\|_\rho^2 = \langle f, f \rangle_\rho = \int_X |f(x)|^2 d\rho(x)$. Since $K(x, x) \leq \kappa$ by assumption, the corresponding integral operator $L_K : L^2(X, \rho) \to L^2(X, \rho)$

$$(L_K f)(x) = \int_X K(x, s) f(s) d\rho(s) \tag{5}$$

is a bounded operator.

Suppose we are now given a set of points $\mathbf{x} = (x_1, \ldots, x_n)$ sampled i.i.d. according to ρ. Many problems in statistical data analysis and machine learning deal with the empirical kernel $n \times n$-matrix $\mathbf{K}$ given by $\mathbf{K}_{ij} = \frac{1}{n} K(x_i, x_j)$. The question we want to discuss is to which extent we can use the kernel matrix $\mathbf{K}$ (and the corresponding eigenvalues, eigenvectors) to estimate L_K (and the corresponding eigenvalues, eigenfunctions). Answering this question is important as it guarantees that the computable empirical proxy is sufficiently close to the ideal infinite sample limit.

The first difficulty in relating L_K and $\mathbf{K}$ is that they operate on different spaces. By default, L_K is an operator on $L^2(X, \rho)$, while $\mathbf{K}$ is a finite dimensional matrix. To overcome this

difficulty we let $\mathcal{H}$ be the RKHS associated with K and define the operators $T_{\mathcal{H}}, T_n : \mathcal{H} \to \mathcal{H}$ given by

$$T_{\mathcal{H}} = \int_X \langle \cdot, K_x \rangle K_x d\rho(x), \tag{6}$$

$$T_n = \frac{1}{n} \sum_{i=1}^{n} \langle \cdot, K_{x_i} \rangle K_{x_i}. \tag{7}$$

Note that $T_{\mathcal{H}}$ is the integral operator with kernel K with range and domain $\mathcal{H}$ rather than in $L^2(X, \rho)$. The reason for writing it in this seemingly complicated form is to make the parallel with (7) clear. To justify the "extension operator" in (7), consider the natural "restriction operator[1] ", $R_n : \mathcal{H} \to \mathbb{C}^n$, $R_n f = (f(x_1), \ldots, f(x_n))$. It is not hard to check that the adjoint operator $R_n^* : \mathbb{C}^n \to \mathcal{H}$ can be written as $R_n^*(y_1, \ldots, y_n) = \frac{1}{n} \sum_{i=1}^n y_i K_{x_i}$. Indeed, we see that

$$\begin{aligned} \langle R_n^*(y_1, \ldots, y_n), f \rangle_{\mathcal{H}} &= \langle (y_1, \ldots, y_n), R_n f \rangle_{\mathbb{C}^n} \\ &= \frac{1}{n} \sum_{i=1}^{n} y_i \overline{f(x_i)} = \frac{1}{n} \sum_{i=1}^{n} y_i \langle K_{x_i}, f \rangle_{\mathcal{H}}, \end{aligned}$$

where $\mathbb{C}^n$ is endowed with $1/n$ times the canonical scalar product. Thus, we observe that $T_n = R_n^* R_n$ is the composition of the restriction operator and its adjoint. On the other hand for the matrix $\mathbf{K}$ we have that $\mathbf{K} = R_n R_n^*$, regarded as operator on $\mathbb{C}^n$. Similarly, if $R_{\mathcal{H}}$ denotes the inclusion $\mathcal{H} \hookrightarrow L^2(X, \rho)$, $T_{\mathcal{H}} = R_{\mathcal{H}}^* R_{\mathcal{H}}$ and $L_K = R_{\mathcal{H}} R_{\mathcal{H}}^*$.

In the next subsection, we discuss a parallel with the Singular Value Decomposition for matrices and show that $T_{\mathcal{H}}$ and L_K have the same eigenvalues (possibly, up to some zero eigenvalues) and the corresponding eigenfunctions are closely related. A similar relation holds for T_n and $\mathbf{K}$. Thus to establish a connection between the spectral properties of $\mathbf{K}$ and L_K, it is sufficient to bound the difference $T_{\mathcal{H}} - T_n$, which is done in the following theorem (De Vito et al., 2005b). While the proof can be found in De Vito et al. (2005b), we provide it for completeness and to emphasize its simplicity.

Theorem 7 *The operators $T_{\mathcal{H}}$ and T_n are Hilbert-Schmidt. Under the above assumption with confidence $1 - 2e^{-\tau}$*

$$\|T_{\mathcal{H}} - T_n\|_{HS} \leq \frac{2\sqrt{2}\kappa\sqrt{\tau}}{\sqrt{n}}.$$

Proof We introduce a sequence $(\xi_i)_{i=1}^n$ of random variables in the Hilbert space of Hilbert-Schmidt operators by

$$\xi_i = \langle \cdot, K_{x_i} \rangle K_{x_i} - T_{\mathcal{H}}.$$

From (6) follows that $E(\xi_i) = 0$. By a direct computation we have that $\|\langle \cdot, K_x \rangle K_x\|_{HS}^2 = \|K_x\|^4 \leq \kappa^2$. Hence, using (6), $\|T_{\mathcal{H}}\|_{HS} \leq \kappa$ and

$$\|\xi_i\|_{HS} \leq 2\kappa, \qquad i = 1, \ldots, n.$$

1. R_n is also called sampling or evaluation operator. We prefer to call it the *restriction operator* since $R_n f$ is the restriction of the function $f : X \to \mathbb{R}$ to the set of points $\{x_1, \ldots, x_n\}$.

From inequality (4) we have with probability $1 - 2e^{-\tau}$

$$\|\frac{1}{n}\sum_{i=1}^{n}\xi_i\|_{HS} = \|T_{\mathcal{H}} - T_n\|_{HS} \leq \frac{2\sqrt{2}\kappa\sqrt{\tau}}{\sqrt{n}},$$

which establishes the result. ■

As an immediate consequence of Theorem 7 we obtain several concentration results for eigenvalues and eigenfunctions discussed in subsection 3.2. However, before doing that we provide an interpretation of the Nyström extension based on SVD and needed to properly compare the empirical operator and his mean.

3.1 Extension operators

We will now briefly revisit the Nyström extension and clarify some connections to the Singular Value Decomposition (SVD) for operators. Recall that applying SVD to a $p \times m$ matrix A produces a *singular system* consisting of singular (strictly positive) values $(\sigma_j)_{j=1}^{k}$ with k being the rank of A, vectors $(u_j)_{j=1}^{m} \in \mathbb{C}^m$ and $(v_j)_{j=1}^{p} \in \mathbb{C}^p$ such that they form orthonormal bases of $\mathbb{C}^m$ and $\mathbb{C}^p$ respectively, and

$$\begin{cases} A^*Au_j = \sigma_j u_j & j = 1, \dots k \\ A^*Au_j = 0 & j = k+1, \dots, m \\ AA^*v_j = \sigma_j v_j & j = 1, \dots k \\ AA^*v_j = 0 & j = k+1, \dots, p. \end{cases}$$

It is not hard to see that the matrix A can be written as $A = V\Sigma U^*$, where U and V are matrices obtained by "stacking" u's and v's in the columns, and Σ is a $p \times m$ matrix having the singular values σ_i on the first k-entries on the diagonal (and zero outside), so that

$$\begin{cases} Au_j = \sqrt{\sigma_j}v_j & j = 1, \dots k \\ Au_j = 0 & j = k+1, \dots, m \\ A^*v_j = \sqrt{\sigma_j}u_j & j = 1, \dots k \\ A^*v_j = 0 & j = k+1, \dots, p, \end{cases}$$

which is the formulation we will use in this paper. The same formalism applies more generally to operators and allows us to connect the spectral properties of L_K and $T_{\mathcal{H}}$ as well as the matrix $\mathbf{K}$ and the operator T_n. The basic idea is that each of these pairs (as shown in the previous subsection) corresponds to a singular system and thus share eigenvalues (up to some zero eigenvalues) and have eigenvectors related by a simple equation. Indeed the following result can be obtained considering the SVD decomposition associated with $R_{\mathcal{H}}$ (and proposition 9 considering the SVD decomposition associated with R_n). The proof of the following proposition can be deduced from the results in De Vito et al. (2005b, 2006)[2].

2. In De Vito et al. (2005b, 2006) the results are stated for real kernels, however the proof does not change if K is complex valued. Moreover, if K is real and L_K is regarded as integral operator on the space of square integrable complex functions, one can easily check that the eigenvalues are positive and, if u is an eigenfunction with eigenvalue $\sigma \geq 0$, then the complex conjugate $\overline{u}$ is also an eigenfunction with the same eigenvalue, so that it is always possible to choose all the eigenfunctionsto be real valued.

Proposition 8 *The following facts hold true.*

1. *The operators L_K and $T_{\mathcal{H}}$ are positive, self-adjoint and trace class. In particular both $\sigma(L_K)$ and $\sigma(T_{\mathcal{H}})$ are contained in $[0, \kappa]$.*

2. *The spectra of L_K and $T_{\mathcal{H}}$ are the same, possibly up to the zero. If σ is a nonzero eigenvalue and u, v are the associated eigenfunctions of L_K and $T_{\mathcal{H}}$ (normalized to norm 1 in $L^2(X, \rho)$ and $\mathcal{H}$) respectively, then*

$$\begin{aligned} u(x) &= \frac{1}{\sqrt{\sigma_j}} v(x) \qquad \text{for } \rho\text{-almost all } x \in X \\ v(x) &= \frac{1}{\sqrt{\sigma_j}} \int_X K(x, s) u(s) d\rho(s) \qquad \text{for all } x \in X \end{aligned}$$

3. *The following decompositions hold:*

$$\begin{aligned} L_K g &= \sum_{j \geq 1} \sigma_j \langle g, u_j \rangle_\rho u_j \qquad g \in L^2(X, \rho) \\ T_{\mathcal{H}} f &= \sum_{j \geq 1} \sigma_j \langle f, v_j \rangle v_j \qquad f \in \mathcal{H}, \end{aligned}$$

   where the eigenfunctions $(u_j)_{j \geq 1}$ of L_K form an orthonormal basis of $\ker L_K{}^{\perp}$ and the eigenfunctions $(v_j)_{j \geq 1}$ of $T_{\mathcal{H}}$ form an orthonormal basis for $ker(T_{\mathcal{H}})^{\perp}$.

If K is real-valued, both the families $(u_j)_{j \geq 1}$ and $(v_j)_{j \geq 1}$ can be chosen as real valued functions.

Note that the RKHS $\mathcal{H}$ does not depend on the measure ρ. If the support of the measure ρ is only a subset of X (e.g., a finite set of points or a submanifold), then functions in $L^2(X, \rho)$ are only defined on the support of ρ whereas functions in $\mathcal{H}$ are defined on the whole space X. The eigenfunctions of L_K and $T_{\mathcal{H}}$ coincide (up-to a scaling factor) on the support of the measure, and v is an *extension* of u outside of the support of ρ. Moreover, the extension/restriction operations preserve both the normalization and orthogonality of the eigenfunctions. In a slightly different context Coifman and Lafon (2006) showed the connection between the Nyström method and the set of eigenfunctions of L_K, which are called *geometric harmonics*. The main difference between our result and the cited paper is that we consider all the eigenfunctions, whereas Coifman and Lafon introduce a threshold on the spectrum to ensure stability since they do not consider a sampling procedure.

An analogous result relates the matrix $\mathbf{K}$ and the operator T_n .

Proposition 9 *The following facts hold:*

1. *The operator T_n is of finite rank, self-adjoint and positive, whereas the matrix $\mathbf{K}$ is conjugate symmetric and semi-positive defined. In particular the spectrum $\sigma(T_n)$ has only finitely many nonzero elements and is contained in $[0, \kappa]$.*

2. *The spectra of* $\mathbf{K}$ *and* T_n *are the same up to the zero, that is,* $\sigma(\mathbf{K})\setminus\{0\} = \sigma(T_n)\setminus\{0\}$. *Moreover, if* $\hat{\sigma}$ *is a nonzero eigenvalue and* $\hat{u}, \hat{v}$ *are the corresponding eigenvector and eigenfunction of* $\mathbf{K}$ *and* T_n *(normalized to norm* 1 *in* $\mathbb{C}^n$ *and* $\mathcal{H}$*) respectively, then*

$$\begin{aligned} \hat{u} &= \frac{1}{\sqrt{\hat{\sigma}}}(\hat{v}(x_1), \dots, \hat{v}(x_n)) \\ \hat{v} &= \frac{1}{\sqrt{\hat{\sigma}}}\left(\frac{1}{n}\sum_{i=1}^{n} K_{x_i}\hat{u}^i\right), \end{aligned}$$

where $\hat{u}^i$ *is the* $i-th$ *component of the eigenvector* $\hat{u}$.

3. *The following decompositions hold:*

$$\begin{aligned} \mathbf{K}w &= \sum_{j=1}^{k} \hat{\sigma}_j \langle w, \hat{u}_j\rangle \hat{u}_j \qquad w \in \mathbb{C}^n \\ T_n f &= \sum_{j=1}^{k} \hat{\sigma}_j \langle f, \hat{v}_j\rangle_{\mathcal{H}} \hat{v}_j \qquad f \in \mathcal{H}, \end{aligned}$$

where k *is the rank of* K *and both sums run over the nonzero eigenvalues, the family* $(\hat{u}_j)_{j\geq 1}$ *is an orthonormal basis for* $\ker\{\mathbf{K}\}^{\perp} \subset \mathbb{C}^n$ *and the family* $(\hat{v}_j)_{j\geq 1}$ *of* T_n *forms an orthonormal basis for the space* $\ker(T_n)^{\perp} \subset \mathcal{H}$, *where*

$$\ker(T_n) = \{f \in \mathcal{H} \mid f(x_i) = 0 \ \forall i = 1, \dots, n\}.$$

If K *is real-valued, both the families* $(\hat{u}_j)_{j\geq 1}$ *and* $(\hat{v}_j)_{j\geq 1}$ *can be chosen as real valued vectors and functions, respectively.*

Note that in this section L_K, T and T_n are self-adjoint operators and $\mathbf{K}$ is a conjugate symmetric matrix. If K is real, we can directly work with real Hilbert spaces. However since we need complex vector spaces in Section 4 for consistency we stated the above results for complex reproducing kernels.

3.2 Bounds on eigenvalues and spectral projections.

Using Theorem 7, we are able to bound the ℓ_2-distance between the spectrum of L_K and the spectrum of $\mathbf{K}$.

Proposition 10 *There exist an extended enumeration* $(\sigma_j)_{j\geq 1}$ *of discrete eigenvalues for* L_K *and an extended enumeration* $(\hat{\sigma}_j)_{j\geq 1}$ *of discrete eigenvalues for* $\mathbf{K}$ *such that*

$$\sum_{j\geq 1}(\sigma_j - \hat{\sigma}_j)^2 \leq \frac{8\kappa^2\tau}{n},$$

with confidence greater than $1 - 2e^{-\tau}$. *In particular* $\sup_{j\geq 1}|\sigma_j - \hat{\sigma}_j| \leq \frac{2\sqrt{2}\kappa\sqrt{\tau}}{\sqrt{n}}$.

Proof By Proposition 8, an extended enumeration $(\sigma_j)_{j\geq 1}$ of discrete eigenvalues for L_K is also an extended enumeration $(\sigma_j)_{j\geq 1}$ of discrete eigenvalues for $T_{\mathcal{H}}$, and a similar relation holds for $\mathbf{K}$ and T_n by Proposition 9. Theorem 5 with $A = T_n$ and $B = T_{\mathcal{H}}$ gives that

$$\sum_{j\geq 1}(\sigma_j - \hat{\sigma}_j)^2 \leq \|T_{\mathcal{H}} - T_n\|_{HS}^2$$

for a suitable extended enumerations $(\sigma_j)_{j\geq 1}$, $(\hat{\sigma}_j)_{j\geq 1}$ of discrete eigenvalues for T and T_n, respectively. Theorem 7 provides us with the claimed bound. ■

Theorem 4.2 and the following corollaries of Koltchinskii and Giné (2000) provide the same convergence rate (in expectation) under a different setting (the kernel K is only symmetric, but with some assumption on the decay of the eigenvalues of L_K).

The following result can be deduced by Theorem 5 with $p = 1$ and Theorem 7, however a simpler direct proof is given below.

Proposition 11 *Under the assumption of Proposition 10 with confidence* $1 - 2e^{-\tau}$

$$|\sum_{j\geq 1}(\sigma_j - \hat{\sigma}_j)| = |\operatorname{Tr}(T_{\mathcal{H}}) - \operatorname{Tr}(T_n)| \leq \frac{2\sqrt{2}\kappa\sqrt{\tau}}{\sqrt{n}}.$$

Proof Note that

$$\operatorname{Tr}(T_n) = \frac{1}{n}\sum_{i=1}^{n} K(x_i, x_i), \qquad \text{and} \qquad \operatorname{Tr}(T_{\mathcal{H}}) = \int_X K(x,x)d\rho(x).$$

Then we can define a sequence $(\xi_i)_{i=1}{}^n$ of real-valued random variables by $\xi_i = K(x_i, x_i) - \operatorname{Tr}(T_{\mathcal{H}})$. Clearly $\mathbb{E}[\xi_i] = 0$ and $|\xi_i| \leq 2\kappa$, $i = 1, \ldots, n$ so that Hoeffding inequality (3) yields with confidence $1 - 2e^{-\tau}$

$$\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i\right| = |\operatorname{Tr}(T_{\mathcal{H}}) - \operatorname{Tr}(T_n)| \leq \frac{2\sqrt{2}\kappa\sqrt{\tau}}{\sqrt{n}}.$$

■

From Proposition 6 and Theorem 7 we directly derive the probabilistic bound on the eigenprojections given by Zwald and Blanchard (2006) – their proof is based on bounded difference inequality for real random variables– see also De Vito et al. (2005a). Assume for the sake of simplicity, that the cardinality of $\sigma(L_K)$ is infinite. Given an integer N, let m be the sum of the multiplicities of the first N distinct eigenvalues of L_K, so that

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_m > \sigma_{m+1},$$

and P_N be the orthogonal projection from $L^2(X, \rho)$ onto the span of the corresponding eigenfunctions. Let k be the rank of $\mathbf{K}$, and $\hat{u}_1 \ldots, \hat{u}_k$ the eigenvectors corresponding to the nonzero eigenvalues of $\mathbf{K}$ in a decreasing order. Denote by $\hat{v}_1 \ldots, \hat{v}_k \in \mathcal{H} \subset L^2(X, \rho)$ the corresponding Nyström extension given by item 2 of Proposition 9.

Theorem 12 *Given $\tau > 0$, if the number n of examples satisfies*

$$n > \frac{128\,\kappa^2\tau}{(\sigma_m - \sigma_{m+1})^2},$$

then

$$\sum_{j=1}^{m} \|(I - P_N)\hat{v}_j\|_\rho^2 + \sum_{j=m+1}^{k} \|P_N \hat{v}_j\|_\rho^2 \leq \frac{32\kappa^2\tau}{(\sigma_m - \sigma_{m+1})^2\, n}\ , \tag{8}$$

with probability greater than $1 - 2e^{-\tau}$.

Proof Let $(u_j)_{j\geq 1}$ be an orthonormal family of eigenfunctions of L_K with strictly positive eigenvalues. Without loss of generality, we can assume that $u_1, \ldots, u_m$ are the eigenfunctions with eigenvalues $\sigma_1, \sigma_2, \ldots, \sigma_m$. Let $(v_j)_{j\geq 1}$ the corresponding family of eigenfunctions of $T_{\mathcal{H}}$ given by Proposition 8 and complete to an orthonormal basis of $\mathcal{H}$. Complete also the family $\hat{v}_1 \ldots, \hat{v}_k$ to an other orthonormal basis of $\mathcal{H}$.

Apply Proposition 6 with $A = T_{\mathcal{H}}$ and $B = T_n$ by taking into account Theorem 7. With high probability

$$\|P^{T_n} - P^{T_{\mathcal{H}}}\|_{\mathrm{HS}}^2 \leq \frac{8\kappa^2\tau}{(\sigma_m - \sigma_{m+1})^2\, n} \leq \frac{a_m - a_{m+1}}{2},$$

where

$$P^{T_{\mathcal{H}}} = \sum_{j=1}^{m} \langle f, v_j\rangle_{\mathcal{H}}\, v_j \qquad P^{T_n} = \sum_{j=1}^{m} \langle f, \hat{v}_j\rangle_{\mathcal{H}}\, \hat{v}_j$$

and the last bound follows from the condition on n. In particular, $\hat{\sigma}_m > \hat{\sigma}_{m+1}$.

Since both $(v_i)_{j\geq 1}$ and $(\hat{v}_i)_{j\geq 1}$ are orthonormal bases for $\mathcal{H}$

$$\begin{aligned}
\|P^{T_n} - P^{T_{\mathcal{H}}}\|_{\mathrm{HS}}^2 &= \sum_{i,j\geq 1} |\big\langle P^{T_n} v_i - P^{T_{\mathcal{H}}} v_i, \hat{v}_j\big\rangle_{\mathcal{H}}|^2 \\
&= \sum_{j=1}^{m} \sum_{i\geq m+1} |\langle v_i, \hat{v}_j\rangle_{\mathcal{H}}|^2 + \sum_{j\geq m+1} \sum_{i=1}^{m} |\langle v_i, \hat{v}_j\rangle_{\mathcal{H}}|^2 \\
&\geq \sum_{j=1}^{m} \sum_{\substack{i\geq m+1 \\ T_{\mathcal{H}} v_i \neq 0}} |\langle v_i, \hat{v}_j\rangle_{\mathcal{H}}|^2 + \sum_{j\geq m+1}^{k} \sum_{i=1}^{m} |\langle v_i, \hat{v}_j\rangle_{\mathcal{H}}|^2
\end{aligned}$$

Since the sum on i is with respect to the eigenfunctions of $T_{\mathcal{H}}$ with nonzero eigenvalue, the Mercer theorem implies that $\langle v_i, \hat{v}_j\rangle_{\mathcal{H}} = \langle u_i, \hat{v_j}\rangle_\rho$. Finally, observe that

$$\begin{aligned}
&\sum_{i=1}^{m} |\langle u_i, \hat{v}_j\rangle_\rho|^2 = \|P_N \hat{v}_j\|_\rho^2 \\
&\sum_{\substack{i\geq m+1 \\ T_{\mathcal{H}} v_i \neq 0}} |\langle u_i, \hat{v}_j\rangle_\rho|^2 = \sum_{\substack{i\geq m+1 \\ L_K u_i \neq 0}} |\langle u_i, \hat{v}_j\rangle_\rho|^2 = \|(I - P_N)\hat{v}_j\|_\rho
\end{aligned}$$

where we used that $\ker T_{\mathcal{H}} \subset \ker T_n$, so that $\hat{v_j} \in \ker L_K{}^{\perp}$ with probability 1. ■

First term in the left side of inequality (19) is the projection of the vector space spanned by the Nyström extension of the first top eigenvectors of the empirical matrix $\mathbf{K}$ onto the orthogonal of the vector space $\mathcal{M}_N$ spanned by the first top eigenfunctions of the integral operator L_K, the second term is the projection of the vector space spanned by the Nyström extensions of the other eigenvectors of $\mathbf{K}$ onto $\mathcal{M}_N$. Both differences are in $L^2(X, \rho)$ norm. A similar result is given in Zwald and Blanchard (2006), however, the role of the Nyström extensions is not considered– they study only the operators $T_{\mathcal{H}}$ and T_n (with our notation). Another result similar to ours is independently given in Smale and Zhou (2008), where the authors considered a single eigenfunction with multiplicity 1.

4. Asymmetric Graph Laplacian

In this section we will consider the case of the so-called asymmetric normalized graph Laplacian, which is the identity matrix minus the transition matrix for the natural random walk on a graph. In such a random walk, the probability of leaving a vertex along a given edge is proportional to the weight of that edge. As before, we will be interested in a specific class of graphs (matrices) associated with data.

Let $W : X \times X \to \mathbb{R}$ be a symmetric continuous weight function such that

$$0 < c \leq W(x, s) \leq C \qquad x, s \in X. \tag{9}$$

Note that we will not require W to be positive definite, but positive.

A set of data points $\mathbf{x} = (x_1, \ldots, x_n) \in X$ defines a weighted undirected graph with the weight matrix $\mathbf{W}$ given by $\mathbf{W}_{ij} = \frac{1}{n} W(x_i, x_j)$. The (asymmetric) normalized graph Laplacian $\mathbf{L} : \mathbb{C}^n \to \mathbb{C}^n$ is an $n \times n$ matrix given by

$$\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1}\mathbf{W},$$

where the *degree* matrix $\mathbf{D}$ is diagonal with

$$\mathbf{D}_{ii} = \frac{1}{n} \sum_{j=1}^{n} W(x_i, x_j) = \sum_{j=1}^{n} \mathbf{W}_{ij},$$

which is invertible since $\mathbf{D} \geq cI$ by (9).

As in the previous section, X is a subset of $\mathbb{R}^d$ endowed with a probability measure ρ and $L^2(X, \rho)$ is the space of square integrable complex functions with respect to ρ.

Let $m(x) = \int_X W(x, s) d\rho(s)$ be the *degree function*, bound (9) implies that the operator $\mathbb{L} : L^2(X, \rho) \to L^2(X, \rho)$

$$(\mathbb{L}f)(x) = f(x) - \int_X \frac{W(x, s) f(s)}{m(x)} d\rho(s),$$

is well defined and continuous. The fact that W is bounded away from zero is essential to control the behavior of the degree function m, however it might be possible to replace this condition with the requirement that $m(x) \geq c$, to consider localized weight functions.

We see that when a set $\mathbf{x} = (x_1, \ldots, x_n) \in X$ is sampled i.i.d. according to ρ, the matrix $\mathbf{L}$ is an empirical version of the operator $\mathbb{L}$. We will view $\mathbf{L}$ as a perturbation of $\mathbb{L}$ due to finite sampling and will extend the approach developed in this paper to relate their spectral properties. Note that the methods described in the previous section are not directly applicable in this setting since W is not necessarily positive definite, so there is no RKHS associated with it. Moreover, even if W is positive definite, $\mathbb{L}$ involves division by a function, and may not be a map from the RKHS to itself. To overcome this difficulty in our theoretical analysis, we will rely on an auxiliary RKHS $\mathcal{H}$ with reproducing kernel K. Interestingly enough, this space will play no role from the algorithmic point of view, but only enters the theoretical analysis.

To state the properties of $\mathcal{H}$ we define the following functions

$$\begin{aligned} K_x : X \to \mathbb{C} \qquad & K_x(t) = K(t,x) \\ W_x : X \to \mathbb{R} \qquad & W_x(t) = W(t,x) \\ m_n : X \to \mathbb{R} \qquad & m_n = \frac{1}{n}\sum_{i=1}^{n} W_{x_i}, \end{aligned}$$

where m_n is the empirical counterpart of the function m and, in particular, $m_n(x_i) = \mathbf{D}_{ii}$.

To proceed we need the following Assumption A, which postulates that the functions $w_x, W_x/m, W_x/m_n$ belong to $\mathcal{H}$. However it is important to note that for W sufficiently smooth (as we expect it to be in most applications) these conditions can be satisfied by choosing $\mathcal{H}$ to be a Sobolev space of sufficiently high degree. This is made precise in the Section 4.2 (see Assumption A1).

Assumption A *Given a continuous weight function W satisfying* (9)*, we assume there exists a RKHS $\mathcal{H}$ with bounded continuous kernel K such that*

$$\begin{aligned} & W_x,\ \frac{1}{m}W_x,\ \frac{1}{m_n}W_x \in \mathcal{H} \\ & \|\frac{1}{m}W_x\|_{\mathcal{H}} \le C, \end{aligned}$$

for all $x \in X$.

Assumption A allows to define the following bounded operators $\mathbb{L}_{\mathcal{H}}, A_{\mathcal{H}} : \mathcal{H} \to \mathcal{H}$

$$\begin{aligned} A_{\mathcal{H}} &= \int_X \langle \cdot, K_x \rangle_{\mathcal{H}} \frac{1}{m} W_x \, d\rho(x) \\ \mathbb{L}_{\mathcal{H}} &= I - A_{\mathcal{H}} \end{aligned} \tag{10}$$

and their empirical counterparts $\mathbb{L}_n, A_n : \mathcal{H} \to \mathcal{H}$

$$\begin{aligned} A_n &= \frac{1}{n}\sum_{i=1}^{n} \langle \cdot, K_{x_i} \rangle_{\mathcal{H}} \frac{1}{m_n} W_{x_i} \\ \mathbb{L}_n &= I - A_n. \end{aligned} \tag{11}$$

Next result will show that $\mathbb{L}_{\mathcal{H}}$, $A_{\mathcal{H}}$ and $\mathbb{L}$ have related eigenvalues and eigenfunctions and that eigenvalues and eigenfunctions (eigenvectors) of $\mathbb{L}_n$, A_n and $\mathbf{L}$ are also closely related. In particular we will see in the following that to relate the spectral properties of $\mathbb{L}$ and $\mathbf{L}$ it suffices to control the deviation $A_{\mathcal{H}} - A_n$. However, before doing this, we make the above statements precise in the following subsection.

4.1 Extension Operators

In analogy to Section 3.1 we consider the relation between the operators we want to study and their extensions. We define the restriction operator $R_n : \mathcal{H} \to \mathbb{C}^n$ and the extension operator $E_n : \mathbb{C}^n \to \mathcal{H}$ as

$$\begin{aligned} R_n f &= (f(x_1), \ldots, f(x_n)) \qquad\qquad f \in \mathcal{H} \\ E_n(y_1, \ldots, y_n) &= \frac{1}{n} \sum_{i=1}^{n} y_i \frac{1}{m_n} W_{x_i} \qquad (y_1, \ldots, y_n) \in \mathbb{C}^n \end{aligned}$$

Clearly the extension operator is no longer the adjoint of R_n but the connection among the operators $\mathbf{L}$ to $\mathbb{L}_n$ and A_n can still be clarified by means of R_n and E_n. Indeed it is easy to check that $A_n = E_n R_n$ and $\mathbf{D}^{-1}\mathbf{W} = R_n E_n$. Similarly the infinite sample restrictions and extension operators can be defined to relate the operators $\mathbb{L}$, $A_{\mathcal{H}}$ and $\mathbb{L}_{\mathcal{H}}$. The next proposition considers such a connection.

Proposition 13 *The following facts hold true.*

1. *The operator $A_{\mathcal{H}}$ is Hilbert-Schmidt, the operators $\mathbb{L}$ and $\mathbb{L}_{\mathcal{H}}$ are bounded and have positive eigenvalues.*

2. *Given $\sigma \in [0, +\infty[$, $\sigma \in \sigma(\mathbb{L}_{\mathcal{H}})$ if and only if $1 - \sigma \in \sigma(A_{\mathcal{H}})$, with the same eigenfunction.*

3. *The spectra of $\mathbb{L}$ and $\mathbb{L}_{\mathcal{H}}$ are the same up to the eigenvalue 1. If $\sigma \neq 1$ is an eigenvalue and u, v associated eigenfunctions of $\mathbb{L}$ and $\mathbb{L}_{\mathcal{H}}$ respectively, then*

$$\begin{aligned} u(x) &= v(x) \qquad \text{for almost all } x \in X \\ v(x) &= \frac{1}{1-\sigma} \int_X \frac{W(x,t)}{m(x)} u(t)\, d\rho(t) \qquad \text{for all } x \in X \end{aligned}$$

4. *Finally the following decompositions hold*

$$\mathbb{L} = \sum_{\substack{j \geq 1 \\ \sigma_j \neq 1}} \sigma_j P_j + P_0, \tag{12}$$

$$\mathbb{L}_{\mathcal{H}} = I - \sum_{\substack{j \geq 1 \\ \sigma_j \neq 1}} (1 - \sigma_j) Q_j + D, \tag{13}$$

where the projections Q_j, P_j are the spectral projections of $\mathbb{L}$ and $\mathbb{L}_{\mathcal{H}}$ associated with the eigenvalue σ_j, P_0 is the spectral projection of $\mathbb{L}$ associated with the eigenvalue 1, and D is a quasi-nilpotent operator such that $\ker D = \ker (I - \mathbb{L}_{\mathcal{H}})$ and $Q_j D = D Q_j = 0$ for all $j \geq 1$.

Furthermore, all the eigenfunctions can be chosen as real-valued.

The proof of the above result is long and quite technical and can be found in Appendix A. Note that, with respect to Proposition 9, neither the normalization nor the orthogonality is preserved by the extension/restriction operations, so that we are free to normalize v with the factor $1/(1-\sigma)$, instead of $1/\sqrt{1-\sigma}$ as in Proposition 8. One can easily show that, if $u_1, \ldots, u_m$ is a linearly independent family of eigenfunctions of $\mathbb{L}$ with eigenvalues $\sigma_1, \ldots, \sigma_m \neq 1$, then the extension $v_1, \ldots, v_m$ is a linearly independent family of eigenfunctions of $\mathbb{L}_{\mathcal{H}}$ with eigenvalues $\sigma_1, \ldots, \sigma_m \neq 1$. Finally, we stress that in item 4 both series converge in the strong operator topology, however, though $\sum_{j\geq 1} P_i = I - P_0$, it is not true that $\sum_{j\geq 1} Q_i$ converges to $I - Q_0$, where Q_0 is the spectral projection of $\mathbb{L}_{\mathcal{H}}$ associated with the eigenvalue 1. This is the reason why we need to write the decomposition of $\mathbb{L}_{\mathcal{H}}$ as in (13) instead of (12). An analogous result allows us to relate $\mathbf{L}$ to $\mathbb{L}_n$ and A_n.

Proposition 14 *The following facts hold true.*

1. *The operator A_n is Hilbert-Schmidt, the matrix $\mathbf{L}$ and the operator $\mathbb{L}_n$ have positive eigenvalues.*

2. *Given $\sigma \in [0, +\infty[$, $\sigma \in \sigma(\mathbb{L}_n)$ if and only if $1 - \sigma \in \sigma(A_n)$, with the same eigenfunction.*

3. *The spectra of $\mathbf{L}$ and $\mathbb{L}_n$ are the same up to the eigenvalue 1, moreover if $\hat{\sigma} \neq 1$ is an eigenvalue and the $\hat{u}, \hat{v}$ eigenvector and eigenfunction of $\mathbf{L}$ and $\mathbb{L}_n$, then*

$$\begin{aligned} \hat{u} &= (\hat{v}(x_1), \ldots, \hat{v}(x_1)) \\ \hat{v}(x) &= \frac{1}{1-\hat{\sigma}} \frac{1}{n} \sum_{i=1}^{n} \frac{W(x, x_i)}{m_n(x)} \hat{u}^i \end{aligned}$$

   where $\hat{u}^i$ is the $i-th$ component of the eigenvector $\hat{u}$.

4. *Finally the following decompositions hold*

$$\begin{aligned} \mathbf{L} &= \sum_{\substack{j\geq 1 \\ \hat{\sigma}_j \neq 1}} \hat{\sigma}_j \hat{P}_j + \hat{P}_0, \\ \mathbb{L}_n &= \sum_{\substack{j\geq 1 \\ \hat{\sigma}_j \neq 1}} \hat{\sigma}_j \hat{Q}_j + \hat{Q}_0 + \hat{D}, \end{aligned}$$

   where the projections Q_j, P_j are the spectral projections of $\mathbf{L}$ and $\mathbb{L}_n$ associated with the eigenvalue σ_j, $\hat{P}_0$ and $\hat{Q}_0$ are the spectral projections of $\mathbf{L}$ and $\mathbb{L}_n$ associated with the eigenvalue 1, and $\hat{D}$ is a quasi-nilpotent operator such that $\ker \hat{D} = \ker (I - \mathbb{L}_n)$ and $\hat{Q}_j \hat{D} = \hat{D} \hat{Q}_j = 0$ for all j with $\hat{\sigma}_j \neq 1$.

The last decomposition is parallel to the Jordan canonical form for (non-symmetric) matrices. Notice that, since the sum is finite, $\sum_{\substack{j\geq 1 \\ \hat{\sigma}_j \neq 1}} \hat{Q}_j + \hat{Q}_0 = I$.

4.2 Graph Laplacian Convergence for Smooth Weight Functions

If the weight function W is sufficiently differentiable, we can choose the RKHS $\mathcal{H}$ to be a suitable Sobolev space. For the sake of simplicity, we assume that X is a bounded open subset of $\mathbb{R}^d$ with a *nice* boundary[3]. Given $s \in \mathbb{N}$, the Sobolev space $\mathcal{H}^s = \mathcal{H}^s(X)$ is

$$\mathcal{H}^s = \{f \in L^2(X, dx) \mid D^\alpha f \in L^2(X, dx) \ \forall |\alpha| = s\},$$

where $D^\alpha f$ is the (weak) derivative of f with respect to the multi-index $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$, $|\alpha| = \alpha_1 + \cdots + \alpha_d$ and $L^2(X, dx)$ is the space of square integrable complex functions with respect to the Lebesgue measure (Burenkov, 1998). The space $\mathcal{H}^s$ is a separable Hilbert space with respect to the scalar product

$$\langle f, g \rangle_{\mathcal{H}^s} = \langle f, g \rangle_{L^2(X,dx)} + \sum_{|\alpha|=s} \langle D^\alpha f, D^\alpha g \rangle_{L^2(X,dx)} .$$

Let $C_b^s(X)$ be the set of continuous bounded functions such that all the (standard) derivatives of order s exists and are continuous bounded functions. The space $C_b^s(X)$ is a Banach space with respect to the norm

$$\|f\|_{C_b^s} = \sup_{x \in X} |f(x)| + \sum_{|\alpha|=s} \sup_{x \in X} |(D^\alpha f)(x)|.$$

Since X is bounded, it is clear that $C_b^s(X) \subset \mathcal{H}^s$ and $\|f\|_{\mathcal{H}^s} \leq d \|f\|_{C_b^s}$, where d is a suitable constant depending only on s. A sort of converse also holds, which will be crucial in our approach, see Corollary 21 of Burenkov (1998). Let $l, m \in \mathbb{N}$ such that $l - m > \frac{d}{2}$, then

$$\mathcal{H}^l \subset C_b^m(X) \qquad \|f\|_{C_b^m} \leq d' \|f\|_{\mathcal{H}^l} \tag{14}$$

where d' is a constant depending only on l and m.

From Eq (14) with $l = s$ and $m = 0$, we see that the Sobolev space $\mathcal{H}^s$, where $s = \lfloor d/2 \rfloor + 1$, is a RKHS with a continuous[4] real valued bounded kernel K^s.

We are ready to state our assumption on the weight function, which implies Assumption A.

Assumption A1 *We assume that $W : X \times X \to \mathbb{R}$ is a positive, symmetric function such that*

$$W(x, t) \geq c > 0 \qquad \forall x, t \in X \tag{15}$$

$$W \in C_b^{d+1}(X \times X). \tag{16}$$

As we will see, condition (16) quantifies the regularity of W we need to use Sobolev spaces as RKHS and, as usual, it critically depends on the dimension of the input space, see also Remark 19 below. By inspecting our proofs, (16) can be replaced by the more technical condition $W \in \mathcal{H}^{d+1}(X \times X)$.

As a consequence of Assumption A1, we are able to control the deviation of $\mathbb{L}_n$ from $\mathbb{L}_{\mathcal{H}}$.

3. The conditions are very technical and we refer to Burenkov (1998) for the precise assumptions.
4. The kernel K^s is continuous on $X \times X$ since the embedding of $\mathcal{H}^s$ into $C_b(X)$ is compact, see Schwartz (1964).

Theorem 15 *Under the conditions of Assumption A1, with confidence* $1 - 2e^{-\tau}$ *we have*

$$\|\mathbb{L}_n - \mathbb{L}_{\mathcal{H}}\|_{HS} = \|A_{\mathcal{H}} - A_n\|_{HS} \leq C\frac{\sqrt{\tau}}{\sqrt{n}}, \tag{17}$$

where $\|\cdot\|_{HS}$ *is the Hilbert-Schimdt norm of an operator in the Sobolev space* $\mathcal{H}^s$ *with* $s = \lfloor d/2 \rfloor + 1$*, and* C *is a suitable constant.*

To prove this result we need some preliminary lemmas. In the following C will be a constant that could change from one statement to the other. The first result shows that Assumption A1 implies Assumption A with $\mathcal{H} = \mathcal{H}^s$ and that the multiplicative operators defined by the degree function, or its empirical estimate, are bounded.

Lemma 16 *There exists a suitable constant* $C > 0$ *such that*

1. *for all* $x \in X$, $W_x, \frac{1}{m}W_x, \frac{1}{m_n}W_x \in \mathcal{H}^{d+1} \subset \mathcal{H}^s$ *and* $\|\frac{1}{m}W_x\|_{\mathcal{H}^s} \leq C$*;*
2. *the multiplicative operators* $M, M_n : \mathcal{H}^s \to \mathcal{H}^s$ *defined by*

$$Mf = mf \qquad M_n f = m_n f \qquad\qquad f \in \mathcal{H}^s$$

are bounded invertible operators satisfying

$$\|M\|_{\mathcal{H}^s,\mathcal{H}^s}, \|M^{-1}\|_{\mathcal{H}^s,\mathcal{H}^s}, \|M_n\|_{\mathcal{H}^s,\mathcal{H}^s}, \|M_n^{-1}\|_{\mathcal{H}^s,\mathcal{H}^s} \leq C \tag{18}$$

$$\|M - M_n\|_{\mathcal{H}^s,\mathcal{H}^s} \leq C\|m - m_n\|_{\mathcal{H}^{d+1}}, \tag{19}$$

where $\|\cdot\|_{\mathcal{H}^s,\mathcal{H}^s}$ *is the operator norm of an operator in the Sobolev space* $\mathcal{H}^s$*.*

Proof Let $C_1 = \|W\|_{C_b^{d+1}(X\times X)}$. Given $x \in X$, clearly $W_x \in C_b^{d+1}(X)$ and, by standard results of differential calculus, both m and $m_n \in C_b^{d+1}(X)$ with

$$\|W_x\|_{C_b^{d+1}(X)}, \|m\|_{C_b^{d+1}(X)}, \|m_n\|_{C_b^{d+1}(X)} \leq C_1.$$

Leibniz rule for the quotient with bound (15) gives that $\frac{1}{m}$ and $\frac{1}{m_n} \in C_b^{d+1}(X)$ with

$$\|\frac{1}{m}\|_{C_b^{d+1}(X)}, \|\frac{1}{m_n}\|_{C_b^{d+1}(X)} \leq C_2,$$

where C_2 is independent both on n and on the sample $(x_1, \ldots, x_n)$. Claim in item 1 is a consequence of the fact that pointwise multiplication is a continuous bilinear map on $C_b^{d+1}(X)$, and $C_b^{d+1}(X) \subset \mathcal{H}^{d+1} \subset \mathcal{H}^s$ with

$$\|f\|_{\mathcal{H}^s} \leq C_3\|f\|_{\mathcal{H}^{d+1}} \leq C_4\|f\|_{C_b^{d+1}(X)}.$$

We claim that if $g \in C_b^{d+1}(X)$ and $f \in \mathcal{H}^s$, then $gf \in \mathcal{H}^s$ and

$$\|gf\|_{\mathcal{H}^s} \leq \|g\|_{\mathcal{H}^{d+1}}\|f\|_{\mathcal{H}^s}.$$

Indeed, Lemma 15 of Section 4 of Burenkov (1998) with $p = p_2 = 2$, $p_1 = \infty$, $l = s$ and $n = d$ ensures that

$$\|gf\|_{\mathcal{H}^s} \leq \|g\|_{C_b^s(X)} \|f\|_{\mathcal{H}^s}.$$

Eq. (14) with $m = s$ and $l = d+1 > d/2+s = d/2+[d/2]+1$ provides us with the claimed bound.

The content of item 2. is a consequence of the above result with $g = m, m_n, \frac{1}{m}, \frac{1}{m_n}$, and $m - m_n$, respectively, satisfying $\|g\|_{\mathcal{H}^{d+1}} \leq C_4 \max\{C_1, C_2\} = C_5$.
The constant C will be the maximum among the constants C_i. ■

Next lemma shows that the integral operator of kernel W and its empirical counterpart are Hilbert-Schmidt operators and it bounds their difference.

Lemma 17 *The operators $L_{W,\mathcal{H}}, L_{W,n} : \mathcal{H}^s \to \mathcal{H}^s$ defined by*

$$\begin{aligned} L_{W,\mathcal{H}} &= \int_X \langle \cdot, K_x^s \rangle_{\mathcal{H}^s} W_x d\rho(x), \\ L_{W,n} &= \frac{1}{n} \sum_{i=1}^n \left\langle \cdot, K_{x_i}^s \right\rangle_{\mathcal{H}^s} W_{x_i}, \end{aligned}$$

are Hilbert-Schmidt. Furthermore, with confidence $1 - 2e^{-\tau}$

$$\|L_{W,\mathcal{H}} - L_{W,n}\|_{HS} \leq C \frac{\sqrt{\tau}}{\sqrt{n}}.$$

for a suitable constant C.

Proof Note that $\|\left\langle \cdot, K_{x_i}^s \right\rangle_{\mathcal{H}^s} W_{x_i}\|_{HS} = \|K_{x_i}^s\|_{\mathcal{H}^s} \|W_{x_i}\|_{\mathcal{H}^s} \leq C_1$ for a suitable constant C_1, which is finite by item 1 of Lemma 16 and the fact that K^s is bounded. Hence $L_{W,n}, L_{W,\mathcal{H}}$ are Hilbert Schmidt operators on $\mathcal{H}^s$. The random variables $(\xi_i)_{i=1}^n$ defined by $\xi_i = \left\langle \cdot, K_{x_i}^s \right\rangle_{\mathcal{H}^s} W_{x_i} - L_{W,\mathcal{H}}$, taking value in the Hilbert space of Hilbert-Schmidt operators, are zero mean and bounded. Applying (4) we have with confidence $1 - 2e^{-\tau}$

$$\|L_{W,\mathcal{H}} - L_{W,n}\|_{HS} \leq C \frac{\sqrt{\tau}}{\sqrt{n}}, \tag{20}$$

for a suitable constant C. ■

We then consider multiplication operators defined by the degree functions.

Lemma 18 *With confidence $1 - 2e^{-\tau}$*

$$\|M - M_n\|_{\mathcal{H}^s, \mathcal{H}^s} \leq C \frac{\sqrt{\tau}}{\sqrt{n}}.$$

for a suitable constant C.

Proof Item 2 of Lemma 16 ensures that M and M_n are bounded operators on $\mathcal{H}^s$ with $\|M - M_n\|_{\mathcal{H}^s, \mathcal{H}^s} \leq C_1 \|m - m_n\|_{\mathcal{H}^{d+1}}$.

The random variables $(\xi_i)_{i=1}^n$, defined by $\xi_i = W_{x_i} - m \in \mathcal{H}^{d+1}$ are zero mean and bounded. Applying (4) we have with high probability

$$\|m - m_n\|_{\mathcal{H}^{d+1}} \leq \frac{C_2\sqrt{\tau}}{\sqrt{n}},$$

so that the claim is proved with a suitable choice for C. ■

Remark 19 *In the above lemma we need to control $m - m_n$ in a suitable Hilbert space in order to use Hoeffding inequality* (4)*. Lemma* 15 *of Burenkov (1998) ensures that $\|M - M_n\|_{\mathcal{H}^s,\mathcal{H}^s}$ is bounded by $\|m - m_n\|_{C_b^s(X)}$. In order to control it with a Sobolev norm by means of* (14)*, we need to require that $m - m_n \in \mathcal{H}^l$ with $l > s + d/2$. Furthermore, the requirement that $\mathcal{H}^s$ is a RKHS with continuous bounded kernel implies that $s > d/2$ so that $l > d$. Hence a natural requirement on the weight function is that $W_x \in \mathcal{H}^l(X)$, which is closely related to Assumption A1 with the minimal choice $l = d + 1$.*

Finally, we can combine the above two lemmas to get the proof of Theorem 15.

Proof [Proof of Theorem 15] By Lemma 16, both M and M_n are invertible operators and

$$A_{\mathcal{H}} = M^{-1}L_{W,\mathcal{H}} \qquad A_n = M_n^{-1}L_{W,n},$$

so that we can consider the following decomposition

$$\begin{aligned} A_n - A_{\mathcal{H}} &= M_n^{-1}L_{W,n} - M^{-1}L_{W,\mathcal{H}} \\ &= (M_n^{-1} - M^{-1})L_{W,\mathcal{H}} + M_n^{-1}(L_{W,n} - L_{W,\mathcal{H}}) \\ &= M_n^{-1}(M - M_n)M^{-1}L_{W,\mathcal{H}} + \\ &\quad + M_n^{-1}(L_{W,n} - L_{W,\mathcal{H}}). \end{aligned} \tag{21}$$

By taking the Hilbert-Schmidt norm of the above expression and using (2) with the bounds provided by Lemma 16, we get

$$\|A_n - A_{\mathcal{H}}\|_{HS} \leq C^2\|M - M_n\|_{\mathcal{H}^s,\mathcal{H}^s}\|L_{W,\mathcal{H}}\|_{HS} + C\|L_{W,n} - L_{W,\mathcal{H}}\|_{HS}$$

The concentration inequalities (17) and (18) give (17), possibly redefining the constant C. ■

In the next section we discuss the implications of the above result in terms of concentration of eigenvalues and spectral projections.

4.3 Bounds on eigenvalues and spectral projections

Since the operators are no longer self-adjoint the perturbation results in Section 3.2 cannot be used. See Appendix B for a short review about spectral theory for compact (not necessarily self-adjoint) operators. The following theorem is an adaptation of results in Anselone (1971), compare with Theorem 4.21.

Theorem 20 *Let A be a compact operator. Given a finite set Λ of non-zero eigenvalues of A, let Γ be any simple rectifiable closed curve (having positive direction) with Λ inside and $\sigma(A) \setminus \Lambda$ outside. Let P be the spectral projection associated with Λ, that is,*

$$P = \frac{1}{2\pi i} \int_\Gamma (\lambda - A)^{-1} \, d\lambda,$$

and define

$$\delta^{-1} = \sup_{\lambda \in \Gamma} \|(\lambda - A)^{-1}\|.$$

Let B be another compact operator such that

$$\|B - A\| \leq \frac{\delta^2}{\delta + \ell(\Gamma)/2\pi} < \delta,$$

where $\ell(\Gamma)$ is the length of Γ, then the following facts hold true.

1. *The curve Γ is a subset of the resolvent set of B enclosing a finite set $\widehat{\Lambda}$ of non-zero eigenvalues of B;*
2. *Denoting by $\widehat{P}$ the spectral projection of B associated with $\widehat{\Lambda}$, then*

$$\|\widehat{P} - P\| \leq \frac{\ell(\Gamma)}{2\pi\delta} \frac{\|B - A\|}{\delta - \|B - A\|};$$

3. *The dimension of the range of P is equal to the dimension of the range of $\hat{P}$.*

Moreover, if $B - A$ is a Hilbert-Schmidt operator, then

$$\|\widehat{P} - P\|_{HS} \leq \frac{\ell(\Gamma)}{2\pi\delta} \frac{\|B - A\|_{HS}}{\delta - \|B - A\|}.$$

We postpone the proof of the above result to Appendix A.
We note that, if A is self-adjoint, then the spectral theorem ensures that

$$\delta = \min_{\lambda \in \Gamma, \sigma \in \Lambda} |\lambda - \sigma|.$$

The above theorem together with the results obtained in the previous section allows to derive several results.

Proposition 21 *If Assumption A1 holds, let σ be an eigenvalue of $\mathbb{L}$, $\sigma \neq 1$, with multiplicity m. For any $\varepsilon > 0$ and $\tau > 0$, there exists an integer n_0 and a positive constant C such that, if the number of examples is greater than n_0, with probability greater than $1 - 2e^{-\tau}$,*

1. *there are $\hat{\sigma}_1, \ldots, \hat{\sigma}_m$ (possibly repeated) eigenvalues of the matrix $\mathbf{L}$ satisfying*

$$|\hat{\sigma}_i - \sigma| \leq \varepsilon \qquad \text{for all } i = 1, \ldots, m.$$

2. *for any normalized eigenvector* $\hat{u} \in \mathbb{C}^n$ *of* $\mathbf{L}$ *with eigenvalue* $\hat{\sigma}_i$ *for some* $i = 1, \dots, m$, *there exists an eigenfunction* $u \in \mathcal{H}^s \subset L^2(X, \rho)$ *of* $\mathbb{L}$ *with eigenvalue* σ, *satisfying*

$$\|\hat{v} - u\|_{\mathcal{H}^s} \leq C \frac{\sqrt{\tau}}{\sqrt{n}},$$

where $\hat{v}$ *is the Nyström extension of the vector* $\hat{u}$ *given in item* 3 *of Proposition* 14.

If $\mathbb{L}_{\mathcal{H}}$ *is self-adjoint, then* $n_0 \geq 4\frac{C^2\tau}{\varepsilon^2}$ *provided that* $\varepsilon < \min_{\sigma' \in \sigma(\mathbb{L}_{\mathcal{H}}), \sigma' \neq \sigma} |\sigma' - \sigma|$.

Proof Theorem 15 gives that, with probability greater than $1 - 2e^{-\tau}$,

$$\|A_n - A_{\mathcal{H}}\| \leq \|A_n - A_{\mathcal{H}}\|_{HS} \leq \frac{C_1\sqrt{\tau}}{\sqrt{n}} \leq \frac{\delta^2}{\delta + \varepsilon}. \tag{22}$$

for all $n \geq n_0$, where C_1 is a suitable constant, $n_0 \in \mathbb{N}$ is such that $\frac{C_1\sqrt{\tau}}{\sqrt{n_0}} \leq \frac{\delta^2}{\delta+\varepsilon}$ and $\delta^{-1} = \sup_{\lambda \in \Gamma} \|(\lambda - A_{\mathcal{H}})^{-1}\|$. Under this conditions, we apply Theorem 20 with $A = A_{\mathcal{H}}$, $B = A_n$ and $\Gamma = \{\lambda \in \mathbb{C} \mid |\lambda - (1 - \sigma)| = \varepsilon\}$, so that $\ell(\Gamma) = 2\pi\varepsilon$. Since $A_{\mathcal{H}}$ is compact and assuming ε small enough, we have that $\Lambda = \{1 - \sigma\}$.
Item 1 of Theorem 20 with Proposition 14 ensures that $\widehat{\Lambda} = \{1 - \hat{\sigma}_1, \dots, 1 - \hat{\sigma}_m\}$, so that $|\hat{\sigma}_i - \sigma| < \varepsilon$ for all $i = 1, \dots, m$. Let now $\hat{u} \in \mathbb{C}^n$ be a normalized vector such that $\mathbf{L}\hat{u} = \hat{\sigma}_i\hat{u}$ for some $i = 1, \dots, m$.
Proposition 14 ensures that $\hat{v}$ is an eigenfunction of A_n with eigenvalue $1 - \hat{\sigma}$, so that $\hat{Q}\hat{v} = \hat{v}$ where $\hat{Q}$ is the spectral projection of A_n associated with $\hat{\Lambda}$. Let Q be the spectral projection of $A_{\mathcal{H}}$ associated with $1 - \sigma$ and define $u = Q\hat{v} \in \mathcal{H}^s$. By definition of Q, $A_{\mathcal{H}}u = (1 - \sigma)u$. Since $\mathcal{H}^s \subset L^2(X, \rho)$, Proposition 13 ensures that $\mathbb{L}u = \sigma u$. Item 2 of Theorem 20 gives that

$$\begin{aligned}\|\hat{v} - u\|_{\mathcal{H}^s} = \|\hat{Q}\hat{v} - Q\hat{v}\|_{\mathcal{H}^s} &\leq \|\hat{Q} - Q\|_{\mathcal{H}^s, \mathcal{H}^s} \|\hat{v}\|_{\mathcal{H}^s} \leq \|\hat{v}\|_{\mathcal{H}^s} \frac{\varepsilon}{\delta} \frac{\|A_n - A_{\mathcal{H}}\|_{HS}}{\delta - \|A_n - A_{\mathcal{H}}\|_{HS}} \\ &\leq \frac{C_2}{1 - (\sigma + \varepsilon)} C_1 \frac{\delta + \varepsilon}{\delta^2} \frac{\sqrt{\tau}}{\sqrt{n}},\end{aligned}$$

where we use (22), the fact that $\|A_n - A_{\mathcal{H}}\| \leq \frac{\delta^2}{\delta+\varepsilon}$ and

$$\|\hat{v}\|_{\mathcal{H}^s} \leq \frac{1}{1 - \hat{\sigma}} \sup_{x \in X} \|\frac{1}{m_n} W_x\|_{\mathcal{H}^s} = \frac{C_2}{1 - \hat{\sigma}} \leq \frac{C_2}{1 - (\sigma + \varepsilon)}$$

with C_2 being the constant in item 1 of Lemma 16.
If $A_{\mathcal{H}}$ is self-adjoint, the spectral theorem ensures that $\delta = \varepsilon$, so that $n_0 \geq 4\frac{C^2\tau}{\varepsilon^2}$. ■

Next we consider convergence of the spectral projections of $A_{\mathcal{H}}$ and A_n associated with the first N-eigenvalues. For sake of simplicity, we assume that the cardinality of $\sigma(A_{\mathcal{H}})$ is infinite.

Proposition 22 *Consider the first* N *eigenvalues of* $A_{\mathcal{H}}$. *There exist an integer* n_0 *and a constant* $C > 0$, *depending on* N *and* $\sigma(A_{\mathcal{H}})$, *such that, with confidence* $1 - 2e^{-\tau}$ *and for any* $n \geq n_0$,

$$\|P_N - \hat{P}_D\|_{HS} \leq \frac{C\sqrt{\tau}}{\sqrt{n}},$$

where $P_N, \hat{P}_D$ are the eigenprojections corresponding to the first N eigenvalues of $A_{\mathcal{H}}$ and D eigenvalues of A_n, and D is such that the sum of the multiplicity of the first D eigenvalues of A_n is equal to the sum of the multiplicity of the first N eigenvalues of $A_{\mathcal{H}}$.

Proof The proof is close to the one of previous proposition. We apply Theorem 20 with $A = A_{\mathcal{H}}$, $B = A_n$ and the curve Γ equal to the boundary of the rectangle

$$\{\lambda \in \mathbb{C} \mid \frac{\alpha_N + \alpha_{N+1}}{2} \leq \Re e(\lambda) \leq \|A\| + 2,\ |\Im m(\lambda)| \leq 1\},$$

where α_N is the N-largest eigenvalue of $A_{\mathcal{H}}$ and α_{N+1} the $N+1$-largest eigenvalue of $A_{\mathcal{H}}$. Clearly Γ encloses the first N largest eigenvalues of $A_{\mathcal{H}}$, but no other points of $\sigma(A)$. Define $\delta^{-1} = \sup_{\lambda\in\Gamma}\|(\lambda - A_{\mathcal{H}})^{-1}\|$ and $n_0 \in \mathbb{N}$ such that

$$\frac{C_1\sqrt{\tau}}{\sqrt{n_0}} \leq \frac{\delta^2}{\delta + \ell(\Gamma)/2\pi} \qquad \text{and} \qquad \frac{C_1\sqrt{\tau}}{\sqrt{n_0}} < 1,$$

where C_1 is the constant in Theorem 15. As in the above corollary, with probability greater than $1 - 2e^{-\tau}$, for all $n \geq n_0$

$$\|A_n - A_{\mathcal{H}}\| \leq \frac{\delta^2}{\delta + \ell(\Gamma)/2\pi} \qquad \text{and} \qquad \|A_n - A_{\mathcal{H}}\| < 1.$$

The last inequality ensures that the largest eigenvalues of A_n is smaller than $1 + \|A_{\mathcal{H}}\|$, so that by Theorem 20, the curve Γ encloses the first D-eigenvalues of A_n, where D is equal to the sum of the multiplicity of the first N eigenvalues of $A_{\mathcal{H}}$. The proof is finished letting $C = \frac{\delta+\ell(\Gamma)/2\pi}{\delta^2} C_1$. ■

Acknowledgments

We would like to thank the referees for many constructive suggestions and comments. Ernesto De Vito and Lorenzo Rosasco have been partially supported by the FIRB project RBIN04PARL and by the EU Integrated Project Health-e-Child IST-2004-027749. Mikhail Belkin is partially supported by the NSF Early Career Award 0643916.

Appendix A. Some Proofs

We start giving the proof of Proposition 13.

Proof [of Proposition 13]

We first need a preliminary fact. The function m is bounded from above and below by a positive constant by (9), so that the measure $\rho_W = m\rho$, having density m w.r.t. ρ, is equivalent[5] to ρ. This implies that the spaces $L^2(X, \rho)$ and $L^2(X, \rho_W)$ are the same vector space and the corresponding norm are equivalent. In this proof, we regard $\mathbb{L}$ as an operator from $L^2(X, \rho_W)$ into $L^2(X, \rho_W)$, observing that its eigenvalues and eigenfunctions are the same as the eigenvalues and eigenfunctions of $\mathbb{L}$, viewed as an operator from $L^2(X, \rho)$ into

5. Two measures are equivalent if they have the same null sets.

$L^2(X,\rho)$ – however, functions that are orthogonal in $L^2(X,\rho_W)$ in general are not orthogonal in $L^2(X,\rho)$.

Note that the operator $I_K : \mathcal{H} \to L^2(X,\rho_W)$ defined by $I_K f(x) = \langle f, K_x\rangle$ is linear and Hilbert-Schmidt since

$$\begin{aligned}\|I_K\|_{HS}^2 &= \sum_{j\geq 1}\|I_K e_j\|_{\rho_W}^2 = \int_X \sum_{j\geq 1}\langle K_x, e_j\rangle^2 \; d\rho_W(x)\\ &= \int_X K(x,x)m(x)\; d\rho(x) \leq \kappa\|m\|_\infty,\end{aligned}$$

where $\kappa = \sup_{x\in X} K(x,x)$. The operator $I_W^* : L^2(X,\rho_W) \to \mathcal{H}$ defined by

$$I_W^* f = \int_X \frac{1}{m} W_x f(x)\, d\rho(x)$$

is linear and bounded since, by Assumption A

$$\|\int_X \frac{1}{m} W_x f(x)\, d\rho(x)\|_{\mathcal{H}} \leq \int_X \|\frac{1}{m} W_x\|_{\mathcal{H}} |f(x)|\, d\rho(x) \leq C\|f\|_\rho \leq \frac{C}{c}\|f\|_{\rho_W}.$$

A direct computation shows that

$$I_W^* I_K = A_{\mathcal{H}} = I - \mathbb{L}_{\mathcal{H}} \qquad \sigma(A_{\mathcal{H}}) = 1 - \sigma(\mathbb{L}_{\mathcal{H}})$$

and

$$I_K I_W^* = I - \mathbb{L} \qquad\qquad \sigma(I_K I_W^*) = 1 - \sigma(\mathbb{L}).$$

Both $I_W^* I_K$ and $I_K I_W^*$ are Hilbert-Schmidt operators since they are composition of a bounded operator and Hilbert-Schmidt operator. Moreover, let $\sigma \neq 1$ and $v \in \mathcal{H}$ with $v \neq 0$ such that $\mathbb{L}_{\mathcal{H}} v = \sigma v$. Letting $u = I_K v$, then

$$\mathbb{L}u = (I - I_K I_W^*) I_K v = I_K \mathbb{L} v = \sigma u \quad \text{and} \quad I_W^* u = I_W^* I_K v = (1-\sigma)v \neq 0,$$

so that $u \neq 0$ and u is an eigenfunction of $\mathbb{L}$ with eigenvalue σ. Similarly we can prove that if $\sigma \neq 1$ and $u \in L^2(X,\rho), u \neq 0$ is such that $\mathbb{L}u = \sigma u$, then $v = \frac{1}{1-\sigma} I_W^* u$ is different from zero and is an eigenfunction of $\mathbb{L}_{\mathcal{H}}$ with eigenvalue σ.

We now show that $\mathbb{L}$ is a positive operator on $L^2(X,\rho_W)$, so that both $\mathbb{L}$ and $\mathbb{L}_{\mathcal{H}}$ have positive eigenvalues. Indeed, let $f \in L^2(X,\rho_W)$,

$$\begin{aligned}\langle \mathbb{L}f, f\rangle_{\rho_W} &= \int_X |f(x)|^2 m(x) d\rho(x) - \int_X \left(\int_X \frac{W(x,s)}{m(x)} f(s) d\rho(s)\right) \overline{f(x)} m(x) d\rho(x)\\ &= \frac{1}{2}\int_X\int_X \left[|f(x)|^2 W(x,s) - 2W(x,s) f(x)\overline{f(s)} + |f(s)|^2 W(x,s)\right] d\rho(s) d\rho(x)\\ &= \frac{1}{2}\int_X\int_X W(x,s)|f(x) - f(s)|^2 d\rho(s) d\rho(x) \geq 0,\end{aligned}$$

where we used that $W(x,s) = W(s,x) > 0$ and $m(x) = \int_X W(x,s)\, d\rho(s)$. Since W is real valued, it follows that we can always choose the eigenfunctions of $\mathbb{L}$ as real valued, and, as a consequence, the eigenfunctions of $\mathbb{L}_{\mathcal{H}}$.

Finally we prove that both $\mathbb{L}$ and $\mathbb{L}_{\mathcal{H}}$ admit a decomposition in terms of spectral projections – we stress that the spectral projection of $\mathbb{L}$ is orthogonal in $L^2(X, \rho_W)$, but not in $L^2(X, \rho)$.
Since $\mathbb{L}$ is a positive operator on $L^2(X, \rho_W)$, it is self-adjoint, as well as $I_K I_W^*$, hence the spectral theorem gives

$$I_K I_W^* = \sum_{j \geq 1} (1 - \sigma_j) P_j$$

where for all j, $P_j : L^2(X, \rho_W) \to L^2(X, \rho_W)$ is the spectral projection of $I_K I_W^*$ associated to the eigenvalue $1 - \sigma_j \neq 0$. Moreover note that P_j is also the spectral projection of $\mathbb{L}$ associated to the eigenvalue $\sigma_j \neq 1$. By definition P_j satisfies:

$$\begin{aligned}
P_j^2 &= P_j, \\
P_j^* &= P_j \qquad \text{the adjoint is in } L^2(X, \rho_W) \\
P_j P_i &= 0, \quad i \neq j, \\
P_j P_{ker(I_K I_W^*)} &= 0 \\
\sum_{j \geq 1} P_j &= I - P_{ker(I_K I_W^*)} = I - P_0
\end{aligned}$$

where $P_{ker(I_K I_W^*)}$ is the projection on the kernel of $I_K I_W^*$, that is, the projection P_0. Moreover the sum in the last equation converges in the strong operator topology. In particular we have

$$I_K I_W^* P_j = P_j I_K I_W^* = (1 - \sigma_j) P_j,$$

so that

$$\mathbb{L} = I - I_K I_W^* = \sum_{j \geq 1} \sigma_j P_j + P_0.$$

Let $Q_j : \mathcal{H} \to \mathcal{H}$ be defined by

$$Q_j = \frac{1}{1 - \sigma_j} I_W^* P_j I_K.$$

Then from the properties of the projections P_j we have,

$$\begin{aligned}
Q_j^2 &= \frac{1}{(1 - \sigma_j)^2} I_W^* P_j I_K I_W^* P_j I_K = \frac{1}{1 - \sigma_j} I_W^* P_j P_j I_K = Q_j, \\
Q_j Q_i &= \frac{1}{(1 - \sigma_j)(1 - \sigma_i)} I_W^* P_j I_K I_W^* P_i I_K = \frac{1}{1 - \sigma_i} I_W^* P_j P_i I_K = 0 \quad i \neq j.
\end{aligned}$$

Moreover,

$$\sum_{j \geq 1} (1 - \sigma_j) Q_j = \sum_{j \geq 1} (1 - \sigma_j) \frac{1}{1 - \sigma_j} I_W^* P_j I_K = I_W^* (\sum_{j \geq 1} P_j) I_K = I_W^* I_K - I_W^* P_{ker(I_K I_W^*)} I_K$$

so that

$$I_K I_W^* = \sum_{j \geq 1} (1 - \sigma_j) Q_j + I_W^* P_{ker(I_K I_W^*)} I_K,$$

where again all the sums are to be intended as converging in the strong operator topology. If we let $D = I_W^* P_{ker(I_K I_W^*)} I_K$ then

$$Q_j D = \frac{1}{1-\sigma_j} I_W^* P_j I_K I_W^* P_{ker(I_K I_W^*)} = I_W^* P_j P_{ker(I_K I_W^*)} = 0,$$

and, similarly $DQ_j = 0$. By construction $\sigma(D) = 0$, that is, D is a quasi-nilpotent operator. Equation (13) is now clear as well as the fact that $\ker D = \ker (I - \mathbb{L}_{\mathcal{H}})$. ■

Proof [Proof of Proposition 14] The proof is the same as the above proposition by replacing ρ with the empirical measure $\frac{1}{n}\sum_{i=1}^n \delta_{x_i}$. ■

Next we prove Theorem 20.

Proof [Proof of Theorem 20] We recall the following basic result. Let S and T two bounded operators acting on $\mathcal{H}$ and defined $C = I - ST$. If $\|C\| < 1$, then T has a bounded inverse and

$$T^{-1} - S = (I - C)^{-1} CS$$

where we note that $\|I - C\|^{-1} \leq \frac{1}{1-\|C\|}$ since $\|C\| < 1$, see Proposition 1.2 of Anselone (1971).

Let A and B two compact operators. Let Γ be a compact subset of the resolvent set of A and define

$$\delta^{-1} = \sup_{\lambda \in \Gamma} \|(\lambda - A)^{-1}\|,$$

which is finite since Γ is compact and the resolvent operator $(\lambda - A)^{-1}$ is norm continuous, see for example Anselone (1971) . Assume that

$$\|B - A\| < \delta,$$

then for any $\lambda \in \Gamma$

$$\|(\lambda - A)^{-1}(B - A)\| \leq \|(\lambda - A)^{-1}\| \|B - A\| \leq \delta^{-1} \|B - A\| < 1.$$

Hence we can apply the above result with $S = (\lambda - A)^{-1}$, $T = (\lambda - B)$, since

$$C = I - (\lambda - A)^{-1}(\lambda - B)^{-1} = (\lambda - A)^{-1}(B - A).$$

It follows that $(\lambda - B)$ has a bounded inverse and

$$(\lambda - B)^{-1} - (\lambda - A)^{-1} = (I - (\lambda - A)^{-1}(B - A))^{-1}(\lambda - A)^{-1}(B - A)(\lambda - A)^{-1}.$$

In particular, Λ is a subset of the resolvent set of B and, if $B - A$ is a Hilbert-Schmidt operator, so is $(\lambda - B)^{-1} - (\lambda - A)^{-1}$.

Let Λ be a finite set of non-zero eigenvalues. Let Γ be any simple closed curve with Λ inside and $\sigma(A) \setminus \Lambda$ outside. Let P be the spectral projection associated with Λ, then

$$P = \frac{1}{2\pi i} \int_\Gamma (\lambda - A)^{-1} \, d\lambda.$$

Applying the above result, it follows that Γ is a subset of the resolvent set of B and we let $\widehat{\Lambda}$ be the subset of $\sigma(B)$ inside Γ and $\widehat{P}$ the corresponding spectral projection, then

$$\begin{aligned}\widehat{P} - P &= \frac{1}{2\pi i}\int_\Gamma (\lambda - B)^{-1} - (\lambda - A)^{-1}\, d\lambda \\ &= \frac{1}{2\pi i}\int_\Gamma (I - (\lambda - A)^{-1}(B - A))^{-1}(\lambda - A)^{-1}(B - A)(\lambda - A)^{-1}\, d\lambda.\end{aligned}$$

It follows that

$$\|\widehat{P} - P\| \leq \frac{\ell(\Gamma)}{2\pi}\frac{\delta^{-2}\|B - A\|}{1 - \delta^{-1}\|B - A\|} = \frac{\ell(\Gamma)}{2\pi\delta}\frac{\|B - A\|}{\delta - \|B - A\|}.$$

In particular if $\|B - A\| \leq \frac{\delta^2}{\delta + \ell(\Gamma)/2\pi} < \delta$, $\|\widehat{P} - P\| \leq 1$ so that the dimension of the range of P is equal to the dimension of the range of $\hat{P}$. It follows that $\widehat{\Lambda}$ is not empty.
If $B - A$ is a Hilbert-Schmidt operator, we can replace the operator norm with the Hilbert-Schmidt norm, and the corresponding inequality is a consequence of the fact that the Hilbert-Schmidt operator are an ideal. ■

Appendix B. Spectral theorem for non-self-adjoint compact operators

Let $A : \mathcal{H} \to \mathcal{H}$ be a compact operator. The spectrum $\sigma(A)$ of A is defined as the set of complex number such that the operator$(A - \lambda I)$ does not admit a bounded inverse, whereas the complement of $\sigma(A)$ is called the resolvent set and denoted by $\rho(A)$. For any $\lambda \in \rho(A)$, $R(\lambda) = (A - \lambda I)^{-1}$ is the resolvent operator, which is by definition a bounded operator. We recall the main results about the spectrum of a compact operator (Kato, 1966)

Proposition 23 *The spectrum of a compact operator A is a countable compact subset of $\mathbb{C}$ with no accumulation point different from zero, that is,*

$$\sigma(A) \setminus \{0\} = \{\lambda_i \mid i \geq 1,\ \lambda_i \neq \lambda_j \text{ if } i \neq j\}$$

with $\lim_{i\to\infty} \lambda_i = 0$ if the cardinality of $\sigma(A)$ is infinite. For any $i \geq 1$, λ_i is an eigenvalue of A, that is, there exists a nonzero vector $u \in \mathcal{H}$ such that $Au = \lambda_i u$. Let Γ_i be a rectifiable closed simple curve (with positive direction) enclosing λ_i, but no other points of $\sigma(A)$, then the operator defined by

$$P_{\lambda_i} = \frac{1}{2\pi i}\int_{\Gamma_i} (\lambda I - A)^{-1} d\lambda$$

satisfies

$$P_{\lambda_i} P_{\lambda_j} = \delta_{ij} P_{\lambda_i} \qquad \text{and} \qquad (A - \lambda_i) P_{\lambda_i} = D_{\lambda_i} \qquad \text{for all } i, j \geq 1,$$

where D_{λ_i} is a nilpotent operator such that $P_{\lambda_i} D_{\lambda_i} = D_{\lambda_i} P_{\lambda_i} = D_{\lambda_i}$. In particular the dimension of the range of P_{λ_i} is always finite.

We notice that P_{λ_i} is a projection onto a finite dimensional space $\mathcal{H}_{\lambda_i}$, which is left invariant by A. A nonzero vector u belongs to $\mathcal{H}_{\lambda_i}$ if and only if there exists an integer $m \leq \dim \mathcal{H}_{\lambda_i}$ such that $(A-\lambda)^m u = 0$, that is, u is a generalized eigenvector of A. However, if A is symmetric, for all $i \geq 1$, $\lambda_i \in \mathbb{R}$, P_{λ_i} is an orthogonal projection and $D_{\lambda_i} = 0$ and it holds that

$$A = \sum_{i \geq 1} \lambda_i P_{\lambda_i}$$

where the series converges in operator norm. Moreover, if $\mathcal{H}$ is infinite dimensional, $\lambda = 0$ is always in $\sigma(A)$, but it can be or not an eigenvalue of A.

If A be a compact operator with $\sigma(A) \subset [0, \|A\|]$, we introduce the following notation. Denoted by p_A the cardinality of $\sigma(A) \setminus \{0\}$ and given an integer $1 \leq N \leq p_A$, let $\lambda_1 > \lambda_2 > \ldots, \lambda_N > 0$ be the first N nonzero eigenvalues of A, sorted in a decreasing way. We denote by P_N^A the spectral projection onto all the generalized eigenvectors corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_N$. The range of P_N^A is a finite-dimensional vector space, whose dimension is the sum of the algebraic multiplicity of the first N eigenvalues. Moreover

$$P_N^A = \sum_{j=1}^{N} P_{\lambda_j} = \frac{1}{2\pi i} \int_\Gamma (\lambda I - A)^{-1} d\lambda$$

where Γ is a rectifiable closed simple curve (with positive direction) enclosing $\lambda_1, \ldots, \lambda_N$, but no other points of $\sigma(A)$.

References

P. M. Anselone. *Collectively compact operator approximation theory and applications to integral equations*. Prentice-Hall Inc., Englewood Cliffs, N. J., 1971. With an appendix by Joel Davis, Prentice-Hall Series in Automatic Computation.

N. Aronszajn. Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68:337–404, 1950.

F. Bauer, S. Pereverzev, and L. Rosasco. On regularization algorithms in learning theory. *J. Complexity*, 23(1):52–72, 2007.

M. Belkin. *Problems of Learning on Manifolds*. PhD thesis, University of Chicago, USA, 2003.

M. Belkin and P. Niyogi. Towards a theoretical foundation for Laplacian-based manifold methods. In *Proceedings of the 18th Conference on Learning Theory (COLT)*, pages 486–500, 2005.

M. Belkin and P. Niyogi. Convergence of Laplacian eigenmaps. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*, pages 129–136. MIT Press, Cambridge, MA, 2007.

V.I. Burenkov. *Sobolev spaces on domains*. B. G. Teubuer, Stuttgart-Leipzig, 1998.

R. R. Coifman and S. Lafon. Geometric harmonics: A novel tool for multiscale out-of-sample extension of empirical functions. *ACHA*, 21:31–52, 2006.

E. De Vito, A. Caponnetto, and L. Rosasco. Model selection for regularized least-squares algorithm in learning theory. *Foundations of Computational Mathematics*, 5(1):59–85, February 2005a.

E. De Vito, L. Rosasco, A. Caponnetto, U. De Giovannini, and F. Odone. Learning from examples as an inverse problem. *Journal of Machine Learning Research*, 6:883–904, May 2005b.

E. De Vito, L. Rosasco, and A. Caponnetto. Discretization error analysis for Tikhonov regularization. *Anal. Appl. (Singap.)*, 4(1):81–99, 2006.

E. Giné and V. Koltchinskii. Empirical graph Laplacian approximation of Laplace-Beltrami operators: Large sample results. *High Dimensional Probability*, 51:238–259, 2006.

M. Hein. Uniform convergence of adaptive graph-based regularization. In *Proceedings of the 19th Annual Conference on Learning Theory (COLT)*, pages 50–64, New York, 2006. Springer.

M. Hein, J-Y. Audibert, and U. von Luxburg. From graphs to manifolds - weak and strong pointwise consistency of graph Laplacians. In *Proceedings of the 18th Conference on Learning Theory (COLT)*, pages 470–485, 2005. Student Paper Award.

T. Kato. *Perturbation theory for linear operators*. Springer, Berlin, 1966.

T. Kato. Variation of discrete spectra. *Commun. Math. Phys.*, III:501–504, 1987.

V. Koltchinskii. Asymptotics of spectral projections of some random matrices approximating integral operators. *Progress in Probabilty*, 43, 1998.

V. Koltchinskii and E. Giné. Random matrix approximation of spectra of integral operators. *Bernoulli*, 6:113–167, 2000.

S. Lafon. *Diffusion Maps and Geometric Harmonics*. PhD thesis, Yale University, USA, 2004.

S. Lang. *Real and Functional Analysis*. Springer, New York, 1993.

S. Mendelson and A. Pajor. Ellipsoid approximation with random vectors. In *Proceedings of the 18th Annual Conference on Learning Theory (COLT)*, pages 429–433, New York, 2005. Springer.

S. Mendelson and A. Pajor. On singular values of matrices with independent rows. *Bernoulli*, 12(5):761–773, 2006.

I. Pinelis. An approach to inequalities for the distributions of infinite-dimensional martingales. *Probability in Banach Spaces, 8, Proceedings of the 8th International Conference*, pages 128–134, 1992.

L. Schwartz. Sous-espaces hilbertiens d'espaces vectoriels topologiques et noyaux associés (noyaux reproduisants). *J. Analyse Math.*, 13:115–256, 1964. ISSN 0021-7670.

J. Shawe-Taylor, N. Cristianini, and J. Kandola. On the concentration of spectral properties. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14*, pages 511–517, Cambridge, MA, 2002. MIT Press.

J. Shawe-Taylor, C. Williams, N. Cristianini, and J. Kandola. On the eigenspectrum of the gram matrix and the generalisation error of kernel pca. *to appear in IEEE Transactions on Information Theory*, 51, 2004. URL `http://eprints.ecs.soton.ac.uk/9779/`.

A. Singer. From graph to manifold Laplacian: The convergence rate. *Appl. Comput. Harmon. Anal.,*, 21:128–134, 2006.

S. Smale and D.X. Zhou. Learning theory estimates via integral operators and their approximations. *submitted*, 2005. retrievable at http://www.tti-c.org/smale.html.

S. Smale and D.X. Zhou. Geometry of probability spaces. *preprint*, 2008. retrievable at http://www.tti-c.org/smale.html.

U. Von Luxburg, O. Bousquet, and M. Belkin. On the convergence of spectral clustering on random samples: the normalized case. In *Proceedings of the 17th Annual Conference on Learning Theory (COLT 2004*, pages 457–471. Springer, 2004.

U. von Luxburg, M. Belkin, and O. Bousquet. Consistency of spectral clustering. *Ann. Statist.*, 36(2):555–586, 2008.

Y. Yao, L. Rosasco, and A. Caponnetto. On early stopping in gradient descent learning. *Constr. Approx.*, 26(2):289–315, 2007.

L. Zwald and G. Blanchard. On the convergence of eigenspaces in kernel principal component analysis. In *NIPS*, 2006.

Laurent Zwald, Olivier Bousquet, and Gilles Blanchard. Statistical properties of kernel principal component analysis. In *Learning theory*, volume 3120 of *Lecture Notes in Comput. Sci.*, pages 594–608. Springer, Berlin, 2004.