Alma Mater Studiorum · Università di Bologna

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI
Corso di Laurea in Matematica

The Karhunen-Loève Theorem

Tesi di Laurea in Analisi Funzionale

Relatore:
Chiar.mo Prof.
Nicola Arcozzi

Correlatore:
Chiar.mo Prof.
Massimo Campanino

Presentata da:
Giordano Giambartolomei

III Sessione
Anno Accademico 2014-2015

The Karhunen-Loève Theorem

Acknowledgments. I would like to express my thanks and my gratitude firstly to my supervisor Nicola Arcozzi, since without his interest and curiosity, without the fruitful conversation with him, my dissertation would not have been so rich. I hardly imagine I could have accomplished it without the support from him.

Many thanks to my cosupervisor Masimo Campanino, for his precious help in the review.

I would like to thank all the professors, I have studied with in these three years and a half, for their precious lectures, the material wisely suggested and the patient consulting hours and exams. Without their guide, I would have never managed to achieve a clear understanding and a good practice in so many branches of Mathemathics.

My special thanks are also to my beloved and supportive family, and to Si Hui Elisa Chen, whose love helped me overcome years of frenzy work.

My warmest thanks to Simone Buoncompagni, for his lifelong friendship and his always prompt assistance.

Abstract

The univariate Karhunen-Loève Expansion is the decomposition of a continuous-parameter second-order stochastic process into uncorrelated random coefficients. In the present dissertation, the expansion is achieved analytically by projecting the process, considered over a finite time interval $[a,\, b]$, onto a deterministic orthonormal basis obtained from the covariance Hilbert-Schmidt operator's eigenfunctions, which correspond to positive eigenvalues.

Basically, the idea of the method is to first find the positive eigenvalues of the Hilbert-Schmidt integral operator, with the covariance function of the process as a kernel. At every time of the interval, the process is projected onto an orthonormal basis for the space spanned by the eigenfunctions corresponding to positive eigenvalues. This procedure generates random coefficients, which turn to be uncorrelated centered random variables.

The expansion series, a countably infinite linear combination of the eigenfunctions together with the random projections, converges in *mean square* to the process, uniformly over the time interval. Furthermore, in the probability space $(\Omega, \mathcal{F}, P)$ the convergence gets *almost surely* if the process is Gaussian.

Many other expantions exist, however the peculiarity of the Karhunen-Loève basis is that the expantion is optimal in terms of the *total mean square error* resulting from the truncation of the series. Such feature has made this method and its generalizations very successful in applied disciplines.

La trasformata di Karhunen-Loève monodimensionale è la decomposizione di un processo stocastico del secondo ordine a parametrizzazione continua in coefficienti aleatori scorrelati. Nella presente dissertazione, la trasformata è ottenuta per via analitica, proiettando il processo, considerato in un intervallo di tempo limitato $[a, b]$, su una base deterministica ottenuta dalle autofunzioni dell'operatore di Hilbert-Schmidt di covarianza corrispondenti ad autovalori positivi.

Fondamentalmente l'idea del metodo è, dal primo, trovare gli autovalori positivi dell'operatore integrale di Hilbert-Schmidt, che ha in kernel la funzione di covarianza del processo. Ad ogni tempo dell'intervallo, il processo è proiettato sulla base ortonormale dello span delle autofunzioni dell'operatore di Hilbert-Schmidt che corrispondono ad autovalori positivi. Tale procedura genera coefficienti aleatori che si rivelano variabili aleatorie centrate e scorrelate.

L'espansione in serie che risulta dalla trasformata è una combinazione lineare numerabile di coefficienti aleatori di proiezione ed autofunzioni convergente in media quadratica al processo, uniformemente sull'intervallo temporale. Se inoltre il processo è Gaussiano, la convergenza è *quasi sicuramente* sullo spazio di probabilità $(\Omega, \mathcal{F}, P)$.

Esistono molte altre espansioni in serie di questo tipo, tuttavia la trasformata di Karhunen-Loève ha la peculiarità di essere ottimale rispetto all'errore totale in media quadratica che consegue al troncamento della serie. Questa caratteristica ha conferito a tale metodo ed alle sue generalizzazioni un notevole successo tra le discipline applicate.

Table of contents

Introduction

The present dissertation is primarily concerned with the analytical proof of the univariate Karhunen-Loève Expansion Theorem, also known as Kosambi-Karhunen-Loève Expansion Theorem. The Karhunen-Loève Expansion is to be seen as the stochastic parallel of one of the most famous analytical methods ever: the Fourier Expansion. In the Fourier Analysis a periodic real function is decomposed into a countably infinite linear combination of an orthonormal basis elements, the Fourier sinusoidal basis of L_2, together with projection coefficients.

Stochastic functions given by similar series were first inquired by Kosambi (1943). However, Karhunen (1946), Karhunen (1947) and Loève (1948) are the earliest works on a proper generalization of such method in the field of Stochastic Analysis, followed by Ash (1965). Historically, it created a bond between Stochastic Analysis and Information Theory.

The Karhunen-Loève Expansion operates pretty much the same as Fourier Expansion, but on $L^2(\Omega,\, \mathcal{F},\, P)$ continuous-parameter random functions $\{X_t,\, t \in [a,\, b]\}$ considered over a finite time interval $[a,\, b]$. The basis, the random function is projected onto, is found in Section 2 through the study of the Hilbert-Schmidt integral operator, which has the covariance of the random function as a kernel. The Karhunen-Loève Expansion decompose the stocastic process by projecting every variable onto an orthonormal basis for the space spanned by the operator's eigenfunctions, which correspond to the positive eigenvalues thereof. The coefficients of the infinite linear combination are therefore expected to be random variables, since, for every $t \in [a,\, b]$, they are the projection of a random variable onto a deterministic orthogonal basis. As a result these random coefficients are also orthogonal in $L^2(\Omega)$, namely they are uncorrelated.

The bond between Functional Analysis and Stochastic Analysis is here represented by the covariance Hilber-Schmidt operator. Being the whole method basically an eigenvalue problem, the Fredholm Theory is particularly involved, as will become clear in the first two sections of the work, which provide the basics of Functional Analysis required for the Mercer's Theorem. In addition, the second-order Fredholm integral equation happens to be the main character of Section 3, due to its connection with the truncation error in the Karhunen-Loève Expansion.

Unlike the majority of the material available, the present dissertation is not only focused on the proof of the expansion and its mean square convergence uniformly on the time interval, argued in the first part of Section 4. The proof of the optimality of the Karhunen-Loève basis with respect to the total mean square error resulting from truncation is very detailed as well. Section 3 and the second part of Section 4 are dedicated to it. The optimality is fundamental, since in computational applications only a finite number of terms of the decomposition are involved, due to the truncation of the expansion series. There are of course different possible expansions of stochastic processes, however the one that minimizes the total mean square error is the Kahrunen-Loève Expansion, because of the choice of the basis.

Eventually, in Section 5 the fundamental property of the Gaussian processes expansion is derived: the random projections are not only uncorrelated, but independent and Gaussian. As a result, the Gaussian processes expansion achieves convergence *almost surely* in the probability space $(\Omega,\, \mathcal{F},\, P)$. Indeed, this class of processes is amongst the most notable and prolific applications of the Karhunen-Loève decomposition. Two analytical examples are given and fully computed with great detail: the Brownian motion and the Brownian bridge.

The univariate Karhunen-Loève Expansion has known various generalizations and applications up to the recent years. Should at least be mentioned: the multivariate Kahrunen-Loève expansion, the conditional Kahrunen-Loève Expansion, the Kahrunen-Loève Expansion of a continuous spectrum operator. As to the last one, a point should me made: in the present dissertation we deal only with the favorable case of discrete spectrum operators. However, the general Spectral Theorem admits self-adjoint operators with continuous spectrum, and therefore the Karhunen-Loève Theorem can be extended in this direction, wich is done in P. E. T. Jorgensen and M.-S. Song (2007). The present work does not go so far. Nonetheless, we would like to conclude this introduction with a few examples of simple implementations.

Many applied areas have confirmed the Karhunen-Loève decomposition as a very useful representation method: numerical methods, finite element methods, model reduction, functional data analysis, finance, pattern recognition, signal detection, machine learning, etc. On a final note, one of the most immediate applications of the Karhunen-Loève Expansion we would like to mention is the Discrete Karhunen-Loève Expansion. There are uncountable versions of it. One of them is the Linear Karhunen-Loève Approximation of signals, very useful to approximate a whole class of signals with a finite number of terms of the Karhunen-Loève basis. Modelling a signal as a finite dimensional random vector, the best finite approximation thereof is provided by the vectors of the Karhunen-Loève basis, which here happen to diagonalize the covariance matrix of the random vector. It must be mentioned that also non-linear approximations exist, where the vectors of the Karhunen-Loève basis are chosen adaptively to the properties of the signals.

As the examples of Section 5 might suggest, the numerical cost of finding eigenvalues and eigenfunctions of the covariance kernel operator may not be little. However, the example above underlines that when the process is not continuous parameter but is a finite and discrete set of random variables, standard algebra is only required, and the numerical cost of the method dips. In Fukunaga and Koontz (1970) it becomes cleare that this results in the Karhunen-Loève Expansion usefulness not only for discrete processes, since the decomposition can be applied to a finite sampled continuous parameter process, so that, under appropriate conditions, the problem ir reduced again to the discrete case. Basically, this is the idea behind Pricipal Component Analysis. Adaptive optics systems sometimes reconstruct wave-front phase information thanks to such method. The Singular Value Decomposition is closely related to the Kahrunen-Loève Expansion as well, andits applications to image processing testify the success of this technique.

1 Operators on Hilbert spaces

The Karhunen-Loève Expansion shall be derived mainly analytically, according to the scheme followed by Ash (1990), pp. 262-281. However, the proves have been often expanded, sometimes completely changed (mostly from Theorem 1.5 to 1.10 in the present section). In doing so, the main references for the basics of Functional Analysis have been primarily Brézis (2010) and Reed and Simon (1980).

In this first section we recall and prove some useful results about linear compact symmetric operators on any Hilbert space H. Actually, we assume H a real Hilbert space, since in Section 2 we shall work on the Hilbert space $L^2[a,b] := \mathcal{L}^2[a,b]/\sim$, where $\mathcal{L}^2[a,b]$ is the space of all the real-valued Borel-measurable functions f on the interval $[a,b] \subset \mathbb{R}$ such that $\int_a^b f^2(t)dt < \infty$; the equivalence relation $\sim$ identifies all the functions that are equal *almost everywhere*.

In those sections, which are primarily focused with $L^2[a,\, b]$, the notation for the inner product $(\cdot,\cdot)$ and for the norm $\|\cdot\|$ shall be adapted by simply adding a subscript 2, so that $(f,g)_2 := \int_a^b f(t)g(t)dt$ and $\|f\|_2 := \left[\int_a^b f^2(t)dt\right]^{\frac{1}{2}}$. We shall use 0 both for the null real scalar and the vector 0_H.

1.1 Continuous compact linear operators

Lemma 1.1. Let $A: H \longrightarrow H$ be a linear operator. Then A is continuous if and only if A is bounded.

Proof. If A is unbounded, $\|A\|_{\text{op}} := \sup_{\|x\|=1} \|Ax\| > M$ for every $M \in \mathbb{R}$. Then we ca define a sequence in H, $\{x_n\}_{n\in\mathbb{N}}$ such that for every n, $\|x_n\| = 1$ and $\lim_{n\to\infty} \|Ax_n\| = \infty$. Indeed, for every n there is an $x_n \in H$ such that $\|Ax_n\| > n$ and by the *axiom of choice* we can pick it up in order to define the sequence. Now, define another sequence in H, $\{y_n\}_{n\in\mathbb{N}}$, where $y_n := \frac{x_n}{\|Ax_n\|}$. Since $\|x_n\| = 1$, $\|y_n\| = \frac{1}{\|Ax_n\|} \longrightarrow 0$, as $n \longrightarrow \infty$, hence $y_n \longrightarrow 0$ as $n \to \infty$. However, by the linearity of A we have that $\|Ay_n\| = \left\|\frac{Ax_n}{\|Ax_n\|}\right\| \equiv 1$. Therefore if the limit of Ay_n existed, it should be non zero. Since $A \lim_{n\to\infty} y_n = A0 = 0$, it turns out that A is discontinuous, because there are only two possibilities: one is that the $\lim_{n\to\infty} Ay_n$ exists, but since it is non-zero, we have discontinuity because $\lim_{n\to\infty} Ay_n \neq A \lim_{n\to\infty} y_n$; the other is that this limit does not exist, but since $\lim_{n\to\infty} Ay_n = \lim_{y_n\to 0} Ay_n$ does not exist, and $A0 = 0$, again we have discontinuity.

Conversely if A is bounded, there is a constant $C > 0$ such that $\|\mathrm{Ax}\| \leqslant C\|x\|$. Thus for every $x, y \in H$, $\|Ax - Ay\| = \|A(x-y)\| \leqslant C\|x-y\|$, which by definition is the lipschitzianity of A. But it is well known that if A is *Lipschitz*, then A is continuous. $\square$

Remark 1.2. We recall that by definition of *operatorial norm* for a linear operator A, it holds that:

$$||Ax|| = \left\|\|x\|A\left(\frac{x}{\|x\|}\right)\right\| = \|x\|\left\|A\left(\frac{x}{\|x\|}\right)\right\| \leqslant \|x\|\sup_{x\in H}\left\|A\left(\frac{x}{\|x\|}\right)\right\| = \|x\|\|A\|_{\text{op}}.$$

Lemma 1.3. Let A: $H \longrightarrow H$ be a linear operator. If A is compact, then A is continuous.

Proof. By definition A is compact if for every bounded sequence $\{x_n\}$ in H, the sequence $\{Ax_n\}$ has a convergent subsequence in H.

Suppose that A is discontinuous. Then by Lemma 1.1 A is unbounded. Thus there is a sequence $\{x_n\}$ in H such that $\|Ax_n\| \longrightarrow \infty$ as $n \to \infty$. Define $y_n := \|Ax_n\|$. Since $\{y_n\}$ is a real-valued divergent sequence, every subsequence $\{y_{n_j}\}$ is divergent, where $y_{n_j} := \|Ax_{n_j}\|$. Since every subsequence $\{\|Ax_{n_j}\|\}$ diverges, there is no possibility for every subsequence $\{Ax_{n_j}\}$ to converge. Therefore there is no convergent subsequence of $\{Ax_n\}$, namely A is not compact. $\square$

For each compact linear operator A: $H \longrightarrow H$ it is useful to consider the associate *Fredholm operator* $A_\lambda := A - \lambda I$, where λ is a real scalar and I is the identity operator on H.

Definition 1.4. $\varrho(A) := \{\lambda \in \mathbb{R}: A_\lambda \, \text{is bijective}\}$ and $\sigma(A) := \mathbb{R} \backslash \varrho(A)$.

The terminology *resolvent set* for $\varrho(A)$ and *spectrum* for $\sigma(A)$ is not appropriate, since we are not dealing with complex scalars.

In addition we remark that equivalently $\varrho(A) := \{\lambda \in \mathbb{R}: A_\lambda^{-1} \, \text{exists}\}$ and $\sigma(A) := \{\lambda \in \mathbb{R}: A_\lambda^{-1} \, \text{does not exist}\}$.

We recall here a standard notation we shall adopt henceforth: $\mathcal{N}(A)$ denotes the *null space* of A, $\mathcal{N}(A) := \{x \in H: Ax = 0\}$, and $\mathcal{R}(A)$ denotes the *range* of A, $\mathcal{R}(A) := \{y \in H: \exists x \in H, y = Ax\}$. Both are trivially subspaces of H.

Theorem 1.5. Let A: $H \longrightarrow H$ be a continuous linear operator. If A is invertible, then A^{-1} is continuous.

Proof. By the *Open Mapping Theorem* every continuous linear surjective operator between two Banach spaces, is an open map. Hilbert spaces are Banach spaces. Then A, which is linear, continuous and bijective on a Hilbert space, is an open map. Thus if $O \subset H$ is open, then $A(O) \subset H$ is open. Now, it is trivial that $(A^{-1})^{-1} = A$. Thus if $O \subset H$ is open, then $(A^{-1})^{-1}(O) \subset H$ is open. This is the topological definition of continuity for A^{-1}: $H \longrightarrow H$. $\square$

We now adapt the famous Riesz's Lemma for Banach spaces to the case of real Hilbert spaces.

Lemma 1.6. (Riesz's Lemma) Let $C \subsetneq H$ be a closed subspace of a Hilbert space. Then there is an $x \in H$ such that $\|x\| = 1$ and $\text{dist}(x, C) = 1$.

Proof. Let $0 \neq y \in C^\perp$. This y exists, since $C^\perp \supsetneq \{0\}$. In fact $C \subsetneq H$, then there is $w \in H \backslash C$. This $w \neq 0$ since $0 \in C$. If p is the projection of w onto C, then we can take $y = w - p \in C^\perp$.

Now, $\operatorname{dist}(y, C) > 0$, since C is closed. If we eventually define $x := \frac{y}{\|y\|}$ the result follows, since $\|x\| = 1$ and for every $z \in C$

$$\|x - z\| = \left\| \frac{y}{\|y\|} - z \right\| = \left\| \frac{y - z\|y\|}{\|y\|} \right\| = \|y - z\|y\|\| \frac{1}{\|y\|} = \frac{\sqrt{\|y\|^2 - 2\|y\|(y, z) + \|y\|^2\|z\|^2}}{\|y\|} = \frac{\sqrt{\|y\|^2 + \|y\|^2\|z\|^2}}{\|y\|} = \sqrt{1 + \|z\|^2}.$$

Thus by definition of $\operatorname{dist}(x, C) := \inf_{z \in C} \|x - z\| = \|x - 0\| = 1$. □

Theorem 1.7. Let B_H be the *unitary ball* of a normed space H. If B_H is compact, then the dimension of H is finite.

Proof. *Per absurdum* assume the dimension of H is infinite. Then we can consider a sequence of finite-dimensional subspaces H_n, such that $H_{n-1} \subsetneq H_n$. By Lemma 1.6, we can define a sequence $\{x_n\}$ such that $x_n \in H_n, \|x_n\| = 1, \operatorname{dist}(x_n, H_{n-1}) = 1$. Then for every $m < n$, $\|x_n - x_m\| \geqslant 1$. Therefore, there is no convergent subsequence of $\{x_n\}$, which is absurd. Indeed, $\{x_n\}$ is a sequence in B_H, which is assumed compact, namely, from every sequence in B_H can be extracted a convergent subsequence. □

Lemma 1.8. Let $A: H \longrightarrow H$ be a compact linear operator and λ a nonzero real number. Then $\mathcal{N}(A_\lambda)$ is a finite-dimensional subspace of H.

Proof. We shall prove the compactness of $B_{\mathcal{N}(A_\lambda)}$. Then by Theorem 1.7 the result shall follow. Now, it is sufficient to notice that $0 = A_\lambda(B_{\mathcal{N}(A_\lambda)}) = A(B_{\mathcal{N}(A_\lambda)}) - \lambda B_{\mathcal{N}(A_\lambda)}$. Thus $\lambda B_{\mathcal{N}(A_\lambda)} = A(B_{\mathcal{N}(A_\lambda)}) \subset A(B_H) \subset \overline{A(B_H)}$. Since A is a compact operator, by definition $\overline{A(B_H)}$ is compact. Thus $\lambda B_{\mathcal{N}(A_\lambda)}$ is a closed subset of a compact set, that is $\lambda B_{\mathcal{N}(A_\lambda)}$ is compact. Since $\lambda B_{\mathcal{N}(A_\lambda)}$ is homeomorfic to $B_{\mathcal{N}(A_\lambda)}$, the result follows. □

Lemma 1.9. Let $A: H \longrightarrow H$ be a compact linear operator and λ a nonzero real number. Then $\mathcal{R}(A_\lambda)$ is a closed subspace of H.

Proof. Let $\{x_n\}$ be a sequence in H and suppose $\mathcal{R}(A_\lambda) \ni y_n := A_\lambda(x_n) \longrightarrow y$ as $n \to \infty$. Since a subset of a normed space is closed if and only if every sequence in the subset, which converges in the space, converges in the subset as well, we need to prove that $y \in \mathcal{R}(A_\lambda)$.

For every x_n, let $d_n := \operatorname{dist}(x_n, \mathcal{N}(A_\lambda))$. By the Lemma 1.8 the dimension of $\mathcal{N}(A_\lambda)$ is finite, therefore there exists $z_n \in \mathcal{N}(A_\lambda)$ such that $d_n = \|x_n - z_n\|$. Indeed, by the definition of distance from a set, $d_n = \inf_{z \in N(A_\lambda)} \|x_n - z\|$, however this *infimum* is actually a *minimum*, since it is achieved at a specific z_n in $\mathcal{N}(A_\lambda)$. Since $\mathcal{N}(A_\lambda)$ is a finite dimensional (say m – dimensional) subspace of H, $\mathcal{N}(A_\lambda)$ happens to be homeomorfic to $\mathbb{R}^m$. Because of the completeness of $\mathbb{R}^m$, $\mathcal{N}(A_\lambda)$ turns to be a *complete* subspace of H, thus it is closed. Having say that, take the *minimizing sequence* $\{z_n^k\}_{k \in \mathbb{N}}$ such that $\|x_n - z_n^k\| \longrightarrow \inf_{z \in \mathcal{N}(A_\lambda)} \|x_n - z\|$, as $k \to \infty$: the minimizing sequence is well defined and converges in H by definition of *infimum*. Since $\mathcal{N}(A_\lambda)$ is closed, it contains all its accumulation points, and the minimizing sequence converges in $\mathcal{N}(A_\lambda)$. Thus the *infimum* is achieved at a point of $\mathcal{N}(A_\lambda)$. We call it z_n.

Now, take the previous sequence y_n: since $z_n \in \mathcal{N}(A_\lambda)$, it holds that $y_n = A_\lambda(x_n - z_n) = A(x_n - z_n) - \lambda(x_n - z_n)$. Assume that $\{x_n - z_n\}$ is bounded (we will prove this in a while), then by the compactness of A we shall extract from $\{A(x_n - z_n)\}$ a convergent subsequence $\{A(x_{n_j} - z_{n_j})\}$. Let the limit of $\{A(x_{n_j} - z_{n_j})\}$ be h. Since $y_{n_j} = A(x_{n_j} - z_{n_j}) - \lambda(x_{n_j} - z_{n_j})$, we have that $(x_{n_j} - z_{n_j}) = \frac{1}{\lambda}[A(x_{n_j} - z_{n_j}) - y_{n_j}]$. Therefore $(x_{n_j} - z_{n_j}) \longrightarrow \frac{1}{\lambda}(h - y)$, as $j \to \infty$. Let $w :=\frac{h-y}{\lambda}$. Then by the continuity of A_λ we have that $A_\lambda(w) = \lim_{j\to\infty} A_\lambda(x_{n_j} - z_{n_j}) = \lim_{j\to\infty} A_\lambda(x_{n_j}) - 0 = \lim_{j\to\infty} A_\lambda(x_{n_j}) = y$, therefore $y \in \mathcal{R}(A_\lambda)$.

The boundness of $\{x_n - z_n\}$ remains to be proved. We show that $\{\|x_n - z_n\|\}$ is non-divergent (which gives us that $\{x_n - z_n\}$ is bounded). It is well known that if a real sequece is divergent, then every subsequence is divergent. Therefore if there is a non-divergent subsequence, the sequence is not divergent. We will show that $\{\|x_n - z_n\|\}$ has no divergent subsequence. Suppose *per absurdum* that there is a subsequence $\{\|x_{n_j} - z_{n_j}\|\}$ such that $\|x_{n_j} - z_{n_j}\| \longrightarrow \infty$, as $j \to \infty$. Then if we consider the trivially bounded sequence $\{w_n\}$ defined by $w_n :=\frac{x_n - z_n}{\|x_n - z_n\|}$, its subsequence $w_{n_j} :=\frac{x_{n_j} - z_{n_j}}{\|x_{n_j} - z_{n_j}\|}$ is such that

$$A_\lambda(w_{n_j}) = \frac{A_\lambda(x_{n_j} - z_{n_j})}{\|x_{n_j} - z_{n_j}\|} = \frac{y_{n_j}}{\|x_{n_j} - z_{n_j}\|} \underset{j\to\infty}{\longrightarrow} \frac{y}{\infty} = 0.$$

Now, A is compact, w_{n_j} (and λw_{n_j} consequently) is bounded by construction, therefore $A w_{n_j}$ admits a convergent subsequence, which implies that it cannot diverge. Since $A w_{n_j} - \lambda w_{n_j} \longrightarrow 0$, as $j \to \infty$, and neither $A w_{n_j}$ nor λw_{n_j} diverge, both $A w_{n_j}$ and λw_{n_j} must converge to a point u. This point u turns to be in $\mathcal{N}(A_\lambda)$, because by continuity on one hand $A w_{n_j} - \lambda w_{n_j} = A_\lambda(w_{n_j}) = \frac{1}{\lambda} A_\lambda(\lambda w_{n_j}) \longrightarrow \frac{1}{\lambda} A_\lambda(u)$, $j \to \infty$, on the other hand by above $A w_{n_j} - \lambda w_{n_j} \longrightarrow 0$, $j \to \infty$, which means $A_\lambda(u) = 0$.

Now, by the construction of z_{n_j}

$$\operatorname{dist}(w_{n_j}, \quad \mathcal{N}(A_\lambda)) \quad = \quad \frac{\operatorname{dist}(x_{n_j} - z_{n_j}, \mathcal{N}(A_\lambda))}{\|x_{n_j} - z_{n_j}\|} \quad = \quad \frac{\inf\limits_{v \in \mathcal{N}(A_\lambda)} \|(x_{n_j} - z_{n_j}) - v\|}{\|x_{n_j} - z_{n_j}\|} \quad =$$

$$\frac{\inf\limits_{v \in N(A_\lambda)} \|x_{n_j} - (z_{n_j} + v)\|}{\|x_{n_j} - z_{n_j}\|} = \frac{\|x_{n_j} - z_{n_j}\|}{\|x_{n_j} - z_{n_j}\|} = 1, \text{for every } j.$$

However, by the continuity of the distance we get an *absurdum*:

$$1 = \lim_{j\to\infty} \operatorname{dist}(w_{n_j}, \mathcal{N}(A_\lambda)) = \frac{1}{\lambda} \operatorname{dist}(u, \mathcal{N}(A_\lambda)) = 0, \text{since } u \in \mathcal{N}(A_\lambda).$$

We must conclude that there is no divergent subsequence of $\{\|x_n - z_n\|\}$, which results consequently non-divergent. Therefore $\{x_n - z_n\}$ is bounded. $\square$

Theorem 1.10. Let $A: H \longrightarrow H$ be a compact linear operator and λ a nonzero real number. Then A_λ is injective (if and) only if A_λ is surjective.

Proof. Only the sufficient implication will be showed. Suppose that A_λ is injective. The range $H_1 := \mathcal{R}(A_\lambda) \subseteq H$ is a closed subspace by Lemma 1.9. Since A is compact on $H \supseteq H_1$, $A|_{H_1}$ is compact on H_1. Therefore by Lemma 1.9 $H_2 := A_\lambda(H_1)$ is a closed subspace of H_1. Trivially, $H_2 \subseteq H_1$, since if $x_2 \in H_2$, then $x_2 = A_\lambda^2 x = A_\lambda(A_\lambda x) = A_\lambda x_1 \in H_1$. By induction on the powers of A_λ, we have that $H_n = A_\lambda^n(H)$ is a sequence of non-strictly decreasing closed subspaces. Our aim is to show somehow that $H = H_1$.

Suppose *per absurdum* that they are all strictly decreasing closed subspaces. We can therefore apply to H_n the Riesz's Lemma and claim that there is x_n such that $\lambda x_n \in H_n$, $\|\lambda x_n\| = 1$ and $\operatorname{dist}(\lambda x_n, H_{n+1}) = 1$. Thus we have defined a bounded sequence $\{x_n\}$ with such property. Now, take $m < n$ and consider $Ax_n - Ax_m = [\lambda x_n + (Ax_n - \lambda x_n) - (Ax_m - \lambda x_m)] - \lambda x_m = [\lambda x_n + A_\lambda x_n - A_\lambda x_m] - \lambda x_m$. Since $H_m \supsetneq H_{m+1} \supsetneq H_n \supsetneq H_{n+1}$, we have that $[\lambda x_n + A_\lambda x_n - A_\lambda x_m] \in H_{m+1}$. This leads to an *absurdum*: indeed, $\|Ax_n - Ax_m\| = \|[\lambda x_n + A_\lambda x_n - A_\lambda x_m] - \lambda x_m\| \geqslant 1$ for $[\lambda x_n + A_\lambda x_n - A_\lambda x_m] \in H_{m+1}$. Hence for every $m < n$, $\|Ax_n - Ax_m\| \geqslant 1$, which means that there is no convergent subsequence of $\{Ax_n\}$. This is a contraddiction, since $\{x_n\}$ was bounded ($\|x_n\| = \frac{1}{\lambda}$). Thus not all the inclusions are proper, and for some n, which we assume to be the first, $H_n = H_{n+1}$.

Let us prove that $n = 0$ and the thesis shall follow. If *per absurdum* $n > 0$, $H_n = H_{n+1}$ for the first time, then an arbitrary point $x \in H_{n-1}$ is such that $x = A_\lambda^{n-1} y$ for some $y \in H$. Thus $A_\lambda x = A_\lambda^n y \in H_n = H_{n+1}$. Therefore $x \in H_n$, because there is a $w \in H$ such that $A_\lambda x = A_\lambda^{n+1} w = A_\lambda A_\lambda^n w$. This yields, by injectivity, that $x = A_\lambda^n w \in H_n$. Therefore $H_{n-1} \subseteq H_n$. But since trivially also $H_{n-1} \supseteq H_n$ holds, $H_{n-1} = H_n$, contrary to the hypothesis that n was the first. Therefore $n = 0$. □

Lemma 1.11. Let $A\colon H \longrightarrow H$ be a compact linear operator and λ a nonzero real number. Then $\lambda \in \sigma(A)$ if and only if there is a sequence in H, $\{x_n\}$ such that $\|x_n\| = 1$ for every n and $A_\lambda x_n \underset{n\to\infty}{\longrightarrow} 0$.

Proof. Assume that there is a sequence $\{x_n\}$ such that $\|x_n\| \equiv 1$ for every n and $A_\lambda\, x_n \longrightarrow 0$, as $n \to \infty$. Suppose *per absurdum* that $\lambda \notin \sigma(A)$, then A_λ^{-1} exists. By Lemma 1.3, A is continuous, then $A_\lambda = A - \lambda I$ is continuous, hence by Theorem 1.5 A_λ^{-1} is continuous as well. However, the continuity of A_λ^{-1} implies that $x_n = A_\lambda^{-1} A_\lambda x_n \longrightarrow 0$, as $n \to \infty$, but this contradicts $\|x_n\| \equiv 1$.

Conversely if $\lambda \in \sigma(A)$ then A_λ^{-1} does not exist.

If the reason is that A_λ is not injective, then there is $0 \neq x \in \mathcal{N}(A_\lambda) \subset H$, i.e. $A_\lambda x = 0$. Then we can trivially define $x_n \equiv \frac{x}{\|x\|}$, that is a sequence $\{x_n\}$ such that $\|x_n\| \equiv 1$ and $\{A_\lambda x_n\}$ vanishes identically.

If the reason is that A_λ is not surjective, then by Theorem 1.10, A_λ is not injective, and we can proceed again as above and define the sequence $\{x_n\}$. □

Proposition 1.12. If $A\colon H \longrightarrow H$ is a compact linear operator and $0 \neq \lambda \in \sigma(A)$, then λ is an eigenvalue of A.

Proof. If $\lambda \in \sigma(A)$ then A_λ^{-1} does not exist.

If the reason is the non-injectivity of A_λ, then there is $0 \neq x \in \mathcal{N}(A_\lambda) \subset H$, i. e. there is $0 \neq x \in H$ such that $0 = A_\lambda x = Ax - \lambda x$. By definition, x is an eigenvector of A corresponding to λ.

If the reason is the non-surjectivity of A_λ, then by Theorem 1.10 A_λ is not injective, and we can proceed again as above to find an eigenvector corresponding to λ. □

1.2 Eigenvalues of linear compact symmetric operators

Definition 1.13. The *bilinear form* of a linear symmetric (or self-adjoint) operator $A\colon H \longrightarrow H$ is a function $B\colon H \times H \longrightarrow \mathbb{R}$, $B(x,y) = (Ax,y) = (x,Ay)$. Analogously the *quadratic form* of a linear symmetric operator A is a function $Q\colon H \longrightarrow \mathbb{R}$, $Q(x) = (Ax, x) = (x, Ax)$. The norm of the quadratic form is defined as any operatorial norm: $\|Q\|_{\text{op}} := \sup\limits_{\|x\|=1} |Q(x)|$.

Lemma 1.14. For every $x, y \in H$, $B(x,y) = \frac{1}{4}[Q(x+y) - Q(x-y)]$.

Proof. $Q(x+y) - Q(x-y) = (x+y, A(x+y)) - (x-y, A(x-y)) = (x, Ax) + (y, Ax) + (x, Ay) + (y, Ay) - (x, Ax) + (x, Ay) + (y, Ax) - (y, Ay) = 4(x, Ay)$ □

Theorem 1.15. Assume Q be the quadratic form of the symmetric linear operator $A\colon H \longrightarrow H$. Then $\|Q\|_{\text{op}} = \|A\|_{\text{op}}$.

Proof. We prove first that $\|Q\|_{\text{op}} \leqslant \|A\|_{\text{op}}$. For every $x \in H$

$|Q(x)| = |(Ax,x)| = \|x\|^2 \left|\left(A\frac{x}{\|x\|}, \frac{x}{\|x\|}\right)\right| = \|x\|^2 \left|Q\left(\frac{x}{\|x\|}\right)\right| \leqslant \|x\|^2 \|Q\|_{\text{op}} \left\|\frac{x}{\|x\|}\right\| = \|x\|^2 \|Q\|_{\text{op}}$.

Then by taking an x such that $\|x\| = 1$ it follows, from the Cauchy-Schwarz inequality, that

$$|Q(x)| = |(Ax,x)| \leqslant \|Ax\| \, \|x\| \leqslant \|A\|_{\text{op}} \|x\|^2 = \|A\|_{\text{op}}.$$

Conversely we prove that $\|Q\|_{\text{op}} \geqslant \|A\|_{\text{op}}$.

In the previous step we have obtained the following inequality: $|Q(x)| \leqslant \|x\|^2 \|Q\|_{\text{op}}$. Thus by the Lemma 1.14 and the *Hilbert's parallelogram law* for every $x, y \in H$

$|B(x,y)| = \frac{1}{4}|Q(x+y) - Q(x-y)| \leqslant \frac{1}{4}\left|Q(x+y)\right| + \frac{1}{4}\left|Q(x-y)\right| \leqslant \frac{1}{4}\|Q\|_{\text{op}} \|x+y\|^2 + \frac{1}{4}\|Q\|_{\text{op}} \|x-y\|^2 = \frac{1}{4}\|Q\|_{\text{op}} (\|x+y\|^2 + \|x-y\|^2) = \frac{1}{4}\|Q\|_{\text{op}} 2(\|x\|^2 + \|y\|^2) = \frac{1}{2}\|Q\|_{\text{op}} (\|x\|^2 + \|y\|^2)$.

Then taking x, y such that $\|x\| = \|y\| = 1$ it follows that $|B(x,y)| \leqslant \|Q\|_{\text{op}}$.

However, it also holds that

$$|B(x,\,y)| = |(x,\,Ay)| = \|x\|\,\|y\|\left|\left(\frac{x}{\|x\|},\,A\,\frac{y}{\|y\|}\right)\right| = \|x\|\,\|y\|\left|B\left(\frac{x}{\|x\|},\,\frac{y}{\|y\|}\right)\right| \leqslant \|x\|\,\|y\|\|Q\|_{\mathrm{op}}.$$

Thus by taking y such that $\|y\|=1$ and $x=Ay$ we obtain that

$$\|Ay\|\,\|y\|\|Q\|_{\mathrm{op}} = \|x\|\,\|y\|\|Q\|_{\mathrm{op}} \geqslant |B(x,y)| = |B(Ay,y)| = |(Ay,Ay)| = \|Ay\|^2$$

that is $\|Ay\| \leqslant \|y\|\|Q\|_{\mathrm{op}} = 1\|Q\|_{\mathrm{op}} = \|Q\|_{\mathrm{op}}$ for every y such that $\|y\|=1$. Then $\|A\|_{\mathrm{op}} := \sup_{\|y\|=1}\|Ay\| \leqslant \|Q\|_{\mathrm{op}}$. □

Theorem 1.16. A compact symmetric linear operator $A\colon H \longrightarrow H$ has at least one eigenvalue.

Proof. We first claim that either $\|A\|_{\mathrm{op}}$ or $-\|A\|_{\mathrm{op}}$ lies in $\sigma(A)$. Then by the Proposition 1.12 the result shall follow.

To prove that either $\|A\|_{\mathrm{op}}$ or $-\|A\|_{\mathrm{op}}$ lies in $\sigma(A)$, we will primarily rely on the Lemma 1.11. First of all, we remark that by the definition of $\|Q\|_{\mathrm{op}} := \sup_{\|x\|=1}|Q(x)|$, we can define on the unitary ball of H a maximizing sequence $\{x_n\}$, that is a sequence such that for every n, $\|x_n\|=1$ and $|Q(x_n)| \longrightarrow \|Q\|_{\mathrm{op}}$ as $n\to\infty$. Now, by the Theorem 1.15 $\|Q\|_{\mathrm{op}} = \|A\|_{\mathrm{op}}$, thus the sequence is such that $|Q(x_n)| \longrightarrow \|A\|_{\mathrm{op}}$ as $n\to\infty$. Thus we have a subsequence $\{x_{n_j}\}$ such that either $Q(x_{n_j}) \longrightarrow \|A\|_{\mathrm{op}}$ or $Q(x_{n_j}) \longrightarrow -\|A\|_{\mathrm{op}}$ as $j\to\infty$.

If $Q(x_{n_j}) \longrightarrow \|A\|_{\mathrm{op}} := \lambda$, as $j\to\infty$, then

$$\|A_\lambda x_{n_j}\|^2 = (A_\lambda x_{n_j},\,A_\lambda x_{n_j}) = (Ax_{n_j} - \lambda x_{n_j},\,Ax_{n_j} - \lambda x_{n_j}) = \|Ax_{n_j}\|^2 - 2\lambda(Ax_{n_j},\,x_{n_j}) + \lambda^2\|x_{n_j}\|^2 \leqslant \lambda^2 - 2\lambda(Ax_{n_j},x_{n_j}) + \lambda^2 \underset{j\to\infty}{\longrightarrow} \lambda^2 - 2\lambda^2 + \lambda^2 = 0$$

since $(Ax_{n_j},Ax_{n_j}) = \|Ax_{n_j}\|^2 \leqslant \|A\|^2\|x_{n_j}\|^2 = \|A\|^2 = \lambda^2$, for $\|x_{n_j}\|=1$. Thus by the Lemma 1.11 $\|A\|_{\mathrm{op}} := \lambda \in \sigma(A)$.

If $Q(x_{n_j}) \underset{j\longrightarrow\infty}{\longrightarrow} -\|A\|_{\mathrm{op}} := -\lambda$, then analogously

$$\|A_{-\lambda}x_{n_j}\|^2 = (A_{-\lambda}x_{n_j},A_{-\lambda}x_{n_j}) = (Ax_{n_j} + \lambda x_{n_j},Ax_{n_j} + \lambda x_{n_j}) = \|Ax_{n_j}\|^2 + 2\lambda(Ax_{n_j},\,x_{n_j}) + \lambda^2\|x_{n_j}\|^2 \leqslant \lambda^2 + 2\lambda(Ax_{n_j},x_{n_j}) + \lambda^2 \underset{j\longrightarrow\infty}{\longrightarrow} \lambda^2 - 2\lambda^2 + \lambda^2 = 0.$$

Thus by the Lemma 1.11 $-\|A\|_{\mathrm{op}} := -\lambda \in \sigma(A)$. □

Theorem 1.17. Let $\lambda \neq 0$ be a fixed eigenvalue of a compact linear operator A: $H \longrightarrow H$. Let $E_\lambda := \{x \in H\colon Ax = \lambda x\}$be the subspace of eigenvectors corresponding to λ. Then E_λ, also called the *eigenspace* of λ, is finite dimensional.

Proof. First step: take a bounded sequence in E_λ, $\{x_n\}$, then $Ax_n = \lambda x_n$, that is $x_n = \lambda^{-1}Ax_n$. Since by the compactness of the operator, $\{Ax_n\}$ has a convergent subsequence, so does $\{x_n\}$. Therefore every bounded sequence in E_λ has a convergent subsequence.

Second step: *per absurdum*. If E_λ were infinite dimensional, by the Gram-Schmidt process we could construct a sequence $\{e_n\}$ of mutually orthonormal vectors in E_λ. But for

$$\|e_n - e_m\|^2 = \|e_n\|^2 + \|e_m\|^2 - 2(e_n, e_m) = \|e_n\|^2 + \|e_m\|^2 = 2 \text{ for every } n \neq m$$

it follows that the distance between e_n and e_m is $\sqrt{2}$ for every $n \neq m$. Therefore there can be no convergent subsequence, against the result of the first step. Then E_λ must be finite dimensional. □

Remark 1.18. Since the *eigenspace* of λ, $E_\lambda = \mathcal{N}(A_\lambda)$, the previous result could have been drawn straightforward by Lemma 1.8. However the prove above has been given too, primarily because of its geometric significance.

Lemma 1.19. Let $\lambda, \mu \neq 0$ be distinct eigenvalues of the compact symmetric linear operator $A: H \longrightarrow H$. Then the corresponding eigenvectors are orthogonal.

Proof. Since A is symmetric, by definition $(Ax, y) = (x, Ay)$. If x is the eigenvector corresponding to λ and y is the eigenvector corresponding to μ, then $(\lambda x, y) = (x, \mu y)$ which gives us $\lambda(x, y) = \mu(x, y)$ that is $(\lambda - \mu)(x, y) = 0$. Since $\lambda - \mu \neq 0$, $(x, y) = 0$ results. □

Theorem 1.20. Let $A: H \longrightarrow H$ be a compact symmetric linear operator. Then the set of eigenvalues of A is either finite or countably infinite. In addition, the only possible limit point of a sequence of eigenvalues is 0.

Proof. Assume that a convergent sequence of eigenvalues $\{\lambda_n\}$ is given, $\lambda_n \longrightarrow \lambda$, as $n \to \infty$. Consider for every λ_n the corresponding eigenvector x_n. Without loss of generality, we can and do assume that $\lambda_n \neq \lambda_m$ for every $n \neq m$, and take the corresponding eigenvectors x_n, $\|x_n\| = 1$. Indeed, every subsequence we could possibly extract is convergent, therefore we can extract a convergent subsequence such that $\lambda_n \neq \lambda_m$ for every $n \neq m$, (consequently the corresponding eigenvectors are already mutually orthogonal by the symmetry of A, according to Lemma 1.19).

We now suppose *per absurdum* that $\lambda \neq 0$. Then

$$\|Ax_n - Ax_m\|^2 = \|\lambda_n x_n - \lambda_m x_m\|^2 = \lambda_n^2 ||x_n||^2 + \lambda_m^2 ||x_m||^2 = \lambda_n^2 + \lambda_m^2 \underset{n,m \to \infty}{\longrightarrow} 2\lambda^2 > 0.$$

Therefore $\{Ax_n\}$ has no convergent subsequence, which means that A is not compact, that is an *absurdum*. Thus $\lambda = 0$ and the second part of the thesis is proved.

Let now be $\Lambda = \{\lambda \in \mathbb{R}: \exists x \neq 0, Ax = \lambda x\}$the set of all the eigenvalues of the symmetric operator A. Since $\Lambda \subset \mathbb{R}$ and $\mathbb{R} = \bigcup_{n=1}^\infty \left[-n, -\frac{1}{n}\right] \cup \{0\} \cup \bigcup_{n=1}^\infty \left[\frac{1}{n}, n\right]$, we can decopose by distributivity as follows:

$$\Lambda = (\Lambda \cap \{0\}) \cup \left(\bigcup_{n=1}^\infty \left(\left[\frac{1}{n}, n\right] \cap \Lambda\right)\right) \cup \left(\bigcup_{n=1}^\infty \left(\left[-n, -\frac{1}{n}\right] \cap \Lambda\right)\right).$$

Suppose *per absurdum* that Λ is uncountably infinite. Since both the unions are on countably infinite indices, then there exists n such that $\left[\frac{1}{n}, n\right] \cap \Lambda$ or $\left[-n, -\frac{1}{n}\right] \cap \Lambda$ is uncountably infinite. Let us suppose, without loss of generality, $\left[\frac{1}{n}, n\right] \cap \Lambda$ be uncountably infinite. Being $\left[\frac{1}{n}, n\right] \subset \mathbb{R}$ a bounded infinite set, it has an accumulation point, which belongs to this closed interval, and therefore this point is non zero. As a result $\left[\frac{1}{n}, n\right] \cap \Lambda$ has a non zero limit point, and this is a contradiction because we have already proved that every sequence of eigenvalues of A has a zero limit. All in all, Λ is a countable set. □

Lemma 1.21. Let $A: H \longrightarrow H$ be a linear operator and λ one of its eigenvalues. Then $|\lambda| \leqslant \|A\|_{\text{op}}$.

Proof. If $x \neq 0, Ax = \lambda x$, then $|\lambda| = \frac{\|Ax\|}{\|x\|} \leqslant \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} =: \|A\|_{\text{op}}$. □

Remark 1.22. By the Theorem 1.16 if $A: H \longrightarrow H$ is also compact and symmetric, actually $\|A\|_{\text{op}}$ or $-\|A\|_{\text{op}}$ is an eigenvalue.

1.3 The Expansion Theorem

Definition 1.23. If $S \subset H$, we call the smallest closed subspace of H containing S the space *spanned* by the subset S and we denote it as $\text{span}(S)$. Equivalently

$$\text{span}(S) := \left\{ \sum_{k=1}^{n} \lambda_k x_k, n \in \mathbb{N}, x_k \in S, \lambda_k \in \mathbb{R} \right\}.$$

Theorem 1.24. Let C be a closed subspace of the Hilbert space H. Then $H = C \oplus C^{\perp}$.

Proof. We have to show that for every $x \in H$, $x = y + z$, $y \in C, z \in C^{\perp}$. Take $x \in H$, by the *Riesz's Lemma* the orthogonal projection z of x on C is the unique nearest element of C. We define $y: = x - z$. If we prove that $y \in C^{\perp}$ the result shall follow. Call $d := \|x - z\|$, then if $t \in \mathbb{R}$ and $0 \neq w \in C$, $d \leq \|x - (z + tw)\|$, then

$$d^2 \leqslant \|x - (z + tw)\|^2 = \|y - tw\|^2 = d^2 + t^2 \|w\|^2 - 2t\,(y, w)$$

therefore $t^2 \|w\|^2 - 2t(y, w) \geqslant 0$ for every t. Taking $t = \frac{(y, w)}{\|w\|^2}$ we obtain $-\frac{(y, w)^2}{\|w\|^2} \geqslant 0$. Therefore $(y, w) = 0$ for every $0 \neq w \in C$, that is $y \in C^{\perp}$. □

Theorem 1.25. Let $A: H \longrightarrow H$ be a compact symmetric linear operator. The eigenvectors of A span the entire space H.

Proof. Let be E spanned by the eigenvectors of A, $E = \text{span}(\{0 \neq e \in H: \exists \lambda \in \mathbb{R}, Ae = \lambda e\})$. The thesis is that $E = H$. We shall proceed *per absurdum.* Assume $E \subsetneq H$. Since E is a closed subspace, by Theorem 1.24 $H = E \oplus E^{\perp}$, and for $E \subsetneq H$, it holds that $E^{\perp} \neq \{0\}$, since it contains at least the difference between an element of $H \backslash E$ and its projection onto E.

$E^{\perp}$ is a closed subspace too, since if a sequence $\{x_n\} \subset E^{\perp}$ converges to a point x, then $x \in E^{\perp}$ since, by the continuity of the inner product,

$$(x, y) = \left(\lim_{n \to \infty} x_n, y\right) = \lim_{n \to \infty} (x_n, y) = \lim_{n \to \infty} 0 = 0.$$

Furthermore $E^{\perp}$ is $A-$invariant, that is if $x \in E^{\perp}$ then $Ax \in E^{\perp}$. Wa shall prove it. Firstly, if $x \in E^{\perp}$ then Ax is orthogonal to every eigenvector of A. Indeed, given an eigenvector y corresponding to the eigenvalue λ,

$$(Ax, y) = (x, Ay) = (x, \lambda y) = \lambda(x, y) = 0$$

Secondly, all the eigenvectors of A span E, which means by definition of spanned subspace, that for every $y \in E$ there is a sequence of finite linear combinations of eigenvalues that converges to y. E is the closed subspace of all those finite linear combinations indeed. Let us call $\{y_n\}$ this convergent sequence of finite linear combinations of eigenvalues of A.

As a result, if $x \in E^{\perp}$ and $y \in E$, $y = \lim_{n \to \infty} y_n$ as above stated, assuming that e_k^n are the eigenvectors involved in the finite linear combination with coefficients μ_k^n that yield y_n, by the fact that Ax is orthogonal to every eigenvector of A, it holds that

$$(Ax, y) = \left(Ax, \lim_{n \to \infty} y_n\right) = \lim_{n \to \infty} (Ax, y_n) = \lim_{n \to \infty} \left(Ax, \sum_{0<k<\infty} \mu_k^n e_k^n\right) = \sum_{0<k<\infty} \mu_k^n \lim_{n \to \infty} (Ax, e_k^n) = \sum_{0<k<\infty} \mu_k^n \lim_{n \to \infty} 0 = 0,$$

therefore $Ax \in E^{\perp}$.

Now, since $E^{\perp}$ is a closed subspace of an Hilbert space H, $E^{\perp}$ is an Hilbert space itself. We are allowed to restrict A and, since $E^{\perp}$ is $A-$invariant, thi restricion is still a compact symmetric linear operator $A|_{E^{\perp}}\colon E^{\perp} \longrightarrow E^{\perp}$. By Theorem 1.16 this restriction has at least one eigenvalue. Therefore there is at least a (non zero) eigenvector $x \in E^{\perp}$. But by definition, $x \in E$ too. Thus $(x, x) = 0$. Which implies that $x = 0$, which is absurd since it is an eigenvector. Therefore $E = H$. □

Remark 1.26. We now focus on the space spanned by the eigenvectors corresponding to the non zero eigenvalues of A. We can always find an orthonormal basis $\{e_n\}_{n \in \mathbb{N}}$ for this space. We can construct it by forming, for each non zero eigenvalue λ of A, an orthonormal basis for the finite-dimensional (Theorem 1.17) eigenspace E_λ via a Gram-Schmidt process, and eventually taking the union of all such basis, which are already mutually orthogonal (Lemma 1.19).

Theorem 1.27. (Expansion Theorem) Let $A\colon H \longrightarrow H$ be a compact symmetric linear operator. Let $\{e_n\}_{n \in \mathbb{N}}$ be an orthonormal basis for the space spanned by the eigenvectors corresponding to the non zero eigenvalues of A. Then if $x \in H$ and h is the projection of x on $\mathcal{N}(A)$, the null space of A, the following representation of x holds

$$x = h + \sum_{n=1}^{\infty} (x, e_n) e_n.$$

Proof. First, we need to show that $\sum_{n=1}^{\infty}(x, e_n)e_n$ is convergent in H. Since the $\{e_n\}_{n\in\mathbb{N}}$ are an orthonormal sequence in H, by the *Bessel's inequality* we have that $\sum_{i=1}^{\infty}|(x, e_i)|^2 \leqslant \|x\|^2 < \infty$. Therefore $\sum_{i=1}^{\infty}|(x, e_i)|^2$ is a real positive valued convergent series. Then its partial sums form a convergent, and therefore Cauchy, sequence, that is, given $m > n$,

$$\sum_{i=1}^{m}|(x, e_i)|^2 - \sum_{i=1}^{n}|(x, e_i)|^2 = \sum_{i=n}^{m}|(x, e_i)|^2 \underset{n,m\to\infty}{\longrightarrow} 0.$$

By the orthonormality of the system $\{e_n\}$,

$$\left\|\sum_{i=n}^{m}(x, e_i)e_i\right\|^2 = \sum_{i=n}^{m}|(x, e_i)|^2 \underset{n,m\to\infty}{\longrightarrow} 0.$$

Therefore the sequence $S_n: =\sum_{i=1}^{n}(x, e_i)e_i$ is a Cauchy sequence in H, since we have just proved that $\|S_m - S_n\| \longrightarrow 0$, as $n, m \to \infty$. Being H complete, $\{S_n\}$ is a convergent sequence, that is $\sum_{i=1}^{\infty}(x, e_i)e_i$ is a convergent series.

We are now allowed to write

$$x = \left[h + \sum_{n=1}^{\infty}(x, e_n)e_n\right] + \left[x - h - \sum_{n=1}^{\infty}(x, e_n)e_n\right].$$

Define $y := x - h - \sum_{n=1}^{\infty}(x, e_n)e_n$. We claim that y is orthogonal to every e_n. Indeed

$$(y, e_n) = (x, e_n) - (h, e_n) - \left(\sum_{i=1}^{\infty}(x, e_i)e_i, e_n\right) = (x, e_n) - (x, e_n) = 0$$

since firstly by Lemma 1.19 $(h, e_n) = 0$, in that $h \in N(A)$ and for every n, $e_n \notin \mathcal{N}(A)$ by construction; secondly by continuity of the inner product

$$\left(\sum_{i=1}^{\infty}(x, e_i)e_i, e_n\right) = \sum_{i=1}^{\infty}(x, e_i)(e_i, e_n) = \sum_{i=1}^{\infty}(x, e_i)\delta_{in} = (x, e_n).$$

Moreover y is orthogonal to $\mathcal{N}(A)$. Indeed if $z \in \mathcal{N}(A)$ then

$$(y, z) = (x, z) - (h, z) - \left(\sum_{i=1}^{\infty}(x, e_i)e_i, z\right) = (x, z) - (h, z) = (x - h, z) = 0$$

since by the continuity of the inner product and Lemma 1.19, given a basis of $\mathcal{N}(A)$, $\{e_k^0\}_{k=1}^{m}$ if $m = \dim\mathcal{N}(A)$,

$$\left(\sum_{i=1}^{\infty}(x, e_i)e_i, z\right) = \sum_{i=1}^{\infty}(x, e_i)(e_i, z) = \sum_{i=1}^{\infty}(x, e_i)\left(e_i, \sum_{k=1}^{m}(z, e_k^0)e_k^0\right) = \sum_{i=1}^{\infty}(x, e_i)\sum_{k=1}^{m}(z, e_k^0)(e_i, e_k^0) = 0.$$

Furthermore $x - h$ is orthogonal to z for h is the projection of x on $\mathcal{N}(A)$ and $z \in \mathcal{N}(A)$.

As a result, we have shown that y is orthogonal to all eigenvectors of A, which span the whole H by Theorem 1.25. Therefore by linearity of the inner product, y is orthogonal to all elements of H, in particular to y itself, which means that $y = 0$. Since $x = [h + \sum_{n=1}^{\infty} (x, e_n) e_n] + y$, the result follows. □

Corollary 1.28. Given the same hypothesis of the Expansion Theorem, if in addition we call λ_n the non zero eigenvalue corresponding to the eigenvector e_n (these λ_n are not all distinct consequently), then by the continuity of A and its linearity we have that

$$Ax = \sum_{n=1}^{\infty} (x, e_n) \lambda_n e_n$$

since $Ah = 0$.

2 Integral operators

In this Section Ash (1990) is strictly followed. We shall firstly derive the properties of the *Hilbert-Schmidt integral operator* A: $L^2[a,b] \longrightarrow L^2[a,b]$, where $a, b < \infty$, defined as

$$(Af)(t) = \int_a^b R(t,s) f(s) ds,$$

with $t \in [a,b]$ and R: $[a,b] \times [a,b] \longrightarrow \mathbb{R}$ assumed continuous in both variables jointly.

2.1 Properties of the Hilbert-Schmidt operator

Remark 2.1. The Hilbert-Schmidt operator A is linear, bounded and such that for every $f \in L^2[a,b]$, $(Af)(t)$ is continuous on $[a,b]$.

Proof. Linearity of A: let $f, g \in L^2[a,b]$ and $\alpha, \beta \in \mathbb{R}$, then

$$(A(\alpha f + \beta g))(t) = \int_a^b R(t,s)(\alpha f + \beta g)(s) ds = \alpha \int_a^b R(t,s) f(s) ds +$$

$$+ \beta \int_a^b R(t,s) g(s) ds = \alpha (Af)(t) + \beta (Ag)(t).$$

Boundness of A: let $f \in L^2[a, b]$ and $M := \max_{t,s \in [a,b]} |R(t, s)|$, which exixts and is finite by the Weierstrass theorem, because R is a continuous function on the compact set $[a, b] \times [a, b]$(therefore $R(t, \cdot) \in L^2[a, b]$ in addition). By the Cauchy-Schwarz inequality

$$|Af|(t) \leqslant \int_a^b |R(t,s) f(s)|\, ds = \|R(t,\cdot) f\|_1 \leqslant \|R(t,\cdot)\|_2 \|f\|_2 =$$

$$= \|f\|_2 \left(\int_a^b R^2(t,s) \mathrm{ds} \right)^{\frac{1}{2}} \leqslant \|f\|_2 M \sqrt{b-a}.$$

Hence

$$\|A\|_{\text{op}} = \sup_{0 \not\equiv f \in L_2[a,b]} \frac{\|Af\|_2}{\|f\|_2} = \sup_{0 \not\equiv f \in L_2[a,b]} \frac{\left[\int_a^b |Af|^2(t)\, dt\right]^{\frac{1}{2}}}{\|f\|_2} \leqslant$$

$$\sup_{0 \not\equiv f \in L_2[a,b]} \frac{\left[\int_a^b M^2 (b-a) \|f\|_2^2 dt\right]^{\frac{1}{2}}}{\|f\|_2} \leqslant \sup_{0 \not\equiv f \in L_2[a,b]} \frac{M(b-a)\|f\|_2}{\|f\|_2} = M(b-a) < \infty.$$

Continuity: we shall prove equivalently that $\lim_{t \to t_0} (Af)(t) = (Af)(t_0)$:

$$\lim_{t \to t_0} |(Af)(t) - (Af)(t_0)| = \lim_{t \to t_0} \left| \int_a^b R(t,s) f(s) ds - \int_a^b R(t_0,s) f(s) ds \right| = \lim_{t \to t_0} \left| \int_a^b [R(t,\right.$$

$$\left. s) - R(t_0,s)] f(s) ds \right| = \left| \lim_{t \to t_0} \int_a^b [R(t,s) - R(t_0,s)] f(s) ds \right|.$$

If the *dominated convergence* conditions are satisfied, then

$$\lim_{t \to t_0} \int_a^b [R(t,s) - R(t_0,s)] f(s) ds = \int_a^b f(s) \lim_{t \to t_0} [R(t,s) - R(t_0,s)] ds = 0$$

and the result shall follow. Now, $R(t,s) - R(t_0,s) \leqslant |R(t,s)| + |R(t_0,s)| \leqslant 2M$; in addition, given $f \in L^2[a,b]$, $|f| \in L^2[a,b] \subset L^1[a,b]$. Thus for every $t \in [a,b]$, $|[R(t,s) - R(t_0,s)] f(s)| \leqslant 2M\,|f| \in L^1[a,b]$, the dominated convergence hypothesis are satisfied and the limit can be taken inside the integral sign as claimed. □

Corollary 2.2. The eigenfunctions of the Hilbert-Schmidt operator A corresponding to non zero eigenvalues are continuous on $[a,b]$.

Proof. Let e be an eigenfunction of A corresponding to the eigenvalue $\lambda \neq 0$. Then by definition $\lambda e(t) = (Ae)(t)$, then $e(t) = \frac{1}{\lambda}(Ae)(t)$. Hence by the continuity of A over $[a,b]$ the result follows. □

Proposition 2.3. The Hilbert-Schmidt operator A is compact.

Proof. By the *Arzelà-Ascoli Theorem*, if $\{y_k\}_{k \in \mathbb{N}}$ is a sequence of continuous functions on $[a,b]$ and they are uniformly bounded (i.e. there exists a finite constant L such that $|y_k(s)| < L$ for every $k \in \mathbb{N}$, $s \in [a,b]$) and equicontinuous (i.e. for every positive ε, there exists a positive δ_ε, such that for every s_1, $s_2 \in [a,b]$ if $|s_{1-s_2}| < \delta_\varepsilon$,then$|y_k(s_1) - y_k(s_2)| < \varepsilon$ for every $k \in \mathbb{N}$) then $\{y_k\}$ has a uniformly convergent subsequence. The Arzelà-Ascoli Theorem helps us to prove that A is compact (i.e. that for every bounded sequence $\{f_k\} \subset L^2[a,b]$, $\{Af_k\} \subset L^2[a,b]$ has a convergent subsequence). Obviously the boundness and the convergence are meant in L^2-norm. Let now define $y_k(t) := (Af_k)(t) = \int_a^b R(t,s) f_k(s)\, ds$.

First we notice that $\{y_k\}_{k \in \mathbb{N}}$ is uniformly bounded: indeed, in the Remark 2.1 we have already shown that $|y_k(t)| \leqslant \|f_k\|_2 M \sqrt{b-a}$. But now by the boundness there exists a constant N not depedent on k such that for every $k \in \mathbb{N}$, $\|f_k\|_2 \leqslant N$. Thus $|y_k(t)| \leqslant \mathrm{NM}\sqrt{b-a} \leqslant \infty$ for every k, t.

Next we prove the equicontinuity of $\{y_k\}_{k \in \mathbb{N}}$: by the Cauchy-Schwarz inequality

$$|y_k(t_1) - y_k(t_2)| = \left| \int_a^b [R(t_1,s) - R(t_2,s)] f_k(s) ds \right| \leqslant \int_a^b |R(t_1,s) - R(t_2,s)| |f_k(s)| ds \leqslant \left(\int_a^b |R(t_1,s) - R(t_2,s)|^2 ds \right)^{\frac{1}{2}} \|f_k\|_2 \leqslant N \left(\int_a^b |R(t_1,s) - R(t_2,s)|^2 ds \right)^{\frac{1}{2}}.$$

Since R is continuous on the compact $[a,b] \times [a,b]$, it is uniformly continuous, which means that given $\varepsilon > 0$, there is a $\delta_\varepsilon > 0$ such that if $\|(t_1,s_1) - (t_2,s_2)\| < \delta_\varepsilon$, then we have $|R(t_1,s_1) - R(t_2,s_2)| < \varepsilon$. Since $\sqrt{|t_1 - t_2|^2 + |s_1 - s_2|^2} = \|(t_1,s_1) - (t_2,s_2)\|$ by the *Pythagorean Theorem*, we could rewrite: given $\varepsilon > 0$, there is a $\delta_\varepsilon > 0$ such that if $|t_1 - t_2| < \delta_\varepsilon$ then $|R(t_1,s) - R(t_2,s)| < \varepsilon$ for every $s \in [a,b]$. Thus if $|t_1 - t_2| < \delta_\varepsilon$ then $|y_k(t_1) - y_k(t_2)| \leqslant N \left(\int_a^b \varepsilon^2 ds \right)^{\frac{1}{2}} = N\sqrt{b-a}\,\varepsilon$. Then it is sufficient to take a δ_ε for the estimate of $|y_k(t_1) - y_k(t_2)|$ defined as a $\delta_{\varepsilon'}$, that estimates $|R(t_1,s_1) - R(t_2,s_2)|$ as above, where $\varepsilon' := \frac{\varepsilon}{N\sqrt{b-a}}$. Then if $|t_1 - t_2| < \delta_\varepsilon$ then $|y_k(t_1) - y_k(t_2)| \leqslant N\sqrt{b-a}\,\varepsilon' = \varepsilon$. For δ_ε does not depend on k, the argument proves the equicontinuity.

Eventually we draw the conclusion by the Ascoli-Arzelà Theorem: by unirfom boundness and equicontinuity of $\{y_k\}_{k\in\mathbb{N}}$ there is a subsequence $\{y_{k_j}\}_{j\in\mathbb{N}}$ uniformly convergent to a function $y\in C[a,b]\subset L_2[a,b]$. Namely, for every $\varepsilon>0$ there is an $N\in\mathbb{N}$ such that if $j>N$ then $|y_{k_j}(s)-y(s)|<\varepsilon$ for every $s\in[a,b]$. The convergence of $\{y_{k_j}\}$ in $L_2[a,b]-$norm follows straightforward by its uniform convergence and the finite measure of the interval: for every $\varepsilon>0$ there is an $N\in\mathbb{N}$ such that if $j>N$ then

$$\|y_{k_j}-y\|_2=\left(\int_a^b |y_{k_j}(s)-y(s)|^2 ds\right)^{\frac{1}{2}}\leqslant\left(\int_a^b \varepsilon^2 ds\right)^{\frac{1}{2}}=\varepsilon\sqrt{b-a}$$

it is then suffcient to consider an appropriate $\varepsilon' := \frac{\varepsilon}{\sqrt{b-a}}$ corresponding to N. □

2.2 The Mercer's Theorem

Definition 2.4. A function $R:[a,b]\times[a,b]\to\mathbb{R}$ is said *symmetric* if $R(t,s)=R(s,t)$ for all $t,s\in[a,b]$.

Remark 2.5. If $R(t,s)$ is continuous and symmetric, then the operator $(Af)(t)=\int_a^b R(t,s)f(s)ds$ is symmetric.

Proof. By definition A is symmetric on L^2if $(Af,g)_2=(f,Ag)_2$. By the symmetry of R, the result follows basically from the *Fubini-Tonelli Theorem*, which can be used, since the functions f,g are in $L^2[a,b]$ and R is bounded by its continuity on a compact set, consequently the product $R(t,s)f(s)g(t)$ is in $L^1[a,b]$:

$$(Af,g)_2=\int_a^b\int_a^b R(t,s)f(s)ds\,g(t)\,dt=\int_a^b\int_a^b R(t,s)f(s)g(t)ds\,dt=$$

$$=\int_a^b\int_a^b R(t,s)f(s)g(t)dt\,ds=\int_a^b f(s)\int_a^b R(s,t)g(t)dt\,ds=(f,Ag)_2. \qquad \square$$

Proposition 2.6. Let R: $[a,\ b]\times[a,\ b]\to\mathbb{R}$ be symmetric and continuous. All eigenvalues of $(Af)(t)=\int_a^b R(t,\ s)f(s)ds$ are non negative if and only if for all functions f continuous on $[a,b]$, $\int_a^b\int_a^b R(t,s)f(t)f(s)dt\,ds\geqslant 0$.

Proof. If $\int_a^b\int_a^b R(t,s)f(t)f(s)dt\,ds\geqslant 0$ for all functions f continuous on $[a,b]$, then $\int_a^b\int_a^b R(t,s)e(t)e(s)dt\,ds\geqslant 0$, where e is the eigenfunction of A corresponding to the non zero eigenvalue λ, in that by the Corollary 2.2 e is continuous on $[a,b]$.

$$0\leqslant\int_a^b\int_a^b R(t,\ s)e(t)e(s)dt\,ds=\int_a^b e(s)\int_a^b R(t,\ s)e(t)dt\,ds=\int_a^b e(s)\lambda e(s)ds=$$

$$\lambda\int_a^b e^2(s)ds=\lambda\|e\|_2^2.$$

Since an eigenfunction corresponding to a non zero eigenvalue is non zero, $e\not\equiv 0$, and it follows from above that $\lambda\geqslant 0$.

Conversely, suppose that every eigenvalue of A is non negative. Then if $f \in C[a, b] \subset L^2[a,b]$ we have, by the Expansion Theorem in a Hilbert space of the previous section, that $f = h + \sum_{n=1}^{\infty} c_n e_n$, where the $\{e_n\}$ are an orthonormal basis for the space spanned by the eigenvectors corresponding to the nonzero eigenvalues of the symmetric compact linear operator A, arranged such that e_n is eigenvector of the eigenvalue λ_n; $c_n = (f, e_n)_2$; $h \in \mathcal{N}(A)$. Now, the double integral can be rewritten as

$$\int_a^b \int_a^b R(t,s) f(t) f(s) dt ds = \int_a^b f(s) \int_a^b R(t,s) f(t) dt ds = \int_a^b f(s)(Af)(s) ds =$$

$$=(f,\, Af)_2 = \left(h + \sum_{n=1}^{\infty} c_n e_n,\, Af\right)_2 = (h,\, Af)_2 + \sum_{n=1}^{\infty} c_n (e_n,\, Af)_2 = (Ah,\, f)_2 +$$

$$\sum_{n=1}^{\infty} c_n (Ae_n,\, f)_2 = 0 + \sum_{n=1}^{\infty} c_n (Ae_n,\, f)_2 = \sum_{n=1}^{\infty} c_n \lambda_n (\, e_n,\, f)_2 = \sum_{n=1}^{\infty} c_n^2 \lambda_n \geqslant 0.$$

□

Definition 2.7. R: $[a,\, b] \times [a,\, b] \to \mathbb{R}$ is said to be *non-negative definite* if $\sum_{i,j=1}^{n} f(t_i) R(t_i,\, t_j) f(t_j) \geqslant 0$ for all possible choices of $t_1,\, \ldots,\, t_n \in [a,\, b]$ and all possible real-valued functions f on $[a,b]$.

Proposition 2.8. Let R: $[a,b] \times [a,b] \to \mathbb{R}$ be continuous symmetric non-negative definite. Then all eigenvalues of the associated integral operator A are non-negative.

Proof. Let $f \in C[a,\, b]$. Then $\int_a^b \int_a^b R(t,\, s) f(t) f(s) dt ds$ is an ordinary Riemann integral, and the approximating sums are $S_n(\Delta) = \sum_{i,j=1}^{n} f(t_i) R(t_i,\, t_j) f(t_j)$, where Δ: $a = t_1 \leqslant \ldots \leqslant t_n = b$ is an $n-$ partition of $[a,\, b]$. Since R is non-negative definite, these approximating sums are non-egative for every $n-$partition Δ. Thus $0 \leqslant \lim_{n\to\infty} S_n(\Delta) = \int_a^b \int_a^b R(t,\, s) f(t) f(s) dt ds$. Then all eigenvalues of A are non-negative by the Proposition 2.6. □

Theorem 2.9. (Mercer's Theorem) Let R: $[a,\, b] \times [a,\, b] \to \mathbb{R}$ be continuous symmetric non-negative definite and let A be the corresponding Hilbert-Schmidt operator. Let $\{e_n\}$be an orthonormal basis for the space spanned by the eigenvectors corresponding to the non-zero eigenvalues of A. If the basis is taken so that e_n is the eigenvector corresponding to the eigenvalue λ_n, then $R(s,t) = \sum_{n=1}^{\infty} \lambda_n e_n(s) e_n(t)$ for every $s, t \in [a,b]$, where:

i. the series converges absolutely in both variables jointly.

ii. the series converges to $R(s,t)$ uniformly in both variables jointly.

iii. the series converges to $R(s,t)$ in $L^2([a,b] \times [a,b])$.

Proof. We shall first prove the following points:

1. the series converges to $R(s,t)$ in $L^2[a,b]$ in each variable separately.
2. $R_n(s,t) := R(s,t) - \sum_{i=1}^{n} \lambda_i e_i(s) e_i(t)$ is such that for every t, $R_n(t,t) \geqslant 0$.

3. the series converges absolutely in both variables jointly (i) and uniformly in both variables separately.
4. the series converges pointwise to $R(s,t)$.
5. (Dini's Theorem) Given a monotone (say increasing) sequence of functions $\{g_n\}, g_n \in C[a,b], g_n(x) \leqslant g_{n+1}(x)$ for every n and x, if $\lim_{n\to\infty} g_n(x) = g(x) \in C[a,b]$ pointwise, then the convergence is uniform.

The overall scheme of the proof is the following:

$$1 \Rightarrow 2 \overset{1}{\Rightarrow} 3 \overset{1}{\Rightarrow} 4 \Rightarrow g_n(s) = \sum_{i=1}^{n} \lambda_i e_i^2(s) \underset{n\to\infty}{\longrightarrow} R(s,s) \text{ pointwise} \overset{5}{\Rightarrow}$$

$$\overset{5}{\Rightarrow} g_n(s) \underset{n\to\infty}{\longrightarrow} R(s,s) \text{ uniformly} \overset{3}{\Rightarrow} ii \Rightarrow iii$$

Step 1: $R(s,\,t) = \sum_{n=1}^{\infty} \lambda_n e_n(s) e_n(t)$, where the convergence is in $L_2[a,\,b]$ in each variable separately.

For every fixed $s \in [a,b]$ by the Expansion Theorem for $R(s,t) \in C[a,b] \subset L^2[a,b]$ as a function of t, we can decompose $R(s,\,t) = h(s,\,t) + \sum_{n=1}^{\infty} c_n(s) e_n(t)$, where $h \in \mathcal{N}(A)$ (i.e. $(Ah)(s) = \int_a^b R(s,\,t) h(s,\,t)\,dt = 0$) and $c_n(s) = (R,\,e_n)_2 = \int_a^b R(s,\,t) e_n(t) dt = \lambda_n e_n(s)$, for the $\{e_n\}$ are an orthonormal basis for the space spanned by the eigenvectors corresponding to the non-zero eigenvalues of the symmetric compact linear operator A, and they can be chosen such that e_n is an eigenvector of the eigenvalue λ_n. By the Expansion Theorem $h(s,t) + \sum_{n=1}^{\infty} c_n(s) e_n(t)$ converges to $R(s,t)$ in $L^2[a,b]$.

Now we prove that $h(s,\cdot) \equiv 0$ in $L^2[a,b]$, since then from the decomposition above the result shall follow. By the continuity and the linearity of the inner product in a Hilbert space, we have that

$$0 = \int_a^b R(s,t) h(s,t) dt = (R,h)_2 = \left(h + \sum_{n=1}^{\infty} c_n e_n, h \right)_2 = (h,h)_2 +$$

$$+ \left(\sum_{n=1}^{\infty} c_n e_n, h \right)_2 = \|h\|_2^2 + \left(\lim_{m\longrightarrow\infty} \sum_{n=1}^{m} c_n e_n, h \right)_2 = \|h\|_2^2 + \lim_{m\longrightarrow\infty} \left(\sum_{n=1}^{m} c_n e_n, h \right)_2 =$$

$$= \|h\|_2^2 + \lim_{m\longrightarrow\infty} \sum_{n=1}^{m} (c_n e_n,\, h)_2 = \|h\|_2^2 + \sum_{n=1}^{\infty} \int_a^b c_n(s) h(s,\,t) e_n(t) dt = \|h\|_2^2 + \sum_{n=1}^{\infty} c_n(s) \int_a^b h(s,t) e_n(t) dt.$$

Now, by Lemma 1.19, the eigenvectors of a linear symmetric operator, corresponding to different eigenvalues, are orthogonal, thus $\int_a^b h(s,\,t) e_n(t) dt = 0$ for every n, for every e_n corresponds to a nonzero eigenvalue, whereas $h(s,\,\cdot)$ corresponds to 0, belonging to $\mathcal{N}(A)$. Therefore from above we have

$$\|h\|_2^2 = -\sum_{n=1}^{\infty} c_n(s) \int_a^b h(s,t) e_n(t) dt = 0.$$

By the symmetry of R in the two variables the same result would have been achieved fixing t instead of s.

Since $\|h(s,\cdot)\|_2 = 0$, $h(s,\cdot) = 0$ in $L^2[a,b]$ for every $s \in [a,b]$, then by the L^2-convergent decomposition yielded by the Expansion Theorem, it holds that separately in the two variables $R(s,t) = h(s,t) + \sum_{n=1}^{\infty} c_n(s)e_n(t) = \sum_{n=1}^{\infty} c_n(s)e_n(t)$ in $L^2[a,b]$.

Step 2: for every t, $R_n(t,t) \geqslant 0$.

By Step 1, we can rewrite the remainder as

$$R_n(s,\ t) = R(s,\ t) - \sum_{i=1}^{n} \lambda_i e_i(s)e_i(t) = \sum_{i=1}^{\infty} \lambda_i e_i(s)e_i(t) - \sum_{i=1}^{n} \lambda_i e_i(s)e_i(t) = \sum_{i=n+1}^{\infty} \lambda_i e_i(s)e_i(t)$$

with L^2- convergence in each variable separately. Now, given $f \in L^2[a,\,b]$ by the continuity of the inner product it holds that

$$\int_a^b \int_a^b R_n(s,t)f(s)f(t)ds\,dt = (f,(R_n,f)_2)_2 = \left(f(t), \left(\sum_{i=n+1}^{\infty} \lambda_i e_i(s)e_i(t), f(s)\right)_2\right)_2 = \left(f(t), \sum_{i=n+1}^{\infty} \lambda_i e_i(t)(e_i(s), f(s))_2\right)_2 = \sum_{i=n+1}^{\infty} \lambda_i (e_i, f)_2 (f, e_i)_2 = \sum_{i=n+1}^{\infty} \lambda_i (e_i, f)_2^2 \geqslant 0$$

since the λ_i are positive. We shall use this inequality in an argument *per absurdum*.

Suppose that there is a t_0 such that $R_n(t_0,t_0) < 0$. Then by continuity there is an $\varepsilon > 0$ such that for (s,t) in some neighborood of (t_0,t_0), $R_n(s,t) \leqslant -\varepsilon < 0$. Suppose this neighborood is a square described by $t_0 - \alpha < s, t < t_0 + \alpha$. If we define

$$w(s) = \begin{cases} 1, s \in [t_0 - \alpha, t_0 + \alpha] \\ 0, \text{elsewhere} \end{cases},$$

we have trivially that $w \in L^2[a,b]$, and finally that

$$\int_a^b \int_a^b R_n(s,\ t)w(s)w(t)ds\,dt = \int_{t_0-\alpha}^{t_0+\alpha} \int_{t_0-\alpha}^{t_0+\alpha} R_n(s,\ t)ds\,dt \leqslant -\varepsilon \int_{t_0-\alpha}^{t_0+\alpha} \int_{t_0-\alpha}^{t_0+\alpha} ds\,dt = -\varepsilon 4\alpha^2 < 0$$

which is an absurdum, since we have previously stated that for every $f \in L^2[a,\,b]$, $\int_a^b \int_a^b R_n(s,t)f(s)f(t)ds\,dt \geqslant 0$. Thus $R_n(t,t)$ must be positive for every $t \in [a,b]$.

Step 3: the series $\sum_{n=1}^{\infty} \lambda_n e_n(s)e_n(t)$ converges absolutely, and this convergence is uniform in t for each fixed s and *viceversa.*

By Step 2, we know that $R_n(t,\,t) = R(t,\,t) - \sum_{i=1}^n \lambda_i e_i^2(t) \geqslant 0$, hence by the continuity of R on the compact $[a,b]\times[a,b]$, defined $M := \max\limits_{t,s\in[a,b]} |R(t,s)|$, we have that

$$\sum_{i=1}^n \lambda_i e_i^2(t) \leqslant R(t,t) \leqslant M < \infty$$

for every n. Then $\lim_{n\to\infty}\sum_{i=1}^n \lambda_i e_i^2(t) = \sum_{i=1}^\infty \lambda_i e_i^2(t) \leqslant M < \infty$. But $\sum_{i=1}^\infty \lambda_i e_i^2(t)$ is a non-negative terms series, and we have shown its boundness. Thus $\sum_{i=1}^\infty \lambda_i e_i^2(t)$ coverges for every t. In particular it is a Cauchy series, which means that $\sum_{i=n}^m \lambda_i e_i^2(s) \longrightarrow 0,$ as $m,n\to\infty$.

Now, by the Cauchy-Schwarz inequality and the positivity of the λ_i, we have the absolute convergence in both variables jointly and the unifornity in each variable separately:

$$\sum_{i=n}^m \lambda_i |e_i(s)e_i(t)| = \left|\sum_{i=n}^m \sqrt{\lambda_i}|e_i(s)|\sqrt{\lambda_i}|e_i(t)|\right| \leqslant \left(\sum_{i=n}^m [\sqrt{\lambda_i}|e_i(s)|]^2\right)^{\frac{1}{2}}\left(\sum_{i=n}^m [\sqrt{\lambda_i}|e_i(t)|]^2\right)^{\frac{1}{2}} = \left(\sum_{i=n}^m \lambda_i e_i^2(s)\right)^{\frac{1}{2}}\left(\sum_{i=n}^m \lambda_i e_i^2(t)\right)^{\frac{1}{2}} \leqslant \left(\sum_{i=n}^m \lambda_i e_i^2(s)\right)^{\frac{1}{2}}\left(\sum_{i=n}^\infty \lambda_i e_i^2(t)\right)^{\frac{1}{2}} \leqslant \sqrt{M}\left(\sum_{i=n}^m \lambda_i e_i^2(s)\right)^{\frac{1}{2}} \underset{m,n\to\infty}{\longrightarrow} 0$$

for each fixed s, uniformly in t. The argument for t fixed is symmetric and is given by applying Step 2 for $R_n(s,s)$.

Step 4: $\sum\limits_{n=1}^\infty \lambda_n e_n(s)e_n(t) = R(s,t)$ pointwise.

By Step 1, $\sum_{n=1}^\infty \lambda_n e_n(s)e_n(t) = R(s,t)$ in $L_2[a,b]$ separately in the two variables. Fix s. By Step 3, there is a function $G(s,t)$ to which $\sum_{n=1}^\infty \lambda_n e_n(s)e_n(t)$ converges uniformly in t, hence in $L_2 -$ norm (see the last paragraph of Proposition 2.3 for this argument). Thus for each fixed s, $R(s,\,t) = G(s,\,t)$ for *almost every* t. But since for Corollary 2.2 $\{e_n\}$ are continuous, $G(s,t)$ is a uniform limit of continuous functions $g_m(t) = \sum_{n=1}^m \lambda_n e_n(s)e_n(t)$. Hence $G(s,\cdot)$ is a continuous function, as well as $R(s,\cdot)$ is, by hypothesis. Thus for every fixed s, $R(s,t) = G(s,t)$ for every t in $[a,b]$, that is $R(s,t) \equiv G(s,t)$. Thus $\sum_{n=1}^\infty \lambda_n e_n(s)e_n(t)$ converges uniformly in t and pointwise in s to $R(s,t)$. In the same way, fixing t and repeating the proof, we can show that $\sum_{n=1}^\infty \lambda_n e_n(s)e_n(t)$ converges uniformly in s and pointwise in t to $R(s,t)$. Thus $\sum_{n=1}^\infty \lambda_n e_n(s)e_n(t)$ converges pointwise to $R(s,t)$ in $[a,b]\times[a,b]$.

Step 5: (Dini's Theorem) Given a sequence of real-valued functions $\{g_n\}_{n=1}^\infty$ continuous on $[a,\,b]$ and a real-valued function g continuous on $[a,\,b]$ such that $g_n(x) \leqslant g_{n+1}(x)$ for all n and all x and $\lim_{n\to\infty} g_n(x) = g(x)$ for all $x\in[a,\,b]$, then $g_n \longrightarrow g$ as $n\to\infty$, uniformly on $[a,b]$.

Given $\varepsilon > 0$, define the open set $U_n := \{x \in [a, b]\colon |g_n(x) - g(x)| < \varepsilon\}$. Since $\lim_{n\to\infty} g_n(x) = g(x)$ for all $x \in [a, b]$, $\bigcup_{n=1}^{\infty} U_n = [a, b]$. For $\{U_n\}$ is an open cover of the compact $[a, b]$, a finite open subcover of $[a, b]$ can be extract, that is there is a finite set of indeces $\{n_1, \ldots, n_m\}$ such that $\bigcup_{i=1}^{m} U_{n_i} = [a, b]$. By the monotonicity of the sequence $\{g_n\}$, we infer that $U_n \subset U_{n+1}$ for all n, and so the U_{n_i} have the same property. Hence $[a, b] = \bigcup_{i=1}^{m} U_{n_i} = U_m$, so that $|g_m(x) - g(x)| < \varepsilon$ for all $x \in [a, b]$. By the monotonicity again, we have that if $n \geqslant m$, $|g_m(x) - g(x)| < \varepsilon$ for all $x \in [a, b]$. Thus it has been shown that for every $\varepsilon > 0$, there is an m such that if $n \geqslant m$, $|g_m(x) - g(x)| < \varepsilon$ for every $x \in [a, b]$, which is the uniformly convergence of the function sequence $\{g_n\}_{n=1}^{\infty}$.

Proof of Mercer's Theorem:

By Step 4, $\sum_{n=1}^{\infty} \lambda_n e_n^2(s) = R(s, s)$ pointwise on $[a, b]$. Let $g_n(s)\colon = \sum_{i=1}^{n} \lambda_i e_i^2(s)$. Then g_n is continuous on $[a, b]$, $g_n(s) \leqslant g_{n+1}(s)$ for all n and for all $s \in [a, b]$ and $\lim_{n\to\infty} g_n(s) = R(s, s)$ for all $s \in [a, b]$. Hence by Step 5, $g_n(s) \longrightarrow R(s, s)$, as $n \to \infty$ uniformly on $[a, b]$.

Now we proceed as in Step 3, to gain uniform convergence in both variables. Indeed, the Cauchy-Schwarz inequality yields

$$\left|\sum_{i=m}^{n} \lambda_i e_i(s) e_i(t)\right|^2 \leqslant \sum_{i=n}^{m} [\sqrt{\lambda_i} e_i(s)]^2 \sum_{i=n}^{m} [\sqrt{\lambda_i} e_i(t)]^2 \Bigg) < M \sum_{i=m}^{n} \lambda_i e_i^2(s) \underset{n,m\longrightarrow\infty}{\longrightarrow} 0$$

uniformly in t for each fixed s, but in addition the convergence is also uniform in s, since here we have proved that, by Step 5, $\sum_{n=1}^{\infty} \lambda_n e_n^2(s) = R(s, s)$ uniformly on $[a, b]$. Therefore the pointwise convergence $\sum_{i=1}^{\infty} \lambda_i e_i(s) e_i(t) = R(s, t)$ on $[a, b] \times [a, b]$ stated in Step 4 is also uniform on $[a, b] \times [a, b]$ in both variables jointly and not only separately.

The uniform convergence of the series to $R(s, t)$ on $[a, b] \times [a, b]$ implies the convergence in $L^2([a, b] \times [a, b])$, by the same argument shown in the last step of the Proposition 2.3, that is $\int_a^b \int_a^b |R(s, t) - \sum_{i=1}^{n} \lambda_i e_i(s) e_i(t)|^2 ds\, dt \longrightarrow 0$ as $n \longrightarrow \infty$.

$\square$

3 Homogeneous linear integral equations

In the present Section we leave for a while the straight path toward the Karhunen-Loève Expansion we were about to follow, in order to achieve a result which will become crucial for the study of the optimality of the mean square error. In doing so, we change reference and adapt to our purposes the methods found in Courant and Hilbert (1989), pp. 122-125.

Definition 3.1. Let $K: [a, b] \times [a, b] \longrightarrow \mathbb{R}$ be a continuous function and let λ be a scalar parameter. Let f, φ be two real-valued continuous functions on $[a, b]$. The integral equation

$$\varphi(s) = f(s) - \lambda \int_a^b K(s, t) f(t) dt$$

is called a *linear integral equation of the second kind with the kernel* K. If $\varphi \equiv 0$ we are dealing with a *homogeneous linear integral equation of the second kind.* In this case we shall call the reciprocal of a value $\lambda \neq 0$, for which the equation possesses non trivial solutions, *eigenvalue of the kernel* (this convention is not universal though). Such eigenvalues actually correspond to the non zero eigenvalues of the integral operator $(A f)(s) = \int_a^b K(s,\ t) f(t) dt$. The corresponding solutions $f_1,\ \ldots,\ f_k$ are called *eigenfunctions of the kernel* for the eigenvalue λ^{-1}, and are the same of the corresponding Hilbert-Schmidt integral operator.

Remark 3.2. If a homogeneous equation possesses a solution other than the trivial one $f \equiv 0$, the solution multiplied by an arbitrary constant remains a solution, therefore it may be assumed normalized. Furthermore, by the linearity of the integral equation, if $f_1, \ldots, f_k$ are solutions, then all the linear combinations $c_1 f_1 + \ldots + c_k f_k$ are solutions. Therefore if several linearly independent solutions are given, they may be assumed orthonormalized (e. g. by the Gram-Schmidt process in $L^2[a, b]$) without ceasing to be solutions.

Remark 3.3. As a consequence of the Theorem 1.17, the eigenvectors corresponding to each eigenvalue of the integral equation are finite.

3.1 Properties of the Hilbert-Schmidt functions

In connection with integral equations it is natural to inquire into the properties of functions, that we shall call Hilbert-Schmidt functions analogously to the operator of the previous section, of the form:

$$g(s) = \int_a^b K(s, t) h(t)\, dt.$$

Proposition 3.4. If there is a constant M such that $h \in L_2[a, b]$, $\|h\|_2 \leqslant M < \infty$, then for every $\varepsilon > 0$ there is a $\delta_\varepsilon > 0$, independent of the particular function h, such that $|g(s + \eta) - g(s)| < \varepsilon$ whenever $|\eta| < \delta_\varepsilon$.

Proof. Let $\|h\|_2 \leqslant M < \infty$. By the Cauchy-Schwarz inequality

$$|g(s+\eta) - g(s)| \leqslant \int_a^b |K(s+\eta,t) - K(s,t)||h(t)|\,dt \leqslant \|h\|_2 \left[\int_a^b |K(s+\eta,t) - K(s,t)|^2\right]^{\frac{1}{2}} \leqslant M\left[\int_a^b |K(s+\eta,t) - K(s,t)|^2\right]^{\frac{1}{2}}.$$

Now, since $K(s,\,t)$ is continuous on the compact $[a,\,b]\times[a,\,b]$, it is clear that it is uniformly continuous, then there is a $\sigma_\eta > 0$ independent of $(s,\,t)$ such that $|K(s+\eta,t) - K(s,t)| < \sigma_\eta$. Then

$$|g(s+\eta) - g(s)| \leqslant M\left[\int_a^b \sigma_\eta^2\right]^{\frac{1}{2}} = M\sqrt{b-a}\,\sigma_\eta$$

and taking η such that $|K(s+\eta,t) - K(s,t)| < \sigma'_\eta := \frac{\sigma_\eta}{M\sqrt{b-a}}$ the result follows. □

Corollary 3.5. If a sequence of kernels is given for which $\lim_{n\longrightarrow\infty} K_n(s,t) = K(s,t)$ uniformly, the set of fuctions $g_n(s) = \int_a^b K_n(s,t)h(t)dt$, $g(s) = \int_a^b K(s,t)h(t)dt$ is an equicontinuous set, as long as $h \in L_2[a,b], \|h\|_2 \leqslant M < \infty$.

Proof. For $h \in L^2[a,b]$, the relation $g(s) = \lim_{n\longrightarrow\infty}\int_a^b K_n(s,t)h(t)dt$ certainly holds, and the convergence is uniform in s since the passage to the limit may be performed under the integral sign. The result follows then by Proposition 3.4. □

Remark 3.6. $\{g_n\}_{n=1}^\infty$ is also an uniformly bounded set.

Proof. Suppose $\|h\|_2 \leqslant M < \infty$, by the Cauchy-Schwarz inequality it holds that there is a common bound for the set: $|g_n(s)| \leqslant \sqrt{M}\left[\int_a^b K_n^2(s,\,t)dt\right]^{\frac{1}{2}}$ and $|g(s)| \leqslant \sqrt{M}\left[\int_a^b K^2(s,t)dt\right]^{\frac{1}{2}}$. The result follows. □

3.2 The upper bound of the bilinear integral form

The theory of integral equations can be developed in great detail if the kernel is symmetric, that is if $K(s,t) = K(t,s)$. The aim of this section is actually to study the bilinear integral form

$$J(f,g) = \int_a^b\int_a^b K(s,t)f(s)g(t)ds\,dt$$

and in particular the quadratic form $J(f,\,f)$ when the kernel is non-negative definite, that is $J(f,\,f) \geqslant 0$, and with $f \not\equiv 0$.

We consider the problem of finding a *normalized* function $f \not\equiv 0$ for which $J(f,f)$ assumes the greatest possible value. The problem is well-posed since $J(f,\,f) < \infty$.

The method we shall follow in the Theorem 3.7 is originally due to Holmgren.

Theorem 3.7. The positive least upper bound of $J(f,f)$ is actually attained for the normalized eigenfunction of the kernel corresponding to its greatest eigenvalue.

Proof. By the Mercer's Theorem a non-negative symmetric kernel is uniformly approximated by degenerate symmetric kernels of the form $K_n(s,\ t) = \sum_{i=1}^{n} \lambda_i e_i(s) e_i(t)$, where *degenerate* means here that they are obtained by a finite sum of functions of s and t. The maximizing problem for

$$J_n(f,f) = \int_a^b \int_a^b K_n(s,t) f(s) f(t) ds dt$$

with the subsidiary condition of normality of the solution f turns out to be equivalent to the corresponding problem for a quadratic form in n variables. Indeed, setting for every $i = 1, \ldots, n$, $(f, e_i)_2 = x_i$, we obtain

$$J_n(f,\ f) = \int_a^b \int_a^b K_n(s,\ t) f(s) f(t) ds dt = \int_a^b \int_a^b \sum_{i=1}^{n} \lambda_i e_i(s) e_i(t) f(s) f(t) ds dt =$$
$$\sum_{i=1}^{n} \lambda_i \int_a^b e_i(s) f(s) ds \int_a^b e_i(t) f(t) dt = \sum_{i=1}^{n} \lambda_i (f, e_i)_2^2 = \sum_{i=1}^{n} \lambda_i x_i^2.$$

Now, by *Bessel's inequality* we know that $1 = (f,\ f)_2 \geqslant \sum_{i=1}^{n} x_i^2$. Therefore the maximum of the form is attained when $\sum_{i=1}^{n} x_i^2 = 1$, since otherwise the value of $J_n(f,f)$ would be increased if we multiply f by a suitable factor. Now, without loss of generality, suppose that $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n$. Then

$$\sum_{i=1}^{n} \lambda_i x_i^2 \leqslant \sum_{i=1}^{n} \lambda_1 x_i^2 = \lambda_1 \sum_{i=1}^{n} x_i^2 = \lambda_1 \cdot 1 = \lambda_1$$

Thus the maximum is attaied by $x_1 = 1, x_2 = 0, \ldots, x_n = 0$, the value of the maximum is $\lambda_{1,n}$ and f_n is given by the relations $(f, e_1)_2 = 1, (f, e_i)_2 = 0 \text{ for } i > 1,$ hence $f_n = e_1$, which means that f_n is the eigenfunction of the kernel K_n corresponding to the eigenvalue $\lambda_{1,n}$. Thus

$$\lambda_{1,n} f_n(s) = \int_a^b K_n(s,t) f_n(t) dt.$$

We now let n increase: $\lambda_{1,n}$ will increase as well, and it ends up to converge to a number λ_1, which is the maximum of $J(f,\ f)$. We prove this. Since $K_n(s,t)$ approximates uniformly $K(s,t)$, for every $\varepsilon > 0$ there is a sufficiently large integer n such that $|K(s,t) - K_n(s,t)| < \varepsilon$ for every $(s,t) \in [a,b] \times [a,b]$. Then by the Cauchy-Schwarz inequality and the normality of the solution f it holds that for n

$$|J(f,\ \ f) - J_n(f,\ \ f)| \leqslant \int_a^b \int_a^b |K(s,\ \ t) - K_n(s,\ \ t)||f(s) f(t)| ds dt <$$
$$\varepsilon \left[\int_a^b |f(s)| ds \right] \left[\int_a^b |f(t)| dt \right] = \varepsilon \left[\int_a^b |f(t)| \cdot 1 dt \right]^2 \leqslant \varepsilon \|f\|_2 \sqrt{b-a} = \varepsilon \sqrt{b-a}.$$

Since the range of $J_n(f, f)$ coincides arbitrarily closely with the range of $J(f, f)$, the same must be true for the upper bounds. Consequently $\lambda_1 = \lim_{n\to\infty}\lambda_{1,n}$ exists, with all $\lambda_{1,n} < \lambda_1$. Therefore $\lambda_{1,n} f_n(s) = \int_a^b K_n(s, t) f_n(t) dt$ defines a sequence of functions $\{f_n\}$ uniformly boundend and equicontinuus on the compact $[a, b]$ by the Proposition 3.4, the Corollary 3.5 and the Remark 3.6. Then by the *Ascoli-Arzelà Theorem* there is a uniformly convergent subsequence $\{f_{n_j}\}$. Let φ be the uniform limit. By the uniform convergence of K_n and f_n the integral relations satisfied by the f_n take the limit inside the integral, hence:

$$\lambda_1 \varphi(s) = \int_a^b K(s, t)\varphi(t) dt$$

$$(\varphi, \varphi)_2 = 1$$

$$J(\varphi, \varphi) = \lambda_1$$

As can be seen, the function φ solves the maximum problem for the form $J(f, f)$; it is an eigenfunction of the kernel K, and the maximum is the greatest eigenvalue of the Hilbert-Schmidt integral operator. $\square$

4 The Karhunen-Loève Theorem

In this section the second-order random processes are briefly introduced and their Karhunen-Loève decomposition derived. In introducing the random processes, we follow Loève (1978), pp. 121-135.

4.1 Second order random variables

We shall consider random functions formed by random variables whose second moment, and therefore their mixed second moments, are finite. Their second order properties are those which can be expressed in terms of these moments. Up to equivalences, the random variables in question can be interpreted as points in a Hilbert space. Thus, it is to be expected that the study of the second order properties will only require analytical tools.

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X, Y, \ldots$ be second order real-valued random variables, that is $E|X|^2 < \infty$, $E|Y|^2 < \infty$, so that by the Cauchy-Schwarz inequality their mixed second moments $E[XY] < \infty$. This hypothesis allow us to center at expectations every variable, since having finite second moments, the first moment results finite as well. In particular we shall assume, without loss of generality, that each random variable is centered. Therefore $E|X|^2$ turns to be the *variance* of X and $E[XY]$ the *covariance* of X and Y. The space $L^2(\Omega, \mathcal{F}, P)$ of equivalence classes of such random variables is a Hilbert space, with the inner product $E[XY]$ and the inducted norm $\|X\|_2 = \sqrt{(E|X|^2)}$ and distance between the points X, Y determined by $\|X - Y\|_2$. Thus the convergence in norm is a convergence in quadratic mean in this space (or *mean square* convergence). Given random variables $\{X_n\}_{n=1}^{\infty}$, we say that X_n converges in norm to X if $\|X_n - X\|_2 \longrightarrow 0$, as $n \to \infty$, which is equivalent to $E|X_n - X|^2 \longrightarrow 0$, as $n \to \infty$. In this space X, Y are said to be orthogonal, and we write $X \perp Y$, if $E[XY] = 0$. In particular $X \perp X$ if and only if $E|X|^2 = 0$, that is $X = 0$ *almost surely*. Moreover, by orthogonality we have that if $X \perp Y$ then $E|X + Y|^2 = E|X|^2 + E|Y|^2$. More generally if $X_1, X_2, \ldots$, X_n are mutually orthogonal random variables, then $E|\sum_{k=1}^n X_k|^2 = \sum_{k=1}^n E|X_k|^2$.

We introduce now the second order random functions. Let τ be an argument varying on a set $T \subseteq \mathbb{R}$. A *random function* (or equivalently a *random process*, although not all authors accept it as synonym) is a family of random variables $\{X_\tau, \tau \in T\}$. A second order random function is a family $\{X_\tau, \tau \in T\}$ such that for every $\tau \in T$, the *variances* $E|X_\tau|^2 < \infty$. As previously stated, we can and do assume that the second order random functions under consideration are centered at expectations. Eventually we say that the function is a *continuous-parameter* process when its set of indices $T \subseteq \mathbb{R}$ is uncountable.

The function defined on $T \times T$, $K_X(t, \tau) := E[X_t X_\tau]$ is called the *covariance* of the random function X_τ on T (some authors refer to it as *autocovariance*).

According to the Cauchy-Schwarz inequality, the covariance of the process exists and is finite. Conversely, if the function defined on $T \times T$, $K_X(t, \tau) := E[X_t X_\tau]$ exists and is finite, then $E|X_\tau|^2 = K_X(\tau, \tau) < \infty$ for every $\tau \in T$. Therefore second order random functions can be defined as those having covariances. Their second order properties are those which can be defined or determined by means of covariances.

Let $X = \{X_t\}$, $t \in [a,b] \subset \mathbb{R}$ be a continuous-parameter real-valued second-order random process with zero mean and covariance function $K_X(t,\tau)$.

Remark 4.1. K_X is a symmetric non-negative definite function.

Proof. First, by definition of $K_X(t,\,\tau)$: $=E[X_tX_\tau]$, its symmetry is trivial. Next, it holds that for all possible choices of $t_1,\ldots t_n \in [a,b]$, and all possible functions f: $[a,b] \longrightarrow \mathbb{R}$,

$$\sum_{i,j=1}^{n} f(t_i)K_X(t_i,\,t_j)f(t_j) = \sum_{i,j=1}^{n} f(t_i)E[X_{t_i}X_{t_j}]f(t_j) = E\left[\sum_{i,j=1}^{n} f(t_i)X_{t_i}X_{t_j}f(t_j)\right] =$$
$$E\left[\sum_{i=1}^{n} f(t_i)X_{t_i}\left(\sum_{j=1}^{n} f(t_j)X_{t_j}\right)\right] = E\left[\sum_{i=1}^{n} f(t_i)X_{t_i}\sum_{j=1}^{n} f(t_j)X_{t_j}\right] =$$
$$E\left[\left(\sum_{i=1}^{n} f(t_i)X_{t_i}\right)^2\right] \geqslant 0,$$

thus K_X is non-negative definite by definition. □

4.2 The Riemann integral of a stochastic process

We now go straight to the proof of the main theorem of this dissertation. In doing so, we shall walk the path traced by Ash (1990) again.

Definition 4.2. Given g: $[a,\,b] \longrightarrow \mathbb{R}$, we define $\int_a^b g(t)X_t$ dt as follows. Let Δ: $a = t_0 < t_1 < \ldots < t_n = b$ be a partition of $[a,b]$ with $|\Delta| := \max_{i=1}^n |t_i - t_{i-1}|$. Define the random variable $I(\Delta) := \sum_{k=1}^n g(t_k)X_{t_k}(t_k - t_{k-1})$. If the family of random variables $\{I(\Delta)\}_\Delta$ converges in mean square to a random variable I as $|\Delta| \longrightarrow 0$, that is $E|I(\Delta) - I|^2 \longrightarrow 0$ as $|\Delta| \longrightarrow 0$, then we say that $g(t)X_t$ is integrable over $[a,b]$ and that $\int_a^b g(t)X_t\,\mathrm{dt}$: $=I$, which is said the *integral of* $g(t)X_t$ *over* $[a,b]$.

Proposition 4.3. If g: $[a,\,b] \longrightarrow \mathbb{R}$ is continuous and K_X: $[a,\,b] \times [a,\,b] \longrightarrow \mathbb{R}$ is continuous, then $g(t)X_t$ is integrable over $[a,b]$.

Proof. We have to prove that $I(\Delta)$ is convergent in mean square as $|\Delta| \longrightarrow 0$. Being X a second-order process, for every t, $X_t \in L_2(\Omega, \mathcal{F}, P)$; hence we shall equivalently prove that $\{I(\Delta)\}_\Delta$ is a Cauchy sequence with respect to the L^2-norm squared, that is to say mean square norm. This means that given two partitios Δ and Δ', $E[(I(\Delta) - I(\Delta'))^2] \longrightarrow 0$, as $|\Delta|, |\Delta'| \longrightarrow 0$. The result will follow by the completeness of L_2. Since $E[(I(\Delta) - I(\Delta'))^2] = E[I^2(\Delta)] + E[I^2(\Delta')] - 2E[I(\Delta)I(\Delta')]$, we shall consider the limit of each block, as $|\Delta|, |\Delta'| \longrightarrow 0$.

Let Δ: $a = t_0 < t_1 < \ldots < t_n = b$, Δ': $a = \tau_0 < \tau_1 < \ldots < \tau_m = b$ be two partitions of $[a, b]$. Then

$I(\Delta)I(\Delta') = \sum_{i=1}^{n} g(t_i) X_{t_i} (t_i - t_{i-1}) \sum_{j=1}^{m} g(\tau_j) X_{\tau_j} (\tau_j - \tau_{j-1}) = \sum_{i=1}^{n} \sum_{j=1}^{m} g(t_i) g(\tau_j) X_{t_i} X_{\tau_j} (t_i - t_{i-1})(\tau_j - \tau_{j-1})$.

Thus

$E[I(\Delta)I(\Delta')] = E\left[\sum_{i=1}^{n} \sum_{j=1}^{m} g(t_i) g(\tau_j) X_{t_i} X_{\tau_j} (t_i - t_{i-1})(\tau_j - \tau_{j-1})\right] = \sum_{i=1}^{n} \sum_{j=1}^{m} g(t_i) g(\tau_j) E[X_{t_i} X_{\tau_j}](t_i - t_{i-1})(\tau_j - \tau_{j-1}) = \sum_{i=1}^{n} \sum_{j=1}^{m} g(t_i) g(\tau_j) K_X(t_i, \tau_j)(t_i - t_{i-1})(\tau_j - \tau_{j-1}) \longrightarrow \int_a^b \int_a^b g(t) g(\tau) K_X(t, \tau)\, \mathrm{d}t \mathrm{d}\tau$ as $|\Delta|, |\Delta'| \longrightarrow 0$.

Indeed $\sum_{i=1}^{n} \sum_{j=1}^{m} g(t_i) g(\tau_j) K_X(t_i, \tau_j)(t_i - t_{i-1})(\tau_j - \tau_{j-1})$ is an *approximating sum* to the two-dimensional Riemann integral above, which is well defined by the continuity of the integrand over $[a, b] \times [a, b]$. It has been proved that

$$\lim_{|\Delta|, |\Delta'| \longrightarrow 0} E[I(\Delta)I(\Delta')] = \int_a^b \int_a^b g(t) g(\tau) K_X(t, \tau)\, \mathrm{d}t \mathrm{d}\tau.$$

Furthermore, taking $\Delta = \Delta'$ we have that

$$\lim_{|\Delta| \longrightarrow 0} E[I^2(\Delta)] = \int_a^b \int_a^b g(t) g(\tau) K_X(t, \tau)\, \mathrm{d}t \mathrm{d}\tau = \lim_{|\Delta'| \longrightarrow 0} E[I^2(\Delta')].$$

As previously stated, in order to have $\{I(\Delta)\}_\Delta$ satisfying the Cauchy property in mean square norm, it has to be shown that $E[(I(\Delta) - I(\Delta'))^2] \longrightarrow 0$ as $|\Delta|, |\Delta'| \longrightarrow 0$.

$E[(I(\Delta) - I(\Delta'))^2] = E[I^2(\Delta)] + E[I^2(\Delta')] - 2E[I(\Delta)I(\Delta')] \longrightarrow \int_a^b \int_a^b g(t) g(\tau) K_X(t, \tau)\, \mathrm{d}t \mathrm{d}\tau + \int_a^b \int_a^b g(t) g(\tau) K_X(t, \tau)\, \mathrm{d}t \mathrm{d}\tau - 2\int_a^b \int_a^b g(t) g(\tau) K_X(t, \tau)\, \mathrm{d}t \mathrm{d}\tau = 0$. $\square$

Lemma 4.4. If $g, h: [a, b] \longrightarrow \mathbb{R}$ and $K_X: [a, b] \times [a, b] \longrightarrow \mathbb{R}$ are continuous, then

$$E\left[\int_a^b g(t) X_t\, \mathrm{d}t \int_a^b h(\tau) X_\tau d\tau\right] = \int_a^b \int_a^b g(t) h(\tau) K_X(t, \tau)\, \mathrm{d}t \mathrm{d}\tau.$$

Furthermore $E\left[\int_a^b g(t) X_t\, \mathrm{d}t\right] = E\left[\int_a^b h(\tau) X_\tau d\tau\right] = 0$.

Proof. As in Proposition 4.3 let Δ and Δ' be two partitions of $[a,b]$. Let

$$I(\Delta)=\sum_{i=1}^{n} g(t_i)X_{t_i}(t_i-t_{i-1}),\ I(\Delta')=\sum_{j=1}^{m} h(\tau_j)X_{\tau_j}(\tau_j-\tau_{j-1}),\ I=\int_a^b g(t)X_t\,\mathrm{dt},$$

$$J=\int_a^b h(\tau)X_\tau d\tau.$$

By Proposition 4.3 $I(\Delta)\longrightarrow I$ and $J(\Delta')\longrightarrow J$ in $L_2(\Omega,\mathcal{F},P)$ as $|\Delta|,|\Delta'|\longrightarrow 0$. Therefore $E[I^2(\Delta)]\longrightarrow E[I^2]$ and $E[J^2(\Delta')]\longrightarrow E[J^2]$ by the continuity of the inner product of $L^2(\Omega)$. Furthermore we know that $\lim_{|\Delta|,|\Delta'|\longrightarrow 0}E[I(\Delta)J(\Delta')]=\int_a^b\int_a^b g(t)h(\tau)K_X(t,\,\tau)\ \mathrm{dt}d\tau$, for the argument of Proposition 4.3 can be applied, even if we have h instead of g in the approximating sum $J(\Delta')$. Thus if we prove that $\lim_{|\Delta|,|\Delta'|\longrightarrow 0}E[I(\Delta)J(\Delta')]=E[IJ]$, the first statement shall follow. By the triangular inequality first and the Cauchy-Schwarz inequality next, it holds that

$|E[I(\Delta)J(\Delta')]-E[IJ]|=|E[I(\Delta)J(\Delta')-IJ]|=|E[I(\Delta)J(\Delta')-I(\Delta)J+I(\Delta)J-IJ]|\leqslant E[|I(\Delta)J(\Delta')-I(\Delta)J|+|I(\Delta)J-IJ|]=E|I(\Delta)(J(\Delta')-J)|+E|(I(\Delta)-I\ )J|\leqslant\{E[I^2(\Delta)]E[(J(\Delta')-J)^2]\}^{\frac{1}{2}}+\{E[J^2]E[(I(\Delta)-I)^2]\}^{\frac{1}{2}}\longrightarrow\sqrt{E[I^2]}\cdot 0+\sqrt{E[J^2]}\cdot 0=0$ as $|\Delta|,|\Delta'|\longrightarrow 0$.

This argument proves the first statement, as mentioned above. Similarly we shall proceed for the latter statement, which is equivalent to $E[I]=E[J]=0$. Let us prove it for I. Since the process X is centered, $E[I(\Delta)]=\sum_{i=1}^n g(t_i)E[X_{t_i}](t_i-t_{i-1})$=0; hence we only need to prove that $E[I(\Delta)]\longrightarrow E[I]$ as $|\Delta|\longrightarrow 0$. This is straightforward: $L^1(\Omega)\supset L^2(\Omega)$, thus if $I(\Delta)\longrightarrow I$ in L^2, then $I(\Delta)\longrightarrow I$ in L^1 which is equivalent to $E|I(\Delta)-I|\longrightarrow 0$. Since $E|I(\Delta)-I|\geqslant|E[I(\Delta)-I]|=|E[I(\Delta)]-E[I]|$, we have proved that $|E[I(\Delta)]-E[I]|\longrightarrow 0$ as $|\Delta|\longrightarrow 0$. Now, for every $\Delta, E[I(\Delta)]=0$, thus $0=E[I(\Delta)]\longrightarrow E[I]$ as $|\Delta|\longrightarrow 0$, which is equivalent to $E[I]=0$. $\square$

Lemma 4.5. If $h\colon[a,b]\longrightarrow\mathbb{R}$ and $K_X\colon[a,b]\times[a,b]\longrightarrow\mathbb{R}$ are continuous, then for every $t\in[a,b]$

$$E\left[X_t\int_a^b h(\tau)X_\tau d\tau\right]=\int_a^b h(\tau)K_X(t,\tau)\,d\tau$$

Proof. As in the Lemma 4.4 let $J(\Delta')=\sum_{j=1}^m h(\tau_j)X_{\tau_j}(\tau_j-\tau_{j-1})$ and $J=\int_a^b h(\tau)X_\tau\ d\tau$. Since $J(\Delta')\longrightarrow J$ in mean square as $|\Delta'|\longrightarrow 0$, on one hand $E[X_tJ(\Delta')]\longrightarrow E[X_tJ]$: indeed, just as in the Lemma 4.4, by replacing I and $I(\Delta)$ by X_t, which is independent from Δ', we have that

$|E[X_tJ(\Delta')]-E[X_tJ]|=E|X_t(J(\Delta')-J)|\leqslant\{E[X_t^2]E[(J(\Delta')-J)^2]\}^{\frac{1}{2}}\longrightarrow\sqrt{E[X_t^2]}\cdot 0=0$ as $|\Delta'|\longrightarrow 0$, since $X_t\in L_2$.

On the other hand, $E[X_t J(\Delta')] \longrightarrow \int_a^b h(\tau) K_X(t, \tau)\, d\tau$ as $|\Delta'| \longrightarrow 0$, since as in Proposition 4.3, by replacing $I(\Delta)$ by X_t, we have an approximating sum to a Riemann integral:

$$E[X_t J(\Delta')] = E\left[\sum_{j=1}^{m} h(\tau_j) X_t X_{\tau_j} (\tau_j - \tau_{j-1})\right] = \sum_{j=1}^{m} h(\tau_j) E[X_t X_{\tau_j}](\tau_j - \tau_{j-1}) = \sum_{j=1}^{m} h(\tau_j) K_X(t, \tau_j)(\tau_j - \tau_{j-1}) \longrightarrow \int_a^b h(\tau) K_X(t, \tau)\, d\tau \text{ as } |\Delta'| \longrightarrow 0.$$

Thus the lemma follows. □

4.3 The Karhunen-Loève Theorem

Theorem 4.6. (Karhunen-Loève Expansion) Let $X = \{X_t\}$, $t \in [a, b] \subset \mathbb{R}$, a, $b < \infty$, be a continuous-parameter real-valued second-order random process with zero mean and continuous covariance function $K_X(t, \tau)$.

Then for every $t \in [a, b]$ we may decompose

$$X_t = \sum_{k=1}^{\infty} Z_k e_k(t)$$

with

$$Z_k = \int_a^b X_t e_k(t)\, \mathrm{dt},$$

where the $\{e_k\}_{k=1}^{\infty}$ are the eigenfunctions of the Hilbert-Schmidt integral operator on $L^2[a, b]$, $(Af)(t) = \int_a^b K_X(t, \tau) f(\tau)\, d\tau$; $\{e_k\}_k$ is the Karhunen-Loève basis, which is an orthonormal basis for the space spanned by the eigenfunctions corresponding to the non-zero eigenvalues of A; each random variable Z_k is the coefficient given by the projection of X_t onto the k-th deterministic element of the Karhunen-Loève basis in $L^2(\Omega, \mathcal{F}, P)$; moreover the $\{Z_k\}$ are pairwise orthogonal random variables with zero mean and variance λ_k, where λ_k is the eigenvalue corresponding to the eigenfunction e_k.

The series $\sum_{k=1}^{\infty} Z_k e_k(t)$ converges to X_t in mean square, uniformly for $t \in [a, b]$.

Proof. By Corollary 2.2 the eigenfunctions corresponding to the non-zero eigenvalues of A are continuous, thus by Proposition 4.3 $Z_k = \int_a^b X_t e_k(t)\, \mathrm{dt}$ is a well defined random variable. By Lemma 4.4 $E[Z_k] = E\left[\int_a^b X_t e_k(t)\, \mathrm{dt}\right] = 0$ and

$$E[Z_i Z_j] = E\left[\int_a^b X_t e_i(t)\ \mathrm{dt} \int_a^b X_\tau e_j(\tau)\ d\tau\right] = \int_a^b \int_a^b e_i(t) e_j(\tau) K_X(t, \tau)\ d\tau dt = \int_a^b e_i(t) \int_a^b e_j(\tau) K_X(t, \tau) d\tau dt = \lambda_j \int_a^b e_i(t) e_j(t) dt = \lambda_j \delta_{\mathrm{ij}}$$

where δ_{ij} is the *Kronecker's delta*, by the orthonormality of $\{e_k\}_k$ in $L_2[a,b]$. Thus, on one hand, since $E[Z_iZ_j]=0, \forall i\neq j$, the random variables $\{Z_k\}_{k\in\mathbb{N}}$ are pairwise orthogonal in $L^2(\Omega)$ (because of this random coefficients's property, the Karhunen-Loève Expansion is usually referred as *bi-orthogonal*); on the other hand, by definition, $\mathrm{var}(Z_k):=E[Z_kZ_k]=\lambda_k\cdot 1=\lambda_k$.

Let us show the mean square convergence of the series. Let $S_n(t):=\sum_{k=1}^n Z_ke_k(t)$. Then

$$E|S_n(t)-X_t|^2=E[S_n^2(t)-2S_n(t)X_t+X_t^2]=E[S_n^2(t)]-2E[S_n(t)X_t]+E[X_t^2].$$

Since

$$E[X_t^2]:=K_X(t,t)$$

$$E[S_n(t)X_t]=E\left[\sum_{k=1}^n Z_ke_k(t)X_t\right]=\sum_{k=1}^n e_k(t)E[Z_kX_t]$$

$$E[S_n^2(t)] = E\left[\sum_{i=1}^n Z_ie_i(t)\sum_{j=1}^n Z_je_j(t)\right] = E\left[\sum_{i,j=1}^n e_i(t)e_j(t)Z_iZ_j\right] = \sum_{i,j=1}^n e_i(t)e_j(t)E[Z_iZ_j]=\sum_{i,j=1}^n e_i(t)e_j(t)\lambda_{ij}\delta_{ij}=\sum_{k=1}^n \lambda_ke_k^2(t)$$

it holds that

$$E|S_n(t)-X_t|^2=\sum_{k=1}^n \lambda_ke_k^2(t)-2\sum_{k=1}^n e_k(t)E[Z_kX_t]+K_X(t,t).$$

By the Lemma 4.5 $E[Z_kX_t]=E\left[X_t\int_a^b X_\tau e_k(\tau)d\tau\right]=\int_a^b e_k(\tau)K_X(t,\tau)d\tau=\lambda_ke_k(t)$.Thus

$$E|S_n(t)-X_t|^2=\sum_{k=1}^n \lambda_ke_k^2(t)-2\sum_{k=1}^n \lambda_ke_k^2(t)+K_X(t,t)=K_X(t,t)-\sum_{k=1}^n \lambda_ke_k^2(t)\underset{n\to\infty}{\longrightarrow}0,$$

uniformly in $t\in[a,b]$ by Mercer's Theorem. □

Remark 4.7. Since we have just shown that $E[Z_k]=0$ and $\mathrm{var}[Z_k]=\lambda_k$, we could define $Z_k^*:=\frac{Z_k}{\sqrt{\lambda}}$ and decompose X_t with the $\{Z_k^*\}$ in a more popular way:

$$X_t=\sum_{k=1}^\infty \sqrt{\lambda_k}Z_k^*e_k(t)$$

with the advantage that not only by linearity, $E[Z_k^*]=0$, but $\mathrm{var}[Z_k^*]=\frac{1}{\lambda_k}\mathrm{var}[Z_k]=1$.

Remark 4.8. If the random process $\{X_t\}$ is a non zero-mean process, we can subtract its mean first before applying the Karhuen-Loève Expansion, and the result is trivially that

$$X_t = E[X_t] + \sum_{k=1}^{\infty} Z_k e_k(t)$$

4.4 Optimality of the Karhunen-Loève basis

Remark 4.9. As a consequence of the Karhunen-Loève Theorem we can easily evaluate the variances:

$$\mathrm{var}[X_t] = E[X_t^2] - E^2[X_t] = K_X(t,t) - 0 = \sum_{k=1}^{\infty} \lambda_k e_k^2(t)$$

and consequently we draw the conclusion that the *total variance* of the process, defined as the integral of $\mathrm{var}[X_t]$ over the indices interval $[a,b]$, is

$$\int_a^b \mathrm{var}[X_t]\,dt = \int_a^b \sum_{k=1}^{\infty} \lambda_k e_k^2(t)\,dt = \sum_{k=1}^{\infty} \int_a^b \lambda_k e_k^2(t)\,dt = \sum_{k=1}^{\infty} \lambda_k \int_a^b e_k^2(t)\,dt = \sum_{k=1}^{\infty} \lambda_k$$

since the convergence of the sum is uniform by Mercer's Theorem. As a result, if the Karhunen-Loève Expansion gets trucated, we have that the $N-$ truncated expansion explains the $\sum_{k=1}^{N} \lambda_k / \sum_{k=1}^{\infty} \lambda_k$ of the total variance of the process.

This fact is somewhat related to the truncation error, the main topic of the remaining part of this section. We first seek a representation for the truncation error, in order to show the optiality of the Karhunen-Loève Expansion. In doing so, useful hints have been found in Brown (1960).

Remark 4.10. Given an orthonormal basis $\{f_k\}_{k\in\mathbb{N}}$ of $L_2[a,\ b]$, let us suppose that the following expansion in $L_2(\Omega)$, $X_t = \sum_{k=1}^{\infty} A_k f_k(t)$, with $A_k = \int_a^b X_t f_k(t) dt$ holds. Then the truncated sum $X_t^n = \sum_{k=1}^{n} A_k f_k(t)$ approximates X_t by an error $\varepsilon_n(t) = |X_t - X_t^n| = |X_t - \sum_{k=1}^{n} A_k f_k(t)|$.

Definition 4.11. Define the *mean square error* as $E[\varepsilon_n^2(t)]$, then $\int_a^b \mathrm{E}[\varepsilon_n^2(t)]\,dt =: \mathcal{E}_n^2(\{f_k\})$ defines the *total mean square error*, also called *integrated mean square error*, with respect to the basis $\{f_k\}_{k\in\mathbb{N}}$ that X_t is projected onto, for every $t\in[a,b]$.

Lemma 4.12.

$$\mathcal{E}_n^2(\{f_k\}) = \int_a^b E[X_t^2]\mathrm{dt} - \sum_{k=1}^{n} \int_a^b \int_a^b K_X(\sigma,\tau) f_k(\sigma) f_k(\tau) d\sigma d\tau$$

Proof. By linearity of the expectation and by orthonormality of $\{f_k\}_k$ in $L_2[a,b]$

$$\mathcal{E}_n^2(\{f_k\}) = \int_a^b E[\varepsilon_n^2(t)]\mathrm{dt} = E\left[\int_a^b\left(X_t - \sum_{k=1}^n A_k f_k(t)\right)^2\mathrm{dt}\right] =$$
$$E\left[\int_a^b X_t^2\mathrm{dt}\right] - 2E\left[\int_a^b X_t\sum_{k=1}^n A_k f_k(t)\mathrm{dt}\right] + E\left[\int_a^b\left(\sum_{k=1}^n A_k f_k(t)\right)^2\mathrm{dt}\right] =$$
$$\int_a^b E[X_t^2]\mathrm{dt} - 2E\left[\int_a^b\sum_{k=1}^n A_k X_t f_k(t)\mathrm{dt}\right] + E\left[\int_a^b\sum_{i,j=1}^n A_i A_j f_i(t) f_j(t)\mathrm{dt}\right] =$$
$$\int_a^b E[X_t^2]\mathrm{dt} - 2E\left[\sum_{k=1}^n A_k\int_a^b X_t f_k(t)\mathrm{dt}\right] + E\left[\sum_{i,j=1}^n A_i A_j\int_a^b f_i(t) f_j(t)\mathrm{dt}\right] =$$
$$\int_a^b E[X_t^2]\mathrm{dt} - 2E\left[\sum_{k=1}^n A_k^2\right] + E\left[\sum_{i,j=1}^n A_i A_j\delta_{ij}\right] = \int_a^b E[X_t^2]\mathrm{dt} -$$
$$2E\left[\sum_{k=1}^n A_k^2\right] + E\left[\sum_{k=1}^n A_k^2\right] = \int_a^b E[X_t^2]\mathrm{dt} - E\left[\sum_{k=1}^n A_k^2\right] = \int_a^b E[X_t^2]\mathrm{dt} -$$
$$\sum_{k=1}^n E[A_k^2] = \int_a^b E[X_t^2]\mathrm{dt} - \sum_{k=1}^n E\left[\int_a^b X_\tau f_k(\tau)d\tau\int_a^b X_\sigma f_k(\sigma)d\sigma\right] =$$
$$\int_a^b E[X_t^2]\mathrm{dt} - \sum_{k=1}^n E\left[\int_a^b\int_a^b X_\tau X_\sigma f_k(\tau) f_k(\sigma)d\tau d\sigma\right] = \int_a^b E[X_t^2]\mathrm{dt} -$$
$$\sum_{k=1}^n\int_a^b\int_a^b E[X_\tau X_\sigma] f_k(\tau) f_k(\sigma)d\tau d\sigma =$$
$$\int_a^b E[X_t^2]\mathrm{dt} - \sum_{k=1}^n\int_a^b\int_a^b K_X(\sigma,\tau) f_k(\tau) f_k(\sigma)d\tau d\sigma$$

□

Theorem 4.13. $\mathcal{E}_n^2(\{f_k\})$ is minimized if and only if $\{f_k\}_{k=1}^n$ are an orthonormalization (via Gram-Schmidt for example) of the eigenfunctions of the Fredholm equation $\lambda\int_a^b K_X(\sigma,\tau)f(\tau)d\tau = f(\sigma)$ and $\{f_k\}_{k=1}^n$ are arranged in order to correspond to the eigenvalues $\{\lambda_k\}_{k=1}^n$ numbered according to decreasing magnitude: $\lambda_1 \geqslant \lambda_2 \geqslant \ldots. > 0$. Equivalently, this arranged, the Kahrunen-Loève basis minimizes the mean square truncation error.

Proof. By Lemma 4.12 $\mathcal{E}_n^2(\{f_k\}) = \int_a^b E[X_t^2]\mathrm{dt} - \sum_{k=1}^n\int_a^b\int_a^b K_X(\sigma,\tau) f_k(\sigma) f_k(\tau)d\sigma d\tau$ but since the first term is independent from the orthonormal system $\{f_k\}$, the minimum of $\mathcal{E}_n^2(\{f_k\})$ is attained with the maximum of

$$\sum_{k=1}^n\int_a^b\int_a^b K_X(\sigma,\tau) f_k(\sigma) f_k(\tau)d\sigma d\tau.$$

In order to minimize $\mathcal{E}_n^2(\{f_k\})$ with respect to the constraint $\int_a^b f_k(t) f_k(t)\mathrm{dt} = 1$, for every k, $\sum_{k=1}^n \int_a^b \int_a^b K_X(\sigma,\tau) f_k(\sigma) f_k(\tau) d\sigma d\tau$ should be maximized. We can therefore solve the equivalent problem of maximizing the quadratic integral form

$$J(f_k, f_k) = \int_a^b \int_a^b K_X(\sigma,\tau) f_k(\sigma) f_k(\tau) d\sigma d\tau$$

for every k, where $K_X(\sigma, \tau)$ is given continuous, simmetric, non-negative definite and $\int_a^b f_k(t) f_k(t)\mathrm{dt} = 1$. It has already been shown in the Theorem 3.7 that the maximum value of $J(f_k, f_k)$ is λ_1, where λ_1 is the eigenvalue of greatest magnitude of the kernel $K_X(\sigma,\tau)$ of the integral equation $\lambda\int_a^b K_X(\sigma,\tau) f(\tau) d\tau = f(\sigma)$, and the function, which attains the maximum, is the corresponding normalized eigenfunction f_1. Since the kernel $K_X(\sigma,\tau)$ is non-negative definite, the eigenvalue is positive. In fact bilinear integral forms non-negative definite have strictly positive eigenvalues, as a result of the *Hilbert's basic formula.* This argument can be reviewed in Tricomi (1957).

Now, if the eigenvalues of K_X are numbered in non increasing order, so that $\lambda_1 \geqslant \lambda_2 \geqslant \dots. > 0$, then it is clear that $\sum_{k=1}^n \int_a^b \int_a^b K_X(\sigma, \tau) f_k(\sigma) f_k(\tau) d\sigma d\tau$ shall be maximized by chosing the corresponding first n orthonormalized eigenfunctions $f_1, \dots, f_n$. When these first n eigenvalues are distinct, a unique solution is provided and

$$\mathcal{E}_n^2(\{f_k\}) = \int_a^b E[X_t^2]\mathrm{dt} - \sum_{k=1}^n \int_a^b \int_a^b K_X(\sigma,\tau) f_k(\sigma) f_k(\tau) d\sigma d\tau = \int_a^b E[X_t^2]\mathrm{dt} - \sum_{k=1}^n \lambda_k$$

is the smallest $n-\mathrm{th}$ mean square error of truncation.

However, if there are indistinct eigenvalues, namely there is an eigenvalue $\tilde{\lambda}$ which has more than one corresponding eigenfunctions (say a finite number k by Remark 3.3: we say the *index of the eigenvalue* is k; let be $1 \leqslant p \leqslant n$ such that $\tilde{\lambda} = \lambda_p = \dots = \lambda_{p+k-1}$ if $n > p+k-1$ or $\tilde{\lambda} = \lambda_p = \dots = \lambda_n$ if $n \leqslant p+k-1$), these k eigenfunctions are not uniquely determined. Anyway, any two competing sets of orthonormal solutions for $\tilde{\lambda}$, $\{\varphi_1, \dots, \varphi_k\}$ and $\{\psi_1, \dots, \psi_k\}$, are related by an orthogonal transformation, that is there exists an orthogonal $k\times k$ matrix $A = (a_{ij})$ (i.e. $\sum_{j=1}^k a_{ij} a_{mj} = \delta_{im}$) such that $\phi_i(t) = \sum_{j=1}^k a_{ij}\psi_j(t)$. Indeed, it is trivial that if we have two orthonormal basis, the matrix of the basis change is orthogonal. Then

$$\sum_{k=1}^n \int_a^b \int_a^b K_X(\sigma, \tau) f_k(\sigma) f_k(\tau) d\sigma d\tau = \sum_{k=1}^{p-1} \int_a^b \int_a^b K_X(\sigma, \tau) f_k(\sigma) f_k(\tau) d\sigma d\tau + \sum_{k=p}^{p+k-1} \int_a^b \int_a^b K_X(\sigma, \tau) \varphi_k(\sigma) \varphi_k(\tau) d\sigma d\tau + \sum_{k=p+k}^{n} \int_a^b \int_a^b K_X(\sigma, \tau) f_k(\sigma) f_k(\tau) d\sigma d\tau = \sum_{k=1}^{p-1} \lambda_k + \sum_{k=p}^{p+k-1} \int_a^b \int_a^b K_X(\sigma,\tau) \sum_{j=1}^k a_{kj}\psi_j(\sigma) \sum_{j=1}^k a_{kj}\psi_j(\tau) d\sigma d\tau + \sum_{k=p+k}^n \lambda_k$$

but

$$\sum_{k=p}^{p+k-1}\int_a^b\int_a^b K_X(\sigma,\tau)\sum_{j=1}^{k}a_{kj}\psi_j(\sigma)\sum_{j=1}^{k}a_{kj}\psi_j(\tau)d\sigma d\tau=$$

$$\sum_{k=p}^{p+k-1}\int_a^b\int_a^b K_X(\sigma,\tau)\sum_{j,i=1}^{k}a_{kj}a_{ki}\psi_j(\sigma)\psi_i(\tau)d\sigma d\tau=$$

$$\sum_{k=p}^{p+k-1}\sum_{j,i=1}^{k}a_{kj}a_{ki}\int_a^b\int_a^b K_X(\sigma,\tau)\psi_j(\sigma)\psi_i(\tau)d\sigma d\tau=$$

$$\sum_{k=p}^{p+k-1}\sum_{j,i=1}^{k}a_{kj}a_{ki}\int_a^b\psi_i(\tau)\int_a^b K_X(\sigma,\tau)\psi_j(\sigma)d\sigma d\tau=$$

$$\tilde{\lambda}\sum_{k=p}^{p+k-1}\sum_{j,i=1}^{k}a_{kj}a_{ki}\int_a^b\psi_i(\tau)\psi_j(\tau)d\tau = \tilde{\lambda}\sum_{k=p}^{p+k-1}\sum_{j,i=1}^{k}a_{kj}a_{ki}\delta_{ij} = \tilde{\lambda}\sum_{k=p}^{p+k-1}\sum_{j=1}^{k}a_{kj}^2 =$$

$$\tilde{\lambda}\sum_{k=p}^{p+k-1}\delta_{kk}=k\tilde{\lambda}=\lambda_p+\ldots+\lambda_{p+k-1}.$$

Thus

$$\sum_{k=1}^{n}\int_a^b\int_a^b K_X(\sigma,\tau)f_k(\sigma)f_k(\tau)d\sigma d\tau=$$

$$\sum_{k=1}^{p-1}\lambda_k+\sum_{k=p}^{p+k-1}\int_a^b\int_a^b K_X(\sigma,\tau)\sum_{j=1}^{k}a_{kj}\psi_j(\sigma)\sum_{j=1}^{k}a_{kj}\psi_j(\tau)d\sigma d\tau+\sum_{k=p+k}^{n}\lambda_k=$$

$$\sum_{k=1}^{n}\lambda_k$$

Hence it turns out that $\mathcal{E}_n^2(\{f_k\})$ is minimized as well. □

5 Analytical applications to Gaussian processes.

The Karhunen-Loève Expansion assumes a special form when the process X is Gaussian. We therefore briefly recall the notions of Gaussian random variable and Gaussian random vector in order to introduce the Gaussian processes.

5.1 Basics of Gaussian processes.

A real-valued *Gaussian random variable* X is completely defined by the following density of its *distribution function*:

$$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

where $m=\int_{\mathbb{R}} x f(x)\,dx$ is the expectation $E[X]$ and $\sigma^2=\int_{\mathbb{R}}(x-m)^2 f(x)\,dx$ is the variance var$[X]$. We usually denote that X is Gaussian distributed as $X\sim\mathcal{N}(m,\sigma^2)$. The *characteristic function*, which uniquely determines the density, turns to be

$$\varphi(\lambda):=E[\exp(i\lambda X)]=\int_{\mathbb{R}}\exp(i\lambda x)f(x)dx=\exp\left(im\lambda-\frac{1}{2}\lambda^2\sigma^2\right)$$

Similarly, a real-valued *Gaussian random vector* $X=(X_1,X_2,\ldots,X_n)$ is completely defined by the following joint density:

$$f(x)=\frac{1}{\sqrt{(2\pi)^n\det\Sigma}}\exp\left(-\frac{1}{2}(x-m)^t\Sigma^{-1}(x-m)\right)$$

where $x=(x_1,x_2,\ldots,x_n)^t$, $m=(m_1,m_2,\ldots,m_n)^t$, $m_i=E[X_i]$ for every $i=1,\ldots,n$. The *covariance matrix* Σ is a symmetric positive-definite matrix

$$\Sigma=\Sigma^t=\begin{pmatrix}\sigma_1^2\,\sigma_{12}\ldots\sigma_{1n}\\ \sigma_{21}\,\sigma_2^2\ldots\sigma_{2n}\\ \\ \sigma_{n1}\,\sigma_{n2}\ldots\sigma_n^2\end{pmatrix}\text{where }\sigma_i^2=E[(X_i-m_i)^2]\text{ and }\sigma_{ij}=E[(X_i-m_i)(X_j-m_j)].$$

For every $a,b\in\{1,\ldots,n\}$ if we block out

$$X=\begin{pmatrix}X_a\\ X_b\end{pmatrix},m=\begin{pmatrix}m_a\\ m_b\end{pmatrix},\Sigma=\begin{pmatrix}\Sigma_a\,\Sigma_{ab}\\ \Sigma_{ba}\,\Sigma_b\end{pmatrix}$$

then for every $i\in\{a,b\}$, X_i is a *multivariate Gaussian* with expectation vector m_i and covariance matrix Σ_i.

The characteristic function turns to be

$$\varphi(\lambda)=\exp\left(i\lambda^t m-\frac{1}{2}\lambda^t\Sigma\lambda\right)$$

where $\lambda=(\lambda_1,\lambda_2,\ldots,\lambda_n)$.

Remark 5.1. Linear transformations preserve the Gaussian property of random vectors.

Proof. If X is a Gaussian random vector and A is an appropriate matrix, $Y = AX$ turns to be a Gaussian vector as well, since

$$\lambda^t Y = \lambda^t A X = A^t \lambda X = \mu^t X$$

where $\mu = A^t \lambda$, which implies that, up to a change of variable, the characteristic function of Y turns to be the same of X:

$$\varphi_Y(\lambda) = \varphi_X(\mu). \qquad \square$$

Definition 5.2. $\{X_t, t \in T\}$ is said to be a Gaussian process if it is a stochastic process such that for all positive integers k and all choices of $\{t_1, \ldots, t_k\} \subset T,$ the random variables $X_{t_1}, X_{t_2}, \ldots, X_{t_k}$ form a Gaussian random vector, which means that are jointly gaussian.

Proposition 5.3. If jointly Gaussian random variables are orthogonal, then they are independent.

Proof. Let $X = (X_1, X_2, \ldots, X_n)$ be a Gaussian vector of orthogonal components, that is $E[X_i X_j] = 0$ for every $i \neq j$. Thus

$$\Sigma = \begin{pmatrix} \sigma_1^2 0 \ldots 0 \\ 0 \sigma_2^2 \ldots 0 \\ \\ 0 \ldots 0 \sigma_n^2 \end{pmatrix}$$

$$\varphi_X(\lambda) = \exp\left(i\lambda^t m - \frac{1}{2}\lambda^t \Sigma \lambda\right) = \exp\left(i\sum_{j=1}^n m_j \lambda_j - \frac{1}{2}\sum_{j=1}^n \sigma_j^2 \lambda_j^2\right) =$$

$$\prod_{j=1}^n \exp\left(i m_j \lambda_j - \frac{1}{2}\sigma_j^2 \lambda_j^2\right) = \prod_{j=1}^n \varphi_{Xj}(\lambda_j).$$

What we have shown is that the joint characteristic function φ_X of the Gaussian variables X_1, X_2, ..., X_n is actually the product of their marginal characteristic functions φ_{X_j}. This fact is well known to be equivalent to the independence of X_1, $X_2, \ldots, X_n$. $\square$

5.2 Expansion and convergence of Gaussian processes

Theorem 5.4. If $\{X_t, t \in [a, b]\}$ is a zero mean Gaussian process, then the Karhunen-Loève Expansion projections Z_k are independent Gaussian random variables.

Proof. By the Karhunen-Loève Theorem, given $\Delta: a = t_1 \leqslant t_2 \leqslant \ldots \leqslant t_n = b$, for every positive integer k

$$Z_k = \int_a^b X_t e_k(t) dt := \lim_{|\Delta| \to 0} \sum_{i=1}^n X_{t_i} e_k(t_i)(t_i - t_{i-1}) = \lim_{|\Delta| \to 0} I_k(\Delta).$$

Since $\{X_t\}$ is Gaussian, $\{X_{t_1},\ X_{t_2},\ \ldots,\ X_{t_n}\}$ are jointly Gaussian. If we take as coefficients of a linear combination $\alpha_{ki} = e_k(t_i)(t_i - t_{i-1})$, we have that $I_k(\Delta) = \sum_{i=1}^{n} \alpha_{ki} X_{t_i}$. As a consequence for every set of indexes $\{k_1, k_2, \ldots, k_n\}$ we have that

$$\begin{pmatrix} I_{k_1}(\Delta) \\ I_{k_2}(\Delta) \\ \vdots \\ I_{k_n}(\Delta) \end{pmatrix} = \begin{pmatrix} \alpha_{k_1 1}\alpha_{k_1 2}\ldots\alpha_{k_1 n} \\ \alpha_{k_2 1}\,\alpha_{k_2 2}\ldots\alpha_{k_2 n} \\ \\ \alpha_{k_n 1}\alpha_{k_n 2}\ldots\alpha_{k_n n} \end{pmatrix} \begin{pmatrix} X_{t_1} \\ X_{t_2} \\ \vdots \\ X_{t_n} \end{pmatrix}$$

thus by Remark 5.1 $\{I_k(\Delta)\}$ is a Gaussian sequence (i.e. a discrete Gaussian process), since every finite set of samples forms jointly a random Gaussian vector. We can therefore write the joint characteristic function of the random vector $I(\Delta) = (I_{k_1}(\Delta), I_{k_2}(\Delta), \ldots, I_{k_n}(\Delta))$. The calculations are greatly simplified by the hypothesis of having zero mean variables X_{t_i}. In fact, by linearity of the expectation, the whole random vector I will have zero mean. In addition, the covariance matrix will result very simple as well: $(E[I_{k_j}(\Delta) I_{k_m}(\Delta)])_{1 \leqslant j,m \leqslant n}$. The characteristic function is therefore

$$\varphi_{I(\Delta)}(\lambda_1, \ldots, \lambda_n) = \exp\left(-\frac{1}{2} \sum_{j,m=1}^{n} \lambda_j E[I_{k_j}(\Delta) I_{k_m}(\Delta)] \lambda_m\right).$$

Since $I_k(\Delta)$ converges in mean square to Z_k as $|\Delta| \longrightarrow 0$ for every k, we have that

$$E[I_{k_j}(\Delta) I_{k_m}(\Delta)] \underset{|\Delta| \to 0}{\longrightarrow} E[Z_{k_j} Z_{k_m}].$$

Indeed

$$|E[I_{k_j}(\Delta) I_{k_m}(\Delta) - Z_{k_j} Z_{k_m}]| = |E[I_{k_j}(\Delta) I_{k_m}(\Delta) + I_{k_j}(\Delta) Z_{k_m} - I_{k_j}(\Delta) Z_{k_m} - Z_{k_j} Z_{k_m}]| \leqslant E[|I_{k_j}(\Delta)[I_{k_m}(\Delta) - Z_{k_m}]|] + E[|Z_{k_m}[I_{k_j}(\Delta) - Z_{k_j}]|] \leqslant \{E[|I_{k_j}(\Delta)|^2] E[|I_{k_m}(\Delta) - Z_{k_m}|^2]\}^{\frac{1}{2}} + \{E[|Z_{k_m}|^2] E[|I_{k_j}(\Delta) - Z_{k_j}|^2]\}^{\frac{1}{2}} \underset{|\Delta| \to 0}{\longrightarrow} 0$$

because firstly

$$E[|I_{k_j}(\Delta)|^2]^{\frac{1}{2}} \underset{|\Delta| \to 0}{\longrightarrow} E[|Z_{k_j}|^2]^{\frac{1}{2}} < \infty$$

given the fact that $I_{k_j}(\Delta) \in L^2(\Omega)$ and therefore $\lim_{|\Delta| \to 0} E[|I_{k_j}(\Delta)|^2] = E[|\lim_{|\Delta| \to 0} I_{k_j}(\Delta)|^2]$; secondly by hypothesis we know that

$$E[|I_{k_m}(\Delta) - Z_{k_m}|^2] \longrightarrow 0 \text{ and } E[|I_{k_j}(\Delta) - Z_{k_j}|^2] \longrightarrow 0 \text{ as } |\Delta| \longrightarrow 0.$$

As a consequence of the fact just proven, we have that, by continuity

$$\varphi_{I(\Delta)} = \exp\left(-\frac{1}{2} \sum_{j,m=1}^{n} \lambda_j E[I_{k_j}(\Delta) I_{k_m}(\Delta)] \lambda_m\right) \underset{|\Delta| \to 0}{\longrightarrow} \exp\left(-\frac{1}{2} \sum_{j,m=1}^{n} \lambda_j E[Z_{k_j} Z_{k_m}] \lambda_m\right).$$

Now, define the random vector $Z = (Z_{k_1}, \ldots, Z_{k_n})$. By construction $I(\Delta) \longrightarrow$ Z as $|\Delta| \to 0$; therefore by continuity $\varphi_{I(\Delta)} \longrightarrow \varphi_Z$, as $|\Delta| \to 0$.

As a result

$$\varphi_Z(\lambda_1, \ldots, \lambda_n) = \exp\left(-\frac{1}{2} \sum_{j,m=1}^{n} \lambda_j E[Z_{k_j} Z_{k_m}] \lambda_m\right),$$

which means that $\{Z_{k_1}, Z_{k_2}, ..., Z_{k_n}\}$are jointly Gaussian, because by the Karhunen-Loève Theorem they are centered. By the arbitrariness of the indices, we have therefore obtained that $\{Z_k\}$ is a Gaussian sequence. Now, by the Karhunen-Loève Theorem we also know that $\{Z_k\}$ are orthogonal. By the Proposition 5.3 we conclude that they are independent. □

We need a few results to achieve more information about the convergence of a Gaussian process expansion in its probability space $(\Omega, \mathcal{F}, P)$. The most iportant are Etemadi's inequality and Theorem 5.10, both read in Billingsley (1995), pp. 288-290.

Definition 5.5. A sequence of real valued random variables $\{S_n\}$ is said to converge to a random variable S *almost surely* (or *with probability one*) when

$$P\Big(\Big\{\omega \in \Omega, \lim_{n\to\infty} S_n(\omega) = S(\omega)\Big\}\Big) = 1, \text{shortly written as } P\left(\lim_{n\to\infty} S_n = S\right) = 1.$$

Definition 5.6. A sequence of real valued random variables $\{S_n\}$ is said to converge to a random variable S *in probability* when for every $\varepsilon > 0$ there is an $n \in \mathbb{N}$ such that

$$P\Big(\Big\{\omega \in \Omega, \Big|S_n(\omega) - S(\omega)\Big| \geqslant \varepsilon\Big\}\Big) = 0, \text{shortly written as } P(|S_n - S| \geqslant \varepsilon) = 0$$

Proposition 5.7. The *mean square* convergence of a real valued random variables sequence $\{S_n\}$ implies its convergence *in probability.*

Proof. We recall the elementary fact that for any event $A \subset \Omega, P(A) = E[\mathbf{1}_A]$, where $\mathbf{1}_A$ is the *indicator function* of the event A. Therefore

$$P\Big(\Big|S_n - S\Big| \geqslant \varepsilon\Big) = E\Big[\mathbf{1}_{\left(\left|S_n - S\right| \geqslant \varepsilon\right)}\Big]$$

and since on the event $\Big(\Big|S_n - S\Big| \geqslant \varepsilon\Big)$ we have trivailly that $\dfrac{\left|S_n - S\right|^2}{\varepsilon^2} \geqslant 1$, then

$$E\Big[\mathbf{1}_{\left(\left|S_n - S\right| \geqslant \varepsilon\right)}\Big] \leqslant E\left[\frac{\Big|S_n - S\Big|^2}{\varepsilon^2}\mathbf{1}_{\left(\left|S_n - S\right| \geqslant \varepsilon\right)}\right] = \frac{1}{\varepsilon^2}E\Big[\Big|S_n - S\Big|^2\mathbf{1}_{\left(\left|S_n(\omega) - S(\omega)\right| \geqslant \varepsilon\right)}\Big] \leqslant \frac{1}{\varepsilon^2}E\Big[\Big|S_n - S\Big|^2\Big]$$

because $\Big|S_n - S\Big|^2 \geqslant 0$ for every $\omega \in \Omega$. Since by hypothesis of mean square convergence $\lim_{n\to\infty} E\Big[\Big|S_n - S\Big|^2\Big] = 0$, the result follows. □

Remark 5.8. Whereas the *almost surely* convergence implies the convergence *in probability*, the *viceversa* is not always true. However series of independent variables make an exception.

Theorem 5.9. (Etemadi's inequality) Suppose that $\{Z_1, ..., Z_n\}$ are independent. For $\alpha \geqslant 0$, given $S_k = \sum_{i=1}^{k} Z_i$ with $k \leqslant n$,

$$P\Big(\max_{1\leqslant k\leqslant n} |S_k| \geqslant 3\alpha\Big) \leqslant 3 \max_{1\leqslant k\leqslant n} P(|S_k| \geqslant \alpha)$$

Proof. Let B_k be the set where $|S_k| \geqslant 3\alpha$ but $|S_j| < 3\alpha$ for $j < k$. Since the B_k are disjoint,

$P\Big(\max_{1\leqslant k\leqslant n} |S_k| \geqslant 3\alpha\Big) \leqslant P(|S_n| \geqslant \alpha) + \sum_{k=1}^{n-1} P(B_k \cap (|S_n| < \alpha)) \leqslant P(|S_n| \geqslant \alpha) + \sum_{k=1}^{n-1} P(B_k \cap (|S_n - S_k| > 2\alpha)) = P(|S_n| \geqslant \alpha) + \sum_{k=1}^{n-1} P(B_k) P(|S_n - S_k| > 2\alpha) \leqslant P(|S_n| \geqslant \alpha) + \max_{1\leqslant k\leqslant n} P(|S_n - S_k| > 2\alpha) \leqslant P(|S_n| \geqslant \alpha) + \max_{1\leqslant k\leqslant n} (P(|S_n| \geqslant \alpha) + P(|S_k| \geqslant \alpha)) \leqslant 3 \max_{1\leqslant k\leqslant n} P(|S_k| \geqslant \alpha)$. □

Theorem 5.10. If $\{Z_j\}$ is a sequence of independent real valued random variables, then the convergence of the series $\sum_{j=1}^{\infty} Z_j$ *in probability* implies its convergence *almost surely.*

Proof. It is enough to prove that if $S_n = \sum_{i=1}^{n} Z_i$ converges *in probability* to S, then $\{S_n\}$ is a Cauchy sequence with probability one. Since

$$P(|S_{n+j} - S_n| \geqslant \varepsilon) \leqslant P\Big(|S_{n+j} - S| \geqslant \frac{\varepsilon}{2}\Big) + P\Big(|S_n - S| \geqslant \frac{\varepsilon}{2}\Big)$$

the hypothesis of convergence in probability, which yields $\lim_{n\to\infty} P(|S_n - S| \geqslant \varepsilon) = 0$, obtains that

$$\lim_{n\to\infty} \sup_{j\geqslant 1} P(|S_{n+j} - S_n| \geqslant \varepsilon) = 0.$$

But by the Etemadi's inequality

$$P\Big(\max_{1\leqslant j\leqslant k} |S_{n+j} - S_n| \geqslant \varepsilon\Big) \leqslant 3 \max_{1\leqslant j\leqslant k} P\Big(|S_{n+j} - S_n| \geqslant \frac{\varepsilon}{3}\Big),$$

therefore

$$P\Big(\sup_{k\geqslant 1} |S_{n+k} - S_n| > \varepsilon\Big) \leqslant 3 \sup_{k\geqslant 1} P\Big(|S_{n+k} - S_n| \geqslant \frac{\varepsilon}{3}\Big).$$

Since by hypothesis $\lim_{n\to\infty} \sup_{j\geqslant 1} P(|S_{n+j} - S_n| \geqslant \varepsilon) = 0$, then

$$\lim_{n\to\infty} P\Big(\sup_{k\geqslant 1} |S_{n+k} - S_n| > \varepsilon\Big) = 0$$

for each $\varepsilon > 0$.

Let now $E_{n,\varepsilon}$ be the event where $\sup_{j,k\geqslant n}|S_j-S_k|>2\varepsilon$, and put

$$E_\varepsilon := \bigcap_{\varepsilon>0} E_{n,\varepsilon}.$$

Then $E_{n,\varepsilon} \searrow E_\varepsilon$ as $n \longrightarrow \infty$, and since $\lim_{n\to\infty} P(\sup_{k\geqslant 1}|S_{n+k}-S_n|>\varepsilon)=0$, the *continuity from above* of the probability measure obtains $P(E_\varepsilon)=0$. Now, the union over the positive rationals

$$\bigcup_{q\in\mathbb{Q}^+} E_q$$

contains the set where the sequence $\{S_n\}$ is not Cauchy. By countability of $\mathbb{Q}$,

$$P\left(\bigcup_{q\in\mathbb{Q}^+} E_q\right) = \sum_{q\in\mathbb{Q}^+} P(E_q) = 0$$

and therefore, passing to the complementary, the set over which $\{S_n\}$ is Cauchy has probability one. □

Corollary 5.11. For each fixed $t\in[a,b]$, $\sum_{k=1}^{\infty} Z_k e_k(t)$ converges to X_t *almost surely.*

Proof. We have shown that the $\{Z_k\}$ are independent, therefore so are $\{Z_k e_k(t)\}$, since the eigenfunctions are deterministic. Thus the Theorem 5.10 can be applyed to the series of independent random variables $\sum_{k=1}^{\infty} Z_k e_k(t)$, that is its convergence *in probability* shall obtains the convergence *almost surely*. We therefore need to prove only the convergence *in probability*. The argument is actually straightforward, since by the Proposition 5.7 the convergence *in mean square* implies the convergence *in probability*, and the mean square convergence of the partial sums $S_n=\sum_{k=1}^{n} Z_k e_k(t)$ is actually a result of the Karhunen-Loève Theorem. □

As notable examples to test the Karhunen-Loève Expansion on, we chose two significant Gaussian processes, the Brownian motion and the Brownian bridge. We have found some help in deriving their expansion from Ash (1990) and Wang (2008). We start with a definition of Brownian motion introduced according to Loève (1978).

5.3 The Brownian motion

The ceaseless and erratic dance of microscopic particles suspended in a fluid is called *Brownian motion*, after the botanist Brown, who first systematically investigated it. Today we know that this motion is due to the bombardament of the particles by the molecules of the medium. In a liquid, under normal conditions, the order of magnitude of the figures of these impacts is of 10^{20} per second. In 1905 kinetic molecular theory led Einstein to the first mathematical model of Brownian motion: the heat partial differential equation for the probability density that the particle be at a particular position at a particular time. However rigorous definition and study of mathematical Brownian motion requires measure theory. Some 20 years after Lebesgue's theory, Wiener (1923) gave its first satisfactory construction. We shall focus here on the basic case, the one-dimensional Brownian motion modelled as a stochastic process $\{W_t, t\in T\subset\mathbb{R}\}$, where the W refers to Wiener.

There are several definitions of the (one-dimensional) Brownian process. As Loève (1978), pp. 235-236, we proceed by successive refinements, based on required or relevant properties. A process $\{W_t\}$ is Brownian distributed if it is *decomposable*, that is its increments $W_{st} := W_t - W_s$, where $s < t$ for $s, t \in T$, are independents on disjointed intervals (s, t), and $W_{st} \sim \mathcal{N}(\alpha(t-s), \sigma^2(t-s))$, where α is called the *drift* and $\sigma^2 > 0$ is called the *diffusion coefficient* of the process. A more restrictive definition obtains by adding the requirements $T = [0, \infty)$ and $W_0 = 0$ *almost surely*. At this point, with no real loss, from now on we restrict *Brownian distributed* to the normalized form obtained taking $\alpha = 0$ and $\sigma^2 = 1$.

Definition 5.12. A random process $\{W_t, t \geqslant 0\}$, where $W_0 = 0$ *almost surely*, is *Brownian distributed* if its disjoint increments W_{st} are independent and centered normal: $W_{st} \sim \mathcal{N}(0, t-s)$.

Proposition 5.13. If a stocastic process is Brownian distributed then it is a Gaussian process such that $W_t \sim \mathcal{N}(0, t)$ and $E[W_s W_t] = \min(s, t)$.

Proof. Let $\{W_t\}$ be Brownian distributed. We have to prove that any finite selection of the process is jointly Gaussian. Fix any finite sampling $\{W_{t_1}, W_{t_2}, \ldots, W_{t_m}\}$ with $t_1 < t_2 < \ldots < t_m$. Denote $u = (u_1, \ldots, u_m)$. By *decomposability* the joint characteristic function of $W = (W_{t_1}, W_{t_2}, \ldots, W_{t_m})$ is

$$\varphi_W(u) = E[\exp(i\,u_1 W_{t_1} + \ldots + i\,u_m W_{t_m}) = E[\exp(i\,u_1 W_{t_1} + \ldots + i(u_{m-1} + u_m) W_{t_{m-1}} + i\,u_m W_{t_{m-1} t_m})] = \ldots = E[\exp(i(u_1 + \ldots + u_m) W_{t_1 t_2} + \ldots + i\,u_m W_{t_{m-1} t_m})] = E[\exp(i\lambda_1 W_{t_1 t_2} + \ldots + i\,\lambda_m W_{t_{m-1} t_m})] = \varphi_{\mathcal{W}}(\lambda) = \varphi_{W_{t_1 t_2}}(\lambda_1) \ldots \varphi_{W_{t_{m-1} t_m}}(\lambda_{m-1})$$

where $\lambda_k = u_k + u_{k+1} + \ldots + u_m$, $\lambda = (\lambda_1, \ldots, \lambda_{m-1})$ and $\mathcal{W} = (W_{t_1 t_2}, \ldots, W_{t_{m-1} t_m})$. The independent increments $W_{t_k t_{k+1}}$ being Guassian by definition, so is $\mathcal{W}$ indeed. As a result W has a multivariate Gaussian distribution. Since any finite sampling of the process is jointly Gaussian, we have that $\{W_t\}$ is a Gaussian process.

We now find its covariance. Given $0 \leqslant s \leqslant t$,

$$E[W_s W_t] = E[W_s((W_t - W_s) + W_s)] = E[W_s(W_t - W_s)] + E[W_s^2] = E[(W_s - W_0)(W_t - W_s)] + E[W_s^2] = E[W_s - W_0] E[W_t - W_s] + E[W_s^2] = E[W_s] E[W_t - W_s] + E[W_s^2] = 0 + E[W_s^2] = E[W_s^2] = E[(W_s - 0)^2] = E[(W_s - E[W_s])^2] = \mathrm{var}[W_s] = s = \min(s, t). \quad \square$$

We can now expand by the Karhuen-Loève Theorem the Brownian motion over a finite interval [0, 1]. By the Proposition 5.13 we know that the covariance is $K_W(s, t) = \min(s, t)$, $s, t \in [0, 1]$. This allow us to apply the Karhunen-Loève Expansion as follows.

Theorem 5.14. The Karhunen-Loève Expansion of the Brownian motion on $[0, 1]$ is

$$W_t = \frac{2\sqrt{2}}{\pi} \sum_{k=1}^{\infty} \frac{Z_k^*}{2k-1} \sin\left[\frac{(2k-1)\pi}{2} t\right]$$

where $Z_k = \int_0^1 W_t \sqrt{2} \sin\left[\frac{(2k-1)\pi}{2} t\right] dt$, $\lambda_k = \frac{4}{(2k-1)^2 \pi^2}$, $k \in \mathbb{N}$, $Z_k^* = \frac{Z_k}{\sqrt{\lambda_k}}$ and the convergence in mean square is almost surely.

Proof. We have first to find the eigenvalues of $(Af)(s) = \int_0^1 K_W(s,t) f(t) dt$.

$$\int_0^1 \min(s,t) e(t) dt = \lambda e(s)$$

$$\int_0^s t e(t) dt + s \int_s^1 e(t)\, dt = \lambda e(s)$$

Differentiating with respect to s both of the sides we obtain

$$s e(s) + \int_s^1 e(t)\, dt - s e(s) = \lambda e'(s)$$

$$\int_s^1 e(t)\, dt = \lambda e'(s)$$

Differentiating again we obtain

$$-e(s) = \lambda e''(s)$$

its characteristic polynomial is $x^2\ \lambda + 1 = 0$ which has roots $x_{1,2} = \frac{\pm i}{\sqrt{\lambda}}$ and the solutions are therefore

$$e(s) = a \sin \frac{s}{\sqrt{\lambda}} + b \cos \frac{s}{\sqrt{\lambda}}$$

and if $\lambda = 0$ then $e(s) = \lambda e''(s) \equiv 0$, therefore 0 is not an eigenvalue and we can proceed to determine a, b and λ.

For $s = 0$ we have $e(0) = b$, but

$$e(0) = \frac{1}{\lambda} \int_0^1 \min(0,t) e(t) dt = 0.$$

Thus $b = 0$ and $e(s) = a \sin \frac{s}{\sqrt{\lambda}}$. For $s = 1$, we have

$$e'(1) = \frac{1}{\lambda} \int_1^1 e(t)\, dt = 0.$$

Since

$$0 = e'(1) = \left[\frac{d}{ds} e(s)\right]_{s=1} = \left[\frac{a}{\sqrt{\lambda}} \cos \frac{s}{\sqrt{\lambda}}\right]_{s=1} = \frac{a}{\sqrt{\lambda}} \cos \frac{1}{\sqrt{\lambda}}$$

we obtain that if there is a non-trivial solution ($a \neq 0$), $\cos \frac{1}{\sqrt{\lambda}} = 0$ must yield. Namely,

$$\frac{1}{\sqrt{\lambda}} = -\frac{\pi}{2} + k\pi = \pi \left(\frac{2k-1}{2}\right), k \in \mathbb{N}.$$

The positive eigenvalues are therefore $\lambda_k = \frac{4}{(2k-1)^2 \pi^2}$, $k \in \mathbb{N}$ and the eigenfunctions are $e_k(s) = a \sin\left[\frac{(2k-1)\pi}{2} s\right]$ where a must be a normalization constant. Then

$$\frac{1}{a^2} = \left\|\sin\left[\frac{(2k-1)\pi}{2} s\right]\right\|_2^2 = \int_0^1 \sin^2\left[\frac{(2k-1)\pi}{2} s\right] dt = \frac{2}{(2k-1)\pi} \int_0^{\frac{(2k-1)\pi}{2}} \sin^2 z\, dz =$$

$$\frac{2}{(2k-1)\pi} \int_0^{\frac{(2k-1)\pi}{2}} \frac{1 - \cos(2z)}{2}\, dz = \frac{1}{(2k-1)\pi} \int_0^{(2k-1)\pi} \frac{1 - \cos(u)}{2}\, du =$$

$$\frac{1}{2(2k-1)\pi} \int_0^{(2k-1)\pi} 1 - \cos u\, du = \frac{1}{2(2k-1)\pi} [(2k-1)\pi - \sin((2k-1)\pi)] = \frac{1}{2}$$

therefore $a = \left\|\sin\left[\frac{(2k-1)\pi}{2}s\right]\right\|_2^{-1} = \sqrt{2}$.

The whole arguent is invertible, by applying A to $e_k(s) = \sqrt{2}\sin\left[\frac{(2k-1)\pi}{2}s\right]$ we obtain

$$(A\,e_k)(s) = \int_0^s t\sqrt{2}\sin\left[\frac{(2k-1)\pi}{2}t\right]d\,t + \sqrt{2}s\int_s^1 \sin\left[\frac{(2k-1)\pi}{2}t\right]d\,t = \frac{4\sqrt{2}}{(2k-1)^2\pi^2}\int_0^{\frac{(2k-1)\pi}{2}s} z\ \sin\ z\,d\,z + \frac{2\sqrt{2}s}{(2k-1)\pi}\int_{s\frac{(2k-1)\pi}{2}}^{\frac{(2k-1)\pi}{2}} \sin\ z\,d\,z = \frac{4\sqrt{2}}{(2k-1)^2\pi^2}\left\{- z\ \cos\ z\Big|_0^{\frac{(2k-1)\pi}{2}s} + \int_0^{\frac{(2k-1)\pi}{2}s}\cos\ z\,d\,z\right\} - \frac{2\sqrt{2}s}{(2k-1)\pi}\cos\ z\Big|_{s\frac{(2k-1)\pi}{2}}^{\frac{(2k-1)\pi}{2}} = -\frac{2\sqrt{2}s}{(2k-1)\pi}\cos\left[\frac{(2k-1)\pi}{2}s\right] + \frac{4\sqrt{2}}{(2k-1)^2\pi^2}\sin\left[\frac{(2k-1)\pi}{2}s\right] + \frac{2\sqrt{2}s}{(2k-1)\pi}\cos\left[\frac{(2k-1)\pi}{2}s\right] = \frac{4\sqrt{2}}{(2k-1)^2\pi^2}\sin\left[\frac{(2k-1)\pi}{2}s\right] = \lambda_k e(s).$$

Eventually, according to the Karhuen-Loève Expansion, given

$$Z_k = \int_0^1 W_t\sqrt{2}\sin\left[\frac{(2k-1)\pi}{2}t\right]dt$$

we have that

$$W_t = \frac{2\sqrt{2}}{\pi}\sum_{k=1}^{\infty}\frac{Z_k^*}{2k-1}\sin\left[\frac{(2k-1)\pi}{2}t\right]$$

where, being $\{W_t\}$ a Gaussian process, by Remark 4.7 and Theorem 5.4 it follows that the random variables $\frac{Z_k}{\sqrt{\lambda_k}} = \frac{(2k-1)\pi}{2}Z_k = Z_k^* \sim \mathcal{N}(0,\,1)$ are independent. By Corollary 5.11 the convergence of the expansion is predicted not only to be in mean square with probability one. □

5.4 The Brownian bridge

The *Brownian bridge* can be derived from the *Brownian motion* $\{W_t, t\in[0,1]\}$ by conditioning $W_1 = 0$. It is therefore a Brownian process tied down at both ends. Hence the evocative name *Brownian bridge*: supported at both ends of the interval by the conditioning, just as a bridge by pylons. It results particularly useful in the Brownian interpolation of generated points by a Brownian motion simulation. An equivalent definition, more close to our aims, is the following:

Definition 5.15. The Brownian bridge is the process $\{B_t,\ t\in[0,\ 1]\}$ such that $B_t = W_t - \mathrm{tW}_1$, where $\{W_t, t\in[0,1]\}$ is the Brownian motion.

Proposition 5.16. The Brownian bridge is a centered Gaussian process and its covariance is $K_B(s,t) = \min(s,t) - st$.

Proof. The Brownian bridge is centered because, by Proposition 5.13, the Brownian motion is centered: $E[B_t] = E[W_t] - \mathrm{tE}[W_1] = 0$.

Let $\{B_t\}$ be a Brownian bridge. We have to prove that any finite selection of the process is jointly Gaussian. Fix any finite sampling $\{B_{t_1}, B_{t_2}, ..., B_{t_m}\}$ and suppose without loss of generality that $t_1 < t_2 < ... < t_m$. By Remark 5.1, $B = (B_{t_1}, B_{t_2}, ..., B_{t_m})$ is a random gaussian vector, since it is the linear transformation of a random gaussian vector:

$$B = \begin{pmatrix} B_{t_1} \\ \vdots \\ B_{t_m} \end{pmatrix} = \begin{pmatrix} W_{t_1} - t_1 W_1 \\ \vdots \\ W_{t_m} - t_m W_1 \end{pmatrix} = \begin{pmatrix} W_{t_1} \\ \vdots \\ W_{t_m} \end{pmatrix} - W_1 \begin{pmatrix} t_1 \\ \vdots \\ t_m \end{pmatrix} =$$

$$\begin{pmatrix} 10...0 - t_1 \\ 01...0 - t_2 \\ \\ 0...01 - t_m \end{pmatrix} \begin{pmatrix} W_{t_1} \\ \vdots \\ W_{t_m} \\ W_1 \end{pmatrix} = \begin{pmatrix} 10...0 - t_1 \\ 01...0 - t_2 \\ \\ 0...01 - t_m \end{pmatrix} \mathcal{W}$$

where $\mathcal{W} = (W_{t_1}, ..., W_{t_m}, W_{t_1})$ is jointly Gaussian, namely a random vector, because by Proposition 5.13 $\{W_t\}$ is a Gaussian process.

We now evaluate the covariance of the Brownian bridge. Given $0 \leqslant s < t \leqslant 1$:

$K_B(s, t) = E[(B_t - E[B_t])(B_s - E[B_s])] = E[B_t B_s - B_t E[B_s] - B_s E[B_t] + E[B_t] E[B_s]] = E[B_t B_s] = E[(W_t - tW_1)(W_s - sW_1)] = E[W_t W_s - sW_t W_1 - tW_1 W_s + stW_1^2] = E[W_t W_s] - \mathrm{sE}[W_t W_1] - tE[W_1 W_s] + stE[W_1^2] = \min(t, s) - s \min(t, 1) - t \min(1, s) + st \min(1, 1) = s - st - ts + st = s - st = \min(s, t) - st$. □

At this point we shall proceed, as previously for the Brownian motion, with the Karhunen-Loève Expansion applied to the linear operator A corresponding to the covariance kernel $K_B(s, t) = \min(s, t) - st$.

Theorem 5.17. The Karhunen-Loève Expansion of the Brownian bridge on $[0, 1]$ is

$$B_t = \frac{\sqrt{2}}{\pi} \sum_{k=1}^{\infty} \frac{Z_k^*}{k} \sin(k\pi t)$$

where $Z_k = \int_0^1 B_t \sqrt{2} \sin(k\pi t) dt,\ \lambda_k = \frac{1}{k^2 \pi^2}, k \in \mathbb{N},\ Z_k^* = k\pi Z_k$ and the convergence in mean square is almost surely.

Proof. We have first to find the eigenvalues of $(Af)(t) = \int_0^1 K_B(s, t) f(s) ds$.

$$\int_0^1 [\min(s, t) - st] e(s) ds = \lambda e(t)$$

$$\int_0^t (s - st)\, e(s) ds + \int_t^1 (t - st)\, e(s)\, ds = \lambda e(t)$$

$$(1 - t) \int_0^t s e(s) ds + t \int_t^1 (1 - s)\, e(s)\, ds = \lambda e(t)$$

Differentiating with respect to t both of the sides we obtain

$$-\int_0^t se(s)ds + (1-t)te(t) + \int_t^1 (1-s)\,e(s)\,ds - (1-t)te(t) = \lambda e'(t)$$

$$\int_t^1 e(s)\,ds - \int_0^1 se(s)\,ds = \lambda e'(s)$$

Differentiating again we obtain

$$-e(t) = \lambda e''(t)$$

which is the same equation obtained for the Brownian motion. Therefore, as previously, the solutions are

$$e(t) = a\sin\frac{t}{\sqrt{\lambda}} + b\cos\frac{t}{\sqrt{\lambda}}$$

and if $\lambda = 0$ then $e(t) = \lambda e''(t) \equiv 0$, which means 0 is not an eigenvalue, and we can proceed to determine a, b and λ.

For $t = 0$ we have $e(0) = b$, but

$$e(0) = \frac{1}{\lambda}\int_0^1 [\min(0,s) - s0]e(s)ds = 0.$$

Thus $b = 0$ and $e(t) = a\sin\frac{t}{\sqrt{\lambda}}$, which we now substitute in $\int_t^1 e(s)\,ds - \int_0^1 se(s)\,ds = \lambda e'(s)$. We obtain

$$a\int_t^1 \sin\frac{s}{\sqrt{\lambda}}\,ds - a\int_0^1 s\sin\frac{s}{\sqrt{\lambda}}\,ds = \lambda\frac{a}{\sqrt{\lambda}}\cos\frac{t}{\sqrt{\lambda}}$$

$$\sqrt{\lambda}\int_{\frac{t}{\sqrt{\lambda}}}^{\frac{1}{\sqrt{\lambda}}} \sin z dz - \lambda\int_0^{\frac{1}{\sqrt{\lambda}}} z\sin z dz = \frac{\lambda}{\sqrt{\lambda}}\cos\frac{t}{\sqrt{\lambda}}$$

$$-\lambda\cos z\Big|_{\frac{t}{\sqrt{\lambda}}}^{\frac{-1}{\sqrt{\lambda}}} - \lambda\sqrt{\lambda}\left\{-z\cos z\Big|_0^{\frac{1}{\sqrt{\lambda}}} + \int_0^{\frac{1}{\sqrt{\lambda}}}\cos z dz\right\} = \lambda\cos\frac{t}{\sqrt{\lambda}}$$

$$\cos\frac{t}{\sqrt{\lambda}} - \cos\frac{-1}{\sqrt{\lambda}} + \cos\frac{1}{\sqrt{\lambda}} - \sqrt{\lambda}\sin\frac{1}{\sqrt{\lambda}} = \cos\frac{t}{\sqrt{\lambda}}$$

$$\cos\frac{t}{\sqrt{\lambda}} - \sqrt{\lambda}\sin\frac{1}{\sqrt{\lambda}} = \cos\frac{t}{\sqrt{\lambda}}$$

For $t = 1$ we have

$$\cos\frac{1}{\sqrt{\lambda}} - \sqrt{\lambda}\sin\frac{1}{\sqrt{\lambda}} = \cos\frac{1}{\sqrt{\lambda}}$$

which yields

$$\sin\frac{1}{\sqrt{\lambda}} = 0$$

Namely,

$$\frac{1}{\sqrt{\lambda}} = k\pi, k \in \mathbb{N}.$$

The positive eigenvalues are therefore $\lambda_k = \frac{1}{k^2\pi^2}$, $k \in \mathbb{N}$ and the eigenfunctions are $e_k(t) = a\ \sin(t\,k\pi)$ where a must be a normalization constant by orthonormality condition on the eigenfunctios. Hence

$$\frac{1}{a^2} = \|\sin(t\,k\pi)\|_2^2 = \int_0^1 \sin^2(t\,k\pi)dt = \frac{1}{k\pi}\int_0^{k\pi}\sin^2 z\,dz = \frac{1}{k\pi}\int_0^{k\pi}\frac{1-\cos(2z)}{2}\ dz = \frac{1}{4k\pi}\int_0^{2k\pi} 1-\cos u\,du = \frac{1}{4k\pi}\int_0^{2k\pi} 1-\cos u\,du = \frac{1}{4k\pi}[2k\pi - \sin(2k\pi)] = \frac{1}{2}$$

therefore $a = \left\|\sin\left[\frac{(2k-1)\pi}{2}s\right]\right\|_2^{-1} = \sqrt{2}$.

The whole argument is invertible, by applying A to $e_k(t) = \sqrt{2}\sin(t\,k\pi)$ we obtain

$$(A\,e_k)(t) = \sqrt{2}\int_0^t (s - s\,t)\sin(s\,k\pi)\ ds + \sqrt{2}\int_t^1 (t - s\,t)\sin(s\,k\pi)\ ds = \sqrt{2}\int_0^t s\ \sin(s\,k\pi)\ ds - \sqrt{2}t\int_0^t s\ \sin(s\,k\pi)\ ds + \sqrt{2}t\int_t^1 \sin(s\,k\pi)\ ds - \sqrt{2}t\int_t^1 s\sin(s\,k\pi)\,ds = \sqrt{2}\left[\int_0^t s\sin(s\,k\pi)\,ds - t\int_0^1 s\sin(s\,k\pi)\,ds + t\int_t^1 \sin(s\,k\pi)\,ds\right] = \frac{\sqrt{2}}{k\pi}\left[\frac{1}{k\pi}\int_0^{t\,k\pi} z\sin z\,dz - \frac{t}{k\pi}\int_0^{k\pi} z\sin z\,dz + t\int_{t\,k\pi}^{k\pi}\sin z\,dz\right] = \frac{\sqrt{2}}{k\pi}\left\{-\frac{1}{k\pi}z\cos z\Big|_0^{t\,k\pi} + \frac{1}{k\pi}\int_0^{t\,k\pi}\cos z\,dz + \frac{t}{k\pi}z\cos z\Big|_0^{k\pi} - \frac{t}{k\pi}\int_0^{k\pi}\cos z\,dz - t\cos z|_{t\,k\pi}^{k\pi}\right\} = \frac{\sqrt{2}}{k\pi}\left\{-t\cos(\,t\,k\pi) + \frac{1}{k\pi}\sin(t\,k\pi) + t\cos(k\pi) - t\cos k\pi + t\cos(t\,k\pi)\right\} = \frac{\sqrt{2}}{k^2\pi^2}\sin(t\,k\pi) = \lambda_k e_k(t).$$

Eventually, according to the Karhuen-Loève Expansion, given

$$Z_k = \int_0^1 B_t\sqrt{2}\sin(k\pi t)\,dt$$

we have that

$$B_t = \frac{\sqrt{2}}{\pi}\sum_{k=1}^{\infty}\frac{Z_k^*}{k}\sin(k\pi t)$$

where, being $\{B_t\}$ a Gaussian process, by Remark 4.7 and Theorem 5.4 it follows that the random variables $\frac{Z_k}{\sqrt{\lambda_k}} = k\pi Z_k = Z_k^* \sim \mathcal{N}(0,1)$ are independent. By Corollary 5.11 the convergence of the expansion is predicted to be in mean square with probability one. $\square$

Bibliography

R. B. Ash. *Information Theory. Corrected reprint of the 1965 original.* Dover, 1990.

P. Billingsley. *Probability and Measure*. Wiley, 1995.

H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer, 2010.

J. L. Brown, Jr. Mean Square Truncation Error in Series Expansions of Random Functions. *Journal of the Society for Industrial and Applied Mathematics* 8.1: 28-32, 1960.

R. Courant and D. Hilbert. *Methods of Mathematical Physics*. Vol. I, Wiley, 1989.

K. Fukunaga and W. L. G. Koontz. Representation of Random Processing Using the Finite Kahrunen-Loève Expansion. *Information and Control* 16: 85-101, 1970.

P. E. T. Jorgensen and M.-S. Song. Entropy Encoding, Hilbert Space and Karhunen-Loève Transforms. *Journal of Mathematical Physics* 48, 2007.

K. Karhunen. Zur Spektraltheorie Stochastiche Prozesse. *Annales Academiae Scientarum Fennicae. Ser. A. I. Math.-Phys.* 34, 1946.

K. Karhunen. Uber lineare methoden in der wahrscheinlichkeitsrechnung. *Annales Academiae Scientarum Fennicae. Ser. A. I. Math.-Phys.* 37: 3-79, 1947.

D. D. Kosambi. Statistics in Function Space. *Journal of the Indian Mathematical Society* 7: 76-88, 1943.

M. Loève. Fonctions aléatoires du second ordre. *Processus Stochastique et Mouvement Brownien, P. Levy (ed.)*, 1948.

M. Loève. *Probability Theory II*. Springer-Verlag, 1978.

M. Reed and B. Simon. *Methods of Modern Mathematical Physics I: Functional Analysis*. Academic Press, 1980.

F. G. Tricomi. *Integral Equations.* Interscience, 1957.

L. Wang. *Karhunen-Loeve Expansions and their Applications*. Dissertation, The London School of Economics and Political Science. Ann Arbor: ProQuest/UMI, 2008.

D. Williams. *Probability with Martingales*. Cambridge University Press, 1991.