Extended Formula for Kriging Interpolation

X. Bay[a,1,*], B. Gauthier[a,2]

[a]*Ecole Nationale Supérieure des Mines de Saint-Etienne*
158 cours Fauriel, 42023 SAINT-ETIENNE, FRANCE

Abstract

In many fields, Kriging interpolation techniques are used within a *finite* discrete set of known values of a real function (no evaluation or measurement error). In this paper, we extend the simple Kriging formula to the case the function is known on a general (infinite) subset of points. This extension is based on the spectral decomposition of a certain nuclear Hilbert-Schmidt operator. As a step by step application, we revisit the problem of prediction for the Brownian sheet knowing its values on a separation line.

Keywords: Kriging, Gaussian process regression, Gaussian processes, Gaussian measure, regular conditional probability, Gaussian Hilbert spaces, Hilbert-Schmidt operator, reproducing kernel Hilbert space (RKHS), Karhunen-Loève expansion, Brownian sheet, separation line

---

1. Introduction

Conditioning Gaussian processes or linear conditioning is very simple because it reduces to orthogonal projections in Hilbert spaces (see [1]). We first review the fundamental result in this direction and fix some notations.

Let $\mathcal{X}$ be a separable metric space and $(Y_x)_{x\in\mathcal{X}}$ be a real centered Gaussian process indexed by the set $\mathcal{X}$. We assume the covariance function $K(x,x') = \mathbb{E}\left(Y_x Y_{x'}\right)$ continuous in both arguments. Equivalently (see appendix Appendix A), we consider a continuous application from $\mathcal{X}$ to a real separable Gaussian Hilbert space $\mathbf{H}$ of centered variables, the complete linear space spanned by the random variables Y_x.

Let $\mathcal{S}$ be a subset of $\mathcal{X}$. Let x be any point in $\mathcal{X}$ and $\mathbb{E}\left(Y_x \middle| Y_s,\ s\in\mathcal{S}\right)$ be the conditional expectation of the random variable Y_x given $Y_s,\ s\in\mathcal{S}$ (or the corresponding σ-subalgebra). As uncorrelation means independence in the Gaussian case, $\mathbb{E}\left(Y_x \middle| Y_s,\ s\in\mathcal{S}\right)$ is the orthogonal projection of Y_x onto the space $\mathbf{H}_{\mathcal{S}}$, the closed linear subspace of $\mathbf{H}$ spanned by the random variables $Y_s,\ s\in\mathcal{S}$ (or $s\in\bar{\mathcal{S}}$ by mean square continuity of the process). This projection can be seen either in the space $\mathbf{H}$ or in the more familiar bigger space $L^2(\mathbb{P})$, where $(\Omega,\ \mathcal{F},\ \mathbb{P})$ is the underlying complete probability space.

---

*Corresponding author
[1]bay@emse.fr, tel:0033-477490023, fax:0033-477496666
[2]gauthier@emse.fr

When $\mathcal{S} = \{x_1, \cdots, x_n\}$ is finite, the conditional expectation $\mathbb{E}\left(Y_x \middle| Y_s,\ s \in \mathcal{S}\right)$ is of the form

$$\mathbb{E}\left(Y_x \middle| Y_s,\ s \in \mathcal{S}\right) = \sum_{k=1}^{n} \lambda_k(x) Y_{x_k}.$$

The n-dimensional vector $\boldsymbol{\lambda}(x)$ of weights $\lambda_k(x)$ is solution of the linear system

$$\mathbf{K}\boldsymbol{\lambda}(x) = \mathbf{k}(x),$$

where $\mathbf{K} = (K(x_i, x_j))_{1 \leqslant i, j \leqslant n}$ is the $n \times n$ covariance matrix of the random variables $Y_s,\ s \in \mathcal{S}$ and $\mathbf{k}(x) = (K(x, x_i))_{1 \leqslant i \leqslant n}$ is the n-dimensional column vector of cross-covariance between Y_x and the $Y_s,\ s \in \mathcal{S}$. If the random variables Y_{x_k} are formally replaced by a prescribed set of values y_k of a real function $y = f(x)$ at the different points x_k, we get the interpolator

$$x \longmapsto \sum_{k=1}^{n} \lambda_k(x) y_k. \tag{1}$$

This is the simple Kriging interpolation formula or Gaussian regression formula with a priori Gaussian probability measure on the unknown function $f(x)$. This interpolator function or best (linear) predictor is also the mean of the conditional Gaussian distribution of the process Y given the values $Y_{x_k} = y_k$. Assuming with no loss of generality that the linear space $\mathbf{H}_{\mathcal{S}}$ generated by the $Y_s,\ s \in\ \mathcal{S}$, is of full dimension n, the conditional covariance of the process Y is

$$K(x,\ x') - \mathbf{k}(x)^T \mathbf{K}^{-1} \mathbf{k}(x'), \tag{2}$$

with the striking property to be independent of the values $Y_{x_k} = y_k$. It is this conditional covariance structure that provides some measure of the certainty or accuracy of the predictions for different location points outside the subset $\mathcal{S}$ (cf. [2, 3]).

In this paper, we study the case the function is known on a general infinite subset $\mathcal{S}$ of points. More precisely, we consider the natural framework of Gaussian Processes (GPs for short) and establish analogues for conditional mean (1) and conditional covariance (2). Theorem 2.1 in section 2 is the probabilistic version of a result concerning optimal interpolation in reproducing kernel Hilbert space (RKHS) (see [4]) and theorem 2.2 is an important extension from the point of view of statistical learning (Kriging or Gaussian process regression). The main idea is to consider the spectral decomposition of a nuclear Hilbert-Schmidt operator described in terms of the covariance function of the GP and an arbitrary (probability) measure on the data subset $\mathcal{S}$. As a detailed illustration, section 3 gives a rather theoretical use of the formula for conditioning the standard Brownian sheet on a separation line.

2. Main results

Fix a non empty set $\mathcal{S}$ of the separable metric space $\mathcal{X}$, which is naturally equipped with its Borel σ-field (as a separable metric subspace). Take any (probability) measure μ on $\mathcal{S}$ with full support $\mathcal{S}$ and such that

$$\int_{\mathcal{S}} K(s,\ s)\mu(ds) < +\infty.$$

Such a measure always exists. For example, let μ_0 be a discrete probability measure on a dense denumerable subset of $\mathcal{S}$ and consider the finite measure absolutely continuous with respect to μ_0 given by the density

$$s \in \mathcal{S} \longmapsto \frac{1}{1 + K(s,\ s)}.$$

Of course, all the results will be independent of the particular choice of the measure μ. We will not discuss this invariance property but it seems clear to be of high interest (see section 3.2). Note also that all the results will be valid with a (positive) measure finite or σ-finite. But, it is more natural to consider a probability measure in the context of design of experiments using randomization (see [5]).

Consider now the usual Hilbert space $L^2(\mathcal{S},\ \mu)$ (not necessary separable) and the (linear) operator T on $L^2(\mathcal{S},\ \mu)$:

$$f \longmapsto Tf(s) \equiv_s \int_{\mathcal{S}} K(s,\ t)f(t)\mu(dt).$$

From the Schwarz inequality $K(s,\ t)^2 \leq K(s,\ s)K(t,\ t)$, we see that T is the (nuclear) Hilbert-Schmidt operator associated to the (covariance) kernel $K(s,\ t)$, i.e.

$$\int_{\mathcal{S}\times\mathcal{S}} K(s,\ t)^2 \mu(ds)\mu(dt) < +\infty.$$

Thus, it is a compact and self-adjoint positive operator. As the null space of T will play no role, we consider only the non-increasing sequence of its strictly positive spectral values (counting with finite multiplicity) and a corresponding sequence of normalized eigenvectors

$$T\phi_n = \lambda_n \phi_n \ \text{ and } \ \lambda_1 \geq \lambda_2 \geq \cdots > 0.$$

Now, we define the regularized eigenfunctions by the relation

$$x \in \mathcal{X}, \quad \lambda_n \phi_n(x) = \int_{\mathcal{S}} K(x,\ s)\phi_n(s)\mu(ds).$$

Remark that those functions are uniquely and well defined on the whole space $\mathcal{X}$ as elements of the functional space $\mathbb{R}^{\mathcal{X}}$. In fact, the functions $s \longmapsto K(x,\ s)$ are in $L^2(\mathcal{S},\ \mu)$ for all x in $\mathcal{X}$ in virtue of the same Schwarz inequality:

$$K(x,\ s)^2 \leq K(x,\ x)K(s,\ s).$$

Note that the notation is correct. As an element of the linear space $L^2(\mathcal{S},\ \mu)$, the restriction to the subset $\mathcal{S}$ of the regularized eigenfunction $\phi_n(x)$ is the normalized eigenvector $\phi_n(s)$ corresponding to the non zero eigenvalue λ_n. Finally, remark that the regularized eigenfunction $\phi_n(x)$ are in fact (at least) continuous as a simple consequence of the following (Schwarz) inequality:

$$\lambda_n \left|\phi_n(x) - \phi_n(x')\right| \leq \sqrt{\int_{\mathcal{S}} K(s,\ s)\mu(ds)}\sqrt{\mathbb{E}\left(Y_x - Y_{x'}\right)^2}.$$

This discrete (finite or infinite) family of linearly independent functions ϕ_n as elements of $\mathbb{R}^{\mathcal{X}}$ will play the fundamental role for conditioning as stated in the following theorem.

Theorem 2.1 (***Projection***)

For all x in $\mathcal{X}$, the orthogonal projection of Y_x onto the closed linear space $\mathbf{H}_{\mathcal{S}}$ spanned by the $Y_s,\ s\in\ \mathcal{S}$, is

$$\mathbb{E}\left(Y_x \middle| Y_s,\ s\in\mathcal{S}\right) = \sum_n \phi_n(x) \int_{\mathcal{S}} \phi_n(s) Y_s \mu(ds) \quad (\textit{in } \mathbf{H} \textit{ or } L^2(\mathbb{P})).$$

Writing for all x in $\mathcal{X}$,

$$Y_x = \mathbb{E}\left(Y_x \middle| Y_s,\ s\in\mathcal{S}\right) + \epsilon_x,$$

the Gaussian process Y is the sum of two independent centered Gaussian processes with covariance functions (series are absolutely convergent)

$$cov\left(\mathbb{E}\left(Y_x \middle| Y_s,\ s\in\mathcal{S}\right),\ \mathbb{E}\left(Y_{x'} \middle| Y_s,\ s\in\mathcal{S}\right)\right) = \sum_n \lambda_n \phi_n(x)\phi_n(x'), \tag{3}$$

$$cov\left(\epsilon_x, \epsilon_{x'}\right) = K(x,\ x') - \sum_n \lambda_n \phi_n(x)\phi_n(x'). \tag{4}$$

We call the covariance of the residual process ϵ the conditional covariance function and the function

$$x\in\ \mathcal{X} \longmapsto \mathbb{E}\left(Y_x \middle| Y_s = y(s),\ s\in\mathcal{S}\right) = \sum_n \phi_n(x) \int_{\mathcal{S}} \phi_n(s) y(s) \mu(ds) \tag{5}$$

the conditional mean function (or, simply, the mean function) when this expression makes sense for a function $y(s)$ in the space $\mathbb{R}^{\mathcal{S}}$. It is clear that the conditional covariance function do not depend of the measure μ on the set $\mathcal{S}$. As usual with Gaussian distribution, this function is also independent of particular observed values $y(s),\ s\in\mathcal{S}$.

The formula (5) for the conditional mean is of great interest because it is exactly the equivalent of the simple Kriging formula or the best linear predictor in Gaussian process regression. We'll see that this interpolator do not depend in a certain sense of the measure μ (see the theorem 2.2). Up to now, we only have

$$\mathbb{E}\left(Y_{s'} \middle| Y_s = y(s),\ s\in\mathcal{S}\right) = y(s') \quad \mu \textit{ almost everywhere for } s' \textit{ in } \mathcal{S},$$

and for $s \longmapsto y(s)$ in a subset of $L^2(\mathcal{S},\ \mu)$, which is of probability 1 for the Gaussian measure induced by the process $Y_s,\ s\in\mathcal{S}$. Note that T as a nuclear operator is the covariance operator of this Gaussian measure.

Remark 1 : (Extended formula for Kriging) In the case $\mathcal{S} = \{x_1, \cdots, x_n\}$, any measure μ on $\mathcal{S}$ is a discrete sum of Dirac distributions

$$\mu = \mu_1 \delta_{x_1} + \cdots + \mu_n \delta_{x_n}.$$

The condition on the support of μ is simply that all the weights μ_i must be strictly positive. In this case, it can be easily seen that the conditional mean function (5) do not depend of the particular choice of the μ_i and is exactly the simple Kriging predictor (1) (see [4] for details). □

To prove the theorem 2.1, we first need two lemmas concerning the random series of functions $\sum_n \phi_n(x) \int_{\mathcal{S}} \phi_n(s) Y_s \mu(ds)$.

Lemma 2.1

The paths $s \longmapsto Y_s$ are in $L^2(\mathcal{S},\ \mu)$ with probability 1.

Proof: Let $(\xi_n)_{n\geq 1}$ be an orthonormal basis of the separable Gaussian Hilbert space $\mathbf{H}$ spanned by the process Y. Note that $(\xi_n)_{n\geq 1}$ is a sequence of independent and identically distributed $N(0,\ 1)$ random variables. We have

$$Y_s = \sum_n c_n(s)\xi_n$$

where $c_n(s) = \mathbb{E}\left(Y_s\xi_n\right)$ are continuous (hence measurable) functions by mean square continuity of the process Y. Furthermore,

$$\sum_n \int_{\mathcal{S}} \mathbb{E}\left(c_n(s)\xi_n\right)^2 \mu(ds) = \int_{\mathcal{S}} K(s,\ s)\mu(ds) < +\infty.$$

Hence the series $Y_s = \sum_n c_n(s)\xi_n$ is convergent in $L^2(\mathcal{S}\times\Omega,\ \mu\times\mathbb{P})$ which implies by Fubini theorem the measurability of the paths $s \longmapsto Y_s$ (with probability 1) and the fact that

$$\mathbb{E}\int_{\mathcal{S}} Y_s^2\mu(ds) = \int_{\mathcal{S}} K(s,\ s)\mu(ds) < +\infty.$$

■

Remark 2 : Proof of lemma 2.1 says more since we have proved that $Y_s = \sum_n c_n(s)\xi_n$ is in $L^2(\mathcal{S}\times\Omega,\ \mu\times\mathbb{P})$. □

Lemma 2.2

For all x in $\mathcal{X}$, we have

$$\sum_n \lambda_n\phi_n(x)^2 < +\infty \text{ and } \sum_n \phi_n(x)\int_{\mathcal{S}} \phi_n(s)Y_s\mu(ds)$$

is a $L^2(\mathbb{P})$-convergent series that defines a Gaussian centered random variable Z_x in the subspace $\mathbf{H}_{\mathcal{S}}$, and with variance $\sum_n \lambda_n\phi_n(x)^2$.

Proof: Consider by the lemma 2.1 (and remark 2) the sequence of random variables $\eta_n = \frac{1}{\sqrt{\lambda_n}}\int_{\mathcal{S}} \phi_n(s)Y_s\mu(ds)$. It can be easily seen to be a sequence of uncorrelated second order centered random variables with variance 1, thus an orthonormal family of random variables in $L^2(\mathbb{P})$. Now, for x in $\mathcal{S}$, the corresponding Fourier coefficients of Y_x are

$$\mathbb{E}\left(Y_x\eta_n\right) = \frac{1}{\sqrt{\lambda_n}}\int_{\mathcal{S}} K(x,\ s)\phi_n(s)\mu(ds) = \sqrt{\lambda_n}\phi_n(x)$$

by definition of the regularized eigenfunctions $\phi_n(x)$. By the Bessel inequality, we have

$$\sum_n \lambda_n\phi_n(x)^2 \leq K(x,\ x) < +\infty.$$

Thus, Z_x is the sum of uncorrelated second order centered random variables with total variance $\sum_n \lambda_n\phi_n(x)^2 < +\infty$. Furthermore, we see that Z_x is uncorrelated with any random variable orthogonal to the $Y_s,\ s\in\mathcal{S}$. Thus Z_x is in the closed linear space spanned by the $Y_s,\ s\in\mathcal{S}$. We conclude that Z_x is Gaussian, which ends the proof of lemma 2.2. ■

Remark 3 : The random variables $\eta_n = \frac{1}{\sqrt{\lambda_n}} \int_{\mathcal{S}} \phi_n(s) Y_s \mu(ds)$ are in the subspace $\mathbf{H}_{\mathcal{S}}$ (hence Gaussian). In fact, the process Z is Gaussian and is defined by a natural Karhunen-Loève expansion associated to the covariance kernel

$$cov\left(Z_x,\ Z_{x'}\right) = \sum_n \lambda_n \phi_n(x) \phi_n(x').$$

$\square$

Proof of theorem 2.1 : Fix $x \in \mathcal{X}$. The function $s \longmapsto K(x,s)$ is in the Hilbert space $L^2(\mathcal{S},\ \mu)$ with Fourier coefficient associated to the function $\phi_n(s)$

$$\int_{\mathcal{S}} K(x,\ s)\phi_n(s)\mu(ds) = \lambda_n \phi_n(x).$$

Let $\phi(s)$ be any function in the null space of T. Then,

$$\mathbb{E}\left(\int_{\mathcal{S}} \phi(s) Y_s \mu(ds)\right)^2 = 0.$$

Hence, $\int_{\mathcal{S}} \phi(s) Y_s \mu(ds) = 0$ and, a fortiori,

$$0 = \mathbb{E}\left(Y_x \int_{\mathcal{S}} \phi(s) Y_s \mu(ds)\right) = \int_{\mathcal{S}} K(x,\ s)\phi(s)\mu(ds).$$

We conclude that the function $s \longmapsto K(x,s)$ is orthogonal to the null space of T and, by Parseval's identity (harmonic synthesis),

$$K(x,\ s) = \sum_n \lambda_n \phi_n(x)\phi_n(s) \quad \textit{in the space } L^2(\mathcal{S},\ \mu).$$

But, using notation of lemma 2.2, we have for s in $\mathcal{S}$:

$$\mathbb{E}(Y_x - Z_x)Y_s = K(x,\ s) - \sum_n \lambda_n \phi_n(x)\phi_n(s).$$

Thus

$$\mathbb{E}(Y_x - Z_x)Y_s = 0 \quad \mu \textit{ almost everywhere in } \mathcal{S}.$$

By mean square continuity of Y and the fact that the support of μ is $\mathcal{S}$, we conclude

$$\mathbb{E}(Y_x - Z_x)Y_s = 0 \quad \textit{for all } s \textit{ in } \mathcal{S},$$

which proves that $Z_x = \sum_n \phi_n(x) \int_{\mathcal{S}} \phi_n(s) Y_s \mu(ds)$ is the orthogonal projection of Y_x onto the closed linear space spanned by the $Y_s,\ s \in \mathcal{S}$. This is the first part of theorem 2.1. For the second part, use the relation

$$cov\left(\mathbb{E}\left(Y_x \middle| Y_s,\ s \in \mathcal{S}\right),\ \mathbb{E}\left(Y_{x'} \middle| Y_s,\ s \in \mathcal{S}\right)\right) = \mathbb{E}\left(Y_{x'} \sum_n \phi_n(x) \int_{\mathcal{S}} \phi_n(s) Y_s \mu(ds)\right)$$

to prove the identity (3) and consequently (4) by orthogonal decomposition. $\blacksquare$

Remark 4 : Let $\mathcal{H}$ be the RKHS (see [6]) with (covariance) kernel $K(x,\ x')$ and $\mathcal{H}_1$ be the closed subspace spanned by the functions $K(s,\ .)$ for $s \in \mathcal{S}$. By the natural isomorphism between $\mathcal{H}$ and the Gaussian space $\mathbf{H}$, the space $\mathcal{H}_1$ is the analogue of $\mathbf{H}_{\mathcal{S}}$, the closed linear space spanned the random variables $Y_s,\ s \in \mathcal{S}$. For $h \in \mathcal{H}$, the orthogonal projection of h onto the space $\mathcal{H}_1$ is (see [4])

$$\mathbb{P}_{\mathcal{H}_1}[h](x) = \sum_n \phi_n(x) \int_{\mathcal{S}} \phi_n(s) h(s) \mu(ds). \tag{6}$$

So, we have the following suggestive relation as in the usual finite case (cf. [7])

$$\mathbb{P}_{\mathcal{H}_1}[h](x) = \mathbb{E}\left(Y_x \middle| Y_s = h(s),\ s \in \mathcal{S}\right).$$

□

It is well known that sample functions of the process Y are in the RKHS $\mathcal{H}$ with the probability zero (in the infinite dimensional case). As a consequence, the relation (6) in remark 4 that gives a sense to the conditional mean function simultaneously for all x is valid on a set of probability zero. So, the important question that arises now is to get a regular conditional probability distribution of the Gaussian process $(Y_x)_{x \in \mathcal{X}}$ given the $(Y_s)_{s \in \mathcal{S}}$ (see appendix Appendix B). Note that this question is not relevant in the case $\mathcal{S}$ finite (usual case) or in the case $\mathcal{X}$ denumerable. The next theorem gives a satisfying answer under extra assumption on the continuity of the process.

For any topological space $\mathcal{K}$ compact, we denote by $C(\mathcal{K})$ the usual Banach space of real continuous functions on $\mathcal{K}$ (uniform norm).

Theorem 2.2 (*Regular Conditional Probability*)

Suppose the process Y admits a version with continuous sample functions and the parameter metric space $\mathcal{X}$ compact. With no loss of generality, assume that $\mathcal{S}$ is a closed subspace of $\mathcal{X}$. Then, the random series of functions $\sum_n \phi_n(x) \int_{\mathcal{S}} \phi_n(s) Y_s \mu(ds)$ is, with probability 1, uniformly convergent on $\mathcal{X}$ and defines a continuous centered Gaussian process

$$x \in \mathcal{X} \longmapsto Z_x = \sum_n \phi_n(x) \int_{\mathcal{S}} \phi_n(s) Y_s \mu(ds).$$

If μ' is a different measure on the set $\mathcal{S}$ with full support $\mathcal{S}$, the corresponding process Z' is a modification of the process $x \longmapsto Z_x$ such that

$$\mathbb{P}\left(\forall x \in \mathcal{X},\ Z'_x = Z_x\right) = 1.$$

For any function $s \in \mathcal{S} \longmapsto y(s)$ into a subset of $C(\mathcal{S})$ corresponding to uniformly convergent series $\sum_n \phi_n(x) \int_{\mathcal{S}} \phi_n(s) y(s) \mu(ds)$ on the parameter space $\mathcal{X}$, we can define the Gaussian probability distribution on the space $C(\mathcal{X})$ of mean

$$x \longmapsto \mathbb{E}\left(Y_x \middle| Y_s = y(s),\ s \in \mathcal{S}\right) = \sum_n \phi_n(x) \int_{\mathcal{S}} \phi_n(s) y(s) \mu(ds),$$

and covariance function

$$cov\left(\epsilon_x, \epsilon_{x'}\right) = K(x,\ x') - \sum_n \lambda_n \phi_n(x) \phi_n(x').$$

Then, this family of Gaussian distributions on $C(\mathcal{X})$ is a regular conditional probability of the process Y given the $(Y_s)_{s\in\mathcal{S}}$.

Remark 5 : (continuing remark 4) Let h in $\mathcal{H}$ and $x \in \mathcal{X}$, we have (see [4]),

$$|\sum_{k\geq n}\phi_k(x)\int_{\mathcal{S}}\phi_k(s)h(s)\mu(ds)|^2 \leq \|h\|^2_{\mathcal{H}}\sum_{k\geq n}\lambda_k\phi_k(x)^2.$$

Since the parameter space $\mathcal{X}$ is compact and $x \longmapsto \sum_n \lambda_n\phi_n(x)^2$ is a continuous function (by mean square continuity of the process Y and continuity of projection), Dini's theorem implies the uniform convergence of the series $\sum_n \phi_n(x)\int_{\mathcal{S}}\phi_n(s)h(s)\mu(ds)$. Remark now that the restriction to the subset $\mathcal{S}$ of the functions in $\mathcal{H}$ is the RKHS generated by the trace kernel $K(s,\ s'),\ s,\ s'\ in\ \mathcal{S}$. Thus, this RKHS is included in the set of continuous functions $s \in \mathcal{S} \longmapsto y(s)$ such that the series $\sum_n \phi_n(x)\int_{\mathcal{S}}\phi_n(s)h(s)\mu(ds)$ is uniformly convergent. Note that such series functions are independent of the representation (the measure μ) since we have

$$\sum_n \phi_n(x)\int_{\mathcal{S}}\phi_n(s)h(s)\mu(ds) = \mathbb{P}_{\mathcal{H}_1}[h](x).$$

Hence, the conditional distribution is unambiguously defined for any h in the RKHS generated by the trace kernel $K(s,\ s'),\ s,\ s'\ in\ \mathcal{S}$ (or h in $\mathcal{H}$). As

$$\mathcal{H}_1^{\perp} = \{h \in \mathcal{H} : \forall s \in \mathcal{S},\ h(s) = 0\},$$

we have the interpolating property

$$\forall s \in \mathcal{S}, \sum_n \phi_n(s)\int_{\mathcal{S}}\phi_n(s')h(s')\mu(ds') = \mathbb{P}_{\mathcal{H}_1}[h](s) = h(s).$$

By continuity of sample functions, we also have

$$\mathbb{E}\left(Y_s \middle| Y_{s'} = y(s'),\ s' \in \mathcal{S}\right) = y(s)$$

for $s \longmapsto y(s)$ in a subset of $C(\mathcal{S})$, which is of probability 1 for the Gaussian measure on $C(\mathcal{S})$ induced by the process $Y_s,\ s \in \mathcal{S}$. □

Proof of theorem 2.2 : The main difficulty is to prove the uniform convergence of the random series of functions $\sum_n \phi_n(x)\int_{\mathcal{S}}\phi_n(s)Y_s\mu(ds)$. We first see that the process $x \longmapsto \mathbb{E}\left(Y_x \middle| Y_s,\ s \in \mathcal{S}\right)$ admits a continuous version and then use a general result of the theory of Gaussian fields. For the first point which seems very intuitive, we use a symmetry argument. Introduce the process $Y^{'}$

$$x \in \mathcal{X},\ Y^{'}_x = \mathbb{E}\left(Y_x \middle| Y_s,\ s \in \mathcal{S}\right) - \epsilon_x.$$

Observe that $Y^{'}$ and Y are equal in distribution (by independence of the two right hand processes and symmetry of the residual process ϵ). Consider now the relation

$$\frac{1}{2}(Y_x + Y^{'}_x) = \mathbb{E}\left(Y_x \middle| Y_s,\ s \in \mathcal{S}\right)$$

on a dense denumerable subset of points of the compact metric set $\mathcal{X}$. By hypothesis of continuity of the sample functions of the process Y, the left hand term is a uniformly continuous function with probability 1. Hence, the process

$$x \longmapsto \mathbb{E}\left(Y_x \middle| Y_s,\ s \in \mathcal{S}\right) = \sum_n \phi_n(x) \int_{\mathcal{S}} \phi_n(s) Y_s \mu(ds)$$

admits a continuous version (and also the process ϵ_x). Remark now that

$$x \longmapsto \sum_n \phi_n(x) \int_{\mathcal{S}} \phi_n(s) Y_s \mu(ds)$$

is a natural Karhunen-Loève decomposition of this process using the basis functions $\sqrt{\lambda_n}\phi_n(x)$ of the reproducing kernel Hilbert space generated by the covariance kernel

$$cov\left(\mathbb{E}\left(Y_x \middle| Y_s,\ s \in \mathcal{S}\right),\ \mathbb{E}\left(Y_{x'} \middle| Y_s,\ s \in \mathcal{S}\right)\right) = \sum_n \lambda_n \phi_n(x)\phi_n(x').$$

Conclude by a general result [8][theorem 3.8]. The rest of the theorem follows from the sample function continuity of the process $x \longmapsto \sum_n \phi_n(x) \int_{\mathcal{S}} \phi_n(s) Y_s \mu(ds)$ and the residual process ϵ. ■

Remark 6 : The theorem 2.2 still holds with sligth modifications for a parameter space $\mathcal{X}$ σ-compact ($\mathcal{X}$ is the increasing reunion of a denumerable set of compacts). □

Remark 7 : As can be shown with the proof of theorem 2.2, it is sufficient to have a measure (finite or σ-finite) with support the closed set $\overline{\{s \in \mathcal{S},\ K(s,\ s) > 0\}}^{\mathcal{S}}$ of $\mathcal{S}$. Note that $K(s,\ s) = 0$ means $Y_s = 0$. □

3. Illustration: Conditioning the Brownian Sheet

In this section $\mathcal{X} = \mathbb{R}_+ \times \mathbb{R}_+$ will denote the positive quadrant in the plane, $x = (x_1, x_2)$ will be a generic element of $\mathcal{X}$. We remind that the Brownian sheet is the centered Gaussian process $(B_x)_{x \in \mathcal{X}}$ with covariance kernel, for x and $y \in \mathcal{X}$,

$$K(x, y) = (x_1 \wedge y_1)(x_2 \wedge y_2),$$

where $x_i \wedge y_i$ $(i = 1, 2)$ denotes the minimum between the two reals x_i and y_i. Alternatively, the Brownian sheet is the two-parameter continuous Gaussian process solution to the stochastic partial differential equation

$$\frac{\partial^2 F(x_1,\ x_2)}{\partial x_1 \partial x_2} = \dot{W}_{(x_1,\ x_2)},$$

with boundary conditions $F(x_1,\ 0) = F(0,\ x_2) = 0$ and where $\dot{W}$ is the standard two-parameter White noise process. Following [9], our aim is to establish a prediction formula for the Brownian sheet, knowing its values on a *separation line* $\mathcal{S}$.

3.1. The quarter-circle case

We first consider the case $\mathcal{S}$ is the unit quarter-circle, i.e. $\mathcal{S} = \left\{x \in \mathcal{X} \ : \ x_1^2 + x_2^2 = 1\right\}$ and assume $\mathcal{S}$ endowed with its natural Lebesgue measure. Throughout section 3.1, we will use the abusive notation

$$\theta \in \left[0, \frac{\pi}{2}\right], \ B_\theta = B_{(\cos\theta, \ \sin\theta)}, \ f(\theta) = f(\cos\theta, \ \sin\theta) \text{ and so on.}$$

Using a polar coordinate system, we have to consider the integral operator T:

$$\alpha \in \left[0, \frac{\pi}{2}\right], \ T[f](\alpha) = \int_0^{\frac{\pi}{2}} (\cos\alpha \wedge \cos\theta)\,(\sin\alpha \wedge \sin\theta)\, f(\theta) d\theta.$$

Straightforward calculations (see appendix Appendix C) show that the spectral decomposition of T (in $L^2(\mathcal{S}, \ \mu)$) is:

$$n \geqslant 1, \ \lambda_n = \frac{1}{4n^2 - 1}, \text{ associated with } \widetilde{\phi}_n(\theta) = \frac{2}{\sqrt{\pi}} \sin(2n\theta).$$

The next step is the calculation of the regularized eigenfunctions defined by the relation

$$\lambda_n \phi_n(x) = \int_0^{\frac{\pi}{2}} (x_1 \wedge \cos\theta)\,(x_2 \wedge \sin\theta)\, \widetilde{\phi}_n(\theta) d\theta. \tag{7}$$

As suggested in figure 1, we define α_1 and $\alpha_2 \in \left[0, \frac{\pi}{2}\right]$ such that $x_1 = \cos\alpha_1$ and $x_2 = \sin\alpha_2$. We have to consider two different cases:

- x lies below the separation line, i.e. x is inside the quarter-disc, i.e. $\|x\| \leq 1$ (with $\|\cdot\|$ the euclidean norm),
- x lies above the separation line, i.e. x is outside the quarter-disc, i.e. $\|x\| \geq 1$.

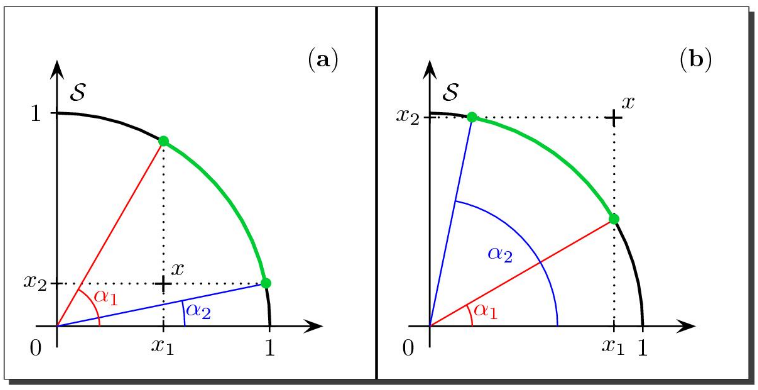

Figure 1: Graphical representation and parametrization of the intercepted arc for x inside the quarter-disc (**a**) and outside the quarter-disc (**b**).

$\mathbf{x}$ *below the separation line:*. In this case $\alpha_1 \geq \alpha_2$ and

$$K\big(x, (\cos\theta, \sin\theta)\big) = \begin{cases} x_1 \sin\theta, \ \text{for } \theta \leqslant \alpha_2, \\ x_1 x_2, \ \text{for } \alpha_2 \leq \theta \leqslant \alpha_1, \\ x_2 \cos\theta, \ \text{for } \alpha_1 \leq \theta. \end{cases}$$

By splitting the integral (7) into three terms and using integration by parts twice, we get:

$$\begin{aligned} \lambda_n \phi_n(x) &= \frac{x_2}{4n^2}\sqrt{1-x_1^2}\widetilde{\phi}_n(\alpha_1) + \frac{x_1}{4n^2}\sqrt{1-x_2^2}\widetilde{\phi}_n(\alpha_2) \\ &\quad + \frac{x_1}{4n^2}\int_0^{\alpha_2} \sin\theta \widetilde{\phi}_n(\theta) d\theta + \frac{x_2}{4n^2}\int_{\alpha_1}^{\frac{\pi}{2}} \cos\theta \widetilde{\phi}_n(\theta) d\theta. \end{aligned}$$

Since $\lambda_n = \frac{1}{4n^2-1}$, we have the interesting relation

$$\begin{aligned} \phi_n(x) &= x_2\sqrt{1-x_1^2}\widetilde{\phi}_n(\alpha_1) + x_1\sqrt{1-x_2^2}\widetilde{\phi}_n(\alpha_2) \\ &\quad + x_1\int_0^{\alpha_2} \sin\theta \widetilde{\phi}_n(\theta) d\theta + x_2\int_{\alpha_1}^{\frac{\pi}{2}} \cos\theta \widetilde{\phi}_n(\theta) d\theta - \lambda_n \phi_n(x). \end{aligned} \tag{8}$$

Remind the notation $B_\theta = B_{(\cos\theta, \sin\theta)}$ to deduce from relation (8) and theorem 2.2 that

$$\begin{aligned} \mathbb{E}\left(B_x \middle| B_s, \ s \in \mathcal{S}\right) &= x_2\sqrt{1-x_1^2}B_{\alpha_1} + x_1\sqrt{1-x_2^2}B_{\alpha_2} \\ &\quad + x_1\int_0^{\alpha_2} \sin\theta B_\theta d\theta + x_2\int_{\alpha_1}^{\frac{\pi}{2}} \cos\theta B_\theta d\theta \\ &\quad - \int_0^{\frac{\pi}{2}} \sum_{n\geqslant 1} \lambda_n \phi_n(x) \widetilde{\phi}_n(\theta) B_\theta d\theta. \end{aligned}$$

From the orthogonal decomposition of the kernel, we have

$$\forall x \in \mathcal{X} \text{ and } \forall s \in \mathcal{S}, \ K(x,s) = \sum_{n\geqslant 1} \lambda_n \phi_n(x) \phi_n(s),$$

i.e., $\forall x \in \mathcal{X}$ and $\forall \theta \in \left[0, \frac{\pi}{2}\right]$,

$$\sum_{n\geqslant 1} \lambda_n \phi_n(x) \widetilde{\phi}_n(\theta) = K\big(x, (\cos\theta, \sin\theta)\big) = \begin{cases} x_1 \sin\theta, \ \text{for } \theta \leqslant \alpha_2, \\ x_1 x_2, \ \text{for } \alpha_2 \leq \theta \leqslant \alpha_1, \\ x_2 \cos\theta, \ \text{for } \alpha_1 \leq \theta. \end{cases}$$

After two simplifications, we finally get

$$\mathbb{E}\left(B_x \middle| B_s, \ s \in \mathcal{S}\right) = x_2\sqrt{1-x_1^2}B_{\alpha_1} + x_1\sqrt{1-x_2^2}B_{\alpha_2} - x_1 x_2 \int_{\alpha_2}^{\alpha_1} B_\theta d\theta.$$

x *above the separation line:*. In this case, $\alpha_1 \leq \alpha_2$, $\alpha_1 = 0$ if $x_1 \geq 1$ and $\alpha_2 = \frac{\pi}{2}$ if $x_2 \geq 1$. For all x such that $\|x\| \geqslant 1$ and $\theta \in \left[0, \frac{\pi}{2}\right]$,

$$K\big(x, (\cos\theta, \sin\theta)\big) = \begin{cases} x_1 \sin\theta, \text{ for } \theta \leqslant \alpha_1, \\ \cos\theta \sin\theta, \text{ for } \alpha_1 \leq \theta \leqslant \alpha_2, \\ x_2 \cos\theta, \text{ for } \alpha_2 \leq \theta. \end{cases}$$

In exactly the same way as in the previous case, we obtain for all x such that $\|x\| \geqslant 1$,

$$\mathbb{E}\left(B_x \middle| B_s,\ s \in \mathcal{S}\right) = \left(1 - x_1^2\right)_+ B_{\alpha_1} + \left(1 - x_2^2\right)_+ B_{\alpha_2} + 3\int_{\alpha_1}^{\alpha_2} \cos\theta \sin\theta B_\theta d\theta,$$

where $t_+ = \max(0,\ t)$ for $t \in \mathbb{R}$.

Now, we can state the following theorem:

Theorem 3.1

Let $(B_x)_{x \in \mathbb{R}_+^2}$ be the Brownian sheet and $\mathcal{S} = \left\{x \in \mathbb{R}_+^2,\ \|x\| = 1\right\}$. Then,

- *for x such that $\|x\| \leqslant 1$ $(x_1 = \cos\alpha_1,\ x_2 = \sin\alpha_2,\ \alpha_2 \leq \alpha_1)$:*

$$\mathbb{E}\left(B_x \middle| B_s,\ s \in \mathcal{S}\right) = x_2\sqrt{1 - x_1^2} B_{\alpha_1} + x_1\sqrt{1 - x_2^2} B_{\alpha_2} - x_1 x_2 \int_{\alpha_2}^{\alpha_1} B_\theta d\theta, \quad (9)$$

- *for x such that $\|x\| \geqslant 1$ $\left(\alpha_1 \leq \alpha_2,\ \alpha_1 = 0 \text{ if } x_1 \geq 1,\ \alpha_2 = \frac{\pi}{2} \text{ if } x_2 \geq 1\right)$:*

$$\mathbb{E}\left(B_x \middle| B_s,\ s \in \mathcal{S}\right) = \left(1 - x_1^2\right)_+ B_{\alpha_1} + \left(1 - x_2^2\right)_+ B_{\alpha_2} + 3\int_{\alpha_1}^{\alpha_2} \cos\theta \sin\theta B_\theta d\theta, \quad (10)$$

with the notation $B_\theta = B_{(\cos\theta, \sin\theta)}$ for any $\theta \in \left[0, \frac{\pi}{2}\right]$.

Proof: (*Direct proof*) With the notation $Z_x = \mathbb{E}\left(B_x \middle| B_s,\ s \in \mathcal{S}\right)$, we only have to prove for any $x \in \mathcal{X}$ and any $\alpha \in \left[0, \frac{\pi}{2}\right]$:

$$\mathbb{E}\left\{B_\alpha \left(B_x - Z_x\right)\right\} = 0,$$

or, equivalently,

$$\mathbb{E}\left\{B_\alpha B_x\right\} = \mathbb{E}\left\{B_\alpha Z_x\right\}.$$

Fix x such that $\|x\| \leqslant 1$. For $\alpha \leq \alpha_2$, we have

$$\mathbb{E}\left\{B_\alpha Z_x\right\} = x_2\sqrt{1 - x_1^2}\cos\alpha_1 \sin\alpha + x_1\sqrt{1 - x_2^2}\cos\alpha_2 \sin\alpha - x_1 x_2 \int_{\alpha_2}^{\alpha_1} \cos\theta \sin\alpha d\theta.$$

Hence,

$$\mathbb{E}\left\{B_\alpha Z_x\right\} = x_1 x_2\sqrt{1 - x_1^2}\sin\alpha + x_1(1 - x_2^2)\sin\alpha - x_1 x_2(\sin\alpha_1 - \sin\alpha_2)\sin\alpha.$$

Finally, using the relations $x_1 = \cos\alpha_1$, $x_2 = \sin\alpha_2$, we get

$$\mathbb{E}\left\{B_\alpha Z_x\right\} = x_1 \sin\alpha = \mathbb{E}\left\{B_\alpha B_x\right\}.$$

If $\alpha_2 \leq \alpha \leq \alpha_1$, we have

$$\begin{aligned}\mathbb{E}\left\{B_\alpha Z_x\right\} &= x_2\sqrt{1-x_1^2}\cos\alpha_1\sin\alpha + x_1\sqrt{1-x_2^2}\sin\alpha_2\cos\alpha \\ &\quad -x_1x_2\left(\int_{\alpha_2}^{\alpha}\cos\alpha\sin\theta d\theta + \int_{\alpha}^{\alpha_1}\cos\theta\sin\alpha d\theta\right).\end{aligned}$$

Hence,

$$\begin{aligned}\mathbb{E}\left\{B_\alpha Z_x\right\} &= x_1x_2\sqrt{1-x_1^2}\sin\alpha + x_1x_2\sqrt{1-x_2^2}\cos\alpha \\ &\quad -x_1x_2(\cos\alpha_2-\cos\alpha)\cos\alpha - x_1x_2(\sin\alpha_1-\sin\alpha)\sin\alpha,\end{aligned}$$

and $\mathbb{E}\left\{B_\alpha Z_x\right\} = x_1x_2 = \mathbb{E}\left\{B_\alpha B_x\right\}$. Finally, the last case $\alpha_1 \leq \alpha$ is symmetric of the first one.

The case x such that $\|x\| \geq 1$ is analogue. ■

Remark 8 : We have found in a different way and under a different form known results of prediction for the Brownian sheet. Let $\varphi(t) = (\varphi_1(t),\ \varphi_2(t)) = (t,\ \sqrt{1-t^2})$ be the usual parametrization of the quarter-circle ($t \in [0,\ 1]$). Let $a = \cos\alpha_1 = x_1$ and $b = \cos\alpha_2 = \sqrt{1-x_2^2}$. By a stochastic integration by parts, the theorem 4.2 (a) in [9] implies, for $\|x\| \leqslant 1$,

$$\mathbb{E}\left(B_x\middle|B_s,\ s\in\mathcal{S}\right) = \varphi_1(a)\varphi_2(b)\left(\frac{B_{\varphi(a)}}{\varphi_1(a)\varphi_2(a)} + \frac{bB_{\varphi(b)}}{\varphi_2(b)} - \frac{aB_{\varphi(a)}}{\varphi_2(a)} - \int_a^b \frac{B_{\varphi(t)}}{\sqrt{1-t^2}}dt\right).$$

Hence,

$$\mathbb{E}\left(B_x\middle|B_s,\ s\in\mathcal{S}\right) = x_2\sqrt{1-x_1^2}B_{\alpha_1} + x_1\sqrt{1-x_2^2}B_{\alpha_2} - x_1x_2\int_a^b \frac{B_{\varphi(t)}}{\sqrt{1-t^2}}dt.$$

By the usual change of variables $t = \cos\theta$, we see that the expression (4.4) of [9, theorem 4.2] leads to the formula (9) in theorem 3.1. In the same way, the expression (4.6) of [9, theorem 4.2] gives for $0 < b \leq a < 1$

$$\mathbb{E}\left(B_x\middle|B_s,\ s\in\mathcal{S}\right) = B_{\varphi(a)} - \left(\frac{a^3B_{\varphi(a)}}{\varphi_1(a)} - \frac{b^3B_{\varphi(b)}}{\varphi_1(b)} - \int_b^a 3tB_{\varphi(t)}dt\right).$$

Thus,

$$\mathbb{E}\left(B_x\middle|B_s,\ s\in\mathcal{S}\right) = \left(1-x_1^2\right)B_{\alpha_1} + \left(1-x_2^2\right)B_{\alpha_2} + 3\int_b^a tB_{\varphi(t)}dt.$$

Use the same change of variables $t = \cos\theta$ to get the formula (10) in theorem 3.1. □

Corollary 3.1

Let $C_0(\mathcal{S})$ be the space of continuous functions f on $\mathcal{S}$ such that $f(0) = f(\frac{\pi}{2}) = 0$. For any f in $C_0(\mathcal{S})$, we can define the continuous conditional mean function $x \longmapsto \mathbb{E}\left(B_x\middle|B_\theta = f(\theta), \theta\in\left[0,\frac{\pi}{2}\right]\right)$ by

- *for x below* $(\|x\| \leq 1)$:

$$x \longmapsto x_2\sqrt{1-x_1^2}f(\alpha_1) + x_1\sqrt{1-x_2^2}f(\alpha_2) - x_1x_2\int_{\alpha_2}^{\alpha_1} f(\theta)d\theta,$$

- *for x above* $(\|x\| \geq 1)$:

$$x \longmapsto \left(1-x_1^2\right)_+ f(\alpha_1) + \left(1-x_2^2\right)_+ f(\alpha_2) + 3\int_{\alpha_1}^{\alpha_2} \cos\theta \sin\theta f(\theta)d\theta.$$

Consider the conditional covariance function given by the relation

$$cov\,(\epsilon_x, \epsilon_{x'}) = K(x,\ x') - cov\left(B_x,\ B_{x'}\middle|B_s,\ s \in \mathcal{S}\right)$$

where the covariance term $cov\left(B_x,\ B_{x'}\middle|B_s,\ s \in \mathcal{S}\right)$ can be analytically calculated using the theorem 3.1.

Then, this family of conditional Gaussian distributions on the space of continuous functions on $\mathcal{X}$ is a regular conditional probability of the Brownian sheet given its values on the unit quarter-circle. The mean function is a weak solution to the partial differential equation

- *(for x below)*

$$\frac{\partial^2}{\partial x_1 \partial x_2}\left(\frac{F(x_1,\ x_2)}{x_1x_2}\right) = 0,$$

- *(for x above)*

$$\frac{\partial^2 F(x_1,\ x_2)}{\partial x_1 \partial x_2} = 0,$$

with the boundary condition $F = f$ on $\mathcal{S} = \{x \in \mathcal{X},\ \|x\| = 1\}$. For f in the RKHS $\mathcal{H}_\mathcal{S}$ generated by the trace kernel $K(s,\ s')$, $s,\ s'$ in $\mathcal{S}$ (i.e. f is the restriction to $\mathcal{S}$ of a function in the RKHS $\mathcal{H}$ associated to $K(\cdot,\cdot)$, see (11)), the corresponding mean function is also a spline interpolation function solution to the following regularization (optimization) problem:

$$h = f \text{ on } \mathcal{S} \text{ and } \iint \left(\frac{\partial^2 h(x)}{\partial x_1 \partial x_2}\right)^2 dx_1dx_2 \text{ minimum on } \mathcal{H}$$

where $$\mathcal{H} = \left\{h : h(x_1, x_2) = \int_0^{x_1}\int_0^{x_2} g(u_1, u_2)du_1du_2 \text{ with } g = \frac{\partial^2 h}{\partial x_1 \partial x_2} \in L^2(dx_1dx_2)\right\} \tag{11}$$

is the RKHS associated to the Brownian sheet.

Proof: Let f in $C(\mathcal{S})$. Observe that the condition $f(0) = f(\frac{\pi}{2}) = 0$ is a necessary (and sufficient) condition for the continuity of the mean function

$$x \longmapsto \mathbb{E}\left(B_x\middle|B_\theta = f(\theta), \theta \in \left[0, \frac{\pi}{2}\right]\right).$$

Since $B_0 = B_{\frac{\pi}{2}} = 0$, the subspace $C_0(\mathcal{S})$ of $C(\mathcal{S})$ is of probability 1 and theorem 2.2 implies the first part of this corollary. For the second part, the interpretation of the mean function in terms of solution of a certain PDE is a direct consequence of its analytical expression. The second interpretation is a consequence of remarks **??** and **??**. ■

Remark 9 : Consider the standard Brownian motion $(W_x)_{x\geq 0}$. Let u be any real number. Conditioning by $W_1 = u$ leads to the Gaussian process of mean $x \longmapsto u \times x$ for $x \leq 1$, and $x \longmapsto u$ for $x \geq 1$ (the mean is constant). Note that this mean function is both solution of a first-order differential equation and an optimization problem. Corollary 3.1 is the 2-parameter version of this observation. But, there is an important difference. For the Brownian motion conditioned by its value in $x = 1$, the conditional covariance function is $x \wedge x' - xx'$ for x and $x' \leqslant 1$ (standard Brownian bridge covariance), $(x-1) \wedge (x'-1)$ for x and $x' \geqslant 1$ and is zero in the other case (x below and x' above)! In other words, the two Gaussian processes $(W_x)_{x\leq 1}$ and $(W_x)_{x\geq 1}$ are independent conditionally to W_1. It is also a simple consequence of the Markov property. In the quarter-circle case, the two Gaussian processes $(B_x)_{\|x\|\leq 1}$ and $(B_x)_{\|x\|\geq 1}$ are *not* independent conditionally to $B_s,\ s \in \mathcal{S}$. Thus, it is necessary to consider the special case x *below and* x' *above* to have a full analytical expression of the conditional covariance function in corollary 3.1 (see figure 2 for a simulation). □

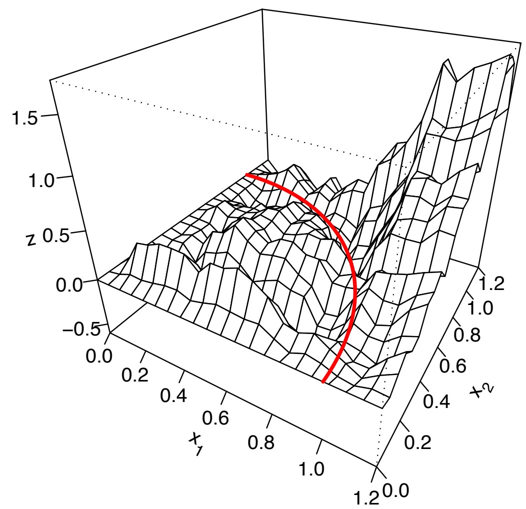

Figure 2: Realization of a Brownian sheet null on $\mathcal{S}$.

3.2. The general case

Let $\mathcal{S}$ be a separation line, i.e. any polar curve

$$\mathcal{S} = \Big\{ x \in \mathcal{X} \ : \ x = \big(r(\theta)\cos\theta, r(\theta)\sin\theta\big), \theta \in \Big[0, \frac{\pi}{2}\Big] \Big\}$$

where $\theta \mapsto r(\theta)$ is a continuous function with values in $]0, +\infty]$ and such that

$$\begin{cases} r(\theta)\cos\theta \ : \ \left[0, \frac{\pi}{2}\right] \to [0, +\infty] \text{ non-increasing,} \\ r(\theta)\sin\theta \ : \ \left[0, \frac{\pi}{2}\right] \to [0, +\infty] \text{ non-decreasing.} \end{cases} \quad (12)$$

Remark 10 : In the case $\theta \mapsto r(\theta)\cos\theta$ (strictly) increasing, a separation line can also be interpreted as the graph of a function of type $x_2 = f(x_1)$ with f non-increasing. □ **Remark 11 :** Observe that this definition also includes the case when $\mathcal{S}$ contains horizontal or vertical segments, or both. But, as an important difference with [9], we do not require that $\mathcal{S}$ meets coordinate axes. □

The monotony-properties of a separation line imply the following relation for the Brownian sheet, θ and $\alpha \in \left[0, \frac{\pi}{2}\right]$:

$$Cov\left(B_{(r(\alpha)\cos\alpha,\, r(\alpha)\sin\alpha)},\ B_{(r(\theta)\cos\theta,\, r(\theta)\sin\theta)}\right) = r(\alpha)r(\theta)Cov\left(B_{(\cos\alpha,\, \sin\alpha)},\ B_{(\cos\theta,\, \sin\theta)}\right).$$

In other words, we have for a separation line:

$$\left(r(\alpha)\cos\alpha \wedge r(\theta)\cos\theta\right)\left(r(\alpha)\sin\alpha \wedge r(\theta)\sin\theta\right) = r(\alpha)r(\theta)\left(\cos\alpha \wedge \cos\theta\right)\left(\sin\alpha \wedge \sin\theta\right). \quad (13)$$

Throughout section 3.2, we will use the abusive notation $B_\theta = B_{(r(\theta)\cos\theta,\, r(\theta)\sin\theta)}$, for $\theta \in \left[0, \frac{\pi}{2}\right]$. For simplicity, we will assume that the function $\theta \mapsto r(\theta)$ is of class C^2. As we are free to use any positive measure on $\mathcal{S}$ and of full support $\mathcal{S}$, we choose $\mu(d\theta) = \frac{1}{r(\theta)^2}d\theta$. We have

$$\int_{\mathcal{S}} K(s,\ s)\mu(ds) = \int_0^{\frac{\pi}{2}} \cos\theta\sin\theta d\theta = \frac{1}{2} < +\infty.$$

Consider the operator

$$T[f](\alpha) = \int_0^{\frac{\pi}{2}} \left(r(\alpha)\cos\alpha \wedge r(\theta)\cos\theta\right)\left(r(\alpha)\sin\alpha \wedge r(\theta)\sin\theta\right) f(\theta)\frac{1}{r(\theta)^2}d\theta.$$

Now, the spectral decomposition of T can be easily deduce from the quarter-circle case. Denote by λ_n and $\widetilde{\phi}_n$, $n \geqslant 1$ the spectral decomposition of section 3.1 integral operator, we have that the eigenvalues of T are λ_n, $n \geqslant 1$, associated to the normalized eigenfunctions:

$$\forall\theta \in \left[0, \frac{\pi}{2}\right],\ \widetilde{\phi}_n^r(\theta) = r(\theta)\widetilde{\phi}_n(\theta).$$

Indeed, we remark by (13) that

$$\begin{aligned} T[\widetilde{\phi}_n^r](\alpha) &= \int_0^{\frac{\pi}{2}} \left(r(\alpha)\cos\alpha \wedge r(\theta)\cos\theta\right)\left(r(\alpha)\sin\alpha \wedge r(\theta)\sin\theta\right)\widetilde{\phi}_n^r(\theta)\frac{1}{r(\theta)^2}d\theta \\ &= \int_0^{\frac{\pi}{2}} r(\alpha)r(\theta)\left(\cos\alpha \wedge \cos\theta\right)\left(\sin\alpha \wedge \sin\theta\right) r(\theta)\widetilde{\phi}_n(\theta)\frac{1}{r(\theta)^2}d\theta \\ &= \lambda_n r(\alpha)\widetilde{\phi}_n(\alpha) = \lambda_n\widetilde{\phi}_n^r(\alpha). \end{aligned}$$

As in section 3.1, the next step is the regularization of the eigenfunctions, i.e. the calculation for $x \in \mathcal{X}$ and $n \geqslant 1$ of:

$$\lambda_n \phi_n^r(x) = \int_0^{\frac{\pi}{2}} (x_1 \wedge r(\theta)\cos\theta)\,(x_2 \wedge r(\theta)\sin\theta)\,\widetilde{\phi}_n^r(\theta)\frac{d\theta}{r(\theta)^2}. \tag{14}$$

We pose $x_1 = r(\alpha_1)\cos\alpha_1$ and $x_2 = r(\alpha_2)\sin\alpha_2$ with α_1 and $\alpha_2 \in \left[0, \frac{\pi}{2}\right]$ and consider the two cases.

x *below the separation line:*. $\alpha_1 \geqslant \alpha_2$ and

$$K\big(x, (r(\theta)\cos\theta, r(\theta)\sin\theta)\big) = \begin{cases} x_1 r(\theta)\sin\theta, \text{ for } \theta \leqslant \alpha_2, \\ x_1 x_2, \text{ for } \alpha_2 \leq \theta \leqslant \alpha_1, \\ x_2 r(\theta)\cos\theta, \text{ for } \alpha_1 \leq \theta. \end{cases}$$

By splitting the integral and using integration by parts twice, we find:

$$\begin{aligned} \lambda_n \phi_n^r(x) &= \frac{x_2}{4n^2}\sin\alpha_1 \widetilde{\phi}_n(\alpha_1) + \frac{x_1}{4n^2}\cos\alpha_2 \widetilde{\phi}_n(\alpha_2) \\ &+ \frac{x_2}{4n^2}\int_{\alpha_1}^{\frac{\pi}{2}} \cos\theta \widetilde{\phi}_n(\theta) d\theta + \frac{x_1}{4n^2}\int_0^{\alpha_2} \sin\theta \widetilde{\phi}_n(\theta) d\theta \\ &+ \frac{x_1 x_2}{4n^2}\left[\frac{d}{d\theta}\left(\frac{1}{r(\theta)}\right)(\alpha_1)\widetilde{\phi}_n(\alpha_1) - \frac{d}{d\theta}\left(\frac{1}{r(\theta)}\right)(\alpha_2)\widetilde{\phi}_n(\alpha_2)\right] \\ &- \frac{x_1 x_2}{4n^2}\int_{\alpha_2}^{\alpha_1} \frac{d^2}{d\theta^2}\left(\frac{1}{r(\theta)}\right)(\theta)\widetilde{\phi}_n(\theta) d\theta. \end{aligned}$$

Following section 3.1 and using the analogue notation $B_\theta = B_{(r(\theta)\cos\theta,\, r(\theta)\sin\theta)}$ for $\theta \in \left[0, \frac{\pi}{2}\right]$, we get:

$$\begin{aligned} \mathbb{E}\left(B_x \middle| B_s,\ s \in \mathcal{S}\right) &= x_2\left[\frac{\sin\alpha_1}{r(\alpha_1)} + \cos\alpha_1 \frac{d}{d\theta}\left(\frac{1}{r(\theta)}\right)(\alpha_1)\right] B_{\alpha_1} \\ &+ x_1\left[\frac{\cos\alpha_2}{r(\alpha_2)} - \sin\alpha_2 \frac{d}{d\theta}\left(\frac{1}{r(\theta)}\right)(\alpha_2)\right] B_{\alpha_2} \\ &- x_1 x_2 \int_{\alpha_2}^{\alpha_1}\left[\frac{d^2}{d\theta^2}\left(\frac{1}{r(\theta)}\right)(\theta) + \frac{1}{r(\theta)}\right] B_\theta \frac{d\theta}{r(\theta)}. \end{aligned}$$

x *above the separation line:*. Use for $\alpha_1 \leqslant \alpha_2$ the expression

$$K\big(x, (r(\theta)\cos\theta, r(\theta)\sin\theta)\big) = \begin{cases} x_1 r(\theta)\sin\theta, \text{ for } \theta \leqslant \alpha_1, \\ r(\theta)^2\cos\theta\sin\theta, \text{ for } \alpha_1 \leq \theta \leqslant \alpha_2, \\ x_2 r(\theta)\cos\theta, \text{ for } \alpha_2 \leq \theta. \end{cases}$$

This leads to:

Theorem 3.2

Let $(B_x)_{x\in\mathcal{X}}$ be the Brownian sheet on $\mathcal{X} = \mathbb{R}_+^2$ and let $\mathcal{S}$ be a separation line verifying conditions (12) and with $r(\theta)$ of class C^2. Then,

- *for x below the separation line* $(x_1 = r(\alpha_1)\cos\alpha_1,\ x_2 = r(\alpha_2)\sin\alpha_2,\ \alpha_2 \leq \alpha_1)$:

$$\begin{aligned}\mathbb{E}\left(B_x \middle| B_s,\ s\in\mathcal{S}\right) &= x_2\left[\frac{\sin\alpha_1}{r(\alpha_1)} + \cos\alpha_1 \frac{d}{d\theta}\left(\frac{1}{r(\theta)}\right)(\alpha_1)\right] B_{\alpha_1}\\ &\quad + x_1\left[\frac{\cos\alpha_2}{r(\alpha_2)} - \sin\alpha_2 \frac{d}{d\theta}\left(\frac{1}{r(\theta)}\right)(\alpha_2)\right] B_{\alpha_2}\\ &\quad - x_1 x_2 \int_{\alpha_2}^{\alpha_1}\left[\frac{d^2}{d\theta^2}\left(\frac{1}{r(\theta)}\right) + \frac{1}{r(\theta)}\right] B_\theta \frac{d\theta}{r(\theta)},\end{aligned}$$

- *for x above the separation line* $(\alpha_1 \leq \alpha_2,\ \alpha_1 = 0$ *if* $x_1 \geq r(0),\ \alpha_2 = \frac{\pi}{2}$ *if* $x_2 \geq r(\frac{\pi}{2}))$:

$$\begin{aligned}\mathbb{E}\left(B_x \middle| B_s,\ s\in\mathcal{S}\right) &= \left[\cos^2\alpha_1 - \frac{1}{r(\alpha_1)}\frac{d}{d\theta}(r(\theta)\cos\theta\sin\theta)(\alpha_1)\right] B_{\alpha_1}\\ &\quad + \left[\sin^2\alpha_2 + \frac{1}{r(\alpha_2)}\frac{d}{d\theta}(r(\theta)\cos\theta\sin\theta)(\alpha_2)\right] B_{\alpha_2}\\ &\quad - \int_{\alpha_1}^{\alpha_2}\left[\frac{d^2}{d\theta^2}\left(r(\theta)\cos\theta\sin\theta\right) + r(\theta)\cos\theta\sin\theta\right] B_\theta \frac{d\theta}{r(\theta)},\end{aligned}$$

with $B_\theta = B_{(r(\theta)\cos\theta,\ r(\theta)\sin\theta)}$ *for* $\theta \in \left[0, \frac{\pi}{2}\right]$.

Proof: With $Z_x = \mathbb{E}\left(B_x \middle| B_s,\ s\in\mathcal{S}\right)$, a direct proof of this theorem consists to verify

$$\forall\alpha \in \left[0, \frac{\pi}{2}\right],\ \mathbb{E}\left\{B_\alpha B_x\right\} = \mathbb{E}\left\{B_\alpha Z_x\right\}.$$

It is a straightforward but long calculation using the relations:

- (for x below)

$$\cos\theta\left(\frac{d^2}{d\theta^2}\left(\frac{1}{r(\theta)}\right) + \frac{1}{r(\theta)}\right) = \frac{d}{d\theta}\left(\frac{\sin\theta}{r(\theta)} + \cos\theta\frac{d}{d\theta}\left(\frac{1}{r(\theta)}\right)\right),$$

$$-\sin\theta\left(\frac{d^2}{d\theta^2}\left(\frac{1}{r(\theta)}\right) + \frac{1}{r(\theta)}\right) = \frac{d}{d\theta}\left(\frac{\cos\theta}{r(\theta)} - \sin\theta\frac{d}{d\theta}\left(\frac{1}{r(\theta)}\right)\right),$$

- (for x above)

$$\cos\theta\left(\frac{d^2}{d\theta^2}\left(r(\theta)\cos\theta\sin\theta\right) + r(\theta)\cos\theta\sin\theta\right) = \frac{d}{d\theta}\left(\sin\theta(r(\theta)\cos\theta\sin\theta) + \cos\theta\frac{d}{d\theta}\left(r(\theta)\cos\theta\sin\theta\right)\right),$$

$$-\sin\theta\left(\frac{d^2}{d\theta^2}\left(r(\theta)\cos\theta\sin\theta\right) + r(\theta)\cos\theta\sin\theta\right) = \frac{d}{d\theta}\left(\sin\theta(r(\theta)\cos\theta\sin\theta) - \sin\theta\frac{d}{d\theta}\left(r(\theta)\cos\theta\sin\theta\right)\right).$$

■

Remark 12 : As we made in corollary 3.1, this last theorem could be reformulated in terms of conditional probability distribution of the process Y knowing its values on a separation line $\mathcal{S}$. In particular, the conditional mean function associated to any continuous function f on $\mathcal{S}$ with certain limit conditions is both solution of a PDE and an optimization problem if f is in a certain RKHS. □

Acknowledgments

We would like to thank all members of the 3MI team for their helpful suggestions. All graphics and computations have been performed with the free software `R`.

Appendix A. Mean square continuity

Let $\mathcal{X}$ be a metric space (or any topological space) and $(Y_x)_{x\in\mathcal{X}}$ be a real centered Gaussian process indexed by the set $\mathcal{X}$.

Proposition Appendix A.1

The covariance function $K(x,x') = \mathbb{E}\,(Y_x Y_{x'})$ is continuous in both arguments if and only if the process Y is mean square continuous (i.e. the application $x \in \mathcal{X} \longmapsto Y_x \in L^2(\mathbb{P})$ is continuous).

Proof: The identity

$$\mathbb{E}\,(Y_x - Y_{x'})^2 = K(x,\ x) + K(x',\ x') - 2K(x,\ x')$$

trivially shows the necessary condition. The sufficient condition is the interesting one. To prove it, we fix the point (x,x') in $\mathcal{X}\times\mathcal{X}$ and we show that the covariance function K is continuous in (x,x'). We have

$$\left|K(y,\ y') - K(x,\ x')\right| \le \left|K(y,\ y') - K(y,\ x')\right| + \left|K(y,\ x') - K(x,\ x')\right|.$$

From the Schwarz inequality, we get

$$\left|K(y,\ x') - K(x,\ x')\right|^2 \le K(x',\ x') \times \mathbb{E}\,(Y_y - Y_x)^2,$$

and also

$$\left|K(y,\ y') - K(y,\ x')\right|^2 \le K(y,\ y) \times \mathbb{E}\,(Y_{y'} - Y_{x'})^2.$$

Note that the mean square continuity of the process implies the continuity of the variance function $y \longmapsto K(y,\ y)$. Now, we easily see that the covariance $K(y,\ y')$ tends to K(x, x') as (y, y') tends to (x, x'). ■

Appendix B. Regular conditional probability

Let $(F,\ \mathcal{F},\ \mu)$ be a probability space and $(G,\ \mathcal{G})$ be a measurable space. Let $T : F \longrightarrow G$ be a measurable application and denote by μ_T the push-forward measure of μ on G (μ_T is the law or distribution of the random element T). A regular conditional probability (or desintegration) of μ with respect to T is defined as a map $\pi : G\times\mathcal{F} \longrightarrow [0,\ 1]$ such that:

1. For all $\mathcal{A}$ in $\mathcal{F}$, $g \in G \longmapsto \pi(g,\ \mathcal{A})$ is $\mathcal{G}$-measurable
2. For all g in G, $\mathcal{A} \in \mathcal{F} \longmapsto \pi(g,\ \mathcal{A})$ is a probability
3. For all $\mathcal{A}$ in $\mathcal{F}$ and $\mathcal{B}$ in G, $\mu\left(\mathcal{A}\ and\ T \in \mathcal{B}\right) = \int_{\mathcal{B}} \pi(g,\ \mathcal{A})\mu_T(dg)$

The probability $\pi(g, \mathcal{A})$ is interpreted as the conditional probability of the event $\mathcal{A}$ given $T = g$. In a more familiar notation, $\pi(g,\ \mathcal{A}) = \mu\left(\mathcal{A}|T = g\right)$. As can be seen in the definition, we only need to define $\pi(g,\ \mathcal{A})$ for points g in a set of probability 1 for μ_T.

Let $\mathcal{X}$ be a compact metric space. Consider $F = C(\mathcal{X})$ the usual separable Banach space (Polish space) of real continuous functions on $\mathcal{X}$ (uniform topology). Note that the Borel σ-algebra $\mathcal{F}$ coincides with the cylindrical σ-algebra and that any probability μ on $\mathcal{F}$ is a Radon measure. In this case, it always exists a regular conditional probability of μ with respect to any measurable application T (see [10]). Let $\mathcal{S}$ be any closed subset of points in $\mathcal{X}$. Consider the application T from $C(\mathcal{X})$ to $G = C(\mathcal{S})$ where $g = T(f)$ is the restriction of f to $\mathcal{S}$. Then, T is (linear) continuous (hence measurable). Theorem 2.2 gives the way to construct a regular conditional probability when μ is the Gaussian measure induced by a continuous centered Gaussian process indexed by the set $\mathcal{X}$. In this special case, the familiar significance of a regular conditional probability is, for $g \in C(\mathcal{S})$ (or for g in a set of probability 1 as in theorem 2.2) and $\mathcal{A} \in \mathcal{F}$,

$$\mu\left(\mathcal{A}|T = g\right) = \mathbb{P}\left(\ (Y_x)_{x\in\mathcal{X}} \in \mathcal{A} \mid Y_s = g(s),\ s \in \mathcal{S}\right).$$

We say that the family of Gaussian measures (or Gaussian probability distributions) $\mu\left(\mathcal{A}|T = g\right)$ is a regular conditional probability of the process $(Y_x)_{x\in\mathcal{X}}$ given $(Y_s)_{s\in\mathcal{S}}$.

Finally, note that such a regular conditional probability is unique up to a set of functions of probability zero for the Gaussian measure on $C(\mathcal{S})$ induced by the process $(Y_s)_{s\in\mathcal{S}}$.

Appendix C. Spectral Decomposition of the Section 3.1 Integral Operator

With the notation of section 3.1, we have for f in $L^2(\left[0, \frac{\pi}{2}\right],\ d\theta)$:

$$\begin{aligned} T[f](\alpha) &= \int_0^{\frac{\pi}{2}} \left(\cos\alpha \wedge \cos\theta\right)\left(\sin\alpha \wedge \sin\theta\right) f(\theta)d\theta \\ &= \int_0^{\alpha} \cos\alpha \sin\theta f(\theta)d\theta + \int_{\alpha}^{\frac{\pi}{2}} \sin\alpha \cos\theta f(\theta)d\theta. \end{aligned} \tag{C.1}$$

Let ϕ in $L^2(\left[0, \frac{\pi}{2}\right],\ d\theta)$ and $\lambda > 0$ such that

$$T[\phi](\alpha) = \lambda\phi(\alpha). \tag{C.2}$$

Combining (C.1) and (C.2), we can choose a representative of ϕ to be the function defined for all α by the relation

$$\lambda\phi(\alpha) = \cos\alpha \int_0^{\alpha} \sin\theta\phi(\theta)d\theta + \sin\alpha \int_{\alpha}^{\frac{\pi}{2}} \cos\theta\phi(\theta)d\theta. \tag{C.3}$$

This last relation shows that ϕ is continuous with $\phi(0) = \phi\left(\frac{\pi}{2}\right) = 0$. But now, relation (C.3) implies that ϕ is of class C^1 with

$$\lambda\phi'(\alpha) = -\sin\alpha \int_0^\alpha \sin\theta\phi(\theta)d\theta + \cos\alpha \int_\alpha^{\frac{\pi}{2}} \cos\theta\phi(\theta)d\theta.$$

Finally, we see that ϕ is of class C^∞ and solution of the following second-order differential equation

$$\text{for } \alpha \in \left[0, \frac{\pi}{2}\right],\ \lambda\phi''(\alpha) + (\lambda+1)\phi(\alpha) = 0 \text{ and } \phi(0) = \phi\left(\frac{\pi}{2}\right) = 0.$$

References

[1] S. Janson. *Gaussian Hilbert Spaces.* Cambridge University Press, 1997.

[2] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning.* The MIT Press, 2006.

[3] M. L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging.* Springer Verlag, 1999.

[4] B. Gauthier and X. Bay. Optimal Interpolation in RKHS, Spectral Decomposition of Integral Operators and Application. *Mathematical and Computer Modelling, submitted April 26th, 2010*, see http://www.emse.fr/~bay/Optimal_Interpolation_And_Operators.pdf.

[5] V. V. Fedorov. *Theory of Optimal Experiments.* New York Academic Press, 1972.

[6] N. Aronszajn. Theory of Reproducing Kernels. *Trans. Amer. Math. Soc.*, 63:337–404, 1950.

[7] G. Wahba. *Spline Models for Observational Data.* SIAM, 1990.

[8] R. J. Adler. *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes.* IMS Lecture Notes-Monograph Series, 1990.

[9] R. C. Dalang and F. Russo. A Prediction Problem for the Brownian Sheet. *Journal of Multivariate Analysis*, 26:16–47, 1988.

[10] K. R. Parthasarathy. *Probability Measures on Metric Spaces.* Academic Press Inc., 1967.