DOI: 10.1111/cgf.14501

COMPUTER GRAPHICS *forum*
Volume 41 (2022), number 6 pp. 67–81

Gaussian Process for Radiance Functions on the $\mathbb{S}^2$ Sphere

R. Marques,[1] C. Bouville[2] and K. Bouatouch[2]

[1]Departament de Matemàtiques i Informàtica, Universitat de Barcelona, Barcelona, Spain
[2]IRISA & Université de Rennes 1, Rennes, France

Abstract

Efficient approximation of incident radiance functions from a set of samples is still an open problem in physically based rendering. Indeed, most of the computing power required to synthesize a photo-realistic image is devoted to collecting samples of the incident radiance function, which are necessary to provide an estimate of the rendering equation solution. Due to the large number of samples required to reach a high-quality estimate, this process is usually tedious and can take up to several days. In this paper, we focus on the problem of approximation of incident radiance functions on the $\mathbb{S}^2$ sphere. To this end, we resort to a Gaussian Process (GP), a highly flexible function modelling tool, which has received little attention in rendering. We make an extensive analysis of the application of GPs to incident radiance functions, addressing crucial issues such as robust hyperparameter learning, or selecting the covariance function which better suits incident radiance functions. Our analysis is both theoretical and experimental. Furthermore, it provides a seamless connection between the original spherical domain and the spectral domain, on which we build to derive a method for fast computation and rotation of spherical harmonics coefficients.

Keywords: rendering, ray tracing, signal processing, methods and applications

CCS Concepts: • Computing methodologies → Rendering; Gaussian processes; Ray tracing

1. Introduction

Gaussian processes (GPs) have proven to be a highly flexible modelling technique for spatial and temporal data or processes. GPs can suit a large variety of problems such as speech waveforms [Wil19], geophysical data [CVVMM*16], wireless communication channels [NRL*20], among others. Many well-known stochastic processes such as Brownian motions, Langevin processes and Wiener processes are all special cases of GP. GPs are particularly well-suited to prior modelling in Bayesian regression problems. It is indeed the simplest way to model priors without parameters since a GP is fully-specified by a mean and a covariance function only. GP-based Bayesian regression (GPR) provides a powerful tool to deal with unknown function approximation and interpolation. Many existing approximation methods such as splines or regularized least-square can be expressed in a GP-based framework [RW06].

GPs present several advantages, which make them appealing for unknown function approximation. One of them is that GPR does not require any precomputed topological data structure, such as mesh or connectivity list. Another one, and perhaps the most important, is the probabilistic characterization of the knowledge and uncertainty about the unknown function. This feature is particularly interesting, since the careful exploitation of this knowledge (which is generally not available when using other techniques) has already opened several promising research avenues such as uncertainty-driven sampling [GOG*14] or even the joint estimation of multiple correlated integrals [XBG18]. However, these breakthroughs face several obstacles in what concerns a direct application to rendering. Indeed, the aforementioned works almost exclusively focus on providing a single or a small number of unknown function approximations. Therefore, in their case, the GP model can easily be adapted to the features of the particular target function(s), by selecting a suitable covariance function and learning the corresponding hyperarameters. Nevertheless, a direct application of this approach to rendering, where millions of function approximations must be performed, would imply an excessive computational overhead. This raises the question of whether there exists a general GP setting which can be learned over a set of different incident radiance functions, and then used for an efficient and accurate approximation of most incident radiance functions.

Our goal in this paper is to answer the above question and set the grounds for a straightforward use of GPs for modelling incident radiance functions on the $\mathbb{S}^2$ sphere. To this end, we tackle the problem of model selection for incident radiance functions by exploring

different potentially suitable covariance functions, and learning their most adequate hyperparameters for a set of distinct incident radiance functions. We show that radiance functions can be modelled with GPs under a very simple and general GP setting. Moreover, to further motivate the use of GPs in rendering and enlarge its future applicability, we also show that GPR can be used to very effectively compute the spherical harmonics (SH) coefficients of the approximated functions. Indeed, another benefit of GPR is that, thanks to its kernel-based approach, it allows an easy connection with the spectral representation of radiance functions, i.e. their SH expansion. We show that this property can be used for a fast computation of all SH coefficients up to a specified degree through a single transform matrix and a set of samples values at specified locations. Furthermore, we also show that this SH expansion can be expressed in a form that enables a fast and easy rotation method of the SH coefficients. Our main contributions can be summarized as follows:

1. An analysis of covariance functions so as to compare their suitability to GP modelling of radiance functions. Both probabilistic and spectral approaches are considered in this analysis;
2. A new simplified likelihood expression for hyperparameter learning, which we use to show how the covariance functions can be parametrized;
3. A performance analysis of GPR applied to radiance function approximation considering various covariance functions;
4. The derivation and experimentation of a new SH computation method using a single transform matrix that enables a fast computation of rotated SH coefficients.

Our paper is structured as follows. We first give an overview of the related work in the field of spherical functions interpolation and approximation with a focus on kernel-based and SH-based methods (Section 2). We then introduce the theory of GPs, covariance functions and Bayesian regression (Section 3). Our contribution begins in Section 4 with the spectral analysis of different covariance functions (Section 4.1), followed by the derivation of our simplified hyperparameters learning method (Section 4.2), and by our new method for enabling fast SH coefficients computation and rotation using GPs (Section 4.3). Then, in Section 5, we present a set of experiments illustrating the benefits of our proposed approach. Finally, we discuss our experimental results in Section 6 and conclude in Section 7, providing also some suggestions for future research.

2. Related Work

Many rendering algorithms require interpolation and approximation of radiance functions. Interpolation allows significant savings in computing time when radiance in some direction can be inferred from neighbouring sampled directions. Indeed, this type of functions are known to be rough and computationally expensive to sample. Furthermore, the conditions for which interpolation can provide a radiance estimate with a sufficient precision are difficult to determine. Various solutions that go beyond the simple bilinear interpolation have been proposed in the rendering literature. For example, in Bala *et al*. [BDT99], the authors propose an interpolation technique specifically dedicated to ray tracers, which enables error control through adaptive sampling. Other interpolation techniques used in global illumination renderers are based on a sum of Gaussian models that locally approximate the radiance function (e.g. [RHJD18, GKMD06]). Although the sum of Gaussian approximation has some benefits when it comes to shading integral computation (e.g. in Tsai and Shih [TS06]), the Gaussian radial basis functions (RBFs) are overly smooth to deal with radiance functions.

Another family of methods are based on SH expansions that provide a compact representation of the radiance function enabling not only fast interpolation but also fast integration. These properties are extensively used in the pre-computed radiance transfer (PRT) rendering method so as to enable global illumination rendering in real time (e.g. Sloan *et al*. [SKS02]). In the radiance or irradiance caching methods [KGPB05, AFO05, ZBN19], SH expansions are used to store the directional distribution of incident radiance or irradiance so as to allow fast interpolation from a scattered set of radiance samples. However, to take full advantage of these benefits, the SH coefficients computation load must not be too heavy. Radiance functions are generally highly irregular with sharp intensity variations, and fast transform methods [DH94, DWD08, GLW21] fail to provide valuable results on this type of functions. It is therefore necessary to resort to the straight computations of the inner products through Monte-Carlo integration, which strongly constrains the degree of the SH expansion. When implementing SH-based methods, we are faced with the problem of rotating the SH coefficients so as to express the SH expansion in different coordinate frames. The theoretical solution to this problem is given by the so-called Wigner D functions but this method is known to be computationally expensive and unstable. Several methods have been proposed in the literature to speed up this computation [IR98, BFB97, KKP*06, LdWF12, NSF12]. In Section 4.3, we propose a GP-based solution to this problem that allows both fast computation and rotation of SH coefficients.

Kernel techniques offer a very general theoretical framework for approximation and interpolation [SW06]. For example, Marques *et al*. [MBR*13b, MBSB15] improve the accuracy of illumination integral computation by using a non-parametric Bayesian method known as Bayesian Quadrature (BQ). In another work [MBB19], the theory of reproducing kernel Hilbert spaces (RKHS) is used to compute optimal weights for the computation of shading integrals from sample sets having different directional distributions. Both BQ and RKHS methods are interpolating quadratures since they implicitly use some form of kernel-based interpolation to approximate the integrand. In Schaback and Wendland [SW06], the authors provide a wide survey of kernel-based tools in various fields of numerical analysis and in particular for approximation and interpolation. As regards the case of spherical functions, the work of Narcowhich and Ward [NW02, NSWW07] is particularly interesting as their approach is based on the theory of Sobolev spaces, which allows to make connections between spectral analysis and approximation error. Our approach in this paper is essentially based on GP [RW06] that is used to model the space of the unknown function. In this probabilistic framework, the kernel function is used to express the covariance function but the theoretical work of Narcowhich and Ward still applies as we will see in Sections 4.1 and 4.3.

A deep understanding of approximation and interpolation of unknown functions involves spectral analysis methods, which again are based on SH. Pilleboue *et al*. [PSC*15] focus on the spectral analysis of stochastic integration of spherical functions. To tackle the problem of the convergence rate of error variance, they

consider different classes of characteristic functions to model the integrand power spectrum. Marques *et al*. [MBB20] propose a theoretical framework based on Sobolev spaces to analyse the integration error of quadrature rules used for shading integrals. Their analysis is also extended to the computation of shading integrals considered as inner products in the SH domain.

In the following, we propose a GP-based method for interpolation that better fits the characteristics of radiance function compared to bilinear interpolation or approximations based on the Gaussian kernel. We show that it is possible to find general hyperparameter settings of the GP, which adapt well to a set of different incidence radiance functions, thus avoiding the need for further parameter learning phases in future applications. Besides, a SH-based representation can easily be derived from our model allowing both fast computation of SH coefficients and fast rotation of these coefficients.

3. Theoretical Background

3.1. Overview

In this section, we provide the necessary background for understanding GP-based regression on the $\mathbb{S}^2$ sphere. We start by providing a formal definition of GP (Section 3.2) which, as detailed below, is specified by a *prior mean function* and a *covariance function*. Intuitively, the GP can be thought of as a probabilistic model over a space of functions, the same way that a pdf characterizes the probability distribution over random variables. A GP is a *non-parametric* model where the output of the model (i.e. the form and complexity of the unknown function approximation) directly depends on the number of observations [RW06, Bis06]. In contrast, in parametric models, such as Gaussian Mixture Models (GMMs) or linear regression with basis functions, a number of parameters must be first estimated from the observations, and the model expressiveness is limited by the (finite) number of used parameters [Bis06, RW06]. An example of model parameters in the case of a Gaussian mixture is the number of used components and their respective weights, mean and lengthscale (or bandwidth). No such constraints are involved in a non-parametric regression.

In a non-parametric Bayesian regression, the prior model is usually defined by a GP for the sake of tractability. The prior mean function of the GP contains our prior knowledge or beliefs regarding the value of the function we wish to approximate. The covariance function, in its turn, captures our knowledge or beliefs regarding the sharpness of the unknown function, and is parametrized by a set of *hyperparameters*. The term 'hyperparmeters' emphasizes the fact that it concerns the prior knowledge. Furthermore, we refer to the set composed of the covariance function, the hyperparameters and the mean function of the prior GP model as the *GP setting*. As we show below, the choice of the covariance function and its hyperparameters is of crucial importance to the quality of the function approximation resulting from the *GP-based regression*. Therefore, in Section 3.3, we present different covariance functions of potential interest for GP-based modelling of radiance functions. Then, in Section 3.4, we provide mathematical details on GP-based regression and hyperparameters learning through likelihood maximization. These details are tightly connected to our contributions regarding a spectral analysis of the considered covariance functions (detailed in Section 4.1), a simplified likelihood expression for hyperparameters learning (Section 4.2), and fast SH coefficients computation and rotation (Section 4.3). The most important mathematical terms used throughout the paper, together with a short description, can be found in Table 1.

Table 1: *Table of notation.*

Symbol	Definition
$\mathbf{x}$	Input vector in $\mathbb{R}^D$
$f(\mathbf{x})$	Unknown function to be estimated
$\bar{f}(\mathbf{x})$	Prior mean function $\mathrm{E}[f(\mathbf{x})]$
$k(\mathbf{x}, \mathbf{x}')$	General covariance function
r	Radial distance $\|\mathbf{x} - \mathbf{x}'\|$
$k_s(r)$	Squared exponential covariance function
$kp_0(r)$	Piecewise polynomial covariance function
$kp_1(r)$	Differentiable piecewise polynomial cov. function
$k_{gd}(r)$	Generalized distance covariance function
l_s	Length-scale hyperparameter for k_s, kp_0 and kp_1
s_g	Smoothness hyperparameter for k_{gd}
ϵ	Sample of an IID Gaussian noise
σ_n^2	Variance of ϵ (hyperparameter)
σ_f^2	Variance of the Gaussian Process (hyperparameter)
σ_n'	Noise ratio σ_n/σ_f
X	Set of sample locations s.t. $X = \{\mathbf{x}_i\}$
$\mathbf{t}$	Vector of observations s.t. $t_i = f(\mathbf{x}_i) + \epsilon$
K	Covariance matrix s.t. $K_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)$
$\mathbf{x}_*$	Test input. Specifies a location at which the value of $f(\mathbf{x})$ will be inferred
$\mathbf{k}(\mathbf{x}_*)$	$\mathbf{k}(\mathbf{x}_*) = [k(\mathbf{x}_*, \mathbf{x}_1), \ldots, k(\mathbf{x}_*, \mathbf{x}_n)]$
$f_*(\mathbf{x}_*)$	Posterior estimate of $f(\mathbf{x}_*) = E[f(\mathbf{x}_*)\|\mathbf{t}, X]$

3.2. Definitions

We restrict our analysis to theoretical elements that we have deemed appropriate for rendering applications. For an exhaustive analysis of GPs theory and applications, refer to Rasmussen and Williams [RW06]. Formally, a GP is a stochastic process such that any finite collection of random variables involved in this process has a multivariate normal distribution. A GP is completely defined by its mean function $\bar{f}(\mathbf{x})$ and its covariance function $k(\mathbf{x}, \mathbf{x}')$:

$$\begin{aligned} \bar{f}(\mathbf{x}) &= \mathrm{E}[f(\mathbf{x})] \\ k(\mathbf{x}, \mathbf{x}') &= \mathrm{E}[(f(\mathbf{x}) - \bar{f}(\mathbf{x}))(f(\mathbf{x}')\bar{f}(\mathbf{x}'))] \end{aligned} \tag{1}$$

and will be denoted by:

$$f(\mathbf{x}) \sim \mathcal{GP}\big(\bar{f}(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\big)$$

In Equation (1), $\mathbf{x} \in \mathcal{X}$ denotes an input vector of dimension D, $\mathcal{X}$ being the set of all possible inputs. Given such an input vector $\mathbf{x}$, $f(\mathbf{x})$ is thus a Gaussian random variable and, more generally, given a set of n input vectors $X = \{\mathbf{x}_i\}_1^n$, the set $\{f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n)\}$ is jointly Gaussian with mean value $m(X) = [\bar{f}(\mathbf{x}_1), \ldots, \bar{f}(\mathbf{x}_n)]$ and covariance matrix $k(X, X)$ the coefficients of which are $k(\mathbf{x}_i, \mathbf{x}_j)$. A realization of such a GP is called a *sample function* or *sample path*. The vector space $\mathcal{X}$ over which the GP is defined may be $\mathbb{R}$, e.g. time, or more generally $\mathbb{R}^D$. In rendering applications, we are mostly

concerned with spherical functions, in which case $\mathcal{X}$ is $\mathbb{S}^2$ the unit sphere in $\mathbb{R}^3$ and more generally a d-fold product of unit spheres $\mathbb{S}^2$.

3.3. Covariance function characteristics

A covariance function $k(\mathbf{x}, \mathbf{x}')$ is a mapping of a product space $\mathcal{X} \times \mathcal{X}$ to $\mathbb{R}$ which, briefly speaking, characterizes the smoothness of the GP model. The covariance function $k(\mathbf{x}, \mathbf{x}')$ must be positive semidefinite (cf. Rasmussen and Williams [RW06]). In most applications, the covariance function is stationary, i.e. it only depends on $\boldsymbol{\tau} = (\mathbf{x} - \mathbf{x}')$ and, more specifically, isotropic (i.e. a function only of $r = \|\mathbf{x} - \mathbf{x}'\|$). The function space to which the sample functions belong is characterized by the covariance function and in particular as regards continuity and differentiability. This feature is important as it characterizes the assumed smoothness of the functions we want to approximate. As the function $f(\mathbf{x})$ is random, continuity and differentiability are defined on a mean square (MS) basis [RW06]. A GP over $\mathbb{R}^D$ with stationary covariance function is continuous if and only if its covariance function is continuous at $\boldsymbol{\tau} = \mathbf{0}$. Furthermore, for such GPs, if a $2k$th-order partial order derivative of the covariance function exists and is finite at $\boldsymbol{\tau} = \mathbf{0}$ then the kth-order MS partial derivative exists for all $\mathbf{x} \in \mathbb{R}^D$. Therefore, the continuity and differentiability properties of stationary GPs are only determined by the properties of the covariance function at $\boldsymbol{\tau} = \mathbf{0}$. Note, however, that MS continuity of the covariance function does not imply that the sample functions of the GP are continuous. A detailed theoretical analysis of this subject matter can be found in textbooks [Abr97, Adl81].

In the following, we only focus on isotropic covariance functions, which are used in most cases. This choice of isotropic covariance function corresponds to the most general setting, where all directions are equivalent a priori and no prior assumption regarding directional knowledge is made. Such functions are met in many other theories and are called kernel functions. Consequently, many positive definite isotropic kernels have been proposed in the literature in various application contexts (see e.g., Rasmussen and Williams [RW06]). All the isotropic kernels presented hereafter are designed such that $k(\mathbf{0}) = 1$. Therefore, covariance functions are derived from these kernels by simply multiplying them by σ_f^2, the pointwise variance of the isotropic GP. The *squared exponential* kernel k_s is certainly the most popular isotropic covariance function. It is defined as

$$k_s(r) = \exp\left(-\frac{r^2}{2l_s^2}\right) \tag{2}$$

with l_s defining the *length-scale* parameter. As this kernel is infinitely differentiable, the GP produced with this covariance function is mean squared (MS) differentiable at all orders, which results in very smooth sample functions. This smoothness property might be unsuited to model the sharp variations of illumination functions. Various other classes of covariance functions have been proposed in the literature so as to allow kernels with different smoothness levels as for example the following piecewise polynomial covariance functions [Wen04] that applies to GP defined on $\mathbb{R}^D$:

$$kp_0(r) = \left(1 - \frac{r}{l_s}\right)_+^j, \quad j = \left\lfloor \frac{D}{2} \right\rfloor + 1 \tag{3}$$

$$kp_1(r) = \left(1 - \frac{r}{l_s}\right)_+^{j+1} \left((j+1)\frac{r}{l_s} + 1\right), \quad j = \left\lfloor \frac{D}{2} \right\rfloor + 2 \tag{4}$$

with a length-scale $l_s \leq 1$. The first one (Equation (3)) is continuous but not differentiable while the second one is twice continuously differentiable. The corresponding GPs is thus not MS differentiable and once MS differentiable, respectively. Our interest in considering these kernels is that they have a *compact support*. This leads to a sparse covariance matrix which, in its turn, enables the use of computationally efficient techniques.

Isotropic covariance functions defined on $\mathbb{R}^3$ are also valid on the $\mathbb{S}^2$ sphere. In this case, the Euclidean distance $r = \|\mathbf{x} - \mathbf{x}'\|$ with $\mathbf{x}, \mathbf{x}' \in \mathbb{S}^2$ is known as the chordal distance on the sphere. Chordal distance r are related to geodesic distance θ through

$$r = \sqrt{2(1 - \cos(\theta))} \quad \text{or :} \tag{5}$$

$$r = 2\sin\left(\frac{\theta}{2}\right). \tag{6}$$

These so-called chordal covariance functions have the same continuity and differentiability properties as when defined on $\mathbb{R}^3$ [GF16]. Therefore, the polynomial covariance functions defined above are valid on $\mathbb{S}^2$ by setting $D = 3$ in Equations (3) and (4). Covariance functions can also be directly defined on $\mathbb{S}^2$ but such covariance functions are generally not valid on $\mathbb{R}^3$. Various isotropic kernels specifically defined for the $\mathbb{S}^2$ sphere have been proposed in the literature [KSKM13, GF16, FS98, BSSW14, SW06]. For example, the generalized distance (GD) kernel [BSSW14]

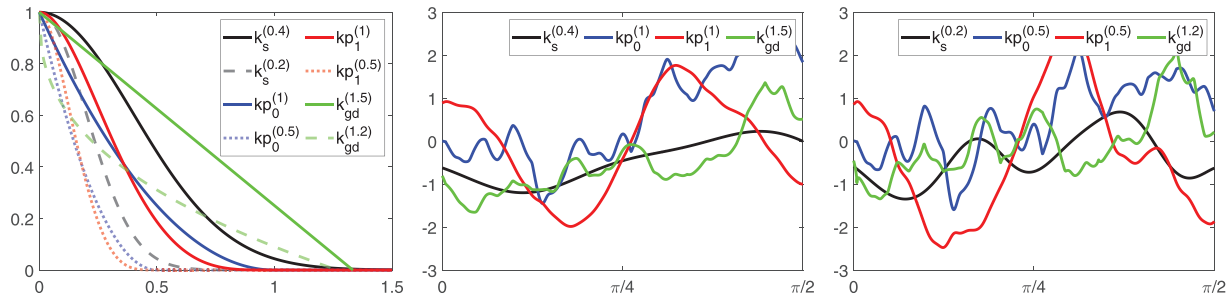
$$k_{gd}(r) = 1 - s_g\left(\frac{r}{2}\right)^{2s_g - 2} \quad \text{with } 1 < s_g < 2, \tag{7}$$

s_g being a smoothness parameter, has been used in Marques *et al.* [MBB19] to derive optimum quadrature weights for shading integrals. The GPs produced with this covariance function are MS continuous but not differentiable. However, k_{gd} is nearly first order differentiability as the smoothness parameter s approaches 2.

Figure 1(a) illustrates the aforementioned covariance functions on the $\mathbb{S}^2$ sphere, using different parameters l_s and s_g (with a large and a small value). Figures 1(b) and 1(c) show spherical sample functions drawn from GPs using the covariance functions shown in Figure 1(a). Note how the shape of the covariance functions is altered by their parameter (l_s or s_g). However the effect of these parameters is very different. The scaling parameter l_s has no effect on the MS differentiability of the GP whereas the parameter s_g, used in the GD kernel (Equation (7)), enables an effective control of the GP smoothness. This property does not clearly appear from the covariance function plots in Figure 1(a), nor from the GP realizations shown in Figures 1(b) and 1(c). However, as we show later, our spectral analysis presented in Section 4.1 makes this difference explicit, which will contribute to understand which kernel is more suitable to model incident radiance functions.

3.4. Regression

Kernel techniques and GPs can be used as a solution to many basic problems such as classification [RW06] or probability density



(a) Covariance functions on $\mathbb{S}^2$ (b) Sample functions on $\mathbb{S}^2$ (c) Sample functions with smaller l_s and s_g

Figure 1: *Left (a): example of covariance functions on the $\mathbb{S}^2$ sphere. The numbers in-between brackets show the used l_s for k_s (Equation (4)), k_{p0} (Equation (3)) and k_{p1} (Equation (2)), and the used s_g for the case of k_{gd} (Equation (7)). Centre (b) and right (c): spherical sample functions $f(\theta, \phi)$ with different l_s and s_g values. The plots show function values for $\phi = 0$ and $\theta = [0, \pi/2]$. Note the smoothness of the spherical function drawn using the squared exponential kernel k_s.*

estimation [SW06]. In this section, we focus on the problem of unknown functions approximation and again, many kernel-based theories such as RKHS, regularization networks, spline models, etc., have been proposed for this purpose in the literature. In this paper, our approach is based on non-parametric Bayesian regression which we think is most appropriate for modelling illumination functions given its efficiency and flexibility. In particular, GP regression can efficiently deal with highly unstructured data such as those produced by adaptive sampling techniques. In Rasmussen and Williams [RW06], the interested reader will find detailed theoretical analysis of the relationships between GP regression and other kernel-based models.

In a Bayesian framework, a regression is basically an inference problem. Given a set $\mathcal{D}$ of n observations such that $\mathcal{D} = \{(t_i, \mathbf{x}_i)|i = 1, \ldots, n\}$, $X = \{\mathbf{x}_i\}$ being the set of input vectors (i.e. the observed locations) and $\mathbf{t} = [t_1, \ldots, t_n]^t$ the corresponding target values (i.e. the observed process values), the inference of the function value $f(x)$ is described by the posterior probability distribution

$$P(f(\mathbf{x})|\mathbf{t}, X) = \frac{P(\mathbf{t}|f(\mathbf{x}), X)\, P(f(\mathbf{x}))}{P(\mathbf{t}|X)}. \tag{8}$$

In the usual parametric regression, the prior on $f(\mathbf{x})$ consists in modelling the function as a weighted sum of some basis functions and the problem amounts to finding the weights that give the most probable estimate of the unknown function. In non-parametric regression, the prior is directly placed on a space of functions, which we choose to be a GP. This model is the simplest to enable an analytic solution to Equation (8). In this way, we have access to a function space of infinite dimension instead of the finite vector space of the parametric model.

Let $\mathcal{GP}(\bar{f}(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$ be our GP prior on $f(\mathbf{x})$. We assume that the target values t_i are noisy values of the unknown function $f(\mathbf{x})$ such that $t_i = f(\mathbf{x}_i) + \epsilon$, ϵ being samples of an additive IID Gaussian noise with variance σ_n^2. With this Gaussian model of the prior, the posterior distribution is then also a GP as a result of the conjugate prior property of Gaussian distributions. Given a test input vector $\mathbf{x}_*$ (i.e. the location at which a prediction is inferred), the MAP (maximum a posteriori) estimate of $f(\mathbf{x}_*)$ is $f_*(\mathbf{x}_*) = E[f(\mathbf{x}_*)|\mathbf{t}, X]$ given by [RW06]

$$f_*(\mathbf{x}_*) = \bar{f}(\mathbf{x}_*) + \mathbf{k}(\mathbf{x}_*)\big(K + \sigma_n^2 I\big)^{-1}\big(\mathbf{t} - \bar{f}(X)\big), \tag{9}$$

where $\mathbf{k}(\mathbf{x}_*) = [k(\mathbf{x}_*, \mathbf{x}_1), \ldots, k(\mathbf{x}_*, \mathbf{x}_n)]$, K is the covariance matrix whose coefficients are the $k(\mathbf{x}_i, \mathbf{x}_j)$ covariances, I is the $n \times n$ identity matrix and $\bar{f}(X) = [\bar{f}(\mathbf{x}_1), \ldots, \bar{f}(\mathbf{x}_n)]^t$. Equation (9) can be rearranged as follows to show that this prediction is built from a weighted sum of kernels function:

$$f_*(x_*) = \bar{f}(\mathbf{x}_*) + \sum_{i=1}^{n} \alpha_i k(\mathbf{x}_*, \mathbf{x}_i), \tag{10}$$

with

$$\boldsymbol{\alpha} = \big(K + \sigma_n^2 I\big)^{-1}\big(\mathbf{t} - \bar{f}(X)\big).$$

When the kernel is isotropic, Equation (10) leads to a RBF expansion of the approximating function. Note, however, that, although this expansion is limited to n terms, the bandwidth of $f_*(x_*)$ in the spectral domain is infinite if the kernel bandwidth is infinite.

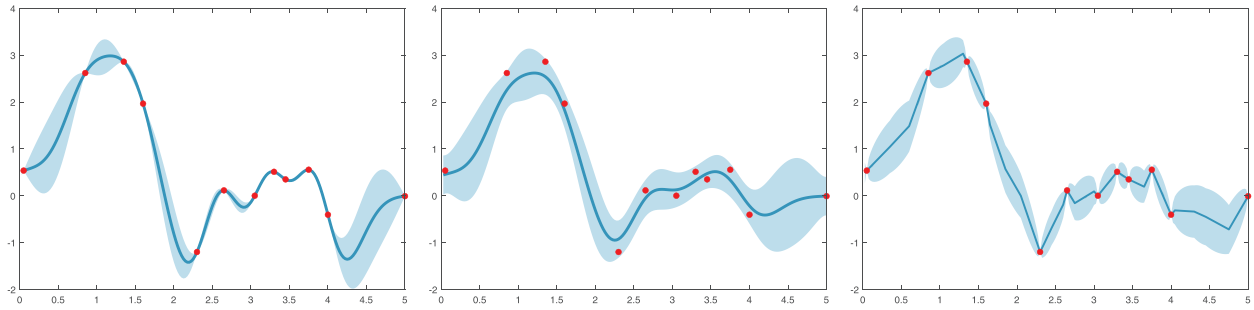
For most applications, a constant GP prior mean value $\bar{f}(\mathbf{x}) = f_c$ is appropriate. Setting $f_c = 0$ is often sufficient, but, in some cases, better results are obtained by inferring the mean value from the observed data. When the observed locations are uniformly distributed among the observation space, it is sufficient to take the arithmetic mean of the observations for f_c, i.e. $f_c = \sum_i t_i/n$. In Rasmussen and Williams [RW06], the authors provide a more general solution that leads to the following prediction of the constant mean value:

$$f_c = \frac{HQ^{-1}\mathbf{t}}{HQ^{-1}H^t} \quad \text{with } H = [1, \ldots, 1] \quad \text{and} \quad Q = K + \sigma_n^2 I. \tag{11}$$

The variance of the $\bar{f}_*$ estimate is given by

$$V(\bar{f}_*, \mathbf{x}_*) = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}(\mathbf{x}_*)(K + \sigma^2 I)^{-1}\mathbf{k}(\mathbf{x}_*)^t. \tag{12}$$

In this equation, the first term $k(\mathbf{x}_*, \mathbf{x}_*)$ represents the prior variance estimate while the second one represents the variance decrease due to observation data. Note that this variance estimate, does not depend on the target values $\mathbf{t}$ but only on the observed locations X.



(a) k_s, $\sigma_n' = 0$ (b) k_s, $\sigma_n' = 0.2$ (c) kp_0, $\sigma_n' = 0$

Figure 2: *Examples of one-dimensional regressions obtained with the squared exponential k_s and polynomial kp_0 covariance functions. Data points are marked in red. The blue shaded areas show the amplitude of the standard-deviation of error. The length scale is $l_s = 0.3$ and $l_s = 1$ for the k_s and kp_0 covariances, respectively.*

Indeed, this estimate assumes that the 'amplitude' information is provided by the prior covariance function. Therefore, this variance can be used to estimate error-bounds insofar as the unknown function belongs to the function space of the prior GP. Most often, the smoothness of the function to be approximated is unknown and error bounds based on Equation (12) may be very inaccurate. In Schaback and Wendland [SW06], the authors propose a more general theoretical analysis of approximation error based on the theory of Sobolev spaces.

Isotropic covariance functions can be written under the form $k_r(r) = \sigma_f^2 k(r)$ with $k(0) = 1$, $k(r)$ being a kernel such as those discussed in section 3.3. In this case, Equations (9) and (12) can be expressed as functions of the normalized kernel as suggested in Marques *et al.* [MBR*13b]:

$$f_*(\mathbf{x}_*) = \bar{f}(\mathbf{x}_*) + \mathbf{k}(\mathbf{x}_*)\Big(K + \sigma_n'^2 I\Big)^{-1}\big(\mathbf{t} - \bar{f}(X)\big) \qquad (13)$$

$$V(f_*, \mathbf{x}_*) = \sigma_f^2\Big(1 - \mathbf{k}(\mathbf{x}_*)\Big(K + \sigma_n'^2 I\Big)^{-1}\mathbf{k}(\mathbf{x}_*)^t\Big), \qquad (14)$$

where $\sigma_n' = \sigma_n/\sigma_f$ is the *noise ratio* and the covariance matrix K and the vector $\mathbf{k}(\mathbf{x}_*)$ are both computed with the normalized kernel instead of k_r.

Once a covariance function has been chosen, given a training set of n data points $\mathcal{D} = \{\mathbf{t}, X\}$, the set of hyperparameters $\boldsymbol{\theta}$ can be estimated by a maximum likelihood (MLE) method applied to the log likelihood function of the GP [RW06]. The log likelihood function is given by

$$\ln p(\mathbf{t}|X, \boldsymbol{\theta}) = -\frac{1}{2}\mathbf{t}^t Q^{-1}\mathbf{t} - \frac{1}{2}\ln|Q| - \frac{n}{2}\ln 2\pi\ , \qquad (15)$$

where $Q = K(X, X) + \sigma_n^2 I$. The likelihood function $p(\mathbf{t}|X, \boldsymbol{\theta})$ expresses the probability of occurrence of the observed target values t given the observed locations X and the hyperparameters $\boldsymbol{\theta}$.

Figure 2 shows examples of regression obtained with various priors and the same set of 1D data points. As expected, the predicted function is much smoother with the squared exponential covariance and this smoothness increases with a noise ratio $\sigma_n' = 0.2$. When $\sigma_n' = 0$, $V(\bar{f}_*, \mathbf{x}_i) = 0$ and $\bar{f}_*(\mathbf{x}_i) = t_i$, that is, the error variance is zero at all observed locations $\mathbf{x}_i$. In this case, $\bar{f}_*$ is an interpolant of f. The kp_0 covariance produces a predicted function made of connected line segments since kp_0 is the hat function in 1D. This figure shows that the predicted function can be very different depending on the prior, i.e. the covariance function and the related hyperparameters $\boldsymbol{\theta} = \{l_s, \sigma_n', \sigma_f\}$. Finding an appropriate set-up for the GP model is thus decisive to correctly approximate incident radiance functions over the $\mathbb{S}^2$ sphere.

4. Our Theoretical Contributions

In this section, we present our three theoretical contributions, which consist of: (i) a spectral analysis of the considered covariance functions; (ii) a simplified likelihood expression for hyperparameters learning and (iii) fast SH coefficients computation and rotation. For the sake of clarity, the new equations, which constitute our most important results, are boxed.

4.1. Spectral analysis of covariance functions

Stationary covariance functions allow spectral representations through Fourier transform. In the following, we analyse the connections between these spectral representations and the smoothness properties of sample functions. In the case of spherical functions, spectral analysis must be performed through SH expansions. The SH expression of an isotropic kernel (i.e. its spectral density) reduces to a Legendre polynomials expansion with [SW06]:

$$k_\theta(\theta) = \sum_{l=0}^{\infty} a_l \frac{2l+1}{4\pi} P_l(\cos\theta), \qquad (16)$$

where $\theta = \arccos(r)$ and $P_l(x)$ is the Legendre polynomial of degree l. Then, from Equation (5), $k_\theta(\theta) = k(2\arcsin(r/2))$. The spectral density of an isotropic covariance function is then characterized by the sequence of a_l coefficients given by

$$a_l = 2\pi \int_{-\pi/2}^{\pi/2} k_\theta(\theta) P_l(\cos\theta)\sin\theta\, d\theta \quad \text{or :} \qquad (17)$$

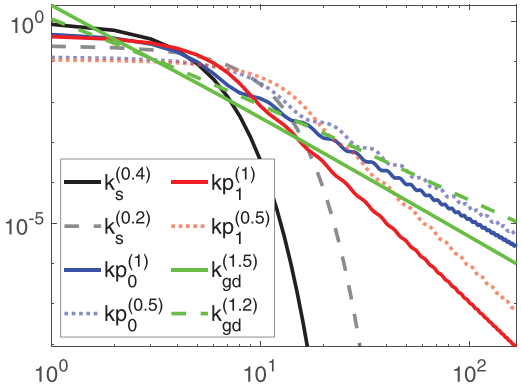


Figure 3: *Spectral densities (a_l) as a function of the SH degree l, for the spherical covariance functions shown in Figure 1(a). The numbers in-between brackets show the used l_s for k_s (Equation (4)), k_{p0} (Equation (3)) and k_{p1} (Equation (2)), and the used s_g for the case of k_{gd} (Equation (7)).*

$$a_l = 2\pi \int_{-1}^{1} k_z(z) P_l(z)\, dz, \tag{18}$$

with $k_z(z) = k_\theta(\arccos(z)) = k(\sqrt{2(1-z)})$ from Equation (5). The a_l coefficients must be non-negative for the covariance function to be positive semi-definite. The Legendre polynomial expansion of some familiar RBFs can be found in Hubbert and Baxter [HB00]. For the squared exponential covariance function, we obtain

$$a_l = 2\pi e^{-\gamma} \sqrt{\frac{2\pi}{\gamma}} I_\nu(\gamma) \quad \text{with } \gamma = \frac{1}{l_s^2}, \quad \nu = l + \frac{1}{2}, \tag{19}$$

where I_ν is the modified Bessel function of the first kind. The expansion of the GD kernel of Equation (7) can be derived from the potential splines case in Hubbert and Baxter [HB00] to give

$$a_0 = 0 \tag{20}$$

$$a_l = -4\pi \frac{\Gamma(1 - s_g + l)\Gamma(1 + s_g)}{\Gamma(1 + s_g + l)\Gamma(1 - s_g)}, \quad l \geq 1, \tag{21}$$

where $\Gamma()$ is the extended factorial function. As regards kp_0 and kp_1, no analytic formulation exists for the spectral density of polynomial covariance functions. However, their a_l coefficients can be obtained by numerical integration of Equation (18).

Figure 3 shows the spectral densities produced by the squared exponential, polynomial and GD kernels illustrated in Figure 1(a). It can be seen that the spectral density of the GD kernels k_{gd} (shown in green) is almost a straight line. In fact, this kernel has been designed for this purpose and it can be shown that the slope of the straight line is exactly $\beta = 2s_g$ [BSSW14]. The slope is thus $\beta = 2.4$ when $s_g = 1.2$ (green dashed line), and $\beta = 3$ when $s_g = 1.5$ (green continuous line). As for the polynomial kernels kp_0 (in blue) and kp_1 (in red), they exhibit an asymptotic slope $\beta = 3$ and $\beta = 5$, respectively. Note that, as opposed to the GD kernels (in green), the asymptotic value of the slopes for the polynomial kernels remain constant when the kernel parameter l_s is changed. Finally, a similar observation can be made regarding the squared exponential kernels k_s (in black), since the asymptotic rate of decay of their SH coefficients (a_l) tends to infinity regardless of the used lengthscale l_s (black continuous line vs. black dashed line).

The importance of the observations made above is given by the connection between the asymptotic slope β and the smoothness of the GP. Let us recall that the MS differentiability properties of the GP depend on the differentiability of the covariance kernel at $\boldsymbol{\tau} = \mathbf{0}$ (cf. Section 3.3). These properties lead to the result that a GP over $\mathbb{S}^2$ is k-times MS continuously differentiable if $k < \beta/2 - 1$ [BSSW14, NW02, LS15]. When applied to Figure 3, this result has the striking consequence that the k_{gd} kernel is the only one which allows an active control of the MS differentiability properties of the GP, since it is the single considered kernel whose asymptotic slope can be explicitly adjusted.

As regards the actual smoothness properties, in the case of the GD kernel k_{gd}, we have that within the range of allowed s_g values ($1 < s_g < 2$), the asymptotic slope satisfies $2 < \beta < 4$, which means that the GPs produced with the GD kernel are MS continuous but not MS differentiable. However, the GP smoothness can vary between these two extremes and the in-between differentiability corresponds to the so-called *weak derivatives* defined in the theory of Sobolev spaces. That is why such a kernel is also called *Sobolev kernel* [SW06]. It is interesting to note that the polynomial kernel kp_0, has a similar slope to that of k_{gd} with $s_g = 1.5$. This feature is more noticeable when l_s increases, since the effect of the compact support diminishes. Therefore the GPs produced with kp_0 and k_{gd} with $s_g = 1.5$ have similar smoothness characteristics. As regards the kp_0 polynomial kernel, its slope $\beta = 5$ means that it is one-time MS continuously differentiable, which confirms the differentiability properties mentioned in Section 3.3. Finally, the squared exponential kernel (k_s), as it is well known, leads to GPs, which are MS derivable at all orders.

Driven by the above analysis, in Section 5, we will show that both the k_s and kp_1 covariance functions are too smooth to model radiance functions, and that the k_{gd} and kp_0 kernels are clearly more appropriate. Indeed, the first degree of differentiability is a too high expectation for radiance functions.

4.2. A simplified likelihood expression

The goal of this section is to derive a new likelihood expression, which reduces the number of hyperparameters to learn from three ($\boldsymbol{\theta} = \{l_s, \sigma_n', \sigma_f\}$) to two ($\boldsymbol{\theta} = \{l_s, \sigma_n'\}$). Indeed, as shown in the following, σ_f can be derived analytically in the case of anisotropic functions as explained below. To this end, we apply the σ_f^2 factorization used for Equations (13) and (14). Equation (15) then becomes

$$\ln p(\mathbf{t}|X, \boldsymbol{\theta}) = -\frac{1}{2\sigma_f^2} \mathbf{t}^t Q'^{-1} \mathbf{t} - \frac{n}{2} \ln \sigma_f^2 - \frac{1}{2} \ln |Q'| - \frac{n}{2} \ln 2\pi \tag{22}$$

where $Q' = K(X, X) + \sigma_n'^2 I$. The partial derivative w.r.t. σ_f^2 is then

$$\frac{\partial}{\partial \sigma_f^2} \ln p(\mathbf{t}|X, \boldsymbol{\theta}) = \frac{1}{2\sigma_f^4} \mathbf{t}^t Q'^{-1} \mathbf{t} - \frac{n}{2\sigma_f^2} \tag{23}$$

which is zero for:

$$\sigma_f^2 = \frac{\mathbf{t}^t Q'^{-1}\mathbf{t}}{n} \tag{24}$$

Integrating this σ_f value in Equation (22), we obtain a new expression of the likelihood function as follows:

$$\ln p(\mathbf{t}|X,\boldsymbol{\theta}) = -\frac{n}{2} - \frac{n}{2}\ln\left(\frac{\mathbf{t}^t Q'^{-1}\mathbf{t}}{n}\right) - \frac{1}{2}\ln|Q'| - \frac{n}{2}\ln 2\pi \tag{25}$$

With this expression, the hyperparameters optimization gets simpler, hence reducing the chances for the optimizer to get stuck in local minima.

4.3. Enabling fast SH coefficients computation and rotation

In the following, we show that a non-parametric Bayesian regression can be as well expressed in the spectral domain thus enabling a direct derivation of the spherical harmonic (SH) expansion of the approximating function. This is possible since an isotropic covariance function can be represented in the spectral domain as we have seen in Section 4.1. The spectral domain analysis of Bayesian regression based on GP has been studied in different works [Pac07, WAP18] but their method uses discrete Fourier transform (DFT) to represent the spectral domain, which is not suited to spherical functions. The spectral analysis of spherical functions is based on the SH expansion:

$$f(\mathbf{x}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{m=l} f_{lm} Y_l^m(\mathbf{x}), \tag{26}$$

where $\mathbf{x} \in \mathbb{S}^2$, $f \in L^2(\mathbb{S}^2)$, $\{f_{lm}\}$ are the SH coefficients and $\{Y_l^m(\mathbf{x})\}$ are the SH basis functions. So as to express this equation in a matrix-vector product form, it is necessary to limit the expansion by setting a cutoff value L to the degree l. Let us thus define the SH coefficient vector $\mathbf{c} = [c_1, \ldots, c_M]^t$ such that $c_i = f_{lm}$ for $i = l^2 + l + m + 1$ and $M = (L+1)^2$. Similarly, the basis function vector $\mathbf{b}(\mathbf{x})$ is defined by $\mathbf{b}(\mathbf{x}) = [b_1, \ldots, b_M]^t$ such that $b_i = Y_l^m(\mathbf{x})$. Equation (26) can be rewritten as follows:

$$f_L(\mathbf{x}) = \mathbf{b}^t(\mathbf{x})\,\mathbf{c}, \tag{27}$$

where $f_L(\mathbf{x})$ is a band-limited approximation of $f(\mathbf{x})$. As for kernels SH expansion, using $\cos\theta = \mathbf{x}\cdot\mathbf{y}$ with $\{\mathbf{x},\mathbf{y}\} \in \mathbb{S}^2$ in Equation (16), we have

$$k(\mathbf{x},\mathbf{y}) = \sum_{l=0}^{\infty} a_l \frac{2l+1}{4\pi} P_l(\mathbf{x}\cdot\mathbf{y}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{m=l} a_l Y_l^m(\mathbf{x}) Y_l^m(\mathbf{y}) \tag{28}$$

in application of the addition theorem of SH. A matrix-vector form of Equation (28) can be built in the same way as for Equation (27). For this purpose, let us define an $M \times M$ diagonal matrix G whose coefficients g_{ij} are given by

$$g_{ii} = a_l \text{ when } i \in [l^2+1, (l+1)^2] \forall i \in [1, M] \tag{29}$$

$$g_{ij} = 0 \text{ when } i \neq j \tag{30}$$

Then, we have

$$k_L(\mathbf{x},\mathbf{y}) = \mathbf{b}^t(\mathbf{x})\,G\,\mathbf{b}(\mathbf{y}), \tag{31}$$

where $k_L(\mathbf{x},\mathbf{y})$ is a band-limited approximation of $k(\mathbf{x},\mathbf{y})$. With the same band-limited approximation, the covariance vector $\mathbf{k}(\mathbf{x}_*)$ in Equation (9) can be expanded as follows:

$$\mathbf{k}_L(\mathbf{x}_*) = \mathbf{b}^t(\mathbf{x}_*)\,GB, \tag{32}$$

where B is a $M \times n$ matrix giving the SH basis function values at each target location $\mathbf{x}_i$, i.e. the ith column of B corresponds to the vector $\mathbf{b}(\mathbf{x}_i)$. Then, using Equations (27) and (32), Equation (9) becomes

$$f_*(\mathbf{x}_*) = \mathbf{b}^t(\mathbf{x}_*)\Big[\bar{\mathbf{c}} + GB\big(K + \sigma_n^2 I\big)^{-1}\big(\mathbf{t} - \bar{f}(X)\big)\Big], \tag{33}$$

where $\bar{\mathbf{c}}$ is the vector of SH coefficients of $\bar{f}$. Considering Equation (27), the SH coefficients of f_* can be identified as follows:

$$\mathbf{c}_* = \bar{\mathbf{c}} + GB\big(K + \sigma_n^2 I\big)^{-1}\big(\mathbf{t} - \bar{f}(X)\big). \tag{34}$$

We will show in the following that Equation (34) provides an efficient method to obtain quite accurate SH coefficients of the unknown function f without any integral computation. The main computing load is due to the $T = GB(K + \sigma_n^2 I)^{-1}$ matrix product. However, this $M \times n$ matrix can be pre-computed for appropriate combinations of sample sets (e.g. Fibonacci grids [MBR*13a]) and cutoff SH degree L, which reduces the computation to a $\mathcal{O}(Mn)$ matrix-vector product.

Another benefit of the SH coefficients computation through Equation (34) is that it enables fast rotation of SH coefficients without the expensive computation of the Wigner D matrix [Aub13, BFB97]. Examining Equation (10) provides an interesting interpretation of Equation (34). Equation (10) represents the approximating function in terms of zonal basis functions expansion. Each of this zonal basis function $k(\mathbf{x}_*, \mathbf{x}_i)$ is rotated so as to be centred on a sampling location $\mathbf{x}_i$. As $k(., \mathbf{z})$ is zonal, the SH coefficients of the rotated function $k(., \mathbf{x}_i)$ is simply $k_{lm}(\mathbf{x}_i) = a_l Y_l^m(\mathbf{x}_i)$ where a_l is the l-degree SH coefficient of the kernel $k(\mathbf{x},\mathbf{y})$ (cf. Equation (28)). Indeed, all the $a_l Y_l^m(\mathbf{x}_i)$ products are performed by the matrix product GB in Equation (34). Therefore, if we want to rotate the whole set of SH coefficients so that the new z axis is z', we simply have to rotate the whole set of sample points $\mathbf{x}_i$ before computing the B matrix. So as to enable fast rotation of SH coefficients, Equation (34) can be rewritten as follows:

$$\mathbf{c}_* = \bar{\mathbf{c}} + GB\boldsymbol{\alpha} \tag{35}$$

where $\boldsymbol{\alpha}$ is given by Equation (10). Once the $\boldsymbol{\alpha}$ matrix is pre-computed, rotated sets of SH coefficients are then easily obtained as explained above. In Lessig *et al.* [LdWF12], a similar use of zonal function expansion is made except that a different kernel is used for each band, i.e. each SH degree. Therefore, a specific kernel-based expansion is computed for each band with a specific sample set whereas our method requires a single kernel and a single sample set. Furthermore, our method enables both the computation of SH coefficients and their rotation whereas the method described in Lessig *et al.* [LdWF12] only applies to SH coefficient rotation.

Table 2: *Average value of the learned hyperparameters for each kernel and for different numbers of samples. Two objective functions were used: likelihood (for maximum likelihood estimation, MLE); and root mean squared error (RMSE).*

	N	MLE		RMSE	
k_s	500	$l_s = 0.26$	$\sigma_n' = 0.69$	$l_s = 0.11$	$\sigma_n' = 0.57$
	1000	$l_s = 0.20$	$\sigma_n' = 0.65$	$l_s = 0.07$	$\sigma_n' = 0.49$
	2000	$l_s = 0.14$	$\sigma_n' = 0.62$	$l_s = 0.05$	$\sigma_n' = 0.43$
	4000	$l_s = 0.09$	$\sigma_n' = 0.55$	$l_s = 0.03$	$\sigma_n' = 0.41$
kp_0	500	$l_s = 1.40$	$\sigma_n' = 0.44$	$l_s = 1.06$	$\sigma_n' = 0.25$
	1000	$l_s = 1.42$	$\sigma_n' = 0.42$	$l_s = 0.96$	$\sigma_n' = 0.16$
	2000	$l_s = 1.23$	$\sigma_n' = 0.36$	$l_s = 0.89$	$\sigma_n' = 0.11$
	4000	$l_s = 1.00$	$\sigma_n' = 0.27$	$l_s = 0.87$	$\sigma_n' = 0.09$
kp_1	500	$l_s = 1.38$	$\sigma_n' = 0.64$	$l_s = 0.85$	$\sigma_n' = 0.36$
	1000	$l_s = 1.14$	$\sigma_n' = 0.55$	$l_s = 0.68$	$\sigma_n' = 0.24$
	2000	$l_s = 0.92$	$\sigma_n' = 0.49$	$l_s = 0.47$	$\sigma_n' = 0.21$
	4000	$l_s = 0.66$	$\sigma_n' = 0.41$	$l_s = 0.39$	$\sigma_n' = 0.16$
k_{gd}	500	$s_g = 1.41$	$\sigma_n' = 0.23$	$s_g = 1.40$	$\sigma_n' = 0.05$
	1000	$s_g = 1.44$	$\sigma_n' = 0.19$	$s_g = 1.41$	$\sigma_n' = 0.04$
	2000	$s_g = 1.41$	$\sigma_n' = 0.18$	$s_g = 1.68$	$\sigma_n' = 0.03$
	4000	$s_g = 1.38$	$\sigma_n' = 0.08$	$s_g = 1.51$	$\sigma_n' = 0.01$

5. Experiments

5.1. Hyperparameter learning through simplified MLE

Experimental set-up. To learn the optimal hyperparameters for each kernel, we have applied our simplified likelihood expression (Equation (25)) to determine the hyperparameter values which maximize the likelihood (*maximum likelihood estimation*, MLE). The incident radiance was simulated using eight different environment maps. The hyperparameters have been learned using sample sets of different sizes, ranging from 500 to 4000. This relatively large number of samples is required to capture the high-frequency features of the typical incident radiance functions encountered in rendering applications, and obtain a representative set of hyperparameter values. For each sample count, several runs of the learning function have been performed, using different sample locations, while keeping the sample distribution uniform over the sphere. To this end, for each learning run with N samples, the sample positions have been determined using a spherical Fibonacci point set of size N [MBR*13a], followed by a random rotation applied to all the point set. Then, we averaged the learned hyperparameters over the different runs. The results of the hyperparameter learning through MLE for each kernel are shown in the left-hand side of Table 2 (the detailed set of results for each environment map can be found in additional material). On the right-hand side of Table 2, the hyperparameters learned by minimizing the *RMSE* (root mean squared error) instead of the log likelihood. In this case, the RMSE was computed by measuring the error between the original (reference) image, and the image reconstructed from the samples using the tested hyperparameters through a GP-based regression (Equation (13)). Figure 4 shows iso-value plots of the log likelihood and RMSE as functions of the hyperparameters.

Comparing k_s and k_{gd}. Let us start by analysing the results in Table 2 for k_s (top) and k_{gd} (bottom), which are the two kernels that do not have a compact support. It is interesting to note that, for both likelihood- and RMSE-based learning, the lengthscale parameter l_s of the squared exponential kernel k_s exhibits a dependence on the number of samples: larger (i.e. denser) point sets lead to smaller l_s values. We can conjecture that this is due to the strong smoothness assumptions of the k_s kernel, which, we recall, is infinitely differentiable. Therefore, when the number of samples increases, the lengthscale of the model needs to be shortened so as to fit sharp variations of the unknown functions that were not detected at lower sampling densities. It is also worth noting that, for k_{gd}, the smoothness parameter remains roughly around $s_g \approx 1.5 \pm 0.1$ using both optimization criteria. The fact that s_g does not show a dependence on the number of samples is a strong indication that the model assumptions properly suit incident radiance functions with no hyperparameter adaptation. Finally, it is interesting to compare the results when using MLE and RMSE as objective function. These criteria are fundamentally different, since the likelihood gives the probability of the observations *given the model assumptions*, whereas these assumptions are not directly considered in the RMSE criterion. However, in the case of the k_{gd} kernel, both criteria lead to very similar objective functions, as shown in Figures 4(d) and 4(h). This result further reinforces the idea that the k_{gd} kernel assumptions properly suit incident radiance functions, and explains why, for this kernel, the hyperparameters found with both criteria are generally similar. In contrast, for the k_s kernel, likelihood- and RMSE-based optimization leads to substantial differences in l_s values due to the difference in the objective functions as shown in Figures 4(a) and 4(e). Indeed, the RMSE criterion favours strict data fitting whereas the likelihood criteria aims at a balanced solution between data fitting and complexity of the model, hence leading to a smoother solution (see Rasmussen and Williams [RW06] for more details).

Results for the kp_0 and kp_1 kernels. To interpret the results for the kp_0 kernel, let us first recall that the rate of decay of its SH coefficients is independent of the value of its lengthscale parameter l_s. Furthermore, this rate of decay is always equal to that of k_{gd} with $s_g = 1.5$, as shown previously in Figure 3 (Section 4.1). Consequently, and given that the optimal s_g value for k_{gd} is around 1.5, this suggests that the kp_0 kernel is also suitable to represent incident radiance functions. The optimization results of Table 2 show that the lengthscale parameter l_s, which controls the kernel support size, is optimal at around $l_s \approx 1.2$. However, a closer analysis shows that this value is not critical, as depicted in Figures 4(b) and 4(f). Indeed, both figures show that there is a large set of l_s values for which both the likelihood and the RMSE are nearly optimal. A relevant question is thus how small can the kernel support be such that, on the one hand, we can have a sparse covariance matrix K leading to computational advantages and, on the other hand, we do not incur in a RMSE or likelihood penalization. Our results suggest that using $l_s = 0.4$ is a good compromise between both constraints. Finally, a similar l_s value can also be identified regarding the kp_1 kernel, as seen in Figures 4(c) and 4(g). Note, however, that, compared to kp_0, the RMSE and likelihood functions exhibit fewer similarities.

Fast hyperparameter set-up. Based on our extensive learning results (see additional material), which are summarized in Table 2, we propose a set of general purpose hyperparameters for each of the considered kernels. Our goal is to provide a set of hyperparameters, which should perform well for most incident radiance functions,

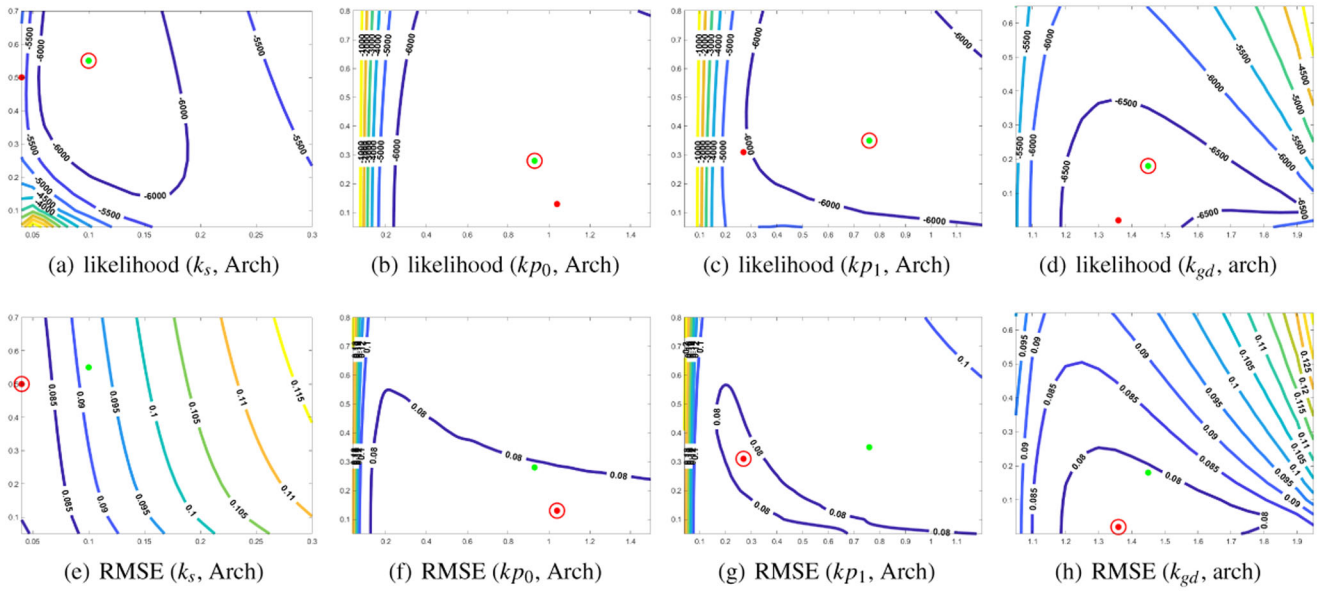


Figure 4: *Contour plots for the likelihood (top row) and RMSE (bottom row) for each kernel as a function of the hyperparameters, using the Arch environment map. The* x*-axis corresponds to s_g in the case of the k_{gd} kernel, and l_s for the other kernels. The* y*-axis shows the σ_n' value. Each curve corresponds to a set of points for which the RMSE or likelihood has the same value (isovalue). The line colour encodes the isovalue, where red represents a high value and blue represents a low value. In each image, the minimum function value found by optimization is marked with a coloured dot, surrounded by a red circle. The other coloured dot in the image marks the minimum according to the other criterion (RMSE or likelihood) for the same kernel. All images have been generated using 4000 samples.*

Table 3: *Recommended hyperparameters for each considered kernel. For k_s, the dependence of l_s on the number of samples N is modelled by the function $f_g(N)$ defined in Equation (36).*

	k_s	kp_0	kp_1	k_{gd}
l_s, s_g	$f_g(N)$	0.4	0.4	1.5
σ_n'	0.5	0.1	0.3	0.1

hence avoiding running an explicit learning step. These proposed values are as shown in Table 3. Note that, for the squared exponential k_s kernel, optimal l_s value can be obtained by the function f_g given below that accounts for the dependence on N:

$$f_g(N) = e^{-0.46 \ln N + 0.56}. \tag{36}$$

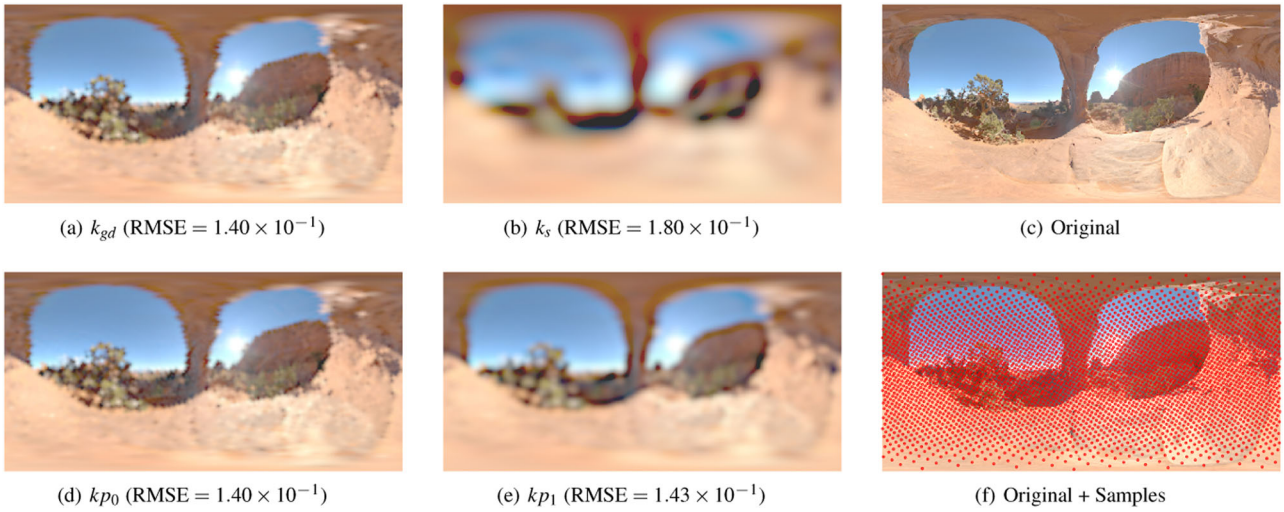
This function results from approximating the dependence of l_s on the number of samples N with a polynomial of 1st degree in log space, using the l_s values of Table 2. In the next section, we will use these hyperparameters to perform GP-based regression and compare the results obtained using the different considered kernels.

5.2. Regression

The results for regression in the original spherical domain using a set of 4000 uniformly distributed samples are shown in Figure 5, for the Arch environment map. It is important to note that when the number of samples is large (e.g. 4000), the sharp transitions of the incident radiance function are, in general, sufficiently well-captured, which allows assessing the capability of the GP model to fit these sharp transitions. Results for various environment maps are given in additional material, as well as regression results for a lower number of samples. We show the results for the four tested kernels parametrized with the hyperparameter settings of Table 3, together with the original environment map and the used sampling pattern. All kernels use exactly the same samples. We can observe that the squared exponential kernel k_s (Equation (2)) is the one that delivers poorer visual results. This is confirmed by the RMSE values achieved using the k_{gd} kernel, which are significantly higher than those using the other kernels. Such a performance is explained by the strong smoothness assumptions when using k_s, which do not fit the typical sharpness of incident radiance functions in practice. The kp_1 kernel, being also a smooth kernel, also suffers from the same problem but to a smaller extent. This is apparent both visually and in terms of RMSE. Finally, the k_{gd} and kp_0 kernels deliver the best results, which confirms the findings of Section 5.1.

5.3. GP-based SH coefficients computation

In Section 4.3, we have seen that the SH coefficients of the approximating function resulting from a Bayesian regression can be directly obtained using Equation (34). This also provides an efficient method to compute the SH coefficients of a radiance functions as mentioned above. Our goal in this section is to put this method in practice and assess its results.



(a) k_{gd} (RMSE $= 1.40 \times 10^{-1}$) (b) k_s (RMSE $= 1.80 \times 10^{-1}$) (c) Original

(d) kp_0 (RMSE $= 1.40 \times 10^{-1}$) (e) kp_1 (RMSE $= 1.43 \times 10^{-1}$) (f) Original + Samples

Figure 5: *Regression results for the Arch environment map using different kernels and 4000 samples uniformly distributed over the sphere. The sample positions are marked in red in (f). The root mean squared error (RMSE) using each kernel is shown in-between brackets. The used hyperparameters for each kernel are the general purpose hyperparameters provided in Table 3.*

Our SH computation method consists in using a precomputed transform matrix T of size $M \times n$, as explained in Section 4.3. This transform matrix enables direct computation of a vector $\mathbf{c}_*$ of $M = (L+1)^2$ SH coefficients from a set of n samples, with location $\mathbf{x}_i$ and values t_i, and L being the maximum degree of the SH coefficients. Recall that

$$T = GB\big(K + \sigma_n^2 I\big)^{-1} \tag{37}$$

and

$$\mathbf{c}_* = \bar{\mathbf{c}} + T\big(\mathbf{t} - \bar{f}(X)\big). \tag{38}$$

These equations express the regression equation (Equation (9)) under a spectral form. The 'spectral' terms are: (i) the G matrix defined by Equations (29) and (30), which is a $M \times M$ diagonal matrix; (ii) the B matrix, which is the matrix of SH basis functions value $Y_l^m(\mathbf{x}_i)$ and (iii) the vector $\bar{\mathbf{c}}$ of SH coefficients of the mean function $\bar{f}(\mathbf{x})$. As our method is to be applied with no prior knowledge on $f(\mathbf{x})$ other than the GP prior model $\mathcal{GP}(\bar{f}(\mathbf{x}) = f_c,\ k(\mathbf{x}, \mathbf{x}'))$, uniformly-distributed sample sets are most appropriate for the n observed locations $\{\mathbf{x}_i\}$. In our experiments, we use Fibonacci sample sets [MBR*13a]. In this case, we can take $f_c = \sum_i f(\mathbf{x}_i)/n$ for the constant mean value and thus the vector $\bar{\mathbf{c}}$ in Equation (38) is reduced to $\bar{c}_1 = f_{00} = 2f_c\sqrt{\pi}$ and $\bar{c}_i = 0$ for $i > 1$. As for the covariance function, we have used the GD kernel k_{gd} (Equation (7)) since its SH coefficients a_l are known under an analytic form (Equation (21)), which facilitates the computation of the G matrix. To assess the accuracy of the SH coefficients $\mathbf{c}_*$ of the approximated function, we use an error criterion based on the $\mathbb{L}^2$ norm: $\|f\| = \sum_l \sum_m |f_{lm}|^2 = \sum_i |c_i|^2$ where the c_i are the exact SH coefficients. The relative SH approximation error E_r is evaluated as follows:

$$E_r = \frac{\sum_{i=1}^{M} |c_{*i} - c_i|^2}{\sum_{i=1}^{M} |c_i|^2}. \tag{39}$$

In this equation, the numerator expresses the $\mathbb{L}^2$ norm of the function $(f_{*L} - f_L)$ giving the deviation between the function f_L reconstructed with the reference SH coefficients c_i and the approximating function f_{*L} reconstructed with the c_{*i} SH coefficients. Only the SH coefficients of the luminance component (i.e. a scalar value) is considered in the following. The reference SH coefficients c_i are computed by the classic inner product integration method (labelled IPI) with quasi Monte-Carlo integration using 1.5×10^6 samples. To eliminate the effect of spurious correlations between the sampling pattern and the radiance function, the relative error E_r is averaged over 30 evaluations corresponding to 30 different rotations of the sampling pattern. Only rotations about the z axis are used to avoid too complex rotation corrections of the SH coefficients (i.e. only the azimuth ϕ of the samples location is modified).

The results for the estimation of the SH coefficients using IPI (QMC-based) and our GP-based approach are shown in Figure 6. The figure shows the relative error $\mathscr{E}$ of the IPI method with respect to the error obtained with the GP-based approach, such that

$$\mathscr{E} = \frac{\left|E_r^{(IPI)} - E_r^{(GP)}\right|}{E_r^{(GP)}} \times 100, \tag{40}$$

where $E_r^{(IPI)}$ and $E_r^{(GP)}$ denote the errors obtained with Equation (39) for the IPI+ and GP-based methods, respectively. The axis of abscissas in Figure 6 represents the number of the maximum degree L of the estimated SH coefficients. For each value of L, the number of samples used for the estimate of the SH coefficients was the same for both methods, and it was set such that the error $E_r^{(GP)}$ (as defined by Equation (39)) of our method is of 0.3%, 0.7% and 1.5% (red, green and blue lines, respectively). The plots of Figure 6 show that the error of the IPI method is always larger than that of our GP-based method. This relative error $\mathscr{E}$ increases with L and, when $L = 30$, it reaches: +18.5% for $E_r^{(GP)} = 0.3\%$; +54.7% for $E_r^{(GP)} = 0.7\%$ and +215% for $E_r^{(GP)} = 1.5\%$. The larger differences are explained by the GP regression effect which compensates for the decreasing

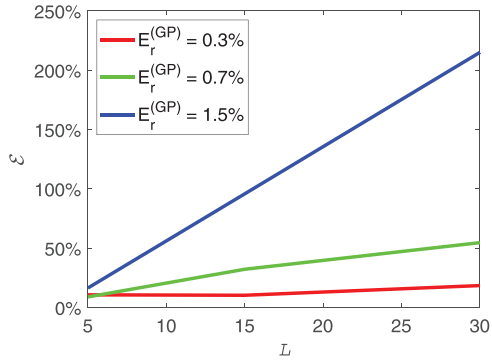


Figure 6: *Relative error $\mathcal{E}$ (Equation (40)) of the IPI method with respect to the error obtained with the GP-based approach, using the same number of samples for both methods. The error $\mathcal{E}$ is plotted as a function of the maximum degree L of the estimated SH coefficients. For each value of L, the number of used samples to estimate the SH coefficients c_{*i} is set such that the error of the GP-based approach ($E_r^{(GP)}$) computed using Equation (39) is* 0.3%, 0.7% *and* 1.5% *(red, green and blue lines, respectively).*

sampling density. When $L \leq 5$, our experiments show that the gain in accuracy achieved by our method is small. However, our method remains beneficial when extensive use of SH coefficients rotation is necessary. Recall that, as explained in Section 4.3, our method enables fast rotation of SH coefficients through the use of Equation (35).

As regards the computing time, both methods require $\mathcal{O}(M \times n)$ operations if, in the IPI method, the SH basis function values are pre-computed at the sample set locations. Indeed, in this case, the inner product computation reduces to the sum of product computations involved in the classic QMC integration method. However, as explained above, its performance in terms of error dependence on the number of samples makes the IPI method much more costly in general. For example, for $L = 50$, our method requires 3300 samples to obtain an error level E_r (Equation (39)) of 0.7%, whereas the IPI method requires $n = 5000$ samples. The ratio between the number of samples required for each of the two methods is even greater for an error level E_r of 1.5% with $n = 3000$ for IPI and $n = 1250$ for our method.

6. Application

Our experiments with HDR environment maps have shown that the GD (k_{gd}) and polynomial (kp_0) covariance functions are the most suited to radiance functions. Moreover, with this covariance functions, no hyperparameters adjustment is necessary for both covariance functions. Taking $\sigma_n = 0.1$, and $s_g = 1.5$ and $l_s = 0.4$ (for k_{gd} and kp_0, respectively) gives good results for the tested environment maps, and for both the original spherical and the spectral domain regressions. In the case of of the squared-exponential covariance function, an appropriate hyperparameter setting is easily obtained with Equation (36). We have seen that GP-based regression results in radiance function estimate that can be expressed in both original spherical (Equation (9) or (10)) and spectral domains (Equation (34)). Both expressions are appropriate for shading integral computations. In the spectral domain, the shading integral amounts to a $\mathbb{L}_2$ inner product computation through a sum of products of SH coefficients.

A direct application of GP-based regression is radiance function approximation/interpolation either by using a whole sphere/hemisphere sample set or a localized subset of samples. PRT [SKS02], path guiding [MGN17, RHJD18] and radiance caching [KGPB05] are certainly the most appropriate rendering methods that could benefit from GP-based regression and its ability to provide fast SH expansion. However, a key issue is how to use such approximations in importance sampling, i.e. how to provide an appropriate probability density function (pdf). Although Equation (10) defines a weighted sum of Gaussians when $k(\mathbf{x}, \mathbf{x}')$ is the squared exponential kernel, it cannot be considered as a Gaussian mixture pdf (such as in Vorba *et al*. [VKv*14]) since the weights α_i are not constrained to be positive. As used in [MGN17], the most appropriate importance sampling method in our case is hierarchical sample warping [CJAMJ05] and especially in its SH-based version [JCJ09] since SH-based expansion are easily obtained with Equation (35) or (38). For this purpose, the SH-expansion can be band-limited to a quite low L degree as a rough approximation is only necessary. Besides, GP can be used to model pdfs as shown in Fradi *et al*. [FFS*21] but this requires different covariance functions from those used in this paper. Another interesting application domain of GP-based regression concerns the case of BRDFs (e.g. [HLW15]) for which the potentialities are yet to be explored. GP-based methods (which are non-parametric) can also be used to define control variates for Monte-Carlo integration [OGC16] or speed up Markov Chain Monte Carlo integration [Ras03] but again, these applications require customized GP models. Finally, in the context of path guiding algorithms [VKv*14, MGN17], the incident radiance function might be noisy, which is not the case of our environment map-based experiments. However, the GP allows to model noisy observations through the σ_n hyperparameter.

7. Conclusion and Future Work

In this paper, we have shown that GP-based regression is a powerful tool for representing radiance functions provided that appropriate covariance functions are used. Furthermore, we have shown that GP-based regression can be easily expressed in the spectral domain in terms of SH expansion. This property led us to an efficient method for computing SH expansion of radiance functions and the kernel-based theoretical support of this method greatly facilitates rotation of the SH coefficients. We believe that GPs have a very rich potential of applications in global illumination rendering. We are currently investigating the use of GP to model BRDF and generate BRDF SH expansions. Future research includes adaptive sampling for shading integral computation and probability density estimation for Markov Chain Monte-Carlo integration.

Acknowledgements

This work was partially funded by TIN2018-095232-B-C21, SGR-2017 1742 and CERCA Programme/Generalitat de Catalunya. We acknowledge the support of NVIDIA Corporation with the donation

of the Titan Xp GPUs. As Serra Húnter Fellow, Ricardo Marques acknowledges the support of the Serra Húnter Programme.

References

[Abr97] Abrahamsen P.: *A Review of Gaussian Random Fields and Correlation Functions*. Oslo, Norway: Norwegian Computing Center, 1997. https://books.google.fr/books?id=InByAAAACAAJ.

[Adl81] Adler R.: *The Geometry of Random Fields. Probability and Statistics Series*. New York, USA: Wiley, 1981. https://books.google.fr/books?id=SxXvAAAAMAAJ.

[AFO05] Arikan O., Forsyth D., O'Brien J.: Fast and detailed approximate global illumination by irradiance decomposition. *ACM Transactions on Graphics 24*, 3 (July 2005), 1108–1114.

[Aub13] Aubert G.: An alternative to wigner d-matrices for rotating real spherical harmonics. *AIP Advances 3*, 6 (2013), 062121.

[BDT99] Bala K., Dorsey J., Teller S.: Radiance interpolants for accelerated bounded-error ray tracing. *ACM Transactions on Graphics 18*, 3 (July 1999), 213–256.

[BFB97] Blanco M., Flórez M., Bermejo M.: Evaluation of the rotation matrices in the basis of real spherical harmonics. *Journal of Molecular Structure: THEOCHEM 419*, 1 (1997), 19–27.

[Bis06] Bishop C. M.: *Pattern Recognition and Machine Learning*. New York, USA: Springer, 2006. http://research.microsoft.com/en-us/um/people/cmbishop/prml/.

[BSSW14] Brauchart J., Saff E., Sloan I., Womersley R.: QMC Designs: Optimal Order Quasi Monte Carlo Integration Schemes on the Sphere. *Mathematics of Computation 83*, 290 (2014). https://doi.org/10.1090/S0025-5718-2014-02839-1.

[CJAMJ05] Clarberg P., Jarosz W., Akenine-Möller T., Jensen H. W.: Wavelet importance sampling: Efficiently evaluating products of complex functions. *ACM Transactions on Graphics 24*, 3 (July 2005), 1166–1175.

[CVVMM*16] Camps-Valls G., Verrelst J., Munoz-Mari J., Laparra V., Mateo-Jimenez F., Gomez-Dans J.: A survey on Gaussian processes for earth-observation data analysis: A comprehensive investigation. *IEEE Geoscience and Remote Sensing Magazine 4*, 2 (2016), 58–78.

[DH94] Driscoll J., Healy D.: Computing fourier transforms and convolutions on the 2-sphere. *Advances in Applied Mathematics 15*, 2 (June 1994), 202–250.

[DWD08] Drake J., Worley P., D'Azevedo E.: Algorithm 888: Spherical harmonic transform algorithms. *ACM Transactions on Mathematical Software 35*, 3 (Oct. 2008). https://doi.org/10.1145/1391989.1404581.

[FFS*21] Fradi A., Feunteun Y., Samir C., Baklouti M., Bachoc F., Loubes J.-M.: Bayesian regression and classification using Gaussian process priors indexed by probability density functions. *Information Sciences 548* (2021), 56–68.

[FS98] Fasshauer G. E., Schumaker L. L.: Scattered Data Fitting on the Sphere. *Proceedings of the International Conference on Mathematical Methods for Curves and Surfaces II Lillehammer*. USA: Vanderbilt University Press, 1998, 117–166.

[GF16] Guinness J., Fuentes M.: Isotropic covariance functions on spheres: Some properties and modeling considerations. *Journal of Multivariate Analysis 143* (2016), 143–152.

[GKMD06] Green P., Kautz J., Matusik W., Durand F.: View-dependent precomputed light transport using nonlinear Gaussian function approximations. In *I3D'06: Proceedings of the 2006 Symposium on Interactive 3D Graphics and Games* (New York, NY, USA, 2006), Association for Computing Machinery, pp. 7–14. https://doi.org/10.1145/1111411.1111413.

[GLW21] Gia Q. T. L., Li M., Wang Y. G.: Algorithm 1018: FaVeST—Fast Vector Spherical Harmonic Transforms. *ACM Transactions on Mathematical Software*, *47*, 4 (2021), 1–24. http://doi.org/10.1145/3458470

[GOG*14] Gunter T., Osborne M. A., Garnett R., Hennig P., Roberts S. J.: Sampling for inference in probabilistic models with fast Bayesian quadrature. In *Advances in Neural Information Processing Systems* (vol. 27). Z. Ghahramani, M. Welling, C. Cortes and N. Lawrence, K. Q. Weinberger (Eds.). Curran Associates, Inc. (2014). https://proceedings.neurips.cc/paper/2014/file/e94f63f579e05cb49c05c2d050ead9c0-Paper.pdf.

[HB00] Hubbert S., Baxter B.: Radial basis functions for the sphere. In *Recent Progress in Multivariate Approximation. ISNM International Series of Numerical Mathematics* (vol. 137). W. Haussmann, K. Jetter and M. Reimer (Eds.). Birkhauser, Basel (2000), pp. 33–47. https://doi.org/10.1007/978-3-0348-8272-9_4.

[HLW15] Hao J., Liu Y., Weng D.: A BRDF representing method based on Gaussian process. In *Proceedinngs of the Computer Vision—ACCV 2014 Workshops* (Cham, 2015), C. Jawahar and S. Shan (Eds.), Springer International Publishing, pp. 542–553.

[IR98] Ivanic J., Ruedenberg K.: Rotation matrices for real spherical harmonics. Direct determination by recursion. *The Journal of Physical Chemistry A 102*, 45 (1998), 9099–9100.

[JCJ09] Jarosz W., Carr N. A., Jensen H. W.: Importance sampling spherical harmonics. *Computer Graphics Forum 28*, 2 (2009), 577–586.

[KGPB05] Křivánek J., Gautron P., Pattanaik S., Bouatouch K.: Radiance caching for efficient global illumination computation. *Visualization and Computer Graphics, IEEE Transactions on 11*, 5 (Sep.–Oct. 2005), 550–561.

[KKP*06] Křivánek J., Konttinen J., Pattanaik S., Bouatouch K., Žára J.: Fast approximation to spherical harmonics rotation. In *SIGGRAPH'06: Proceedings of the ACM*

SIGGRAPH 2006 Sketches (New York, NY, USA, 2006), Association for Computing Machinery, pp. 154–es. https://doi.org/10.1145/1179849.1180042.

[KSKM13] Kennedy R. A., Sadeghi P., Khalid Z., McEwen J. D.: Classification and construction of closed-form kernels for signal representation on the 2-sphere. In *Proceedings of the SPIE 8858, Wavelets and Sparsity XV* (2013), SPIE, pp. 88580M–88580M–15. https://doi.org/10.1117/12.2026126.

[LdWF12] Lessig C., de Witt T., Fiume E.: Efficient and accurate rotation of finite spherical harmonics expansions. *Journal of Computational Physics 231*, 2 (2012), 243–250.

[LS15] Lang A., Schwab C.: Isotropic Gaussian random fields on the sphere: Regularity, fast simulation and stochastic partial differential equations. *The Annals of Applied Probability 25*, 6 (2015), 3047–3094.

[MBB19] Marques R., Bouville C., Bouatouch K.: Optimal sample weights for hemispherical integral quadratures. *Computer Graphics Forum 38*, 1 (2019), 59–72.

[MBB20] Marques R., Bouville C., Bouatouch K.: Spectral analysis of quadrature rules and fourier truncation-based methods applied to shading integrals. *IEEE Transactions on Visualization and Computer Graphics 26*, 10 (2020), 3022–3036.

[MBR*13a] Marques R., Bouville C., Ribardière M., Santos L. P., Bouatouch K.: Spherical Fibonacci point sets for illumination integrals. *Computer Graphics Forum 32*, 8 (Dec. 2013), 134–143.

[MBR*13b] Marques R., Bouville C., Ribardière M., Santos L. P., Bouatouch K.: A spherical Gaussian framework for Bayesian Monte Carlo rendering of glossy surfaces. *IEEE Transactions on Visualization and Computer Graphics 19*, 10 (Oct. 2013), 1619–1632.

[MBSB15] Marques R., Bouville C., Santos L., Bouatouch K.: *Efficient Quadrature Rules for Illumination Integrals: From Quasi Monte Carlo to Bayesian Monte Carlo*. Synthesis Lectures on Computer Graphics and Animation. USA: Morgan & Claypool Publishers, 2015. https://books.google.fr/books?id=p0vrCQAAQBAJ.

[MGN17] Müller T., Gross M., Novák J.: Practical path guiding for efficient light-transport simulation. *Computer Graphics Forum 36*, 4 (2017), 91–100.

[NRL*20] Nguyen A. H., Rath M., Leitinger E., Nguyen K. V., Witrisal K.: Gaussian process modeling of specular multipath components. *Applied Sciences 10*, 15 (2020). https://doi.org/10.3390/app10155216.

[NSF12] Nowrouzezahrai D., Simari P., Fiume E.: Sparse zonal harmonic factorization for efficient sh rotation. *ACM Transactions on Graphics 31*, 3 (June 2012). https://doi.org/10.1145/2167076.2167081.

[NSWW07] Narcowich F. J., Sun X., Ward J. D., Wendland H.: Direct and inverse Sobolev error estimates for scattered data interpolation via spherical basis functions. *Foundations of Computational Mathematics 7*, 3 (2007), 369–390.

[NW02] Narcowich F. J., Ward J. D.: Scattered data interpolation on spheres: Error estimates and locally supported basis functions. *SIAM Journal on Mathematical Analysis 33*, 6 (2002), 1393–1410.

[OGC16] Oates C. J., Girolami M.: Chopin N. Control Functionals for Monte Carlo Integration. *Journal of the Royal Statistical Society*. Series B (Statistical Methodology) 79, no. 3 (2017): 695–718. http://www.jstor.org/stable/44681807

[Pac07] Paciorek C.: Bayesian smoothing with Gaussian processes using Fourier basis functions in the spectralGP package. *Journal of Statistical Software, Articles 19*, 2 (2007), 1–38.

[PSC*15] Pilleboue A., Singh G., Coeurjolly D., Kazhdan M., Ostromoukhov V.: Variance Analysis for Monte Carlo Integration. *ACM Transactions on Graphics 34*, 4 (July 2015), 124:1–124:14.

[Ras03] Rasmussen C.: Gaussian processes to speed up hybrid Monte Carlo for expensive Bayesian integrals. *Bayesian Statistics 7* (2003), 651–659.

[RHJD18] Reibold F., Hanika J., Jung A., Dachsbacher C.: Selective guided sampling with complete light transport paths. *ACM Transactions on Graphics 37*, 6 (Dec. 2018). https://doi.org/10.1145/3272127.3275030.

[RW06] Rasmussen C. E., Williams C. K. I.: *Gaussian Process for Machine Learning*. US: MIT Press, 2006.

[SKS02] Sloan P., Kautz J., Snyder J.: Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. In *SIGGRAPH'02: Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 2002), Association for Computing Machinery, pp. 527–536. https://doi.org/10.1145/566570.566612.

[SKS02] Sloan P.-P., Kautz J., Snyder J.: Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. *ACM Transaction on Graphics 21*, 3 (2002), 527–536.

[SW06] Schaback R., Wendland H.: Kernel techniques: From machine learning to meshless methods. *Acta Numerica 15* (May 2006), 543–639.

[TS06] Tsai Y., Shih Z.: All-frequency precomputed radiance transfer using spherical radial basis functions and clustered tensor approximation. *ACM Transactions on Graphics 25*, 3 (July 2006), 967–976.

[VKv*14] Vorba J., Karlík O., Šik M., Ritschel T., Křivánek J.: On-line learning of parametric mixture models for light trans-

port simulation. *ACM Transactions on Graphics 33*, 4 (July 2014). https://doi.org/10.1145/2601097.2601203.

[WAP18] Wu A., Aoi M., Pillow J.,: Exploiting gradients and Hessians in Bayesian optimization and Bayesian quadrature. arXiv preprint: 1704.00060v2 (2018). http://arxiv.org/abs/1704.00060.

[Wen04] Wendland H.: *Scattered Data Approximation. Cambridge Monographs on Applied and Computational Mathematics*. UK: Cambridge University Press, 2004. https://doi.org/10.1017/CBO9780511617539.

[Wil19] Wilkinson W. J.: Gaussian Process Modelling for Audio Signals. PhD thesis, School of Electronic Engineering and Computer Science, Queen Mary University of London, 2019.

[XBG18] Xi X., Briol F.-X., Girolami M.: Bayesian quadrature for multiple related integrals. In *Proceedings of the 35th International Conference on Machine Learning, PMLR* (July 2018), J. Dy and A. Krause (Eds.), vol. 80, pp. 5373–5382. https://proceedings.mlr.press/v80/xi18a.html.

[ZBN19] Zhao Y., Belcour L., Nowrouzezahrai D.: View-dependent radiance caching. In *GI'19: Proceedings of the 45th Graphics Interface Conference on Proceedings of Graphics Interface 2019* (Waterloo, Canada, 2019), Canadian Human-Computer Communications Society. https://doi.org/10.20380/GI2019.22.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supporting Information