Supplementary: Sparse Gaussian Processes with Spherical Harmonic Features

Vincent Dutordoir [1 2] **Nicolas Durrande** [1] **James Hensman** [3]

A. Spherical Harmonics on $\mathbb{S}^{d-1}$

This section gives a brief overview of some of the useful properties of spherical harmonics. We refer the interested reader to Dai & Xu (2013); Efthimiou & Frye (2014) for an in-depth overview.

Spherical harmonics are special functions defined on a hypersphere and originate from solving Laplace's equation in the spherical domain. They form a complete set of orthogonal functions, and any sufficiently regular function defined on the sphere can be written as a sum of these spherical harmonics, similar to the Fourier series with sines and cosines. Spherical harmonics have a natural ordering by increasing angular frequency. In the next paragraphs we introduce these concepts more formally.

We adopt the usual L_2 inner product for functions $f : \mathbb{S}^{d-1} \to \mathbb{R}$ and $g : \mathbb{S}^{d-1} \to \mathbb{R}$ restricted to the sphere

$$\langle f, g \rangle_{L_2(\mathbb{S}^{d-1})} = \frac{1}{\Omega_{d-1}} \int_{\mathbb{S}^{d-1}} f(x)\, g(x)\, \mathrm{d}\omega(x), \tag{1}$$

where $\mathrm{d}\omega(x)$ is the surface area measure such that Ω_{d-1} denotes the surface area of $\mathbb{S}^{d-1}$

$$\Omega_{d-1} = \int_{\mathbb{S}^{d-1}} \mathrm{d}\omega(x) = \frac{2\pi^{d/2}}{\Gamma(d/2)}. \tag{2}$$

Definition 1. *Spherical harmonics of degree (or level) ℓ, denoted as ϕ_ℓ, are defined as the restriction to the unit hypersphere $\mathbb{S}^{d-1}$ of the harmonic homogeneous polynomials (with d variables) of degree ℓ. It is the map $\phi_\ell : \mathbb{S}^{d-1} \to \mathbb{R}$ with ϕ_ℓ a homogeneous polynomial and $\Delta \phi_\ell = 0$.*

For a specific dimension d and degree ℓ there exist

$$N_\ell^d := (2\ell + d - 2)\frac{\Gamma(\ell + d - 2)}{\Gamma(\ell + 1)\Gamma(d - 2)}$$

different linearly independent spherical harmonics on $\mathbb{S}^{d-1}$. We refer to them as the set $\{\phi_{\ell,k}^d\}_{k=1}^{N_\ell^d}$. but in the subsequent we will drop the dependence on dimension d. The set is ortho-normal:

$$\langle \phi_{\ell,k}, \phi_{\ell',k'} \rangle_{L_2(\mathbb{S}^{d-1})} = \delta_{\ell\ell'}\delta_{kk'}. \tag{3}$$

Theorem 1. *Since the spherical harmonics form an orthonormal basis, every function $f : \mathbb{S}^{d-1} \to \mathbb{R}$ can be decomposed as*

$$f = \sum_{\ell=0}^{\infty} \sum_{k=1}^{N_\ell^d} \widehat{f}_{\ell,k}\phi_{\ell,k}, \text{ with } \widehat{f}_{\ell,k} = \langle f, \phi_{\ell,k} \rangle_{L_2(\mathbb{S}^{d-1})}. \tag{4}$$

Which can be seen as the spherical analogue of the Fourier decomposition of a periodic function in $\mathbb{R}$ onto a basis of sines and cosines.

Theorem 2. *The spherical harmonics are the eigenfunctions of the Laplace-Beltrami operator with eigenvalues $\lambda_\ell = \ell(\ell + d - 2)$ so that*

$$\Delta^{\mathbb{S}^{d-1}} \phi_{\ell,k} = \ell(\ell + d - 2)\phi_{\ell,k}. \tag{5}$$

In the experiments we used Dai & Xu (2013, Theorem 5.1) for an explicit expression of ϕ. We note that while this expression gives us a general form of the spherical harmonics in any dimension, we found that it becomes numerically unstable for $d \geq 10$.

A.1. Gegenbauer polynomials

Gegenbauer polynomials $C_\ell^{(\alpha)} : [-1, 1] \to \mathbb{R}$ are orthogonal polynomials with respect to the weight function $(1 - z^2)^{\alpha - 1/2}$. A variety of characterizations of the Gegenbauer polynomials are available. We use the polynomial characterisation for its numerical stability. It is given by

$$C_\ell^{(\alpha)}(z) = \sum_{k=0}^{\lfloor \ell/2 \rfloor} \frac{(-1)^k\, \Gamma(\ell - k + \alpha)}{\Gamma(\alpha)\Gamma(k + 1)\Gamma(\ell - 2k + 1)} (2z)^{\ell - 2k}. \tag{6}$$

The polynomials normalise by

$$\int_{-1}^{1} \left[C_\ell^{(\alpha)}(z)\right]^2 (1 - z^2)^{\alpha - \frac{1}{2}} \mathrm{d}z = \frac{\Omega_{d-1}}{\Omega_{d-2}} \frac{\alpha}{\ell + \alpha} C_\ell^{(\alpha)}(1), \tag{7}$$

[1]PROWLER.io, Cambridge, United Kingdom [2]Department of Engineering, University of Cambridge, Cambridge, United Kingdom [3]Amazon Research, Cambridge, United Kingdom (work done while JH was affiliated to PROWLER.io). Correspondence to: Vincent Dutordoir <vincent@prowler.io>.

Proceedings of the 37^{th} International Conference on Machine Learning, Online, PMLR 119, 2020.

with $C_\ell^{(\alpha)}(1) = \frac{\Gamma(2\alpha+\ell)}{\Gamma(2\alpha)\,\ell!}$.

There exists a close relationship between Gegenbauer polynomials (also known as *generalized Legendre polynomials*) and spherical harmonics, as we will show in the next theorems.

Theorem 3 (Addition)**.** *Between the spherical harmonics of degree ℓ in dimension d and the Gegenbauer polynomials of degree ℓ there exists the relation*

$$\sum_{k=1}^{N_\ell^d} \phi_{\ell,k}(\mathbf{x})\phi_{\ell,k}(\mathbf{x}') = \frac{\ell+\alpha}{\alpha}\, C_\ell^{(\alpha)}(\mathbf{x}^\top\mathbf{x}'), \tag{8}$$

with $\alpha = \frac{d-2}{2}$.

As a illustrative example, this property is analogues to the trigonometric addition formula: $\sin(x)\sin(x') + \cos(x)\cos(x') = \cos(x-x')$.

Theorem 4 (Funk-Hecke)**.** *Let $s(\cdot)$ be an integrable function such that $\int_{-1}^{1}\|s(t)\|(1-t^2)^{(d-3)/2}\mathrm{d}t$ is finite and $d \geq 2$. Then for every $\phi_{\ell,k}$*

$$\frac{1}{\Omega_{d-1}}\int_{\mathbb{S}^{d-1}} s(\mathbf{x}^\top\mathbf{x}')\,\phi_{\ell,k}(\mathbf{x}')\,\mathrm{d}\omega(\mathbf{x}') = \widehat{a}_\ell\,\phi_{\ell,k}(\mathbf{x}), \tag{9}$$

where $\widehat{a}_\ell$ is a constant defined by

$$\widehat{a}_\ell = \frac{\omega_d}{C_\ell^{(\alpha)}(1)}\int_{-1}^{1} s(t)\,C_\ell^{(\alpha)}(t)\,(1-t^2)^{\frac{d-3}{2}}\mathrm{d}t, \tag{10}$$

with $\alpha = \frac{d-2}{2}$ *and* $\omega_d = \frac{\Omega_{d-2}}{\Omega_{d-1}}$.

Funk-Hecke simplifies a $(d-1)$-variate surface integral on $\mathbb{S}^{d-1}$ to a one-dimensional integral over $[-1,1]$. This theorem gives us a practical way of computing the Fourier coefficients for any zonal kernel. In section B.3, we use it to compute the coefficients of the arc-cosine kernel. Notice how the Fourier coefficients $\widehat{a}_\ell$ only depend on the level ℓ (or degree) of the spherical harmonic and not the orientation (denoted by the k index).

B. Zonal kernels

B.1. Mercer's decomposition

Zonal kernels can be seen as the spherical counterpart of stationary kernels. Stationary kernels are a function of $\mathbf{x}-\mathbf{x}'$ and are thus invariant to translations in the input space (Rasmussen & Williams, 2006). Zonal kernels (defined on $\mathbb{S}^{d-1}\times\mathbb{S}^{d-1}$) are a function of $\mathbf{x}^\top\mathbf{x}'$ and are thus invariant to rotations.

The spherical harmonics are the eigenfunctions of the Laplace-Beltrami operator (Dai & Xu, 2013; Efthimiou & Frye, 2014). In the main paper we show the commutativity of the Laplace-Beltrami operator and the kernel operator of zonal kernels. This means that the spherical harmonics are also the eigenfunctions of zonal kernels, as commuting operators share the same eigenfunctions.

Mercer's theorem allows us to express the kernel in terms of its eigenvalues and eigenfunctions.

Theorem 5 (Mercer representation)**.** *Any zonal kernel k on the hypersphere can be decomposed as*

$$k(\mathbf{x},\mathbf{x}') = \sum_{\ell=0}^{\infty}\sum_{k=1}^{N_\ell^d} \widehat{a}_{\ell,k}\phi_{\ell,k}(\mathbf{x})\phi_{\ell,k}(\mathbf{x}'), \tag{11}$$

where $\mathbf{x},\mathbf{x}' \in \mathbb{S}^{d-1}$ and $\widehat{a}_\ell$ are the positive Fourier coefficients, $\phi_{\ell,k}$ denote the elements of the spherical harmonic basis in $\mathbb{S}^{d-1}$, and N_ℓ^d corresponds to the number of spherical harmonics for a given level ℓ.

For zonal kernels the Fourier coefficients within a level are equal: $\widehat{a}_{\ell,k} = \widehat{a}_\ell$ for $1 \leq k \leq N_\ell^d$. This allows us to simplify the Mercer decomposition of a zonal kernel using theorem 3 to

$$k(\mathbf{x},\mathbf{x}') = \sum_{\ell=0}^{\infty}\widehat{a}_\ell\frac{\ell+\alpha}{\alpha}\, C_\ell^{(\alpha)}(\mathbf{x}^\top\mathbf{x}'), \tag{12}$$

with $\alpha = \frac{d-2}{2}$.

B.2. RKHS

Given the Mercer representation of a zonal kernel, its RKHS can be characterised by

$$\mathcal{H} = \left\{ g = \sum_{\ell=0}^{\infty}\sum_{k=1}^{N_\ell^d}\widehat{g}_{\ell,k}\phi_{\ell,k} : \sum_{\ell=0}^{\infty}\sum_{k=1}^{N_\ell^d}\frac{|\widehat{g}_{\ell,k}|^2}{\widehat{a}_\ell} < \infty \right\}$$

with a reproducing inner product between two functions $g(\mathbf{x}) = \sum_{\ell,k}\widehat{g}_{\ell,k}\phi_{\ell,k}(\mathbf{x})$ and $h(\mathbf{x}) = \sum_{\ell,k}\widehat{h}_{\ell,k}\phi_{\ell,k}(\mathbf{x})$ defined as

$$\langle g,h\rangle_{\mathcal{H}} = \sum_{\ell=0}^{\infty}\sum_{k=1}^{N_\ell^d}\frac{\widehat{g}_{\ell,k}\widehat{h}_{\ell,k}}{\widehat{a}_\ell}. \tag{13}$$

Proof. (Reproducing property). The Fourier coefficients for $k(\mathbf{x},\cdot):\mathbb{S}^{d-1}\to\mathbb{R}$ and $f:\mathbb{S}^{d-1}\to\mathbb{R}$ are $\widehat{a}_{\ell,k}\phi_{\ell,k}(\mathbf{x})$ and $\widehat{f}_{\ell,k}$, respectively. Substituting these coefficients in eq. (13) gives:

$$\langle k(\mathbf{x},\cdot),f\rangle_{\mathcal{H}} = \sum_{\ell=0}^{\infty}\sum_{k=1}^{N_\ell^d}\frac{\widehat{a}_\ell\phi_{\ell,k}(\mathbf{x})\widehat{f}_{\ell,k}}{\widehat{a}_\ell} \tag{14}$$

$$= \sum_{\ell=0}^{\infty}\sum_{k=1}^{N_\ell^d}\widehat{f}_{\ell,k}\phi_{\ell,k}(\mathbf{x}) = f(\mathbf{x}) \tag{15}$$

which proofs the reproducing property. □

In the next sections we address the computation of the Fourier coefficients (eigenvalues) of the kernels. In section B.3 for the Arc-Cosine kernel and in section B.4 for the Matérn family.

B.3. Fourier coefficients for the Arc-Cosine kernel

The Fourier coefficients are computed using theorem 4, where the shape function of the Arc-Cosine kernel of the first order (Cho & Saul, 2009) is given by:

$$s(x) = \sin x + (\pi - x)\cos x. \tag{16}$$

Notice that we expressed the shape function as a function of the angle between the two inputs $s : [0, \pi] \mapsto \mathbb{R}$, rather than the great-circle distance, as it simplifies the subsequent computations.

Using a change of variables we also rewrite theorem 4

$$\widehat{a}_\ell = c_{d,\ell} \int_0^\pi s(x)\, C_\ell^{\frac{d-2}{2}}(\cos x) \sin^{d-2} x \,\mathrm{d}x, \tag{17}$$

with $c_{d,\ell} = \frac{\omega_d}{C_\ell^{(\alpha)}(1)}$. This one dimensional integral can be solved in closed-form for any setting of d and ℓ. Filling in the definition for the Gegenbauer polynomial (eq. (6)), we observe that we need a general solution of the integral $\int_0^\pi [\sin(x) + (\pi - x)\cos(x)] \cos^n(x) \sin^m(x) dx$ for $n, m \in \mathbb{N}$.

The first term can be computed with the well-known result:

$$\int_0^\pi \sin^n(x) \cos^m(x) dx \tag{18}$$

$$= \begin{cases} 0 & \text{if } m \text{ odd} \\ \frac{(n-1)!!\,(m-1)!!}{(n+m)!!}\pi & \text{if } m \text{ even and } n \text{ odd,} \\ \frac{(n-1)!!\,(m-1)!!}{(n+m)!!}2 & \text{if } n, m \text{ even.} \end{cases} \tag{19}$$

The second term is harder:

$$I = \int_0^\pi (\pi - x) \sin^n(x) \cos^m(x) dx \tag{20}$$

which we solved using integration by parts with $u = \pi - x$ and $dv = \sin^n(x)\cos^m(x)dx$, so that

$$I = u(0)v(0) - u(\pi)v(\pi) + \int_0^\pi v(x')dx', \tag{21}$$

where $v(x') = \int_0^{x'} \sin^n(x)\cos^m(x)dx$. This gives $v(0) = 0$ and $u(0) = 0$, simplifying $I = \int_0^\pi v(x')dx'$. We first focus on $v(x')$: for $\underline{n \text{ odd}}$, there exists a $n' \in \mathbb{N}$ so that $n = 2n' + 1$, resulting

$$v(x') = \int_0^{x'} \sin^{2n'}(x) \cos^m(x) \sin(x) dx \tag{22}$$

$$= -\int_0^{\cos(x')} (1 - u^2)^{n'} u^m du \tag{23}$$

Where we used $\sin^2(x) + \cos^2(x) = 1$ and the substitution $u = \cos(x) \implies du = -\sin(x)dx$. Using the binomial expansion, we get

$$v(x') = -\int_0^{\cos(x')} \sum_{i=0}^{n'} \binom{k}{i} (-u^2)^i u^m du \tag{24}$$

$$= \sum_{i=0}^{n'} (-1)^{i+1} \binom{k}{i} \frac{\cos(x')^{2i+m+1} - 1}{2i + m + 1}. \tag{25}$$

Similarly, for $\underline{m \text{ odd}}$, we have $m = 2m' + 1$ and use the substitution $u = \sin(x)$, to get

$$v(x') = \sum_{i=0}^{m'} (-1)^i \binom{k}{i} \frac{\sin(x')^{2i+n+1}}{2i + n + 1}. \tag{26}$$

For $\underline{n \text{ and } m \text{ even}}$, we have $n' = n/2$ and $m' = m/2$, we use double-angle identities to get

$$v(x') = \int_0^{x'} \left(\frac{1 - \cos(2x)}{2}\right)^{n'} \left(\frac{1 + \cos(2x)}{2}\right)^{m'} dx \tag{27}$$

Making use of the binomial expansion twice, we get

$$v(x') = 2^{-(n'+m')} \sum_{i,j=0}^{n',m'} (-1)^i \binom{n'}{i}\binom{m'}{j} \int_0^{x'} \cos(2x)^{i+j} dx. \tag{28}$$

Returning back to the original problem $I = \int_0^\pi v(x')dx'$. Depending on the parity of n and m we need to evaluate:

$$\int_0^\pi \cos(x')^p dx' = \begin{cases} \frac{(p-1)!!}{p!!}\pi & \text{if } p \text{ even} \\ 0 & \text{if } p \text{ odd} \end{cases} \tag{29}$$

and

$$\int_0^\pi \sin(x')^p dx' = \begin{cases} \frac{(p-1)!!}{p!!}\pi & \text{if } p \text{ even} \\ \frac{(p-1)!!}{p!!}2 & \text{if } p \text{ odd.} \end{cases} \tag{30}$$

For m and n even we need to solve the double integral

$$\int_0^\pi \int_0^{x'} \cos(2x)^p dx dx' = \begin{cases} \frac{(p-1)!!}{p!!}\frac{\pi^2}{2} & \text{if } p \text{ even} \\ 0 & \text{if } p \text{ odd.} \end{cases} \tag{31}$$

Combining these results gives us the solution for the integral $\int_0^\pi s(x)\cos^n(x)\sin^m(x)dx$ for any $n, m \in \mathbb{N}$, which is necessary to compute $\widehat{a}_\ell$ for the arc-cosine kernel.

B.4. Fourier coefficients for the Matérn family kernels

The Matérn covariance between two points x, x' separated by $r = x - x'$ distance units is given by (Rasmussen & Williams, 2006):

$$k_\nu(r) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{r}{\rho}\right)^\nu K_\nu\left(\sqrt{2\nu}\frac{r}{\rho}\right), \tag{32}$$

where Γ is the gamma function, K_ν is the modified Bessel function of the second kind, and ρ (lengthscale) and ν (differentiability) are non-negative parameters of the covariance. The covariance has a spectral density defined on $\mathbb{R}^d$

$$S(\omega) = \frac{2^d \pi^{\frac{d}{2}} \Gamma(\nu + \frac{d}{2})(2\nu)^\nu}{\Gamma(\nu)\rho^{2\nu}} \left(\frac{2\nu}{\rho^2} + 4\pi^2\omega^2 \right)^{-\left(\nu + \frac{d}{2}\right)}. \tag{33}$$

A key result from Solin & Särkkä (2014, eq. 20) is to show that the coefficients $\widehat{a}_{\ell,k}$ have a simple expression that depends on the kernel spectral density S and the eigenvalues of the Laplace-Beltrami operator (theorem 2). For GPs on $\mathbb{S}^{d-1}$ the coefficients boil down to

$$\widehat{a}_{\ell,k} = S\left(\sqrt{\ell(\ell + d - 2)}\right). \tag{34}$$

References

Cho, Y. and Saul, L. K. Kernel methods for deep learning. In *Advances in Neural Information Processing Systems*, pp. 342–350, 2009.

Dai, F. and Xu, Y. *Approximation theory and harmonic analysis on spheres and balls*. Springer, 2013.

Efthimiou, C. and Frye, C. *Spherical harmonics in p dimensions*. World Scientific Publishing, 2014.

Rasmussen, C. E. and Williams, C. K. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

Solin, A. and Särkkä, S. Hilbert space methods for reduced-rank Gaussian process regression. *Statistics and Computing*, pp. 1–28, 2014.