A spectral mixture representation of isotropic kernels to generalize random Fourier features

Nicolas Langrené*[1], Xavier Warin[2], and Pierre Gruet[2]

[1]Guangdong Provincial/Zhuhai Key Laboratory of Interdisciplinary Research and Application for Data Science, Beijing Normal-Hong Kong Baptist University
[2]EDF Lab Paris-Saclay, FiME (Laboratoire de Finance des Marchés de l'Énergie)

April 9, 2025

Rahimi and Recht [37] introduced the idea of decomposing positive definite shift-invariant kernels by randomly sampling from their spectral distribution. This famous technique, known as Random Fourier Features (RFF), is in principle applicable to any such kernel whose spectral distribution can be identified and simulated. In practice, however, it is usually applied to the Gaussian kernel because of its simplicity, since its spectral distribution is also Gaussian. Clearly, simple spectral sampling formulas would be desirable for broader classes of kernels. In this paper, we show that the spectral distribution of positive definite isotropic kernels in $\mathbb{R}^d$ for all $d \geq 1$ can be decomposed as a scale mixture of α-stable random vectors, and we identify the mixing distribution as a function of the kernel. This constructive decomposition provides a simple and ready-to-use spectral sampling formula for many multivariate positive definite shift-invariant kernels, including exponential power kernels, generalized Matérn kernels, generalized Cauchy kernels, as well as newly introduced kernels such as the Beta, Kummer, and Tricomi kernels. In particular, we retrieve the fact that the spectral distributions of these kernels are scale mixtures of the multivariate Gaussian distribution, along with an explicit mixing distribution formula. This result has broad applications for support vector machines, kernel ridge regression, Gaussian processes, and other kernel-based machine learning techniques for which the random Fourier features technique is applicable.

Keywords: random Fourier features, random projections, spectral distribution, stable distribution, isotropic kernels, generalized Matérn kernel, generalized Cauchy kernel, Tricomi kernel, spectral Monte Carlo, Gaussian processes.

*Corresponding author, nicolaslangrene@uic.edu.cn

1 Introduction

We start by fixing some notation and terminology regarding kernel functions. Let $\tilde{K} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a kernel function. It is said to be *shift-invariant* (a.k.a. translation-invariant, radially-symmetric, or stationary) if for any $\mathbf{x}_i \in \mathbb{R}^d$ and $\mathbf{x}_j \in \mathbb{R}^d$, $\tilde{K}(\mathbf{x}_i, \mathbf{x}_j) = K(\mathbf{x}_i - \mathbf{x}_j)$ only depends on $\mathbf{x}_i$ and $\mathbf{x}_j$ through the difference $\mathbf{x}_i - \mathbf{x}_j$. Moreover, the shift-invariant kernel $K : \mathbb{R}^d \to \mathbb{R}$ is said to be *isotropic* [14] if it only depends on $\mathbf{x}_i$ and $\mathbf{x}_j$ through the Euclidean norm $\|\mathbf{x}_i - \mathbf{x}_j\|$ of the difference $\mathbf{x}_i - \mathbf{x}_j$.

Next, we say that the shift-invariant kernel $K : \mathbb{R}^d \to \mathbb{R}$ is *positive definite* if for any $N \geq 1$, $(\mathbf{x}_1, \dots, \mathbf{x}_N) \in \mathbb{R}^{d \times N}$ and $(z_1, \dots, z_N) \in \mathbb{R}^N$,

$$\sum_{i=1}^{N} \sum_{j=1}^{N} z_i z_j K(\mathbf{x}_i - \mathbf{x}_j) \geq 0. \tag{1}$$

Denote by Φ_∞ the set of all continuous shift-invariant kernels which are positive definite in $\mathbb{R}^d$ for all $d \geq 1$. According to Bochner's theorem [1, 2], a continuous, shift-invariant kernel $K : \mathbb{R}^d \to \mathbb{R}$ is positive definite if and only if there exists a finite positive measure μ such that

$$K(\mathbf{u}) = \int_{\mathbb{R}^d} \exp(i\mathbf{x}^\top \mathbf{u}) d\mu(\mathbf{x}), \quad \mathbf{u} \in \mathbb{R}^d.$$

In other words, K is proportional to a characteristic function. If the kernel is scaled such that $K(\mathbf{0}) = 1$, then μ is a probability measure. Suppose that it admits a density f (see [40, Theorem 1.8.16] for a characterization). Then

$$K(\mathbf{u}) = K(\mathbf{0}) \int_{\mathbb{R}^d} \exp(i\mathbf{x}^\top \mathbf{u}) f(\mathbf{x}) d\mathbf{x}, \quad \mathbf{u} \in \mathbb{R}^d. \tag{2}$$

This shows that K is the multivariate Fourier transform of f. According to Bochner's theorem, f is nonnegative if and only if K is positive definite. In other words, f is a probability density function, known as the *spectral density* of K, if and only if K is positive definite.

In general, equation (2) is complex-valued. Since a characteristic function is real-valued if and only if f is symmetric around zero ([11, Lemma 1 page 499], [40, Theorem 1.3.13]), we can further explicit equation (2) by assuming that $f(\mathbf{x}) = f(-\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^d$, in which case $K(\mathbf{u}) = K(-\mathbf{u})$ for all $\mathbf{u} \in \mathbb{R}^d$ and

$$K(\mathbf{u}) = K(\mathbf{0}) \int_{\mathbb{R}^d} \cos(\mathbf{x}^\top \mathbf{u}) f(\mathbf{x}) d\mathbf{x}, \quad \mathbf{u} \in \mathbb{R}^d. \tag{3}$$

Moreover, K and f are Fourier duals:

$$f(\mathbf{x}) = \frac{1}{K(\mathbf{0})(2\pi)^d} \int_{\mathbb{R}^d} \cos(\mathbf{x}^\top \mathbf{u}) K(\mathbf{u}) d\mathbf{u}, \quad \mathbf{x} \in \mathbb{R}^d. \tag{4}$$

The interesting aspect of equation (3) is that if K is positive definite, then f is a probability density function, and the following probabilistic representation holds

$$K(\mathbf{u}) = K(\mathbf{0}) \mathbb{E}\left[\cos(\boldsymbol{\eta}^\top \mathbf{u})\right] \tag{5}$$

where $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_d)$ is a continuous random vector with density f. The random vector $\boldsymbol{\eta}$ is known as *random projection*. The probabilistic representation (5) suggests that the kernel K can be approximated by Monte Carlo simulations:

$$K(\mathbf{u}) \simeq K_M(\mathbf{u}) := \frac{K(\mathbf{0})}{M} \sum_{m=1}^{M} \cos(\boldsymbol{\eta}_m^\top \mathbf{u}), \quad \mathbf{u} \in \mathbb{R}^d. \tag{6}$$

This approach is known as *random Fourier features* [37]. It provides an explicit feature map to approximate kernel functions in an efficient way[1], with broad applications in machine learning and statistical learning, for example for algorithms such as kernel ridge regression, Gaussian process inference, kernel principal component analysis, support vector machines, and other kernel-based methods [13].

In order to implement equation (6), the spectral distribution of K needs to be precomputed. Rahimi and Recht [37] provide three examples of shift-invariant multivariate kernels amenable to this method: the Gaussian kernel, the Laplace kernel and the Cauchy kernel:

- The Gaussian kernel admits a Gaussian spectral density

$$K(\mathbf{u}) = \exp\left(-\frac{1}{2}\|\mathbf{u}\|^2\right) \quad , \quad f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}\|\mathbf{x}\|^2\right) , \tag{7}$$

- The Laplace kernel admits a Cauchy spectral density

$$K(\mathbf{u}) = \exp\left(-\sum_{\ell=1}^{d} |u_\ell|\right) \quad , \quad f(\mathbf{x}) = \frac{1}{\pi^d} \prod_{\ell=1}^{d} \frac{1}{1+x_\ell^2} , \tag{8}$$

- The Cauchy kernel admits a Laplace spectral density

$$K(\mathbf{u}) = \prod_{\ell=1}^{d} \frac{1}{1+u_\ell^2} \quad , \quad f(\mathbf{x}) = \frac{1}{2^d} \exp\left(-\sum_{\ell=1}^{d} |x_\ell|\right). \tag{9}$$

where $\mathbf{u} = (u_1, \ldots, u_d) \in \mathbb{R}^d$ and $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$. Remark that the two kernels (8)-(9) are not isotropic kernels but tensor kernels, obtained as the product of univariate kernels; for the Gaussian kernel (7), the isotropic and tensor formulations coincide. Remark also how the roles of K and f can be swapped: if K is a nonnegative integrable kernel with spectral density f, then $f/f(\mathbf{0})$ is a kernel with spectral density $K/((2\pi)^d f(\mathbf{0}))$. This is a particular case of the duality theorem in Harrar et al. [17]; see also [15].

In principle, the random Fourier features formula (6) can be applied to any positive definite shift-invariant kernel for which the spectral density can be computed and simulated. In practice, the simplicity and convenience of the three examples (7)-(8)-(9), especially the Gaussian kernel (7), means that there is not much work available on the application of random Fourier features to other multivariate parametric kernels. One clear challenge is the ability to compute the multivariate inverse Fourier transform (4) analytically, and to find a suitable sampling algorithm for the corresponding distribution.

[1] $K(\mathbf{x}_i - \mathbf{x}_j) \simeq \varphi(\mathbf{x}_i)^\top \varphi(\mathbf{x}_j)$ where for any $\mathbf{x} \in \mathbb{R}^d$,
$\varphi(\mathbf{x}) := \frac{\sqrt{K(\mathbf{0})}}{\sqrt{M}} \begin{bmatrix} \cos(\boldsymbol{\eta}_1^\top \mathbf{x}) & \ldots & \cos(\boldsymbol{\eta}_M^\top \mathbf{x}) & \sin(\boldsymbol{\eta}_1^\top \mathbf{x}) & \ldots & \sin(\boldsymbol{\eta}_M^\top \mathbf{x}) \end{bmatrix}^\top \in \mathbb{R}^{2M}$

This is the task that we propose to address in this paper, with a focus on isotropic kernels which are positive definite in $\mathbb{R}^d$ for all $d \geq 1$ ($K \in \Phi_\infty$). First, we characterize the spectral distributions of such kernels as convenient scale mixtures for simulation. Since Schoenberg's theorem characterizes kernels in Φ_∞ as Gaussian mixtures [40, Theorem 3.8.5] and the Fourier transform of a Gaussian mixture is also a Gaussian mixture [15], one can readily deduce that the random projections of kernels in Φ_∞ are Gaussian mixtures too. This means that multiplying multivariate Gaussian simulations by simulations from a specific nonnegative distribution conveniently extends the random Fourier features approach from Gaussian kernels to any positive definite isotropic kernel in $\mathbb{R}^d$ for all $d \geq 1$ (same mixture simulations in every dimension) or tensor kernel (independent mixture simulations in each dimension) of interest (Remark 4).

Then, using results from the theories of multivariate stable distributions [39, 7, 30] and multivariate characteristic functions [40], we show that the spectral distributions of kernels in Φ_∞ can more generally be characterized as scale mixtures of multivariate symmetric stable distributions. The scaling distribution is obtained explicitly from the inverse Laplace transform of the kernel function (Theorem 1, Corollary 1). We provide explicit formulas to simulate these spectral distributions using known results about stable distributions [3, 30].

Finally, we apply our decomposition of random projections to several shift-invariant kernels of interest in machine learning, including Matérn kernels, exponential power kernels and generalized Cauchy kernels. Our mixture representation of random projections makes it straightforward to generalize and create new multivariate isotropic kernels for which the random Fourier features methodology is readily applicable. Examples of newly introduced kernels include the generalized Matérn, Beta, Kummer and Tricomi kernels (Table 1). The availability of new kernel functions with additional parameters is of great practical interest, especially in the situation when classical parametric kernels are too rigid to properly capture the information available in the data. We implement and illustrate these kernels along with their random Fourier features approximations. Our numerical experiments confirm the exactness of our theoretical results.

The rest of the paper is organized as follows. Section 2 describes our characterization of random projections as scale mixtures, and applies it to several examples of multivariate isotropic kernels. Section 3 describes how to sample such random projections in practice. Section 4 provides numerical examples, and finally Section 5 concludes the paper.

2 Mixture representation of spectral distributions

In this section, we show that the spectral distribution of (multivariate) isotropic kernels can always be characterized as a scale mixture of symmetric stable distributions (Theorem 1). We characterize this scaling distribution and use it to decompose the random projections of several multivariate isotropic kernels of interest, and easily create new ones (Table 1).

In order to describe this characterization of multivariate spectral distributions, we need the following definition.

Definition 1. For any $\alpha \in (0, 2]$, let $\boldsymbol{S}_\alpha$ be a d-dimensional random vector with characteristic function ϕ_α given by

$$\phi_\alpha(\mathbf{u}) = \mathbb{E}\left[e^{i\boldsymbol{S}_\alpha^\top \mathbf{u}}\right] = e^{-\|\mathbf{u}\|^\alpha} \ , \ \mathbf{u} \in \mathbb{R}^d \tag{10}$$

where $\|\mathbf{u}\| = \sqrt{u_1^2 + \ldots + u_d^2}$ is the Euclidean norm of $\mathbf{u} = (u_1, \ldots, u_d) \in \mathbb{R}^d$. The vector $\boldsymbol{S}_\alpha$ is called *symmetric stable* (see for example [39]).

Lemma 1. *For any $\alpha \in (0, 2]$, let $\boldsymbol{S}_\alpha$ be a d-dimensional symmetric stable random vector (Definition 1), let R be a real-valued nonnegative random variable, independent of $\boldsymbol{S}_\alpha$, with Laplace transform $\mathcal{L}$, and let $\lambda > 0$ be a positive constant. Then, the random projection vector defined by*

$$\boldsymbol{\eta} = (\lambda R)^{\frac{1}{\alpha}} \boldsymbol{S}_\alpha \tag{11}$$

spans the following isotropic kernel $K : \mathbb{R}^d \to \mathbb{R}$:

$$K(\mathbf{u}) = K(\mathbf{0})\mathbb{E}\left[e^{i\boldsymbol{\eta}^\top \mathbf{u}}\right] = K(\mathbf{0})\mathcal{L}(\lambda \|\mathbf{u}\|^\alpha) \ , \ \mathbf{u} \in \mathbb{R}^d \tag{12}$$

Proof. Recall that the Laplace transform $\mathcal{L}$ of a nonnegative random variable R is defined by $\mathcal{L}(s) = \mathbb{E}\left[e^{-sR}\right]$ for $s \geq 0$. Then, the characteristic function of $\boldsymbol{\eta}$ is given by

$$\begin{aligned} \mathbb{E}\left[e^{i\boldsymbol{\eta}^\top \mathbf{u}}\right] &= \mathbb{E}\left[\mathbb{E}\left[\exp\left(i\boldsymbol{S}_\alpha^\top \left((\lambda R)^{\frac{1}{\alpha}}\mathbf{u}\right)\right) | R\right]\right] \\ &= \mathbb{E}\left[\exp\left(-\lambda R \|\mathbf{u}\|^\alpha\right)\right] \\ &= \mathcal{L}(\lambda \|\mathbf{u}\|^\alpha) \end{aligned}$$

which proves equation (12). □

Lemma 1 has previously been used for example in [34] in the case $d = 1$. Now, the following Theorem 1 provides a characterization of the spectral distribution of continuous, positive definite, isotropic kernels in Φ_∞ as a scale mixture of α-stable random vectors.

Theorem 1. *Let $K(\mathbf{u}) = k(\|\mathbf{u}\|^2)$, $\mathbf{u} \in \mathbb{R}^d$, be a kernel function in Φ_∞. Then*

a) there exists a nonnegative random variable R such that its Laplace transform is equal to $k/k(0)$,

b) for every $\alpha \in (0, 2]$, the isotropic kernel $K_\alpha(\mathbf{u}) := k(\|\mathbf{u}\|^\alpha)$ is also positive definite in $\mathbb{R}^d$ for every $d \geq 1$,

c) for every $\alpha \in (0, 2]$, the unique random projection vector $\boldsymbol{\eta}_\alpha$ of the kernel K_α admits the representation $\boldsymbol{\eta}_\alpha = R^{\frac{1}{\alpha}} \boldsymbol{S}_\alpha$, where $\boldsymbol{S}_\alpha$ is a symmetric stable random vector independent of R.

Proof. Part a) is a rephrasing of Schoenberg's theorem for kernel functions in Φ_∞ [40, Theorem 3.8.5] and the Hausdorff-Bernstein-Widder theorem [40, Theorem 3.9.6], which we need to introduce the random variable $R \geq 0$ whose existence is needed in parts b) and c). As clearly stated in [40, Theorem 3.9.8], the fact that the isotropic kernel $K(\mathbf{u}) = k(\|\mathbf{u}\|^2)$, $\mathbf{u} \in \mathbb{R}^d$, is continuous and positive definite for every $d \geq 1$ is equivalent to the fact that there exists a finite nonnegative measure μ on $[0, \infty)$ such that $k(u) = \int_0^\infty e^{-us} d\mu(s)$ for all $u \in [0, \infty)$. This means that $k(0)$ is finite and there exists a nonnegative random variable $R \geq 0$ such that $k(u) = k(0)\mathbb{E}\left[e^{-uR}\right]$. In other words, $k/k(0)$ is the Laplace transform of R.

Then, using Lemma 1 with $\lambda = 1$ and this random variable R, the random projection vector $\boldsymbol{\eta}_\alpha = R^{\frac{1}{\alpha}} \boldsymbol{S}_\alpha$ spans the isotropic kernel $K_\alpha(\mathbf{u}) = k(\|\mathbf{u}\|^\alpha)$, where we used the fact that $K(\mathbf{0}) = k(0)$. The existence of $\boldsymbol{\eta}_\alpha$ proves b) by Bochner's theorem, and the Uniqueness Theorem 1.3.3 in [40] completes the proof of c). □

It is sometimes more convenient to work with a kernel defined as $k(\|\mathbf{u}\|)$ (as in the definition of isotropic kernels) instead of $k(\|\mathbf{u}\|^2)$. A simple change of variable in Theorem 1 gives the following Corollary:

Corollary 1. *Let $K(\mathbf{u}) = k(\|\mathbf{u}\|)$, $\mathbf{u} \in \mathbb{R}^d$, be a kernel function in Φ_∞. Then*

a) there exists a nonnegative random variable R such that its Laplace transform is equal to $k(\sqrt{\cdot})/k(0)$,

b) for every $\alpha \in (0, 1]$, the isotropic kernel $K_\alpha(\mathbf{u}) := k(\|\mathbf{u}\|^\alpha)$ is also positive definite in $\mathbb{R}^d$ for every $d \geq 1$,

c) for every $\alpha \in (0, 1]$, the unique random projection vector $\boldsymbol{\eta}_\alpha$ of the kernel K_α admits the representation $\boldsymbol{\eta}_\alpha = R^{\frac{1}{2\alpha}} \boldsymbol{S}_{2\alpha}$, where $\boldsymbol{S}_{2\alpha}$ is a symmetric stable random vector independent of R.

In the case $\alpha = 2$ in Theorem 1, one can easily retrieve the fact that the random projection is a scale mixture of a Gaussian vector:

Corollary 2. *In the case $\alpha = 2$, the continuous, positive definite, isotropic kernel $K(\mathbf{u}) = k(\|\mathbf{u}\|^2)$ admits the random projection vector $\boldsymbol{\eta} = \sqrt{2R}\boldsymbol{N}$, where the distribution of R is the inverse Laplace transform of $k/k(0)$, and $\boldsymbol{N}$ is a d-dimensional standard Gaussian vector independent of R.*

Proof. Apply Theorem 1 with $\alpha = 2$ and use the fact that $\boldsymbol{S}_2 \stackrel{d}{=} \sqrt{2}\boldsymbol{N}$ where $\boldsymbol{N}$ is a d-dimensional standard Gaussian vector (from Definition 1). □

Remark 1. Corollary 2 shows that the random projection vector of the continuous, positive definite, isotropic kernel $K(\mathbf{u}) = k(\|\mathbf{u}\|^2)$ is a scale mixture of a standard Gaussian vector. In fact, the same is true for all the kernels $K_\alpha(\mathbf{u}) = k(\|\mathbf{u}\|^\alpha)$, $\alpha \in (0, 2]$, as pointed out later in Proposition 1 (see Corollary 3).

Remark 2. The distribution of the nonnegative random variable R in Theorem 1 is the inverse Laplace transform of the scaled kernel $k/k(0)$. In other words, the Laplace transform $\mathcal{L}$ of R is equal to $\mathcal{L}(s) = k(s)/k(0)$ for every $s \geq 0$. Equivalently, R can be characterised by its characteristic function $\phi(s) = \mathbb{E}\left[e^{isR}\right]$. Indeed, for a nonnegative random variable the equality $\mathcal{L}(s) = \phi(is)$ holds for every $s \geq 0$. This means that for every $\alpha \in (0, 2]$, the continuous, positive definite, isotropic kernel $K_\alpha(\mathbf{u}) = k(\|\mathbf{u}\|^\alpha)$ can be written as

$$K_\alpha(\mathbf{u}) = k(\|\mathbf{u}\|^\alpha) = k(0)\mathcal{L}(\|\mathbf{u}\|^\alpha) = k(0)\phi(i\,\|\mathbf{u}\|^\alpha) \tag{13}$$

for every $\mathbf{u} \in \mathbb{R}^d$, where the distribution of the nonnegative random variable R is the inverse Laplace transform of $k/k(0)$, $\mathcal{L}$ is the Laplace transform of R, and ϕ is the characteristic function of R. The characteristic function formulation (13) is more convenient to construct new positive definite multivariate isotropic kernels from a given nonnegative distribution for R, since characteristic functions are much more commonly precomputed and available for a vast range of distributions [31, 32]. This is the approach adopted in Table 1.

In the following, we provide several examples of interest for which both the kernel K and the distribution of the random variable $R \geq 0$ from Theorem 1 are known analytically. We also use the approach described in Remark 2 to create new multivariate positive definite kernels. These examples are listed in Table 1. Other valid kernel examples can be deduced for example from [15] and [29], for continuous and discrete spectral mixing distributions respectively. Table 1 makes use of the following special functions:

- Γ is the gamma function [8, 5.2]
- $\mathcal{B}(a, b)$ is the beta function [8, 5.12]
- $\mathcal{K}_\beta$ is the modified Bessel function [8, 10.25],
- $\mathcal{M}(a, b, z)$ is the Kummer confluent hypergeometric function [8, 13.2], also denoted as ${}_1F_1(a, b, z)$ in some references,
- $\mathcal{U}(a, b, z)$ is the Tricomi confluent hypergeometric function, a.k.a. Kummer's function of the second kind [8, 13.2].

Table 1 also uses the following random variables:

- $\boldsymbol{N}$ is a standard multivariate Gaussian random vector,
- G_β, $\beta > 0$, is a Gamma random variable, with density $f(x) = \frac{1}{\Gamma(\beta)} x^{\beta-1} e^{-x}$, $x > 0$,
- $B_{\beta,\gamma}$, $\beta > 0$, $\gamma > 0$, is a Beta random variable, with density $f(x) = \frac{x^{\beta-1}(1-x)^{\gamma-1}}{\mathcal{B}(\beta,\gamma)}$, $x \in (0, 1)$. It can be obtained from two independent Gamma random variables G_β and G_γ as $B_{\beta,\gamma} \stackrel{d}{=} \frac{G_\beta}{G_\beta + G_\gamma}$,
- $F_{2\beta,2\gamma}$, $\beta > 0$, $\gamma > 0$, is a Fisher-Snedecor random variable, with density $f(x) = \frac{1}{x\mathcal{B}(\beta,\gamma)} \frac{(\beta x)^\beta \gamma^\gamma}{(\beta x+\gamma)^{\beta+\gamma}}$, $x > 0$. It can be obtained from two independent Gamma random variables G_β and G_γ as $F_{2\beta,2\gamma} \stackrel{d}{=} \frac{\gamma G_\beta}{\beta G_\gamma}$.

Particular cases of covariance kernels of interest from Table 1 include:

- The Laplace kernel $K(\mathbf{u}) = e^{-\|\mathbf{u}\|}$ ($R = 1$, $\lambda = 1$, $\alpha = 1$), also known as exponential kernel [38] (see for example [13, equation (4)]). The random projection $\boldsymbol{\eta} = \boldsymbol{S}_1$ follows a standard multivariate Cauchy distribution ([40, Lemma 3.7.3], [7]).
- The Gaussian kernel $K(\mathbf{u}) = e^{-\frac{\|\mathbf{u}\|^2}{2}}$ ($R = 1$, $\lambda = 1/2$, $\alpha = 2$), also known as squared exponential kernel [38]. The random projection $\boldsymbol{\eta} = \boldsymbol{S}_2(1/2)^{\frac{1}{2}} = \boldsymbol{N}$ follows a standard multivariate Gaussian distribution.

	Name	Formula	
$R =$	Constant 1	$\phi(x) = e^{ix}$	
$K =$	Exponential power	$K(\mathbf{u}) = e^{-\|\mathbf{u}\|^\alpha}$	$\lambda = 1$
$R =$	Gamma G_β , $\beta > 0$	$\phi(x) = (1 - ix)^{-\beta}$	
$K =$	Generalized Cauchy	$K(\mathbf{u}) = \frac{1}{\left(1+\frac{\|\mathbf{u}\|^\alpha}{2\beta}\right)^\beta}$	$\lambda = \frac{1}{2\beta}$
$R =$	Inverse Gamma $1/G_\beta$, $\beta > 0$	$\phi(x) = \frac{2(-ix)^{\beta/2}}{\Gamma(\beta)} \mathcal{K}_\beta(\sqrt{-4ix})$	
$K =$	Generalized Matérn	$K(\mathbf{u}) = \frac{(\sqrt{2\beta}\|\mathbf{u}\|^{\frac{\alpha}{2}})^\beta}{\Gamma(\beta)2^{\beta-1}} \mathcal{K}_\beta(\sqrt{2\beta}\,\|\mathbf{u}\|^{\frac{\alpha}{2}})$	$\lambda = \frac{\beta}{2}$
$R =$	Beta $B_{\beta,\gamma}$, $\beta > 0$, $\gamma > 0$	$\phi(x) = \mathcal{M}(\beta, \beta+\gamma, ix)$	
$K =$	Kummer	$K(\mathbf{u}) = \mathcal{M}(\beta, \beta+\gamma, -\|\mathbf{u}\|^\alpha)$	$\lambda = 1$
$R =$	Beta-exponential $-\log(B_{\beta,\gamma})$, $\beta > 0$, $\gamma > 0$	$\phi(x) = \frac{\mathcal{B}(\beta-ix,\gamma)}{\mathcal{B}(\beta,\gamma)}$	
$K =$	Beta	$K(\mathbf{u}) = \frac{\mathcal{B}(\beta+\|\mathbf{u}\|^\alpha,\gamma)}{\mathcal{B}(\beta,\gamma)}$	$\lambda = 1$
$R =$	F-distribution $F_{2\beta,2\gamma}$, $\beta > 0, \gamma > 0$	$\phi(x) = \frac{\Gamma(\beta+\gamma)}{\Gamma(\gamma)} \mathcal{U}\left(\beta, 1-\gamma, -i\frac{\gamma}{\beta}x\right)$	
$K =$	Tricomi	$K(\mathbf{u}) = \frac{\Gamma(\beta+\gamma)}{\Gamma(\gamma)} \mathcal{U}\left(\beta, 1-\gamma, \frac{\gamma}{\beta}\|\mathbf{u}\|^\alpha\right)$	$\lambda = 1$

Table 1: Characteristic functions ϕ of random mixture distributions $R \geq 0$ and resulting covariance kernels $K(\mathbf{u}) = K(\mathbf{0})\phi(i\,\|\mathbf{u}\|^\alpha)$. The notations R, K and λ are defined in Lemma 1 and Theorem 1.

- The exponential power kernel $K(\mathbf{u}) = e^{-\|\mathbf{u}\|^\alpha}$ ($R = 1$, $\lambda = 1$, $\alpha \in (0, 2]$) [4, (21.4)], also known as generalized Gaussian [9], generalized normal [35], γ-exponential [38] or Subbotin [42] kernel. The random projection $\boldsymbol{\eta} = \boldsymbol{S}_\alpha$ follows a symmetric stable distribution ([6], [7]).

- The Matérn-ν kernel $K(\mathbf{u}) = \frac{(\sqrt{2\nu}\|\mathbf{u}\|)^\nu}{\Gamma(\nu)2^{\nu-1}} \mathcal{K}_\nu(\sqrt{2\nu}\,\|\mathbf{u}\|)$, $\nu > 0$ ($R = 1/G_\nu$, $\lambda = \frac{\nu}{2}$, $\alpha = 2$) [36], also known as multivariate symmetric Laplace kernel [20, equation (5.2.2)]. The random projection $\boldsymbol{\eta} = \boldsymbol{S}_2(\nu/(2G_\nu))^{\frac{1}{2}} = \boldsymbol{N}/\sqrt{G_\nu/\nu}$ follows a standard multivariate Student $t_{2\nu}$ distribution with 2ν degrees of freedom [21, equation (1.2)].

- The power kernel $K(\mathbf{u}) = \frac{1}{1+\|\mathbf{u}\|^\alpha}$ ($R = G_1$, $\lambda = 1$, $\alpha \in (0, 2]$). The random projection $\boldsymbol{\eta} = \boldsymbol{S}_\alpha E^{\frac{1}{\alpha}}$, where E is a standard exponential random variable, follows a Linnik distribution [26], [23], [5], also known as Linnik-Laha distribution.

- The Student t kernel $K(\mathbf{u}) = \left(1 + \frac{\|\mathbf{u}\|^2}{2\beta}\right)^{-\beta}$, $\beta > 0$, with $2\beta - 1$ degrees of freedom ($R = G_\beta$, $\lambda = \frac{1}{2\beta}$, $\alpha = 2$), also known as rational quadratic kernel [38] or generalized inverse multiquadric kernel [18]. The random projection $\boldsymbol{\eta} = \boldsymbol{S}_2(G_\beta/(2\beta))^{\frac{1}{2}} = \boldsymbol{N}\sqrt{\frac{G_\beta}{\beta}}$ follows a Matérn distribution, also known as generalized Laplace distribution [20, Definition 4.1.1] (see [20, Proposition 4.1.2]), [40, Theorem 3.7.5], which is a particular case of the variance-gamma distribution [12], also known as Bessel function distribution [28], see [15, equation (10)].

- The generalized Cauchy kernel $K(\mathbf{u}) = \frac{1}{(1+\|\mathbf{u}\|^\alpha)^\beta}$, $\beta > 0$ ($R = G_\beta$, $\lambda = 1$, $\alpha \in (0, 2]$) [16], also known as generalized Pearson VII kernel [4, page 144]. The random projection $\boldsymbol{\eta} = \boldsymbol{S}_\alpha G_\beta^{\frac{1}{\alpha}}$ follows a generalized Linnik distribution [6, 7]. The scaling $\lambda = \frac{1}{2\beta}$ proposed in Table 1 (by analogy with Matérn kernels) is such that $K(\mathbf{u}) = \frac{1}{\left(1+\frac{\|\mathbf{u}\|^\alpha}{2\beta}\right)^\beta} \underset{\beta\to\infty}{\longrightarrow} e^{-\frac{\|\mathbf{u}\|^\alpha}{2}}$, which is an exponential power kernel (with scaling $\lambda = 1/2$) which contains the Gaussian kernel as the particular case $\alpha = 2$.

- Since $F_{2\beta,2\gamma} \overset{d}{=} (G_\beta/\beta)/(G_\gamma/\gamma)$ and $\lim_{a\to\infty} \frac{G_a}{a} = 1$, the proposed Tricomi kernel $K(\mathbf{u}) = \frac{\Gamma(\beta+\gamma)}{\Gamma(\gamma)} \mathcal{U}\left(\beta, 1-\gamma, \frac{\gamma}{\beta}\|\mathbf{u}\|^\alpha\right)$ contains generalized Matérn kernels (when $\beta \to \infty$) and generalized Cauchy kernels (when $\gamma \to \infty$) as limit cases, and therefore contains all the classical stationary kernels (Laplace, Gaussian, Matérn, Student, Power, Exponential Power, etc.) as particular limit cases.

Remark 3. Symmetric stable random vectors (Definition 1) only exist when $\alpha \in (0, 2]$. As a result, for every kernel in Table 1, the parameter α is restricted to the interval $(0, 2]$ to enforce positive definiteness. When setting α to a value larger than 2, the inverse Fourier transform (4) of the kernel is not a density anymore as it can take negative values. Explicit examples from the literature include the exponential power kernel with $\alpha = 4$ and $\alpha = 6$ [19], and $\alpha = 3$ [9], as well as the generalized Cauchy kernel with $\alpha = 4$ (Laha distribution [22], its inverse Fourier transform is the Silverman kernel [43]) and $\alpha = 6$ [19].

Remark 4. This section focuses on isotropic kernels, which can be written as $K(\mathbf{u}) = k(\|\mathbf{u}\|)$ where k is a univariate kernel and $\|\mathbf{u}\|$ is the Euclidean norm of the vector $\mathbf{u} \in \mathbb{R}^d$. Another classical way to construct multivariate kernels is the tensor approach, where $K(\mathbf{u}) = \prod_{\ell=1}^d k(u_\ell)$ is the product of univariate kernels, as shown with the examples (8) and (9) from Rahimi and Recht [37]. The random projection $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_d)$ of a tensor kernel is a vector with i.i.d. components with distribution equal to the spectral distribution of the univariate kernel k. In view of this, the results from this section, in particular Table 1, can also be used to generate new tensor kernels, by setting $d = 1$ in equation (11) and simulating the resulting univariate random projection $\eta = (\lambda R)^{\frac{1}{\alpha}} \boldsymbol{S}_\alpha$ d times independently. However, the results of this paper suggest that there is little reason to favour tensor kernels over isotropic kernels when resorting to the random Fourier feature approximation, if only because simulating the spectral distribution of isotropic kernels is faster than for tensor kernels, as shown in Remark 7 in the next section.

Remark 5. Theorem 1 and Corollary 1 suggest that, in order to compare two positive definite isotropic kernels K_1 and K_2 in Φ_∞, it is sufficient to compare the distributions of their respective scale mixtures $R_1 \geq 0$ and $R_2 \geq 0$. This is a much simpler task, since kernels are defined in $\mathbb{R}^d$ while scale mixtures are univariate. As an illustration, the result in [10] connecting Cauchy and Matérn covariance functions can be proved alternatively without working with the non-analytical density of the generalized Linnik distribution [25], by using the fact that the scale mixture of generalized Cauchy kernels has a Gamma distribution, and the fact that $G_\beta/\beta \overset{a.s.}{\to} 1$ when $\beta \to \infty$ for a Gamma random variable G_β. This alternative approach also implies that no other connection between Cauchy and Matérn covariance functions can exist besides the limiting case identified in [10], since there is no way to transform a Gamma distribution into an inverse Gamma distribution by simple parameter rescaling.

Remark 6. Let $\Sigma \in \mathbb{R}^{d\times d}$ be a symmetric, positive definite matrix. By noting that $\|\mathbf{u}\|^\alpha = (\mathbf{u}^\top \mathbf{u})^{\frac{\alpha}{2}}$, one can replace $\|\mathbf{u}\|^\alpha$ by $(\mathbf{u}^\top \Sigma \mathbf{u})^{\frac{\alpha}{2}}$ in the definition of all the kernels from Table 1, by changing the random projection formula (11) to $\boldsymbol{\eta} = (\lambda R)^{\frac{1}{\alpha}} \boldsymbol{S}_\alpha \Sigma^{\frac{1}{2}}$. This is another possible approach to introduce additional parameters into the formula of a multivariate parametric kernel.

3 Sampling spectral distributions

In order to simulate the random projections from Table 1, we need to be able to simulate multivariate symmetric stable random vectors $\boldsymbol{S}_\alpha$, $\alpha \in (0, 2]$, as well as the nonnegative random variables R.

Gamma random variables All the nonnegative random variables R in Table 1 can be obtained from independent simulations of Gamma random variables, which are elementary distributions for which simulation routines are widely available. Popular approaches to simulate Gamma random variables include acceptance-rejection and numerical inversion, see Luengo [27].

Symmetric stable random vectors The following Proposition describes how to obtain simulations of symmetric stable random vectors, in the form of scale mixtures of multivariate Gaussian random vectors.

Proposition 1. *Let* $\boldsymbol{N}$ *be a d-dimensional standard Gaussian vector, let* U_1 *and* U_2 *be two independent standard uniform random variables, independent of* $\boldsymbol{N}$*, let* $W = -\log(U_1)$ *be a standard exponential random variable, and let* $\Theta = \pi\left(U_2 - \frac{1}{2}\right)$ *be a uniform random variable in* $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$*. Then, for any* $\alpha \in (0, 2]$*, the multivariate symmetric stable distribution* $\boldsymbol{S}_\alpha$ *(Definition 1) admits the following decomposition*

$$\boldsymbol{S}_\alpha \overset{d}{=} \sqrt{2A_\alpha}\boldsymbol{N} \tag{14}$$

where

$$A_\alpha := \frac{\sin\left(\frac{\alpha\pi}{4} + \frac{\alpha}{2}\Theta\right)}{(\cos(\Theta))^{2/\alpha}} \left(\frac{\cos\left(\frac{\alpha\pi}{4} + \left(\frac{\alpha}{2} - 1\right)\Theta\right)}{W}\right)^{\frac{2}{\alpha}-1} \tag{15}$$

Proof. To obtain the Gaussian mixture representation (14)-(15) of symmetric stable vectors, apply [39, Proposition 2.5.2], use the fact that $\boldsymbol{S}_2 \sim \mathcal{N}(\mathbf{0}, 2\mathbf{I}_d)$ is a multivariate Gaussian random vector with independent components with mean zero and variance 2 (where $\mathbf{I}_d$ is the d-dimensional identity matrix), and use the simulation formula for nonsymmetric stable random variable from [30, Theorem 1.3] with maximum skew $\beta = 1$. □

Corollary 3. *A consequence of Theorem 1 ($\boldsymbol{\eta}_\alpha = R^{\frac{1}{\alpha}}\boldsymbol{S}_\alpha$) and Proposition 1 ($\boldsymbol{S}_\alpha = \sqrt{2A_\alpha}\boldsymbol{N}$) is that any continuous, positive definite, isotropic kernel $K(\mathbf{u}) = k(\|\mathbf{u}\|^\alpha)$ in Φ_∞, $\mathbf{u} \in \mathbb{R}^d$, where $\alpha \in (0, 2]$, admits a representation of its random projections as a scale mixture of Gaussians, explicitly given by*

$$\boldsymbol{\eta}_\alpha = (R^{\frac{1}{\alpha}}\sqrt{2A_\alpha})\boldsymbol{N}, \tag{16}$$

where the distribution of the nonnegative random variable $R \geq 0$ is given by the inverse Laplace transform of $k/k(0)$, A_α is defined by equation (15), and R, A_α and $\boldsymbol{N}$ are independent.

This corollary has useful practical implications. For example, it suggests that a task such as kernel learning via learning spectral distributions can be brought down to learning the parameter $\alpha \in (0, 2]$ and the distribution of the univariate nonnegative random radius R. It can also suggest variance reduction techniques by splitting the effort between the Gaussian vector $\boldsymbol{N}$ [33] and the random scaling factor $R^{\frac{1}{\alpha}}\sqrt{2A_\alpha}$.

Remark 7. In view of equation (16), simulating the random projection vector of an isotropic kernel in Φ_∞ requires to simulate $d + 2$ independent random variables (the vector $\boldsymbol{N}$ of size d, the random variable A_α and the random variable R). By contrast, simulating the random projection vector of a tensor kernel requires to simulate $3d$ independent random variables (d independent simulations of the scalar random variable $\eta = (R^{\frac{1}{\alpha}}\sqrt{2A_\alpha})N$ which involves the three random variables N, A_α and R). This shows that the isotropic formulation of multivariate kernels is almost three times more efficient than the tensor formulation when using the random Fourier features approach with the scale mixture representation formula (16).

4 Numerical examples

This section gives several examples of isotropic kernels from Table 1 along with their random Fourier features approximation (6), using the random projection representation (11) and the simulation algorithms described above.

Remark that all the known theoretical results about the convergence of the random Fourier features approximation with respect to the number of random projections, such as [41], [24] or [44], still apply to the kernels discussed in this paper.

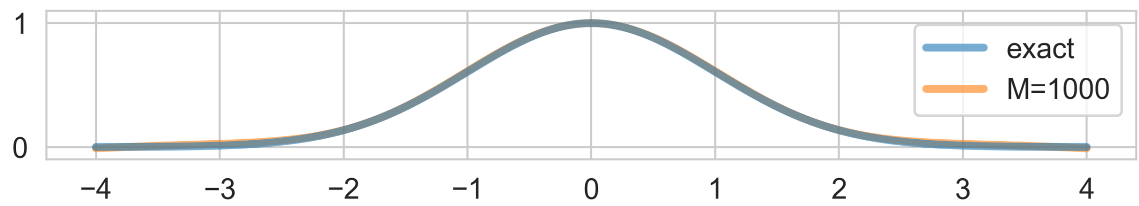

Figure 1: Univariate Gaussian kernel and its random Fourier features approximation (6) using $M = 1000$ random projections with Gaussian distribution.

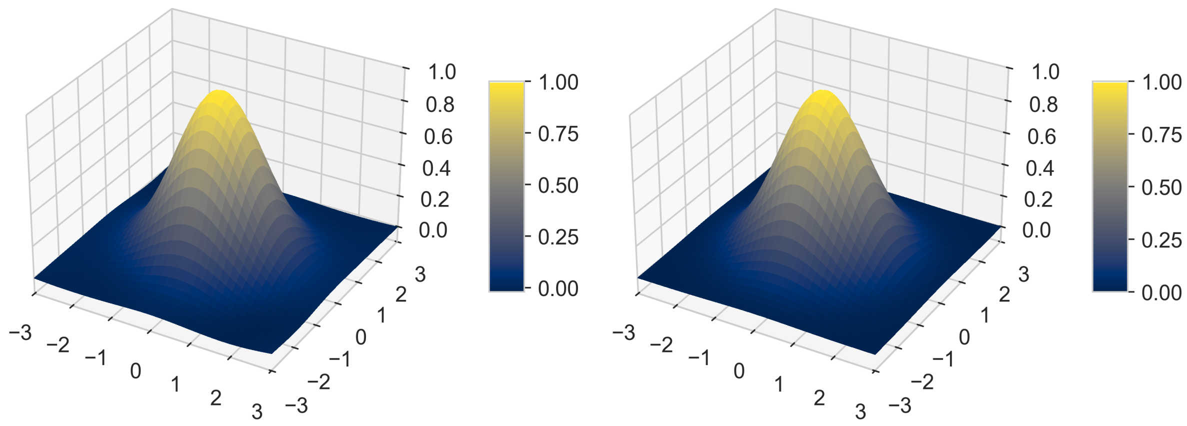

Figure 2: Bivariate Gaussian kernel (right) and its random Fourier features approximation (6) (left) using $M = 4000$ random projections.

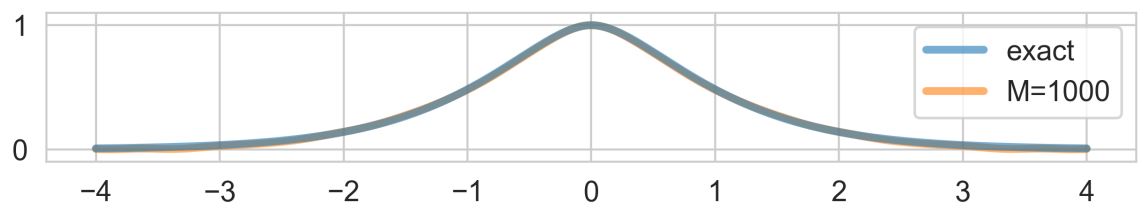

Figure 3: Univariate Matérn-3/2 kernel and its random Fourier features approximation (6) using $M = 1000$ random projections.

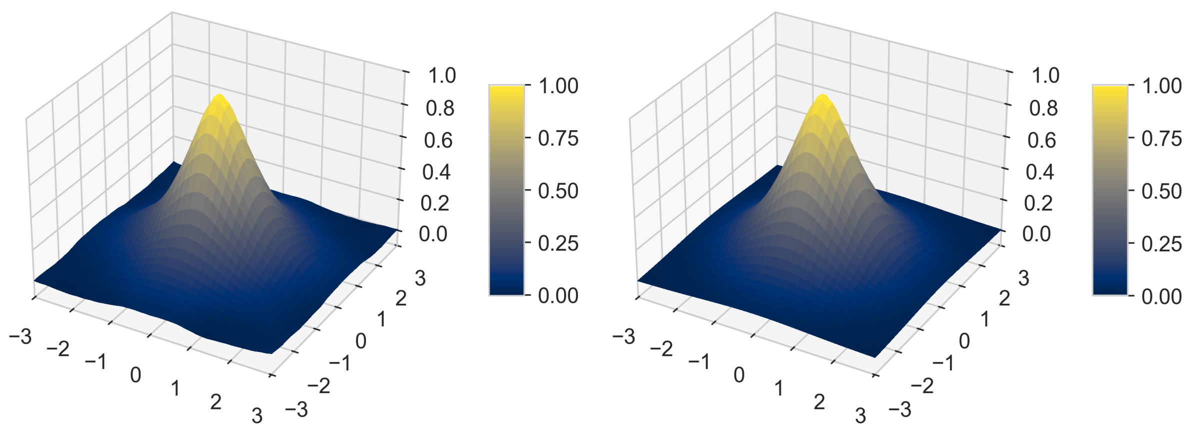

Figure 4: Bivariate Matérn-3/2 kernel (right) and its random Fourier features approximation (6) (left) using $M = 4000$ random projections.

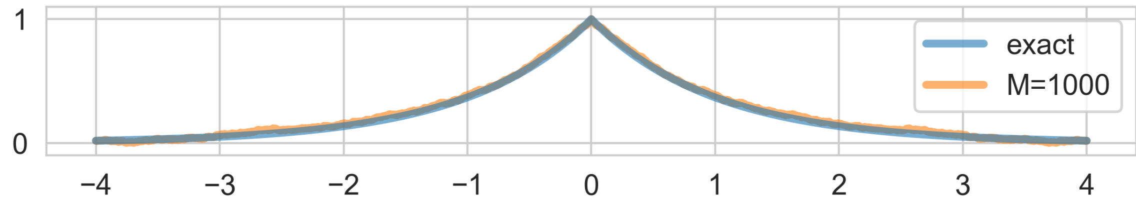

Figure 5: Univariate Laplace (a.k.a. Matérn-1/2) kernel and its random Fourier features approximation (6) using $M = 1000$ random projections.

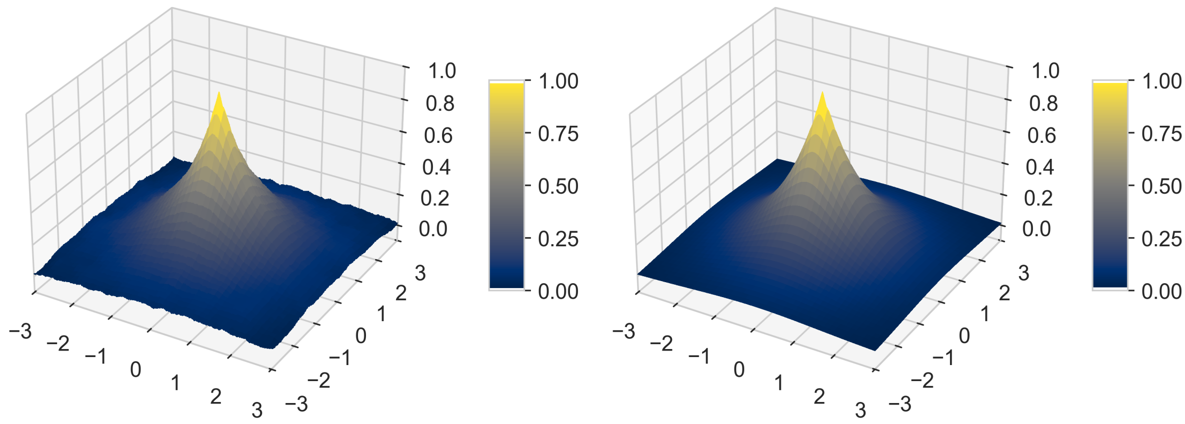

Figure 6: Bivariate Laplace (a.k.a. Matérn-1/2) kernel (right) and its random Fourier features approximation (6) (left) using $M = 4000$ random projections.

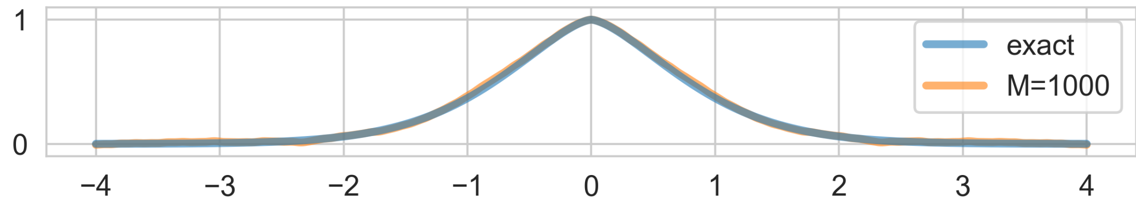

Figure 7: Univariate exponential power kernel with $\alpha = 1.5$ and its random Fourier features approximation (6) using $M = 1000$ random projections.

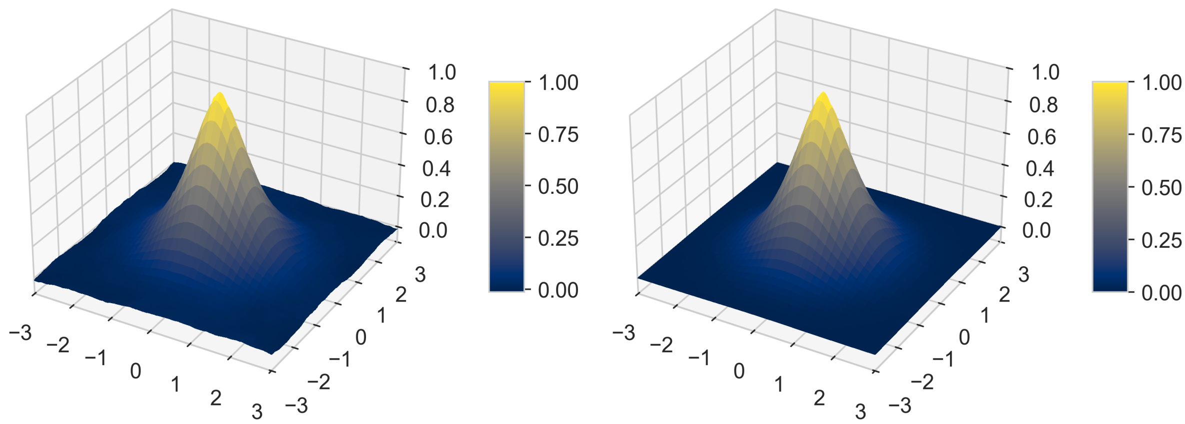

Figure 8: Bivariate exponential power kernel with $\alpha = 1.5$ (right) and its random Fourier features approximation (6) (left) using $M = 4000$ random projections.

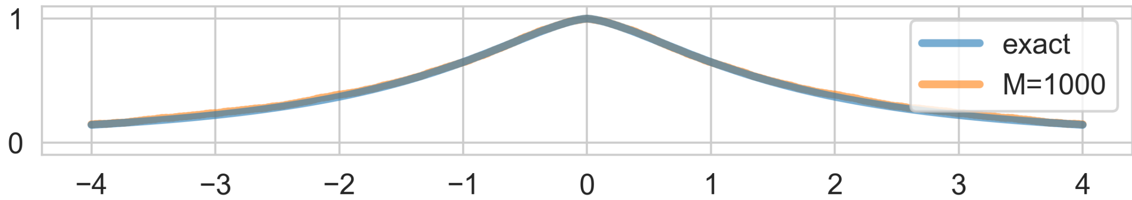

Figure 9: Univariate generalized Cauchy kernel with $(\alpha, \beta) = (1.5, 1.5)$ and its random Fourier features approximation (6) using $M = 1000$ random projections.

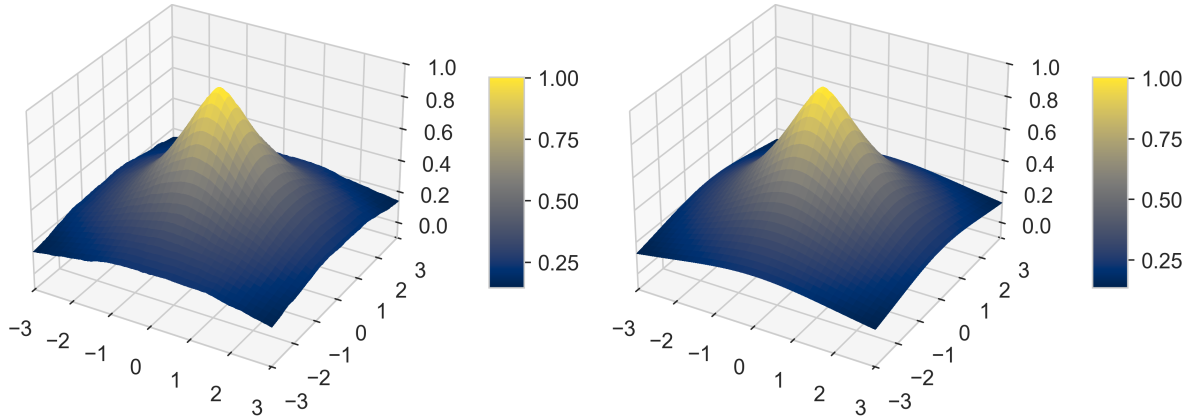

Figure 10: Bivariate generalized Cauchy power kernel with $(\alpha, \beta) = (1.5, 1.5)$ (right) and its random Fourier features approximation (6) (left) using $M = 4000$ random projections.

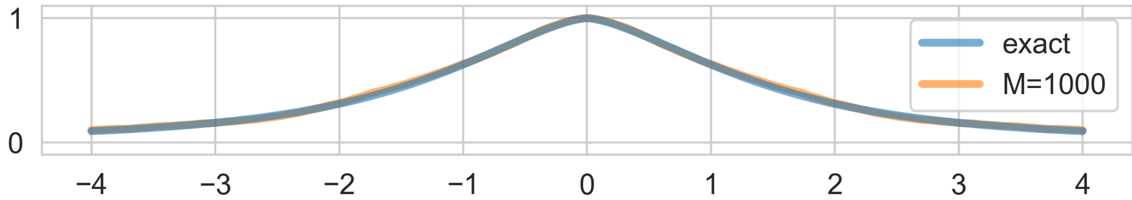

Figure 11: Univariate Kummer kernel with $(\alpha, \beta, \gamma) = (1.5, 1.5, 1.5)$ and its random Fourier features approximation (6) using $M = 1000$ random projections.

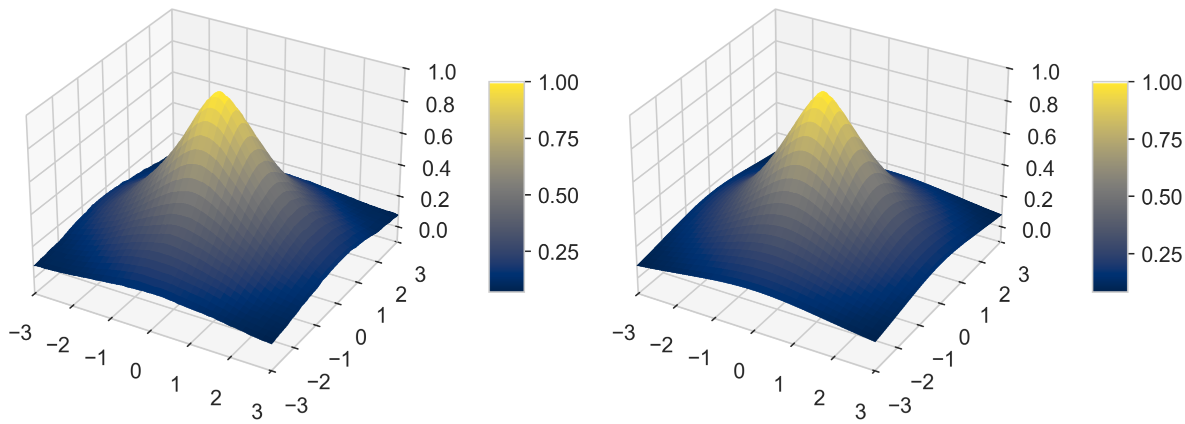

Figure 12: Bivariate Kummer kernel with $(\alpha, \beta, \gamma) = (1.5, 1.5, 1.5)$ (right) and its random Fourier features approximation (6) (left) using $M = 4000$ random projections.

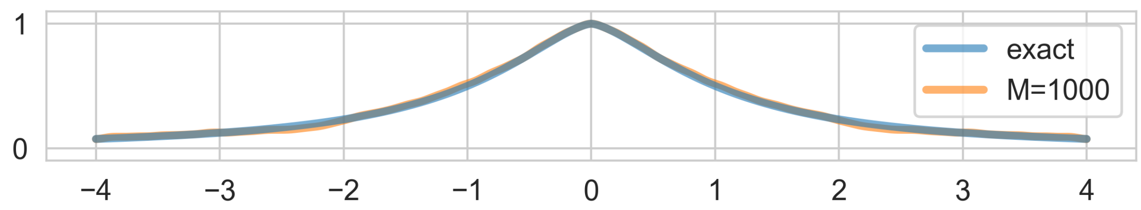

Figure 13: Univariate beta kernel with $(\alpha, \beta, \gamma) = (1.5, 1.5, 1.5)$ and its random Fourier features approximation (6) using $M = 1000$ random projections.

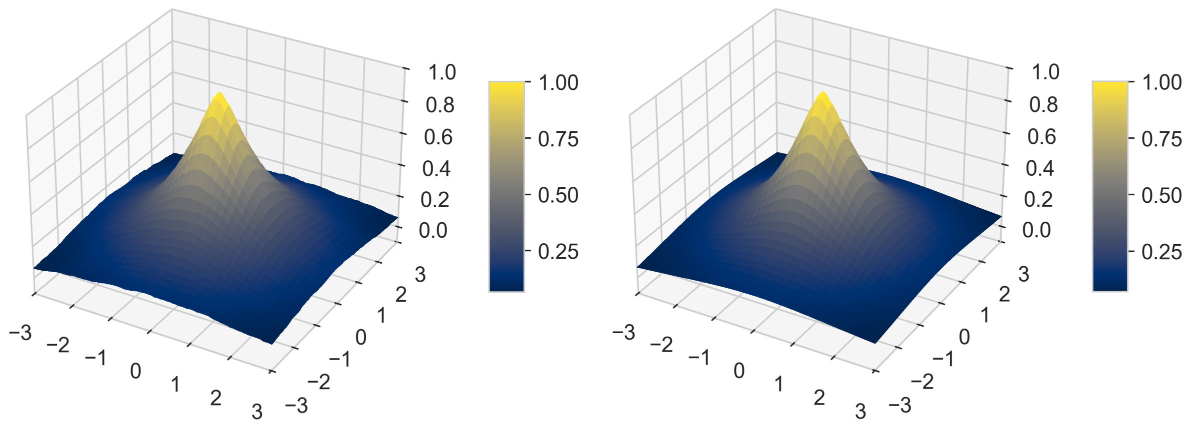

Figure 14: Bivariate beta kernel with $(\alpha, \beta, \gamma) = (1.5, 1.5, 1.5)$ (right) and its random Fourier features approximation (6) (left) using $M = 4000$ random projections.

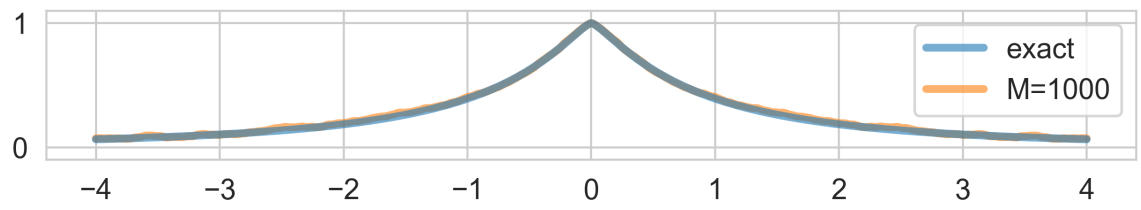

Figure 15: Univariate Tricomi kernel with $(\alpha, \beta, \gamma) = (1.5, 1.5, 1.5)$ and its random Fourier features approximation (6) using $M = 1000$ random projections.

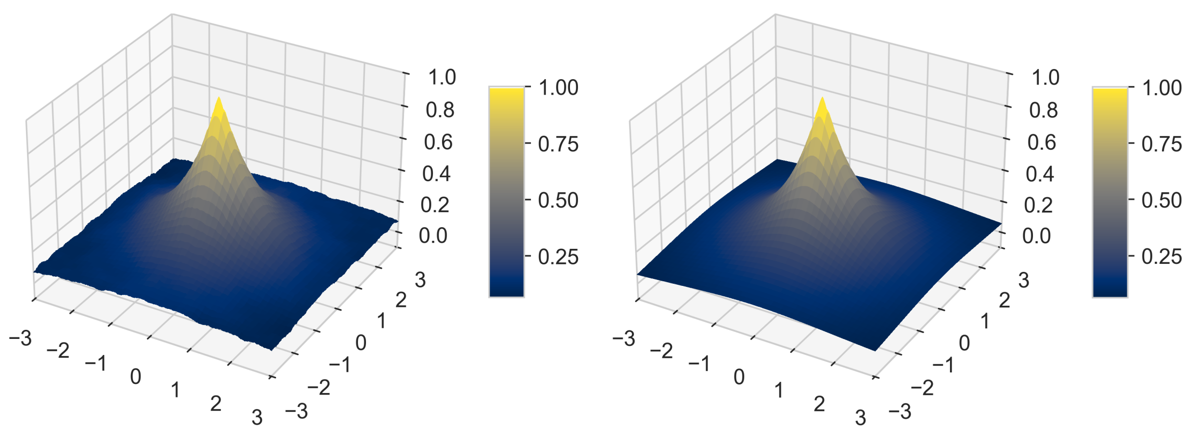

Figure 16: Bivariate Tricomi kernel with $(\alpha, \beta, \gamma) = (1.5, 1.5, 1.5)$ (right) and its random Fourier features approximation (6) (left) using $M = 4000$ random projections.

5 Conclusion

In this paper, we proved that if the isotropic kernel $K(\mathbf{u}) = k(\|\mathbf{u}\|^2)$, $\mathbf{u} \in \mathbb{R}^d$, is continuous and positive definite for all $d \geq 1$, then the same is true for all the kernels $K_\alpha(\mathbf{u}) = k(\|\mathbf{u}\|^\alpha)$, $\mathbf{u} \in \mathbb{R}^d$, $\alpha \in (0, 2]$, and their random projections $\boldsymbol{\eta}_\alpha$ (the random vector whose distribution is the spectral distribution of K_α) admits the mixture representation $\boldsymbol{\eta}_\alpha = R^{\frac{1}{\alpha}} \boldsymbol{S}_\alpha$, where $\boldsymbol{S}_\alpha$ is a symmetric stable random vector (with characteristic function $\mathbb{E}\left[e^{i\boldsymbol{S}_\alpha^\top \mathbf{u}}\right] = e^{-\|\mathbf{u}\|^\alpha}, \mathbf{u} \in \mathbb{R}^d$, see Definition 1), and R is a non-negative random variable, independent of $\boldsymbol{S}_\alpha$, whose distribution is defined by the inverse Laplace transform of $k/k(0)$ (Theorem 1). This mixture representation can be further decomposed into $\boldsymbol{\eta}_\alpha = R^{\frac{1}{\alpha}} \sqrt{2A_\alpha} \boldsymbol{N}$, where the random variable A_α is defined by equation (15), and $\boldsymbol{N}$ is a standard multivariate Gaussian vector (Proposition 1). This theoretical result has several interesting implications.

First, it improves our understanding of the spectral distributions of kernels in Φ_∞. Many classical parametric kernels are such that the distribution of the random radius R is analytical and easy to simulate. This includes the Gaussian, Laplace, Matérn, Generalized Cauchy, and several other kernels of interest in machine learning and statistical learning. Then, it makes it very easy to create new multivariate isotropic kernels which are guaranteed to be positive definite in $\mathbb{R}^d$ for all $d \geq 1$, by simply changing the distribution of R. We used this approach to introduce several new multivariate kernels, including the Beta, Kummer, and Tricomi kernels (Table 1). Finally, this result makes it very easy to implement the random Fourier features methodology, which is based on the simulation of spectral distributions, not only for isotropic kernels but also for tensor kernels (Remark 4). In particular, our mixture representation suggests that a Gaussian kernel implementation of the random Fourier features method can easily be adapted to any other compatible kernel by simply multiplying the Gaussian simulations by independent simulations of the scaling random variable $R^{\frac{1}{\alpha}} \sqrt{2A_\alpha}$. We implemented these random projections for each kernel in Table 1, and illustrated the resulting kernels in dimension 1 and 2. The approximation error and convergence are in line with the known theoretical results about random Fourier features.

Our work provides substantial benefits to the random Fourier features methodology: it simplifies its practical implementation, readily adapts it to every positive definite kernels in Φ_∞, simplifies the creation of new multivariate positive definite kernels, and simplifies kernel learning to the identification of the distribution of the random radius R. We believe that further developments ought to be achievable at this intersection between the theories of spectral distributions and stable distributions. Future work could include the extension of this mixture representation of spectral distributions to more general classes of kernels, including asymmetric, non positive definite, and perhaps non-stationary kernels, with useful implications for kernel-based machine learning.

Acknowledgements

The authors would like to thank Dr. Christian Walder (Google DeepMind) for fruitful discussions, and Prof. Emilio Porcu for letting us know about an error in a previous version of this work. Nicolas Langrené acknowledges the partial support of the

Guangdong Provincial/Zhuhai Key Laboratory IRADS (2022B1212010006) and the BNBU Start-up Research Fund UICR0700041-22. Pierre Gruet and Xavier Warin acknowledge support from the FiME Lab.

References

[1] Bochner, S. (1933). Monotone Funktionen, Stieltjessche Integrale und harmonische Analyse. *Mathematische Annalen*, 108(1):378–410. (Cited on page 2)

[2] Bochner, S. (1959). *Lectures on Fourier integrals*, volume 42. Princeton University Press. (Cited on page 2)

[3] Chambers, J., Mallows, C., and Stuck, B. (1976). A method for simulating stable random variables. *Journal of the American Statistical Association*, 71(354):340–344. (Cited on page 4)

[4] Crooks, G. E. (2019). *Field guide to continuous probability distributions*. Berkeley Institute for Theoretical Science. (Cited on pages 8 and 9)

[5] Devroye, L. (1990). A note on Linnik's distribution. *Statistics & Probability Letters*, 9(4):305–306. (Cited on page 9)

[6] Devroye, L. (2006). Nonuniform random variate generation. In *Handbooks in Operations Research and Management Science*, volume 13, chapter 4, pages 83–121. Elsevier. (Cited on pages 8 and 9)

[7] Devroye, L. and James, L. (2014). On simulation and properties of the stable law. *Statistical Methods & Applications*, 23(3):307–343. (Cited on pages 4, 7, 8, and 9)

[8] DLMF (2024). NIST Digital Library of Mathematical Functions. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds. (Cited on page 7)

[9] Dytso, A., Bustin, R., Poor, H. V., and Shamai, S. (2018). Analytical properties of generalized Gaussian distributions. *Journal of Statistical Distributions and Applications*, 5(1):1–40. (Cited on pages 8 and 9)

[10] Faouzi, T., Porcu, E., Kondrashuk, I., and Bevilacqua, M. (2024). Convergence arguments to bridge Cauchy and Matérn covariance functions. *Statistical Papers*, 65(2):645–660. (Cited on page 10)

[11] Feller, W. (1971). *An introduction to probability theory and its applications*, volume II. John Wiley & Sons. (Cited on page 2)

[12] Fischer, A., Gaunt, R., and Sarantsev, A. (2024). The Variance-Gamma distribution: a review. *Statistical Science*. (Cited on page 9)

[13] Francis, D. P. and Raimond, K. (2021). Major advancements in kernel function approximation. *Artificial Intelligence Review*, 54:843–876. (Cited on pages 3 and 7)

[14] Genton, M. (2001). Classes of kernels for machine learning: a statistics perspective. *Journal of Machine Learning Research*, 2:299–312. (Cited on page 2)

[15] Gneiting, T. (1997). Normal scale mixtures and dual probability densities. *Journal of Statistical Computation and Simulation*, 59(4):375–384. (Cited on pages 3, 4, 7, and 9)

[16] Gneiting, T. and Schlather, M. (2004). Stochastic models that separate fractal dimension and the Hurst effect. *SIAM Review*, 46(2):269–282. (Cited on page 9)

[17] Harrar, S. W., Seneta, E., and Gupta, A. K. (2006). Duality between matrix variate t and matrix variate VG distributions. *Journal of Multivariate Analysis*, 97(6):1467–1475. (Cited on page 3)

[18] Hu, X.-G., Ho, T.-S., and Rabitz, H. (1998). The collocation method based on a generalized inverse multiquadric basis for bound-state problems. *Computer Physics Communications*, 113(2-3):168–179. (Cited on page 9)

[19] Jones, M. and Wand, M. (1992). Asymptotic effectiveness of some higher order kernels. *Journal of Statistical Planning and Inference*, 31(1):15–21. (Cited on page 9)

[20] Kotz, S., Kozubowski, T., and Krzysztof, P. (2001). *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*. Springer. (Cited on pages 8 and 9)

[21] Kotz, S. and Nadarajah, S. (2004). *Multivariate t-distributions and their applications*. Cambridge University Press. (Cited on page 8)

[22] Laha, R. (1958). An example of a nonnormal distribution where the quotient follows the Cauchy law. *Proceedings of the National Academy of Sciences*, 44(2):222–223. (Cited on page 9)

[23] Laha, R. G. (1961). On a class of unimodal distributions. *Proceedings of the American Mathematical Society*, 12(2):181–184. (Cited on page 9)

[24] Li, Z., Ton, J.-F., Oglic, D., and Sejdinovic, D. (2021). Towards a unified analysis of random Fourier features. *Journal of Machine Learning Research*, 22(108):1–51. (Cited on page 11)

[25] Lim, S. and Teo, L. P. (2010). Analytic and asymptotic properties of multivariate generalized Linnik's probability densities. *Journal of Fourier Analysis and Applications*, 16(5):715–747. (Cited on page 10)

[26] Linnik, Y. (1953). Linear forms and statistical criteria. *Ukrainian Mathematical Journal*, 5(2):207–243. (Cited on page 9)

[27] Luengo, E. A. (2022). Gamma pseudo random number generators. *ACM Computing Surveys*, 55(4):1–33. (Cited on page 10)

[28] McKay, A. T. (1932). A Bessel function distribution. *Biometrika*, 24(1–2):39–44. (Cited on page 9)

[29] McNolty, F. (1973). Some probability density functions and their characteristic functions. *Mathematics of Computation*, 27(123):495–504. (Cited on page 7)

[30] Nolan, J. (2020). *Univariate stable distributions: models for heavy tailed data*, volume 10 of *Springer Series in Operations Research and Financial Engineering*. Springer. (Cited on pages 4 and 11)

[31] Oberhettinger, F. (1973). *Fourier transforms of distributions and their inverses: a collection of tables*, volume 16 of *Probability and Mathematical Statistics*. Academic Press. (Cited on page 7)

[32] Oberhettinger, F. (1990). *Tables of Fourier transforms and Fourier transforms of distributions*. Springer-Verlag. (Cited on page 7)

[33] Pagès, G. and Printems, J. (2003). Optimal quadratic quantization for numerics: the Gaussian case. *Monte Carlo Methods and Applications*, 9(2):135–165. (Cited on page 11)

[34] Pakes, A. (1998). Mixture representations for symmetric generalized Linnik laws. *Statistics and Probability Letters*, 37(3):213–221. (Cited on page 5)

[35] Pogány, T. K. and Nadarajah, S. (2010). On the characteristic function of the generalized normal distribution. *Comptes Rendus Mathématique*, 348(3-4):203–206. (Cited on page 8)

[36] Porcu, E., Bevilacqua, M., Schaback, R., and Oates, C. (2024). The Matérn model: a journey through statistics, numerical analysis and machine learning. *Statistical Science*, 39(3):469–492. (Cited on page 8)

[37] Rahimi, A. and Recht, B. (2007). Random features for large-scale kernel machines. In *Advances in Neural Information Processing Systems*, volume 20, pages 1177–1184. (Cited on pages 1, 3, and 9)

[38] Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian processes for machine learning*. The MIT Press. (Cited on pages 7, 8, and 9)

[39] Samorodnitsky, G. and Taqqu, M. (1994). *Stable non-Gaussian processes: stochastic models with infinite variance*. Stochastic modeling. Chapman & Hall. (Cited on pages 4, 5, and 11)

[40] Sasvári, Z. (2013). *Multivariate characteristic and correlation functions*, volume 50 of *De Gruyter Studies in Mathematics*. De Gruyter. (Cited on pages 2, 4, 5, 6, 7, and 9)

[41] Sriperumbudur, B. and Szabó, Z. (2015). Optimal rates for random Fourier features. *Advances in Neural Information Processing Systems*, 28:1144–1152. (Cited on page 11)

[42] Subbotin, M. F. (1923). On the law of frequency of error. *Matematicheskiĭ Sbornik*, 31(2):296–301. (Cited on page 8)

[43] Tsybakov, A. (2009). *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer. (Cited on page 9)

[44] Yao, J., Erichson, N. B., and Lopes, M. E. (2023). Error estimation for random Fourier features. In *International Conference on Artificial Intelligence and Statistics*, pages 2348–2364. PMLR. (Cited on page 11)