



HAL
open science

Mesh-free error integration in arbitrary dimensions: A numerical study of discrepancy functions

Philippe G. Lefloch, Jean-Marc Mercier

► **To cite this version:**

Philippe G. Lefloch, Jean-Marc Mercier. Mesh-free error integration in arbitrary dimensions: A numerical study of discrepancy functions. *Computer Methods in Applied Mechanics and Engineering*, 2020, 369, pp.113245 -. <10.1016/j.cma.2020.113245>. <hal-03491562>

HAL Id: hal-03491562

<https://hal.science/hal-03491562v1>

Submitted on 18 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

Mesh-free error integration in arbitrary dimensions: a numerical study of discrepancy functions

Philippe G. LeFloch* and Jean-Marc Mercier†

Abstract

We are interested in mesh-free formulas based on the Monte-Carlo methodology for the approximation of multi-dimensional integrals, and we investigate their accuracy when the functions belong to a reproducing-kernel space. A kernel typically captures regularity and qualitative properties of functions “beyond” the standard Sobolev regularity class. We are interested in the issue whether quantitative error bounds can be a priori guaranteed in applications (including mathematical finance, scientific computing, and machine learning). Our main contribution is a numerical study of the error discrepancy function based on a comparison between several numerical strategies, when one varies the choice of the kernel, the number of approximation points, and the dimension of the problem. We consider two strategies in order to localize to a bounded set the standard kernels defined in the whole Euclidian space (exponential, multiquadric, Gaussian, truncated), namely, on one hand the class of *periodic kernels* defined via a discrete Fourier transform on a lattice and, on the other hand, a class of *transport-based kernels*. Relying on the Poisson formula on a lattice together with heuristic arguments, we study the derivation of theoretical bounds for the discrepancy function of periodic kernels. Then, for four kernels of particular interest we perform extensive numerical experiments and generate the optimal distributions of points and the discrepancy error functions. Our numerical results validate our theoretical observations and, importantly, provide us with quantitative estimates for the error made with a kernel-based strategy as opposed to a purely random strategy.

Contents

1	Introduction	1
2	Functional framework based on a reproducing kernel	4
3	Methodology for defining classes of kernels	6
4	Designing kernels on a bounded domain	10
5	Three formulation of the discrepancy function	18
6	Controlling the discrepancy function	20
7	A comparative study for a selection of kernels	24

1 Introduction

An error approximation formula. We are motivated here by applications to partial differential equations arising continuum physics, including the development of mesh-free methods in fluid dynamics and material sciences [2, 6, 7, 12, 13, 17, 18, 24]. Specifically, we are interested in approximating multi-dimensional integrals via Monte-Carlo formula and deriving error estimates, in which the dependency with respect to the dimension of the problem and other important parameters is specified in a quantitative manner. By

*Laboratoire Jacques-Louis Lions, Centre National de la Recherche Scientifique, Sorbonne Université, 4 Place Jussieu, 75252 Paris, France. Email: contact@philippelefloch.org.

†MPG-Partners, 136 Boulevard Haussmann, 75008 Paris, France. Email: jean-marc.mercier@mpg-partners.com.

revisiting this problem of multivariate integration, our purpose is to clarify the derivation and validity of such estimates whose importance has been highlighted in recent years in artificial intelligence, for mesh-free computations of partial differential equations, and mathematical finance. The existing literature emphasizes the role of Sobolev-type spaces, while we would like here to stress the importance of kernel-based Hilbert spaces. In many applications, one is interested in preserving certain a priori structure that are available a priori and the choice of a kernel is dictated by properties (symmetry, scaling, regularity, decay, etc.) that should be incorporated in the approximation algorithms. Therefore, it is desirable to have a flexible framework that encompasses a wide class of kernels, as we consider in the present paper.

More specifically, within a given Hilbert (or Banach) space we seek to optimize the choice of the interpolation points in an integral approximation formula and establish a sharp error estimate within the chosen class of regularity and decay. Two parameters are of primary interest, namely, the dimension $D \geq 1$ of the problem and the number of interpolation points $N \geq 1$, and it is essential to have quantitative estimates with a specified dependency in N, D that can be determined from the kernels of interest. In the present paper, we contribute to this general objective and provide a systematic study and comparison of several classes of kernels, which we refer to as periodic kernels and transported kernels. Our periodic framework for periodic kernels is motivated by work by Cohn and Elkies [3] who studied the problem of sphere packing. Our result depends upon a “kernel density” function which arises as a key factor in a quantitative bound. We build here on many earlier works on the subject, including contributions in approximation theory [1, 4, 5, 15, 19, 20, 22, 23].

The discrepancy function associated with a kernel. To any kernel $K : \Omega \times \Omega \rightarrow \mathbb{R}$ defined on a bounded and open subset $\Omega \subset \mathbb{R}^D$ and satisfying a positivity condition (see Section 2.1), we associate a Banach space $\mathcal{H}_K^{s,p}(\Omega)$ of real-valued functions defined on Ω with regularity exponent $s > 0$ and integrability exponent $p \in [1, +\infty)$. (More generally, the Lebesgue measure could be replaced by a probability measure.) Then, an “abstract” error integration estimate reads, for any $N \geq 1$ and any function $\varphi \in \mathcal{H}_K^{s,p}(\Omega)$,

$$\sup_{\|\varphi\|_{\mathcal{H}_K^{s,p}(\Omega)}=1} \inf_{x^1, \dots, x^N \in \Omega} \left| \frac{1}{|\Omega|} \int_{\Omega} \varphi(x) dx - \frac{1}{N} \sum_{1 \leq n \leq N} \varphi(x^n) \right| \leq E_K^{s,p}(N, D), \quad (1.1)$$

in which the *discrepancy function* $E_K^{s,p}(N, D)$ is independent of φ . Hence, (1.1) provides us—in the class of functions under consideration—a factorization of the error in two contributions: $\|\varphi\|_{\mathcal{H}_K^{s,p}(\Omega)}$ measures the regularity of the function while the discrepancy function is related to the best distribution of N points in Ω . The challenge is to control $E_K^{s,p}(N, D)$, which can be expressed in several forms:

1. In the physical space Ω , the function $E_K^{s,p}(N, D)$ can be formulated with a pseudo-distance associated with the kernel.
2. In suitable spectral variables determined from an operator naturally associated with the kernel, the function $E_K^{s,p}(N, D)$ takes a rather explicit form involving the eigenfunctions and eigenvectors of this operator.

However, both formulations are difficult to work with directly—except in dimension $D = 1$. So, we introduce below a third standpoint which is more efficient in order to control and its dependency with respect to D and N , that is, we introduce the class of “lattice-based” kernels (in a tensorial form), as we call them. In this context, we can express the function $E_K^{s,p}(N, D)$ via:

3. a Poisson formula in dual discrete Fourier variables associated with a lattice (see next section).

Interestingly, for this latter class of kernels, quantitative estimates can be established that involve the the notion of a “lattice density” function, as we explain it in this paper, and shed some light on the problem of the curse of dimensionality. A priori and quantitative error bounds are obtained at any order of accuracy at the expense of possibly increasing the regularity of the functions under consideration. Importantly for the applications, the error function is controlled quantitatively in a given functional framework.

Evaluation of the discrepancy function. We focus attention on a selected list of kernels which we construct by a nonlinear transformation of four translation-invariant kernels, that is, of the form $K(x, y) = \varphi(x - y)$. We choose kernels that are commonly used in the applications, namely the exponential¹, multiquadric, Gaussian, and truncated kernels. Their Fourier transform $\hat{\varphi}$ defined on \mathbb{R}^D is known explicitly and is listed in Table 1.1.

¹which is sometimes also referred to as the Matérn kernel

	exponential (E)	multiquadric (M)	Gaussian (G)	truncated (T)
χ	$\exp(- x _1)$	$(1 + x ^2)^{-(D+1)/2}$	$\exp(- x ^2/2)$	$\sup(1 - x , 0)^D$
$\hat{\chi}$	$(1 + \xi ^2)^{-(D+1)/2}$	$\exp(- \xi)$	$\exp(- \xi ^2/2)$	(see [23])

Table 1.1: Four kernels and their Fourier transforms on \mathbb{R}^D

These kernels are defined in \mathbb{R}^D and we proceed by “localizing” them to a bounded domain Ω , taken to be the unit cube $[0, 1]^D$ for simplicity in the presentation. We propose two methods for such a localization of a kernel K defined on $\mathbb{R}^D \times \mathbb{R}^D$:

- The *periodic version* K^{per} of K defined from a discrete Fourier transform.
- The *transported version* K^{tran} of K defined via a nonlinear transport map.

This provides us with eight kernels (listed in Table 1.2) and our main purpose is to investigate the discrepancy function associated with each of them.

In principle, we could use numerically any one of the three expressions of the error function which we derive below and attempt to minimize it over the set of Y . In most cases, this requires a computation which, in general, cannot be done explicitly and a numerical integration of this function would be very costly, especially in large dimensions. We discuss this below. In particular, due to the (non-convex) form of the kernel, minimizing the error function in the physical space is computationally challenging. By introducing a periodic version based on the discrete Fourier transform, we arrive at an expression that is computationally tractable. We are able to make comparisons between these kernels and investigate the rate of convergence while comparing with the case when the points are randomly chosen. Our numerical results confirm and support our theoretical discussion.

Applications and perspectives. The material in this paper should be useful for analyzing mesh-free methods for computing solutions to partial differential equations and deriving quantitative bounds for algorithms used in pattern recognition and artificial intelligence. The estimate discussed here provides us with a key building block in order to establish an error analysis of the transported mesh-free method presented in the companion paper [11]. The method therein can be regarded as a generalization of the Lagrangian mesh-free method use in computational fluid dynamics, but also allows to include Navier-Stokes-type diffusive terms.

Most of the literature on error integration estimates is focused on functions with Sobolev regularity while we are interested here in functions with regularity adapted to specific applications. For instance, the standard choice of radially-symmetric kernels leads to functional spaces that are variants of Sobolev spaces and, in particular, are invariant by translations. Allowing more general kernels allows one to describe local (direction-dependent) properties of functions. For instance, a kernel we discuss below is adapted to measure the regularity of functions of the form $\varphi = \sum_{0 < n_1 < \dots < n_k \leq D} \varphi_{n_1, \dots, n_k}(x_{n_1}, \dots, x_{n_k})$, relevant in mathematical finance. The strategy in [9, 11] is now applied in industrial applications [16] and its accuracy can be explained in the light of the present study. This is relevant when considering the valuations of complex financial products (including the so-called American exercising) written on a large number of underlyings, and aiming at computing rapidly complex risk measures; see [9].

Outline of this paper. In Section 2, we present some basic material on reproducing kernel spaces. In Section 3, we discuss our methodology for constructing the two classes of kernels of main interest. In Section 4, the kernels studied in the present paper are presented and some their properties discussed. In Section 5, we derive several expressions of the discrepancy function, depending whether physical, spectral, or Fourier variables are used and, next, in Section 6 we derive estimates on the discrepancy error. In Section 7, we present and discuss our numerical results for each of the kernels of interest.

	exponential (E)	multiquadric (M)	Gaussian (G)	truncated (T)
Tra	$\exp(- \operatorname{erf}^{-1}(x) _1)$	$(1 + \operatorname{erf}^{-1}(x) ^2)^{-(D+1)/2}$	$\exp(- \operatorname{erf}^{-1}(x) ^2/2)$	$\sup(1 - \operatorname{erf}^{-1}(x) , 0)^D$
Per	$\sum \exp(- x + \alpha _1)$	$\sum (1 + x + \alpha ^2)^{-(D+1)/2}$	$\sum \exp(- x + \alpha ^2/2)$	$\sum \sup(1 - x + \alpha , 0)^D$

Table 1.2: Transported (Tra) and periodic (Per) kernels on $[-1, 1]^D$ (sum over $\alpha \in \mathbb{Z}^D$)

2 Functional framework based on a reproducing kernel

2.1 Discrete setup

The class of admissible kernels. Since we are primarily interested in kernels defined on a bounded set, in the present section we restrict attention to this class —although we will allow ourselves to manipulate kernels defined on the whole Euclidian space and treated as “seed data” in order to generate the kernels of actual interest. A reproducing kernel provides a convenient way to generate a broad class of Hilbert spaces (or, more generally, Banach space); cf. [4, 22]. A bounded and continuous function $K : \Omega \times \Omega \rightarrow \mathbb{R}$ on a bounded open set $\Omega \subset \mathbb{R}^D$ is called an *admissible kernel* if it satisfies (1) the *symmetry property*: $K(x, y) = K(y, x)$ for all $x, y \in \Omega$, and (2) the *positivity property*: for any collection of N distinct points $Y = (y^1, \dots, y^N)$ in Ω , the symmetric matrix $K(Y, Y) = (K(y^m, y^n))_{1 \leq m, n \leq N}$ is positive definite in the sense that $a^T K(Y, Y) a > 0$ for all $a \in \mathbb{R}^N \setminus \{0\}$. It is said to be *uniformly positive* if there exists a uniform constant $c > 0$ such that for any collection of distinct points Y one has $a^T K(Y, Y) a \geq c |a|^2$ for all $a \in \mathbb{R}^N$.

Clearly, any admissible kernel also satisfies

$$K(x, x) \geq 0, \quad K(x, y)^2 \leq K(x, x) K(y, y), \quad x, y \in \Omega. \quad (2.1)$$

This implies that $2K(x, y) \leq K(x, x) + K(y, y)$ and, therefore, the non-negative function

$$\mathbf{D}(x, y) := K(x, x) + K(y, y) - 2K(x, y) \geq 0, \quad x, y \in \Omega, \quad (2.2)$$

can be interpreted as a “pseudo-distance” in view of the properties $\mathbf{D}(x, x) = 0$ and $\mathbf{D}(x, y) = \mathbf{D}(y, x)$. (The triangle inequality need not hold.) Many examples of admissible kernels will be presented in the next two sections.

Finite dimensional framework. Given any finite collection of points $Y = (y^1, \dots, y^N)$ chosen in Ω , we introduce the (finite dimensional) vector space $\mathcal{H}_K^Y(\Omega)$ consisting of all linear combinations of the *basis functions* $x \mapsto K(x, y^n)$. In other words, we set

$$\mathcal{H}_K^Y(\Omega) := \left\{ \sum_{1 \leq m \leq N} a_m K(\cdot, y^m) / a = (a_1, \dots, a_N) \in \mathbb{R}^N \right\}. \quad (2.3)$$

Since K is continuous, $\mathcal{H}_K^Y(\Omega) \subset \mathcal{C}(\Omega)$ embeds into the space $\mathcal{C}(\Omega)$ of all continuous functions on Ω . To any two functions $\varphi = \sum_{1 \leq m \leq N} a_m K(\cdot, y^m)$ and $\psi = \sum_{1 \leq n \leq N} b_n K(\cdot, y^n)$, we associate the bilinear expression

$$\langle \varphi, \psi \rangle_{\mathcal{H}_K^Y(\Omega)} := a^T K(Y, Y) b = \sum_{1 \leq m \leq N} \sum_{1 \leq n \leq N} a_m b_n K(y^m, y^n) \quad (2.4)$$

(with $a = (a_m)$, etc.), which endows the space $\mathcal{H}_K^Y(\Omega)$ with a Hilbertian structure with norm $\|\varphi\|_{\mathcal{H}_K^Y(\Omega)}^2 := a^T K(Y, Y) a$. Now, the so-called *reproducing kernel property* (immediate from (2.4))

$$\langle K(\cdot, y^m), K(\cdot, y^n) \rangle_{\mathcal{H}_K^Y(\Omega)} = K(y^m, y^n), \quad (2.5)$$

allows one to relate the coefficients of the decomposition of a function $\varphi = \sum_{1 \leq m \leq N} a_m K(\cdot, y^m)$ to its scalar product with the basis functions, namely

$$\langle \varphi, K(\cdot, y^n) \rangle_{\mathcal{H}_K^Y(\Omega)} = \sum_{1 \leq m \leq N} a_m \langle K(\cdot, y^m), K(\cdot, y^n) \rangle_{\mathcal{H}_K^Y(\Omega)} = \sum_{1 \leq m \leq N} a_m K(y^m, y^n) = a^T K(Y, y^n). \quad (2.6)$$

Discrete spectral decomposition. Since $K(Y, Y)$ is a symmetric and positive definite matrix, it admits real and positive eigenvalues, denoted by $\lambda_Y^n > 0$, together with a basis of right-eigenvectors $\zeta_Y^n \in \mathbb{R}^D$ (with $n = 1, 2, \dots, N$) chosen to be unitary, and satisfying

$$K(Y, Y)\zeta_Y^n = \lambda_Y^n \zeta_Y^n, \quad n = 1, 2, \dots, N. \quad (2.7)$$

This decomposition is useful in order to define, for any $s \geq 0$ and $p \geq 1$, the finite dimensional Banach space $\mathcal{H}_K^{Y, s, p}(\Omega)$ of all functions of the form $\sum_{1 \leq m \leq N} a_m K(\cdot, y^m)$ with finite norm

$$\|\varphi\|_{\mathcal{H}_K^{Y, s, p}(\Omega)}^p := \sum_{1 \leq n \leq N} (\lambda_Y^n)^{-sp} \langle \varphi, \zeta_Y^n \rangle_{\mathcal{H}_K^Y(\Omega)}^p. \quad (2.8)$$

Projection operator. Consider a function $f \in \mathcal{C}(\Omega)$ and introduce the vector $f(Y) = (f(y^1), \dots, f(y^N))$ consisting of the values of this function at the given points. We define its projection $P_Y(f)$ into the discrete space $\mathcal{H}_K^Y(\Omega)$ by setting

$$P_Y(f) := a^T K(\cdot, Y), \quad a := K(Y, Y)^{-1} f(Y). \quad (2.9a)$$

Clearly, this defines a projection since $P_Y \circ P_Y(f) = P_Y(f)$ and, in fact, $P_Y(\varphi) = \varphi$ for any function $\varphi = a^T K(\cdot, Y)$ belonging to the space $\mathcal{H}_K^Y(\Omega)$ (where we used that K is symmetric). Moreover, the norm of this projection reads

$$\|P_Y(f)\|_{\mathcal{H}_K^Y(\Omega)}^2 = f(Y)^T K(Y, Y)^{-1} f(Y). \quad (2.9b)$$

The partition of unity. A basis is naturally associated with the discrete space $\mathcal{H}_K^Y(\Omega)$, that is, N functions $\theta_Y^n : \Omega \rightarrow \mathbb{R}$ taking the values 0 or 1 at the points of the set Y . Precisely, writing $\theta_Y := (\theta_Y^1, \dots, \theta_Y^N)$, we define

$$\theta_Y := K(Y, Y)^{-1} K(Y, \cdot). \quad (2.10a)$$

It follows that $(\theta_Y(y^m))_{1 \leq n, m \leq N} = K(Y, Y)^{-1} K(Y, Y) = \text{Id}$ (the identity matrix), and using the Kronecker symbol, we have $\theta_Y^n(y^m) = \delta_{nm}$, while the scalar product of any two basis functions is

$$\langle \theta_Y^m, \theta_Y^n \rangle_{\mathcal{H}_K^Y(\Omega)} = K^{-1}(y^m, y^n). \quad (2.10b)$$

This partition of unity is useful in expressing the projection of a function $f \in \mathcal{C}(\Omega)$, namely $P_Y(f) = \sum_{1 \leq n \leq N} f(y^n) \theta_Y^n$.

2.2 Continuous setup

Functional spaces. Given an admissible kernel $K : \Omega \times \Omega \rightarrow \mathbb{R}$, we now introduce the infinite dimensional space $\tilde{\mathcal{H}}_K(\Omega)$ consisting of all finite linear combinations of the functions $K(x, \cdot)$ parametrized by $x \in \Omega$, that is, $\tilde{\mathcal{H}}_K(\Omega) := \text{Span}\{K(\cdot, x) / x \in \Omega\}$, which we endow with the scalar product and norm defined in the finite dimensional setup; see (2.1) (where now Y and N are no longer fixed). By construction, the reproducing kernel property (2.5) also holds in $\tilde{\mathcal{H}}_K(\Omega)$, i.e.

$$\langle K(\cdot, x), K(\cdot, y) \rangle_{\tilde{\mathcal{H}}_K(\Omega)} = K(x, y), \quad x, y \in \Omega. \quad (2.11a)$$

From the Cauchy-Schwarz inequality, it follows that for any $\varphi \in \tilde{\mathcal{H}}_K(\Omega)$ and $x \in \Omega$

$$|\varphi(x)| = |\langle K(\cdot, x), \varphi \rangle_{\tilde{\mathcal{H}}_K(\Omega)}| \leq \|K(\cdot, x)\|_{\tilde{\mathcal{H}}_K(\Omega)} \|\varphi\|_{\tilde{\mathcal{H}}_K(\Omega)} = \sqrt{K(x, x)} \|\varphi\|_{\tilde{\mathcal{H}}_K(\Omega)}. \quad (2.11b)$$

Since the kernel is continuous and bounded, the ‘‘point evaluation’’ $\varphi \mapsto \varphi(x)$ is thus a linear and bounded functional on $\tilde{\mathcal{H}}_K(\Omega)$ (for any $x \in \Omega$).

We have defined a pre-Hilbert space, that is, a vector space endowed with a scalar product and, in order to obtain a complete metric space, the completion of the pre-Hilbert space $\tilde{\mathcal{H}}_K(\Omega)$ is considered by taking all linear combinations based on countably many points $Y = (y^1, y^2, \dots)$ in Ω . The corresponding space is denoted by $\mathcal{H}_K(\Omega)$ and is the *reproducing Hilbert space* generated from the kernel K .

Clearly, both properties (2.11) remain true in $\mathcal{H}_K(\Omega)$. Observe also that

$$\begin{aligned} |\varphi(x) - \varphi(y)| &= |\langle K(\cdot, x) - K(\cdot, y), \varphi \rangle_{\mathcal{H}_K(\Omega)}| \\ &\lesssim \|K(\cdot, x) - K(\cdot, y)\|_{\mathcal{H}_K(\Omega)} \|\varphi\|_{\mathcal{H}_K(\Omega)} = \mathbf{D}(x, y) \|\varphi\|_{\mathcal{H}_K(\Omega)}, \quad x, y \in \Omega. \end{aligned} \quad (2.12)$$

Since K and thus \mathbf{D} are continuous in Ω , we have the embedding $\mathcal{H}_K(\Omega) \subset C(\Omega)$, that is, all of the functions are continuous, at least.

Mercer decomposition. We now consider the linear operator $T_K : L^2(\Omega) \rightarrow L^2(\Omega)$ defined by $T_K(a) := \int_{\Omega} K(\cdot, x)a(x) dx$ (for $a \in L^2(\Omega)$) on the Hilbert space $L^2(\Omega)$ endowed with its standard inner product. We have

$$\|T_K(a)\|_{L^2(\Omega)} \leq |\Omega|^{1/2} \|K\|_{L^2(\Omega \times \Omega)} \|a\|_{L^2(\Omega)},$$

so this operator is continuous and self-adjoint. It is easily checked to be compact: if a sequence $a^p \rightharpoonup a$ weakly in the L^2 norm then $T_K(a^p) \rightarrow T_K(a)$ strongly in the L^2 norm. The classical spectral decomposition applies and the operator T_K admits an (at most countable) non-increasing sequence of eigenvalues $\lambda_i > 0$ and a corresponding set of eigenfunctions ζ_i such that $T_K(\zeta_i) = \lambda_i \zeta_i$ and the family $\{\zeta_1, \zeta_2, \dots\}$ forms an orthonormal basis of $L^2(\Omega)$ for the L^2 inner product. Furthermore, since the kernel is continuous and bounded, the eigenfunctions ζ_i are continuous, at least.

We then introduce the kernel

$$L(x, y) = \sum_{j=1,2,\dots} \lambda_j \zeta_j(x) \zeta_j(y), \quad x, y \in \Omega, \quad (2.13)$$

in which the sum converges in the L^2 sense. This kernel is admissible since, for any collection of points $Y = (y^1, \dots, y^N)$ in \mathbb{R}^N , the bilinear form $L^Y(x, x') = \sum_{1 \leq n, m \leq N} L(y^n, y^m) x_n x'_m$ (with $x, x' \in \mathbb{R}^N$), satisfies

$$L^Y(x, x) = \sum_{j=1,2,\dots} \lambda_j \sum_{1 \leq n, m \leq N} \zeta_j(y^n) \zeta_j(y^m) x_n x_m = \sum_{j=1,2,\dots} \lambda_j \left| \sum_{1 \leq n \leq N} \zeta_j(y^n) x_n \right|^2 \geq 0.$$

In fact, one can check that L coincides with the given kernel K and (2.13) represents its spectral decomposition. The family of functions $\{\lambda_i^{1/2} \zeta_i\}_{i \geq 1}$ is an orthonormal basis of the space $\mathcal{H}_K(\Omega)$, as follows from the defining relation $T_K \zeta_i = \lambda_i \zeta_i$, namely²

$$\langle \lambda_i^{1/2} \zeta_i, \lambda_j^{1/2} \zeta_j \rangle_{\mathcal{H}_K(\Omega)} = \lambda_i^{1/2} \lambda_j^{1/2} \frac{1}{\lambda_i} \langle T_K \zeta_i, \zeta_j \rangle_{\mathcal{H}_K(\Omega)} = \frac{\lambda_j^{1/2}}{\lambda_i^{1/2}} \langle \zeta_i, \zeta_j \rangle_{L^2(\Omega)} = \delta_{ij}. \quad (2.14)$$

This series representation of the elements in $\mathcal{H}_K(\Omega)$ is referred to as the *Mercer representation* (which may be non-unique). In short, Mercer theorem states that any admissible kernel $K : \Omega \times \Omega \rightarrow \mathbb{R}$ can be viewed as a positive, self-adjoint and compact operator on $L^2(\Omega)$.

Banach spaces. Based on the Mercer representation, we can define the L^p -based spaces at any order of differentiability. Namely, for any $s \geq 0$ and $p \in [1, +\infty)$ we consider the Banach space

$$\mathcal{H}_K^{s,p}(\Omega) := \left\{ \varphi \in L^p(\Omega) / \sum_{i=1,2,\dots} \lambda_i^{-ps/2} |\langle \varphi, \zeta_i \rangle_{L^2(\Omega)}|^p < +\infty \right\} \quad (2.15)$$

endowed with the norm $(\|\varphi\|_{\mathcal{H}_K^{s,p}(\Omega)})^p := \sum_{i=1,2,\dots} \lambda_i^{-ps/2} |\langle \varphi, \zeta_i \rangle_{L^2(\Omega)}|^p$. When p is chosen to be 2, the space $\mathcal{H}_K^s(\Omega) := \mathcal{H}_K^{s,2}(\Omega)$ is a Hilbert space endowed with the inner product

$$\langle f, g \rangle_{\mathcal{H}_K^s(\Omega)} = \sum_{i=1,2,\dots} (\lambda_i)^{-s} \langle f, \zeta_i \rangle_{L^2(\Omega)} \langle g, \zeta_i \rangle_{L^2(\Omega)}. \quad (2.16)$$

In particular, $\mathcal{H}_K(\Omega) = \mathcal{H}_K^1(\Omega) = \mathcal{H}_K^{1,2}(\Omega)$ and we recover the Hilbert space defined earlier.

3 Methodology for defining classes of kernels

3.1 Translation-invariant and radially-symmetric kernels on \mathbb{R}^D

Translation invariant kernels. Our first task now is to provide some preliminary material and a classification involving the notions of translation-invariant kernel, periodic kernel, tensorial kernel, and radially-symmetric kernel. In the present section we allow ourselves to introduce kernels defined on the whole of \mathbb{R}^D ,

² δ_{ij} being the Kronecker symbol

although Section 2 was restricted to kernels defined on a bounded set; namely, we will use kernel defined on \mathbb{R}^D only for defining the kernels of interest defined on a bounded set.

We begin with the class of the form $K(x, y) = \chi(x - y)$ (the examples in Table 1.1 being of this type), which we refer to as *translation-invariant kernels*

$$K(x, y) = \chi(x - y), \quad x, y \in \mathbb{R}^D. \quad (3.1a)$$

This family is parametrized by a *generating function* $\chi : \mathbb{R}^D \rightarrow \mathbb{R}$, which (after normalization) must satisfy (as we check below)

$$\begin{aligned} \chi(0) &= 1, & \chi(-x) &= \chi(x), & x &\in \mathbb{R}^D, \\ \widehat{\chi}(\xi) &\geq 0 & & & \xi &\in \mathbb{R}^D. \end{aligned} \quad (3.1b)$$

Positivity property. Under the assumption (3.1b), let us consider a collection $Y = (y^1, \dots, y^N)$ in \mathbb{R}^D and the associated bilinear form $K(Y, Y)(\xi, \xi) = \sum_{1 \leq n, m \leq N} K(y^n, y^m) \xi_n \xi_m$ for $\xi \in \mathbb{R}^N$. We obtain the positivity property

$$\sum_{1 \leq n, m \leq N} \langle \widehat{\chi}, e^{-i \langle y^n - y^m, \cdot \rangle} \rangle_{\mathcal{D}', \mathcal{D}} \xi_n \xi_m = \langle \widehat{\chi}, \left| \sum_n \xi_n e^{-i \langle y^n, \cdot \rangle} \right|^2 \rangle_{\mathcal{D}', \mathcal{D}} \geq 0,$$

in which $\langle \cdot, \cdot \rangle_{\mathcal{D}', \mathcal{D}}$ is the duality bracket for distributions. We thus find

$$\begin{aligned} \iint_{\mathbb{R}^D \times \mathbb{R}^D} K(x, y) \varphi(x) \varphi(y) dx dy &= \iint_{\mathbb{R}^D \times \mathbb{R}^D} \chi(x - y) \varphi(x) \varphi(y) dx dy \\ &= \int_{\mathbb{R}^D} (\chi \star \varphi)(x) \varphi(x) dx = \int_{\mathbb{R}^D} \widehat{(\chi \star \varphi)}(\xi) \widehat{\varphi}(\xi) d\xi = \int_{\mathbb{R}^D} |\widehat{\varphi}|^2 \widehat{\chi} d\xi \geq 0, \end{aligned}$$

or equivalently $\iint_{\mathbb{R}^D \times \mathbb{R}^D} K(x, y) \varphi(x) \varphi(y) dx dy \geq 0$.

Radially-symmetric kernels. Among translation-invariant kernels, the class of *radially-symmetric kernels* is an important subclass and corresponds to the case where the function χ that only depends on the modulus of its argument only, that is,

$$K(x, y) = \chi(|x - y|), \quad x, y \in \mathbb{R}^D, \quad (3.2)$$

where the generating function $\chi : \mathbb{R} \rightarrow \mathbb{R}$ is now regarded as a function of a real variable.

3.2 Localization via scaling: the transported kernels

The translation-invariant fail to be sufficiently localized (say in the sense that K fails to be $L^1(\mathbb{R}^D \times \mathbb{R}^D)$), and we now explain how a transportation map can be applied in order to “localize” a kernel to a bounded set. Many proofs will be postponed to the Appendix, including the one of the following result.

Proposition 3.1. *Let $K^{seed} : \mathbb{R}^D \times \mathbb{R}^D \rightarrow \mathbb{R}$ be a bounded admissible kernel and μ be a sufficiently regular, probability measure such that $\text{supp}(\mu)$ is a convex set (with, therefore, $\mu \geq 0$ and $\mu(\mathbb{R}^D) = 1$). Then the kernel*

$$K^{loc}(x, x') := K^{seed}(x, x') \mu(x) \mu(x'), \quad x, x' \in \mathbb{R}^D, \quad (3.3a)$$

is admissible and belongs to $L^1(\mathbb{R}^D \times \mathbb{R}^D)$. Moreover, let $\Omega \subset \mathbb{R}^D$ be any open and convex subset with normalized volume $|\Omega| = 1$, and consider a transportation map for the measure μ , that is, a one-to-one map $S : \Omega \rightarrow \text{supp}(\mu)$ such that $S = \nabla h$ for some convex function $h : \Omega \rightarrow \mathbb{R}$ and $S_{\#} \mu = dy$ (the Lebesgue measure). Then,

$$K^{tran}(y, y') := K^{seed}(S(y), S(y')), \quad y, y' \in \Omega \quad (3.3b)$$

*defines an admissible kernel, referred to as the **transported kernel** associated with K^{seed} and μ . Furthermore, for any $\varphi \in L^1_{\mu}(\mathbb{R}^D)$ (the set of functions that are integrable for the measure μ) and any choice of points $x^n = S(y^n)$ one has*

$$\int_{\mathbb{R}^D} \varphi(x) d\mu(x) - \frac{1}{N} \sum_{1 \leq n \leq N} \varphi(x^n) = \int_{\Omega} (\varphi \circ S)(y) dy - \frac{1}{N} \sum_{1 \leq n \leq N} (\varphi \circ S)(y^n). \quad (3.3c)$$

In the statement above, the transportation maps satisfies $\int_{\mathbb{R}^D} \varphi d\mu = \int_{\Omega} (\varphi \circ S) dy$ for any continuous function $\varphi \in L^1_{\mu}(\mathbb{R}^D)$. Provided S is sufficiently regular, one has $S_{\#}\mu = |\det \nabla S| \circ S^{-1} dy$, where $|\det \nabla S|$ is the Jacobian of S . The convexity of $\text{supp}(\mu)$ ensures that S is one-to-one and continuous from Ω onto the support set $\text{supp}(\mu)$. Furthermore, thanks to the localization argument above, deriving an error estimate associated with the left-hand side of (3.3c) reduces to (1.1), and the role of the measure μ is eliminated. Within the framework of kernel spaces, the relation between the μ -weighted norm on \mathbb{R}^D and the unweighted norm on the bounded set Ω is as follows (with obvious notation):

$$\|\varphi\|_{\mathcal{H}_{K^{\text{seed},\mu}^{s,p}}(\mathbb{R}^D)} = \|\varphi\|_{\mathcal{H}_{K^{\text{loc},s,p}}(\mathbb{R}^D)} = \|\varphi \circ S\|_{\mathcal{H}_{K^{\text{tran},s,p}}(\Omega)}.$$

3.3 Localization via periodization: the periodic kernels

A discrete lattice. Motivated by Cohn and Elkies's work [3] on the sphere packing problem, we propose to embed the support of a general kernel in a periodic lattice. By periodicity, we can always extend to \mathbb{R}^D the kernel defined to an elementary cell. Importantly, the terms arising in the spectral decomposition can be controlled almost explicitly, thanks to a Poisson decomposition formula associated with the lattice.

A family of D vectors $l_1, l_2, \dots, l_D \in \mathbb{R}^D$ being given, we consider their convex hull $\mathbf{C} \subset \mathbb{R}^D$ which serves as the fundamental cell of our discrete lattice and whose volume is denoted by $|\mathbf{C}|$. By suitably translating \mathbf{C} , we thus generate the periodic lattice $\mathbf{L} := \left\{ \sum_{1 \leq d \leq D} \alpha_d l_d / \alpha = (\alpha_1, \dots, \alpha_D) \in \mathbb{Z}^D \right\}$, and we denote its dual by $\mathbf{L}^* := \left\{ \alpha^* \in \mathbb{R}^D / \langle \alpha, \alpha^* \rangle \in \mathbb{Z} \text{ for all } \alpha \in \mathbf{L} \right\}$. We denote by l_1^*, \dots, l_D^* the vectors generating the elementary cell $\mathbf{C}^* \subset \mathbb{R}^D$ of the dual lattice. Next, we introduce the discrete Hilbert space $l^2(\mathbf{L}^*)$ consisting of all functions defined on the dual lattice, endowed with the inner product

$$\langle f, g \rangle_{l^2(\mathbf{L}^*)} := \frac{1}{|\mathbf{C}|} \sum_{\alpha^* \in \mathbf{L}^*} f(\alpha^*) g(\alpha^*), \quad f, g : \mathbf{L}^* \rightarrow \mathbb{R}. \quad (3.4)$$

Poisson formula. Consider the Fourier transform of a real-valued function $\varphi : \mathbb{R}^D \rightarrow \mathbb{R}$ defined on \mathbb{R}^D and let us restrict it to the dual lattice, that is, consider the discrete values

$$\widehat{\varphi}(\alpha^*) := \int_{\mathbb{R}^D} \varphi(x) e^{-2i\pi \langle x, \alpha^* \rangle} dx. \quad (3.5)$$

Then, the so-called Poisson formula reads

$$\sum_{\alpha \in \mathbf{L}} \varphi(\alpha + x) = \frac{1}{|\mathbf{C}|} \sum_{\alpha^* \in \mathbf{L}^*} e^{2i\pi \langle x, \alpha^* \rangle} \widehat{\varphi}(\alpha^*), \quad x \in \mathbb{R}^D. \quad (3.6)$$

Provided the function φ is supported on the cell \mathbf{C} , the sum in the left-hand side contains a *single term* and, therefore,

$$\varphi(x) = \frac{1}{|\mathbf{C}|} \sum_{\alpha^* \in \mathbf{L}^*} e^{2i\pi \langle x, \alpha^* \rangle} \widehat{\varphi}(\alpha^*), \quad x \in \mathbf{C}, \text{ provided } \text{supp}(\varphi) \subset \mathbf{C} \quad (3.7a)$$

or, with our notation,

$$\varphi(x) = \langle e^{2i\pi \langle x, \cdot \rangle}, \widehat{\varphi} \rangle_{l^2(\mathbf{L}^*)}, \quad x \in \mathbf{C}, \text{ provided } \text{supp}(\varphi) \subset \mathbf{C}. \quad (3.7b)$$

Hence, a function defined on the cell \mathbf{C} can be recovered (via an discrete inverse Fourier transform) from the values of its Fourier transform on the dual lattice \mathbf{L}^* .

More generally, let us derive an identity that will be useful to us later on. Consider now a collection (y^1, \dots, y^N) of points in \mathbb{R}^D and, in view of (3.7), let us write

$$\sum_{1 \leq n, m \leq N} \sum_{\alpha \in \mathbf{L}} \varphi(\alpha + y^n - y^m) = \frac{1}{|\mathbf{C}|} \sum_{1 \leq n, m \leq N} \sum_{\alpha^* \in \mathbf{L}^*} e^{2i\pi \langle y^n - y^m, \alpha^* \rangle} \widehat{\varphi}(\alpha^*) = \frac{1}{|\mathbf{C}|} \sum_{\alpha^* \in \mathbf{L}^*} \left| \sum_{1 \leq n \leq N} e^{2i\pi \langle y^n, \alpha^* \rangle} \right|^2 \widehat{\varphi}(\alpha^*).$$

We can arrange that, in the left-hand side, the sum over $\alpha \in \mathbf{L}$ reduces to a single term obtain when $\alpha = 0$. We reach the following conclusion.

Lemma 3.2. For any function φ supported on the elementary cell \mathbf{C} of a lattice \mathbf{L} , that is, $\text{supp}(\varphi) \subset \mathbf{C}$ and for any finite collection (y^1, \dots, y^N) satisfying the “localization property” $y^n - y^m \in \mathbf{C}$ for all $n, m = 1, 2, \dots, N$, the following identity holds:

$$\sum_{1 \leq n, m \leq N} \varphi(y^n - y^m) = \frac{1}{|\mathbf{C}|} \sum_{\alpha^* \in \mathbf{L}^*} \left| \sum_{1 \leq n \leq N} e^{2i\pi \langle y^n, \alpha^* \rangle} \right|^2 \widehat{\varphi}(\alpha^*). \quad (3.8)$$

Periodic kernels. The interest of the following class of kernels lies in the fact that their spectral decomposition can be determined almost explicitly, in terms of exponential functions defined on the dual lattice. Namely, Lemma 3.2 allows us to pass from the continuous physical variables on \mathbf{C} to the discrete Fourier variables on \mathbf{L}^* . The role of the function ρ introduced below is going to be played by (the restriction to the lattice of) the Fourier transform of an arbitrary kernel. Indeed, our definition below provides a way to transform a “seed” kernel K^{seed} defined on \mathbb{R}^D to a periodic kernel K^{per} defined on the lattice cell \mathbf{C} . (See the proof in the appendix.)

Proposition 3.3 (Periodic kernels associated with a generating function). Consider a discrete lattice $\mathbf{L} \subset \mathbb{R}^D$ generated from an elementary cell \mathbf{C} , and let $\rho : \mathbf{L}^* \rightarrow (0, +\infty)$ be a positive, integrable, and even function, that is,

$$\rho^{1/2} \in \ell^2(\mathbf{L}^*), \quad \rho(-\alpha_*) = \rho(\alpha_*) \geq 0 \quad (\text{with } \alpha_* \in \mathbf{L}^*). \quad (3.9a)$$

Then, the discrete Fourier transform of ρ extended to the whole of \mathbb{R}^D , that is,

$$K^{\text{per}}(x, y) := \langle e^{2i\pi \langle x-y, \cdot \rangle}, \rho \rangle_{\ell^2(\mathbf{L}^*)}, \quad x, y \in \mathbb{R}^D \quad (3.9b)$$

defines an admissible kernel on \mathbb{R}^D which is periodic with period \mathbf{C} and its associated Hilbert space is

$$\mathcal{H}_{K^{\text{per}}}(\mathbf{C}) = \left\{ \varphi \in C(\mathbb{R}^D) / \mathbf{C}\text{-periodic} / \widehat{\varphi} \rho^{-1/2} \in \ell^2(\mathbf{L}^*) \right\} \quad (3.9c)$$

endowed with the norm $\langle f, g \rangle_{\mathcal{H}_{K^{\text{per}}}(\mathbf{C})} = \langle \widehat{f} \rho^{-1/2}, \widehat{g} \rho^{-1/2} \rangle_{\ell^2(\mathbf{L}^*)}$.

More generally, provided $\rho^s \in \ell^p(\mathbf{L}^*)$ for some $p \in [1, +\infty)$ and $s \geq 0$, we can also introduce the Banach space

$$\mathcal{H}_{K^{\text{per}}}^{s,p}(\mathbf{C}) = \left\{ \varphi \in C(\mathbb{R}^D) / \mathbf{C}\text{-periodic} / \widehat{\varphi} \rho^{-s/p} \in \ell^{p'}(\mathbf{L}^*) \right\}. \quad (3.10)$$

3.4 Further generating techniques

The tensor technique. A broad class of examples on \mathbb{R}^D can be obtained by tensor decomposition from an admissible kernel in one dimension, say $K^{\text{seed}} : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$, namely

$$K(x, y) = \prod_{1 \leq d \leq D} K^{\text{seed}}(x_d, y_d), \quad x, y \in \mathbb{R}^D. \quad (3.11a)$$

This applies, particularly, to a translation-invariant kernel, in which case we choose any even function $\chi^{\text{seed}} : \mathbb{R} \rightarrow \mathbb{R}$ and set

$$K(x, y) = \prod_{1 \leq d \leq D} \chi^{\text{seed}}(x_d - y_d), \quad x, y \in \mathbb{R}^D. \quad (3.11b)$$

An example: the tensorial truncated kernel. The kernel $K_T(x, y) = (1 - |x - y|)^+$ has the form above and can also be expressed as a convolution, namely

$$K_T(x, y) = \chi(x - y) = (\lambda * \lambda)(x - y), \quad \lambda(x) = 1_{[-1/2, 1/2]}(x), \quad x \in \mathbb{R}. \quad (3.12a)$$

Here, we have $\widehat{\chi}(\xi) = \widehat{\lambda}^2(\xi)$ and $\widehat{\lambda}(\xi) = \frac{\sin(2\pi\xi)}{2\pi\xi}$. More generally, in dimension D we consider

$$K(x, y) = \prod_{1 \leq d \leq D} K^{\text{seed}}(x_d, y_d) = \prod_{1 \leq d \leq D} (1 - |x_d - y_d|)^+, \quad (3.12b)$$

written also as a convolution $K(x, y) = (\lambda * \lambda)(x - y)$ with $\lambda(x) = 1_{[0, 1]^D}(x)$. Again, we have $K(x, y) = \chi(x - y)$ with $\widehat{\chi}(\xi) = \widehat{\lambda}^2(\xi)$, and now $\widehat{\lambda}(\xi) = \prod_{1 \leq d \leq D} \frac{\sin(2\pi\xi_d)}{2\pi\xi_d}$.

The normalization technique. The transformations below can also serve as building blocks in order to adapt existing examples to a particular application. If K is an admissible kernel on \mathbb{R}^D we can consider

$$K^{\text{norm}}(x, y) = \frac{K(x, y)}{K(x, x)^{1/2}K(y, y)^{1/2}}, \quad x, y \in \mathbb{R}^D. \quad (3.13)$$

Clearly, we have $K^{\text{norm}}(x, x) = 1$ and the required admissibility conditions are easily checked. The coefficients a_n and a_n^{norm} of the corresponding decomposition (2.3) of a function in $\mathcal{K}_K(\mathbb{R}^D)$ and $\mathcal{K}_{K^{\text{norm}}}(\mathbb{R}^D)$, respectively, are related by $a_n^{\text{norm}} = a_n K(y^n, y^n)^{1/2}$, so that the two spaces are quite similar from the application standpoint. We thus regard (3.13) as a normalization procedure.

Taking sums and products. If K_1, K_2 are admissible kernels in \mathbb{R}^D , then it is also easily checked that, $a, b > 0$ being some given constants,

$$K_3(x, y) = aK_1(x, y) + bK_2(x, y), \quad K_4(x, y) = K_1(x, y)K_2(x, y), \quad (3.14)$$

With the notation used in Section 2 the matrices $K_3(Y, Y)$ and $K_4(Y, Y)$ are symmetric positive definite, as follows easily from a standard linear algebra argument.

Zonal kernels. If a function ϕ is such that $(x, x') \mapsto \phi(xx')$ is a one-dimensional kernel, then the following formula

$$K_{\text{zonal}}(x, y) = \phi(\langle x, y \rangle), \quad x, y \in \mathbb{R}^D \quad (3.15)$$

(where $\langle x, y \rangle$ stands for the Euclidian inner product) defines an admissible kernel. This class is often used by the artificial intelligence community.

Convolution kernels. Another class of translation-invariant kernels can be generated by choosing a function

$$\lambda \in L^2(\mathbb{R}^D), \quad \int_{\mathbb{R}^D} \lambda(x)\lambda(-x) dx \neq 0, \quad \int_{\mathbb{R}^D} (\widehat{\lambda})^2 d\xi = 1 \quad (3.16a)$$

and then defining our generating function χ by the convolution formula

$$K(x, y) := \chi(x - y) = (\lambda \star \lambda)(x - y), \quad (\lambda \star \lambda)(y) = \frac{\int_{\mathbb{R}^D} \lambda(x)\lambda(y - x) dx}{\int_{\mathbb{R}^D} \lambda(x)\lambda(-x) dx}, \quad y \in \mathbb{R}^D. \quad (3.16b)$$

Indeed, it is clear that $\chi(0) = 1$ and $\widehat{\chi} = (\widehat{\lambda})^2 \geq 0$, so that the Fourier transform of χ is a probability measure. This class of kernels is used in machine learning, for instance in combination of multi-layer neural networks.

4 Designing kernels on a bounded domain

4.1 Standard kernels taken as seed data

We focus our attention to the standard examples listed in Table 1.1, that are radially-symmetric, translation-invariant kernels $K(x, y) = \chi(x - y)$.

Exponential kernel K_E . The choice $\widehat{\chi}_E(\xi) = (1 + |\xi|^2)^{-m}$ (with $m > D/2$) leads to the standard Sobolev space $\mathcal{H}_{K_E}(\mathbb{R}^D) = W^{m,2}(\mathbb{R}^D)$, and is a standard choice in the numerical analysis literature.

Multiquadric kernel K_M . The choice $\widehat{\chi}_M(\xi) = e^{-|\xi|}$ is only Lipschitz continuous at the origin and is relevant for representing sufficiently smooth functions with polynomial decay, hence provides (slightly) more information than the Gaussian one (below).

Gaussian kernel K_G . The choice $\widehat{\chi}_G(\xi) = e^{-|\xi|^2}$ provides an exponential decay in both the Fourier and the physical spaces. Both functions $\widehat{\chi}_G \geq 0$ and $\chi_G \geq 0$ are globally positive. The Gaussian kernel is adapted to the description of smooth and fast decaying functions which have “almost” compact support in physical and Fourier variables. Hence, the function space $\mathcal{H}_{K_G}(\mathbb{R}^D)$ is “small” and, in term, provides limited “information” on the functions.

Truncated kernel K_T . A more interesting and also quite standard choice is obtained by truncation in the physical space, namely, $\chi_T(x) = (1 - |x|_1)_+^l$ (with $l \geq D/2$) where the notation $a_+ := \sup(a, 0)$ stands for the positive part. This kernel is only Lipschitz continuous in the physical space.

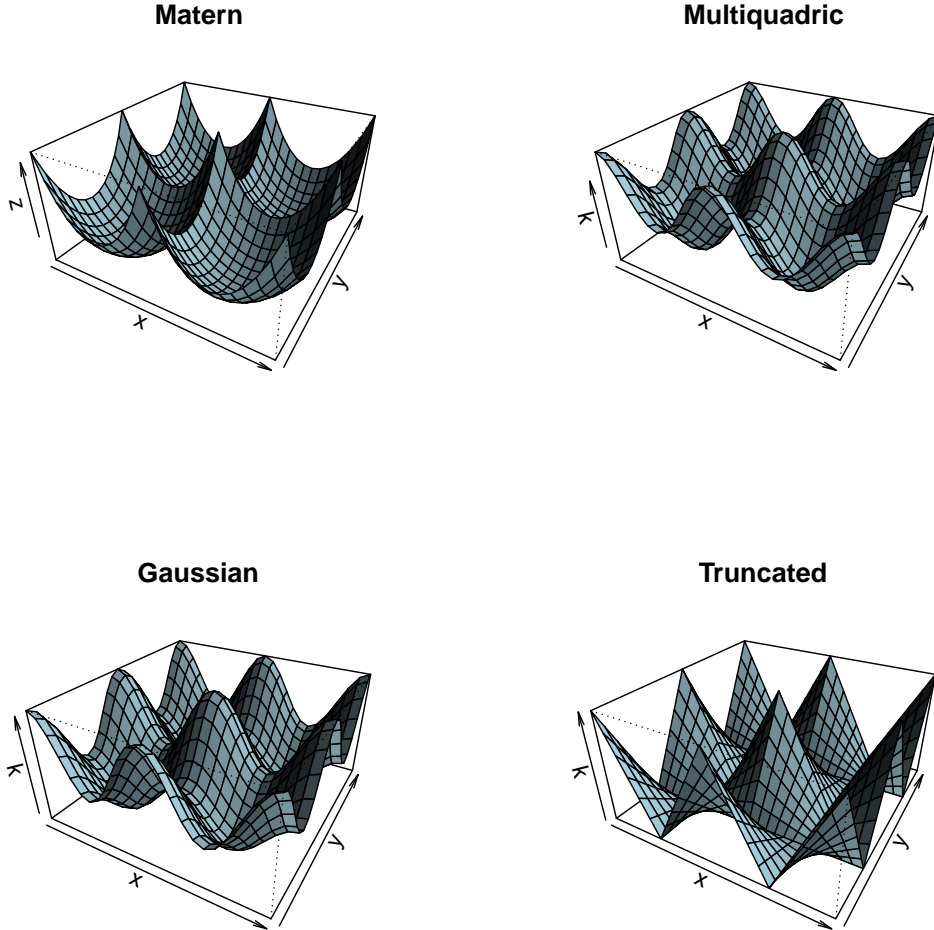


Figure 4.1: Plots of periodic kernels in two dimensions: exponential, multiquadric, Gaussian, and truncated, respectively.

4.2 Periodic kernels of interest

Objective. We now present, in their rescaled form, the four periodic kernels already listed in Table 1.2 (and associated with each of the examples in Table 1.1). Periodic kernels are translation-invariant, i.e. $K^{\text{per}}(x, y) = \chi^{\text{per}}(x - y)$ and we plot the corresponding functions χ^{per} in Figure 4.1. Moreover, we also plot the level set of the Fourier transform of $\widehat{\chi}$ in Figure 4.2 for the two-dimensional case. For the sake of simplicity in the notation, since the lattice and the dual lattice coincide we simply write α (instead of α^*) for a general element of the lattice or dual lattice.

Periodic tensorial exponential kernel K_E^{per} . Consider the one-dimensional exponential kernel given by

$$K_E(x, y) = \exp(-|x - y|) = \chi_E(|x - y|), \quad x, y \in \mathbb{R}, \quad (4.1a)$$

with $\widehat{\chi}_E(\xi) = \frac{2}{1 + 4\pi^2 \xi^2}$. We make this kernel tensorial and periodic using the localization method in Section 3.3 (see also (6.8) below), that is, for $x, y \in [0, 1]^D$

$$K_E^{\text{per}}(x, y) = \chi_E^{\text{per}}(x - y) = \sum_{\alpha \in \mathbb{Z}^D} \frac{e^{2i\pi \langle x - y, \alpha \rangle}}{\prod_{d=1}^D (1 + 4\pi^2 \alpha_d^2 / \tau_D^2)} = \prod_{d=1}^D \sum_{\alpha_d \in \mathbb{Z}} \frac{e^{2i\pi (x_d - y_d) \alpha_d}}{1 + 4\pi^2 \alpha_d^2 / \tau_D^2}. \quad (4.1b)$$

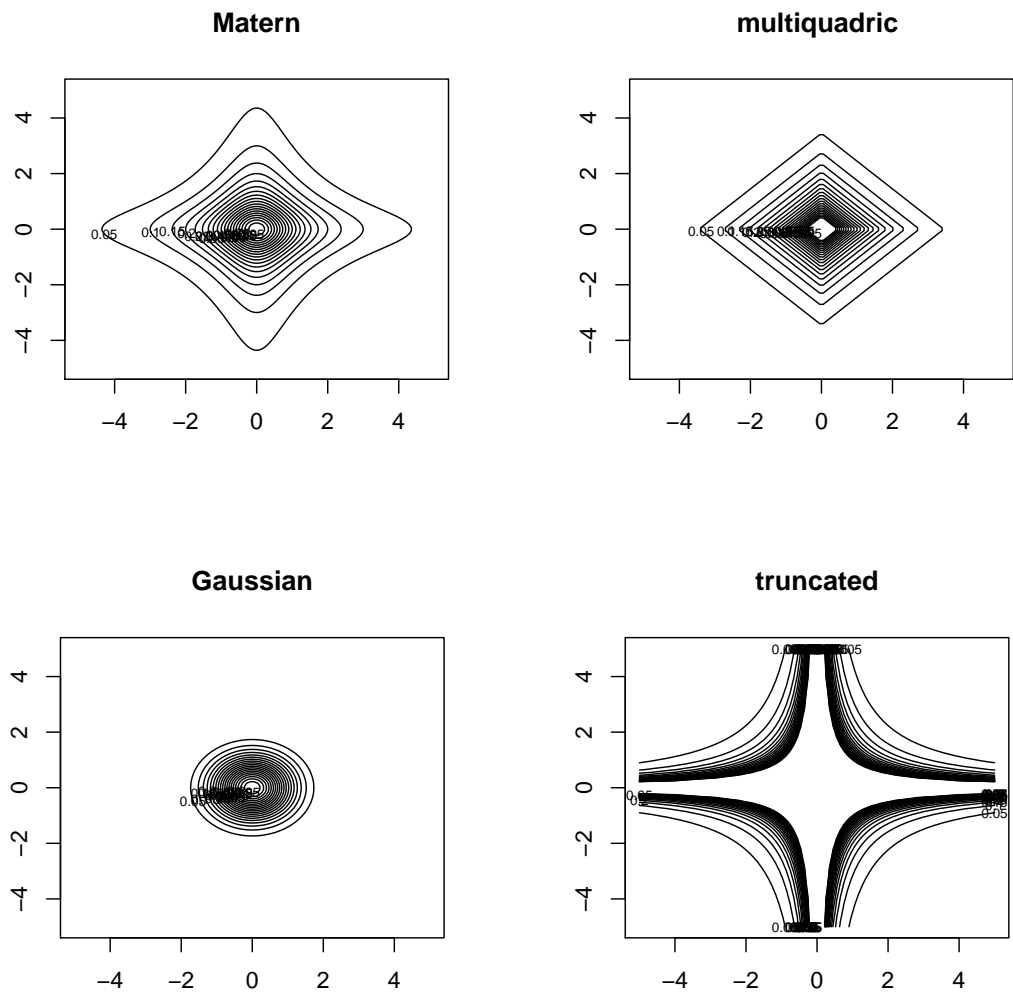


Figure 4.2: Level sets of periodic kernels in two dimensions: exponential, multiquadric, Gaussian, and truncated, respectively.

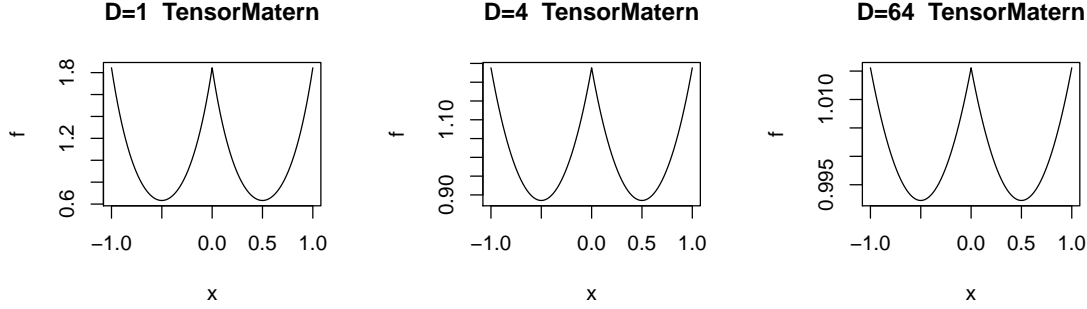


Figure 4.3: Periodic exponential kernel in dimension $D = 1, 4, 64$.

Here, we have also introduced a parameter $0 < \tau_D \rightarrow 0$ as $D \rightarrow +\infty$. Observe that the Fourier transform of $\tau_D e^{-\tau_D |x|}$ is $\frac{2}{1+4\pi^2 \alpha_d^2 / \tau_D^2}$. Thus, thanks to the Poisson formula (3.6), our kernel coincides with

$$\begin{aligned} K_E^{\text{per}}(x, y) &= \sum_{\alpha \in \mathbb{Z}^D} \tau_D \exp(-\tau_D |x - y + \alpha|_1) = \tau_D \sum_{\alpha \in \mathbb{Z}^D} \prod_{d=1}^D \exp(-\tau_D |x_d - y_d + \alpha_d|) \\ &= \prod_{d=1}^D \frac{2\tau_D}{e^{\tau_D} - 1} \left(\exp(\tau_D |x_d - y_d|) + \exp(\tau_D (1 - |x_d - y_d|)) \right). \end{aligned} \quad (4.1c)$$

We plot the corresponding function χ_E^{per} in Figure 4.3 for several dimensions. The Banach space associated to this kernel (defined in (3.10)) reads

$$\mathcal{H}_{K_E^{\text{per}}}^{s,p}([0, 1]^D) = \left\{ \varphi \text{ is periodic on } [0, 1]^D \mid \left\| \left(\prod_{d=1}^D (1 + 4\pi^2 \alpha_d^2 / \tau_D^2) \right)^{s/p} \widehat{\varphi} \right\|_{\ell^{p'}(\mathbb{Z}^D)} < +\infty \right\}. \quad (4.1d)$$

In particular, the space $\mathcal{H}_{K_E^{\text{per}}}^{1,1}([0, 1]^D)$ has been found to be relevant in mathematical finance.

Periodic Gaussian kernel. The Gaussian kernel is translation-invariant, radial, and smooth, and reads

$$K_G(x, y) = \exp(-|x - y|^2) = \prod_{d=1}^D \exp(-|x_d - y_d|^2) = \chi_G(x - y), \quad (4.2)$$

with $\widehat{\chi}_G(\xi) = 2^{-D/2} \exp(-|\xi|^2/4)$. We define its periodic version as

$$K_G^{\text{per}}(x, y) := \frac{1}{\tau_D} \sum_{\alpha \in \mathbb{Z}^D} \prod_{d=1}^D \exp\left(-\frac{|x_d - y_d + \alpha_d|^2}{4\tau_D^2}\right) = \prod_{d=1}^D \vartheta_3(i\pi |x_d - y_d|, \tau_D), \quad (4.3)$$

where $\vartheta_3(z; \tau_D) = \sum_{n \in \mathbb{Z}} (\exp(-\pi^2 \tau_D))^{n^2} \cos(2\pi n z)$ is nothing but the third Jacobi-theta function. Here, τ_D is a numerical parameter. The diagonal term enjoys the following decay $K_G^{\text{per}}(x, x) = \vartheta_3(0; \tau_D)^D \simeq (1 + 2e^{-\pi^2 \tau_D})^D$, and we thus choose $\tau_D = \frac{\ln(2D)}{\pi^2}$ to ensure $K_G^{\text{per}}(x, x) \leq e$. Let us denote (with some abuse of notation) $K_G^{\text{per}}(x, y) = \chi_G^{\text{per}}(x - y) = \prod_d \chi_G^{\text{per}}(x_d - y_d)$. We plot the function χ_G^{per} in Figure 4.4 for several dimensions.

Periodic multiquadric kernel. The one-dimensional multiquadric kernel is translation-invariant, radially-symmetric, and smooth, and reads

$$K_M(x, y) = \frac{1}{1 + |x - y|^2} = \chi_M(|x - y|), \quad x, y \in \mathbb{R}, \quad (4.4a)$$

with $\widehat{\chi}_M(\xi) = \exp(-|\xi|)$, so this kernel is nothing but the Fourier transform of the exponential kernel. above. We define its periodic version by setting, for all $x, y \in [0, 1]^D$,

$$K_M^{\text{per}}(x, y) := \sum_{\alpha \in \mathbb{Z}^D} \prod_{d=1}^D e^{2i\pi \langle x - y, \alpha \rangle - \tau_D |\alpha|_1} = \prod_{d=1}^D \sum_{\alpha_d \in \mathbb{Z}} e^{2i\pi (x_d - y_d) \alpha_d - \tau_D |\alpha_d|}, \quad (4.4b)$$

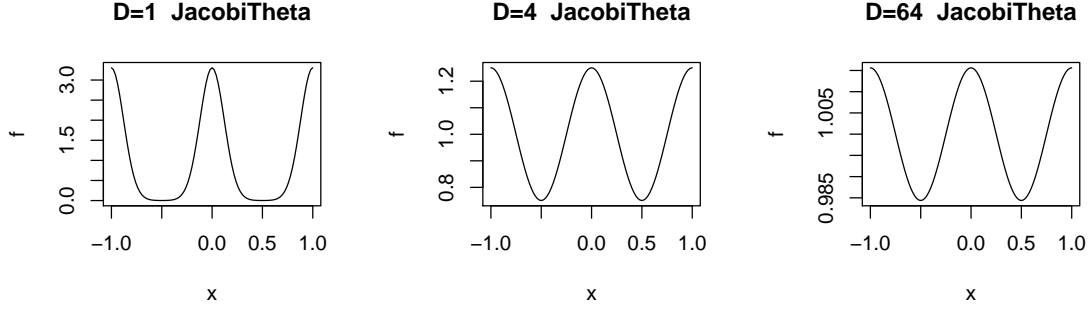


Figure 4.4: Periodic Gaussian kernel in dimension $D = 1, 4, 64$

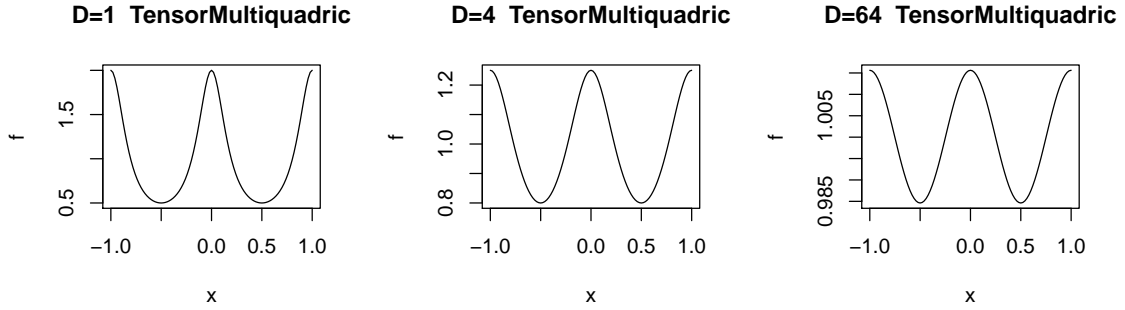


Figure 4.5: Periodic multiquadric kernel in dimension $D = 1, 4, 64$.

where τ_D is a parameter. Observe that the inverse Fourier transform of $\tau_D \exp(-\tau_D|x|)$ is $\frac{1}{1+\alpha_d^2/\tau_D^2}$. Thanks to the Poisson formula, the kernel K_M^{per} coincides with (for $x, y \in [0, 1]^D$)

$$K_M^{\text{per}}(x, y) = \prod_{d=1}^D \sum_{\alpha_d \in \mathbb{Z}} \frac{1}{1 + (x_d - y_d + \alpha_d)^2 / \tau_D^2} = \prod_{d=1}^D \frac{\sinh(2\pi\tau_D)}{\cosh(2\pi\tau_D) - \cos(2\pi(x_d - y_d))}. \quad (4.4c)$$

The diagonal contribution enjoys the following asymptotics

$$K_M^{\text{per}}(x, x) = \prod_{d=1}^D \sum_{\alpha_d \in \mathbb{Z}} \frac{1}{1 + \alpha_d^2 / \tau_D^2} = \left(\frac{\sinh(2\pi\tau_D)}{\cosh(2\pi\tau_D) - 1} \right)^D = \coth(\pi\tau_D/2)^D.$$

Hence we choose $\tau_D = \frac{2 \coth^{-1}(1+1/D)}{\pi} = \frac{2}{\pi} (\ln(2 + 1/D) - \ln(1/(D + 1)))$ to ensure that $K_M^{\text{per}}(x, x) \leq e$ uniformly for any dimension. Finally, we write $K_M^{\text{per}}(x, y) = \chi_M^{\text{per}}(x - y) = \prod_d \chi_M^{\text{per}}(x_d - y_d)$ (with some abuse of notation), and we plot the function χ_M^{per} in Figure 4.5 for several dimensions.

Periodic truncated kernel. The truncated kernel is translation-invariant and Lipschitz continuous only, and reads

$$K_T(x, y) = \sup(1 - |x - y|, 0) = \chi(x - y), \quad \widehat{\varphi}(\xi) = \frac{\sin^2(\pi\xi/2)}{(\pi\xi/2)^2}. \quad (4.5a)$$

We emphasize that its tensor product

$$\prod_{d=1}^D \sup(1 - 2|x_d - y_d|, 0) = \prod_{d=1}^D \chi(x_d - y_d), \quad \widehat{\varphi}(\xi) = \prod_{d=1}^D \frac{\sin^2(\pi\xi_d/2)}{(\pi\xi_d/2)^2} \quad (4.5b)$$

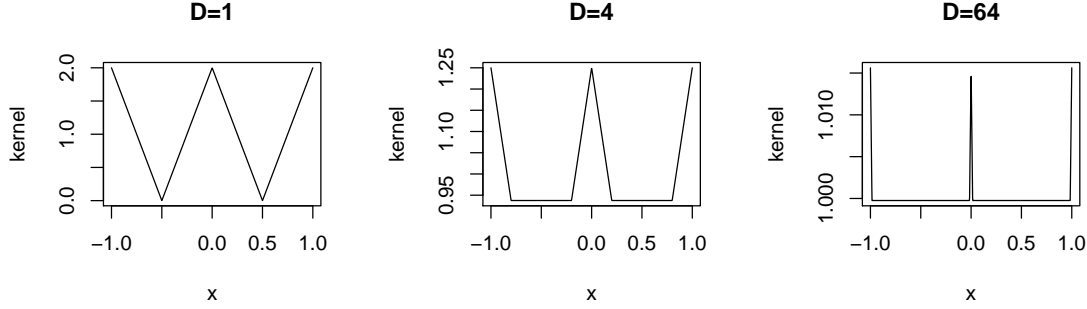


Figure 4.6: Periodic truncated kernels in dimension $D = 1, 4, 64$

typically arises when designing a finite difference scheme on a Cartesian grid (say of the type $z_d^n = n/N$). Using again the Poisson formula (3.6) we define its periodic version as

$$K_T^{\text{per}}(x, y) := \sum_{\alpha \in \mathbb{Z}^D} \prod_{d=1}^D \tau_D \sup(1 - \tau_D |x_d - y_d + \alpha_d|, 0) = \sum_{\alpha^* \in \mathbb{Z}^D} \prod_{d=1}^D \tau_D \frac{\sin^2(\pi \alpha_d / \tau_D)}{(\pi \alpha_d^* / \tau_D)^2} e^{2i\pi(x_d - y_d)\alpha_d^*}, \quad (4.5c)$$

where $1 \leq \tau_D \leq 2$ is a parameter, chosen so that the sum is finite, thus

$$K_T^{\text{per}}(x, y) = \prod_{d=1}^D \sum_{\alpha_d \in \{-1, 0, 1\}} \tau_D \sup(1 - \tau_D |x_d - y_d + \alpha_d|, 0). \quad (4.5d)$$

The diagonal term is $K_T^{\text{per}}(x, x) = \tau_D^D$, and we thus choose $\tau_D = 1 + 1/D$ in order to ensure that $K_T^{\text{per}}(x, x) \leq e$. Writing $K_T^{\text{per}}(x, y) = \chi_T^{\text{per}}(x - y) = \prod_d \chi_T^{\text{per}}(x_d - y_d)$, we plot χ_T^{per} in Figure 4.6 for several dimensions.

4.3 Transported kernels of interest

Objective. An admissible kernel K defined on an open set Ω is said to be a *compactly supported* if it extends to a continuous function on the closure $\bar{\Omega}$ and this extension vanishes on the boundary $\partial\Omega$. For our numerical experiments, we design four compactly supported kernels defined on the unit cube $[0, 1]^D$, determined from the four examples in Table 1.1. Figure 4.7 displays $K(x, 1/2)$ for the two-dimensional case.

Transported tensorial exponential kernel. In view of the standard expression of the exponential kernel, we introduce the following transported version defined on the unit cube $[0, 1]^D$

$$K_E^{\text{tran}}(x, y) = \frac{1}{\beta_D} \exp\left(-\tau_D \left| \text{erf}^{-1}(2x - 1) - \text{erf}^{-1}(2y - 1) \right|_1\right), \quad (4.6a)$$

where $\text{erf}^{-1}(x) = (\text{erf}^{-1}(x_1), \dots, \text{erf}^{-1}(x_D))$ and $|x|_1 = \sum_d |x_d|$ and we the erf function reads $\text{erf}(x) = (2/\sqrt{\pi}) \int_0^x e^{-y^2} dy$. By Proposition 3.1, this kernel corresponds to the transport of the localized kernel

$$K_E^{\text{loc}}(x, y) = \frac{1}{\beta_D} \exp(-\tau_D |x - y|_1) \exp(-|x|^2 - |y|^2), \quad (4.6b)$$

obtained from the map $S(x) = \text{erf}^{-1}(2x - 1)$. Here, we choose

$$\tau_D = \frac{\sqrt{\pi}}{D}, \quad \beta_D = \left(e^{\tau_D^2/4} (1 - \text{erf}(\tau_D/2)) \right)^D$$

with, independently of the dimension D ,

$$K_E^{\text{tran}}(x, x) = \frac{1}{\beta_D} \simeq \left(1 + \frac{\tau_D}{\sqrt{\pi}}\right)^D \leq e, \quad \iint_{[0, 1]^D \times [0, 1]^D} K_E^{\text{tran}}(x, y) dx dy = 1.$$

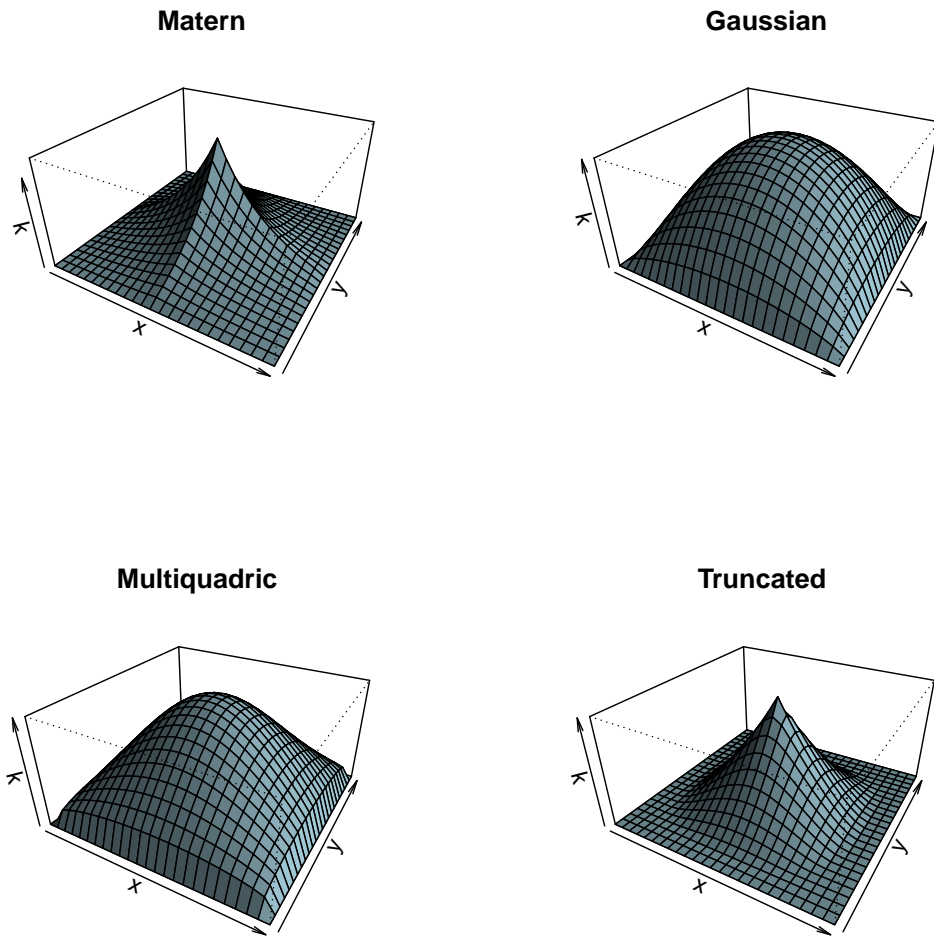


Figure 4.7: Transported kernels in two dimensions: exponential, multiquadric, Gaussian, and truncated, respectively.

Transported multiquadric kernel. In view of the expression of multiquadric kernel, we consider the following localized version on the unit cube $[0, 1]^D$

$$K_M^{\text{tran}}(x, y) = \frac{\beta_D}{(1 + \tau_D^2 |\text{erf}^{-1}(2x - 1) - \text{erf}^{-1}(2y - 1)|^2)^{(D+1)/2}}, \quad (4.7a)$$

corresponding to the following transported kernel

$$K_M^{\text{loc}}(x, y) = \frac{\beta_D \exp(-|x|^2 - |y|^2)}{(1 + \tau_D^2 |x - y|^2)^D}, \quad x, y \in \mathbb{R}^D, \quad (4.7b)$$

where we choose

$$\tau_D = \sqrt{\frac{2}{D}}, \quad \beta_D = \left(\frac{1}{\sqrt{\pi} \exp(1/\tau_D^2) (1 - \text{erf}(1/\tau_D)) (1/\tau_D)} \right)^D \simeq \left(1 + \frac{\tau_D^2}{2} \right)^D,$$

as $\tau_D \rightarrow 0$, determined so that

$$K_M^{\text{tran}}(x, x) = \beta_D \leq e, \quad \int_{[0,1]^D} \int_{[0,1]^D} K_M^{\text{tran}}(x, y) dx dy \simeq \frac{\beta_D}{\left(\sqrt{\pi} \exp(\tau_D^2/2) (1 - \text{erf}(1/\tau_D)) (1/\tau_D) \right)^D} = 1.$$

Transported Gaussian kernel. In view the expression of the Gaussian kernel, we introduce the following localized version on the unit cube $[0, 1]^D$

$$K_G(x, y) = \phi(x - y) = \beta_D \exp\left(-\tau_D^2 \sum_{d=1}^D (\text{erf}^{-1}(2x_d - 1) - \text{erf}^{-1}(2y_d - 1))^2 \right), \quad x, y \in [0, 1]^D, \quad (4.8a)$$

where we choose $\tau_D = \sqrt{\frac{2}{D}}$ and $\beta_D = (1 + \tau_D^2)^{D/2}$, determined so that

$$K_G^{\text{tran}}(x, x) = \beta_D \leq e, \quad \int_{[0,1]^D} \int_{[0,1]^D} K_G^{\text{tran}}(x, y) dx dy \simeq \frac{\beta_D}{(1 + \tau_D^2)^{D/2}} = 1.$$

Transported truncated kernel. In view of the expression of the truncated kernel, we consider the following localized version on the unit cube $[0, 1]^D$

$$K_T^{\text{tran}}(x, y) = \frac{\beta_D}{(1 + \tau_D |\text{erf}^{-1}(2x_d - 1) - \text{erf}^{-1}(2y_d - 1)|^2)^D}, \quad x, y \in [0, 1]^D, \quad (4.9a)$$

corresponding to the following transported kernel

$$K_T^{\text{loc}}(x, y) = \beta_D \frac{\exp(-|x|^2 - |y|^2)}{(1 + \tau_D |x - y|)^D}, \quad x, y \in \mathbb{R}^D. \quad (4.9b)$$

4.4 Further constructions

A compactly supported, tensor product kernel. The example given now is not translation-invariant and is supported on the unit interval $\Omega = [0, 1]$:

$$K^{\text{seed}}(x, y) := y(1 - x)1_{y \leq x} + x(1 - y)1_{x \leq y}, \quad x, y, \in \mathbb{R}. \quad (4.10a)$$

Observe that $K^{\text{seed}}(0, y) = K^{\text{seed}}(1, y) = 0$, and that the Fourier technique above does not apply. For future reference we compute

$$\begin{aligned} \partial_y K^{\text{seed}}(x, y) &= (1 - x)1_{y < x} - x1_{y > x}, & \partial_y^2 K^{\text{seed}}(x, y) &= -\delta_x \\ \int_{[0,1]} K^{\text{seed}}(x, y) dy &= \frac{1}{2}x(1 - x), & \iint_{[0,1]^2} K^{\text{seed}}(x, y) dx dy &= \frac{1}{12}, \end{aligned} \quad (4.10b)$$

together with the following integration by part formula:

$$\int_{[0,1]} \partial_2 K^{\text{seed}}(\cdot, x) \partial_2 K^{\text{seed}}(\cdot, y) dz = \int_{[0,1]} \partial_2 (K^{\text{seed}}(\cdot, x) \partial_2 K^{\text{seed}}(\cdot, y)) dz + K^{\text{seed}}(x, y) = K^{\text{seed}}(x, y). \quad (4.10c)$$

Hence, the space $\mathcal{H}_{K^{\text{seed}}}(0, 1)$ is the homogeneous Sobolev space $\dot{H}_0^1([0, 1])$ of functions that vanish at the boundary, defined by density from the set of smooth functions supported on $(0, 1)$ in the $\ell^2(0, 1)$ -norm. The eigenvalues and eigenfunctions of the spectral decomposition is given by solving $\int_{[0,1]} K^{\text{seed}}(\cdot, x) \zeta_i(x) dx = \lambda_i \zeta_i$ or, equivalently, $\partial_x^2 \zeta_i = -\frac{1}{\lambda_i} \zeta_i$. We find $\zeta_i(x) = \sin(i\pi x)$, $\lambda_i = \frac{1}{(i\pi)^2}$ for $i = 1, 2, \dots$

A multi-dimensional version. Using the kernel K^{seed} in (4.10a), in general dimensions we consider $\Omega = [0, 1]^D$ and the following kernel

$$K(x, y) = \prod_{1 \leq d \leq D} K^{\text{seed}}(x_d, y_d) = \prod_{1 \leq d \leq D} \left(y_d(1 - x_d) 1_{y_d \leq x_d} + x_d(1 - y_d) 1_{x_d \leq y_d} \right). \quad (4.11a)$$

In view of

$$\int_{\Omega} (\partial_1 \dots \partial_D) K(\cdot, x) (\partial_1 \dots \partial_D) K(\cdot, y) dz = \prod_{1 \leq d \leq D} \left(\int_{[0,1]} \partial_d K^{\text{seed}}(x_d, \cdot) \partial_d K^{\text{seed}}(y_d, \cdot) dz \right) = K(x, y), \quad (4.11b)$$

the space generated by this kernel corresponds to the norm $\int_{(0,1)^D} (|\varphi(x)|^2 + |\partial_1 \dots \partial_D \varphi(x)|^2) dx$. Similarly to the one-dimensional case, we compute

$$\begin{aligned} \partial_d K(x, y) &= (\partial K^{\text{seed}})(x_d, y_d) \prod_{e \neq d} K^{\text{seed}}(x_e, y_e), & \prod_{1 \leq d \leq D} \partial_d^2 K(x, y) &= (-1)^D \delta_x \\ \int_{[0,1]^D} K(x, y) dy &= \prod_{1 \leq d \leq D} \frac{1}{2} x_d (1 - x_d), & \iint_{[0,1]^D \times [0,1]^D} K(x, y) dx dy &= \frac{1}{12^D}. \end{aligned} \quad (4.11c)$$

The eigenvalues and eigenfunctions of the the spectral decomposition are now

$$\zeta_{\alpha}(x) = \prod_{1 \leq d \leq D} \sin(\alpha_d \pi x_d), \quad \lambda_{\alpha} = \frac{1}{\prod_{1 \leq d \leq D} (\alpha_d \pi)^2}, \quad \alpha = (\alpha_1, \dots, \alpha_D) \in \mathbb{N}^D. \quad (4.11d)$$

5 Three formulation of the discrepancy function

5.1 Formulation in physical variables

We now factor out the integration error into (1) a factor depending upon the *regularity* of the function φ under consideration, measured in the \mathcal{H}_K -norm, which is independent of the choice of the interpolation points, and (2) a factor depending solely upon the kernel K and the mesh points, which is independent of the choice of the function. Three equivalent formulations of the second term are now derived in the physical, spectral, or discrete Fourier variables. Although these formulations are in principle equivalent, they shed a very different light on the problem of interest. We begin with an expression in the physical variables, which we will not use directly for the derivation of actual estimates, as it appears to be difficult to work with.

Proposition 5.1 (Factorization in physical variables). *Consider an admissible kernel $K = K(x, y)$ defined on an open set $\Omega \subset \mathbb{R}^D$. Then, for any function φ in the Hilbert Space $\mathcal{H}_K(\Omega)$, one has*

$$\left| \int_{\Omega} \varphi(x) dx - \frac{1}{N} \sum_{1 \leq n \leq N} \varphi(y^n) \right| \leq E_K(Y, N, D) \|\varphi\|_{\mathcal{H}_K(\Omega)} \quad (5.1)$$

for any set $Y = (y_1, \dots, y_N)$ of points in Ω , with the error function

$$E_K(Y, N, D)^2 := \frac{1}{N^2} \sum_{n, m=1}^N \iint_{\Omega \times \Omega} \left(K(x, y) + K(y^n, y^m) - K(y^n, y) - K(x, y^m) \right) dx dy \quad (5.2a)$$

or, equivalently,

$$E_K(Y, N, D) = \|e_K(Y)\|_{\mathcal{H}_K(\Omega)}, \quad e_K(Y) := \int_{\Omega} \left(K(\cdot, x) - \frac{1}{N} \sum_{1 \leq n \leq N} K(\cdot, y^n) \right) dx. \quad (5.2b)$$

Remark 5.2. We observe that the integrand in (5.2a) is non-negative and can be also written in terms of the pseudo-distance \mathbf{D} (see (2.2)), that is,

$$E_K(Y, N, D)^2 = -\frac{1}{2N^2} \sum_{n,m=1}^N \iint_{\Omega \times \Omega} \left(\mathbf{D}(x, y) + \mathbf{D}(y^n, y^m) - \mathbf{D}(y^n, y) - \mathbf{D}(x, y^m) \right) dx dy. \quad (5.3)$$

Recall that (2.12) shows that $\mathbf{D}(x, y)$ measures the “distance” between $K(\cdot, x)$ and $K(\cdot, y)$, and (5.3) provides an equivalent representation of the error.

5.2 Formulation in spectral variables

Proposition 5.3 (Minimizing the error function in spectral variables). *Consider an admissible kernel $K : \Omega \times \Omega \rightarrow \mathbb{R}$ together with its Mercer representation $(\lambda_i, \zeta_i)_{i \geq 1}$. Then the error function³*

$$E_K^{s,p}(N, D) := \inf_{y^1, \dots, y^N \in \Omega} \left(\sum_{i \geq 1} \lambda_i^{sp'/2} \left(\int_{\Omega} \zeta_i(x) dx - \frac{1}{N} \sum_{1 \leq n \leq N} \zeta_i(y^n) \right)^{p'} \right) \quad (5.4)$$

(for $s \geq 0$ and p' begin the dual exponent to $p \in [1, +\infty)$) satisfies

$$E_K^{s,p}(N, D) \leq \left(\sum_{i > N} \lambda_i^{sp'/2} \left(\int_{\Omega} \zeta_i(x) dx - \frac{1}{N} \sum_{1 \leq n \leq N} \zeta_i(y^n) \right)^{p'} \right), \quad (5.5a)$$

where y^1, \dots, y^N solve the system of equations

$$\int_{\Omega} \zeta_i(x) dx - \frac{1}{N} \sum_{1 \leq n \leq N} \zeta_i(y^n) = 0. \quad (5.5b)$$

Whenever the eigenvalues and functions are known explicitly, the spectral decomposition associated with the kernel can be used in combination with the following formula for the error function. Observe that we can now deal with the spaces $\mathcal{H}_K^{s,p}(\Omega)$ with general integrability and differentiability exponents, and recall that $E_K(Y, N, D) = E_K^{1,2}(Y, N, D)$.

Proposition 5.4 (Factorization in spectral variables). *Consider an admissible kernel $K = K(x, y)$ defined on an open set $\Omega \subset \mathbb{R}^D$. Then, for any function φ in the corresponding Banach space $\mathcal{H}_K^{s,p}(\Omega)$ associated with exponents $s > 0$ and $p \in [1, +\infty)$. Then, for any set $Y = (y^1, \dots, y^N)$ of points in Ω and any function $\varphi \in \mathcal{H}_K^{s,p}(\Omega)$, one has*

$$\left| \int_{\Omega} \varphi(x) dx - \frac{1}{N} \sum_{1 \leq n \leq N} \varphi(y^n) \right| \leq E_K^{s,p}(Y, N, D) \|\varphi\|_{\mathcal{H}_K^{s,p}(\Omega)}, \quad (5.6)$$

where the error function (with $1/p + 1/p' = 1$)

$$E_K^{s,p}(Y, N, D) := \left(\sum_{i \geq 1} \lambda_i^{sp'/2} \left(\int_{\Omega} \zeta_i(x) dx - \frac{1}{N} \sum_{1 \leq n \leq N} \zeta_i(y^n) \right)^{p'} \right)^{1/p'} = \|e_K(Y)\|_{\mathcal{H}_K^{s,p'}(\Omega)} \quad (5.7)$$

is expressed in terms of the eigenfunctions and eigenvectors ζ_i, λ_i of the Mercer representation of the kernel K .

³This expression is equivalent to the one given in the introduction.

5.3 Formulation in discrete Fourier variables

Assuming now more structure on the kernel and, specifically, assuming that it is based on a discrete lattice, we express the error function in a third form based on the Poisson formula.

Proposition 5.5 (Factorization in discrete Fourier variables). *Consider a kernel based on a discrete lattice \mathbf{L} with elementary cell \mathbf{C} , say $K^{per}(x, y) := \frac{1}{|\mathbf{C}|} \langle e^{2i\pi \langle x-y, \cdot \rangle}, \rho \rangle_{\ell^2(\mathbf{L}^*)}$, determined from a generating function ρ defined on the dual lattice \mathbf{L}^* . Then, for any set $Y = (y^1, \dots, y^N)$ of points in \mathbf{C} and any \mathbf{L} -periodic function $\varphi \in \mathcal{H}_{K^{per}}^{s,p}(\Omega)$ (for some $s > 0$ and $p \in [1, +\infty)$ with $1/p + 1/p' = 1$), one has (5.6) with*

$$E_{K^{per}}^{s,p}(Y, N, D) = \left(\sum_{\alpha^* \in \mathbf{L}^*} \rho^s \left| \int_{\mathbf{C}} e^{2i\pi \langle x, \alpha^* \rangle} dx - \frac{1}{N} \sum_n e^{2i\pi \langle y^n, \alpha^* \rangle} \right|^{p'} \right)^{1/p'}. \quad (5.8)$$

6 Controlling the discrepancy function

6.1 Physical variables

We seek for a set of points minimizing the error function and, first in the physical variables, we obtain the following result whose proof is given in the appendix.

Proposition 6.1 (Minimizing the error function in physical variables). *Let $K : \Omega \times \Omega \rightarrow \mathbb{R}$ be an admissible kernel defined on some open set Ω , and consider the error function (5.1). Provided the kernel is convex with respect to each variable, the functional $Y = (y^1, \dots, y^N) \mapsto (E_K(Y, N, D))^2$ is a positive (non-strictly) convex with positive infimum*

$$E_K(N, D) := \inf_{Y=(y^1, \dots, y^N)} E_K(Y, N, D) > 0.$$

Moreover, if Y be a minimizer, then the gradient of the functional vanishes at Y , namely

$$\int_{\Omega} \nabla K(x, y^n) dx = \frac{1}{N} \sum_{m=1}^N (\nabla K)(y^n, y^m), \quad n = 1, 2, \dots, N. \quad (6.1)$$

Furthermore, for all functions $\mathcal{H}_K^Y(\Omega)$ (as defined in (2.3)), the following integration formula holds (without error term):

$$\int_{\Omega} \varphi(x) dx = \frac{1}{N} \sum_{1 \leq n \leq N} \varphi(y^n), \quad \varphi \in \mathcal{H}_K^Y(\Omega). \quad (6.2)$$

Here, $\nabla K(x, y)$ stands here for any of $\nabla_x K(x, y)$ or $\nabla_y K(x, y)$, since the kernel K is symmetric in its two arguments. The following comments are in order:

- The functional is be *totally symmetric* with respect to its arguments so, clearly, a minimizer is never unique. Yet, suppose that our kernel $K(x, y)$ is concave, then the following semi-discrete algorithm

$$\frac{d}{dt} y^n = \int_{\Omega} \nabla K(x, y^n) dx - \frac{1}{N} \sum_{m,n=1}^N (\nabla K)(y^n, y^m), \quad n = 1, 2, \dots, N \quad (6.3)$$

converges toward a minimum of the functional E_K . Observe also that the dynamical system under consideration involves two terms: $\nabla K(y^n, y^m)$ tends to push points away from each other, while the integral term tends to attract the points toward the mass-center of Ω . These two competitive effects leads to a non-trivial distribution of the points.

- With the partition of unity defined in (2.10b), any minimizer must satisfy $\int_{\Omega} \delta^n(x) dx = \frac{1}{N}$, for $n = 1, 2, \dots, N$.

Best discrepancy sequences in one dimension. We are now in a position to analyze an example in dimension $D = 1$, where we have fully explicit expressions.

Proposition 6.2 (The case of dimension $D = 1$). *For the kernel (4.10), the equally-spaced sequence*

$$y^n = \frac{2n-1}{2N}, \quad n = 1, 2, \dots, N \quad (6.4)$$

is the unique solution (up to re-ordering) that minimizes the error function $Y \mapsto E_K(Y)$ introduced in (5.1) and, with this sequence, the following estimate holds:

$$\left| \int_{[0,1]} \varphi(x) dx - \frac{1}{N} \sum_{1 \leq n \leq N} \varphi(y^n) \right| \leq \frac{1}{\sqrt{6N}} \|\varphi\|_{\mathcal{H}_K^{1,2}([0,1])}, \quad \varphi \in \mathcal{H}_K^{1,2}([0,1]).$$

6.2 Spectral variables

For clarity in the presentation, let us repeat here our previous observation. Since the following results will be of a direct application in the numerical section, we include the remaining proofs in the main text.

Proposition 6.3 (Minimizing the error function in spectral variables). *Consider an admissible kernel $K : \Omega \times \Omega \rightarrow \mathbb{R}$ and denote by $(\lambda_i, \zeta_i)_{i \geq 1}$ its Mercer representation. Then for the error function (5.7) after minimization*

$$E_K^{s,p}(N, D) := \inf_{y^1, \dots, y^N \in \Omega} E_K^{s,p}(Y, N, D) \quad (6.5)$$

one has

$$E_K^{s,p}(N, D) \leq \left(\sum_{i > N} \lambda_i^{sp'/2} \left(\int_{\Omega} \zeta_i(x) dx - \frac{1}{N} \sum_{1 \leq n \leq N} \zeta_i(y^n) \right)^{p'} \right)^{1/p'}, \quad (6.6a)$$

where $(y^1, \dots, y^N) \in \mathbb{R}^{N \times D}$ is given by the following system of equations

$$\int_{\Omega} \zeta_i(x) dx - \frac{1}{N} \sum_{1 \leq n \leq N} \zeta_i(y^n) = 0. \quad (6.6b)$$

Proof. The expression $K_N(x, y) = \sum_{1 \leq n \leq N} \lambda_n \zeta_n(x) \zeta_n(y)$ defines an admissible kernel, for which we have an orthogonal decomposition of the form $\mathcal{H}_K(\Omega) = \mathcal{H}_{K_N}(\Omega) \oplus \mathcal{H}_{K_N}^{\perp}(\Omega)$. It suffices to consider the minimization problem

$$E_{K_N}(N, D) = \inf_{y^1, \dots, y^N \in \Omega} \left(\sum_{1 \leq i \leq N} \lambda_i \left(\int_{\Omega} \zeta_i(x) dx - \frac{1}{N} \sum_{1 \leq n \leq N} \zeta_i(y^n) \right)^2 \right)^{1/2}$$

and

$$\sum_{1 \leq i \leq N} \lambda_i \left(\int_{\Omega} \zeta_i(x) dx - \frac{1}{N} \sum_{1 \leq n \leq N} \zeta_i(y^n) \right) (\nabla \zeta_i)(y^n) = 0, \quad n = 1, \dots, N. \quad \square$$

Best discrepancy sequences in one dimension. We can now revisit the example in dimension $D = 1$ treated in Proposition 6.2.

Proposition 6.4 (The case of dimension $D = 1$). *For the kernel (4.10) and the equally-spaced sequence*

$$y^n = \frac{2n-1}{2N}, \quad n = 1, 2, \dots, N, \quad (6.7)$$

the error function $Y \mapsto E_K^{s,p}(Y, N, 1)$ introduced in (5.7) can be bounded by, for $s > 1 - 1/p$ and some constant $C_{s,p} > 0$:

$$\left| \int_{[0,1]} \varphi(x) dx - \frac{1}{N} \sum_{1 \leq n \leq N} \varphi(y^n) \right| \leq \frac{C_{s,p}}{N^s} \|\varphi\|_{\mathcal{H}_K^{s,p}([0,1])}, \quad \varphi \in \mathcal{H}_K^{s,p}([0,1]).$$

Proof. The general case follows using the spectral formulation (5.7), where $\zeta_i(x) = \sin(i\pi x)$ and $\lambda_i = \frac{1}{(i\pi)^2}$. We introduce

$$\begin{aligned} c_i &:= \int_{[0,1]} \zeta_i(x) dx - \frac{1}{N} \sum_{1 \leq n \leq N} \zeta_i(y^n) \\ &= \int_{[0,1]} \sin(i\pi x) dx - \frac{1}{N} \sum_{1 \leq n \leq N} \sin\left(i\pi \frac{2n-1}{2N}\right) = \frac{1 - \cos(i\pi)}{i\pi} - \frac{\sin^2\left(\frac{i\pi}{2}\right)}{N \sin\left(\frac{i\pi}{2N}\right)}. \end{aligned}$$

Hence, $c_{2i} = 0$, $c_{2i+1} = \frac{2}{(2i+1)\pi} - \frac{1}{N \sin(\frac{(2i+1)\pi}{2N})}$ and, in view of (5.7),

$$E_K^{s,p}(Y, N, 1)^{p'} = \sum_{i \geq 0} \frac{1}{((2i+1)\pi)^{sp'}} \left(\frac{2}{(2i+1)\pi} - \frac{1}{N \sin(\frac{(2i+1)\pi}{2N})} \right)^{p'} = \sum_{i \geq 0} \lambda_{2i+1}^{sp'/2} \left(2\lambda_{2i+1}^{1/2} - \frac{1}{N \sin(\frac{1}{2N\lambda_{2i+1}^{1/2}})} \right)^{p'}$$

where $1/p + 1/(p') = 1$. In particular, $E_K^{1,2}(Y, N, 1) = \frac{1}{\sqrt{6N}}$, as computed in Proposition 6.2. For general s, p , we get the expansion $\left(2\lambda_{2i+1}^{1/2} - \frac{1}{N \sin(\frac{1}{2N\lambda_{2i+1}^{1/2}})} \right) \simeq \frac{\lambda_{2i+1}^{-1/2}}{N^2}$ as $\frac{\lambda_{2i+1}^{-1/2}}{2N} < 1$, that is $i\pi < N$, and is bounded elsewhere.)

The error term in (5.7) is bounded as follows:

$$E_K^{s,p}(Y, N, 1)^{p'} \leq \sum_{\pi i \leq N} \frac{C}{N^{2p'}} \lambda_{2i-1}^{(s-1)p'/2} + C \sum_{i\pi \geq N} \lambda_{2i+1}^{sp'/2},$$

where the first term in the right hand-side is uniformly bounded by $\frac{C}{N^{2p'}}$ provided $(s-1)p' > 1$, that is, $s > 1/p$. The second term is bounded by the first term $\frac{C}{N^{sp'}}$ provided $sp' > 1$. Thus we can control $E_K^{s,p}(Y, N, 1)$ by C/N^s in the range $sp' > 1$ at least. Observe finally that we are led to the same estimate, since

$$(E_K^{s,p}(N, 1))^{p'} \leq \sum_{i > N} \lambda_i^{sp'/2} \left(\int_{[0,1]^D} \zeta_i(x) dx - \frac{1}{N} \sum_{1 \leq n \leq N} \zeta_i(y^n) \right)^{p'} \lesssim \frac{1}{N^{sp'}}. \quad \square$$

6.3 Discrete Fourier variables: toward a sharp estimate

Our third formulation provides us with the most practical setup for the applications, when the spectral decomposition may not be explicitly available.

Proposition 6.5. *Let \mathbf{L} a lattice with elementary cell \mathbf{C} normalized such that $|\mathbf{C}| = 1$, and consider \mathbf{L}^* its dual lattice. Let $\alpha^* \in \mathbf{L}^* \mapsto \rho(\alpha^*)$ be any discrete function satisfying $\rho \in \ell^1(\mathbf{L}^*)$, $\rho \geq 0$, and $\rho(0) = 1$. Then the kernel*

$$K^{\text{per}}(x, y) = \sum_{\alpha^* \in \mathbf{L}^*} \rho(\alpha^*) e^{2i\pi \langle x-y, \alpha^* \rangle} \quad (6.8)$$

is an admissible kernel with period \mathbf{C} and in the Hilbert space $\mathcal{H}_{K^{\text{per}}}(\mathbf{C})$ one has

$$\left| \int_{\Omega} \varphi(x) dx - \frac{1}{N} \sum_{n=1}^N \varphi(y^n) \right| \leq E_{K^{\text{per}}}(Y, N, D) \|\varphi\|_{\mathcal{H}_K}, \quad (6.9)$$

where

$$E_{K^{\text{per}}}(Y, N, D)^2 = \frac{1}{N^2} \sum_{n,m=1}^N K^{\text{per}}(y^n, y^m) - 1, \quad (6.10)$$

where Y is any sequence of N points in \mathbf{C} . Moreover, setting $\tilde{Y} := \arg \inf_{Y \in \mathbf{C}^N} E_K(Y, N, D)$, one has

$$E_K(N, D)^2 \leq \sum_{n > N} \frac{\rho(\alpha^{*n})}{N^2} \left| \sum_{n=1}^N e^{2i\pi \langle y^n, \alpha^{*n} \rangle} \right|^2, \quad (6.11)$$

where the ordering $n \mapsto \alpha^n$ is defined so that $n \mapsto \rho(\alpha^{*n})$ is a decreasing sequence along the points $\alpha \in L^*$ while y^1, \dots, y^N are defined by solving the set of equations (when $\alpha^{*n} \neq 0$)

$$\sum_{m=1}^N e^{2i\pi \langle y^m, \alpha^{*n} \rangle} = 0, \quad 1 \leq n \leq N. \quad (6.12)$$

We conjecture that the system (6.12) does admit a solution for any set $\{\alpha^{*n} \in L^*\}_{1 \leq n \leq N}$. Numerically, for this system for each of the periodic kernels of interest we computed (see below) a numerical approximation for a broad range of dimensions D and integers N . Moreover, the existence of such points can be

established rigorously for several examples of kernels. Importantly, we tested numerically (see below) that the following estimate is sharp:

$$E_K(N, D) \simeq \frac{1}{N} \sum_{n>N} \rho(\alpha^{*n}). \quad (6.13)$$

In our examples, we will be able to compute the above sum, both numerically and analytically, and it will be proven to provide us with a *sharp estimate* of the discrepancy error, although we cannot establish rigorously the validity of this estimate. In the context of Proposition 5.5, we will compute numerically

$$\sum_{n>N} \rho(\alpha^{*n}) = \sum_{n \geq 0} \rho(\alpha^{*n}) - \sum_{n \leq N} \rho(\alpha^{*n}) = K^{\text{per}}(0, 0) - \sum_{n \leq N} \rho(\alpha^{*n}). \quad (6.14)$$

Another theoretical standpoint is obtained by the following “level-set argument”, namely by relying on the approximation

$$\sum_{n>N} \rho(\alpha^{*n}) \simeq \int_{\xi / \rho(\xi) < m^{-1}(N)} \rho(\xi) d\xi, \quad (6.15)$$

where m^{-1} is the inverse of the function $m(\epsilon) = |\{\rho(\xi) > \epsilon\}|$, where $\rho(\xi)$ is a smooth extension of ρ on the whole space \mathbb{R}^D satisfying $\int_{\mathbb{R}^D} \rho(\xi) d\xi = 1$ with $\rho(\xi) \geq 0$ and $\rho(0) = 1$. Unfortunately, this approximation is very inaccurate in the examples we have considered.

In particular, if a transport map of the function is known, that is a one-to-one map $y : \mathbb{R}^D \rightarrow \Omega$ such that $dy = \rho(\xi) d\xi$, then, denoting its inverse as $\xi(y)$, the last integral reduces to measuring the following set

$$\sum_{n>N} \rho(\alpha^{*n}) \simeq \int_{y: \rho(\xi(y)) < m^{-1}(N)} dy = |\{y : \rho(\xi(y)) < m^{-1}(N)\}|. \quad (6.16)$$

Let us make some further remarks:

- If one want to study a general kernel $K(x, y) = \chi(x - y)$ with $x, y \in \Omega$, then one can compute an upper-bound using 6.11 with the Fourier coefficients on the doubled lattice L_2 , generated by the cell $\Omega_2 := \{x - y / x, y \in \Omega\}$. This is done using the Poisson formula, i.e.

$$K(x, y) = \chi(x - y) = \frac{\langle e^{2i\pi \langle x-y, \cdot \rangle}, \rho \rangle_{l^2(\mathbf{L}_2^*)}}{|\Omega_2|}, \quad \rho(\xi) = \frac{1}{(2\pi)^{D/2}} \int_{\Omega_2} \chi(\zeta) e^{-2i\pi \langle \xi, \zeta \rangle} d\zeta. \quad (6.17)$$

However, in practical terms, computing Fourier coefficients can be a quite difficult task.

- We can specify directly the Fourier coefficients $\chi(\alpha^*)$ in the light of Proposition 5.5. This obviously simplifies the problem of finding suitable parameters since they are automatically set by the two conditions $\chi(0) = 1$ and $\chi \in \ell^1(\mathbf{L}^*)$.
- It is clear from this result that the price to pay, when the dimension of the problem increases, is increasing also the *regularity* of the kernel, and the space \mathcal{H}_K contains functions that are more regular as the dimension increases.
- The lattice \mathbf{L} certainly plays an important role in finding the best constant $E_K(N, D)$. If χ is symmetric, the present discussion connects with the sphere packing problem. For instance, for $D = 2$, the best lattice is the hexagonal one, for $D = 3$ the dodecahedral one (Kepler conjecture), while $D = 8, 24$ was treated in [21]. Other cases remain opened.

Proof of Proposition 5.5. To derive (6.9), we recall from (5.2a) that

$$E_K(Y, N, D)^2 = \iint_{\mathbf{C} \times \mathbf{C}} K(x, y) dx dy + \frac{1}{N^2} \sum_{n, m=1}^N K(y^n, y^m) - \frac{2}{N} \sum_n \int_{\Omega} K(x, y^n) dx.$$

We use the fact that $K(x, y) = \chi(x - y)$ is \mathbf{L} -periodic and deduce that the right-hand side is a constant, and does not depend on y^n , therefore

$$\int_{\mathbf{C}} K(x, y) dx = C_0 = \iint_{\mathbf{C} \times \mathbf{C}} K(x, y) dx dy, \quad y \in \mathbf{C},$$

since $|\mathbf{C}| = 1$. Consider (6.8) and the error function expressed in Fourier variable (5.8) with $s = 1$ and $p = 2$, that is,

$$E_K(Y, N, D)^2 = \sum_{\alpha^* \in \mathbf{L}^*} \chi(\alpha^*) \left| \int_{\mathbf{C}} e^{2i\pi \langle x, \alpha^* \rangle} dx - \frac{1}{N} \sum_{n=1}^N e^{2i\pi \langle y^n, \alpha^* \rangle} \right|^2.$$

We have

$$\int_{\mathbf{C}} e^{2i\pi \langle x, \alpha^{*n} \rangle} dx = 1 \text{ if } \alpha^{*n} = 0, \quad \text{while it is 0 otherwise.}$$

When $\alpha^* \neq 0$ we consider

$$E_K(Y, N, D)^2 = \sum_{\alpha^* \in \mathbf{L}^*, \alpha^* \neq 0} \frac{\chi(\alpha^*)}{N^2} \left| \sum_{n=1}^N e^{2i\pi \langle y^n, \alpha^* \rangle} \right|^2. \quad (6.18)$$

Using the ordering $n \mapsto \alpha^n$ associated with the function $\chi(\alpha^{*n})$ we arrive at the system of equations (6.12). \square

The existence of solutions to (6.12) can for instance be established for the canonical lattice $\mathbf{C} = [0, 1]^D$ and $\mathbf{L} = \mathbf{L}^* = \mathbb{Z}^D$, and any set of points $\{\alpha^* : L_d \leq \alpha_d^* < R_d\}$, with size $N = \prod_{d=1}^D (R_d - L_d)$. Without loss of generality we assume that $L_d = 0$ and we set $N_e = \prod_{d=1}^{e-1} R_d$ for $e = 1, \dots, D$ (with the convention that $N_1 = 1$). We define a one-to-one map from the set $\{1 \leq n \leq N\}$ to $\{0 \leq n_d < R_d\}_{d=1, \dots, D}$ by using the map in

$$n(n_1, \dots, n_D) = \sum_{d \leq D} N_d n_d, \quad \alpha^{*n} = (n_1, \dots, n_D), \quad y^n = \left(\frac{n_d(n)}{R_d} \right)_{d=1, \dots, D}.$$

We compute

$$\begin{aligned} \sum_{n=1}^N e^{2i\pi \langle y^n, \alpha^{*n} \rangle} &= \sum_{n=1}^N e^{2i\pi \sum_d \frac{n_d(n)}{R_d} n_d(n)} = \sum_{n_1=L_1}^{R_1-1} e^{2i\pi \frac{n_1}{R_1} n_1^m} \dots \sum_{n_D=L_D}^{R_D-1} e^{2i\pi \frac{n_D}{R_D} n_D^m} \\ &= \left(\frac{1 - e^{2i\pi \frac{n_1^m}{R_1}}}{1 - e^{2i\pi \frac{n_1^m}{R_1}}} \right) \dots \left(\frac{1 - e^{2i\pi \frac{n_D^m}{R_D}}}{1 - e^{2i\pi \frac{n_D^m}{R_D}}} \right) = 0. \end{aligned}$$

Moreover, for computing the solutions to (6.12) for an arbitrary set $\{\alpha^{*n} \in \mathbb{Z}^D\}$ with size N , we can consider the functional $I(Y) := \sum_{n=1}^N \left| \sum_{m=1}^N e^{2i\pi \langle y^m, \alpha^{*n} \rangle} \right|^2$. It is regular and admits the derivatives $\partial_{y^k} I(Y) = \sum_{n=1}^N 4i\pi \alpha^{*n} \sum_{m=1}^N e^{2i\pi \langle y^m - y^k, \alpha^{*n} \rangle}$. Therefore, we can use a gradient-based method in order to numerically compute the solutions, and we expect that any local minimum of the functional $I(Y)$ will also be a global minimum.

7 A comparative study for a selection of kernels

7.1 Our strategy for the numerical study

The aim of this section is to numerically compute the optimal sequences \bar{Y} and the discrepancy functions sdefined by (with $\Omega = [0, 1]^D$)

$$\bar{Y} = \arg \inf_{Y \in \Omega^N} E_K(Y, N, D), \quad E_K(N, D) = E_K(\bar{Y}, N, D). \quad (7.1)$$

We treat two classes of interest:

- Transported kernel: we recall (see (5.2a)) that the following quantity provides the worst Monte-Carlo-type integration error for any admissible kernel $K = K^{\text{loc}}$ localized to the set Ω :

$$E_K(Y, N, D)^2 = \frac{1}{N^2} \sum_{n, m=1}^N \iint_{\Omega \times \Omega} \left(K^{\text{loc}}(x, y) + K^{\text{loc}}(y^n, y^m) - K^{\text{loc}}(y^n, y) - K^{\text{loc}}(x, y^m) \right) dx dy. \quad (7.2)$$

- Periodic kernels: in the context of Proposition 5.5, the expression of the discrepancy function takes a simpler form and in the limit $N, D \mapsto \infty$ we can compare our results to the asymptotic expression

$$E_K(Y, N, D)^2 = \frac{1}{N^2} \sum_{n,m=1}^N K^{\text{per}}(y^n, y^m) - 1, \quad E_K(N, D) \leq \sum_{n>N} \rho(\alpha^{*n}). \quad (7.3)$$

We recall that α^n are chosen on the dual lattice so that $n \mapsto \chi(\alpha^{*n})$ is decreasing. We choose here the canonical lattice $\mathbf{L} = \mathbb{Z}^D$ for simplicity.

We start from the four examples in Table 1.1, that are defined in the whole space \mathbb{R}^D , and design the eight kernels listed in Table 1.2. A normalization parameter, depending on the dimension was determined explicitly for each case. In each case, we compute $E_K(Y, N, D)$ for the following two sequences:

- Y is a randomly chosen sequence.
- Y is a numerical sequence that approximates the optimal solution \bar{Y} .

The numerical results for $E_K(Y, N, D)$ are displayed as tables for $N = 16, 32, \dots, 512$ and $D = 1, 2, 4, \dots, 128$. and we can compare the error obtained for various values N, D and various kernels. The kernels in the second line of Table 1.2 are periodic, so that we can compute the upper bounds given by the formula (6.11) and we can compare it to the numerically computed results. As a general rule, the computation time grows like $C DN^3$ for some rather large constant $C > 0$. For instance, on a machine Intel I7 7700K, each result in the tables presented below took between 10 secondes and up to two minutes for the tests in dimension $D = 128$ and with $N = 512$. In order to ease the comparison between the various tests and kernels we have considered, we are collecting our results in Tables 7.1, 7.2, and 7.3 when $N = 512$ and in dimensions 1, 16, 128. Remarkably, our asymptotic analysis provides an accurate prediction of the numerical error (except for the truncated kernel in the largest dimension $D = 512$).

Table 7.1: $E_K(Y, N, D)$ for each choice of periodic tensorial kernels when $N = 512$ in dimension $D = 1$

	random	numerical	asymptotics
exponential	0.035	0.002	0.002
multiquadric	0.036	0.000	0.000
Gaussian	0.036	0.000	0.000
truncated	0.036	0.002	0.002

Table 7.2: $E_K(Y, N, D)$ for each choice of periodic tensorial kernels when $N = 512$ in dimension $D = 16$

	random	numerical	asymptotics
exponential	0.054	0.034	0.034
multiquadric	0.055	0.021	0.021
Gaussian	0.055	0.018	0.018
truncated	0.056	0.036	0.038

Table 7.3: $E_K(Y, N, D)$ for each choice of periodic tensorial kernels when $N = 512$ in dimension $D = 128$

	random	numerical	asymptotics
exponential	0.059	0.049	0.043
multiquadric	0.059	0.037	0.037
Gaussian	0.059	0.037	0.037

truncated	0.058	0.042	0.057
-----------	-------	-------	-------

7.2 Periodic kernels

Strategy for periodic kernels. For each of the four periodic kernels in Table 1.2, we present three tables:

- (A) $E_K(Y, N, D)$ for a randomly chosen sequence Y .
- (B) $E_K(Y, N, D)$ for a numerical sequence that approximates the optimal solution \bar{Y} .
- (C) $\sqrt{\frac{1}{N} \sum_{n>N} \rho(\alpha^{*n})}$ which is our theoretical (but yet heuristic) asymptotic rate of convergence (6.13).

From a computational point of view, some remarks are in order:

- In item (A) above, the random sequences Y are computed using the standard Mersenne-Twister mt19937. The error bound $E_K(Y, N, D)$ can be computed using the left-hand formula in (7.3).
- In item (B) above, in order to compute the optimal Y we have two strategies:
 - We can either minimizing directly the left-hand side of (7.3). This optimization problem can be solved easily by a gradient descent and is computationally tractable. However, depending upon the choice of the kernel this method can show poor performances (i presence of almost vanishing gradients).
 - Or else, we can compute the first values α^{*n} for $n = 0, \dots, N - 1$ and then solve the equation (6.12). We recall here that $\sum_{m=1}^N e^{2i\pi \langle y^m, \alpha^{*n} \rangle} = 0$ for $1 \leq n \leq N$. We used this method and it always gave in principle good numerical results. However, to solve the relevant system we must use an algorithm whose complexity is of order N^3 , that leads to extremely long execution times as N becomes large.
 - Once the sequence Y is computed, the error $E_K(Y, N, D)$ is obtained using the left-hand side of (7.3).
- Finally, in item (C) above, we compute first the set of values $\chi(\alpha^{*n})$ for $n = 1, \dots, N$ by using a graph-search type algorithm. Observe that this step can be challenging in the case of a rather oscillatory function χ such as the truncated kernel.

For each periodic kernel we provide below the asymptotic formula derived from our level-set arguments, although it is not very accurate within the range of N, D under consideration. We plot in this section the distribution of sequences of points approximating the best discrepancy sequences \bar{Y} . These figures corresponds to $N = 256$ and $D = 2$. Observe that none of these distributions is radially-symmetric, as the corresponding periodic kernels are not.

Numerical results for the periodic tensorial exponential kernel. For the periodic kernel (4.1a) the level-set method (6.15) applies and lead us to the bound $E_K(N, D) \lesssim \frac{\log(N)^{D-1}}{N}$, which is probably far from being optimal but coincides with the Koksma-Hlawka inequality for low-discrepancy sequences. Here, the space $\mathcal{H}_K^{1,1}([0, 1]^D)$ (see (4.1d)) is similar to the space of functions with bounded variation $BV([0, 1]^D)$. The three relevant tables are given here. Observe that the discrepancy error for the numerical points is always much smaller than the discrepancy error for the random points, as was expected.

Note that the asymptotic convergence rate is quite close to the one obtained with the numerical sequence in Table 7.5. There exist cases for which this convergence rate is smaller than the theoretical bound. This might seem to be a contradiction, since the latter bound is supposed to be the minimum over all possible sequences. However, several sources of (yet small) numerical error arise due to the search algorithm of the decreasing sequence α^{*n} . In addition, a second (small) error arises from the estimate 6.14, which is probably sharp but yet only an approximation of the theoretical bound.

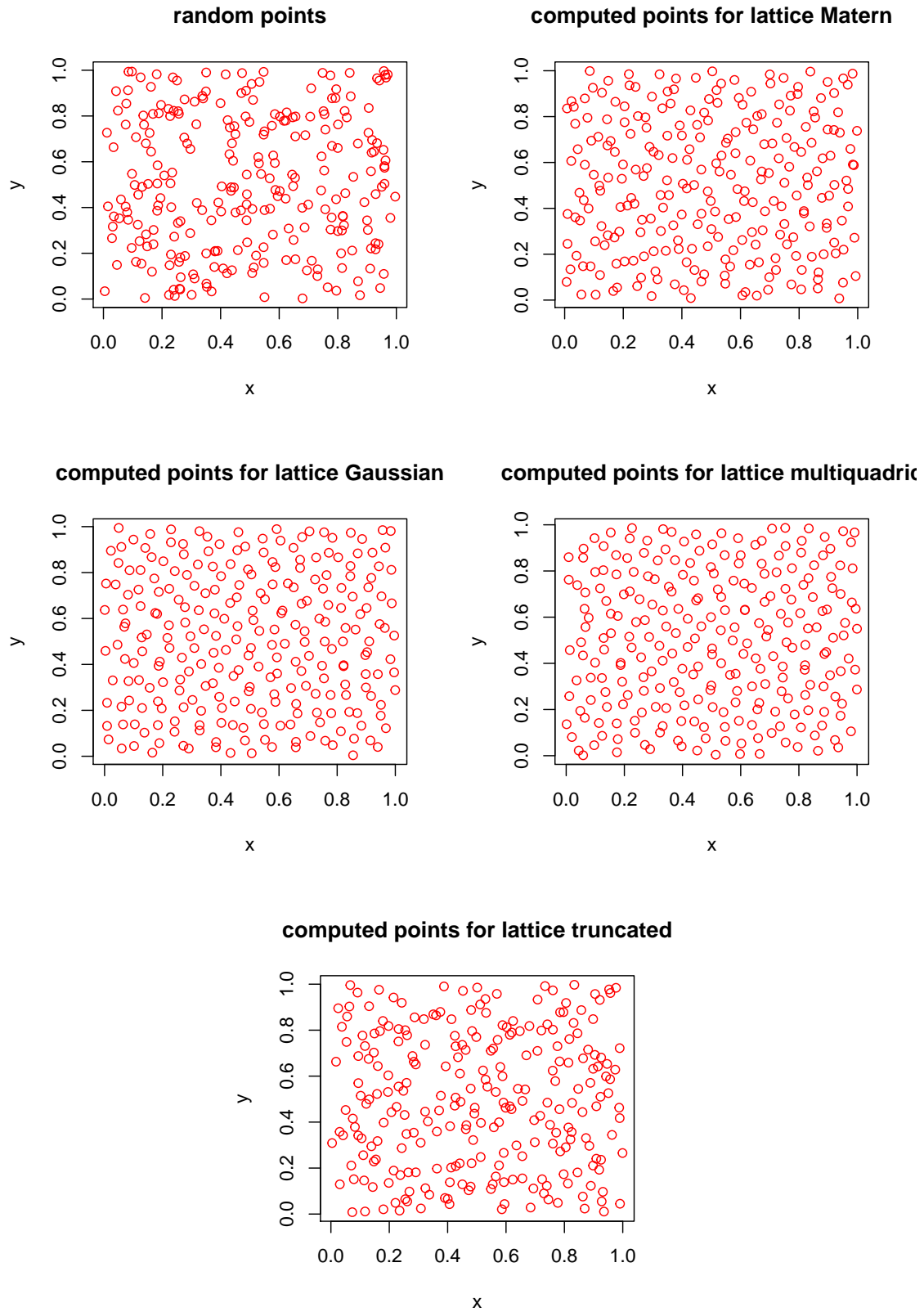


Figure 7.1: Random and numerical sequences for periodic kernels with $N = 256$

Table 7.4: $E_K(Y, N, D)$ for periodic tensorial exponential with random points Y

	D= 1	D= 2	D= 4	D= 8	D= 16	D= 32	D= 64	D= 128
N= 16	0.228	0.365	0.355	0.312	0.304	0.308	0.326	0.319
N= 32	0.307	0.222	0.216	0.199	0.210	0.228	0.231	0.234
N= 64	0.154	0.142	0.138	0.139	0.150	0.168	0.163	0.165
N= 128	0.117	0.088	0.087	0.097	0.111	0.116	0.114	0.115
N= 256	0.060	0.053	0.061	0.077	0.075	0.081	0.081	0.082
N= 512	0.035	0.038	0.050	0.049	0.054	0.057	0.057	0.059

Table 7.5: $E_K(Y, N, D)$ for periodic tensorial exponential with numerical points Y

	D= 1	D= 2	D= 4	D= 8	D= 16	D= 32	D= 64	D= 128
N= 16	0.062	0.126	0.172	0.195	0.211	0.217	0.221	0.223
N= 32	0.031	0.075	0.114	0.131	0.143	0.149	0.151	0.153
N= 64	0.016	0.049	0.076	0.090	0.099	0.103	0.106	0.107
N= 128	0.008	0.030	0.051	0.063	0.069	0.073	0.074	0.077
N= 256	0.004	0.020	0.034	0.043	0.048	0.051	0.054	0.061
N= 512	0.002	0.012	0.022	0.030	0.034	0.037	0.042	0.049

Table 7.6: $E_K(N, D)$ for periodic tensorial exponential with the asymptotic formula

	D= 1	D= 2	D= 4	D= 8	D= 16	D= 32	D= 64	D= 128
N= 16	0.069	0.143	0.202	0.245	0.288	0.308	0.318	0.323
N= 32	0.034	0.082	0.129	0.157	0.179	0.207	0.220	0.226
N= 64	0.017	0.046	0.078	0.102	0.116	0.129	0.147	0.156
N= 128	0.009	0.026	0.048	0.067	0.077	0.084	0.092	0.105
N= 256	0.004	0.014	0.029	0.042	0.052	0.056	0.060	0.066
N= 512	0.002	0.008	0.018	0.027	0.034	0.038	0.040	0.043

Numerical results for the periodic multiquadric kernel. For the periodic multiquadric kernel and using the level-set argument in (6.15) we can derive the bound $E_K(N, D) \lesssim \frac{2^D}{\tau^D} \left(1 - \exp\left(-\frac{N^{1/D}D}{2\tau D}\right)\right)$. The three tables are now presented and the same observations as above for the exponential kernel can be made. Note again the existence of a small numerical error for the two-dimensional case, as the error discrepancy for the optimized sequence is not decreasing, namely for $D = 2$ and $N = 256, 512$. Observe that the error vanishes for the one-dimensional case provides $N \geq 64$. This was expected since we are dealing with a kernel generating a space consisting of periodic function, whose Fourier coefficient decreases at exponential rate, as predicted by our theoretical convergence rate above. From a numerical point of view, it is expected to be a finite dimensional space, having few basis functions in low dimensions.

Table 7.7: $E_K(Y, N, D)$ for periodic tensorial multiquadric with random points Y

	D= 1	D= 2	D= 4	D= 8	D= 16	D= 32	D= 64	D= 128
N= 16	0.249	0.410	0.392	0.315	0.302	0.301	0.325	0.311
N= 32	0.349	0.245	0.230	0.196	0.201	0.227	0.226	0.234
N= 64	0.171	0.150	0.140	0.136	0.143	0.169	0.163	0.167
N= 128	0.130	0.090	0.086	0.093	0.108	0.117	0.115	0.114
N= 256	0.066	0.050	0.059	0.075	0.075	0.081	0.081	0.082
N= 512	0.036	0.036	0.049	0.048	0.055	0.057	0.057	0.059

Table 7.8: $E_K(Y, N, D)$ for periodic tensorial multiquadric with numerical points Y

	D= 1	D= 2	D= 4	D= 8	D= 16	D= 32	D= 64	D= 128
N= 16	0.002	0.078	0.172	0.204	0.261	0.277	0.313	0.306
N= 32	0.000	0.030	0.095	0.128	0.149	0.198	0.210	0.227
N= 64	0.000	0.005	0.045	0.081	0.103	0.106	0.140	0.157
N= 128	0.000	0.001	0.018	0.044	0.067	0.074	0.075	0.098
N= 256	0.000	0.002	0.007	0.024	0.042	0.051	0.052	0.053
N= 512	0.000	0.007	0.003	0.014	0.021	0.034	0.037	0.037

Table 7.9: $E_K(N, D)$ for periodic tensorial multiquadric with the asymptotic formula

	D= 1	D= 2	D= 4	D= 8	D= 16	D= 32	D= 64	D= 128
N= 16	0.004	0.081	0.171	0.207	0.272	0.301	0.314	0.321
N= 32	0.000	0.027	0.092	0.134	0.148	0.194	0.213	0.223
N= 64	0.000	0.005	0.044	0.085	0.100	0.105	0.137	0.151
N= 128	0.000	0.001	0.017	0.043	0.067	0.073	0.075	0.097
N= 256	0.000	0.000	0.008	0.025	0.043	0.050	0.052	0.053
N= 512	0.000	0.000	0.003	0.014	0.021	0.034	0.036	0.037

Numerical results for the periodic Gaussian kernel. For the periodic kernel (4.3) we can use our level-set arguments in (6.15) and we can arrive at the bound $E_K(N, D) \leq N^{1-1/D} \exp(-N^{2/D})$. We now present the three tables of interest and again the results are quite similar to the exponential and multiquadric cases.

Table 7.10: $E_K(Y, N, D)$ for periodic Gaussian with random points Y

	D= 1	D= 2	D= 4	D= 8	D= 16	D= 32	D= 64	D= 128
N= 16	0.279	0.445	0.400	0.316	0.301	0.301	0.325	0.311
N= 32	0.415	0.267	0.235	0.196	0.200	0.227	0.226	0.234
N= 64	0.205	0.159	0.138	0.136	0.142	0.168	0.163	0.167
N= 128	0.152	0.090	0.085	0.092	0.107	0.117	0.115	0.114
N= 256	0.075	0.047	0.059	0.074	0.075	0.081	0.081	0.082
N= 512	0.036	0.034	0.046	0.048	0.055	0.057	0.057	0.059

Table 7.11: $E_K(Y, N, D)$ for periodic Gaussian with numerical points Y

	D= 1	D= 2	D= 4	D= 8	D= 16	D= 32	D= 64	D= 128
N= 16	0	0.008	0.164	0.191	0.258	0.276	0.313	0.306
N= 32	0	0.000	0.051	0.123	0.145	0.197	0.209	0.227
N= 64	0	0.000	0.013	0.075	0.099	0.105	0.140	0.157
N= 128	0	0.000	0.003	0.033	0.064	0.072	0.075	0.098
N= 256	0	0.002	0.000	0.019	0.041	0.050	0.052	0.053
N= 512	0	0.008	0.000	0.008	0.018	0.033	0.036	0.037

Table 7.12: $E_K(N, D)$ for periodic Gaussian with the asymptotic formula

	D= 1	D= 2	D= 4	D= 8	D= 16	D= 32	D= 64	D= 128
N= 16	0	0.018	0.145	0.198	0.270	0.300	0.314	0.321

	D= 1	D= 2	D= 4	D= 8	D= 16	D= 32	D= 64	D= 128
N= 32	0	0.000	0.052	0.126	0.145	0.193	0.213	0.223
N= 64	0	0.000	0.012	0.077	0.097	0.104	0.137	0.151
N= 128	0	0.000	0.002	0.032	0.065	0.072	0.074	0.097
N= 256	0	0.000	0.000	0.020	0.041	0.050	0.052	0.053
N= 512	0	0.000	0.000	0.008	0.018	0.033	0.036	0.037

Numerical results for the periodic truncated kernel. For the periodic kernel (4.3) we can use our level-set method arguments and we arrive at the bound $E_K(N, D) \leq \frac{\log(N)^{D-1}}{N^2}$. We now present the three relevant tables as above and the observations we made concerning these results are quite similar.

Table 7.13: $E_K(Y, N, D)$ for periodic tensorial truncated with random points Y

	D= 1	D= 2	D= 4	D= 8	D= 16	D= 32	D= 64	D= 128
N= 16	0.279	0.398	0.354	0.327	0.315	0.317	0.324	0.328
N= 32	0.382	0.240	0.216	0.210	0.225	0.225	0.232	0.231
N= 64	0.189	0.148	0.142	0.149	0.159	0.163	0.163	0.163
N= 128	0.142	0.092	0.090	0.105	0.113	0.116	0.115	0.115
N= 256	0.071	0.051	0.063	0.081	0.079	0.082	0.082	0.082
N= 512	0.038	0.040	0.055	0.052	0.056	0.057	0.057	0.058

Table 7.14: $E_K(Y, N, D)$ for periodic tensorial truncated with numerical points Y

	D= 1	D= 2	D= 4	D= 8	D= 16	D= 32	D= 64	D= 128
N= 16	0.062	0.100	0.176	0.209	0.233	0.276	0.303	0.315
N= 32	0.031	0.058	0.116	0.140	0.156	0.173	0.196	0.214
N= 64	0.016	0.035	0.079	0.096	0.107	0.119	0.125	0.139
N= 128	0.007	0.021	0.049	0.067	0.075	0.081	0.085	0.090
N= 256	0.004	0.013	0.030	0.046	0.052	0.056	0.058	0.061
N= 512	0.002	0.010	0.021	0.032	0.036	0.039	0.040	0.042

Table 7.15: $E_K(N, D)$ for periodic tensorial truncated with the asymptotic formula

	D= 1	D= 2	D= 4	D= 8	D= 16	D= 32	D= 64	D= 128
N= 16	0.062	0.127	0.217	0.289	0.314	0.322	0.325	0.327
N= 32	0.031	0.077	0.133	0.188	0.218	0.227	0.230	0.231
N= 64	0.016	0.042	0.086	0.114	0.148	0.159	0.162	0.163
N= 128	0.007	0.023	0.054	0.073	0.096	0.110	0.114	0.115
N= 256	0.004	0.013	0.034	0.050	0.059	0.075	0.080	0.081
N= 512	0.002	0.007	0.022	0.034	0.038	0.049	0.055	0.057

7.3 Transported kernels

The numerical results we performed on transported kernel kerne are very similar to the one we just described for periodic kernels. We content with pointing out here significant differences between periodic and transported kernels:

- We did not numerically compute the theoretical convergence rates for the four transported kernels. Indeed, we would need to compute the Fourier coefficients appearing in Proposition 5.5. This appears to be too costly and computationally out of reach with the existing techniques.

- On the other hand, we did compute the error function thanks to its expression (7.2), evaluated via a direct Monte-Carlo method. This is still computationally expensive, and adds some additional white noise to the final results.

Thus, in this section, we restrict ourselves with presenting the plot in Figure 7.2 displaying the distribution of points that approximates the best discrepancy sequences \bar{Y} . This corresponds to the choice $N = 256$ and the dimension $D = 2$. Some observations are in order:

- Among our four transportation-based kernels, three are radially-symmetric, that is, Gaussian, multiquadric, and truncated. Their best discrepancy sequences are expected to be radially-symmetric. Indeed, our numerical tests confirm this property for the multiquadric and truncated kernels, which enjoy similar properties.
- However the Gaussian kernel did not lead us to a radially-symmetric result. This is due to a numerical challenge we can refer to as the “vanishing gradient problem” (a terminology used in the artificial intelligence community), due to the fact that the functional we want to minimize is almost flat. This problem did arise also with the multiquadric kernel, but we solved it by using the same trick discussed already in Section 7.2.
- The exponential kernel is not radially-symmetric so that its best discrepancy sequence is not expected to be so, and this is fully consistent with our results in Figure 7.2.

Acknowledgments. The first author (PLF) gratefully acknowledges support from the Simons Center for Geometry and Physics, Stony Brook University, as well as from the Innovative Training Network (ITN) grant 642768 (ModCompShocks). This paper was written during the Academic year 2018-2019 when he was visiting the Courant Institute for Mathematical Sciences, New York University.

References

- [1] I. BABUSKA, U. BANERJEE, AND J.E. OSBORN, Survey of mesh-less and generalized finite element methods: a unified approach, *Acta Numer.* 12 (2003), 1–125.
- [2] M.A. BESSA, AND J.T. FOSTER, T. BELYTSCHKO, AND W.K. LIU, A meshfree unification: reproducing kernel peridynamics, *Comput. Mech.* 53 (2014), 1251–1264.
- [3] H. COHN AND N. ELKIES, New upper bounds on sphere packings, *Ann. of Math.* 157 (2003), 689–714.
- [4] G.E. FASSHAUER, *Mesh-free methods*, in “Handbook of Theoretical and Computational Nanotechnology”, Vol. 2, 2006.
- [5] E.G. FASSHAUER, *Mesh-free approximation methods with Matlab*, Interdisciplinary Math. Sciences, Vol. 6, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2007.
- [6] F.C. GÜNTHER AND W.K. LIU, Implementation of boundary conditions for mesh-less methods, *Comput. Methods Appl. Mech. Engrg.* 163 (1998), 205–230.
- [7] J.J. KOESTER AND J.-S. CHEN, Conforming window functions for mesh-free methods, *Comm. Numer. Methods Engrg.* 347 (2019), 588–621.
- [8] P.G. LEFLOCH AND J.-M. MERCIER, Revisiting the method of characteristics via a convex hull algorithm, *J. Comput. Phys.* 298 (2015), 95–112.
- [9] P.G. LEFLOCH AND J.-M. MERCIER, A new method for solving Kolmogorov equations in mathematical finance, *C. R. Math. Acad. Sci. Paris* 355 (2017), 680–686.
- [10] P.G. LEFLOCH AND J.-M. MERCIER, The Transport-based Mesh-free Method (TMM). A short review, *The Wilmott journal* (2020). See ArXiv:1911.00992.
- [11] P.G. LEFLOCH AND J.-M. MERCIER, The Transport-based Mesh-free Method (TMM) and its applications in mathematical finance and fluid dynamics, in preparation.
- [12] S.F. LI AND W.K. LIU, *Mesh-free particle methods*, Springer Verlag, Berlin, 2004.
- [13] G.R. LIU, *Mesh-free methods: moving beyond the finite element method*, CRC Press, Boca Raton, FL, 2003.
- [14] G.R. LIU, An overview on mesh-free methods for computational solid mechanics, *Int. J. Comp. Methods* 13 (2016), 1630001.
- [15] M. MATSUMOTO AND T. NISHIMURA, Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator, *ACM Trans. Model. Comput. Simul.* 8 (1998), 3–30.

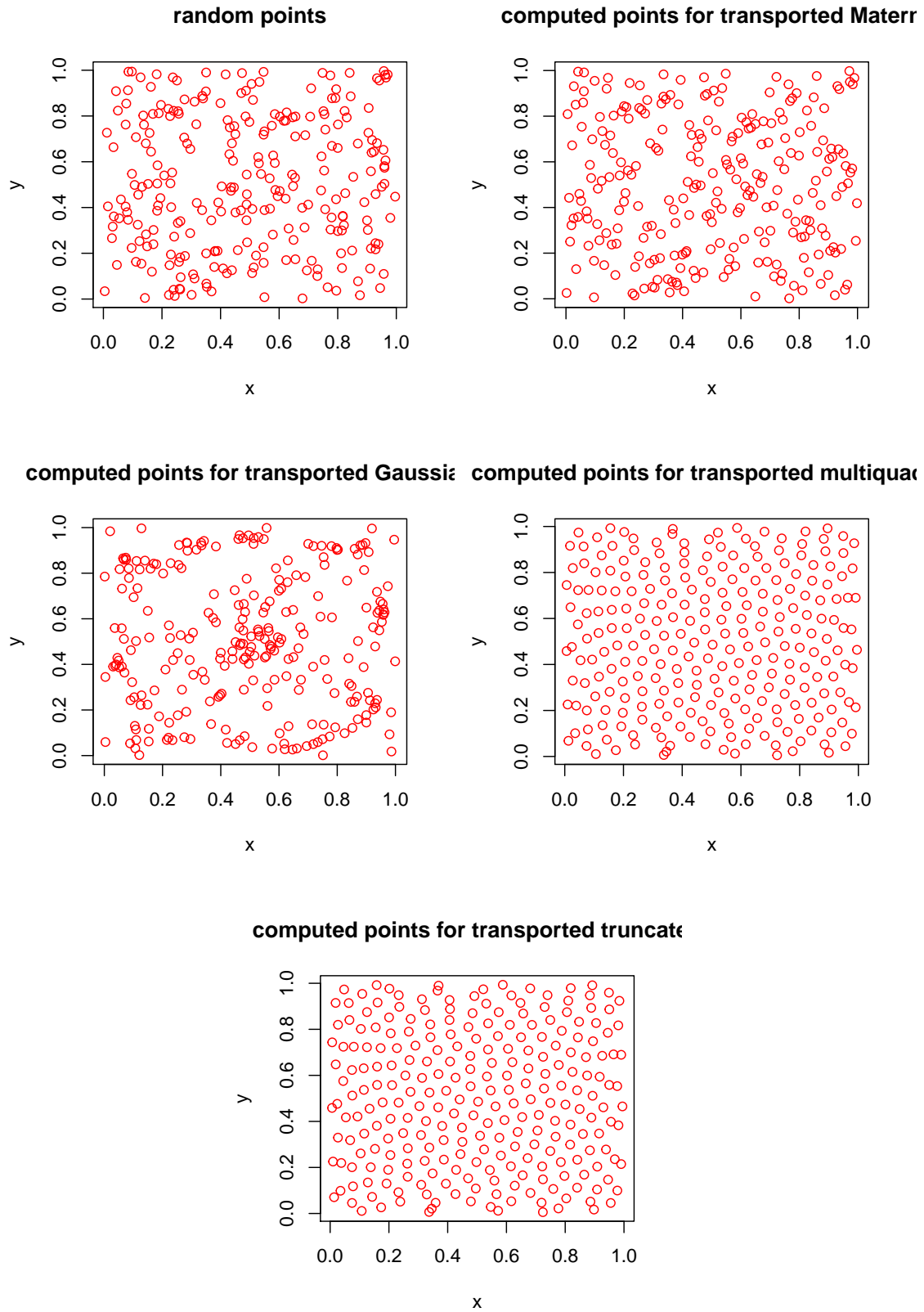


Figure 7.2: Random and numerical sequences for transported kernels with $N = 256$

- [16] J.-M. MERCIER AND S. MIRYUSUPOV, Hedging strategies for net interest income and economic values of equity, Sept. 2019, available at <https://ssrn.com/abstract=3454813>.
- [17] Y. NAKANO, Convergence of meshfree collocation methods for fully nonlinear parabolic equations, *Numer. Math.* 136 (2017), 703–723.
- [18] H.S. OH, C. DAVIS, AND J.W. JEONG, Mesh-free particle methods for thin plates, *Comput. Methods Appl. Mech. Engrg.* 209/212 (2012), 156–171.
- [19] R. OPFER, Multiscale kernels, *Adv. Computat. Math.* 25 (2006), 357–380.
- [20] R. SALEHI AND M. DEGHAN, A moving least square reproducing polynomial mesh-less method, *Appl. Numer. Math.* 69 (2013), 34–58.
- [21] M.S. VIAZOVSKA, The sphere packing problem in dimension 8, *Ann. of Math.* 185 (2017), 991–1015.
- [22] H. WENDLAND, Sobolev-type error estimates for interpolation by radial basis functions, in “Surface fitting and multiresolution methods” (Chamonix-Mont-Blanc, 1996), Vanderbilt Univ. Press, Nashville, TN, 1997, pp. 337–344.
- [23] H. WENDLAND, *Scattered data approximation*, Cambridge Monograph Appl. Compu. Math., Cambridge Univ., 2005.
- [24] J.X. ZHOU AND M.E. LI, Solving phase field equations using a mesh-less method, *Comm. Numer. Methods Engrg.* 22 (2006), 1109–1115.

Proof of Proposition 3.1. For any function $\varphi \in L^1_\mu(\mathbb{R}^D)$ we have $\iint_{\mathbb{R}^D \times \mathbb{R}^D} \varphi(x)\varphi(y)d\mu(x)d\mu(y) = \left(\int_{\mathbb{R}^D} \varphi d\mu\right)^2$, which is positive. Hence, being the product of two admissible kernels, we see that the kernel K^{loc} is admissible. The transported kernel K^{tran} is also admissible since for all relevant ψ

$$\iint_{\Omega \times \Omega} K^{\text{tran}}(y, y')\psi(y)\psi(y')dydy' = \iint_{\mathbb{R}^D \times \mathbb{R}^D} K^{\text{loc}}(x, x')(\psi \circ S^{-1})(x)(\psi \circ S^{-1})(x')d\mu(x)d\mu(x') \geq 0. \quad \square$$

Proof of Proposition 3.3. Since $\rho^{1/2} \in \ell^2(\mathbf{L}^*)$ it is clear that the expression

$$K^{\text{per}}(x, y) = \frac{1}{|\mathbf{C}|} \sum_{\alpha^* \in \mathbf{L}^*} e^{2i\pi \langle x-y, \alpha^* \rangle} \rho(\alpha^*)$$

is finite and, in fact, globally bounded on \mathbb{R}^D . It is symmetric since χ is even. The positivity condition follows from

$$\sum_{n, m} a_n a_m K^{\text{per}}(y^n, y^m) = \sum_{n, m} a_n a_m \langle e^{2i\pi \langle y^n - y^m, \cdot \rangle}, \rho \rangle_{\ell^2(\mathbf{L}^*)} = \sum_{\alpha^* \in \mathbf{L}^*} \left| \sum_n e^{2i\pi \langle y^n, \alpha^* \rangle} a_n \rho(\alpha^*) \right|^2 \geq 0. \quad \square$$

Proof of Proposition 5.1. In view of (2.4) and the identity $\varphi(x) = \langle \varphi, K(\cdot, x) \rangle_{\mathcal{H}_K(\Omega)}$, we find

$$\frac{1}{N} \sum_{1 \leq n \leq N} \varphi(y^n) = \frac{1}{N} \sum_{1 \leq n \leq N} \langle \varphi, K(\cdot, y^n) \rangle_{\mathcal{H}_K(\Omega)} = \langle \varphi, \frac{1}{N} \sum_{1 \leq n \leq N} K(\cdot, y^n) \rangle_{\mathcal{H}_K(\Omega)}.$$

In agreement with the statement of the proposition, we write

$$\int_{\Omega} \varphi(x)dx - \frac{1}{N} \sum_{1 \leq n \leq N} \varphi(y^n) = \langle \varphi, e_K(Y) \rangle_{\mathcal{H}_K(\Omega)}, \quad e_K(Y) = \int_{\Omega} \left(K(\cdot, x) - \frac{1}{N} \sum_{1 \leq n \leq N} K(\cdot, y^n) \right) dx,$$

and, by the Cauchy-Schwarz inequality, $|E_Y(\varphi)| \leq \|\varphi\|_{\mathcal{H}_K(\Omega)} \|e\|_{\mathcal{H}_K(\Omega)}$. Hence, the integration error splits into two contributions, as expected. By expanding the integrand using (2.4) we find the equivalent expression

$$\|e_K(Y)\|_{\mathcal{H}_K(\Omega)}^2 = \left\| \int_{\Omega} K(\cdot, x)dx \right\|_{\mathcal{H}_K(\Omega)}^2 + \frac{1}{N^2} \sum_{n, m=1}^N K(y^n, y^m) - \frac{2}{N} \sum_{1 \leq n \leq N} \int_{\Omega} K(x, y^n)dx$$

and, using (2.4), $\left\| \int_{\Omega} K(\cdot, x)dx \right\|_{\mathcal{H}_K(\Omega)}^2 = \int_{\Omega \times \Omega} K(x, y)dxdy$. Moreover, in view of $K(x, y) = \frac{1}{2}(K(x, x) + K(y, y) - \mathbf{D}(x, y))$, the derivation of (5.2a) is completed. \square

Proof of Proposition 5.3. The expression $K_N(x, x') = \sum_{1 \leq n \leq N} \lambda_n \zeta_n(x)\zeta_n(x')$ defines an admissible kernel, for which we have the following orthogonal decomposition: $\mathcal{H}_K(\Omega) = \mathcal{H}_{K_N}(\Omega) \oplus \mathcal{H}_{K_N}^\perp(\Omega)$. Consider the minimization problem

$$E_{K_N} = \inf_{y^1, \dots, y^N \in \Omega} \left(\sum_{1 \leq i \leq N} \lambda_i \left(\int_{\Omega} \zeta_i(x)dx - \frac{1}{N} \sum_{1 \leq n \leq N} \zeta_i(y^n) \right)^2 \right)$$

and

$$\sum_{1 \leq i \leq N} \lambda_i \left(\int_{\Omega} \zeta_i(x)dx - \frac{1}{N} \sum_{1 \leq n \leq N} \zeta_i(y^n) \right) (\nabla \zeta_i)(y^n) = 0, \quad n = 1, \dots, N. \quad \square$$

Proof of Proposition 5.4. We consider the Mercer representation and use the relations $T_K \zeta_i = \lambda_i \zeta_i$ and with $e_K = \sum_i \langle e_K, \zeta_i \rangle_{\ell^2(\Omega)} \zeta_i$. We have

$$\langle e_K, \zeta_i \rangle_{\ell^2(\Omega)} = \int_{\Omega} \int_{\Omega} \left(K(y, x) dx - \frac{1}{N} \sum_{1 \leq n \leq N} K(y, y^n) dx \right) \zeta_i(y) dy = \lambda_i \left(\int_{\Omega} \zeta_i(x) dx - \frac{1}{N} \sum_{1 \leq n \leq N} \zeta_i(y^n) \right)$$

and, by Hölder inequality, we obtain

$$\begin{aligned} \langle \varphi, e_K \rangle_{\mathcal{H}_K(\Omega)} &= \sum_{i=1,2,\dots} \lambda_i^{-1} \langle \varphi, \zeta_i \rangle_{\ell^2} \lambda_i \left(\int_{\Omega} \zeta_i(x) dx - \frac{1}{N} \sum_{1 \leq n \leq N} \zeta_i(y^n) \right) \\ &= \left\langle \left(\lambda_i^{-s/2} \langle \varphi, \zeta_i \rangle_{\ell^2} \right)_{i \geq 1}, \lambda_i^{s/2} \left(\int_{\Omega} \zeta_i(x) dx - \frac{1}{N} \sum_{1 \leq n \leq N} \zeta_i(y^n) \right)_{i \geq 0} \right\rangle_{\ell^p(\mathbb{N}), \ell^{p'}(\mathbb{N})} \\ &\leq \|\varphi\|_{\mathcal{H}_K^{s,p}(\Omega)} \left(\sum_i \lambda_i^{s p' / 2} \left(\int_{\Omega} \zeta_i(x) dx - \frac{1}{N} \sum_{1 \leq n \leq N} \zeta_i(y^n) \right)^{p'} \right)^{1/p'}. \quad \square \end{aligned}$$

Proof of Proposition 5.5 . Provided $\text{supp}(\varphi) \subset \mathbf{C}$, we have

$$\begin{aligned} \int_{\mathbf{C}} \varphi(x) dx - \frac{1}{N} \sum_n \varphi(y^n) &= \left\langle \int_{\mathbf{C}} e^{2i\pi \langle x, \cdot \rangle} dx - \frac{1}{N} \sum_n e^{2i\pi \langle y^n, \cdot \rangle}, \widehat{\varphi} \right\rangle_{\ell^2(\mathbf{L}^*)} := \langle \widehat{e}, \widehat{\varphi} \rangle_{\ell^2(\mathbf{L}^*)} \\ &= \left\langle \frac{\widehat{\varphi}}{\rho^{s/p}}, \rho^{s/p} \widehat{e} \right\rangle_{\ell^2(\mathbf{L}^*)}, \end{aligned}$$

where the error function \widehat{e} is defined as the Fourier transform of $e := 1_{\mathbf{C}} - \frac{1}{N} \sum_{1 \leq n \leq N} \delta_{y^n}$, hence

$$\left| \int_{\mathbf{C}} \varphi(x) dx - \frac{1}{N} \sum_n \varphi(y^n) \right| \leq \left\| \rho^{s/p} \left(\int_{\mathbf{C}} e^{2i\pi \langle x, \cdot \rangle} dx - \frac{1}{N} \sum_n e^{2i\pi \langle y^n, \cdot \rangle} \right) \right\|_{\ell^p(\mathbf{L}^*)} \|\varphi\|_{\mathcal{H}_K^{s,p}(\mathbf{C})}. \quad \square$$

Proof of Proposition 6.1. Let us set $e := \frac{1}{N} \sum_{1 \leq n \leq N} \int_{\Omega} \left(K(\cdot, x) dx - K(\cdot, y^n) \right) dx$. We have denoted by $e_Y := \sum_n e_Y(y^n) \delta_{y^n}^n$ its projection on the approximation space $\mathcal{H}_K^Y(\Omega)$. It satisfies

$$E_K(Y, N, D)^2 = \|e\|_{\mathcal{H}_K(\Omega)}^2 = \|e_Y\|_{\mathcal{H}_K^Y(\Omega)}^2 + \|e - e_Y\|_{\mathcal{H}_K}^2.$$

Thus $E_K(Y, N, D)^2 \geq \|e - e_Y\|_{\mathcal{H}_K}^2 > 0$, and, as we are going to check, a minimizer makes the term $\|e_Y\|_{\mathcal{H}_K^Y(\Omega)}^2$ to vanish.

In view of the definition (5.7) and the symmetry property $\partial_{y^n} K(y^n, y^m) = \partial_{y^m} K(y^m, y^n)$, we obtain ($n \in \{1, 2, \dots, N\}$, $d \in \{1, \dots, D\}$)

$$\partial_{y_d^n} (E_K(Y, N, D))^2 = \frac{2}{N^2} \sum_n \partial_{y_d^n} K(y^n, y^k) - \frac{2}{N} \int_{\Omega} \partial_{y_d^n} K(x, y^n) dx,$$

establishing (6.1). By computing second-order derivatives, we see that Hessian $\partial_{y_d^n} \partial_{y_e^n} (E_K(Y, N, D))^2$ reads as follows ($d, e \in \{1, \dots, D\}$):

$$\frac{2}{N^2} \sum_{m,n} \partial_{y_d^n} \partial_{y_e^n} K(y^n, y^m) - \frac{2}{N} \int_{\Omega} \partial_{y_d^n} \partial_{y_e^n} K(x, y^n) dx. \quad (.4)$$

Moreover, recalling (2.10b), we compute

$$e_Y = \sum_n e_Y(y^n) \delta_{y^n}^n = \left\langle \int_{\Omega} K(Y, Y)(x) dx, \delta_Y \right\rangle_{\mathbb{R}^N} - \frac{1}{N} \sum_n K(\cdot, y^n) = \left\langle \int_{\Omega} \delta_Y(x) dx, K_Y \right\rangle_{\mathbb{R}^N} - \frac{1}{N} \sum_n K(\cdot, y^n).$$

Hence, we arrive at

$$\begin{aligned} \|e_Y\|_{\mathcal{H}_K(\Omega)}^2 &= \left\langle \sum_n \left(\int_{\Omega} \delta^n(x) dx - \frac{1}{N} \right) \delta_{y^n}, \left\langle \int_{\Omega} \delta_Y(x) dx, K_Y \right\rangle_{\mathbb{R}^N} - \frac{1}{N} \sum_n K(\cdot, y^n) \right\rangle_{\mathcal{D}', \mathcal{D}} \\ &= \sum_{m,n} K(y^n, y^m) \left(\int_{\Omega} \delta^n(x) dx - \frac{1}{N} \right) \left(\int_{\Omega} \delta^m(x) dx - \frac{1}{N} \right) \geq C \left\| \int_{\Omega} \delta^n(x) dx - \frac{1}{N} \right\|_{\mathbb{R}^N}^2. \end{aligned}$$

In particular, the minimum vanishes when $\int_{\Omega} \delta^n(x) dx = \frac{1}{N}$, and this establishes (6.2). Hence, suppose that this relation is true, then from (2.10b) we deduce

$$\frac{1}{N} = \int_{\Omega} \delta_Y(x) dx = K(Y, Y)^{-1} \int_{\Omega} K(Y, Y)(x) dx,$$

implying $K(Y, Y) \frac{1}{N} = \int_{\Omega} K(Y, Y)(x) dx$ or, equivalently, $K(Y, Y)^m$ is a stochastic matrix, which is (6.2). \square

Proof of Proposition 6.2. We begin with the case $s =$ and $p = 2$, and consider an arbitrary (ordered) sequence of points, say $0 < y^1 < \dots < y^N < 1$. We rewrite our functional (5.2a) as

$$\begin{aligned} (E_K(Y, N, D))^2 &= \int_{([0,1])^2} K(x, y) dx dy - \frac{1}{N} \sum_{1 \leq n \leq N} (y^n - 1) y^n + \frac{1}{N^2} \sum_{n > m, n=1, \dots, N} y^m (y^n - 1) \\ &\quad + \frac{1}{N^2} \sum_{n \leq m, n=1, \dots, N} (y^m - 1) y^n \\ &= 1/12 - \frac{1}{N} \sum_{1 \leq n \leq N} (y^n - 1) y^n + \frac{1}{N^2} \left(\sum_{1 \leq n \leq N} y^n \right)^2 - \frac{1}{N^2} \sum_{n, m=1}^N y^n (1_{m > n} + 1_{n \leq m}) \end{aligned}$$

or, equivalently,

$$(E_K(Y, N, D))^2 = 1/12 - \frac{1}{N} \sum_{1 \leq n \leq N} (y^n)^2 + \frac{1}{N^2} \left(\sum_{1 \leq n \leq N} y^n \right)^2 + \frac{1}{N^2} \sum_{1 \leq n \leq N} y^n (2n - N - 1).$$

In view of Proposition 6.1, any minimizer satisfies the system of equations ($n = 1, 2, \dots, N$)

$$\partial_{y^n} (E_K(Y, N, D))^2 = -\frac{2}{N} y^n + \frac{2}{N^2} \sum_{1 \leq n \leq N} y^n + \frac{1}{N^2} (2n - N - 1) = 0,$$

which we put in the form $\frac{1}{N} \sum_{1 \leq n \leq N} y^n - \frac{1}{2} + \frac{1}{N} (n - 1/2) = y^n$. This is a linear algebraic system, which is readily solved explicitly; this leads us to (6.4). We check immediately that $\frac{1}{N} \sum_{1 \leq n \leq N} y^n = \frac{1}{2}$ and $\sum_{1 \leq n \leq N} (y^n)^2 = \frac{N}{3} - \frac{1}{12N}$, and, finally, we evaluate the functional to be

$$(E_K(Y, D, N))^2 = \frac{1}{12} + \left(\frac{1}{3} - \frac{1}{12N^2} \right) - \frac{1}{4} - \frac{2}{N^2} \sum_{1 \leq n \leq N} \frac{(n - 1/2)^2}{N} + \frac{2}{N^2} \sum_{1 \leq n \leq N} \frac{n - 1/2}{2} = \frac{1}{6N^2}. \quad \square$$