Matrix-valued kernels
for shape deformation analysis[*]

Mario Micheli and Joan A. Glaunès

Dedicated to David Mumford on the occasion of his 75th birthday

The main purpose of this paper is providing a systematic study and classification of non-scalar positive definite kernels for Reproducing Kernel Hilbert Spaces (RKHS), to be used in the analysis of deformations in shape spaces endowed with metrics induced by the action of groups of diffeomorphisms. After providing an introduction to matrix-valued kernels and their relevant differential properties, we explore extensively those, that we call TRI kernels, that induce a metric on the corresponding Hilbert spaces of vector fields that is both translation- and rotation-invariant. These are analyzed in a very effective manner in the Fourier domain, where the characterization of RKHS of curl-free and divergence-free vector fields is particularly natural. A technique for constructing generic matrix-valued kernels from scalar kernels is also developed. The exposition of the theory is supported by several examples. We also provide numerical results that show the dynamics induced by different choices of TRI kernels on the manifold of labeled landmark points, and illustrate one application in computational anatomy.

1. Introduction

Recent years have seen the rapid development of acquisition techniques for medical data, such as magnetic resonance imaging, that allow the identification and very detailed visualization of anatomical structures *in vivo* and the quantitative description of their shape. The key challenge for researchers is to devise statistical methods that assess normal and abnormal variations of such shapes across subjects, with the goal of gaining understanding of several pathologies and ultimately providing practitioners with shape-based diagnostic tools. The problems of comparing geometrical objects of the same nature, establishing correspondences that are anatomically justified, and

quantifying deformation (i.e. computing *distances* between shapes, so to be able to perform, for example, shape classification) are central in addressing the above issues. The emerging discipline of computational anatomy thus lies at the interface of geometry, imaging science, statistics, and numerical analysis.

The group action approach. It was Ulf Grenander [31] that introduced the notion of group action in this field. The main idea is to compare anatomical objects through the estimation and the analysis of a deformation of their entire space where such objects are located. The theoretical model thus consists in defining (i) the mathematical space of the observed objects (e.g. the images, surfaces, curves, point sets—in one word, the *shapes*), (ii) an appropriate group of transformations, and (iii) a group action on the objects. This approach stems from the observation that the visual comparison of medical images coming from different individuals suggests the existence of an underlying deformation of the ambient space.

The formulation of precise registration algorithms requires rigorous mathematical modeling of the deformations of such space. Rigid or affine transformations alone are insufficient to describe the complexity of the observed deformations, and it is in fact appropriate to consider functional spaces of infinite dimension. In the diffeomorphic approach introduced in [18, 22] the transformations are obtained by the temporal integration of a family of vector fields—this was partly inspired by Vladimir Arnold's seminal paper [2], in which he proved that incompressible fluid dynamics can be characterized as geodesic flow in the group of volume preserving diffeomorphisms, with respect to the kinetic energy metric (i.e. the L^2 norm of the fluid velocity). A rigorous mathematical foundation for this new model and the construction of invariant metrics for the groups of diffeomorphisms was established by Alain Trouvé [65]: in this approach one chooses a Hilbert space V of vector fields whose norm $\|\cdot\|_V$ defines the cost of infinitesimal deformations, and time-dependent diffeomorphisms $\varphi_t : \Omega \to \Omega$, with $\Omega \subseteq \mathbb{R}^d$, are obtained by the integration of a family of vector fields $v(t, \cdot) : \Omega \to \mathbb{R}^d$ in V, $t \in [0, 1]$, via the differential equation $\partial_t \varphi_t(x) = v(t, \varphi_t(x))$, with $\varphi_0(x) = x$. Given two shapes (geometric objects or images), among all diffeomorphisms of such type that perform the registration, the one that is generated by a time-dependent vector field that minimizes the kinetic energy $\int_0^1 \|v(t, \cdot)\|_V^2 dt$ is chosen, and the square root of the minimal energy is in fact a *distance* between the two shapes.

There are several advantages to this group action-based approach. First of all, it fits into the shape space idea which views shapes as points in a infinite dimensional curved manifold, and the estimated deformations, when restricted to the shapes themselves, correspond to geodesics in this shape space. Therefore the method generates smooth transitions between shapes

which are in fact all plausible shapes of the considered population, at least from a geometrical point of view. This also enables one to employ mathematical notions and algorithms provided by Riemannian geometry [36, 37], such as Fréchet means or medians [25] or parallel translation [68]. Secondly, the generated diffeomorphisms can be applied to any geometrical structure in the ambient space. Whence one may estimate a deformation induced by a particular anatomical structure, for instance the cortical surface of the brain, and then apply this deformation to inner brain structures. Thirdly, the group action-based approach gives a unified framework for shape analysis based on various geometrical objects: deformations estimated from the raw images, or from segmented anatomical structures such as landmarks points or surfaces, all belong to the same space and can be compared; moreover, deformations may be estimated from combinations of such different structures. This is of course of primary importance for statistical purposes, as it allows one to detach the deformation analysis from the particular modality which was used to estimate the deformations.

This framework, often referred to as Large Deformation Diffeomorphic Metric Mapping (LDDMM), has received considerable interest and is now popular in the fields of computational anatomy, morphometry, and shape analysis. The method may be applied for performing registration and computing distances in different kinds of "shape spaces", such as the manifolds curves [29, 45], surfaces [71], images [6, 46, 47], vector fields [15], diffusion tensor images [14], measures [27], currents [28, 66], and labeled landmark points [30, 35, 42, 47]. We should note that in this context the formulation of the registration problem requires the vector fields to be L^1 or L^2 in the time parameter t; these and other notions, that deviate from the classical theory of dynamical systems, have been expanded in a book by L. Younes [69].

David Mumford, to whom this work is dedicated, and his collaborators have given considerable contributions to the understanding of the differential geometry of shape spaces endowed with a Riemannian metric, including several instances of metrics induced by the group action of diffeomorphisms. The knowledge of such geometry (and in particular, of curvature) is fundamental as it allows one to infer, for example, about the uniqueness of geodesics between shapes, the existence of conjugate points, and the well-posedness of the problem of computing the intrinsic (or Karcher's) mean [54] of a database of shapes. See, for example, the papers [42, 43, 44, 45, 48, 70] and references therein.

Reproducing Kernel Hilbert Spaces. In LDDMM the Hilbert space of vector fields V has a reproducing kernel, whose choice uniquely determines the model in use and the properties of the deformation maps; moreover, from a computational point of view, the most demanding operations

in registration algorithms usually consist in convolutions of data points with these kernels or their derivatives. For example, when dealing with the shape manifold $\mathcal{L}^N(\mathbb{R}^d)$ of N labeled landmarks in $\mathbb{R}^d$ (typically, $d = 2$ or 3), the registration constraints are described in terms of the displacements of such points. The mathematical object of interest is the reproducing kernel $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{d \times d}$ of the space V of deformation fields. One can in fact show that the vector fields that minimize the kinetic energy for such registration constraints are sums of spline functions centered on the landmark trajectories $x_a(t)$, $a = 1, \ldots, N$, $t \in [0, 1]$, i.e.

$$(1) \qquad v(t, x) = \sum_{a=1}^{N} K\big(x, x_a(t)\big)\alpha_a(t), \qquad x \in \mathbb{R}^d,\ t \in [0, 1],$$

and that the associated cost of the infinitesimal deformation for vector fields of this type (i.e. their norm in the Hilbert space V) is also expressed in terms of the reproducing kernel, as follows:

$$\|v(t, \cdot)\|_V^2 = \sum_{a,b=1}^{N} \alpha_a(t) \cdot K\big(x_a(t), x_b(t)\big)\alpha_b(t), \qquad t \in [0, 1];$$

the vectors $\alpha_a : [0, 1] \to \mathbb{R}^d$, $a = 1, \ldots, N$, are called *momenta*, in analogy with analytical mechanics [3], and they completely parametrize the search space of an optimal (i.e. energy-minimizing) solution.

In this framework it is therefore natural to consider *the kernel* as the starting point for modeling the linear deformation space V and the group of diffeomorphisms that it generates. The theory of Reproducing Kernel Hilbert Spaces (RKHS) was developed starting in the 1940s, mostly by Aronszajn [4, 5] and Schwartz [62], who built on previous studies by Bergman [7, 8], Bochner [10], Schoenberg [59, 60], and others. Such a theory is mostly used in complex and functional analysis, and in more recent years it has found numerous applications and developments in statistics and machine learning [51, 61]. Reproducing kernels are also used as spline functions for performing data interpolation [67]; their interest in this framework stems from the fact that functions expressed in terms of kernels solve the minimal norm interpolation problem in the corresponding Hilbert space. Kernels are also commonly used to interpolate data with values in the Euclidean space $\mathbb{R}^d$ (mostly with $d = 2$ or 3), such as vector fields in fluid dynamics (see [11] and references therein).

The kernels that are most commonly used in applications are either scalar-valued functions or (in the case of RKHS of $\mathbb{R}^d$-valued functions) they are given by a scalar-valued function multiplied by the $d \times d$ identity

matrix $\mathbb{I}_d$. However, employing kernels that are truly non-scalar allows one to obtain desirable properties of the vector fields that are not achievable otherwise. To our knowledge a systematic study and classification of such kernels have not yet taken place. Here we focus our attention on kernels that induce translation- and rotation-invariant norms in the corresponding RKHS, since this a common requirement for the interpolation of geometric data. We shall examine in detail the ties between the properties of the kernels and the corresponding deformation spaces, with the goal of providing a large class of kernels that may be used to perform shape analysis. In particular, matrix-valued kernels that induce divergence-free vector fields are desirable in analyzing volume-preserving transformations, and those that induce divergence-free vector fields can be used for deformations that are caused by the pure growth or loss of mass. However, curl-free and divergence-free vector fields are only achievable with non-scalar reproducing kernels.

Additional related work. As we said, it was Laurent Schwartz in his seminal work [62] that built the foundations of the theory of (non-scalar) reproducing kernels. Much more recently, Carmeli *et al.* [16] derived regularity results of vector-valued RKHS and analyzed some of their properties from the point of view of the machine learning community. De Vito *et al.* [19] extend Mercer's theorem to matrix-valued measurable kernels, whereas Micchelli and Pontil [41] use such kernels for learning vector-valued functions. Cachier and Ayache [13] consider a class of matrix-valued kernels that generalize thin-plate splines for the interpolation of dense and sparse vector fields, for the purposes of image registration. The work by Dodu and Rabut [21] is related to ours in that it introduces a class of irrotational and divergence-free kernels that minimize Beppo Levi-type seminorms, for the interpolation of vector fields in two and three dimensions; our study is restricted to the case of positive definite kernels but is more general in that it characterizes the *entire* class of kernels that induce translation- and rotation-invariant norms. Last, but not least, we should certainly mention a very recent paper by Mumford and Michor [50], where they study an approximation to Euler's equation with EPDiff, i.e. the geodesic equation in the group of diffeomorphisms [49], by choosing kernels that are Green's functions of a specific class of differential operators. The latter depend on two positive parameters, and their limit behavior yields precisely Euler's equation for fluid flow; such kernels, as well as their limit cases, fall within our study.

Paper organization. In section 2 we introduce notation and definitions for Reproducing Kernel Hilbert Spaces of vector-valued functions, state and prove existence and uniqueness theorems for such spaces, and investigate relevant differential properties of the corresponding matrix-valued kernels. Section 3 explores the characterization of what we call *TRI kernels*, i.e. those that induce a translation- and rotation-invariant metric on the corresponding

RKHS. This is also performed in the Fourier domain, and constitutes the core of the present contribution: in fact, it turns out that the characterization of kernels Hilbert spaces of curl-free or divergence-free vector fields is described in a more natural manner in the Fourier domain than it is in the spatial domain. Section 4 details methods for building TRI kernels from the more commonly used scalar kernels, which is of course crucial for applications; in fact we also show that all matrix-valued kernels may be obtained with such constructive procedures. Section 5 develops the equations of the dynamics induced by the LDDMM approach on the manifold of landmark points, such as equation (1) introduced eariler. In section 6 we illustrate numerical results for both landmarks and anatomical surfaces, comparing results for different choices of TRI kernels. We summarize conclusions and describe directions of potential future research developments of our research in section 7.

We would like to point out that the length of our paper is justified by our desire to make it self-contained, exhaustive, and accessible to the largest possible audience—in the hope to attract more researchers (mathematicians, engineers, statisticians, computer scientists) to this interdisciplinary field.

2. Reproducing Kernel Hilbert Spaces of vector-valued functions

The theory of Reproducing Kernel Hilbert Spaces of scalar-valued functions [4, 5, 67] may be extended to vector-valued functions [16, 32, 62] and has recently been adapted to the study of deformation [69]. Here we provide notions and extensions of results that are relevant precisely for this purpose.

Notation. If S is a subset of a vector space then $\mathrm{span}(S)$ denotes the set of finite linear combinations of elements of S. If $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$ are two normed spaces with X continuously embedded in Y (i.e. $X \subseteq Y$ and there there is a constant $C > 0$ such that $\|x\|_Y \leq C\|x\|_X$ for all $x \in X$) we write $X \hookrightarrow Y$. If $(H, \langle \cdot, \cdot \rangle_H)$ is an inner product space and U is a subspace of H we indicate its orthogonal complement with $U^\perp$. If H is a Hilbert space its dual space is written as H^*; if a sequence $\{u_n\}$ in H converges weakly to some $u \in H$, i.e. $\langle u_n - u, h \rangle_H \to 0$ for all $h \in H$, we write $u_n \rightharpoonup u$ in H. The dot product of two vectors $a, b \in \mathbb{R}^d$ is denoted with $\langle a, b \rangle_{\mathbb{R}^d}$ or $a \cdot b$. Vectors in $\mathbb{R}^d$ are normally treated as column vectors, and ^{T} indicates the transpose of a vector or a matrix.

Definition 2.1. *Let $(H, \langle \cdot, \cdot \rangle_H)$ be a Hilbert space of $\mathbb{R}^d$-valued functions defined on a set Ω. We call H a* Reproducing Kernel Hilbert Space (RKHS) *if the evaluation functionals $\delta_x^\alpha : u \mapsto \alpha \cdot u(x)$ are linear and continuous for all $x \in \Omega$ and $\alpha \in \mathbb{R}^d$; that is, if $\delta_x^\alpha \in H^*$.*

By the Riesz Representation Theorem [26], in a Reproducing Kernel Hilbert Space H, for all $x \in \Omega$ and $\alpha \in \mathbb{R}^d$ there exists a unique function $K_x^\alpha(\cdot) \in H$ such that

$$\left\langle K_x^\alpha, u \right\rangle_H = \alpha \cdot u(x) \tag{2}$$

for all $u \in H$. Such function is called the *representer* of the evaluation functional δ_x^α, and relation (2) is referred to as the *reproducing property* of K_x^α. The map $\alpha \mapsto K_x^\alpha(\cdot)$ is linear in α, i.e. $K_x^{a\alpha+b\beta} = aK_x^\alpha + bK_x^\beta$ for all $x \in \Omega$, $\alpha, \beta \in \mathbb{R}^d$, and $a, b \in \mathbb{R}$, by the uniqueness of the representer. So for any pair of points $x, y \in \Omega$ there exists a matrix $K(y, x) \in \mathbb{R}^{d \times d}$ such that $K_x^\alpha(y) = K(y, x)\alpha$ for all $\alpha \in \mathbb{R}^d$; the matrix-valued function $K : \Omega \times \Omega \to \mathbb{R}^{d \times d}$ is called the *reproducing kernel*, or simply the *kernel*, of the space H.

For any pair of points $x, y \in \Omega$ and any pair of vectors $\alpha, \beta \in \mathbb{R}^d$ we have that, by the reproducing property (2), $\langle K_x^\alpha, K_y^\beta \rangle_H = \alpha \cdot K_y^\beta(x) = \alpha \cdot K(x, y)\beta$ but also $\langle K_y^\beta, K_x^\alpha \rangle_H = \beta \cdot K_x^\alpha(y) = \beta \cdot K(y, x)\alpha = \alpha \cdot K(y, x)^T \beta$, so that the symmetry $K(x, y) = K(y, x)^T$ holds for all $x, y \in \Omega$.

Definition 2.2. *A Reproducing Kernel Hilbert Space with kernel K is called* non-degenerate *when it has the following property: for any $N \in \mathbb{N}$ and any distinct points $x_1, \ldots, x_N \in \Omega$, if the vectors $\alpha_1, \ldots, \alpha_N \in \mathbb{R}^d$ are such that $\sum_{a=1}^N \alpha_a \cdot u(x_a) = 0$ for all $u \in H$, then $\alpha_1 = \cdots = \alpha_N = 0$.*

Non-degeneracy establishes a certain richness of functions in Reproducing Kernel Hilbert Spaces. Rather obviously, we will call a RKHS *degenerate* if it is not non-degenarate.

Definition 2.3. *A generic matrix-valued function $K : \Omega \times \Omega \to \mathbb{R}^{d \times d}$ is* positive definite *if for abitrary $N \in \mathbb{N}$, vectors $\alpha_1, \ldots, \alpha_N \in \mathbb{R}^d$ and points $x_1, \ldots, x_N \in \Omega$, it is the case that*

$$\sum_{a,b=1}^N \alpha_a \cdot K(x_a, x_b)\alpha_b \geq 0. \tag{3}$$

Moreover, K is strictly *positive definite if for abitrary $N \in \mathbb{N}$, vectors $\alpha_1, \ldots, \alpha_N \in \mathbb{R}^d$, and distinct points $x_1, \ldots, x_N \in \Omega$ the above inequality holds, with equality if and only if $\alpha_1 = \cdots = \alpha_N = 0$.*

Proposition 2.4. *The kernel K of a RKHS is a positive definite matrix-valued function. Moreover, the RKHS is non-degenerate if and only if its kernel is* strictly *positive definite.*

Proof. By the reproducing property (2) of the representer function we have

$$0 \leq \big\| \textstyle\sum_{i=a}^{N} K_{x_a}^{\alpha_a} \big\|_H^2 = \big\langle \textstyle\sum_{a=1}^{N} K_{x_a}^{\alpha_a}, \sum_{b=1}^{M} K_{x_b}^{\alpha_b} \big\rangle_H$$
$$= \textstyle\sum_{a,b=1}^{M} \big\langle K_{x_a}^{\alpha_a}, K_{x_b}^{\alpha_b} \big\rangle_H = \sum_{a,b=1}^{M} \alpha_a \cdot K(x_a, x_b) \alpha_b.$$

Assuming that the RKHS is non-degenerate, if for $x_1, \ldots, x_N \in \Omega$ (distinct), $\alpha_1, \ldots, \alpha_N \in \mathbb{R}^d$ we have $\| \sum_a K_{x_a}^{\alpha_a} \|_H^2 = \sum_{a,b} \alpha_a \cdot K(x_a, x_b) \alpha_b = 0$, then $\sum_{a=1}^{N} K_{x_a}^{\alpha_a} = 0$, i.e. $\langle \sum_{a=1}^{N} K_{x_a}^{\alpha_a}, u \rangle_H = 0$ for all $u \in H$; this is equivalent to $\sum_a \alpha_a \cdot u(x_a) = 0$ for all $u \in H$, therefore $\alpha_1 = \cdots = \alpha_N = 0$ by Definition 2.2. On the other hand, assuming that the kernel K is strictly positive definite, if at least one of the vectors $\alpha_1, \ldots, \alpha_N \in \mathbb{R}^d$ is non-zero take $u := \sum_a K_{x_a}^{\alpha_a}$, with distinct points $x_1, \ldots, x_N \in \Omega$; it is the case that $\sum_a \alpha_a \cdot u(x_a) = \sum_{a,b} \alpha_a \cdot K(x_a, x_b) \alpha_b > 0$ by strict positive definiteness. □

Example. All *finite*-dimensional Reproducing Kernel Hilbert Spaces are degenerate. To see this, consider a finite-dimensional space H of $\mathbb{R}^d$-valued functions with a given inner product $\langle \cdot, \cdot \rangle_H$. We may assume without loss of generality that $H = \text{span}\{f_1, \ldots, f_M\}$ with $\langle f_i, f_j \rangle_H = \delta_{ij}$, where δ_{ij} is Kronecker's delta. The reproducing kernel of H is the matrix-valued function $K(x, y) = \sum_{i=1,\ldots,M} f_i(x) f_i(y)^T$. Now fix arbitrarily $N \in \mathbb{N}$ and the distinct points $x_1, \ldots, x_N \in \Omega$. Define:

$$\boldsymbol{\alpha} := \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} \in \mathbb{R}^{Nd} \quad \text{and} \quad f_i(\boldsymbol{x}) := \begin{bmatrix} f_i(x_1) \\ \vdots \\ f_i(x_N) \end{bmatrix} \in \mathbb{R}^{Nd}, \quad i = 1, \ldots, M,$$

all intended as column vectors. Having $\sum_{i=1}^{N} \alpha_i \cdot u(x_i) = 0$ for all $u \in H$ is equivalent to $\boldsymbol{\alpha}$ being orthogonal to $U := \text{span}\{f_1(\boldsymbol{x}), \ldots, f_M(\boldsymbol{x})\}$, i.e. we have $\boldsymbol{\alpha} \in U^\perp$. But $\dim U \leq M$, therefore $\dim U^\perp \geq \max\{Nd - M, 0\}$. So, if N was chosen large enough, $\boldsymbol{\alpha}$ is not necessarily zero.

We have seen above that Reproducing Kernel Hilbert Spaces have a unique reproducing kernel. The next definition of "kernel" is justified by the theorem that immediately follows.

Definition 2.5. *A positive definite function $K : \Omega \times \Omega \to \mathbb{R}^{d \times d}$ such that $K(y, x) = K(x, y)^T$ for all $x, y \in \Omega$ is called a* positive kernel, *or simply a* kernel, *of dimension d. If, in addition, K is a strictly positive definite function, then we shall call it a* strictly positive kernel *of dimension d.*

Theorem 2.6. *Given a positive kernel $K : \Omega \times \Omega \to \mathbb{R}^{d \times d}$ of dimension d, there exists a* unique *RKHS of $\mathbb{R}^d$-valued functions defined on Ω that has K as its reproducing kernel.*

Proof. For arbitrary $x \in \Omega$ and $\alpha \in \mathbb{R}^d$ we define the function $K_x^\alpha(\cdot) := K(\cdot, x)\alpha$ and the linear space $H_0 := \operatorname{span}\{K_x^\alpha \mid x \in \Omega, \alpha \in \mathbb{R}^d\}$. We also define the inner product $\langle K_x^\alpha, K_y^\beta \rangle_{H_0} := \alpha \cdot K(x, y)\beta$. In order to extend it to all of H_0 by bilinearity it is sufficient to show that if

$$ f = \sum_{i=1}^{n} K_{x_i}^{\alpha_i} = \sum_{i=1}^{n'} K_{x_i'}^{\alpha_i'} $$

(i.e. there are two different representations of $f \in H_0$) then for all $g = K_y^\beta \in H_0$ we have $\langle f, g \rangle_{H_0} := \sum_{i=1}^{n} \langle K_{x_i}^{\alpha_i}, K_y^\beta \rangle_{H_0} = \sum_{i=1}^{n'} \langle K_{x_i'}^{\alpha_i'}, K_y^\beta \rangle_{H_0}$. To see this take $m := n + n'$, $\gamma_i := \alpha_i$ for $1 \le i \le n$ and $\gamma_i := -\alpha_i'$ for $n + 1 \le i \le n + n'$; now observe that since $\sum_{i=1}^{m} K_{z_i}^{\gamma_i} = 0$ we also have $\sum_{j=1}^{m} \langle K_{z_j}^{\gamma_j}, K_y^\beta \rangle_{H_0} = \sum_{j=1}^{m} \gamma_j \cdot K(z_j, y)\beta = (\sum_{j=1}^{m} K_{z_j}^{\gamma_j}(y))\beta = 0$.

It is immediate to verify that the inner product $\langle \cdot, \cdot \rangle_{H_0}$ is symmetric; also, if $u \in H_0$ then it is the case that $\langle u, u \rangle_{H_0} \ge 0$ by the positive definiteness of K. By construction $\langle K_x^\alpha, u \rangle_{H_0} = \alpha \cdot u(x)$ for all $u \in H_0$, so by Cauchy-Schwarz we have $|\alpha \cdot u(x)| \le \|K_x^\alpha\|_{H_0} \|u\|_{H_0}$ for all $u \in H_0$, $x \in \Omega$ and $\alpha \in \mathbb{R}^d$; whence if $\langle u, u \rangle_{H_0} = 0$ then $u = 0$. Whence H_0 is an inner product space.

Now one can follow the standard procedure for the completion of metric spaces [58] with some modifications. Letting $\mathcal{C}$ be *the set of Cauchy sequences in H_0*, one first partitions it into equivalence classes (we call $\{u_n\}, \{v_n\} \in \mathcal{C}$ equivalent and write $\{u_n\} \sim \{v_n\}$ if $\lim_{n \to \infty} \|u_n - v_n\|_{H_0} = 0$). We denote with $H_e := \mathcal{C}/\sim$ the set of equivalence classes, and with $[\{u_n\}]$ the generic element of H_e. Then one defines the inner product between $\mathcal{U}, \mathcal{V} \in H_e$ as $\langle \mathcal{U}, \mathcal{V} \rangle_{H_e} := \lim_{n \to \infty} \langle u_n, v_n \rangle_{H_0}$, where $\{u_n\} \in \mathcal{U}$ and $\{v_n\} \in \mathcal{V}$. The map $\varphi : H_0 \to H_e$ given by $\varphi(u) = [\{u, u, u, \ldots\}]$ is an isometry, and $\varphi(H_0)$ is in fact dense in H_e. Any Cauchy sequence in $\varphi(H_0)$ converges to some $\mathcal{U} \in H_e$ (in fact if $\{\varphi(u_n)\}$ is Cauchy in H_e then by isometry $\{u_n\}$ is Cauchy in H_0, therefore $\{u_n\}$ belongs to some $\mathcal{U} \in H_e$; one can prove that $\varphi(u_n) \to \mathcal{U}$ in H_e). Together with the density of $\varphi(H_0)$ in H_e, this can be used to show that H_e is complete. As we mentioned above, this procedure was rather standard; what follows is specific to our setting.

It is in fact the case that any Cauchy sequence in H_0 converges pointwise to some function $\Omega \to \mathbb{R}^d$. To see this, if $\{u_n\} \in \mathcal{C}$ then it is the case that $|(u_m(x) - u_n(x)) \cdot \alpha| = |\langle u_m - u_n, K_x^\alpha \rangle_{H_0}| \le \|u_m - u_n\|_{H_0} \sqrt{\alpha \cdot K(x, x)\alpha}$ for all $x \in \Omega$, $\alpha \in \mathbb{R}^d$, and $m, n \in \mathbb{N}$, by the Cauchy-Schwarz inequality. Whence $\{u_n(x)\}$ is Cauchy in $\mathbb{R}^d$ for all $x \in \Omega$ and there exists a function $u : \Omega \to \mathbb{R}^d$ such that $u_n \to u$ pointwise.

On the other hand Cauchy sequences in H_0 from distinct equivalence classes converge, pointwise, to distinct functions. In fact, if $\{u_n\}, \{v_n\} \in \mathcal{C}$

converge pointwise to the same $u : \Omega \to \mathbb{R}^d$ then the sequence $\{h_n\}$, with $h_n := u_n - v_n$, is also Cauchy in H_0 and it converges to 0 pointwise. Therefore, since $\langle h_n, K_x^\alpha \rangle_{H_0} = h_n(x) \cdot \alpha$ for all $x \in \Omega$ and $\alpha \in \mathbb{R}^d$, we have that $h_n \rightharpoonup 0$ in H_0. By the Cauchy property of $\{h_n\}$ there exists an integer M such that $\|h_n\|_{H_0}^2 - 2\langle h_n, h_M \rangle_{H_0} \leq \|h_n - h_M\|_{H_0}^2 \leq \varepsilon$ for all $n \geq M$. But we have $\lim \langle h_n, h_M \rangle_{H_0} = 0$ by weak convergence, whence for sufficiently large n we have $\|h_n\|_{H_0}^2 = \|u_n - v_n\|_{H_0}^2 \leq 2\varepsilon$. So it is the case that $\{u_n\} \sim \{v_n\}$.

This suggests that we may realize the completion H_0 as follows: we associate to each equivalence class in $\mathcal{C}/\sim$ of Cauchy sequences in H_0 the function $u : \Omega \to \mathbb{R}^d$ to which they converge pointwise, and let H be the space of such functions. If $\{u_n\}, \{v_n\} \in \mathcal{C}$ and $u_n \to u$, $v_n \to v$ pointwise, we say that $u = \lim_n u_n$ in H and set $\langle u, v \rangle_H := \lim_{n \to \infty} \langle u_n, v_n \rangle_{H_0}$, so H is isometric to H_e. By construction the function space H is complete, H_0 is embedded in H, and $H = \overline{H_0}$. Moreover K is a reproducing kernel for H, since if $\{u_n\} \in \mathcal{C}$ and $u_n \to u \in H$ pointwise then $\langle K_x^\alpha, u \rangle_H = \lim_n \langle K_x^\alpha, u_n \rangle_{H_0} = \lim_n \alpha \cdot u_n(x) = \alpha \cdot u(x)$, for all $x \in \Omega$ and $\alpha \in \mathbb{R}^d$.

Finally, assume that H_1 is another RKHS with the same kernel K. So it must be the case that $H_0 \subseteq H_1$, and the inner products $\langle \cdot, \cdot \rangle_H$ and $\langle \cdot, \cdot \rangle_{H_1}$ must coincide on H_0. If $\{u_n\} \in \mathcal{C}$ is such that $u_n \to u \in H$ pointwise, by the completeness of H_1 there exists $v \in H_1$ with $u_n \to v$ in H_1, and

$$|(u(x) - v(x)) \cdot \alpha| = \lim_n |(u_n(x) - v(x)) \cdot \alpha| = \lim_n |\langle u_n - v, K_x^\alpha \rangle_{H_1}|$$
$$\leq \lim_n \|u_n - v\|_{H_1} \|K_x^\alpha\|_{H_0} = 0,$$

for all $x \in \Omega$ and $\alpha \in \mathbb{R}^d$, so it is the case that $u = v$. Therefore $H \subseteq H_1$. Also $\langle \cdot, \cdot \rangle_{H_1} = \langle \cdot, \cdot \rangle_H$ on H, whence H is a closed subset of H_1 and $H_1 = H \oplus H^\perp$. If $u \in H_1$ is orthogonal to H then $\alpha \cdot u(x) = \langle u, K_x^\alpha \rangle_{H_1} = 0$ for all $x \in \Omega$ and $\alpha \in \mathbb{R}^d$, whence $u = 0$. In conclusion $H = H_1$. $\qquad \square$

Corollary 2.7. *Let H be a RKHS of $\mathbb{R}^d$-valued functions that are defined on Ω, with reproducing kernel K. It is the case that $H = \overline{H_0}$, where $H_0 = \text{span}\{K(\cdot, x)\alpha \mid x \in \Omega, \alpha \in \mathbb{R}^d\}$.*

Corollary 2.8. *Let H be a RKHS, and $\{u_n\}$ a sequence in H. If $u_n \rightharpoonup u$ in H (weak convergence) then it converges pointwise to u. On the other hand, if $\{u_n\}$ is bounded in H and it converges pointwise to some function u, then $u \in H$ and $u_n \rightharpoonup u$ in H.*

Proof. Let $x \in \Omega$ and $\alpha \in \mathbb{R}^d$ be arbitrary. If $u_n \rightharpoonup u$ in H then $\lim_n \alpha \cdot u_n(x) = \lim_n \langle K_x^\alpha, u_n \rangle_H = \langle K_x^\alpha, u \rangle_H = \alpha \cdot u(x)$. Vice versa, if the sequence $\{u_n\}$ is bounded in H then it has a subsequence $\{u_{\psi(n)}\}$ that converges

weakly to some $w \in H$; however, $\alpha \cdot w(x) = \langle K_x^\alpha, w \rangle_H = \lim_n \langle K_x^\alpha, u_{\psi(n)} \rangle_H = \lim_n \alpha \cdot u_{\psi(n)}(x) = \alpha \cdot u(x)$, whence $u = w$ and $u \in H$. Now, we have that $\lim_n \langle K_x^\alpha, u_n \rangle_H = \lim_n \alpha \cdot u_n(x) = \alpha \cdot u(x) = \langle K_x^\alpha, u \rangle_H$, so it is also the case that $\lim_n \langle u_n, h_0 \rangle_H = \langle u, h_0 \rangle_H$ for all $h_0 \in H_0$. If $M := \sup_n \|u_n - u\|_H$ then, since $H = \overline{H}_0$, for arbitrary $\varepsilon > 0$ and $h \in H$ we can choose $h_0 \in H_0$ such that $\|h - h_0\| < \varepsilon/M$, and we have $|\langle u_n - u, h \rangle_H| \leq |\langle u_n - u, h_0 \rangle_H| + \|u_n - u\|_H \|h - h_0\|_H < 2\varepsilon$ for sufficiently large n. $\qquad\square$

Proposition 2.9 (Kernel of separable RKHS)**.** *Let H be a separable RKHS, K its kernel and $\{u_i\}_{i \in \mathbb{N}}$ an orthonormal basis of H. Then $K(x, y) = \lim_{n \to \infty} \sum_{i=1}^n u_i(x) \, u_i(y)^T$ pointwise, for all $x, y \in \Omega$.*

Proof. For all $y \in \Omega$ and $\alpha \in \mathbb{R}^d$ we have $K_y^\alpha = \lim_{n \to \infty} \sum_{i=1}^n \langle K_y^\alpha, u_i \rangle_H u_i$ (limit in H and therefore pointwise), whence it must be the case that: $K(x, y)\alpha = \lim_{n \to \infty} \sum_{i=1}^n \langle K_y^\alpha, u_i \rangle_H u_i(x) = \lim_{n \to \infty} \sum_{i=1}^n \alpha \cdot u_i(y) \, u_i(x)$. $\quad\square$

We now introduce notions of *regularity* of $\mathbb{R}^d$-valued functions in RKHS, for which we assume that Ω is an open, connected subset of $\mathbb{R}^m$. Denote with $C_0(\Omega, \mathbb{R}^d)$ the space of continuous functions $u : \Omega \to \mathbb{R}^d$ that vanish at infinity (i.e. such that for every $\varepsilon > 0$ the set $\{x \in \Omega : \|u(x)\|_{\mathbb{R}^d} \geq \varepsilon\}$ is compact) which is Banach with the norm $\|u\|_\infty := \sup_{x \in \Omega} \|u(x)\|_{\mathbb{R}^d}$. For any integer $s \geq 0$ we define

$$C_0^s(\Omega, \mathbb{R}^d) := \left\{ u \in C^s(\Omega, \mathbb{R}^d) : \partial^p u \in C_0(\Omega, \mathbb{R}^d) \text{ for } |p| \leq s \right\},$$

where $C^s(\Omega, \mathbb{R}^d)$ is the space of functions that are continuously differentiable s times. (We have used multi-index notation [24] for partial derivatives $\partial^p = \partial_1^{p_1} \partial_2^{p_2} \ldots \partial_m^{p_m}$: i.e. if $p = (p_1, \ldots, p_m)$ then we set $|p| := p_1 + \cdots + p_m$. Also, we have used the symbol s instead of the more customary k to avoid confusion with a symbol that we will use later for kernels). The space $C_0^s(\Omega, \mathbb{R}^d)$ is Banach with the $W^{s,\infty}$ (Sobolev) norm

$$\|u\|_{s,\infty} := \sum_{|p| \leq s} \sup_{x \in \Omega} \|\partial^p u(x)\|_{\mathbb{R}^d},$$

and $C_0^s(\Omega, \mathbb{R}^d) \hookrightarrow C_0^{s-1}(\Omega, \mathbb{R}^d)$ for any $s \geq 1$. Also, $C(\Omega, \mathbb{R}^d)$ and $C^s(\Omega, \mathbb{R}^d)$, respectively, have the topology induced by the convergence with respect to the norms $\| \cdot \|_\infty$ and $\| \cdot \|_{s,\infty}$ on compact sets.

Definition 2.10. *Let $s \geq 0$ be an integer. A Hilbert space H of $\mathbb{R}^d$-valued functions defined on a set Ω is called s-admissible if $H \hookrightarrow C_0^s(\Omega, \mathbb{R}^d)$.*

Given a s-admissible space H, for a fixed point $x \in \Omega$ and a fixed vector $\alpha \in \mathbb{R}^d$ the evaluation functional $\delta_x^\alpha : H \to \mathbb{R} : u \mapsto \alpha \cdot u(x)$ is

obviously linear in u but also bounded: in fact, for all $u \in H$,

$$|\delta_x^\alpha(u)| = |\alpha \cdot u(x)| \leq \|\alpha\|_{\mathbb{R}^d}\|u(x)\|_{\mathbb{R}^d} \leq \|\alpha\|_{\mathbb{R}^d}\|u\|_{s,\infty} \leq C\|\alpha\|_{\mathbb{R}^d}\|u\|_H \,;$$

we have used the Cauchy-Schwarz inequality, the definition of $\|\cdot\|_{s,\infty}$, and Definition 2.10. Therefore if H is a s-admissible Hilbert space then it is also a Reproducing Kernel Hilbert Space, and admits a reproducing kernel such that $K(\cdot,x)\alpha \in C_0^s(\Omega, \mathbb{R}^d)$ for all $x \in \Omega$, $\alpha \in \mathbb{R}^d$.

Remark. By the symmetry $K(x,y) = K(y,x)^T$ it is the case that if K is continuous (respectively, differentiable) in one of the two variables then it is continuous (differentiable) in the other one too.

If the kernel $K : \Omega \times \Omega \to \mathbb{R}^{d \times d}$ is smooth enough and $p = (p_1, \ldots, p_m)$ and $q = (p_1, \ldots, p_m)$ are multi-indices, we indicate with $\partial_1^p \partial_2^q K(\cdot, \cdot)$ the partial derivative of K where ∂^p and ∂^q are taken with respect to the first and second sets of variables of $K(\cdot, \cdot)$, respectively.

Theorem 2.11. *Let H be a RKHS with kernel $K : \Omega \times \Omega \to \mathbb{R}^{d \times d}$, and $s \geq 0$ be an integer. The following two statements are equivalent:*

- A. $H \hookrightarrow C^s(\Omega, \mathbb{R}^d)$;
- B. *the function $\partial_1^p \partial_2^p K$ exists for all multi-indices p with $0 \leq |p| \leq s$, it is continuous in each of the two variables (separately), and it is locally bounded.*

Under the above assumptions, the following also holds:

- C. *for all $x \in \Omega$, $\alpha \in \mathbb{R}^d$, and multi-index p such that $0 \leq |p| \leq s$, we have $\partial_2^p K(\cdot, x)\alpha \in H$ and*

$$(4) \qquad \langle \partial_2^p K(\cdot, x)\alpha, u \rangle_H = \alpha \cdot \partial^p u(x), \quad \text{for all } u \in H.$$

If, furthermore, we have $K(\cdot, x)\alpha \in C_0^s(\Omega, \mathbb{R}^d)$ for all $x \in \Omega$ and $\alpha \in \mathbb{R}^d$, then $H \hookrightarrow C_0^s(\Omega, \mathbb{R}^d)$.

Proof. We shall prove that $(\text{A} \Rightarrow \text{B} \,\&\, \text{C})$ and $(\text{B} \Rightarrow \text{A} \,\&\, \text{C})$; in particular, this will imply that $(\text{A} \Leftrightarrow \text{B})$.

• $(\text{A} \Rightarrow \text{B} \,\&\, \text{C})$. We will prove by induction on $|p|$ that for all multi-indices p with $0 \leq |p| \leq s$, the function $\partial_2^p K(\cdot, x)\alpha$, with fixed $x \in \Omega$ and $\alpha \in \mathbb{R}^d$, is in H and (4) holds, and that the function $\partial_1^p \partial_2^p K$ exists and is continuous in each of its two variables (its local boundedness is proven separately). If $|p| = 0$ then $\partial_2^p K(\cdot, x)\alpha = K(\cdot, x)\alpha \in H$ and (4) is simply the reproducing property (2); also, since $H \subset C^s(\Omega, \mathbb{R}^d)$ we have that $\partial_1^p \partial_2^p K(\cdot, x) = K(\cdot, x)$ is continuous in its first variable: the continuity in the second one derives from the symmetry $K(x,y) = K(y,x)^T$, $x, y \in \Omega$.

We now fix $|p|$ and assume that for any multi-index r with $0 \leq |r| \leq |p|$ we have that $\partial_1^r \partial_2^r K$ exists and it is continuous in each of its two variables; also, $\partial_2^r K(\cdot, x)\alpha \in H$ and $\langle \partial_2^r K(\cdot, x)\alpha, u \rangle_H = \alpha \cdot \partial^r u(x)$, for all $u \in H$. For fixed $|p|$ take $p = (p_1, \ldots, p_m)$ and $q = p + e_\ell = (p_1, \ldots, p_\ell + 1, \ldots, p_m)$ for some arbitrary index $\ell \in \{1, \ldots, m\}$. For any $x \in \Omega$ and $\alpha \in \mathbb{R}^d$ we have that $K(\cdot, x)\alpha \in H \subset C^s(\Omega, \mathbb{R}^d)$; by the symmetry $K(x, y) = K(y, x)^T$ we also have that $K(x, \cdot)\alpha$ is in $C^s(\Omega, \mathbb{R}^d)$. Whence for any sequence $\{\varepsilon_n\}$ in $\mathbb{R}$ with $|\varepsilon_n| \to 0$, the sequence $\Delta_n(\cdot) := (\partial_2^p K(\cdot, x + \varepsilon_n e_\ell)\alpha - \partial_2^p K(\cdot, x)\alpha)/\varepsilon_n \in H$ converges pointwise to $\partial_2^q K(\cdot, x)\alpha$. For all $u \in H$ we have $\langle \Delta_n, u \rangle_H = \alpha \cdot (\partial^p u(x + \varepsilon_n e_\ell) - \partial^p u(x))/\varepsilon_n$, which converges to $\alpha \cdot \partial^q u(x)$ because $H \subset C^s(\Omega, \mathbb{R}^d)$. Therefore, for a fixed $u \in H$, we have $\sup_n |\langle \Delta_n, u \rangle_H| < \infty$, so by the Uniform Boundedness Principle [26] the sequence $\|\Delta_n\|_H$ is bounded. Together with the pointwise convergence of Δ_n, by Corollary 2.8 this implies:

(i) $\partial_2^q K(\cdot, x)\alpha \in H$. Since $H \subset C^s(\Omega, \mathbb{R}^d)$, $\partial_1^q \partial_2^q K(\cdot, x)$ exists and it is continuous in its first variable; by the symmetry $K(x, y) = K(y, x)^T$ it is also continuous in the second variable.

(ii) for all $u \in H$ the sequence $\langle \Delta_n, u \rangle_H$ converges to $\langle \partial_2^q K(\cdot, x)\alpha, u \rangle_H$. But we saw above that it also converges to $\alpha \cdot \partial^q u(x)$, whence we have $\langle \partial_2^q K(\cdot, x)\alpha, u \rangle_H = \alpha \cdot \partial^q u(x)$.

By the arbitrariness of $\ell \in \{1, \ldots, m\}$, this concludes the induction argument. To prove local boundedness of $\partial_1^p \partial_2^p K$ for an arbitrary multi-index p with $0 \leq |p| \leq s$, fix a compact subset $D \subset \Omega$. For an arbitrary $\alpha \in \mathbb{R}^d$ consider the maps $\{\Lambda_x : H \to \mathbb{R} : u \mapsto \langle \partial_2^p K(\cdot, x)\alpha, u \rangle_H \,|\, x \in D\} \subset H^*$, parameterized by $x \in D$. For a fixed $u \in H$ it is the case that $\sup_{x \in D} |\Lambda_x u| = \sup_{x \in D} |\langle \partial_2^p K(\cdot, x)\alpha, u \rangle_H| = \sup_{x \in D} |\alpha \cdot \partial^p u(x)| < \infty$, by the continuity of the derivative $\partial^p u$. Whence $\sup_{x \in D} \|\Lambda_x\|_{H^*}^2 = \sup_{x \in D} \alpha \cdot \partial_1^p \partial_2^p K(x, x)\alpha < \infty$, again by the Uniform Boundedness Principle; i.e., $x \mapsto \alpha \cdot \partial_1^p \partial_2^p K(x, x)\alpha$ is locally bounded. Therefore $\partial_1^p \partial_2^p K$ is locally bounded, since for all $x, y \in \Omega$ and $\alpha, \beta \in \mathbb{R}^d$ we have

$$\alpha \cdot \partial_1^p \partial_2^p K(x, y)\beta = \left\langle \partial_2^p K(\cdot, x)\alpha, \partial_2^p K(\cdot, y)\beta \right\rangle_H$$

$$\leq \left\| \partial_2^p K(\cdot, x)\alpha \right\|_H \left\| \partial_2^p K(\cdot, y)\beta \right\|_H = \sqrt{\alpha \cdot \partial_1^p \partial_2^p K(x, x)\alpha} \; \sqrt{\beta \cdot \partial_1^p \partial_2^p K(y, y)\beta}.$$

• (B $\Rightarrow$ A & C). Similarly, we will prove by induction on $|p|$ that for all multi-indices p with $0 \leq |p| \leq s$ we have that $H \subset C^{|p|}(\Omega, \mathbb{R}^d)$, that the function $\partial_2^p K(\cdot, x)\alpha$, with fixed $x \in \Omega$ and $\alpha \in \mathbb{R}^d$, is in H and that (4) holds (the continuity of the inclusion, $H \hookrightarrow C^s(\Omega, \mathbb{R}^d)$, will be proven separately). If $|p| = 0$ then $\partial_2^p K(\cdot, x)\alpha = K(\cdot, x)\alpha \in H$ and (4) is simply the reproducing property (2). To prove the continuity of the functions in H, fix $x_0 \in \Omega \subset \mathbb{R}^m$ and let $\{x_n\}$ be a sequence in Ω with $x_n \to x_0$. Let $\alpha \in \mathbb{R}^d$ be arbitrary and

define $v_n := K(\cdot, x_n)\alpha \in H$. It is the case that (i) $\|v_n\|_H^2 = \alpha \cdot K(x_n, x_n)\alpha$ is bounded, by the local boundedness of K, and (ii) $v_n(y) = K(y, x_n)\alpha \to K(y, x_0)\alpha$, for all $y \in \Omega$, by the continuity of K in each of its variables. Since v_n is bounded and it converges pointwise, by Corollary 2.8 it converges weakly to the same function: for any $u \in H$ we have $\langle K(\cdot, x_N)\alpha, u \rangle_H \to \langle K(\cdot, x_0)\alpha, u \rangle_H$, that is $\alpha \cdot u(x_n) \to \alpha \cdot u(x_0)$. Therefore $H \subset C(\Omega, \mathbb{R}^d)$ by the arbitrariness of u.

We now fix $|p|$ and assume that $H \subset C^{|p|}(\Omega, \mathbb{R}^d)$ and that for any multi-index r with $0 \leq |r| \leq |p|$ we have $\partial_2^r K(\cdot, x)\alpha \in H$ and $\langle \partial^r K(\cdot, x)\alpha, u \rangle_H = \alpha \cdot \partial^r u(x)$, for all $u \in H$. So, for fixed $|p|$, assume that $p = (p_1, \ldots, p_m)$ and $q = p + e_\ell = (p_1, \ldots, p_\ell + 1, \ldots, p_m)$, for some $\ell \in \{1, \ldots, m\}$. Take an arbitrary $u \in H$ and a sequence $\{\varepsilon_n\}$ in $\mathbb{R}$, with $|\varepsilon_n| \to 0$. By (4), for all $x \in \Omega$ and $\alpha \in \mathbb{R}^d$ we have

(5)
$$\psi_n := \alpha \cdot \frac{\partial^p u(x + \varepsilon_n e_\ell) - \partial^p u(x)}{\varepsilon_n} = \left\langle \frac{\partial_2^p K(\cdot, x + \varepsilon_n e_\ell)\alpha - \partial_2^p K(\cdot, x)\alpha}{\varepsilon_n}, u \right\rangle_H.$$

For fixed $x \in \Omega$ and $\alpha \in \mathbb{R}^d$ the sequence of functions $\Delta_n(\cdot) := (\partial_2^p K(\cdot, x + \varepsilon_n e_\ell)\alpha - \partial_2^p K(\cdot, x)\alpha)/\varepsilon_n$ converges pointwise to $\partial_2^q K(\cdot, x)\alpha$. Again by (4),

$$\|\Delta_n\|_H^2 = \frac{1}{\varepsilon_n^2}\Big\{ \alpha \cdot \partial_1^p \partial_2^p K(x + \varepsilon_n e_\ell, x + \varepsilon_n e_\ell)\alpha$$
$$- 2\alpha \cdot \partial_1^p \partial_2^p K(x, x + \varepsilon_n e_\ell)\alpha + \alpha \cdot \partial_1^p \partial_2^p K(x, x)\alpha \Big\},$$

which converges to $\alpha \cdot \partial_1^q \partial_2^q K(x, x)\alpha$ as $n \to \infty$, with q as above. Therefore Δ_n is bounded in H. By Corollary 2.8 it converges weakly in H to $\partial_2^q K(\cdot, x)\alpha$, which must be an element of H. Whence $\psi_n = \langle \Delta_n, u \rangle_H$ converges, i.e. the derivative $\partial^q u(x)$ exists and $\alpha \cdot \partial^q u(x) = \langle \partial_2^q K(\cdot, x)\alpha, u \rangle_H$.

To prove the continuity of $\partial^q u$, we proceed as in the case $|p| = 0$. Fix $x_0 \in \Omega \subset \mathbb{R}^m$ and let $\{x_n\}$ be a sequence in Ω with $x_n \to x_0$. Let $\alpha \in \mathbb{R}^d$ be arbitrary and define $w_n := \partial_2^q K(\cdot, x_n)\alpha \in H$. It is the case that (i) $\|w_n\|_H^2 = \alpha \cdot \partial_1^q \partial_2^q K(x_n, x_n)\alpha$ is a bounded sequence, by the local boundedness of K, and (ii) $w_n(y) = \partial_2^q K(y, x_n)\alpha \to \partial_1^q K(y, x_0)\alpha$, for all $y \in \Omega$, by the continuity of $\partial_2^q K$ in each of its variables. By Corollary 2.8, w_n converges weakly to $\partial_2^q K(\cdot, x)\alpha$, i.e. for all $u \in H$ we have $\langle \partial_2^q K(\cdot, x_N)\alpha, u \rangle_H \to \langle \partial_2^q K(\cdot, x_0)\alpha, u \rangle_H$, or $\alpha \cdot \partial^q u(x_n) \to \alpha \cdot \partial^q u(x_0)$; whence $\partial^q u \in C(\Omega, \mathbb{R}^d)$. By the arbitrariness of u and index ℓ we have $H \subset C^{|p|+1}(\Omega, \mathbb{R}^d)$. This concludes the induction.

Consider now the map $\iota : H \to C^s(\Omega, \mathbb{R}^d) : v \mapsto v$, let $\{v_n\}$ be a sequence in H and assume that $(v_n, v_n) \to (u, v)$ in $H \times C^s(\Omega, \mathbb{R}^d)$. Since $v_n \to u$ in H

it converges to it pointwise (by Corollary 2.8) and since $v_n \to v$ in $C^s(\Omega, \mathbb{R}^d)$ it also converges to it pointwise: whence $u = v$. So the graph of ι is a closed subspace of $H \times C^s(\Omega, \mathbb{R}^d)$, and by the Closed Graph Theorem [26] the map ι is bounded. Whence the inclusion is continuous, i.e. $H \hookrightarrow C^s(\Omega, \mathbb{R}^d)$.

• We have thus proven that ($\text{A} \Leftrightarrow \text{B}$). We now introduce the further assumption that $K_x^\alpha \in C_0^s(\Omega, \mathbb{R}^d)$. Let $\{u_n\}$ be a sequence in H_0 that converges, in the H norm, to an arbitrary $u \in H$; therefore $u_n \rightharpoonup u$ in H. For any multi-index p such that $0 \le |p| \le s$ we have $\langle \partial_2^p K(\cdot, x)\alpha, u_n \rangle_H \to \langle \partial_2^p K(\cdot, x)\alpha, u \rangle_H$ i.e. $\alpha \cdot \partial^p u_n(x) \to \alpha \cdot \partial^p u(x)$. Therefore $u_n \to u$ in the topology of $C^s(\Omega, \mathbb{R}^d)$; but $C_0^s(\Omega, \mathbb{R}^d)$ is closed in $C^s(\Omega, \mathbb{R}^d)$, whence $u \in C_0^s(\Omega, \mathbb{R}^d)$. In conclusion, $H \hookrightarrow C_0^s(\Omega, \mathbb{R}^d)$. □

Example. An example of non-degenerate, s-admissible RKHS that is used in applications [69] is the Sobolev space of vector fields $H := W^{\ell,2}(\mathbb{R}^d, \mathbb{R}^d) = H^\ell(\mathbb{R}^d, \mathbb{R}^d)$ with the inner product:

$$(6) \qquad \langle u, v \rangle_H := \int_{\mathbb{R}^d} Lu(x) \cdot v(x)\, dx = \int_{\mathbb{R}^d} \sum_{j=0}^\ell \binom{\ell}{j} \sigma^{2j} \sum_{|p|=j} \partial^p u \cdot \partial^p v\, dx.$$

In (6) the differential operator $L := (1 - \sigma^2 \Delta)^\ell$ (where $\sigma > 0$ is a scaling factor, $\ell \ge 0$ is an integer, and Δ is the Laplace operator) is applied to each component of the vector field $u = (u^1, \dots, u^d)$; if $\ell > s + d/2$ then $H \hookrightarrow C_0^s(\mathbb{R}^d, \mathbb{R}^d)$ by the Sobolev Embedding Theorem [26], i.e. H is s-admissible. Regarding its reproducing kernel, first note that for any $u \in H$ and $x, \alpha \in \mathbb{R}^d$ we must have

$$\alpha \cdot u(x) = \langle K_x^\alpha, u \rangle_H = \int_{\mathbb{R}^d} K_x^\alpha(y) \cdot Lu(y)\, dy$$

$$= \int_{\mathbb{R}^d} \big(K(y, x)\alpha \big)^T Lu(y)\, dy = \alpha \cdot \int_{\mathbb{R}^d} K(x, y)\, Lu(y)\, dy,$$

therefore $u(x) = \int_{\mathbb{R}^d} K(x, y)\, Lu(y)\, dy$ for all $u \in H$. Since L is applied to each component of the vector field u (i.e. the differential operator does not mix the components of u) we must have $K(x, y) = G(x, y)\mathbb{I}_d$, where G is the *Green's function* of L and $\mathbb{I}_d$ is the $d \times d$ identity matrix. As reported in [33], since $L = (1 - \sigma^2 \Delta)^\ell$ we have that $K(x, y)\alpha = k(\|x - y\|_{\mathbb{R}^d})\mathbb{I}_d$, with

$$(7) \qquad k(r) = C(\sigma, d, \ell) \left(\frac{r}{\sigma}\right)^{\ell - \frac{d}{2}} K_{\ell - \frac{d}{2}}\left(\frac{r}{\sigma}\right), \qquad r \ge 0,$$

where $C(\sigma, d, \ell) := \big(2^{\ell + \frac{d}{2} - 1} \pi^{\frac{d}{2}} \Gamma(\ell)\, \sigma^d \big)^{-1}$ and K_ν, with $\nu = \ell - \frac{d}{2}$, is a modified Bessel function [1] of order ν (this should not be confused with the

symbol K that we use for kernels). Kernels of the type $K(x,y) = k(\|x-y\|)\mathbb{I}_d$ with k given by (7) are in fact referred to as *Bessel kernels*.

We have just seen an example of a RKHS with a "scalar" kernel, in that K is given by a scalar-valued function multiplied by the identity matrix. In the next section we shall explore a large class of (non-scalar) reproducing kernels, study their properties, and provide some significant examples.

3. Translation- and rotation-invariant metrics in RKHS

In the current and following sections we shall restrict our attention to RKHS whose kernel induces an inner product that is translation- and rotation-invariant; we shall call these TRI kernels. In this case it is natural to assume that $\Omega = \mathbb{R}^d$, so that the representer functions $K_x^\alpha(\cdot)$ are in fact vector fields in $\mathbb{R}^d$. From now on, unless otherwise specified, we shall assume that $d \geq 2$.

3.1. TRI kernels

In $\Omega = \mathbb{R}^d$ we indicate the generic translation with the symbol $\tau : x \mapsto x+t$, for some fixed $t \in \mathbb{R}^d$, and the generic rotation with $\rho : x \mapsto Rx$, for some fixed $R \in \mathrm{O}(d, \mathbb{R})$ (the orthogonal group over $\mathbb{R}$ of dimension d).

Theorem 3.1 (translation invariance). *Let H be a RKHS with kernel $K :$ $\Omega \times \Omega \to \mathbb{R}^{d \times d}$. The map $u \mapsto u \circ \tau$ is an isometry in H for any translation τ if and only if there exists a function $\mathbf{k} : \mathbb{R}^d \to \mathbb{R}^{d \times d}$ such that $K(x,y) = \mathbf{k}(x-y)$ for all $x,y \in \mathbb{R}^d$.*

Proof. Assume first that $u \mapsto u \circ \tau$ is an isometry for all translations $\tau : x \mapsto x + t$. Let $f := K_x^\alpha$ and $g := K_y^\beta$; their product is $\langle f, g \rangle_H = \alpha \cdot K(x,y)\beta$, while $\langle f \circ \tau, g \circ \tau \rangle_H = \langle K(\cdot, x+t)\alpha, K(\cdot, y+t)\beta \rangle_H = \alpha \cdot K(x+t, y+t)\beta$. Therefore it must be the case that $K(x,y) = K(x+t, y+t)$ for all x, y and $t \in \mathbb{R}^d$; in particular $K(x,y) = K(x-y, 0) = \mathbf{k}(x-y)$ for all $x,y \in \mathbb{R}^d$, where $\mathbf{k}(z) := K(z,0)$, $z \in \mathbb{R}^d$. Conversely if $K(x,y) = \mathbf{k}(x-y)$ for all $x,y \in \mathbb{R}^d$ then it must be the case that $\langle K_{x+t}^\alpha, K_{y+t}^\beta \rangle_H = \alpha \cdot K(x+t, y+t)\beta = \alpha \cdot \mathbf{k}(x+t-y-t)\beta = \alpha \cdot \mathbf{k}(x-y)\beta = \alpha \cdot K(x,y)\beta = \langle K_x^\alpha, K_y^\beta \rangle_H$ for all $t \in \Omega$. Defining the set $H_0 := \mathrm{span}\{K_x^\alpha \,|\, x, \alpha \in \mathbb{R}^d\}$ we have proven that the linear map $T : H_0 \to H_0 : u \mapsto u \circ \tau$, with $\tau : x \mapsto x + t$, is an isometry in H_0 for arbitrary translations τ.

This can be extended to an isometry $\overline{T}$ in H as follows. If $u \in H$, let $\{u_n\}$ be a Cauchy sequence such that $u_n \to u$ in H and therefore point-wise. The map T is an isometry, whence $\{Tu_n\}$ is also Cauchy in H_0 and therefore it converges to some element of H which we call $\overline{T}u$. If $\{u_n\}, \{v_n\}$ are Cauchy in H_0 and respectively converge to $u, v \in H$ then $\langle \overline{T}u, \overline{T}v \rangle_H =$

$\lim_n \langle \overline{T} u_n, \overline{T} v_n \rangle_{H_0} = \lim_n \langle u_n, v_n \rangle_{H_0} = \langle u, v \rangle_H$, so $\overline{T} : H \to H$ is also an isometry. Finally,

$$\alpha \cdot \overline{T} u(x) = \langle K_x^\alpha, \overline{T} u \rangle_H = \lim_n \langle K_x^\alpha, T u_n \rangle_{H_0}$$
$$= \lim_n \alpha \cdot T u_n(x) = \lim_n \alpha \cdot u_n(x+t) = \alpha \cdot u(x+t),$$

for all $x, \alpha \in \mathbb{R}^d$, therefore $\overline{T} u = u \circ \tau$. This completes the proof. $\square$

We call a kernel with the properties described in Theorem 3.1 *translation invariant*; with a small abuse of terminology we shall also call $\mathbf{k}$ "kernel". We now introduce the following matrices in $\mathbb{R}^{d \times d}$:

$$(8) \qquad \mathrm{Pr}_x^\parallel := \frac{xx^T}{\|x\|^2} \qquad \text{and} \qquad \mathrm{Pr}_x^\perp := \mathbb{I}_d - \frac{xx^T}{\|x\|^2}, \qquad x \in \mathbb{R}^d \setminus \{0\},$$

which are, respectively, the projection operators onto vector x and onto the plane that is perpendicular to x (again, we indicate with $\mathbb{I}_d$ the $d \times d$ identity matrix). In (8) we have indicated with $\|x\|$ the Euclidean norm of a point in $x \in \mathbb{R}^d$, as we shall do, for simplicity, throughout the rest of the paper. The following lemma does not require $\mathbf{k}$ to be the kernel of a RKHS.

Lemma 3.2. *Let* $\mathbf{k} : \mathbb{R}^d \to \mathbb{R}^{d \times d}$ *be a generic matrix-valued function. The following are equivalent:*

i. *for any* $x \in \mathbb{R}^d$ *and* $R \in \mathrm{O}(d, \mathbb{R})$ *it is the case that* $\mathbf{k}(-x) = \mathbf{k}(x)^T$ *and*

$$(9) \qquad\qquad\qquad \mathbf{k}(Rx) = R\,\mathbf{k}(x)R^{-1};$$

ii. *there exist a scalar* $k_0 \in \mathbb{R}$ *such that* $\mathbf{k}(0) = k_0 \mathbb{I}_d$ *and two scalar functions* $k^\parallel, k^\perp : \mathbb{R}^+ \to \mathbb{R}$ *such that for all* $x \in \mathbb{R}^d$, $x \neq 0$, *we have*

$$(10) \qquad\qquad\qquad \mathbf{k}(x) = k^\parallel(\|x\|)\,\mathrm{Pr}_x^\parallel + k^\perp(\|x\|)\,\mathrm{Pr}_x^\perp.$$

Proof. (i. $\Rightarrow$ ii.) By choosing $R = -\mathbb{I}_d$ in (9) we have $\mathbf{k}(x) = \mathbf{k}(-x)$ for all $x \in \mathbb{R}^d$; combining the latter with $\mathbf{k}(-x) = \mathbf{k}(x)^T$ yields the symmetry $\mathbf{k}(x) = \mathbf{k}(x)^T$. So $\mathbf{k}(x)$ can be diagonalized, i.e. there exist matrices $U(x) \in \mathrm{O}(d, \mathbb{R})$ and $\Sigma(x) = \mathrm{diag}(\sigma_1(x), \dots, \sigma_d(x))$ such that $\mathbf{k}(x)U(x) = U(x)\Sigma(x)$. Since $\mathbf{k}(0)R = R\,\mathbf{k}(0)$ for all $R \in \mathrm{O}(d, \mathbb{R})$ by (9), we have $\mathbf{k}(0) = \Sigma(0)$ (diagonal matrix). If $R \in \mathrm{O}(d, \mathbb{R})$ is such that $Re_i = e_j$ (for fixed indices $i \neq j$) then $\sigma_j(0)e_j = \mathbf{k}(0)e_j = \mathbf{k}(0)Re_i = R\,\mathbf{k}(0)e_i = R\sigma_i(0)e_i = \sigma_i(0)e_j$, so $\sigma_j(0) = \sigma_i(0)$ (all eigenvalues are the same). Therefore $\mathbf{k}(0) = k_0 \mathbb{I}_d$, for some $k_0 \in \mathbb{R}$.

Now fix $x \in \mathbb{R}^d$, $x \neq 0$, and $R \in \mathrm{O}(d, \mathbb{R})$ such that $Rx = x$ and $\dim \{v \in \mathbb{R}^d \,|\, Rv = v\} = 1$. By (9) we have $\mathbf{k}(x)R = R\,\mathbf{k}(x)$ for such

choice of R, so $R\mathbf{k}(x)x = \mathbf{k}(x)Rx = \mathbf{k}(x)x$, i.e. both $\mathbf{k}(x)x$ and x are eigenvectors of R with eigenvalue 1. Whence they $\mathbf{k}(x)x = \lambda^{\|}(x)x$ for some scalar $\lambda^{\|}(x)$. Therefore $\lambda^{\|}(x)$ is an eigenvalue of $\mathbf{k}(x)$; without loss of generality we assume that $\sigma_1(x) = \lambda^{\|}(x)$. We now claim that $\sigma_i(x) = \sigma_j(x)$ for $i, j \geq 2$. Denoting the orthonormal columns of $U(x)$, i.e. the eigenvectors of $\mathbf{k}(x)$, with $u_1(x), \ldots, u_d(x)$, for any pair of indices $i, j \geq 2$ with $i \neq j$ there exists a matrix $R \in \mathrm{O}(d, \mathbb{R})$ such that $x = Rx$ and $u_i(x) = Ru_j(x)$. By equation (9), $\mathbf{k}(Rx)R\,u_j(x) = R\,\mathbf{k}(Rx)\,u_j(x)$, whence $\mathbf{k}(Rx)\,u_i(x) = R\,\mathbf{k}(Rx)\,u_j(x)$, that is $\sigma_i(x)\,u_i(x) = R\,\sigma_j(x)\,u_j(x) = \sigma_j(x)\,u_i(x)$; therefore $\sigma_i(x) = \sigma_j(x)$. So there is a scalar $\lambda^{\perp}(x)$ such that $\mathbf{k}(x)\beta = \lambda^{\perp}(x)\beta$ for all vectors $\beta \perp x$.

We claim that $\lambda^{\|}$ and $\lambda^{\perp}$ only depend on $\|x\|$. For all $x \in \mathbb{R}^d$ and $R \in \mathrm{O}(d, \mathbb{R})$ we have, by (9), that $\mathbf{k}(Rx)Rx = R\,\mathbf{k}(x)x$, whence $\lambda^{\|}(Rx)Rx = R\lambda^{\|}(x)x$, i.e. $\lambda^{\|}(Rx) = \lambda^{\|}(x)$ for all x and R, which implies that $\lambda^{\|}(x) = k^{\|}(\|x\|)$ for some $k^{\|} : \mathbb{R}^+ \to \mathbb{R}$. Similarly for any $\alpha \in \mathbb{R}^d$ we have, again by (9), that $\mathbf{k}(Rx)R\alpha = R\,\mathbf{k}(x)\alpha$; if $\alpha \perp x$ then also $R\alpha \perp Rx$, whence $\lambda^{\perp}(Rx)R\alpha = R\lambda^{\perp}(x)\alpha$ and in fact $\lambda^{\|}(Rx) = \lambda^{\|}(x)$ for all x and R. This again implies that $\lambda^{\perp}(x) = k^{\perp}(\|x\|)$ for some $k^{\perp} : \mathbb{R}^+ \to \mathbb{R}$.

(ii. $\Rightarrow$ i.) The implication is obvious in the case $x = 0$. When $x \neq 0$, for any $R \in \mathrm{O}(d, \mathbb{R})$ it is the case that $\mathrm{Pr}_{Rx}^{\|} = R\,\mathrm{Pr}_x^{\|}\,R^{-1}$ and $\mathrm{Pr}_{Rx}^{\perp} = R\,\mathrm{Pr}_x^{\perp}\,R^{-1}$, and we conclude immediately. $\qquad\square$

Definition 3.3. *When a function* $\mathbf{k} : \mathbb{R}^d \to \mathbb{R}^{d \times d}$ *may be written in the form* (10)*, we call the functions* $k^{\|}, k^{\perp} : \mathbb{R}^+ \to \mathbb{R}$ *the* coefficients *of* $\mathbf{k}$.

Corollary 3.4. *If the function* $\mathbf{k} : \mathbb{R}^d \to \mathbb{R}^{d \times d}$ *is such that* (10) *holds, then for* $x \neq 0$ *the eigenvalues of* $\mathbf{k}(x)$ *are* $k^{\|}(\|x\|)$*, with multiplicity* 1 *and eigenvector* x*, and* $k^{\perp}(\|x\|)$*, with multiplicity* $d - 1$.

Theorem 3.5 (rotation invariance)**.** *Let* H *be a RKHS with a translation-invariant reproducing kernel, i.e.* $K(x, y) = \mathbf{k}(x - y)$ *for all* $x, y \in \mathbb{R}^d$*. The map* $v \mapsto \rho v \circ \rho^{-1}$ *is an isometry in* H *for any rotation* $\rho : x \mapsto Rx$*, with* $R \in \mathrm{O}(d, \mathbb{R})$*, if and only if* (9) *holds for all* $R \in \mathrm{O}(d, \mathbb{R})$.

Proof. Assume that $v \mapsto \rho v \circ \rho^{-1}$ is an isometry in H for any rotation $\rho : x \mapsto Rx$, with $R \in \mathrm{O}(d, \mathbb{R})$. For fixed $x, y, \alpha, \beta \in \mathbb{R}^d$ and $R \in \mathrm{O}(d, \mathbb{R})$, define the functions $f := R^{-1}K_{Rx}^{R\alpha} \circ R$ and $g := K_y^{\beta} \in H$; then we have

$$\langle f, g \rangle_H = \left\langle R^{-1}K(R\,\cdot\,, Rx)R\alpha, K(\,\cdot\,, y)\beta \right\rangle_H = \beta^T R^{-1} K(Ry, Rx)R\alpha,$$

by the reproducing property of the *second* factor. But for all $\alpha, \beta \in \mathbb{R}^d$ this must be equal to

$$\langle Rf \circ R^{-1}, Rg \circ R^{-1}\rangle_H = \left\langle K(\,\cdot\,, Rx)R\alpha, RK(R^{-1}\,\cdot\,, y)\beta\right\rangle_H$$

$$\overset{(*)}{=} (R\alpha)^T RK(R^{-1}Rx, y)\beta = \alpha^T K(x,y)\beta = \beta^T K(x,y)^T \alpha = \beta^T K(y,x)\alpha$$

in $(*)$ we have used the reproducing property of the *first* factor. Whence we have $R^{-1}K(Ry, Rx)R = K(y,x)$ for all $x, y \in \mathbb{R}^d$ and $R \in \mathrm{O}(d, \mathbb{R})$, and (9) follows from translation invariance (Proposition 3.1).

Now assume that (9) holds for all $R \in \mathrm{O}(d, \mathbb{R})$. Take $f := K_x^\alpha$ and $g := K_y^\beta$, for arbitrary points x, y and and vectors α, β. For any $R \in \mathrm{O}(d, \mathbb{R})$ we have $(Rf \circ R^{-1})(\cdot) = R\mathbf{k}(R^{-1}\cdot - x)\alpha = R\mathbf{k}(R^{-1}(\cdot - Rx))\alpha = \mathbf{k}(\cdot - Rx)R\alpha$ by (9), and a similar expression holds for $Rg \circ R^{-1}$; therefore

$$\begin{aligned}
\left\langle Rf \circ R^{-1}, Rg \circ R^{-1}\right\rangle_H &= \left\langle \mathbf{k}(\cdot - Rx)R\alpha, \mathbf{k}(\cdot - Ry)R\beta\right\rangle_H \\
&= \beta^T R^T \mathbf{k}(R(y-x))R\alpha \overset{(**)}{=} \beta^T R^T R\,\mathbf{k}(y-x)\alpha \\
&= \beta^T \mathbf{k}(y-x)\alpha = \left\langle \mathbf{k}(\cdot - x)\alpha, \mathbf{k}(\cdot - y)\beta\right\rangle_H = \left\langle f, g\right\rangle_H,
\end{aligned}$$

where we have used (9) in $(**)$. Whence rotations are isometries $H_0 = \mathrm{span}\{K_x^\alpha \,|\, x, \alpha \in \mathbb{R}^d\}$. An argument that is in all similar to the last part of the proof of Theorem 3.1 can be employed to prove that the rotations $\rho : x \mapsto Rx$, with $R \in \mathrm{O}(d, \mathbb{R})$, are in fact isometries on all of $H = \overline{H}_0$. $\quad\square$

Note that by "rotations" we intend *all* the elements of $\mathrm{O}(d, \mathbb{R})$ and not just the special orthogonal group $\mathrm{SO}(d, \mathbb{R})$; therefore we also include e.g. all permutations of the coordinates and reflections. By Definition 2.5, translation-invariant *kernels* $K(x,y) = \mathbf{k}(x-y)$ have the property that $\mathbf{k}(-x) = \mathbf{k}(x)^T$ for all $x \in \mathbb{R}^d$, whence they may be written in the form (10).

Definition 3.6 (TRI kernels). *The kernels of RKHS with translation- and rotation-invariant inner products, that may therefore be written in the form (10), are called* TRI kernels.

Not all matrix-valued functions $\mathbf{k}$ of the form (10) are positive definite, i.e. kernels. In the next section we will characterize the functions $(k^\parallel, k^\perp)$ that give rise to TRI kernels, and later in the paper we shall provide a method to construct such coefficients. Introducing the auxiliary function

$$(11) \qquad\qquad \tilde{k}(r) := \frac{k^\parallel(r) - k^\perp(r)}{r^2}, \qquad r > 0,$$

we may write the generic element of a matrix-valued function $\mathbf{k} = (\mathbf{k}_{ij})_{i,j=1,\ldots,d}$ of the type (10) as follows, for $x \neq 0$:

$$(12) \qquad \mathbf{k}_{ij}(x) = \left[k^{\|}(\|x\|) - k^{\perp}(\|x\|)\right] \frac{x_i x_j}{\|x\|^2} + k^{\perp}(\|x\|)\,\delta_{ij}$$

$$= \tilde{k}(\|x\|)\,x_i x_j + k^{\perp}(\|x\|)\,\delta_{ij}, \qquad i,j = 1,\dots,d.$$

If $\mathbf{k}(0) = k_0\mathbb{I}_d$ we may define $k^{\|}$ and $k^{\perp}$ at zero by setting $k^{\|}(0) = k^{\perp}(0) = k_0$; in fact, when $\mathbf{k}$ is continuous, $\lim_{r\to 0^+} k^{\|}(r) = \lim_{r\to 0^+} k^{\perp}(r) = k_0$, thus this choice is justified.

The result that follows states a property of TRI kernels that is a generalization of a well-known property of scalar-valued positive definite functions.

Proposition 3.7. *Let H be a RKHS with a TRI kernel $\mathbf{k}$. Its coefficients $k^{\|}$, $k^{\perp}$ and the number k_0, introduced in Lemma 3.2, have the properties: $k_0 \geq 0$, $k_0 \geq |k^{\|}(r)|$ and $k_0 \geq |k^{\perp}(r)|$ for all $r > 0$. If H is non-degenerate then such inequalities are strict.*

Proof. By Proposition 2.4, when expression (10) holds we must have, for all $N \in \mathbb{N}$, points $x_1,\dots,x_N \in \mathbb{R}^d$ and vectors $\alpha_1,\dots,\alpha_N \in \mathbb{R}^d$,

$$2\sum_{1 \leq a < b \leq N} \left\{ k^{\|}(\|x_a - x_b\|)\alpha_a^T \mathrm{Pr}^{\|}(x_a - x_b)\alpha_b + k^{\perp}(\|x_a - x_b\|)\alpha_a^T \mathrm{Pr}^{\perp}(x_a - x_b)\alpha_b \right\}$$

$$+ k_0 \sum_{a=1}^{N} \|\alpha_a\|^2 \geq 0.$$

For now we shall not assume the non-degeneracy of H. By choosing $N = 1$ we have $k_0\|\alpha_1\|^2 \geq 0$ for all $\alpha_1 \in \mathbb{R}^d$; therefore $k_0 \geq 0$. We now fix $N = 2$. If we take $\alpha_1 = \alpha_2 \perp (x_1 - x_2)$ then the above expression yields $2[k^{\perp}(\|x_1 - x_2\|) + k_0]\|\alpha_1\|^2 \geq 0$ for all $x_1 \neq x_2$ and α_1, so that $k_0 \geq -k^{\perp}(r)$ for all $r > 0$; similarly, if we take $\alpha_1 = -\alpha_2 \perp (x_1 - x_2)$ then we get $2[-k^{\perp}(\|x_1 - x_2\|) + k_0]\|\alpha_1\|^2 \geq 0$ for all $x_1 \neq x_2$ and α_1, which implies $k_0 \geq k^{\perp}(r)$ for all $r > 0$. Combining the two results yields $k_0 \geq |k^{\perp}(r)|$. To prove that $k_0 \geq |k^{\|}(r)|$ one follows an analogous argument, by choosing $\alpha_1 = \alpha_2$ parallel to $(x_1 - x_2)$ first, and $\alpha_1 = -\alpha_2$ parallel to $(x_1 - x_2)$ later. It is immediate to see that when H is non-degenerate the inequalities ($\geq$) become strict ($>$). $\qquad\square$

Remark. So far in this section we have not made any assumption of regularity. For now we shall limit ourselves to observing that if H is a RKHS with a translation-invariant kernel $\mathbf{k}$ and we also have $H \hookrightarrow C^s(\Omega, \mathbb{R}^d)$, then by Theorem 2.11 it is the case that $\mathbf{k} \in C^{2s}(\Omega, \mathbb{R}^{d\times d})$; that is, the kernel is "twice as smooth" as the functions of the space $C_0^s(\Omega, \mathbb{R}^d)$ into which H is embedded. Also, since $\partial_2^q K(\cdot, x)\alpha = (-1)^{|q|}\partial^q \mathbf{k}(\cdot - x)\alpha$ and

$\partial_1^p \partial_2^q K(\cdot, x)\alpha = (-1)^{|q|}\partial^{p+q}\mathbf{k}(\cdot - x)\alpha$, property (4) implies that

$$(13) \qquad \alpha \cdot \partial^{p+q}\mathbf{k}(x - y)\beta = (-1)^{|p|}\langle \partial^p\mathbf{k}(\cdot - x)\alpha, \partial^q\mathbf{k}(\cdot - y)\beta\rangle_H,$$

for all $x, y, \alpha, \beta \in \mathbb{R}^d$ and multi-indices p and q with $0 \leq |p| \leq s$, $0 \leq |q| \leq s$.

3.2. Characterization of TRI kernels

We shall now find conditions on $k^\|$ and $k^\perp$ for a function $\mathbf{k}$ for the type (10) to be positive definite, whence a kernel. The following auxiliary result does not require $\mathbf{k}$ to be of the type (10).

Theorem 3.8 (Bochner's theorem). *Consider a matrix-valued function $\mathbf{k} \in L^1(\mathbb{R}^d, \mathbb{R}^{d \times d})$, such that its Fourier transform $\widehat{\mathbf{k}}$ is also in $L^1(\mathbb{R}^d, \mathbb{R}^{d \times d})$. Then $K(x, y) := \mathbf{k}(x - y)$ is positive definite if and only if $\widehat{\mathbf{k}}(\xi)$ is a self-adjoint positive definite matrix for all $\xi \in \mathbb{R}^d$.*

The classical version of the above theorem holds for $d = 1$ and states that a function is positive definite if and only if its Fourier transform is non-negative [10]. The proof of Theorem 3.8 is reported in Appendix B, and the result can be extended to L^2 functions with the usual density arguments [64].

In the particular case $k^\| = k^\perp =: k$ we have the class of kernels that are simply referred to as *scalar*: $\mathbf{k}(x) = k(\|x\|)\mathbb{I}_d$, $x \in \mathbb{R}^d$. In such a case positive definiteness (Definition 2.3) is obviously equivalent to the following: for abitrary $N \in \mathbb{N}$, $x_1, \ldots, x_N \in \mathbb{R}^d$ and $r_1, \ldots, r_N \in \mathbb{R}$,

$$(14) \qquad \sum_{a,b=1}^N r_a r_b\, k(\|x_a - x_b\|) \geq 0;$$

whence the scalar-valued function $k(\|\cdot\|)$ must be positive definite; to apply Bochner's theorem we must compute its Fourier transform. We shall employ the next proposition (see Appendix C for a simple proof).

Theorem 3.9. *If $f \in L^1(\mathbb{R}^d)$ is a radial function, i.e. $f(x) = g(\|x\|)$ for some $g : \mathbb{R}^+ \to \mathbb{R}$, then so is its Fourier transform $\widehat{f}$. It is in fact that case that $\widehat{f}(\xi) = G(\|\xi\|)$ with*

$$G(\varrho) = \frac{2\pi}{\varrho^\mu} \int_0^\infty r^{\mu+1} g(r)\, J_\mu(2\pi\varrho r)\, dr, \qquad \varrho > 0,$$

where $\mu := \frac{d}{2} - 1$.

We remind the reader that the *Hankel transform* [12, 53, 63] of order ν of a function $f : \mathbb{R}^+ \to \mathbb{R}$ is defined as $\mathcal{H}_\nu[f](\varrho) := \int_0^\infty r f(r)\, J_\nu(\varrho r)\, dr$, where J_ν is the Bessel function of the first kind [1] of order ν (references [20, 23, 52] provide tables of Hankel transforms); in Appendix A we list relevant properties of J_ν that we will be using throughout the rest of the paper. Here we note that if the function $f : \mathbb{R}^d \to \mathbb{R}$ is radial, i.e. $f(x) = g(\|x\|)$ for some $g : \mathbb{R}^+ \to \mathbb{R}$, then

$$\int_{\mathbb{R}^d} |f(x)|^p\, dx = \sigma(\mathbb{S}^{d-1}) \int_0^\infty |g(r)|^p\, r^{d-1} dr,$$

where $\sigma(\mathbb{S}^n)$ is the surface area of the unit n-sphere; therefore $f \in L^p(\mathbb{R}^d)$ if and only if $g \in L^p(\mathbb{R}^+, r^{d-1})$, i.e. g is p^{th}-power integrable with respect to the measure $r^{d-1} dr$.

An immediate consequence of the above proposition and of Bochner's theorem is that a scalar kernel $\mathbf{k}(\cdot) = k(\|\cdot\|)\mathbb{I}_d \in L^1(\mathbb{R}^d, \mathbb{R}^{d\times d})$ is positive definite if and only if the function

$$(15) \qquad h(\varrho) := \frac{2\pi}{\varrho^\mu} \int_0^\infty r^{\mu+1} k(r)\, J_\mu(2\pi\varrho r)\, dr, \qquad \text{defined for } \varrho > 0,$$

is nonnegative, where we have set $\mu := \frac{d}{2} - 1$. Also, we will have that $\widehat{\mathbf{k}}(\xi) = h(\|\xi\|)\mathbb{I}_d$, with $\xi \in \mathbb{R}^d$. We should note that if $\mathbf{k}$ is in $L^1(\mathbb{R}^d, \mathbb{R}^{d\times d})$ then the function h is (uniformly) continuous and vanishes at infinity, by the well-known properties of Fourier transforms [26, 64]. Also, by applying Hankel's integral formula (73) in Appendix A to (15), it is immediate to verify that

$$(16) \qquad k(r) = \frac{2\pi}{r^\mu} \int_0^\infty \varrho^{\mu+1} h(\varrho)\, J_\mu(2\pi\varrho r)\, dr, \qquad r > 0,$$

i.e. the map $L^1(\mathbb{R}^+, r^{d-1}) \to C_0(\mathbb{R}^+) : k \mapsto h$ given by formula (15) is in fact an *involution* (it is equal to its inverse, when inversion makes sense); here we have used the symbol $C_0(\mathbb{R}^+)$ to indicate the set of real-valued continuous functions defined on $\mathbb{R}^+$ that vanish at $+\infty$.

Examples of scalar kernels. For a fixed $\sigma > 0$, three scalar kernels $\mathbf{k}(x) = k(\|x\|)\mathbb{I}_d$, $x \in \mathbb{R}^d$, that are popular in applications are the following:

Gaussian kernels:

$$(17) \quad k(r) = \exp\left(-\frac{1}{2}\frac{r^2}{\sigma^2}\right), \qquad\qquad h(\varrho) = \left(2\pi\sigma^2\right)^{\mu+1} e^{-2\pi^2\sigma^2\varrho^2};$$

Cauchy kernels:

$$(18) \quad k(r) = \frac{1}{1 + r^2/\sigma^2}, \qquad\qquad h(\varrho) = 2\pi\sigma^2 \Big(\frac{\sigma}{\varrho}\Big)^\mu K_\mu(2\pi\sigma\varrho);$$

Bessel kernels:

$$(19) \quad k(r) = C(\sigma, d, \ell) \Big(\frac{r}{\sigma}\Big)^{\ell - \frac{d}{2}} K_{\ell - \frac{d}{2}}\Big(\frac{r}{\sigma}\Big), \quad h(\varrho) = \frac{1}{(1 + 4\sigma^2\pi^2\varrho^2)^\ell},$$

where K_ν is a modified Bessel function of order ν. In cases (17) and (18), we have $\mu := d/2 - 1$ and the functions h defined by (15) are computed using [52, §1.5.9] and [52, §1.4.13], respectively. In (19), which we had already seen at the end of section 2, $C(\sigma, d, \ell) := (2^{\ell + \frac{d}{2} - 1} \pi^{\frac{d}{2}} \Gamma(\ell) \sigma^d)^{-1}$, with $\ell \geq s + d/2$ (for the RKHS to be s-admissible). To compute h in (19) we observe that since $k(\|\cdot\|)$ is the Green's function of the differential operator $L = (1 - \sigma^2\Delta)^\ell$, then its Fourier transform $h(\|\cdot\|)$ is $1/\widehat{L}$; here $\widehat{L}(\xi) = (1 + 4\sigma^2\pi^2\|\xi\|^2)^\ell$, $\xi \in \mathbb{R}^d$, is the Fourier transform of L, intended as a distribution. Note that in all three cases above we have $h \geq 0$. We refer the reader to [69] for further examples and techniques to build scalar kernels.

To generalize the above result we first compute the Fourier transform of matrix-valued functions $\mathbf{k}$ of the type (10), that are not necessarily kernels. First note that for all $R \in \mathrm{O}(d, \mathbb{R})$ we have

$$\widehat{\mathbf{k}}(R\xi) = \int_{\mathbb{R}^d} \mathbf{k}(x)\, e^{-2\pi i x \cdot R\xi}\, dx = \int_{\mathbb{R}^d} \mathbf{k}(Ry)\, e^{-2\pi i y \cdot \xi}\, dy$$

$$\overset{(*)}{=} \int_{\mathbb{R}^d} R\,\mathbf{k}(y) R^T\, e^{-2\pi i y \cdot \xi}\, dy = R\,\widehat{\mathbf{k}}(\xi) R^T;$$

in $(*)$ we have used the property (9), which therefore also holds for $\widehat{\mathbf{k}}$. Whence we may write $\widehat{\mathbf{k}}$ as

$$(20) \qquad \widehat{\mathbf{k}}(\xi) = h^\|(\|\xi\|) \Pr_\xi^\| + h^\perp(\|\xi\|) \Pr_\xi^\perp, \quad \xi \in \mathbb{R}^d \setminus \{0\},$$

for some functions $h^\|, h^\perp : \mathbb{R}^+ \to \mathbb{R}$, by Lemma 3.2. When $\mathbf{k} \in L^1(\mathbb{R}^d, \mathbb{R}^{d\times d})$ is of the type (10) we have that the coefficients $k^\|, k^\perp$ are in $L^1(\mathbb{R}^+, r^{d-1})$ whereas, once again by the properties of Fourier transforms, the functions $h^\|$ and $h^\perp$ are continuous and vanish at infinity.

Definition 3.10. *We call the functions $h^\|, h^\perp : \mathbb{R}^+ \to \mathbb{R}$ in expression (20) the* coefficients *of $\widehat{\mathbf{k}}$.*

The above definition is in line with Definition 3.3; while this may be a bit redundant, we want to associate the symbols $(k^\parallel, k^\perp)$ with $\mathbf{k}$ and $(h^\parallel, h^\perp)$ with its Fourier transform $\widehat{\mathbf{k}}$. The coefficients of $\widehat{\mathbf{k}}$ are expressed in terms of those of $\mathbf{k}$ by the following result, which generalizes formula (15).

Theorem 3.11. *Consider a function* $\mathbf{k} \in L^1(\mathbb{R}^d, \mathbb{R}^{d \times d})$ *of the type* (10). *Its Fourier transform* $\widehat{\mathbf{k}}$ *is a matrix-valued function of the form* (20), *where* $h^\parallel, h^\perp : \mathbb{R}^+ \to \mathbb{R}$ *are the scalar functions:*

$$(21a) \qquad h^\parallel(\varrho) = \frac{2\pi}{\varrho^\mu} \int_0^\infty r^{\mu+1} k^\parallel(r) J_\mu(2\pi\varrho r)\, dr$$

$$- \frac{2\mu+1}{\varrho^{\mu+1}} \int_0^\infty r^\mu \big(k^\parallel(r) - k^\perp(r)\big) J_{\mu+1}(2\pi\varrho r)\, dr,$$

$$(21b) \qquad h^\perp(\varrho) = \frac{2\pi}{\varrho^\mu} \int_0^\infty r^{\mu+1} k^\perp(r) J_\mu(2\pi\varrho r)\, dr$$

$$+ \frac{1}{\varrho^{\mu+1}} \int_0^\infty r^\mu \big(k^\parallel(r) - k^\perp(r)\big) J_{\mu+1}(2\pi\varrho r)\, dr,$$

both defined for $\varrho > 0$, *where* $\mu := \frac{d}{2} - 1$.

Proof. Define $\tilde{k}$ as in (11). The Fourier transform of the element (12) of $\mathbf{k}$ is

$$\widehat{\mathbf{k}}_{j\ell}(\xi) = \int_{\mathbb{R}^d} \mathbf{k}_{j\ell}(x) e^{-2\pi i x \cdot \xi} dx$$

$$= \int_{\mathbb{R}^d} \tilde{k}(\|x\|)\, x_j x_\ell\, e^{-2\pi i x \cdot \xi} dx + \delta_{j\ell} \int_{\mathbb{R}^d} k^\perp(\|x\|)\, e^{-2\pi i x \cdot \xi} dx$$

$$(22) \qquad = -\frac{1}{(2\pi)^2} \frac{\partial^2}{\partial \xi_\ell \partial \xi_j} \widetilde{F}(\xi) + \delta_{j\ell} F^\perp(\xi), \qquad \xi \in \mathbb{R}^d,$$

where $\widetilde{F}, F^\perp : \mathbb{R}^d \to \mathbb{C}$ are, respectively, the Fourier transforms of $\tilde{k}(\|\cdot\|)$ and $k^\perp(\|\cdot\|)$. It is convenient at this point to introduce the auxiliary function $\widetilde{J}_\nu(z) := z^{-\nu} J_\nu(z)$, which has the property $\frac{d}{dz}\widetilde{J}_\nu(z) = -z\widetilde{J}_{\nu+1}(z)$ (see [1, §9.1.30]). By Theorem 3.9 we have

$$\widetilde{F}(\xi) = \frac{2\pi}{\|\xi\|^\mu} \int_0^\infty r^{\mu+1} \tilde{k}(r) J_\mu(2\pi r \|\xi\|)\, dr = (2\pi)^{\mu+1} \int_0^\infty r^{2\mu+1} \tilde{k}(r) \widetilde{J}_\mu(2\pi r \|\xi\|)\, dr,$$

so that

$$\frac{\partial \widetilde{F}}{\partial \xi_j}(\xi) = (2\pi)^{\mu+2}\frac{\xi_j}{\|\xi\|}\int_0^\infty r^{2\mu+2}\tilde{k}(r)\widetilde{J}'_\mu(2\pi r\|\xi\|)\,dr$$

$$= -(2\pi)^{\mu+3}\xi_j\int_0^\infty r^{2\mu+3}\tilde{k}(r)\widetilde{J}_{\mu+1}(2\pi r\|\xi\|)\,dr.$$

Therefore the second partial derivatives of $\widetilde{F}$ are given by:

$$\frac{\partial^2\widetilde{F}}{\partial\xi_\ell\partial\xi_j}(\xi) = -\delta_{j\ell}(2\pi)^{\mu+3}\int_0^\infty r^{2\mu+3}\tilde{k}(r)\widetilde{J}_{\mu+1}(2\pi r\|\xi\|)\,dr$$

$$-(2\pi)^{\mu+4}\frac{\xi_j\xi_\ell}{\|\xi\|}\int_0^\infty r^{2\mu+4}\tilde{k}(r)\widetilde{J}'_{\mu+1}(2\pi r\|\xi\|)\,dr$$

$$= -\delta_{j\ell}(2\pi)^{\mu+3}\int_0^\infty r^{2\mu+3}\tilde{k}(r)\widetilde{J}_{\mu+1}(2\pi r\|\xi\|)\,dr$$

$$+ (2\pi)^{\mu+5}\xi_j\xi_\ell\int_0^\infty r^{2\mu+5}\tilde{k}(r)\widetilde{J}_{\mu+2}(2\pi r\|\xi\|)\,dr,$$

which, inserting $\widetilde{J}_\nu(z) = z^{-\nu}J_\nu(z)$, may be rewritten as

$$\frac{\partial^2\widetilde{F}}{\partial\xi_\ell\partial\xi_j}(\xi) = -\delta_{j\ell}\frac{(2\pi)^2}{\|\xi\|^{\mu+1}}\int_0^\infty r^{\mu+2}\,\tilde{k}(r)J_{\mu+1}(2\pi r\|\xi\|)\,dr$$

$$+ \frac{(2\pi)^3}{\|\xi\|^\mu}\frac{\xi_j\xi_\ell}{\|\xi\|^2}\int_0^\infty r^{\mu+3}\,\tilde{k}(r)J_{\mu+2}(2\pi r\|\xi\|)\,dr.$$

Computing $F^\perp$ via Proposition 3.9 and inserting it into (22) yields:

$$\widehat{\mathbf{k}}_{j\ell}(\xi) = C(\|\xi\|)\frac{\xi_j\xi_\ell}{\|\xi\|^2} + D(\|\xi\|)\delta_{j\ell}$$

(23)
$$= \Big(C(\|\xi\|) + D(\|\xi\|)\Big)\frac{\xi_j\xi_\ell}{\|\xi\|^2} + D(\|\xi\|)\Big(\delta_{j\ell} - \frac{\xi_j\xi_\ell}{\|\xi\|^2}\Big),$$

with $C(\varrho) := -\dfrac{2\pi}{\varrho^\mu}\displaystyle\int_0^\infty r^{\mu+3}\,\tilde{k}(r)J_{\mu+2}(2\pi r\varrho)\,dr$, and $D(\varrho) := h^\perp(\varrho)$, i.e. ex-

pression (21b). Finally,

$$
\begin{aligned}
C(\varrho) &+ D(\varrho) \\
&\overset{(**)}{=} -\frac{2\pi}{\varrho^\mu}\int_0^\infty r^{\mu+1}\big(k^\parallel(r)-k^\perp(r)\big)\Big\{\frac{2(\mu+1)}{2\pi r\varrho}J_{\mu+1}(2\pi r\varrho)-J_\mu(2\pi r\varrho)\Big\}\,dr \\
&+\frac{2\pi}{\varrho^\mu}\int_0^\infty r^{\mu+1}k^\perp(r)J_\mu(2\pi\varrho r)\,dr+\frac{1}{\varrho^{\mu+1}}\int_0^\infty r^\mu\big(k^\parallel(r)-k^\perp(r)\big)J_{\mu+1}(2\pi\varrho r)\,dr \\
&=\frac{2\pi}{\varrho^\mu}\int_0^\infty r^{\mu+1}\,k^\parallel(r)J_\mu(2\pi\varrho r)\,dr-\frac{2\mu+1}{\varrho^{\mu+1}}\int_0^\infty r^\mu\big(k^\parallel(r)-k^\perp(r)\big)J_{\mu+1}(2\pi\varrho r)\,dr,
\end{aligned}
$$

which is $h^\parallel(\varrho)$ by (21a); in $(**)$ we have used the property (74), reported in Appendix A, of Bessel functions of the first kind. Rewriting equation (23) in matrix form using definitions (8) completes the proof. $\qquad\square$

Corollary 3.12. *Under the assumptions of Theorem 3.11, by Bochner's theorem we have that* **k** *is positive definite, i.e. it is a kernel, if and only if* $h^\parallel$ *and* $h^\perp$ *are nonnegative functions.*

As we said above, formulae (21a) and (21b) generalize (15): in fact the three coincide when $k^\parallel = k^\perp$, i.e. for scalar kernels, in which case we also have $h^\parallel = h^\perp$. By the form of equation (20), we have that, for a fixed $\xi \in \mathbb{R}^d \setminus \{0\}$, the numbers $h^\parallel(\|\xi\|)$ and $h^\perp(\|\xi\|)$ are the eigenvalues of the matrix $\widehat{\mathbf{k}}(\xi)$; the former has multiplicity 1 and eigenvector ξ, while the latter has multiplicity $d-1$. While the Fourier transform of a kernel **k** may be written in the form (20) it is not itself a "kernel", in that it is not positive definite (but just the Fourier transform of a positive definite function).

The two functions $h^\parallel$ and $h^\perp$ may be expressed in terms of Hankel transforms, as it is the case for formula (15) in the scalar case. In fact in the examples that we shall work out we will write:

$$
(24)\qquad h^\parallel(\varrho)=\frac{2\pi}{\varrho^\mu}\,H^\parallel(2\pi\varrho),\qquad\text{and}\qquad h^\perp(\varrho)=\frac{2\pi}{\varrho^\mu}\,H^\perp(2\pi\varrho),
$$

where:

$$
(25a)\quad H^\parallel(\varrho):=\int_0^\infty r^{\mu+1}k^\parallel(r)J_\mu(\varrho r)\,dr-\frac{2\mu+1}{\varrho}\int_0^\infty r^{\mu+2}\tilde{k}(r)J_{\mu+1}(\varrho r)\,dr,
$$

$$
(25b)\quad H^\perp(\varrho):=\int_0^\infty r^{\mu+1}k^\perp(r)J_\mu(\varrho r)\,dr+\frac{1}{\varrho}\int_0^\infty r^{\mu+2}\tilde{k}(r)J_{\mu+1}(\varrho r)\,dr.
$$

The following proposition, proven in Appendix B, ensures that for any given TRI kernel **k** we have that as long as either $h^\parallel$ or $h^\perp$ are strictly

positive somewhere (i.e. we are not dealing with the trivial case $\mathbf{k} = 0$) the corresponding RKHS is in fact *non-degenerate* (see Definition 2.2).

Proposition 3.13. *Let* $\mathbf{k} \in L^1(\mathbb{R}^d, \mathbb{R}^{d \times d})$ *be a TRI kernel such that* $\widehat{\mathbf{k}} \in L^1(\mathbb{R}^d, \mathbb{R}^{d \times d})$. *Then* $\mathbf{k}$ *is* strictly *positive definite if and only if there exists* $r_0 > 0$ *such that either* $h^{\parallel}(r_0) > 0$ *or* $h^{\perp}(r_0) > 0$.

A non-example: Gaussian $k^{\parallel}$ and $k^{\perp}$. One may ask whether we can construct a positive definite translation- and rotation-invariant kernel $\mathbf{k}$ of the type (10) with coefficients given by $k^{\parallel}(r) = \exp(-\frac{1}{2} r^2/\sigma_1^2)$ and $k^{\perp}(r) = \exp(-\frac{1}{2} r^2/\sigma_2^2)$, where $\sigma_1 > 0$ and $\sigma_2 > 0$. The answer is that a kernel of this type is positive definite if an only if $\sigma_1 = \sigma_2 =: \sigma$, which makes it of the form $\mathbf{k}(x) = \exp(-\frac{1}{2}\|x\|/\sigma^2)\mathbb{I}_d$, $x \in \mathbb{R}^d$, i.e. *scalar* and of the type (17). To see this, we compute $h^{\parallel}$, $h^{\perp}$ and impose their nonnegativity.

We shall use the following Hankel transforms [52, §1.5.9 & §1.5.10]:

$$(26) \quad \int_0^\infty r^{\nu+1} e^{-cr^2} J_\nu(\varrho r)\, dr = \frac{\varrho^\nu}{(2c)^{\nu+1}} \exp\left(-\frac{\varrho^2}{4c}\right), \quad \text{for } \Re\nu > -1,\ \Re c > 0,$$

$$(27) \quad \int_0^\infty r^{\nu-1} e^{-cr^2} J_\nu(\varrho r)\, dr = \frac{2^{\nu-1}}{\varrho^\nu} \gamma\left(\nu, \frac{\varrho^2}{4c}\right), \qquad \text{for } \Re\nu > 0,\ \Re c > 0,$$

where $\gamma(\nu, x) := \int_0^x e^{-t}\, t^{\nu-1} dt$, $\Re\nu > 0$, is the *lower* incomplete gamma function [1, §6.5.2]. Therefore when the coefficients are $k^{\parallel}(r) = \exp(-\frac{1}{2} r^2/\sigma_1^2)$ and $k^{\perp}(r) = \exp(-\frac{1}{2} r^2/\sigma_2^2)$ the expressions (25a) and (25b) become

$$H^{\parallel}(\varrho) = (\sigma_1^2)^{\mu+1} \varrho^\mu \exp\left(-\frac{\sigma_1^2 \varrho^2}{2}\right) - \frac{2^\mu (2\mu+1)}{\varrho^{\mu+2}} \left[\gamma\left(\mu+1, \frac{\sigma_1^2 \varrho^2}{2}\right) - \gamma\left(\mu+1, \frac{\sigma_2^2 \varrho^2}{2}\right)\right],$$

$$H^{\perp}(\varrho) = (\sigma_2^2)^{\mu+1} \varrho^\mu \exp\left(-\frac{\sigma_2^2 \varrho^2}{2}\right) + \frac{2^\mu}{\varrho^{\mu+2}} \left[\gamma\left(\mu+1, \frac{\sigma_1^2 \varrho^2}{2}\right) - \gamma\left(\mu+1, \frac{\sigma_2^2 \varrho^2}{2}\right)\right].$$

Using formulae (24) and introducing for later convenience the *upper* incomplete gamma function $\Gamma(\nu, x) := \Gamma(\nu) - \gamma(\nu, x) = \int_x^\infty e^{-t} t^{\nu-1} dt$, $\Re\nu > 0$ (see [1, §6.5.3]), we finally get:

$$(28a) \quad h^{\parallel}(\varrho) = (2\pi\sigma_1^2)^{\mu+1} \exp\left(-2\pi^2 \sigma_1^2 \varrho^2\right)$$
$$- \frac{2\mu+1}{2\pi^{\mu+1} \varrho^{2\mu+2}} \left[\Gamma\left(\mu+1, 2\pi^2 \sigma_2^2 \varrho^2\right) - \Gamma\left(\mu+1, 2\pi^2 \sigma_1^2 \varrho^2\right)\right],$$

$$(28b) \quad h^{\perp}(\varrho) = (2\pi\sigma_2^2)^{\mu+1} \exp\left(-2\pi^2 \sigma_2^2 \varrho^2\right)$$
$$+ \frac{1}{2\pi^{\mu+1} \varrho^{2\mu+2}} \left[\Gamma\left(\mu+1, 2\pi^2 \sigma_2^2 \varrho^2\right) - \Gamma\left(\mu+1, 2\pi^2 \sigma_1^2 \varrho^2\right)\right].$$

We first consider the case $d = 2$ (i.e. $\mu = 0$) since it is particularly simple. In fact $\Gamma(1, x) = e^{-x}$ and elementary manipulations lead to the following

expressions for $h^\parallel$ and $h^\perp$:

$$h^\parallel(\varrho) = \tfrac{1}{2\pi\varrho^2}\big[\big(4\pi^2\sigma_1^2\varrho^2 + 1\big)\exp\big(-2\sigma_1^2\pi^2\varrho^2\big) - \exp\big(-2\sigma_2^2\pi^2\varrho^2\big)\big],$$
$$h^\perp(\varrho) = \tfrac{1}{2\pi\varrho^2}\big[\big(4\pi^2\sigma_2^2\varrho^2 + 1\big)\exp\big(-2\sigma_2^2\pi^2\varrho^2\big) - \exp\big(-2\sigma_1^2\pi^2\varrho^2\big)\big],$$

i.e. each is obtained from the other by exchanging σ_1 and σ_2. The condition $h^\parallel \geq 0$ is equivalent to

$$(29) \qquad 4\pi^2\sigma_1^2\varrho^2 + 1 \geq \exp\big(2\pi^2(\sigma_1^2 - \sigma_2^2)\varrho^2\big) \qquad \text{for all } \varrho \geq 0.$$

This is certainly true if $\sigma_1 \leq \sigma_2$ (in which case the graph of the function of ϱ on the right-hand side of the above inequality is a Gaussian curve); in fact this is also necessary for (29) to hold, since if $\sigma_1 > \sigma_2$ then the condition breaks down for large values of ϱ. Similarly, $h^\perp \geq 0$ if and only if $\sigma_2 \leq \sigma_1$, so the functions $h^\parallel$ and $h^\perp$ are both nonnegative if and only if $\sigma_1 = \sigma_2$.

To prove that this is also true in higher dimensions ($d > 2$, or $\mu > 0$) we rewrite (28a) as follows:

$$h^\parallel(\varrho) = (2\pi\sigma_1^2)^{\mu+1}\exp\big(-2\pi^2\sigma_1^2\varrho^2\big) - \frac{2\mu+1}{2}\pi^{\mu+1}\int_{2\sigma_2^2}^{2\sigma_1^2}\tau^\mu e^{-\tau(\pi\varrho)^2}d\tau,$$

which is obviously nonnegative if $\sigma_1 \leq \sigma_2$. The latter condition is in fact also necessary for the nonnegativity of $h^\parallel$. To see this, we use the asymptotic expansion of the upper incomplete gamma function:

$$\Gamma(\nu, x) \sim x^{\nu-1}e^{-x}\Big[1 + \frac{\nu-1}{x} + \frac{(\nu-1)(\nu-2)}{x^2} + \cdots\Big], \qquad \text{as } x \to \infty,$$

(see [1, §6.5.31]) so that for any $\sigma > 0$ it is the case that

$$\frac{1}{2\pi^{\mu+1}\varrho^{2\mu+2}}\Gamma\big(\mu+1, 2\pi^2\sigma^2\varrho^2\big)$$
$$\sim \frac{(2\sigma^2)^\mu\pi^{\mu+1}}{2\varrho^2}\exp\big(-2\pi^2\sigma^2\varrho^2\big)\Big[1 + \frac{\mu}{2\pi^2\sigma^2\varrho^2} + \frac{\mu(\mu-1)}{(2\pi^2\sigma^2\varrho^2)^2} + \cdots\Big]$$

as $\varrho \to \infty$ (incidentally, note that if the dimension d is even then μ is integer and the asymptotic expansion on the right-hand side has a finite number of

terms), so (28a) can be expanded at infinity as follows:

$$h^{\parallel}(\varrho) \sim (2\mu+1)\frac{(2\sigma_1^2)^{\mu}\pi^{\mu-1}}{2\varrho^2}\exp\left(-2\pi^2\sigma_1^2\varrho^2\right)\left\{\left[\frac{4\pi^2\sigma_1^2\varrho^2}{2\mu+1}+1+\frac{\mu}{2\pi^2\sigma_1^2\varrho^2}+\cdots\right]\right.$$
$$\left.-\left(\frac{\sigma_2^2}{\sigma_1^2}\right)^{\mu}\exp\left(2\pi^2(\sigma_1^2-\sigma_2^2)\varrho^2\right)\left[1+\frac{\mu}{2\pi^2\sigma_2^2\varrho^2}+\cdots\right]\right\}, \quad \text{as } \varrho\to\infty.$$

If $\sigma_1 > \sigma_2$ the second exponential function diverges faster than ϱ^2, implying that $h^{\parallel}$ is negative for large values of ϱ. Therefore $h^{\parallel} \geq 0$ if and only if $\sigma_1 \leq \sigma_2$. Manipulating (28b) in a similar manner proves that $h^{\perp} \geq 0$ if and only if $\sigma_2 \leq \sigma_1$. In conclusion, a kernel $\mathbf{k}$ with Gaussian $k^{\parallel}$ and $k^{\perp}$ is positive definite if and only if $\sigma_1 = \sigma_2$, i.e. if and only if it is *scalar*.

Example 1. We now try a different route and verify for which values of $(a,b) \in \mathbb{R}^2$ the functions

$$(30) \quad k^{\parallel}(r) = b\exp\left(-\frac{1}{2}\frac{r^2}{\sigma^2}\right) \text{ and } k^{\perp}(r) = (b-ar^2)\exp\left(-\frac{1}{2}\frac{r^2}{\sigma^2}\right), \quad r > 0,$$

with $\sigma > 0$, define a *positive definite* $\mathbf{k}$ of the type (10), i.e. a TRI kernel. In other words we assume that the function (11) is given by $\tilde{k}(r) = a\exp(-\frac{1}{2}r^2/\sigma^2)$, so that both $\tilde{k}$ and $k^{\parallel}$ are Gaussian with the *same* variance, but different values at $r = 0$ (on the other hand $k^{\parallel}$ and $k^{\perp}$ must have the same value at 0 by Proposition 3.7). In this example we have that

$$H^{\parallel}(\varrho) = (\sigma^2)^{\mu+1}\varrho^{\mu}\left[b-(2\mu+1)a\sigma^2\right]\exp\left(-\frac{\sigma^2\varrho^2}{2}\right),$$
$$H^{\perp}(\varrho) = (\sigma^2)^{\mu+1}\varrho^{\mu}\left[b-(2\mu+1)a\sigma^2+a(\sigma^2\varrho)^2\right]\exp\left(-\frac{\sigma^2\varrho^2}{2}\right).$$

In order to obtain the above expressions we have used the Hankel transform:

$$\int_0^{\infty} r^{\mu+3}e^{-cr^2}J_{\mu}(\varrho r)\,dr = \frac{\varrho^{\mu}}{(2c)^{\mu+3}}\left[4c(\mu+1)-\varrho^2\right]\exp\left(-\frac{\varrho^2}{4c}\right),$$

valid for $\Re\mu > -2$ and $\Re c > 0$ (it is computed by inserting in the left-hand side the relation $J_{\mu}(z) = 2(\mu+1)J_{\mu+1}(z)/z - J_{\mu+2}(z)$, given by (74), and applying (26) twice, with $\nu = \mu+1$ and $\nu = \mu+2$ respectively). By (24),

$$h^{\parallel}(\varrho) = (2\pi\sigma^2)^{\mu+1}\left[b-(2\mu+1)a\sigma^2\right]\exp\left(-2\pi^2\sigma^2\varrho^2\right),$$
$$h^{\perp}(\varrho) = (2\pi\sigma^2)^{\mu+1}\left[b-(2\mu+1)a\sigma^2+a(2\pi\sigma^2\varrho)^2\right]\exp\left(-2\pi^2\sigma^2\varrho^2\right).$$
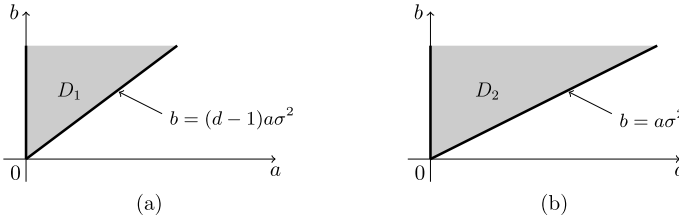
Figure 1: Domains D_1 and D_2 of positive definiteness for the kernels in Examples 1 and 2.

So $h^{\parallel} \geq 0$ if and only if $b - (2\mu + 1)a\sigma^2 \geq 0$, and $h^{\perp} \geq 0$ if and only if $a(2\pi\sigma^2\varrho)^2 \geq -[b - (2\mu + 1)a\sigma^2]$ for all ϱ. Whence $h^{\parallel}$ and $h^{\perp}$ are *simultaneously* nonnegative, i.e. $\mathbf{k}$ is positive definite, if and only if (a, b) is in

$$D_1 := \{(a, b) \in \mathbb{R}^2 \,|\, b \geq (d - 1)a\sigma^2, a \geq 0\}$$

(since $2\mu + 1 = d - 1$) which is the wedge-shaped domain illustrated in Figure 1(a). Note that the slope of the slanted boundary depends on the dimension d. The vertical boundary ($a = 0$) corresponds to scalar kernels, while the meaning of the other one shall be explored in later sections.

The graphs $k^{\parallel}$, $k^{\perp}$, $h^{\parallel}$ and $h^{\perp}$ (symmetrized with respect to $r = 0$ for clarity) are shown in Figure 2 for $d = 2$, $a = 1.5$ and $b = 2\sigma^2 = 1$. Note that $h^{\parallel} \geq 0$ and $h^{\perp} \geq 0$. The same figure shows the corresponding vector field $x \mapsto \mathbf{k}(x)\alpha$, $x \in \mathbb{R}^2$, with $\mathbf{k}$ given by (10) and $\alpha = e_1 \in \mathbb{R}^2$, also shown in the figure (the importance of vector fields of the type $x \mapsto \mathbf{k}(x)\alpha$ is apparent from equation (1), on which we shall return in section 5). For $\alpha = e_1$ the two "vortices" of the vector field $x \mapsto \mathbf{k}(x)\alpha$ are located at points $(0, \pm\sqrt{b/a})$; so for positive definite $\mathbf{k}$ with coefficients (30) their distance is at least 2σ.

Example 2. We now modify the previous example and switch the roles of $k^{\parallel}$ and $k^{\perp}$. In other words we want to find conditions on $(a, b) \in \mathbb{R}^2$ such that the functions

$$(31) \quad k^{\parallel}(r) = (b - ar^2) \exp\left(-\frac{1}{2}\frac{r^2}{\sigma^2}\right) \quad \text{and} \quad k^{\perp}(r) = b \exp\left(-\frac{1}{2}\frac{r^2}{\sigma^2}\right), \quad r > 0,$$

with fixed $\sigma > 0$, define a positive definite $\mathbf{k}$ of the type (10), i.e. a TRI kernel. So we assume that $\tilde{k}(r) = -a \exp(-\frac{1}{2}r^2/\sigma^2)$, hence both $\tilde{k}$ and $k^{\perp}$ are Gaussian, with the same variance. Calculations that are similar to those
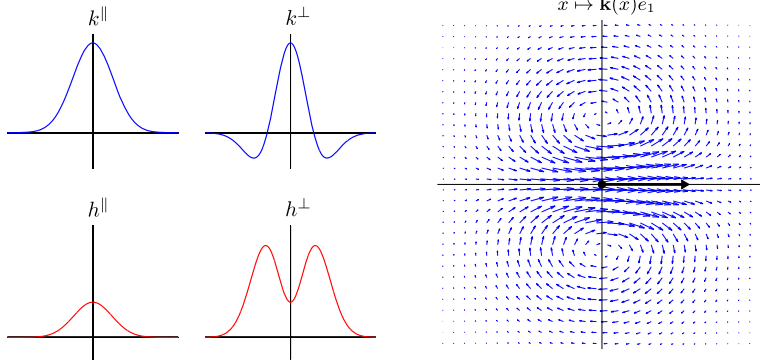
Figure 2: Graphs of $k^{\parallel}$, $k^{\perp}$, $h^{\parallel}$, $h^{\perp}$ for Example 1 as well as the vector field $x \mapsto \mathbf{k}(x)\alpha$, with $\alpha = e_1$ (also shown). Note that $h^{\parallel} \geq 0$ and $h^{\perp} \geq 0$.

of the previous example lead to the following expressions for $h^{\parallel}$ and $h^{\perp}$:

$$h^{\parallel}(\varrho) = (2\pi\sigma^2)^{\mu+1}[b - a\sigma^2 + a(2\pi\sigma^2\varrho)^2]\exp\!\big(-2\pi^2\sigma^2\varrho^2\big),$$
$$h^{\perp}(\varrho) = (2\pi\sigma^2)^{\mu+1}(b - a\sigma^2)\exp\!\big(-2\pi^2\sigma^2\varrho^2\big).$$

The above are simultaneously nonnegative ϱ, i.e. $\mathbf{k}$ is a TRI kernel, if and only if (a, b) is in

$$D_2 := \big\{(a, b) \in \mathbb{R}^2 \,\big|\, b \geq a\sigma^2, a \geq 0\big\}$$

which is the wedge-shaped domain shown in Figure 1(b). The vertical boundary $(a = 0)$ again corresponds to scalar kernels, while the meaning of the other boundary (whose slope this time does *not* depend on the dimension d of the domain) will be explored later.

The graphs of the functions $k^{\parallel}$, $k^{\perp}$, $h^{\parallel}$ and $h^{\perp}$ (again, symmetrized with respect to $r = 0$) are shown in Figure 3 for $d = 2$, $a = 1.5$ and $b = 2\sigma^2 = 1$, as well as the corresponding vector field $x \mapsto \mathbf{k}(x)\alpha$, with $\alpha = e_1 \in \mathbb{R}^2$. For such α the apparent "sink" and "source" of the vector field (i.e. where $\mathbf{k}(x)\alpha = 0$) are located at points $(\pm\sqrt{b/a}, 0)$; that is, their mutual distance is at least 2σ for a positive definite $\mathbf{k}$ with coefficients (31).

3.3. Inversion formulae

Let M be the functional that maps the coefficients $(k^{\parallel}, k^{\perp})$ of a matrix-valued function $\mathbf{k} \in L^1(\mathbb{R}^d, \mathbb{R}^{d \times d})$ of the type (10) to the coefficients $(h^{\parallel}, h^{\perp})$
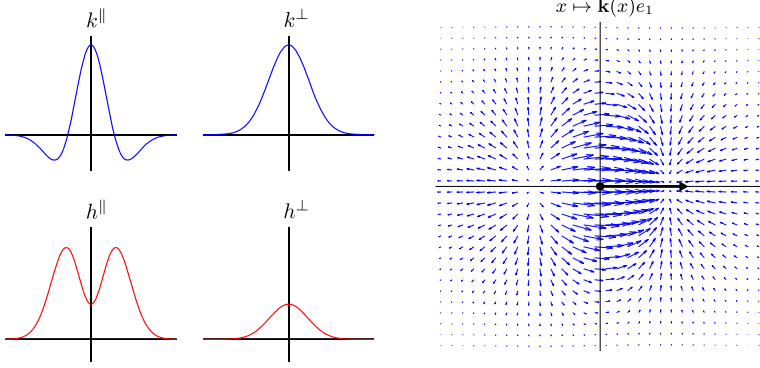
Figure 3: Graphs of $k^{\parallel}$, $k^{\perp}$, $h^{\parallel}$, $h^{\perp}$ for Example 2 as well as the vector field $x \mapsto \mathbf{k}(x)\alpha$, with $\alpha = e_1$ (also shown). Again, $h^{\parallel} \geq 0$ and $h^{\perp} \geq 0$.

of its Fourier transform. That is,

$$
(32) \qquad
\begin{aligned}
M : &\ L^1(\mathbb{R}^+, r^{d-1}) \times L^1(\mathbb{R}^+, r^{d-1}) \longrightarrow C_0(\mathbb{R}^+) \times C_0(\mathbb{R}^+), \\
\text{with} \quad &\ M(k^{\parallel}, k^{\perp}) = (h^{\parallel}, h^{\perp}),
\end{aligned}
$$

where $(h^{\parallel}, h^{\perp})$ are expressed in terms of $(k^{\parallel}, k^{\perp})$ precisely by formulae (21a) and (21b). The symbol $C_0(\mathbb{R}^+)$ indicates the set of real-valued continuous functions defined on $\mathbb{R}^+$ that vanish at $+\infty$. The map M is *linear* and it is an *involution*, i.e. $M^{-1} = M$, in the sense that is specified by the following proposition (this generalizes the fact that (15) is an involution).

Proposition 3.14. *Let* $\mathbf{k} \in L^1(\mathbb{R}^d, \mathbb{R}^{d \times d})$ *be of the type* (10), *and also assume* $\widehat{\mathbf{k}} \in L^1(\mathbb{R}^d, \mathbb{R}^{d \times d})$. *Then the functions* $(k^{\parallel}, k^{\perp})$ *can be computed from the pair* $(h^{\parallel}, h^{\perp})$ *as follows:*

$$
(33\text{a}) \qquad k^{\parallel}(r) = \frac{2\pi}{r^{\mu}} \int_0^{\infty} \varrho^{\mu+1}\, h^{\parallel}(\varrho) J_{\mu}(2\pi\varrho r)\, d\varrho
$$

$$
\qquad\qquad - \frac{2\mu+1}{r^{\mu+1}} \int_0^{\infty} \varrho^{\mu}\big(h^{\parallel}(\varrho) - h^{\perp}(\varrho)\big) J_{\mu+1}(2\pi\varrho r)\, d\varrho,
$$

$$
(33\text{b}) \qquad k^{\perp}(r) = \frac{2\pi}{r^{\mu}} \int_0^{\infty} \varrho^{\mu+1}\, h^{\perp}(\varrho) J_{\mu}(2\pi\varrho r)\, d\varrho
$$

$$
\qquad\qquad + \frac{1}{r^{\mu+1}} \int_0^{\infty} \varrho^{\mu}\big(h^{\parallel}(\varrho) - h^{\perp}(\varrho)\big) J_{\mu+1}(2\pi\varrho r)\, d\varrho;
$$

i.e. $(k^{\parallel}, k^{\perp})$ *are obtained from* $(h^{\parallel}, h^{\perp})$ *by applying formulae* (21a) *and* (21b).

Proof. The claim may be proven by direct computation, i.e. by manipulating formulae (21a)–(21b) and by inverting Hankel transforms (using formula (73) in Appendix A). However, one can simply observe that the Fourier transform, when restricted to functions that are symmetric in their argument, i.e. $\mathbf{k}(x) = \mathbf{k}(-x)$ for all x (and TRI kernels belong to this class), is an involution. But M computes the functions $(h^\parallel, h^\perp)$, which provide the eigenvalues of $\widehat{\mathbf{k}}$ in terms of those of $\mathbf{k}$, therefore it must be itself an involution. $\qquad\square$

The above proposition allows one to construct positive kernels by choosing nonnegative functions $h^\parallel$ and $h^\perp$ in $L^1(\mathbb{R}^+, r^{d-1})$ and applying the inversion formulae (33a) and (33b). This involves computing Hankel transforms, which may be done either by employing tables of transforms [20, 23, 52], or numerically. However, in section 4 we shall illustrate a very simple, constructive method for building arbitrary TRI kernels from *scalar* kernels.

Remark: L^1 vs. L^2. While it is well known that the Fourier transform is an isometry between L^2 and itself, we have assumed that the kernels are in L^1 (and that, sometimes, so are their Fourier transforms) to be able to use and manipulate the integral formulas in the definitions and proofs. We should note that one could employ the usual argument that L^1 is dense in L^2 [64] and extend the results, *mutatis mutandis*, to square integrable kernels. However, for computational convenience, we will keep assuming that the kernels are in L^1 when Fourier transforms are needed throughout the rest of the paper.

3.4. Divergence-free and curl-free kernels

In this section we explore some properties of the vector fields $x \mapsto \mathbf{k}(x)\alpha$ generated by the kernels $\mathbf{k}$ of RKHS. In particular we shall compute the functions $\mathrm{div}(\mathbf{k}(\cdot)\alpha)$ and $\mathrm{curl}(\mathbf{k}(\cdot)\alpha)$, and relate them to the coefficients $h^\parallel$ and $h^\perp$ of the Fourier transform of the kernel $\mathbf{k}$.

Proposition 3.15. *Let $\mathbf{k} \in C^1(\mathbb{R}^d, \mathbb{R}^{d \times d})$ be a TRI kernel. We have that*

$$\mathrm{div}\big(\mathbf{k}(x)\alpha\big) = \frac{\alpha \cdot x}{\|x\|}\left((d-1)\frac{k^\parallel(\|x\|) - k^\perp(\|x\|)}{\|x\|} + \frac{dk^\parallel}{dr}(\|x\|)\right), \quad x \in \mathbb{R}^d \setminus \{0\}.$$

Proof. For $x \neq 0$ define $\alpha^\parallel(x) := \mathrm{Pr}_x^\parallel \alpha$ and $\alpha^\perp(x) := \mathrm{Pr}_x^\perp \alpha = \alpha - \alpha^\parallel(x)$, so that $\alpha = \alpha^\parallel(x) + \alpha^\perp(x)$. It is the case that:

$$\operatorname{div}\big(\mathbf{k}(x)\alpha\big) = \operatorname{div}\big(k^{\|}(\|x\|)\,\alpha^{\|}(x)\big) + \operatorname{div}\big(k^{\perp}(\|x\|)\,\alpha^{\perp}(x)\big)$$

$$= \nabla\big(k^{\|}(\|x\|)\big)\cdot\alpha^{\|}(x) + k^{\|}(\|x\|)\operatorname{div}\big(\alpha^{\|}(x)\big)$$

$$+ \nabla\big(k^{\perp}(\|x\|)\big)\cdot\alpha^{\perp}(x) + k^{\perp}(\|x\|)\operatorname{div}\big(\alpha^{\perp}(x)\big)$$

$$(34) \quad = \frac{dk^{\|}}{dr}(\|x\|)\,\frac{x}{\|x\|}\cdot\alpha^{\|}(x) + k^{\|}(\|x\|)\operatorname{div}\big(\alpha^{\|}(x)\big) + k^{\perp}(\|x\|)\operatorname{div}\big(\alpha^{\perp}(x)\big),$$

where we have used $\frac{x}{\|x\|}\cdot\alpha^{\perp}(x) = 0$. Also, $\frac{x}{\|x\|}\cdot\alpha^{\|}(x) = \frac{x}{\|x\|}\cdot\alpha$ and

$$\operatorname{div}\big(\alpha^{\|}(x)\big) = \operatorname{div}\Big(\frac{\alpha\cdot x}{\|x\|^2}\,x\Big)$$

$$= \frac{1}{\|x\|^2}\big(\nabla(\alpha\cdot x)\big)\cdot x + (\alpha\cdot x)\Big(\nabla\frac{1}{\|x\|^2}\Big)\cdot x + \frac{\alpha\cdot x}{\|x\|^2}\operatorname{div}x$$

$$= \frac{\alpha\cdot x}{\|x\|^2} + (\alpha\cdot x)\Big(\frac{-2}{\|x\|^3}\frac{x}{\|x\|}\Big)\cdot x + \frac{\alpha\cdot x}{\|x\|^2}d = (d-1)\frac{\alpha\cdot x}{\|x\|^2}.$$

Moreover, $\operatorname{div}(\alpha^{\perp}(x)) = -\operatorname{div}(\alpha^{\|}(x))$. Inserting these expressions into (34) completes the proof. $\qquad\square$

Corollary 3.16. *Let $\mathbf{k} \in C^1(\mathbb{R}^d, \mathbb{R}^{d\times d})$ be a TRI kernel. The vector field $x \mapsto \mathbf{k}(x)\alpha$ is divergence-free if and only if*

$$(35) \qquad (d-1)\frac{k^{\|}(r)-k^{\perp}(r)}{r} + \frac{dk^{\|}}{dr}(r) = 0, \qquad \text{for all } r > 0.$$

We want to find a similar result for the *curl* of the vector field $x \mapsto \mathbf{k}(x)\alpha$. Indicating with $\Omega^k\mathbb{R}^d$ the space of differential k-forms in $\mathbb{R}^d$ [38], it is convenient to identify a vector field $(v_1, \ldots, v_d)$ in $\mathbb{R}^d$ with the 1-form $v = \sum v_i\, d_0 x^i$, where d_0 is the differential of a function. In fact indicating with d_k the exterior (Cartan) derivative of a k-form we may write the de Rahm complex for $\mathbb{R}^d$ as follows:

$$\Omega^0\mathbb{R}^d \xrightarrow{d_0} \Omega^1\mathbb{R}^d \xrightarrow{d_1} \Omega^2\mathbb{R}^d \xrightarrow{d_2} \cdots \xrightarrow{d_{d-2}} \Omega^{d-1}\mathbb{R}^d \xrightarrow{d_{d-1}} \Omega^d\mathbb{R}^d.$$

We define $\boxed{\operatorname{curl} := d_1}$ (it corresponds to the classical "curl" when $d = 3$).

Proposition 3.17. *Let $\mathbf{k} \in C^1(\mathbb{R}^d, \mathbb{R}^{d\times d})$ be a TRI kernel. We have*
$$(36)$$
$$\operatorname{curl}\big(\mathbf{k}(x)\alpha\big) = \frac{\alpha\wedge x}{\|x\|}\left(\frac{k^{\|}(\|x\|)-k^{\perp}(\|x\|)}{\|x\|} - \frac{dk^{\perp}}{dr}(\|x\|)\right), \quad x \in \mathbb{R}^d\setminus\{0\},$$

where $\wedge$ indicates the wedge product between differential forms.

Proof. As in the previous proof, let $\alpha^{\|}(x) := \mathrm{Pr}_x^{\|}\,\alpha$ and $\alpha^{\perp}(x) := \mathrm{Pr}_x^{\perp}\,\alpha = \alpha - \alpha^{\|}(x)$. We consider them 1-forms, i.e. let us write $\alpha^{\|}(x) = \sum_i \alpha^{\|}(x)_i\, d_0 x^i$ and $\alpha^{\perp}(x) = \sum_i \alpha^{\perp}(x)_i\, d_0 x^i$. It is the case that

$$\mathrm{curl}\big(\mathbf{k}(x)\alpha\big) = d_1\big(\mathbf{k}(x)\alpha\big) = d_0\big(k^{\|}(\|x\|)\big) \wedge \alpha^{\|}(x) + k^{\|}(\|x\|)\,d_1\big(\alpha^{\|}(x)\big)$$
$$\text{(37)} \qquad\qquad + d_0\big(k^{\perp}(\|x\|)\big) \wedge \alpha^{\perp}(x) + k^{\perp}(\|x\|)\,d_1\big(\alpha^{\perp}(x)\big).$$

The first term is zero since $d_0(k^{\|}(\|x\|)) = \frac{dk^{\|}}{dr}(\|x\|)\frac{x}{\|x\|}$ and $x \wedge \alpha^{\|}(x) = 0$. We also have

$$d_1\big(\alpha^{\|}(x)\big) = d_1\Big(\frac{\alpha \cdot x}{\|x\|^2}\,x\Big) = d_0\Big(\frac{\alpha \cdot x}{\|x\|^2}\Big) \wedge x + \frac{\alpha \cdot x}{\|x\|^2}\,d_1 x,$$

where $d_1 x = d_1\big(\sum_i x^i d_0 x^i\big) = \sum_i d_0 x^i \wedge d_0 x^i = 0$ and

$$d_0\Big(\frac{\alpha \cdot x}{\|x\|^2}\Big) = \frac{1}{\|x\|^2}d_0(\alpha \cdot x) + (\alpha \cdot x)\,d_0 \frac{1}{\|x\|^2} = \frac{\alpha}{\|x\|^2} - 2\frac{\alpha \cdot x}{\|x\|^4}\,x.$$

Since $x \wedge x = 0$ we conclude that $d_1\big(\alpha^{\|}(x)\big) = \frac{\alpha \wedge x}{\|x\|^2}$. Moreover $x \wedge \alpha^{\perp}(x) = x \wedge \alpha = -\alpha \wedge x$, whence

$$d_0\big(k^{\perp}(\|x\|)\big) \wedge \alpha^{\perp}(x) = \frac{dk^{\perp}}{dr}(\|x\|)\,\frac{x}{\|x\|} \wedge \alpha^{\perp}(x) = -\frac{dk^{\perp}}{dr}(\|x\|)\,\frac{\alpha \wedge x}{\|x\|}.$$

Finally, $d_1\big(\alpha^{\perp}(x)\big) = -d_1\big(\alpha^{\|}(x)\big)$. Insertion of such expressions into (37) completes the proof. $\qquad\square$

Corollary 3.18. *Let $\mathbf{k} \in C^1(\mathbb{R}^d, \mathbb{R}^{d\times d})$ be a TRI kernel. The vector field $x \mapsto \mathbf{k}(x)\alpha$ is curl-free in $\mathbb{R}^d$ if and only if*

$$\text{(38)} \qquad \frac{k^{\|}(r) - k^{\perp}(r)}{r} - \frac{dk^{\perp}}{dr}(r) = 0, \qquad \text{for all } r > 0.$$

Note that while condition (35) depends on the dimension d of the space, condition (38) does not. The following fundamental theorem relates the incompressibility and irrotationality of vector fields of the type $x \mapsto \mathbf{k}(x)\alpha$ with the coefficients $h^{\|}$ and $h^{\perp}$ of the Fourier transform of the kernel $\mathbf{k}$.

Theorem 3.19. *Let $\mathbf{k} \in C^1(\mathbb{R}^d, \mathbb{R}^{d\times d}) \cap L^1(\mathbb{R}^d, \mathbb{R}^{d\times d})$ be a TRI kernel with Fourier transform $\widehat{\mathbf{k}}$ that is also in $L^1(\mathbb{R}^d, \mathbb{R}^{d\times d})$. Then the vector field $x \mapsto \mathbf{k}(x)\alpha$ is divergence-free for all $\alpha \in \mathbb{R}^d$ if and only if $h^{\|} = 0$. On the other hand, $x \mapsto \mathbf{k}(x)\alpha$ is curl-free for all $\alpha \in \mathbb{R}^d$ if and only if $h^{\perp} = 0$.*

Proof. It is the case that $h^{\parallel}(\varrho) = 2\pi H^{\parallel}(2\pi\varrho)/\varrho^{\mu}$ and $h^{\perp}(\varrho) = 2\pi H^{\parallel}(2\pi\varrho)/\varrho^{\mu}$, with $H^{\parallel}$ and $H^{\perp}$ respectively given by (25a) and (25b), and with $\mu = \frac{d}{2} - 1$. Inserting (35) into (25a) yields

$$H^{\parallel}(\varrho) = \int_0^{\infty} r^{\mu+1} k^{\parallel}(r) J_{\mu}(\varrho r) \, dr + \frac{1}{\varrho} \int_0^{\infty} r^{\mu+1} \frac{dk^{\parallel}}{dr}(r) J_{\mu+1}(\varrho r) \, dr$$

(note that $2\mu + 1 = d - 1$). We use integration by parts to compute the second term on the right:

$$\int_0^{\infty} r^{\mu+1} \frac{dk^{\parallel}}{dr}(r) J_{\mu+1}(\varrho r) \, dr$$
$$= \left[r^{\mu+1} k^{\parallel}(r) J_{\mu+1}(\varrho r) \right]_{r=0}^{\infty} - \int_0^{\infty} k^{\parallel}(r) \frac{d}{dr} \left\{ r^{\mu+1} J_{\mu+1}(\varrho r) \right\} dr$$
$$= -\varrho \int_0^{\infty} r^{\mu+1} k^{\parallel}(r) \left\{ \frac{\mu+1}{\varrho r} J_{\mu+1}(\varrho r) + J'_{\mu+1}(\varrho r) \right\} dr$$
$$\text{(39)} \qquad = -\varrho \int_0^{\infty} r^{\mu+1} k^{\parallel}(r) J_{\mu}(\varrho r) \, dr,$$

where we have used the property (75) of Bessel functions. Therefore if (35) holds then $h^{\parallel} = 0$. Similarly, inserting (38) into (25b) gives

$$H^{\perp}(\varrho) = \int_0^{\infty} r^{\mu+1} k^{\perp}(r) J_{\mu}(\varrho r) \, dr + \frac{1}{\varrho} \int_0^{\infty} r^{\mu+1} \frac{dk^{\perp}}{dr}(r) J_{\mu+1}(\varrho r) \, dr.$$

and an identical integration by parts allows us to conclude that if (38) holds then $h^{\perp} = 0$.

Assume now $h^{\parallel} = 0$. In this case $k^{\parallel}$ and $k^{\perp}$, given by (33a), (33b), are

$$k^{\parallel}(r) = \frac{2\mu+1}{r^{\mu+1}} \int_0^{\infty} \varrho^{\mu} h^{\perp}(\varrho) J_{\mu+1}(2\pi\varrho r) \, d\varrho,$$
$$k^{\perp}(r) = \frac{2\pi}{r^{\mu}} \int_0^{\infty} \varrho^{\mu+1} h^{\perp}(\varrho) J_{\mu}(2\pi\varrho r) \, d\varrho - \frac{1}{r^{\mu+1}} \int_0^{\infty} \varrho^{\mu} h^{\perp}(\varrho) J_{\mu+1}(2\pi\varrho r) \, d\varrho,$$

therefore the auxiliary function (11) is given by

$$\tilde{k}(r) = \frac{2\pi}{r^{\mu+2}} \int_0^{\infty} \varrho^{\mu+1} h^{\perp}(\varrho) \left\{ 2 \frac{\mu+1}{2\pi\varrho r} J_{\mu+1}(2\pi\varrho r) - J_{\mu}(2\pi\varrho r) \right\} d\varrho$$
$$= \frac{2\pi}{r^{\mu+2}} \int_0^{\infty} \varrho^{\mu+1} h^{\perp}(\varrho) J_{\mu+2}(2\pi\varrho r) \, d\varrho,$$

where we have used property (74) in Appendix A. On the other hand,

$$\frac{1}{r}\frac{dk^{\|}}{dr}(r) = -(2\mu+1)\frac{\mu+1}{r^{\mu+3}}\int_0^\infty \varrho^\mu h^\perp(\varrho)J_{\mu+1}(2\pi\varrho r)\,d\varrho$$

$$+ \frac{2\mu+1}{r^{\mu+2}}\int_0^\infty 2\pi\varrho^{\mu+1}h^\perp(\varrho)J'_{\mu+1}(2\pi\varrho r)\,d\varrho$$

$$= -2\pi\frac{2\mu+1}{r^{\mu+2}}\int_0^\infty \varrho^{\mu+1}h^\perp(\varrho)\Big\{\frac{\mu+1}{2\pi\varrho r}J_{\mu+2}(2\pi\varrho r)-J'_{\mu+1}(2\pi\varrho r)\Big\}d\varrho$$

$$= -2\pi\frac{2\mu+1}{r^{\mu+2}}\int_0^\infty \varrho^{\mu+1}h^\perp(\varrho)J_{\mu+2}(2\pi\varrho r)d\varrho,$$

where we have used property (76) in Appendix A. Since $2\mu + 1 = d - 1$ we have that (35) holds and the vector field $x \mapsto \mathbf{k}(x)\alpha$ is divergence-free. Similarly, when $h^\perp = 0$ it is the case that

$$k^{\|}(r) = \frac{2\pi}{r^\mu}\int_0^\infty \varrho^{\mu+1}h^{\|}(\varrho)J_\mu(2\pi\varrho r)\,d\varrho - \frac{2\mu+1}{r^{\mu+1}}\int_0^\infty \varrho^\mu h^{\|}(\varrho)J_{\mu+1}(2\pi\varrho r)\,d\varrho,$$

$$k^\perp(r) = \frac{1}{r^{\mu+1}}\int_0^\infty \varrho^\mu h^{\|}(\varrho)J_{\mu+1}(2\pi\varrho r)\,d\varrho,$$

so that $$\tilde{k}(r) = \frac{2\pi}{r^{\mu+2}}\int_0^\infty \varrho^{\mu+1}h^{\|}(\varrho)\Big\{J_\mu(2\pi\varrho r) - 2\frac{\mu+1}{2\pi\varrho r}J_{\mu+1}(2\pi\varrho r)\Big\}\,d\varrho$$

$$= -\frac{2\pi}{r^{\mu+2}}\int_0^\infty \varrho^{\mu+1}h^\perp(\varrho)J_{\mu+2}(2\pi\varrho r)\,d\varrho,$$

by property (74) in Appendix A. On the other hand,

$$\frac{1}{r}\frac{dk^\perp}{dr}(r)$$

$$= -\frac{\mu+1}{r^{\mu+3}}\int_0^\infty \varrho^\mu h^{\|}(\varrho)J_{\mu+1}(2\pi\varrho r)\,d\varrho + \frac{1}{r^{\mu+2}}\int_0^\infty 2\pi\varrho^{\mu+1}h^{\|}(\varrho)J'_{\mu+1}(2\pi\varrho r)\,d\varrho$$

$$= \frac{2\pi}{r^{\mu+2}}\int_0^\infty \varrho^{\mu+1}h^{\|}(\varrho)\Big\{-\frac{\mu+1}{2\pi\varrho r}J_{\mu+2}(2\pi\varrho r) + J'_{\mu+1}(2\pi\varrho r)\Big\}\,d\varrho$$

$$= -\frac{2\pi}{r^{\mu+2}}\int_0^\infty \varrho^{\mu+1}h^{\|}(\varrho)J_{\mu+2}(2\pi\varrho r)\,d\varrho,$$

again by (76). Whence (38) holds and $x \mapsto \mathbf{k}(x)\alpha$ is curl-free. $\square$

Proposition 3.20. *Let $H \hookrightarrow C^1(\mathbb{R}^d, \mathbb{R}^d)$ be a RKHS with a given TRI kernel $\mathbf{k}$. It is the case that $\mathrm{div}(\mathbf{k}(\cdot)\alpha) = 0$ for all $\alpha \in \mathbb{R}^d$ if and only*

Table 1: Conditions on the coefficients $\mathbf{k}$ and $\widehat{\mathbf{k}}$ for the kernel to be divergence-free or curl-free

	Conditions on $(k^{\parallel}, k^{\perp})$	Conditions on $(h^{\parallel}, h^{\perp})$
$\mathbf{k}$ is **div-free**	$(d-1)\dfrac{k^{\parallel}(r) - k^{\perp}(r)}{r} + \dfrac{dk^{\parallel}}{dr}(r) = 0$	$h^{\parallel} = 0$
$\mathbf{k}$ is **curl-free**	$\dfrac{k^{\parallel}(r) - k^{\perp}(r)}{r} - \dfrac{dk^{\perp}}{dr}(r) = 0$	$h^{\perp} = 0$

if $\operatorname{div} u = 0$ *for all* $u \in H$. *On the other hand, we have that* $\operatorname{curl}(\mathbf{k}(\cdot)\alpha) = 0$ *for all* $\alpha \in \mathbb{R}^d$ *if and only if* $\operatorname{curl} u = 0$ *for all* $u \in H$.

Proof. The "if" part is obvious. Conversely, let $H_0 := \operatorname{span}\{\mathbf{k}(\cdot - x)\alpha \mid x, \alpha \in \mathbb{R}^d\}$. Take any $u \in H$ and a sequence $\{u_n\}$ in H_0 with $u_n \to u$. For all indices $i, j \in \{1, \ldots, d\}$, by formula (13) it is the case that $e_i \cdot \partial_j u_n(x) = \langle -\partial_j \mathbf{k}(\cdot - x)e_i, u_n \rangle_V$, which converges to $\langle -\partial_j \mathbf{k}(\cdot - x)e_i, u \rangle_V = e_i \cdot \partial_j u(x)$. That is, the sequence of functions $\partial_j u_n$ converges to $\partial_j u$ pointwise. Therefore we have $\operatorname{div} u_n \to \operatorname{div} u$ and $\operatorname{curl} u_n \to \operatorname{curl} u$ pointwise. Since $u_n \in H_0$ for all n, if $\operatorname{div}(\mathbf{k}(\cdot)\alpha) = 0$ for all $\alpha \in \mathbb{R}^d$ then $\operatorname{div} u_n = 0$, whence $\operatorname{div} u = 0$. A similar argument holds in the case $\operatorname{curl}(\mathbf{k}(\cdot)\alpha) = 0$. $\qquad\square$

The above results justify the following definition.

Definition 3.21. *Let* $H \hookrightarrow C^1(\mathbb{R}^d \mathbb{R}^d)$ *be a RKHS with a TRI kernel* $\mathbf{k}$. *We call the space* H *and its kernel* $\mathbf{k}$ *divergence-free (or curl-free) if* $\operatorname{div} u = 0$ *(respectively, if* $\operatorname{curl} u = 0$*) for all* $u \in H$.

Table 1 summarizes the conditions on the coefficients $(k^{\parallel}, k^{\perp})$ of $\mathbf{k}$ and on the coefficients $(h^{\parallel}, h^{\perp})$ of its Fourier transform for a Reproducing Kernel Hilbert Space H with a TRI kernel to be either divergence-free or curl-free. We should note that if $\mathbf{k}$ is *scalar*, i.e. of the type $\mathbf{k}(x) = k(\|x\|)\mathbb{I}_d$, then $k^{\parallel} = k^{\perp}$ and $h^{\parallel} = h^{\perp}$; therefore by Corollaries 3.16 and 3.18 the corresponding RKHS can be *neither* divergence-free *nor* curl-free (unless we are dealing with the trivial case $\mathbf{k} = 0$).

Examples 1 and 2, revisited. In light of the above results, the boundaries of the domains D_1 and D_2 of Examples 1 and 2 in this section have an interesting interpretation. We already noted that in both cases choosing $a = 0$ (vertical boundary of the domains) yields *scalar* kernels, i.e. of the type $\mathbf{k}(x) = k(\|x\|)\mathbb{I}_2$ with $k(r) = b\exp(-\frac{1}{2}r^2/\sigma^2)$ (Gaussian).

In Example 1, when $b = (d-1)a\sigma^2$ we have $\tilde{k}(r) = a\exp(-\frac{1}{2}r^2/\sigma^2)$, $k^{\parallel}(r) = (d-1)a\sigma^2\exp(-\frac{1}{2}r^2/\sigma^2)$; therefore equation (35) holds and the
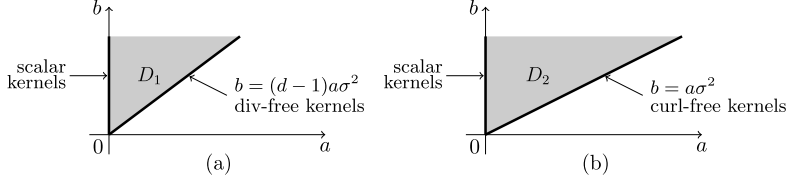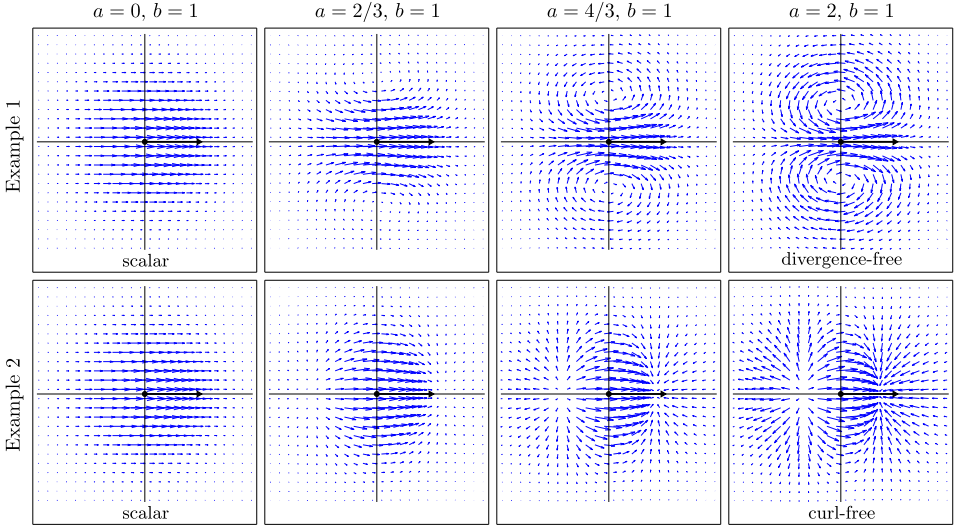
Figure 4: Interpretation of the domain boundaries in Examples 1 and 2.

Figure 5: Transition from a scalar kernel to divergence-free (Example 1) and curl-free (Example 2) kernels by changing the parameter a in the range $[0, 2]$, with $b = 2\sigma^2 = 1$, in $d = 2$ dimensions. The vector fields $x \mapsto \mathbf{k}(x)\alpha$ are illustrated for $\alpha = e_1$, which in also shown in each graph.

kernel is divergence-free. Similarly, in the case of Example 2 when $b = a\sigma^2$ we have $\tilde{k}(r) = -a\exp(-\frac{1}{2}r^2/\sigma^2)$ and $k^\perp(r) = a\sigma^2\exp(-\frac{1}{2}r^2/\sigma^2)$; whence (38) holds and the kernel is curl-free. Thus the boundaries of the two domains may be interpreted as illustrated in Figure 4.

In the case $d = 2$ we have that $D_1 = D_2$. Figure 5 shows the vector field $x \mapsto \mathbf{k}(x)\alpha$, $\alpha = e_1$, for both Example 1 and Example 2, with $b = 2\sigma^2 = 1$ and different choices of the parameter a. In particular, for $a = 0$ the point (a, b) is on the vertical part of the boundary of D_1 and D_2; this corresponds, in both cases, to a scalar kernel. As we move along the line $b = 1$ towards higher values of a, we note that we transition to a vector field $x \mapsto \mathbf{k}(x)\alpha$ that is divergence-free (Example 1) or curl-free (Example 2);

this occurs when (a, b) hits the slanted part of the boundaries of D_1 and D_2, i.e. the line $b = a\sigma^2$.

Hodge decomposition. Let $\mathbf{k} \in L^1(\mathbb{R}^d, \mathbb{R}^{d \times d})$ be a TRI kernel with coefficients $(k^\parallel, k^\perp)$, and let $(h^\parallel, h^\perp)$ be the coefficients of its Fourier transform, which we also assume integrable. Using the functional M introduced in (32), we have $(k^\parallel, k^\perp) = M^{-1}(h^\parallel, h^\perp)$. By the linearity of M^{-1} we also have

$$(40) \quad (k^\parallel, k^\perp) = (k_{\mathrm{cf}}^\parallel, k_{\mathrm{cf}}^\perp) + (k_{\mathrm{df}}^\parallel, k_{\mathrm{df}}^\perp), \ \text{where:} \ \begin{cases} (k_{\mathrm{cf}}^\parallel, k_{\mathrm{cf}}^\perp) := M^{-1}(h^\parallel, 0) \\ (k_{\mathrm{df}}^\parallel, k_{\mathrm{df}}^\perp) := M^{-1}(0, h^\perp) \end{cases}$$

(subscripts "cf" and "df" will be justified soon). With the above definitions, we let $\mathbf{k} = \mathbf{k}_{\mathrm{cf}} + \mathbf{k}_{\mathrm{df}}$, where

$$(41) \quad \mathbf{k}_{\mathrm{cf}}(x) := k_{\mathrm{cf}}^\parallel(x) \operatorname{Pr}_x^\parallel + k_{\mathrm{cf}}^\perp(x) \operatorname{Pr}_x^\parallel \ \text{ and } \ \mathbf{k}_{\mathrm{df}}(x) := k_{\mathrm{df}}^\parallel(x) \operatorname{Pr}_x^\parallel + k_{\mathrm{df}}^\perp(x) \operatorname{Pr}_x^\parallel;$$

by Theorem 3.19 we have that for all $\alpha \in \mathbb{R}^d$ we may write $\mathbf{k}(x)\alpha = \mathbf{k}_{\mathrm{cf}}(x)\alpha + \mathbf{k}_{\mathrm{df}}(x)\alpha$, $x \in \mathbb{R}^d$, where *the first term is a curl-free vector field* and *the second term is instead div-free*. In other words, for any integrable TRI kernel $\mathbf{k}$ and any $\alpha \in \mathbb{R}^d$ this procedure allows one to perform the *Hodge decomposition* [17, 34] of the vector field $x \mapsto \mathbf{k}(x)\alpha$: the two terms, curl-free and divergence-free, may be respectively computed from the coefficients $h^\parallel$ and $h^\perp$ of the Fourier transform of $\mathbf{k}$ precisely by using the inversion formulae (33a) and (33b) twice, the first time with $h^\perp = 0$ (to compute $\mathbf{k}_{\mathrm{cf}}$) and then with $h^\parallel = 0$ (to compute $\mathbf{k}_{\mathrm{df}}$); note that since $\mathbf{k}$ vanishes at infinity *there is no harmonic component in the Hodge decomposition*. We shall call $\mathbf{k}_{\mathrm{cf}}$ and $\mathbf{k}_{\mathrm{df}}$ the *curl-free* and *divergence-free* components of the kernel $\mathbf{k}$.

Incidentally, we note that if $\mathbf{k} \in L^1 \cap L^2$ then for any $\alpha \in \mathbb{R}^d$ the orthogonality in $L^2(\mathbb{R}^d, \mathbb{R}^d)$ of the Hodge components $\mathbf{k}_{\mathrm{cf}}(\cdot)\alpha$ and $\mathbf{k}_{\mathrm{df}}(\cdot)\alpha$ may be immediately verified via Plancherel's theorem as follows:

$$\langle \mathbf{k}_{\mathrm{cf}}\alpha, \mathbf{k}_{\mathrm{df}}\alpha \rangle_{L^2} = \langle \widehat{\mathbf{k}}_{\mathrm{cf}}\alpha, \widehat{\mathbf{k}}_{\mathrm{df}}\alpha \rangle_{L^2} = \int_{\mathbb{R}^d} \left(h^\parallel(\|\xi\|)\operatorname{Pr}_\xi^\parallel \alpha \right) \cdot \left(h^\perp(\|\xi\|)\operatorname{Pr}_\xi^\perp \alpha \right) d\xi = 0.$$

The computation of $\mathbf{k}_{\mathrm{cf}}$ and $\mathbf{k}_{\mathrm{df}}$ from $\mathbf{k}$ via formulae (40) and (41) may be viewed as an orthogonal projection (in the L^2 sense) of $\mathbf{k}$ onto the spaces of curl-free and divergence-free kernels, respectively.

Example 3. We consider the Gaussian scalar kernel $\mathbf{k}(x) = k(\|x\|)\mathbb{I}_d$, $x \in \mathbb{R}^d$, with $k(r) = \exp(-\frac{1}{2}r^2/\sigma^2)$ for some $\sigma > 0$, and compute its Hodge decomposition. We saw earlier (17) that its Fourier transform is

$$(42) \quad \widehat{\mathbf{k}}(\xi) = h(\|\xi\|)\mathbb{I}_d, \ \xi \in \mathbb{R}^d \quad \text{with} \quad h(\varrho) = \left(2\pi\sigma^2 \right)^{\mu+1} \exp\left(-2\pi^2\sigma^2\varrho^2 \right),$$

where as usual $\mu = \frac{d}{2} - 1$. We let $\mathbf{k} = \mathbf{k}_{\mathrm{cf}} + \mathbf{k}_{\mathrm{df}}$, with $\widehat{\mathbf{k}}_{\mathrm{cf}}(\xi) = h(\|\xi\|)\,\mathrm{Pr}_\xi^\|$ and $\widehat{\mathbf{k}}_{\mathrm{df}}(\xi) = h(\|\xi\|)\,\mathrm{Pr}_\xi^\perp$. The coefficients $(k_{\mathrm{cf}}^\|, k_{\mathrm{cf}}^\perp)$ of $\mathbf{k}_{\mathrm{cf}}$ are given by inversion formulae (33a) and (33b) with $h^\| = h$ and $h^\perp = 0$, i.e.

$$(43a) \qquad k_{\mathrm{cf}}^\|(r) = \frac{2\pi}{r^\mu} \int_0^\infty \varrho^{\mu+1}\, h(\varrho)\, J_\mu(2\pi\varrho r)\, d\varrho$$

$$- \frac{2\mu+1}{r^{\mu+1}} \int_0^\infty \varrho^\mu\, h(\varrho)\, J_{\mu+1}(2\pi\varrho r)\, d\varrho,$$

$$(43b) \qquad k_{\mathrm{cf}}^\perp(r) = \frac{1}{r^{\mu+1}} \int_0^\infty \varrho^\mu\, h(\varrho)\, J_{\mu+1}(2\pi\varrho r)\, d\varrho,$$

We compute the integral (43b) with h as in (42) using the Hankel transform (27), which yields

$$(44) \qquad k_{\mathrm{cf}}^\perp(r) = \frac{(2\sigma^2)^{\mu+1}}{2r^{2\mu+2}}\, \gamma\!\left(\mu+1, \frac{1}{2}\frac{r^2}{\sigma^2}\right), \qquad r > 0,$$

where $\gamma(\nu, x) = \int_0^x e^{-t}\, t^{\nu-1} dt$, $\Re\nu > 0$, is the lower incomplete gamma function [1, §6.5.2]. The other coefficients of the kernels $\mathbf{k}_{\mathrm{cf}}$ and $\mathbf{k}_{\mathrm{df}}$ are obtained from (44) simply as follows:

$$(45) \quad k_{\mathrm{cf}}^\| = k - (2\mu+1)k_{\mathrm{cf}}^\perp, \quad k_{\mathrm{df}}^\| = k - k_{\mathrm{cf}}^\| = (2\mu+1)k_{\mathrm{cf}}^\perp, \quad k_{\mathrm{df}}^\perp = k - k_{\mathrm{cf}}^\perp;$$

the first one is derived from (43a), while the other two follow from $k = k_{\mathrm{cf}}^\| + k_{\mathrm{df}}^\|$ and $k = k_{\mathrm{cf}}^\perp + k_{\mathrm{df}}^\perp$ (which, in turn, are a consequence of $\mathbf{k} = \mathbf{k}_{\mathrm{cf}} + \mathbf{k}_{\mathrm{df}}$ and the fact that $\mathbf{k}$ is scalar).

When $d = 2$ (i.e. $\mu = 0$), since $\gamma(1, x) = 1 - e^{-x}$ we simply have $k_{\mathrm{cf}}^\perp(r) = \frac{\sigma^2}{r^2}[1 - \exp(-\frac{1}{2}r^2/\sigma^2)]$, and the other coefficients are again derived from formulae (45) with $\mu = 0$. It is interesting to note that $\mathbf{k}_{\mathrm{cf}}$ and $\mathbf{k}_{\mathrm{df}}$ are *not* Gaussian like the curl-free and divergence-free kernels that we derived in Examples 2 and 1, respectively—in fact they have much heavier "tails" (they go to zero like $1/r^2$ as $r \to \infty$) than the original scalar and Gaussian kernel $\mathbf{k}$; however, these heavy tails cancel each other when added. In the case $2\sigma^2 = 1$ and $d = 2$ the Hodge decomposition of $x \mapsto \mathbf{k}(x)e_1$ into its curl-free and divergence-free terms is illustrated in Figure 6; as observed above, the two vector fields are orthogonal in the L^2 sense.

The discussion above also suggests that a seemingly simple way to generate curl-free (or divergence-free) kernels is to choose functions $h^\|$ (respectively, $h^\perp$) in $L^1(\mathbb{R}^+, r^{d-1})$ and apply formulae (33a) and (33b) with $h^\perp = 0$ (respectively, $h^\| = 0$). This is rather cumbersome because it involves the
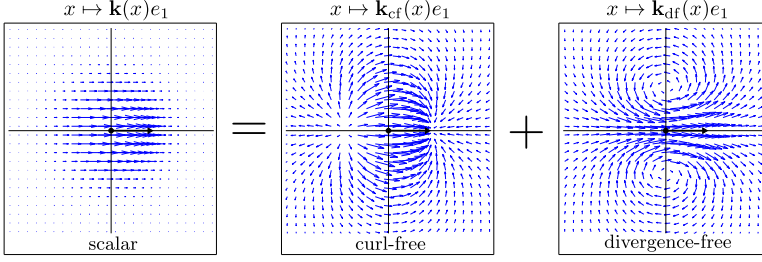
Figure 6: Hodge decomposition of $x \mapsto \mathbf{k}(x)e_1$, for $\mathbf{k}$ scalar and Gaussian.

computation of Hankel transforms, analytically or numerically; in the next section we shall introduce a simple technique for generating all TRI kernels of interest while avoiding such tedious calculations.

4. Construction of matrix-valued kernels from scalar kernels

In this section we illustrate a method for constructing matrix-valued kernels from scalar kernels—in fact, it will turn out that all matrix-valued kernels of interest can be built this way. We remind the reader that a s-admissible Hilbert space of $\mathbb{R}^d$-valued functions defined in Ω is also a RKHS, and its kernel is such that $K(\cdot, x)\alpha \in C_0^s(\Omega, \mathbb{R}^d)$ for all $x \in \Omega$, $\alpha \in \mathbb{R}^d$ (see section 2).

4.1. Curl-free component

Given a s-admissible Hilbert space H of *scalar*-valued differentiable functions defined on $\mathbb{R}^d$, one can define the space $V := \{\nabla f \mid f \in H\}$. Note that (i) if $s \geq 2$ all the elements of V are curl-free and (ii) for all $v \in V$ there is a unique $f \in H$ such that $v = \nabla f$ (in fact if $v = \nabla f_1 = \nabla f_2$ then $f_1 - f_2 \in H$ is a constant that must be zero because $H \subset C_0^s(\mathbb{R}^d, \mathbb{R})$, i.e. it vanishes at infinity). The proposition that follows shows that the kernel of H induces a reproducing kernel on V.

Notation. Given a *scalar*-valued differentiable function $F : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ we indicate with $\partial_{n,i}F$, $n \in \{1, 2\}$, $i \in \{1, \ldots, d\}$ the derivative of F with respect to the ith variable of the nth set of variables; with $\nabla_n F$, where $n \in \{1, 2\}$, the gradient of F with respect to the nth set of variables; and with $\nabla_1 \nabla_2^T F$ the matrix-valued function whose $(i, j)^{\text{th}}$ entry is $\partial_{1,i}\partial_{2,j}F$, for $1 \leq i, j \leq d$. Finally, for a vector-valued function $\mathbf{f} : \mathbb{R}^d \to \mathbb{R}^d$, we indicate with $\nabla \mathbf{f}$ the matrix whose $(i, j)^{\text{th}}$ entry is $\partial_i \mathbf{f}_j$, for $1 \leq i, j \leq d$.

Proposition 4.1. *For fixed $s \geq 1$, consider a Hilbert space $H \hookrightarrow C_0^s(\mathbb{R}^d, \mathbb{R})$ of scalar-valued functions, with kernel $K_H : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. Let the space*

$V := \{\nabla f \mid f \in H\}$ *be endowed with the inner product* $\langle v_1, v_2 \rangle_V := \langle f_1, f_2 \rangle_H$, *where* $v_1 = \nabla f_1$ *and* $v_2 = \nabla f_2$. *Then* $V \hookrightarrow C_0^{s-1}(\mathbb{R}^d, \mathbb{R}^d)$, *i.e. it is a* $(s-1)$-*admissible Hilbert space of* $\mathbb{R}^d$-*valued functions, and its kernel is given by* $K_V = \nabla_1 \nabla_2^T K_H$.

Proof. One can easily verify that $\langle \cdot, \cdot \rangle_V$ is an inner product, that V is complete and continuously embedded in $C_0^{s-1}(\mathbb{R}^d, \mathbb{R}^d)$. So V is a RKHS, and for all $x, \alpha \in \mathbb{R}^d$ and $v = \nabla f \in V$, its kernel K_V must satisfy

$$
\begin{aligned}
\langle K_V(\cdot, x)\alpha, v \rangle_V &= \alpha \cdot v(x) = \alpha \cdot \nabla f(x) = \sum_{i=1}^d \alpha_i \, \partial_i f(x) = (\text{by Theorem } 2.11) \\
&= \sum_{i=1}^d \alpha_i \langle \partial_{2,i} K_H(\cdot, x), f \rangle_H = \langle \alpha \cdot \nabla_2 K_H(\cdot, x), f \rangle_H \\
&= \langle \nabla(\alpha \cdot \nabla_2 K_H(\cdot, x)), v \rangle_V,
\end{aligned}
$$

therefore $K_V(\cdot, x)\alpha = \nabla(\nabla_2 K_H(\cdot, x) \cdot \alpha) = (\nabla_1 \nabla_2^T K_H(\cdot, x))\alpha$; we have used the the property that $\nabla(\mathbf{f} \cdot \alpha) = (\nabla \mathbf{f})\alpha$ for any differentiable vector-valued function $\mathbf{f} : \mathbb{R}^d \to \mathbb{R}^d$. By the arbitrariness of $\alpha \in \mathbb{R}^d$ we conclude that we must have $K_V(y, x) = \nabla_1 \nabla_2^T K_H(y, x)$, for $y, x \in \mathbb{R}^d$. $\qquad \square$

Remark. Before proceeding to the case of translation- and rotation-invariant kernels we note that if the kernel K of a s-admissible Hilbert space is translation-invariant, i.e. $K(x, y) = \mathbf{k}(x-y)$, then $\mathbf{k}$ is differentiable $2s$ times (see the remark at the end of section 3.1).

Proposition 4.2. *Under the assumptions of Proposition 4.1, if* $K_H(x, y) = \mathbf{k}_H(x-y)$ *for some function* $\mathbf{k}_H : \mathbb{R}^d \to \mathbb{R}$, *then* $K_V(x, y) = \mathbf{k}_V(x-y)$, *with* $\mathbf{k}_V = -\text{Hess}\, \mathbf{k}_H$ *(i.e.* $\mathbf{k}_V^{ij} = -\partial_i \partial_j \mathbf{k}_H$*).*

Proof. We have that $\partial_{2,j} K_H(x, y) = -\partial_j \mathbf{k}_H(x - y)$ and $\partial_{1,i} \partial_{2,j} K_H(x, y) = -\partial_i \partial_j \mathbf{k}_H(x - y)$, for all $x, y \in \mathbb{R}^d$. The result follows by applying Proposition 4.1. $\qquad \square$

Corollary 4.3. *Under the assumptions of Propositions 4.1 with* $s \geq 2$ *and 4.2, for any* $\alpha \in \mathbb{R}^d$ *the vector field* $x \mapsto \mathbf{k}_V(x)\alpha$ *is curl-free.*

Proof. Using the fact that $(\nabla \mathbf{f})\alpha = \nabla(\mathbf{f} \cdot \alpha)$ for any differentiable vector-valued function $\mathbf{f} : \mathbb{R}^d \to \mathbb{R}^d$, it is the case that $\mathbf{k}_V(x)\alpha = (-\text{Hess}\, \mathbf{k}_H(x))\alpha = -(\nabla \nabla^T \mathbf{k}_H(x))\alpha = -\nabla(\nabla \mathbf{k}_H(x) \cdot \alpha)$, and we conclude immediately. $\qquad \square$

Proposition 4.4. *Under the assumptions of Propositions 4.1 and 4.2, if* $\mathbf{k}_H(x) = k_H(\|x\|)$ *for some* $k_H : \mathbb{R}^+ \to \mathbb{R}$ *then the kernel of* V *is TRI, i.e.* $\mathbf{k}_V(x) = k_V^{\parallel}(\|x\|) \Pr_x^{\parallel} + k_V^{\perp}(\|x\|) \Pr_x^{\perp}$ *with*

$$
(46) \qquad k_V^{\parallel}(r) = -\frac{d^2 k_H}{dr^2}(r) \qquad and \qquad k_V^{\perp}(r) = -\frac{1}{r}\frac{dk_H}{dr}(r), \quad r > 0.
$$

Proof. First note that we have $\partial_j \mathbf{k}_H(x) = k_H'(\|x\|)\, x_j/\|x\|$ and

$$-\partial_i\partial_j \mathbf{k}_H(x) = -k_H''(\|x\|)\frac{x_i x_j}{\|x\|^2} - k_H'(\|x\|)\frac{\partial}{\partial x^i}\frac{x_j}{\|x\|}$$

$$= -k_H''(\|x\|)\frac{x_i x_j}{\|x\|^2} - k_H'(\|x\|)\left(\frac{\delta_{ij}}{\|x\|} - \frac{x_i x_j}{\|x\|^3}\right),$$

where δ_{ij} is Kronecker's symbol. By Proposition 4.2 it must be the case that

$$\mathbf{k}_V(x) = -k_H''(\|x\|)\frac{xx^T}{\|x\|} - \frac{k_H'(\|x\|)}{\|x\|}\left(\mathbb{I}_d - \frac{xx^T}{\|x\|^2}\right)$$

$$= -k_H''(\|x\|)\,\mathrm{Pr}_x^{\|} - \frac{k_H'(\|x\|)}{\|x\|}\,\mathrm{Pr}_x^{\perp}. \qquad \square$$

Corollary 4.5. *Under the assumptions of Proposition 4.4, $\mathbf{k}_V$ is curl-free.*

When $s \geq 2$, the above follows from Corollary 4.3. However, one can verify that $k_V^{\|}$ and $k_V^{\perp}$, given by (46), satisfy equation (38), and that the latter operation only involves *two* derivatives of k_H (and k_H is twice differentiable when $s = 1$, by Theorem 2.11 and the remark at the end of section 3.1). Therefore Corollary 4.5 also holds when $s = 1$. The curl-free property may be seen in the Fourier domain as follows.

Proposition 4.6. *Under the assumptions of Proposition 4.4, if $\mathbf{k}_H$ is integrable then $\widehat{\mathbf{k}}_H(\xi) = h_H(\|\xi\|)$ and $\widehat{\mathbf{k}}_V(\xi) = h_V^{\|}(\|\xi\|)\,\mathrm{Pr}_\xi^{\|} + h_V^{\perp}(\|\xi\|)\,\mathrm{Pr}_\xi^{\perp}$, where $h_V^{\perp} = 0$ and*

$$(47)\quad h_V^{\|}(\varrho) = (2\pi\varrho)^2\, h_H(\varrho), \quad\text{with}\quad h_H(\varrho) = \frac{2\pi}{\varrho^\mu}\int_0^\infty r^{\mu+1}\, k_H(r)\, J_\mu(2\pi\varrho r)\, dr$$

and, as usual, $\mu = \frac{d}{2} - 1$.

Proof. Assuming just translation invariance, the $(j, \ell)^{\text{th}}$ element of $\widehat{\mathbf{k}}_V(\xi)$ is

$$(\widehat{\mathbf{k}}_V)_{j\ell}(\xi) = -\int_{\mathbb{R}^d}\partial_j\partial_\ell \mathbf{k}_H(x)\,e^{-2\pi i\xi\cdot x}dx$$

$$= -(2\pi i\xi_j)(2\pi i\xi_\ell)\int_{\mathbb{R}^d}\mathbf{k}_H(x)\,e^{-2\pi i\xi\cdot x}dx = (2\pi)^2\xi_j\xi_\ell\,\widehat{\mathbf{k}}_H(\xi),$$

from which we have $\widehat{\mathbf{k}}_V(\xi) = (2\pi\|\xi\|)^2\,\widehat{\mathbf{k}}_H(\xi)\,\mathrm{Pr}_\xi^{\|}$. When $\mathbf{k}_H$ is also rotation-invariant, by Proposition 3.9 we must have $\widehat{\mathbf{k}}_H(\xi) = h_H(\|\xi\|)$ with $h_H(\varrho)$ as in (47), and we conclude immediately. $\qquad \square$

$$\mathbf{k}_H(x) = k_H(\|x\|) \qquad \overset{\mathcal{F}}{\longmapsto} \qquad \widehat{\mathbf{k}}_H(x) = h_H(\|x\|)$$

$$\Big\downarrow {\scriptstyle -\mathrm{Hess}} \qquad\qquad\qquad\qquad \Big\downarrow {\scriptstyle (-\mathrm{Hess})^\wedge}$$

$$\mathbf{k}_V : (k_V^\|, k_V^\perp) = \big(-k_H'', -\tfrac{1}{r}k_H'\big) \quad \overset{\mathcal{F}}{\longmapsto} \quad \widehat{\mathbf{k}}_V : (h_V^\|, h_V^\perp) = \big((2\pi\varrho)^2 h_H, 0\big)$$

Figure 7: Commutative diagram illustrating the construction, from the scalar-valued kernel $\mathbf{k}_H : \mathbb{R}^d \to \mathbb{R}$, of the matrix-valued, curl-free kernel $\mathbf{k}_V : \mathbb{R}^d \to \mathbb{R}^{d\times d}$, with interpretation in the Fourier domain. Note that while $H \hookrightarrow C_0^s(\mathbb{R}^d, \mathbb{R})$, we have $V \hookrightarrow C_0^{s-1}(\mathbb{R}^d, \mathbb{R}^d)$.

The construction of the curl-free kernel $\mathbf{k}_V$ from the scalar-valued kernel $\mathbf{k}_H$ is summarized in Figure 7. We have used the notation $(k_V^\|, k_V^\perp)$ to compactly represent the coefficients of $\mathbf{k}_V$ and $(h_V^\|, h_V^\perp)$ to represent those of its Fourier transform $\widehat{\mathbf{k}}_V$. The dual operator of the negative Hessian in the Fourier domain is the map $h_H \mapsto ((2\pi\varrho)^2 h_H, 0)$. Under the assumptions of Proposition 4.6 we may actually compute the corresponding "initial" kernel $\mathbf{k}_H$ in terms of the non-zero coefficient $h_V^\|$ of $\widehat{\mathbf{k}}_V$; in fact, using the inversion formula (16) and the fact that $h_H(\varrho) = (2\pi\varrho)^{-2} h_V^\|(\varrho)$ yields:

$$(48) \quad \mathbf{k}_H(x) = k_H(\|x\|), \quad \text{with} \ \ k_H(r) = \frac{1}{2\pi r^\mu} \int_0^\infty \varrho^{\mu-1} \, h_V^\|(\varrho) \, J_\mu(2\pi r\varrho) \, d\varrho,$$

defined for $r > 0$, where as usual $\mu = \frac{d}{2} - 1$.

We conclude that any integrable, TRI, curl-free kernel $\mathbf{k}_V$ (i.e. with nonnegative $h_V^\|$, and $h_V^\perp = 0$) of a $(s-1)$-admissible Hilbert space V may be derived with the procedure described in Propositions 4.1 and 4.4 (ultimately, by applying formulae (46)) from the kernel $\mathbf{k}_H(x) = k_H(\|x\|)$ of a s-admissible Hilbert space of scalar-valued functions, where k_H is given by formula (48). This proves the generality of the above method for constructing RKHS of curl-free vector fields with TRI kernels.

Example 2, revisited again. Consider a RKHS of scalar valued functions defined in $\mathbb{R}^d$, denoted by H, with kernel $\mathbf{k}_H(x) = k_H(\|x\|)$, where $k_H(r) = b\sigma^2 \exp(-\frac{1}{2}r^2/\sigma^2)$. As long as $b > 0$ it is the case that the Hilbert space is s-admissible, for all s. Formulae (46) yield

$$(49) \quad k_V^\|(r) = \Big(b - \frac{b}{\sigma^2}r^2\Big) \exp\Big(-\frac{1}{2}\frac{r^2}{\sigma^2}\Big) \quad \text{and} \quad k_V^\perp(r) = b \exp\Big(-\frac{1}{2}\frac{r^2}{\sigma^2}\Big),$$

which are precisely the equations of the coefficients (31) of the kernel in Example 2 with $a = b/\sigma^2$; that is, along the diagonal line of the domain D_2, which corresponds to curl-free kernels. See Figure 4(b).

Example 4 (Bessel-type curl-free kernels)**.** Consider the scalar-valued kernel $\mathbf{k}_H(x) = k_H(\|x\|)$, with k_H given by

$$(50) \qquad k_H(r) = C_0 \left(\frac{r}{\sigma}\right)^{\ell - \frac{d}{2}} K_{\ell - \frac{d}{2}}\left(\frac{r}{\sigma}\right), \qquad r \geq 0,$$

where K_ν is a modified Bessel function of order ν. When $C_0 = C(\sigma, d, \ell) := (2^{\ell + \frac{d}{2} - 1} \pi^{\frac{d}{2}} \Gamma(\ell) \sigma^d)^{-1}$, as in (7) and (19), it is the Green's function of the differential operator $L = (1 - \sigma^2 \Delta)^\ell$. The corresponding Hilbert space H is s-admissible, i.e. $H \hookrightarrow C_0^s(\mathbb{R}^d, \mathbb{R})$, if $\ell \geq s + d/2$. To apply our procedure, we must have at least $s \geq 1$. By Lemma A.1 in Appendix A, applying (46) to the kernel (50) yields

$$(51a) \qquad k_V^\parallel(r) = \frac{C_0}{\sigma^2} \left(\frac{r}{\sigma}\right)^{\nu - 1} \left\{ (2\nu - 1) K_{\nu - 1}\left(\frac{r}{\sigma}\right) - \frac{r}{\sigma} K_\nu\left(\frac{r}{\sigma}\right) \right\},$$

$$(51b) \qquad k_V^\perp(r) = \frac{C_0}{\sigma^2} \left(\frac{r}{\sigma}\right)^{\nu - 1} K_{\nu - 1}\left(\frac{r}{\sigma}\right),$$

where $\nu = \ell - \frac{d}{2}$. From the asymptotic expansion for large arguments of K_ν, given by (78) in Appendix A, one sees that the kernel with coefficients (51a) and (51b) has heavier tails than the Gaussian case (49).

4.2. Divergence-free component

We have seen how one can create the matrix-valued kernel a curl-free RKHS by taking the "double gradient" of a scalar-valued kernel. We will now illustrate how a the kernel of a divergence-free RKHS may be created by taking the "double curl" of a matrix-valued scalar kernel (i.e. with equal diagonal entries). Initially we shall limit ourselves to the case $d = 3$, which is the most relevant in applications.

Lemma 4.7. *A 1-admissible Hilbert space $W \hookrightarrow C_0^1(\mathbb{R}^3, \mathbb{R}^3)$ is divergence-free if and only if the linear map $\mathrm{curl} : W \to C_0(\mathbb{R}^3, \mathbb{R}^3)$ is injective.*

Proof. First assume that W is divergence-free. If $\mathrm{curl}\, w = 0$ then $w = \nabla f$ for some scalar function f; therefore $\Delta f = \mathrm{div}\, w = 0$ because W is divergence-free, and in fact f must be zero because it is harmonic and it vanishes at infinity. Vice versa, assume that curl is injective on W. Take an arbitrary $w \in W$ and let $w = w_1 + w_2$ be its Hodge decomposition (i.e. $\mathrm{curl}\, w_1 = 0$ and

div $w_2 = 0$). We have $\operatorname{curl} w = \operatorname{curl} w_2$, so by the injectivity assumption $w = w_2$, whence $w_1 = 0$. Therefore $\operatorname{div} w = \operatorname{div} w_1 = 0$. □

Let $s \geq 1$ and consider a s-admissible Hilbert space $W \hookrightarrow C_0^s(\mathbb{R}^3, \mathbb{R}^3)$, and the space of vector fields $V := \{\operatorname{curl} w \mid w \in W\}$. By Lemma (4.7) the "primitive" $w \in W$ of $v = \operatorname{curl} w \in V$ is unique if and only if W is divergence-free. With this condition, we can certainly follow a path that is similar to the one described in the previous section and endow V with a reproducing kernel induced by the kernel of W.

Notation. Given a differentiable matrix-valued function $G : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^{3\times 3}$ we indicate with $\operatorname{curl}_n^{\mathrm{C}} G$ the matrix-valued funcion whose i^{th} column (with $i = 1, 2, 3$) is the curl of the ith column of G; the curl is computed with respect to the nth set of variables of G (with $n = 1$ or 2). The definition of $\operatorname{curl}_n^{\mathrm{R}} G$ is obtained by substituting the word "column" with the word "row". If the matrix-valued functions of only three variables $\mathbf{g} : \mathbb{R}^3 \to \mathbb{R}^{3\times 3}$ is differentiable, $\operatorname{curl}^{\mathrm{C}}\mathbf{g}$ (respectively, $\operatorname{curl}^{\mathrm{R}}\mathbf{g}$) simply indicates the matrix whose i^{th} column (row) is the curl of the ith column (row) of $\mathbf{g}$, for $i = 1, 2, 3$.

Proposition 4.8. *For fixed $s \geq 1$, consider a Hilbert space $W \hookrightarrow C_0^s(\mathbb{R}^3, \mathbb{R}^3)$ of divergence-free vector fields, with kernel $K_W : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$. Let $V := \{\operatorname{curl} w \mid w \in W\}$ be endowed with the inner product $\langle v_1, v_2 \rangle_V := \langle w_1, w_2 \rangle_W$, where $v_1 = \operatorname{curl} w_1$ and $v_2 = \operatorname{curl} w_2$. Then $V \hookrightarrow C_0^{s-1}(\mathbb{R}^3, \mathbb{R}^3)$, i.e. it is a $(s-1)$-admissible Hilbert space of $\mathbb{R}^3$-valued functions, and its kernel is given by $K_V = \operatorname{curl}_1^{\mathrm{C}} \operatorname{curl}_2^{\mathrm{R}} K_W$.*

Proof. We indicate with $K_W^{i;}$ and $K_W^{;j}$ the ith row and the jth column of K, respectively. One can easily verify that $\langle \cdot, \cdot \rangle_V$ is an inner product, V is complete, and $V \hookrightarrow C_0^{s-1}(\mathbb{R}^3, \mathbb{R}^3)$; therefore V is a RKHS. For arbitrary $w \in W$ let $v = \operatorname{curl} w$; for all $x, \alpha \in \mathbb{R}^3$ the kernel K_V must be such that

$$\langle K_V(\cdot, x)\alpha, v \rangle_V = \alpha \cdot v(x) = \alpha \cdot \operatorname{curl} w(x)$$
$$= \alpha_1\big(\partial_2 w_3(x) - \partial_3 w_2(x)\big) + \alpha_2\big(\partial_3 w_1(x) - \partial_1 w_3(x)\big) + \alpha_3\big(\partial_1 w_2(x) - \partial_2 w_1(x)\big)$$
$$\overset{(*)}{=} \big\langle \alpha_1\big(\partial_{2,2} K_W^{;3}(\cdot, x) - \partial_{2,3} K_W^{;2}(\cdot, x)\big) + \alpha_2\big(\partial_{2,3} K_W^{;1}(\cdot, x) - \partial_{2,1} K_W^{;3}(\cdot, x)\big)$$
$$+ \alpha_3\big(\partial_{2,1} K_W^{;2}(\cdot, x) - \partial_{2,2} K_W^{;1}(\cdot, x)\big), w \big\rangle_W = \big\langle \big(\operatorname{curl}_2^{\mathrm{R}} K_W(\cdot, x)\big)\alpha, w \big\rangle_W$$
$$= \big\langle \operatorname{curl}\big[\big(\operatorname{curl}_2^{\mathrm{R}} K_W(\cdot, x)\big)\alpha\big], v \big\rangle_V \overset{(**)}{=} \big\langle \big[\operatorname{curl}_1^{\mathrm{C}} \operatorname{curl}_2^{\mathrm{R}} K_W(\cdot, x)\big]\alpha, v \big\rangle_V,$$

where in step $(*)$ we have used Theorem 2.11, while in $(**)$ the fact that $\operatorname{curl}(\mathbf{g}\alpha) = (\operatorname{curl}^{\mathrm{C}}\mathbf{g})\alpha$ for any differentiable function $\mathbf{g} : \mathbb{R}^3 \to \mathbb{R}^{3\times 3}$ and any $\alpha \in \mathbb{R}^3$. This concludes the proof. □

In the translation-invariant case we have an immediate consequence.

Proposition 4.9. *Under the assumptions of Proposition 4.8, if $K_W(x, y) =$ $\mathbf{k}_W(x - y)$ for some function $\mathbf{k}_W : \mathbb{R}^3 \to \mathbb{R}^{3\times 3}$, then $K_V(x, y) = \mathbf{k}_V(x - y)$ with $\mathbf{k}_V = -\mathrm{curl}^C \mathrm{curl}^R \mathbf{k}_W$.*

Corollary 4.10. *Under the assumptions of Propositions 4.8 with $r \geq 2$ and 4.9, for any $\alpha \in \mathbb{R}^d$ the vector field $x \mapsto \mathbf{k}_V(x)\alpha$ is divergence-free.*

Proof. Since $(\mathrm{curl}^C \mathbf{g})\alpha = \mathrm{curl}(\mathbf{g}\alpha)$ for any differentiable matrix-valued function $\mathbf{g} : \mathbb{R}^3 \to \mathbb{R}^{3\times 3}$, it is the case that $\mathbf{k}_V(x)\alpha = (-\mathrm{curl}^C \mathrm{curl}^R \mathbf{k}_W(x))\alpha = -\mathrm{curl}[(\mathrm{curl}^R \mathbf{k}_W(x))\alpha]$, and we conclude immediately. □

Lemma 4.11. *For any matrix-valued function $\mathbf{g} = [g_{ij}]_{1\leq i,j\leq 3} : \mathbb{R}^3 \to \mathbb{R}^{3\times 3}$ that is twice differentiable we have:*

$$\mathrm{curl}^C \mathrm{curl}^R \mathbf{g} = \mathrm{curl}^R \mathrm{curl}^C \mathbf{g}$$

$$= \begin{bmatrix} \partial_2^2 g_{33} - \partial_2\partial_3(g_{23} + g_{32}) + \partial_3^2 g_{22} & -\partial_1\partial_2 g_{33} + \partial_1\partial_3 g_{23} + \partial_2\partial_3 g_{31} - \partial_3^2 g_{21} \\ -\partial_1\partial_2 g_{33} + \partial_1\partial_3 g_{32} + \partial_2\partial_3 g_{13} - \partial_3^2 g_{12} & \partial_1^2 g_{33} - \partial_1\partial_3(g_{13} + g_{31}) + \partial_3^2 g_{11} \\ \partial_1\partial_2 g_{23} - \partial_1\partial_3 g_{22} - \partial_2^2 g_{13} + \partial_2\partial_3 g_{12} & -\partial_1^2 g_{23} + \partial_1\partial_2 g_{13} + \partial_1\partial_3 g_{21} - \partial_2\partial_3 g_{11} \end{bmatrix}$$

$$\begin{matrix} \partial_1\partial_2 g_{32} - \partial_1\partial_3 g_{22} - \partial_2^2 g_{31} + \partial_2\partial_3 g_{21} \\ -\partial_1^2 g_{32} + \partial_1\partial_2 g_{31} + \partial_1\partial_3 g_{12} - \partial_2\partial_3 g_{11} \\ \partial_1^2 g_{22} - \partial_1\partial_2(g_{12} + g_{21}) + \partial_2^2 g_{11} \end{matrix} \Big].$$

The above lemma may be proven by direct computation. We note that curl^R and curl^C *commute*, and if $\mathbf{g}$ is symmetric (i.e. $g_{ij} = g_{ji}$ for $1 \leq i, j \leq 3$) then so is the matrix-valued function $\mathrm{curl}^C \mathrm{curl}^R \mathbf{g}$.

Remark. Before proceeding further we should note that the above procedure and results require the *s*-admissible space of vector fields W (that we use to build the new space V) to be divergence-free. This sounds like a severe limitation because, as we said at the beginning of the section, our goal here is to formulate a procedure that *generates* RKHS of divergence-free vector fields. Having to start from such a space does not seem very useful. The next results, that refer to the TRI case, solve this problem. In order to interpret the results in the Fourier domain, we shall assume that the kernels are integrable.

Proposition 4.12. *Let $\mathbf{k}_1$, $\mathbf{k}_2 : \mathbb{R}^3 \to \mathbb{R}^{3\times 3}$ be the TRI, integrable kernels of two s-admissible Hilbert spaces, with $s \geq 2$. If their Fourier transforms have the the same divergence-free coefficient $h^\perp$, i.e.*

$$\widehat{\mathbf{k}}_1(\xi) = h_1^\parallel(\|\xi\|)\,\mathrm{Pr}_\xi^\parallel + h_1^\perp(\|\xi\|)\,\mathrm{Pr}_\xi^\perp \quad and \quad \widehat{\mathbf{k}}_2(\xi) = h_2^\parallel(\|\xi\|)\,\mathrm{Pr}_\xi^\parallel + h_2^\perp(\|\xi\|)\,\mathrm{Pr}_\xi^\perp,$$

with $h_1^\perp = h_2^\perp := h^\perp$, then $-\mathrm{curl}^C \mathrm{curl}^R \mathbf{k}_1 = -\mathrm{curl}^C \mathrm{curl}^R \mathbf{k}_2 =: \mathbf{k}$, and $\mathbf{k}$ is a divergence-free kernel.

Proof. We first observe that if the Fourier transform a kernel $\mathbf{k}$ has coefficient $h^\perp = 0$ then, for all $\alpha \in \mathbb{R}^3$, it is the case that $\mathrm{curl}(\mathbf{k}\alpha) = 0$; but $\mathrm{curl}(\mathbf{k}\alpha) = (\mathrm{curl}^\mathrm{C}\mathbf{k})\alpha$, therefore $\mathrm{curl}^\mathrm{C}\mathbf{k} = 0$ and $\mathrm{curl}^\mathrm{C}\mathrm{curl}^\mathrm{R}\mathbf{k} = 0$ (by the commutativity of curl^C and curl^R, see Lemma 4.11). By the linearity of the operator M (and its inverse), introduced by (32), for $i = 1, 2$ we have

$$(52) \qquad \begin{aligned} \mathbf{k}_i(x) &= k_i^\|(\|x\|)\mathrm{Pr}_\xi^\| + k_i^\perp(\|x\|)\mathrm{Pr}_x^\perp, \quad \text{with} \\ (k_i^\|, k_i^\perp) &= M^{-1}(h_i^\|, h^\perp) = M^{-1}(h_i^\|, 0) + M^{-1}(0, h^\perp); \end{aligned}$$

note that $h_1^\perp = h_2^\perp = h^\perp$. We now apply $\mathrm{curl}^\mathrm{C}\mathrm{curl}^\mathrm{R}$ to both $\mathbf{k}_1$ and $\mathbf{k}_2$ and, by the first part of the proof, we have that the component that corresponds to the first term on the right-hand side of (52) vanishes for both $\mathbf{k}_1$ and $\mathbf{k}_2$; what remains is the second component, which is the same for both kernels. Therefore $\mathrm{curl}^\mathrm{C}\mathrm{curl}^\mathrm{R}\mathbf{k}_1 = \mathrm{curl}^\mathrm{C}\mathrm{curl}^\mathrm{R}\mathbf{k}_2$. The resulting kernel is divergence-free because, for $i = 1, 2$, we have $(\mathrm{curl}^\mathrm{C}\mathrm{curl}^\mathrm{R}\mathbf{k}_i)\alpha = \mathrm{curl}[(\mathrm{curl}^\mathrm{R}\mathbf{k}_i)\alpha]$, for all $\alpha \in \mathbb{R}^d$. $\qquad\square$

Consequently, if the coefficients of the Fourier transform $\widehat{\mathbf{k}}$ of a kernel $\mathbf{k}$ are $(h^\|, h^\perp)$, then the coefficients of the Fourier transform of $\mathrm{curl}^\mathrm{C}\mathrm{curl}^\mathrm{R}\mathbf{k}$ must be of the type $(0, h_*)$, where h_* only depends on $h^\perp$. Going back to our original problem of inducing a RKHS structure on $V := \{\mathrm{curl}\, w \,|\, w \in W\}$ from the one on W (Propositions 4.8 and 4.9), this suggests an alternative way of computing the kernel $\mathbf{k}_V$.

Namely, we build a *new* auxiliary Hilbert space $W_0 \hookrightarrow C_0^s(\mathbb{R}^3, \mathbb{R}^3)$, with a *scalar* kernel, i.e. of the form $\mathbf{k}_{W_0}(x) = k_{W_0}(\|x\|)\mathbb{I}_3$, $x \in \mathbb{R}^3$. Assuming $\mathbf{k}_W$ is integrable and divergence-free, the coefficients of its Fourier transform $\widehat{\mathbf{k}}_W$ are $(h_W^\|, h_W^\perp) = (0, h)$, for some nonnegative function $h : \mathbb{R}^+ \to \mathbb{R}$. We define the scalar kernel $\mathbf{k}_{W_0}$ via its Fourier transform, as

$$(53) \qquad \widehat{\mathbf{k}}_{W_0}(\xi) := h_{W_0}(\|\xi\|)\mathbb{I}_3, \;\; \xi \in \mathbb{R}^3, \quad \text{with} \quad h_{W_0} := h;$$

in other words, the coefficients of $\mathbf{k}_{W_0}$ and $\widehat{\mathbf{k}}_{W_0}$ are respectively $k_{W_0}^\| = k_{W_0}^\perp = k_{W_0}$ and $h_{W_0}^\| = h_{W_0}^\perp = h_{W_0}$, where we have set $h_{W_0} = h$, i.e. equal to the non-zero coefficient of the Fourier transform of $\mathbf{k}_W$. Equivalently, and perhaps more interestingly, $\mathbf{k}_{W_0}$ is *the only scalar kernel whose divergence-free component is precisely* $\mathbf{k}_W$. The next proposition relates the kernels $\mathbf{k}_W$ and $\mathbf{k}_{W_0}$ explicitly in the spatial domain (as opposed to the Fourier domain); this is summarized by the first and third rows of the diagram of Figure 8.

Proposition 4.13. *Under the assumptions of Proposition 4.8, if the kernel of W is TRI and integrable, with Fourier transform $\widehat{\mathbf{k}}_W = h(\|\xi\|)\,\mathrm{Pr}_\xi^\perp$, $\xi \in$*

$$\mathbf{k}_W : (k_W^{\parallel}, k_W^{\perp}) \xrightarrow{\;\mathcal{F}\;} \widehat{\mathbf{k}}_W : (h_W^{\parallel}, h_W^{\perp}) = (0, h)$$

$$\Big\downarrow {\scriptstyle -\mathrm{curl}^{\mathrm{C}}\mathrm{curl}^{\mathrm{R}}} \qquad\qquad \Big\downarrow {\scriptstyle (-\mathrm{curl}^{\mathrm{C}}\mathrm{curl}^{\mathrm{R}})^{\wedge}}$$

$$\mathbf{k}_V : (k_V^{\parallel}, k_V^{\perp}) = \big(-\tfrac{2}{r}k'_{W_0},\, -\tfrac{1}{r}k'_{W_0} - k''_{W_0}\big) \xrightarrow{\;\mathcal{F}\;} \widehat{\mathbf{k}}_V : (h_V^{\parallel}, h_V^{\perp}) = (0, (2\pi\varrho)^2 h)$$

$$\Big\uparrow {\scriptstyle -\mathrm{curl}^{\mathrm{C}}\mathrm{curl}^{\mathrm{R}}} \qquad\qquad \Big\uparrow {\scriptstyle (-\mathrm{curl}^{\mathrm{C}}\mathrm{curl}^{\mathrm{R}})^{\wedge}}$$

$$\mathbf{k}_{W_0}(x) = k_{W_0}(\|x\|)\mathbb{I}_3, \quad k_{W_0} = \tfrac{1}{2}k_W^{\parallel} + k_W^{\perp} \xrightarrow{\;\mathcal{F}\;} \widehat{\mathbf{k}}_{W_0} : (h_{W_0}^{\parallel}, h_{W_0}^{\perp}) = (h, h)$$

Figure 8: Construction of the divergence-free kernel $\mathbf{k}_V$ from the divergence-free kernel $\mathbf{k}_W$, and (equivalently) from the auxiliary scalar kernel $\mathbf{k}_{W_0}$, with interpretation in the Fourier domain. Note that while $W \hookrightarrow C_0^s(\mathbb{R}^3, \mathbb{R}^3)$ and $W_0 \hookrightarrow C_0^s(\mathbb{R}^3, \mathbb{R}^3)$, we have $V \hookrightarrow C_0^{s-1}(\mathbb{R}^3, \mathbb{R}^3)$.

$\mathbb{R}^d$, then the kernel $\mathbf{k}_{W_0}$, defined in the Fourier domain by (53), has the form $\mathbf{k}_{W_0} = k_{W_0}(\|x\|)\mathbb{I}_3$, $x \in \mathbb{R}^3$, with

$$k_{W_0} = \frac{1}{2}\, k_W^{\parallel} + k_W^{\perp}.$$

Proof. We have $k_{W_0}(r) = \frac{2\pi}{r^{\mu}} \int_0^{\infty} \varrho^{\mu+1} h(\varrho)\, J_{\mu}(2\pi\varrho r)\, d\varrho$, with $\mu = \frac{d}{2} - 1 = \frac{1}{2}$ (since $d = 3$). In addition, since $(h_W^{\parallel}, h_W^{\perp}) = (0, h)$, by Proposition 3.14

$$k_W^{\parallel}(r) = \frac{2\mu + 1}{r^{\mu+1}} \int_0^{\infty} \varrho^{\mu} h(\varrho)\, J_{\mu+1}(2\pi\varrho r)\, d\varrho \quad (\text{where } 2\mu+1 = d-1 = 2)$$

$$\text{and } \ k_W^{\perp}(r) = \frac{2\pi}{r^{\mu}} \int_0^{\infty} \varrho^{\mu+1} h(\varrho)\, J_{\mu}(2\pi\varrho r)\, d\varrho$$

$$- \frac{1}{r^{\mu+1}} \int_0^{\infty} \varrho^{\mu} h(\varrho)\, J_{\mu+1}(2\pi\varrho r)\, d\varrho \ = \ k_{W_0}(r) - \frac{1}{2}k_W^{\parallel}(r). \qquad \square$$

By Proposition 4.12, since $\mathbf{k}_W$ and $\mathbf{k}_{W_0}$ have the same coefficient $h^{\perp}$ in the Fourier domain, we must have $\mathrm{curl}^{\mathrm{C}}\mathrm{curl}^{\mathrm{R}}\mathbf{k}_W = \mathrm{curl}^{\mathrm{C}}\mathrm{curl}^{\mathrm{R}}\mathbf{k}_{W_0}$; the right-hand side is simpler to compute because $\mathbf{k}_{W_0}$ is characterized by only one coefficient (i.e. the function k_{W_0}) in the spatial domain. This is done next.

Proposition 4.14. *If $W_0 \hookrightarrow C_0^s(\mathbb{R}^3, \mathbb{R}^3)$, with $s \geq 1$, has scalar kernel of the type $\mathbf{k}_{W_0}(x) = k_{W_0}(\|x\|)\mathbb{I}_3$, $x \in \mathbb{R}^3$, and we define $\mathbf{k}_V := -\mathrm{curl}^{\mathrm{C}}\mathrm{curl}^{\mathrm{R}}\mathbf{k}_{W_0}$, then $\mathbf{k}_V(x) = k_V^{\parallel}(\|x\|)\mathrm{Pr}_x^{\parallel} + k_V^{\perp}(\|x\|)\mathrm{Pr}_x^{\perp}$ with*

$$(54) \quad k_V^{\parallel}(r) = -\frac{2}{r}\frac{dk_{W_0}}{dr}(r) \ \text{ and } \ k_V^{\perp}(r) = -\frac{1}{r}\frac{dk_{W_0}}{dr}(r) - \frac{d^2 k_{W_0}}{dr^2}(r), \quad r > 0.$$

Proof. We first compute the curl "by rows" of $\mathbf{k}_{W_0}$:

$$\mathrm{curl}^R\,\mathbf{k}_{W_0}(x) = \frac{k'_W(\|x\|)}{\|x\|}\begin{bmatrix} 0 & x_3 & -x_2 \\ -x_3 & 0 & x_1 \\ x_2 & -x_1 & 0 \end{bmatrix} = f(x)\left[\,v_1(x)\,\middle|\,v_2(x)\,\middle|\,v_3(x)\,\right],$$

where $f(x) := k'_{W_0}(\|x\|)/\|x\|$, and the v_i's are the columns of the above matrix. Now we have

$$\nabla f(x) = \Big(\frac{k''_{W_0}(\|x\|)}{\|x\|^2} - \frac{k'_{W_0}(\|x\|)}{\|x\|^3}\Big)x \quad \text{and} \quad \mathrm{curl}\,v_i(x) = 2e_i, \text{ for } i = 1, 2, 3,$$

so that

$$\mathrm{curl}(f v_1) = \nabla f \times v_1 + f\,\mathrm{curl}\,v_1$$

$$= \Big(\frac{k''_{W_0}(\|x\|)}{\|x\|^2} - \frac{k'_{W_0}(\|x\|)}{\|x\|^3}\Big)\begin{bmatrix} x_2^2 + x_3^2 \\ -x_1 x_2 \\ -x_1 x_3 \end{bmatrix} + 2\frac{k'_{W_0}(\|x\|)}{\|x\|}e_1.$$

Similar computations hold for $\mathrm{curl}(f v_2)$ and $\mathrm{curl}(f v_3)$, which lead to

$$\mathbf{k}_V(x) = -\mathrm{curl}^C\,\mathrm{curl}^R\,\mathbf{k}_{W_0}$$

$$= -\Big(\frac{k''_{W_0}(\|x\|)}{\|x\|^2} - \frac{k'_{W_0}(\|x\|)}{\|x\|^3}\Big)\Big(\|x\|^2\mathbb{I}_3 - xx^T\Big) - 2\frac{k'_{W_0}(\|x\|)}{\|x\|}\mathbb{I}_3$$

$$= -2\frac{k'_{W_0}(\|x\|)}{\|x\|}\mathrm{Pr}_x^{\|} - \Big(k''_{W_0}(\|x\|) + \frac{k'_{W_0}(\|x\|)}{\|x\|}\Big)\mathrm{Pr}_x^{\perp}. \qquad \square$$

Corollary 4.15. *Under the assumptions of Proposition 4.14, the elements of V are divergence-free.*

When $s \geq 2$, we already knew that the above was true from Corollary 4.10. However (similarly to what happens in the construction of curl-free kernel, described in section 4.1) one can easily verify that $k_V^{\|}$ and $k_V^{\perp}$, given by (54), satisfy equation (35) for $d = 3$, and that once again the latter operation only involves *two* derivatives of k_{W_0}, which is twice differentiable when $s = 1$. Therefore Corollary 4.15 also holds when $s = 1$. The proposition that follows characterizes the kernel $\mathbf{k}_V$ in the Fourier domain.

Lemma 4.16. *If $g : \mathbb{R}^3 \to \mathbb{R}$ is a twice-differentiable scalar-valued function and we define $\mathbf{g} = g\mathbb{I}_3$, then it is the case that $\mathrm{curl}^C\mathrm{curl}^R\mathbf{g} = (\Delta\mathbb{I}_3 - \mathrm{Hess})g$.*

Proof. By Lemma 4.11, with $g_{ij} = g\delta_{ij}$, we see immediately that

$$\mathrm{curl}^{\mathrm{C}}\mathrm{curl}^{\mathrm{R}}\mathbf{g} = \begin{bmatrix} (\partial_2^2 + \partial_3^2)g & -\partial_1\partial_2 g & -\partial_1\partial_3 g \\ -\partial_1\partial_2 g & (\partial_1^2 + \partial_3^2)g & -\partial_2\partial_3 g \\ -\partial_1\partial_3 g & -\partial_2\partial_3 g & (\partial_1^2 + \partial_2^2)g \end{bmatrix} = (\Delta\mathbb{I}_3 - \mathrm{Hess})g. \quad \square$$

Proposition 4.17. *Under the assumptions of Proposition 4.14, if $\mathbf{k}_{W_0}$ is integrable and $\widehat{\mathbf{k}}_{W_0}(\xi) = h(\|\xi\|)\mathbb{I}_3$, $\xi \in \mathbb{R}^3$, then $\widehat{\mathbf{k}}_V(\xi) = h_V^{\parallel}(\|\xi\|)\Pr_\xi^{\parallel} + h_V^{\perp}(\|\xi\|)\Pr_\xi^{\perp}$, with $h_V^{\parallel} = 0$ and $h_V^{\perp}(\varrho) = (2\pi\varrho)^2 h(\varrho)$.*

Proof. We have $\mathbf{k}_{W_0}(x) = g(x)\mathbb{I}_3$ with $g(x) := k_{W_0}(\|x\|)$, $x \in \mathbb{R}^3$; by Lemma 4.16, it is the case that $\mathbf{k}_V = -\mathrm{curl}^{\mathrm{C}}\mathrm{curl}^{\mathrm{R}}\mathbf{k}_{W_0} = (\mathrm{Hess} - \Delta\mathbb{I}_3)g$. Now, since $\int_{\mathbb{R}^d} \partial_j\partial_\ell g(x) e^{-2\pi i\xi\cdot x}dx = -(2\pi)^2\xi_j\xi_\ell\,\widehat{g}(\xi)$ and $\widehat{g}(\xi) = h(\|\xi\|)$, we compute

$$\widehat{\mathbf{k}}_V(\xi) = (2\pi)^2\big(\|\xi\|^2\mathbb{I}_3 - \xi\xi^T\big)\,h(\|\xi\|) = (2\pi\|\xi\|)^2\,h(\|\xi\|)\Pr_\xi^{\perp}. \quad \square$$

The situation is completely illustrated by the commutative diagram in Figure 8. In summary, given an arbitrary s-admissible Hilbert space W of divergence-free vector fields with kernel $\mathbf{k}_W$ (i.e. given the corresponding coefficient in the frequency domain $h_W^{\perp} = h$), we can always build another space W_0 with a *scalar* kernel $\mathbf{k}_{W_0}$ such that $\mathrm{curl}^{\mathrm{C}}\mathrm{curl}^{\mathrm{R}}\mathbf{k}_W = \mathrm{curl}^{\mathrm{C}}\mathrm{curl}^{\mathrm{R}}\mathbf{k}_{W_0}$. Then by Proposition 4.9 the kernel of the space $V := \{\mathrm{curl}\,w \,|\, w \in W\}$ may simply be computed as $\mathbf{k}_V = -\mathrm{curl}^{\mathrm{C}}\mathrm{curl}^{\mathrm{R}}\mathbf{k}_{W_0}$, i.e. by equations (54).

In fact, given the arbitrariness of $h_W^{\perp} = h$, we may start from any Hilbert space $W_0 \hookrightarrow C_0^s(\mathbb{R}^3, \mathbb{R}^3)$ with a scalar kernel, compute the new coefficients via (54) and these will correspond to the kernel of $V := \{\mathrm{curl}\,w \,|\, w \in W\}$ for *some* s-admissible space W of divergence-free vector fields, whose kernel $\mathbf{k}_W$ *we do not need to make explicit*. This provides us with a practical procedure for building kernels of divergence-free vector fields. Since $h(\varrho) = (2\pi\varrho)^{-2}h_V^{\parallel}(\varrho)$, we finally note that by formula (16) we have

$$(55) \quad \mathbf{k}_{W_0}(x) = k_{W_0}(\|x\|)\mathbb{I}_3, \quad \text{with} \quad k_{W_0}(r) = \frac{1}{2\pi r^\mu}\int_0^\infty \varrho^{\mu-1}\,h_V^{\parallel}(\varrho)\,J_\mu(2\pi r\varrho)\,d\varrho$$

for $r > 0$ ($\mu = \frac{d}{2} - 1 = \frac{1}{2}$). So the procedure is general, in the sense that any integrable, TRI, divergence-free kernel $\mathbf{k}_V$ of a $(s-1)$-admissible Hilbert space V may be derived by applying formulae (54) to the a scalar kernel $\mathbf{k}_{W_0}$ of a certain Hilbert space $W_0 \hookrightarrow C_0^s(\mathbb{R}^3, \mathbb{R}^3)$, in fact given by formula (55).

Generalization to arbitrary dimensions. Thanks to the interpretation of the differential operator $-\mathrm{curl}^{\mathrm{C}}\mathrm{curl}^{\mathrm{R}}$ in the Fourier domain provided by

Proposition [4.17](#), we may actually extend the above procedure, and formulae [(54)](#), to $\mathbb{R}^{d\times d}$-valued TRI kernels, with arbitrary d. The following holds:

Proposition 4.18. *Let* $\mathbf{k}_{W_0} \hookrightarrow C_0^s(\mathbb{R}^d, \mathbb{R}^d)$ *be integrable and of the form* $\mathbf{k}_{W_0}(x) = k_{W_0}(\|x\|)\mathbb{I}_d$, *with Fourier transform given by* $\widehat{\mathbf{k}}_{W_0}(\xi) = h(\|\xi\|)\mathbb{I}_d$. *If we define* $\mathbf{k}_V$ *via its Fourier transform as* $\widehat{\mathbf{k}}_V(\xi) = h_V^\perp(\|\xi\|)\mathrm{Pr}_\xi^\perp$ *with* $h_V^\perp(\varrho) = (2\pi\varrho)^2 h(\varrho)$, $\varrho > 0$, *then the coefficients of* $\mathbf{k}_V$ *are given, for* $r > 0$, *by*

$$(56) \quad k_V^\parallel(r) = -\frac{d-1}{r}\frac{dk_{W_0}}{dr}(r) \ \text{ and } \ k_V^\perp(r) = -\frac{d-2}{r}\frac{dk_{W_0}}{dr}(r) - \frac{d^2 k_{W_0}}{dr^2}(r),$$

Proof. By [(33a)](#) we have $k_V^\parallel(r) = \frac{2\mu+1}{r^{\mu+1}}(2\pi)^2 \int_0^\infty \varrho^{\mu+2} h(\varrho) J_{\mu+1}(2\pi\varrho r)\, d\varrho$, since $h^\parallel = 0$. We note that

$$\frac{2\pi}{\varrho^{\mu+1}} \int_0^\infty r^{\mu+2}\Big(\frac{1}{r}\frac{dk_{W_0}}{dr}(r)\Big) J_{\mu+1}(2\pi\varrho r)\, d\varrho$$
$$\overset{(*)}{=} -\frac{(2\pi)^2}{\varrho^\mu} \int_0^\infty r^{\mu+1} k_{W_0}(r)\, J_\mu(2\pi\varrho r)\, d\varrho = -2\pi h(\varrho);$$

in step $(*)$ one uses an argument that is in all similar to the computation in [(39)](#). Since the transformation [(15)](#), with $\mu+1$ instead of μ, is an involution, it must be the case that

$$\frac{1}{r}\frac{dk_{W_0}}{dr}(r) = \frac{2\pi}{r^{\mu+1}} \int_0^\infty \varrho^{\mu+2}\left(-2\pi h(\varrho)\right) J_{\mu+1}(2\pi\varrho r)\, d\varrho,$$

which allows us to complete the computation of $k_V^\parallel$ (note that $2\mu+1 = d-1$). To calculate $k_V^\perp$ we simply use formula [(35)](#), which is valid for divergence-free kernels, rewritten as $k_V^\perp = k_V^\parallel + \frac{r}{d-1}\, dk_V^\parallel/dr$. $\qquad\square$

By considerations that are in all similar to those above, we conclude that all divergence-free kernels, in arbitrary dimensions, may be obtained from scalar ones via formulae [(56)](#), which generalize [(54)](#).

Example 1, revisited again. Let $W_0 \hookrightarrow C_0^s(\mathbb{R}^d, \mathbb{R}^d)$ be a RKHS and $\mathbf{k}_{W_0}(x) = k_{W_0}(\|x\|)\mathbb{I}_d$, where $k_{W_0}(r) = \frac{b\sigma^2}{d-1}\exp(-\frac{1}{2}\frac{r^2}{\sigma^2})$, with $b > 0$; W_0 is s-admissible, for all s. Formulae [(56)](#) yield

$$(57) \quad k_V^\parallel(r) = b\,\exp\left(-\frac{1}{2}\frac{r^2}{\sigma^2}\right) \ \text{ and } \ k_V^\perp(r) = \left(b - \frac{b}{(d-1)\sigma^2}\, r^2\right)\exp\left(-\frac{1}{2}\frac{r^2}{\sigma^2}\right),$$

which are precisely the equations of the coefficients (30) of the kernel in Example 1 with $a = b \left[(d-1)\sigma^2\right]^{-1}$, which is the line on the boundary of D_1 that corresponds to divergence-free kernels. See Figure 4(a).

Example 5 (Bessel-type divergence-free kernels)**.** Consider the scalar kernel $\mathbf{k}_{W_0}(x) = k_{W_0}(\|x\|)\mathbb{I}_d$, with k_{W_0} given by the right-hand side of (50); if $\ell > s + d/2$ then $W_0 \hookrightarrow C_0^s(\mathbb{R}^d, \mathbb{R}^d)$. Using Lemma A.1, if we apply formulae (56) to such kernel k_{W_0} we get:

$$(58a) \qquad k_V^{\|}(r) = \frac{C_0}{\sigma^2}(d-1)\left(\frac{r}{\sigma}\right)^{\nu-1} K_{\nu-1}\left(\frac{r}{\sigma}\right),$$

$$(58b) \qquad k_V^{\perp}(r) = \frac{C_0}{\sigma^2}\left(\frac{r}{\sigma}\right)^{\nu-1}\left\{(2\nu + d - 3)\, K_{\nu-1}\left(\frac{r}{\sigma}\right) - \frac{r}{\sigma}\, K_\nu\left(\frac{r}{\sigma}\right)\right\},$$

where $\nu = \ell - \frac{d}{2}$. Once again, the kernel with coefficients (58a) and (58b) has heavier tails than the Gaussian case (57).

5. Application: matching of landmark points

To illustrate an application of the tools that we developed thus far, in this section we shall study the problem of matching feature points, or "landmarks" [30, 35, 42, 69]. Let Ω be an open subset of $\mathbb{R}^m$. The set of N *labeled landmark points in* Ω is defined as:

$$(59) \qquad \mathcal{L}^N(\Omega) := \left\{(P_1, \ldots, P_N) \,:\, P_a \in \Omega,\, P_a \neq P_b \text{ for } a \neq b\right\},$$

which is in fact a manifold of dimension $n = Nm$. The generic element of $\mathcal{L}^N(\Omega)$ is called *landmark set*. For any pair of elements of $\mathcal{L}^N(\Omega)$ we will look for a time-dependent velocity field of "minimal energy" (to be defined) that deforms the first landmark set into the other; we shall consider velocity fields in RKHS of the type described in this paper. This will result in the formulation of a Riemannian distance in $\mathcal{L}^N(\Omega)$. Before proceeding, we first review some some known results.

5.1. Ordinary differential equations and groups of diffeomorphisms

We briefly summarize some facts on ordinary differential equations where the time-dependent velocity fields take their values in RKHS of the type described in this paper, and on the deformations that such dynamical systems generate. The reader is referred to [69] for further details and results.

Consider a 1-admissible Hilbert space $V \hookrightarrow C_0^1(\Omega, \mathbb{R}^d)$, where Ω is an open subset of $\mathbb{R}^m$; for now we shall not introduce further assumptions of

invariance. We indicate with $\mathcal{X}_V^p := L^p([0,1], V)$ the space of V-valued time-dependent vector fields $t \mapsto v(t, \cdot) \in V$ such that

$$\|v\|_{\mathcal{X}_V^p} := \Big(\int_0^1 \|v(t, \cdot)\|_V^p \, dt \Big)^{\frac{1}{p}} < \infty.$$

In particular, $\mathcal{X}_V^2$ is a subset of $\mathcal{X}_V^1$ and is Hilbert itself with the inner product $\langle u, v \rangle_{\mathcal{X}_V^2} := \int_0^1 \langle u, v \rangle_V dt$.

If $V \hookrightarrow C_0^1(\Omega, \mathbb{R}^d)$ then it is the case [69, §8.2] that for any $v \in \mathcal{X}_V^1$, $t_0 \in [0,1]$ and $x_0 \in \Omega$ the initial value problem $\dot{z} = v(t, z)$, $z(t_0) = x_0$ has a unique solution of the type $z(t) = \psi(t, t_0, x)$, for $t \in [0,1]$. We define the function $\varphi_{st}^v(x) := \psi(t, s, x)$ and call it the *flow* associated to v; it has the group property $\varphi_{st}^v = \varphi_{rt}^v \circ \varphi_{sr}^v$ for all $r, s, t \in [0,1]$ and $v \in \mathcal{X}_V^1$. In fact, under the assumption that $V \hookrightarrow C_0^1(\Omega, \mathbb{R}^d)$, we have that for all $v \in \mathcal{X}_V^1$ and $s, t \in [0,1]$ the map $\varphi_{st}^v : \Omega \to \Omega$ is a *diffeomorphism* of Ω.

If $V \hookrightarrow C_0^1(\Omega, \mathbb{R}^d)$, we denote with $\mathcal{G}_V := \{\varphi_{01}^v : v \in \mathcal{X}_V^1\}$ the set of diffeomorphisms provided by flows associated to elements of $v \in \mathcal{X}_V^1$ at time 1. For any $\psi_1, \psi_2 \in \mathcal{G}_V$ we define

$$d_V(\psi_1, \psi_2) := \inf_{v \in \mathcal{X}_V^1} \Big\{ \|v\|_{\mathcal{X}_V^1} : \psi_2 = \psi_1 \circ \varphi_{01}^v \Big\};$$

it was proven by A. Trouvé [69, Theorem 8.15] that the function d_V is a *distance* on $\mathcal{G}_V$, and that $(\mathcal{G}_V, d_V)$ is a complete metric space. For any $\psi_1, \psi_2 \in \mathcal{G}_V$, their distance may in fact be computed as follows: $d_V(\psi_1, \psi_2) = \inf_{v \in \mathcal{X}_V^2} \{\|v\|_{\mathcal{X}_V^2} : \psi_2 = \psi_1 \circ \varphi_{01}^v\}$; that is, the search of v may be restricted to the set $\mathcal{X}_V^2 \subset \mathcal{X}_V^1$. The infimum attained in $\mathcal{X}_V^2$: more precisely, for all $\psi_1, \psi_2 \in \mathcal{G}_V$ there exists a time-dependent vector field $v \in \mathcal{X}_V^2$ such that $d_V(\psi_1, \psi_2) = \|v\|_{\mathcal{X}_V^1} = \|v\|_{\mathcal{X}_V^2}$, and it is such that $\|v(t, \cdot)\|_V$ is constant with respect to $t \in [0,1]$. See [69, Chapter 8] for details.

5.2. Interpolation of vector fields

Let V be a non-degenerate RKHS with kernel $K : \Omega \times \Omega \to \mathbb{R}^{d \times d}$, where Ω is an open subset of $\mathbb{R}^m$; for now we shall introduce no further assumption of translation- and rotation-invariance, or differentiability, of the kernel. We consider the following problem.

Interpolation Problem 1 (IP$_1$). *For any landmark set $S = (x_1, \ldots, x_N) \in \mathcal{L}^N(\Omega)$ and any N-tuple of vectors $(\beta_1, \ldots, \beta_N) \in (\mathbb{R}^d)^N$, find a vector field $u \in V$ such that*

$$\begin{cases} u(x_a) = \beta_a \text{ for } 1 \le a \le N, \text{ and} \\ \|u\|_V \text{ is minimal.} \end{cases}$$

For this purpose, for a fixed $S = (x_1, \ldots, x_N) \in \mathcal{L}^N(\Omega)$ we define the linear subspace of V:

$$V_S := \{u \in V \mid u(x_a) = 0 \text{ for all } a = 1, \ldots, N\}.$$

Its orthogonal complement with respect to the inner product $\langle \cdot, \cdot \rangle_V$ is given by the following result.

Lemma 5.1. *If $S = (x_1, \ldots, x_N) \in \mathcal{L}^N(\Omega)$, then $V_S^\perp = \text{span}\{K(\cdot, x_a)\alpha \mid a = 1, \ldots, N, \alpha \in \mathbb{R}^d\}$.*

Proof. We have that $u \in V_S$ if and only if

$$\sum_{a=1}^N \alpha_a \cdot u(x_a) = \left\langle \sum_{a=1}^N K(\cdot, x_a)\alpha_a, u \right\rangle_V = 0$$

for all $\alpha_1, \ldots, \alpha_N \in \mathbb{R}^d$. So $V_S = \text{span}\{K(\cdot, x_a)\alpha \mid a = 1, \ldots, N, \alpha \in \mathbb{R}^d\}^\perp$. But $\{K(\cdot, x_a)\alpha \mid a = 1, \ldots, N, \alpha \in \mathbb{R}^d\}$ has finite dimension (and whence it is closed), so the claim follows immediately. $\square$

Proposition 5.2. *Fix $S = (x_1, \ldots, x_N) \in \mathcal{L}^N(\Omega)$. For any $u \in V$, there exists a unique $u_* \in V_S^\perp$ such that $u_*(x_i) = u(x_i)$ for all $i = 1, \ldots, N$; it is given by the orthogonal projection of u onto $V_S^\perp$.*

Proof. Let u_* be the orthogonal projection of u onto $V_S^\perp$; then $u - u_* \in V_S$, i.e. $u(x_a) = u_*(x_a)$ for $a = 1, \ldots, N$. On the other hand, if $u_* \in V_S^\perp$ is such that $u(x_a) = u_*(x_a)$ for $a = 1, \ldots, N$, then it is the case that $u - u_* \in V_S$, i.e. u_* is the orthogonal projection of u onto $V_S^\perp$. $\square$

Corollary 5.3. *Let V be a RKHS. If a solution u to IP_1 exists, then $u \in V_S^\perp$. If $u \in V_S^\perp$ is a solution to IP_1 restricted to $V_S^\perp$, then it is a solution to IP_1 on all of V.*

Therefore if the solution to IP_1 exists then it must be in $V_S^\perp$, i.e. of the form

(60)
$$u(\cdot) = \sum_{b=1}^N K(\cdot, x_b)\alpha_b,$$

for some choice of vectors $\alpha_1, \ldots, \alpha_N \in \mathbb{R}^d$. The squared norm of such vector fields may be written as $\|u\|_V^2 = \sum_{a,b=1}^N \alpha_a \cdot K(x_a, x_b)\alpha_b$. Inserting $x_1, \ldots, x_N$ in the above expression and imposing the conditions $u(x_a) =$

β_a, $a = 1, \ldots, N$, we see that we are looking for vectors $\alpha_1, \ldots, \alpha_N$ such that

$$(61) \qquad \beta_a = \sum_{b=1}^{N} K(x_a, x_b)\alpha_b, \qquad a = 1, \ldots, N.$$

Introducing now the vectors in $\mathbb{R}^{Nd}$ and the matrix in $\mathbb{R}^{Nd \times Nd}$

$$(62)$$
$$\boldsymbol{\alpha} := \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix}, \ \boldsymbol{\beta} := \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_N \end{bmatrix}, \ \mathbf{K}(\boldsymbol{x}) := \begin{bmatrix} K(x_1, x_1) & K(x_1, x_2) & \cdots & K(x_1, x_N) \\ K(x_2, x_1) & K(x_2, x_2) & \cdots & K(x_2, x_N) \\ \vdots & \vdots & \ddots & \vdots \\ K(x_N, x_1) & K(x_N, x_2) & \cdots & K(x_N, x_N) \end{bmatrix},$$

we may rewrite (61) simply as $\boldsymbol{\beta} = \mathbf{K}(\boldsymbol{x})\boldsymbol{\alpha}$, while $\|u\|_V^2 = \boldsymbol{\alpha} \cdot \mathbf{K}(\boldsymbol{x})\boldsymbol{\alpha}$. Note that the matrix $\mathbf{K}(\boldsymbol{x})$ is invertible (in fact if $\mathbf{K}(\boldsymbol{x})\boldsymbol{\alpha} = 0$ for some $\boldsymbol{\alpha} \in \mathbb{R}^{Nd}$ then $\boldsymbol{\alpha} \cdot \mathbf{K}(\boldsymbol{x})\boldsymbol{\alpha} = 0$; but $\boldsymbol{\alpha} \cdot \mathbf{K}(\boldsymbol{x})\boldsymbol{\alpha}$ is the left-hand side of (3), so if V is non-degenerate we must have $\boldsymbol{\alpha} = 0$). Therefore the only solution to $\boldsymbol{\beta} = \mathbf{K}(\boldsymbol{x})\boldsymbol{\alpha}$ is $\boldsymbol{\alpha} = \mathbf{K}(\boldsymbol{x})^{-1}\boldsymbol{\beta}$. In conclusion, there exists a unique solution to IP$_1$ when this is restricted to $V_S^\perp$, which is the vector field (60) with $\boldsymbol{\alpha} = \mathbf{K}(\boldsymbol{x})^{-1}\boldsymbol{\beta}$. By Corollary 5.3 this is the only solution to IP$_1$ (on all of V).

5.3. Landmark matching via diffeomorphisms, with differential geometric interpretation

We will now formulate and provide a method for solving the more difficult problem of matching landmark sets with flows of diffeomorphisms. For $\Omega \subseteq \mathbb{R}^d$ open and connected, we introduce the further assumption that the non-degenerate RKHS is 1-admissible, i.e. $V \hookrightarrow C_0^1(\Omega, \mathbb{R}^d)$.

Interpolation Problem 2 (IP$_2$). *For any two sets $I = (x_1, \ldots, x_N)$ and $J = (y_1, \ldots, y_N)$ in $\mathcal{L}^N(\Omega)$, find a time-dependent vector field $v \in \mathcal{X}_V^2$ such that*

$$\begin{cases} \varphi_{01}^v(x_a) = y_a \text{ for } 1 \le a \le N, \text{ and} \\ \|v\|_{\mathcal{X}_V^2} \text{ is minimal.} \end{cases}$$

For $v \in \mathcal{X}_V^2$, $S = (x_1, \ldots, x_N) \in \mathcal{L}^N(\Omega)$ and $s, t \in [0, 1]$, we let $\varphi_{st}^v(S) := (\varphi_{st}^v(x_1), \ldots, \varphi_{st}^v(x_N))$, where φ_{st}^v is the flow associated to v. Since φ_{st}^v is a diffeomorphism we have $\varphi_{st}^v(x_a) \ne \varphi_{st}^v(x_b)$ for $a \ne b$, whence $\varphi_{st}^v(S) \in \mathcal{L}^N(\Omega)$. We also introduce the subset of $\mathcal{X}_V^2$:

$$\mathcal{X}_V^2(S) := \left\{ v \in \mathcal{X}_V^2 \,\middle|\, v(t, \cdot) \in V_{\varphi_{0t}^v(S)}^\perp, \text{ for all } t \in [0, 1] \right\}.$$

Theorem 5.4. *Assume that V is a 1-admissible Hilbert space, and let $S = (x_1, \ldots, x_N) \in \mathcal{L}^N(\Omega)$. For any $v \in \mathcal{X}_V^2$ there exists a vector field $v_* \in \mathcal{X}_V^2(S)$ such that $\varphi_{0t}^{v_*}(x_a) = \varphi_{0t}^v(x_a)$ for all $t \in [0,1]$ and $a = 1 \ldots, N$, and it is such that $\|v_*(t, \cdot)\|_V \leq \|v(t, \cdot)\|_V$ for all $t \in [0,1]$.*

Proof. Fix $v \in \mathcal{X}_V^2$, and for all $t \in [0,1]$ let $v_*(t, \cdot)$ be the projection of $v(t, \cdot) \in V$ onto $V_{\varphi_{0t}^v(S)}^\perp$; by Corollary 5.3 it is in fact the vector field of minimal norm $\|\cdot\|_V$ such that, for all $a = 1, \ldots, N$, $v_*(t, \varphi_{0t}^v(x_a)) = v(t, \varphi_{0t}^v(x_a))$. Therefore, by the uniqueness of solutions to ordinary differential equations (discussed in §5.1), it must be the case that $\varphi_{0t}^{v_*}(x_a) = \varphi_{0t}^v(x_a)$, for all $t \in [0,1]$ and $a = 1, \ldots, N$. This also implies that $V_{\varphi_{0t}^{v_*}(S)} = V_{\varphi_{0t}^v(S)}$, therefore $v_*(t, \cdot) \in V_{\varphi_{0t}^{v_*}(S)}$ for all $t \in [0,1]$, i.e. $v_* \in \mathcal{X}_V^2(S)$. Also, by construction we have that $\|v_*(t, \cdot)\|_V \leq \|v(t, \cdot)\|_V$ for all $t \in [0,1]$. $\square$

Whence the solutions of the problem IP_2 must be searched among the vector fields of the form

$$(63) \qquad v(t, \cdot) = \sum_{b=1}^N K\big(\cdot, x_b(t)\big)\alpha_b(t), \qquad t \in [0,1],$$

for some set of functions $\alpha_a : [0,1] \to \mathbb{R}^d$, $a = 1, \ldots, N$, where we have introduced for simplicity the notation $x_a(t) := \varphi_{0t}^v(x_a)$, $a = 1, \ldots, N$ (the above equation is precisely expression (1) in the introduction). Note therefore that expression (63) for v is implicit, in that the trajectories x_a, $a = 1, \ldots, N$, also depend on v. We can parametrize the search space $\mathcal{X}_V^2(S)$ of the solution to IP_2 by the functions α_a, $a = 1, \ldots, N$. Similarly to what was done earlier, for all $t \in [0,1]$ we introduce the vectors in $\mathbb{R}^{Nd}$ and matrix in $\mathbb{R}^{Nd \times Nd}$:

$$(64) \quad \boldsymbol{\alpha}(t) := \begin{bmatrix} \alpha_1(t) \\ \vdots \\ \alpha_N(t) \end{bmatrix}, \quad \boldsymbol{x}(t) := \begin{bmatrix} x_1(t) \\ \vdots \\ x_N(t) \end{bmatrix},$$

$$\mathbf{K}\big(\boldsymbol{x}(t)\big) := \begin{bmatrix} K\big(x_1(t), x_1(t)\big) & \cdots & K\big(x_1(t), x_N(t)\big) \\ K\big(x_2(t), x_1(t)\big) & \cdots & K\big(x_2(t), x_N(t)\big) \\ \vdots & \ddots & \vdots \\ K\big(x_N(t), x_1(t)\big) & \cdots & K\big(x_N(t), x_N(t)\big) \end{bmatrix}.$$

Note that since we assumed that $V \hookrightarrow C_0^1(\Omega, \mathbb{R}^d)$ is non-degenerate, the matrix $\mathbf{K}(\boldsymbol{x}(t))$ is invertible for all $t \in [0,1]$. Note also that the norm of the vector fields of the type (63) may be written as $\|v(t, \cdot)\|_V = \boldsymbol{\alpha}(t) \cdot \mathbf{K}(\boldsymbol{x}(t))\boldsymbol{\alpha}(t)$, by the reproducing property of K. Therefore the solutions

to IP_2 must be searched in the space $\mathcal{X}_V^2(S)$, whose elements have norm that may also be parametrized by the functions α_a, $a = 1, \ldots, N$, as follows:
(65)
$$\|v\|_{\mathcal{X}_V^2}^2 = \int_0^1 \sum_{a,b=1}^N \alpha_a(t) \cdot K\big(x_a(t), x_b(t)\big)\alpha_b(t)\, dt = \int_0^1 \boldsymbol{\alpha}(t) \cdot \mathbf{K}(\boldsymbol{x}(t))\boldsymbol{\alpha}(t)\, dt,$$

where, we remind the reader, $x_a(t) := \varphi_{0t}^v(x_a)$, for $t \in [0, 1]$ and $a = 1, \ldots, N$. Therefore, by definition, it must be the case that $\dot{x}_a(t) = v(t, x_a(t))$; by expression (63), we must have

$$\dot{x}_a(t) = \sum_{b=1}^N K\big(x_a(t), x_b(t)\big)\alpha_b(t), \qquad t \in [0, 1], \quad a = 1, \ldots, N,$$

or, with the compact notation introduced in (64), $\dot{\boldsymbol{x}}(t) = \mathbf{K}(\boldsymbol{x}(t))\boldsymbol{\alpha}(t)$, for $t \in [0, 1]$. We conclude that the squared norm (65) of vector fields in $\mathcal{X}_V^2$ of the type (63) may be rewritten as:

(66)
$$\|v\|_{\mathcal{X}_V^2}^2 = \int_0^1 \dot{\boldsymbol{x}}(t) \cdot \mathbf{K}(\boldsymbol{x}(t))^{-1}\dot{\boldsymbol{x}}(t)\, dt,$$

where $x_a : [0, 1] \to \mathcal{L}^N(\Omega)$, and $a = 1, \ldots, N$, are the trajectories determined by the vector field v.

We note that the right-hand side of (66) may be interpreted in differential geometric terms [36, 37] as the *energy of a curve* $t \mapsto \boldsymbol{x}(t)$ on the Nd-dimensional manifold $\mathcal{L}^N(\Omega)$ with respect to a Riemannian metric (remember that $\Omega \subseteq \mathbb{R}^d$, as specified earlier). In the coordinates of the landmark points, which we indicate here generically as $(q^1, \ldots, q^N) \in \mathcal{L}^N(\Omega)$ with $q^a = (q^{a1}, \ldots, q^{ad}) \in \mathbb{R}^d$ for $a = 1, \ldots, N$, the corresponding *metric tensor* is written as the $Nm \times Nm$ matrix-valued function:

(67) $\quad g(q^1, \ldots, q^N) := \begin{bmatrix} K(q^1, q^1) & K(q^1, q^2) & \cdots & K(q^1, q^N) \\ K(q^2, q^1) & K(q^2, q^2) & \cdots & K(q^2, q^N) \\ \vdots & \vdots & \ddots & \vdots \\ K(q^N, q^1) & K(q^N, q^2) & \cdots & K(q^N, q^N) \end{bmatrix}^{-1}.$

We now introduce notation from mechanics [3], namely $q^a(t) := x_a(t) \in \mathbb{R}^{Nd}$ and $p_a(t) := \alpha_a(t) \in \mathbb{R}^{Nd}$, with $t \in [0, 1]$ and $a = 1, \ldots, N$, for landmark *positions* and *momenta*, respectively; also, we shall write $q(t) := \boldsymbol{x}(t)$

and $p(t) := \boldsymbol{\alpha}(t)$, $t \in [0, 1]$. Whence $\dot{\boldsymbol{x}}(t) = \mathbf{K}(\boldsymbol{x}(t))\boldsymbol{\alpha}(t)$ translates into

$$(68) \qquad \dot{q}^a(t) = \sum_{b=1}^{N} K\big(q^a(t), q^b(t)\big)\, p_b(t), \qquad t \in [0, 1], \quad a = 1, \ldots, N,$$

which is nothing but the "sharp" isomorphism of differential geometry (raising of the indices) with respect to the metric tensor (67), that maps cotangent vectors into tangent vectors (this justifies the upper location of the indices in q^a, $a = 1, \ldots, N$). From the energy of the curve (66) we deduce immediately the form of the Lagrangian function [3, 36] for landmarks, that is $L(q, \dot{q}) = \frac{1}{2}\dot{q} \cdot g(q)\dot{q}$, and the Hamiltonian function which we may write as:

$$(69) \qquad H(p, q) = \frac{1}{2} p \cdot \big(g(q)\big)^{-1} p = \frac{1}{2} \sum_{a,b=1}^{N} p_a \cdot K(q^a, q^b) p_b$$

$$= \frac{1}{2} \sum_{a,b=1}^{N} \sum_{i,j=1}^{d} K_{ij}(q^a, q^b) p_{ai} p_{bj},$$

where we have explicitly written the components of the momenta $p_a = (p_{a1}, \ldots, p_{ad})$, $a = 1, \ldots, N$, and we have indicated with K_{ij} the $(i, j)^{\text{th}}$ element of the matrix-valued kernel K. We shall use (69) in the next section to study the dynamics induced by the landmark matching problem.

Discussion. In light of the above, the landmark trajectories resulting from vector fields $v \in \mathcal{X}_V^2$ that solve IP$_2$ will be length-minimizing geodesics in $\mathcal{L}^N(\Omega)$, with respect to the metric tensor (67), that connect the landmark sets $I = (x_1, \ldots, x_N)$ and $J = (y_1, \ldots, y_N)$. Once such length-minimizing geodesics $q : [0, 1] \to \mathcal{L}^N(\Omega)$ are known, the momenta are computed via the inversion of the equations (68) (i.e. by $p(t) = \mathbf{K}(q(t))^{-1}\dot{q}(t)$, $t \in [0, 1]$), while the corresponding $v \in \mathcal{X}_V^2$ is obtained via formula (63). We also have, by the results summarized in §5.1, that the resulting geodesic distance between the landmark sets $I = (x_1, \ldots, x_N)$ and $J = (y_1, \ldots, y_N)$ in $\mathcal{L}^N(\Omega)$ is equal to $d_V(\varphi_{01}^v, \mathrm{id})$, i.e. the distance in the group of diffeomorphisms $\mathcal{G}_V$ between the φ_{01}^v and the identity diffeomorphism.

Finally, we note that the shape of the geodesics in $\mathcal{L}^N(\Omega)$ must clearly depend on the initial choice of the Hilbert space $V \hookrightarrow C_0^1(\Omega, \mathbb{R}^d)$, i.e. on its kernel; some examples will be seen later. We refer the reader to [42] for a study of the geometry (in particular, of the sectional curvature) of $\mathcal{L}^N(\mathbb{R}^d)$ equipped with the metric (67) in the case of scalar kernels, i.e. of the form $K(x, y) = k(\|x - y\|)\mathbb{I}_d$.

5.4. Hamiltonian dynamics

In the previous section we derived the Hamiltonian induced by the matching problem for landmark sets. We shall now develop Hamilton's equations, i.e. the geodesic equations in $\mathcal{L}^N(\Omega)$ with respect to the metric tensor (67). Also, we will consider the particular case of TRI kernels.

Proposition 5.5. *Let $V \hookrightarrow C_0^1(\Omega, \mathbb{R}^d)$, with $\Omega \subseteq \mathbb{R}^d$ open and connected, be a RKHS with kernel K. Hamilton's equations for the manifold $\mathcal{L}^N(\mathbb{R}^d)$ equipped with the metric tensor* (67) *are given by:*

$$
\text{(70a)} \qquad\qquad \dot{q}^a = \sum_{b=1}^{N} K(q^a, q^b)\, p_b
$$

$$
\text{(70b)} \qquad\qquad \dot{p}_{ai} = -\sum_{b=1}^{N} \sum_{j,k=1}^{d} \partial_{1,i} K_{jk}(q^a, q^b)\, p_{aj} p_{bk}
$$

for $a = 1, \ldots, N$ and $i = 1, \ldots, d$.

As in section 4, $\partial_{n,i} K$ in (70b) indicates the partial derivative of the map $(x_1, \ldots, x_d, y_1, \ldots, y_d) \mapsto K(x, y)$ with respect to the i-th component $(i = 1, \ldots, d)$ of the n-th variable $(n = 1, 2)$.

Proof. Computing $\dot{q}^{ai} = \frac{\partial H}{\partial p_{ai}}$ yields the first equation immediately. Also:

$$
\frac{\partial}{\partial q^{ai}} K_{jk}\big(q^{b1}, \ldots, q^{bD}, q^{c1}, \ldots, q^{cD}\big)
$$
$$
= \sum_{\ell=1}^{D} \big(\partial_{1,\ell} K_{jk}(q^b, q^c)\, \tfrac{\partial q^{b\ell}}{\partial q^{ai}} + \partial_{2,\ell} K_{jk}(q^b, q^c)\, \tfrac{\partial q^{c\ell}}{\partial q^{ai}}\big)
$$
$$
= \sum_{\ell=1}^{D} \big(\partial_{1,\ell} K_{jk}(q^b, q^c)\, \delta_a^b \delta_i^\ell + \partial_{2,\ell} K_{jk}(q^b, q^c)\, \delta_a^c \delta_i^\ell\big)
$$
$$
= \partial_{1,i} K_{jk}(q^b, q^c)\, \delta_a^b + \partial_{2,i} K_{jk}(q^b, q^c)\, \delta_a^c.
$$

Since $K_{jk}(x, y) = K_{kj}(y, x)$ we also have $\partial_{2,i} K_{jk}(x, y) = \partial_{1,i} K_{kj}(y, x)$, thus

$$
\dot{p}_{ai} = -\frac{\partial H}{\partial q^{ck}} = -\tfrac{1}{2} \sum_{b,c=1}^{N} \sum_{j,k=1}^{D} \big(\partial_{1,i} K_{jk}(q^b, q^c)\, \delta_a^b + \partial_{1,i} K_{kj}(q^c, q^b)\, \delta_a^c\big) p_{bj} p_{ck}
$$
$$
= -\tfrac{1}{2} \sum_{j,k=1}^{D} \big(\sum_{c=1}^{N} \partial_{1,i} K_{jk}(q^a, q^c) p_{aj} p_{ck} + \sum_{b=1}^{N} \partial_{1,i} K_{kj}(q^a, q^b) p_{bj} p_{ak} \big)
$$
$$
= - \sum_{j,k=1}^{D} \sum_{c=1}^{N} \partial_{1,i} K_{jk}(q^a, q^c) p_{aj} p_{ck}
$$

where, in the last step, we have simply relabeled the indices. $\qquad\qquad \square$

We now assume that $\Omega = \mathbb{R}^d$ and that the kernel of V is translation-invariant, i.e. $K(x, y) = \mathbf{k}(x - y)$, for some function $\mathbf{k} : \mathbb{R}^d \to \mathbb{R}^{d \times d}$. In this case Hamilton's equations obviously become:

$$(71\text{a}) \qquad \dot{q}^a = \sum_{b=1}^{N} \mathbf{k}(q^a - q^b)\, p_b$$

$$(71\text{b}) \qquad \dot{p}_{ai} = -\sum_{b=1}^{N} p_a \cdot \frac{\partial \mathbf{k}}{\partial x_i}(q^a - q^b)\, p_b$$

with $a = 1, \ldots, N$ and $i = 1, \ldots, d$. In the translation- *and* rotation-invariant case (TRI kernels) the equations take a form that is determined by differentiating the general expression (10).

Proposition 5.6. *Consider a Hilbert space $V \hookrightarrow C_0^1(\mathbb{R}^d, \mathbb{R}^d)$ with a TRI kernel. Then*

$$(72) \qquad \frac{\partial \mathbf{k}}{\partial x_i}(x) = \frac{x_i}{\|x\|}\left[\frac{dk^\|}{dr}(\|x\|)\Pr^\|(x) + \frac{dk^\perp}{dr}(\|x\|)\Pr^\perp(x)\right]$$
$$+ \|x\|\,\tilde{k}(\|x\|)\left[\frac{e_i x^T + x e_i^T}{\|x\|} - 2\frac{x_i}{\|x\|}\Pr^\|(x)\right],$$

for $i = 1, \ldots, d$, where $\tilde{k}$ is given by (11).

Proof. We differentiate the generic matrix element (12) with respect to x_i:

$$\frac{\partial \mathbf{k}_{jk}}{\partial x_i}(x)$$
$$= \frac{x_j x_k}{\|x\|^2}\frac{\partial}{\partial x_i}\left[k^\|(\|x\|) - k^\perp(\|x\|)\right] + \left[k^\|(\|x\|) - k^\perp(\|x\|)\right]\frac{\partial}{\partial x_i}\frac{x_j x_k}{\|x\|^2} + \frac{\partial}{\partial x_i}k^\perp(\|x\|)\,\delta_{jk}$$
$$= \left[\frac{dk^\|}{dr} - \frac{dk^\perp}{dr}\right]\frac{x_i x_j x_k}{\|x\|^3} + \|x\|^2\left(\frac{\delta_{ij}x_k + \delta_{ik}x_j}{\|x\|^2} - \frac{2x_i x_j x_k}{\|x\|^4}\right)\tilde{k} + \frac{dk^\perp}{dr}\frac{x_i}{\|x\|}\delta_{jk},$$

where δ_{ij} is Kronecker's delta. We note that we have $(e_i x^T)_{jk} = x_k \delta_{ij}$ and $(x e_i^T)_{jk} = x_j \delta_{ik}$, so we can rewrite the above in matrix form as follows:

$$\frac{\partial \mathbf{k}}{\partial x_i}(x) = \frac{x_i}{\|x\|}\left[\frac{dk^\|}{dr}(\|x\|) - \frac{dk^\perp}{dr}(\|x\|)\right]\frac{x x^T}{\|x\|^2}$$
$$+ \|x\|\,\tilde{k}(\|x\|)\left(\frac{e_i x^T + x e_i^T}{\|x\|} - \frac{2x_i}{\|x\|}\frac{x x^T}{\|x\|^2}\right) + \frac{dk^\perp}{dr}(\|x\|)\frac{x_i}{\|x\|}\,\mathbb{I}_d;$$

the proof is completed by using the the projection operators (8). $\qquad\square$

Corollary 5.7. *Under the assumptions of Proposition 5.6, if V is divergence-free, i.e. if condition (35) also holds, the partial derivatives (72) be-*

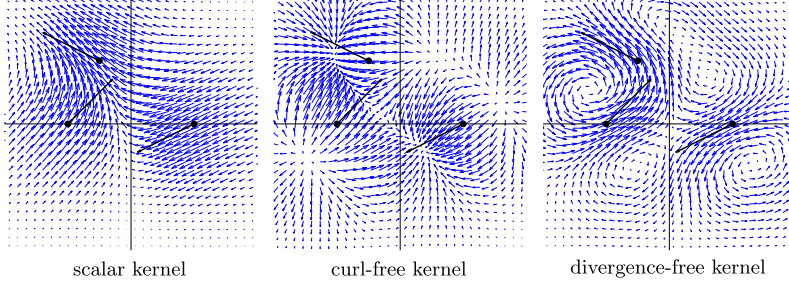scalar kernel   curl-free kernel   divergence-free kernel

Figure 9: Plots of the vector field $v(x) = \sum_{a=1}^{3} \mathbf{k}(x - q^a)\alpha_a$ for three different kernels, with the same landmarks q^a and momenta α_a, $a = 1, 2, 3$.

come

$$\frac{\partial \mathbf{k}}{\partial x_i}(x) = \frac{x_i}{\|x\|} \left[\frac{d+1}{d-1} \frac{dk^{\|}}{dr}(\|x\|) \operatorname{Pr}^{\|}(x) + \frac{dk^{\perp}}{dr}(\|x\|) \operatorname{Pr}^{\perp}(x) \right]$$
$$- \frac{1}{d-1} \frac{dk^{\|}}{dr}(\|x\|) \frac{e_i x^T + x e_i^T}{\|x\|}.$$

Corollary 5.8. *Under the assumptions of Proposition 5.6, if V is curl-free, i.e. if condition (38) also holds, the partial derivatives (72) become*

$$\frac{\partial \mathbf{k}}{\partial x_i}(x) = \frac{x_i}{\|x\|} \left[\left(\frac{dk^{\|}}{dr}(\|x\|) - 2\frac{dk^{\perp}}{dr}(\|x\|) \right) \operatorname{Pr}^{\|}(x) \right.$$
$$\left. + \frac{dk^{\perp}}{dr}(\|x\|) \operatorname{Pr}^{\perp}(x) \right] + \frac{dk^{\perp}}{dr}(\|x\|) \frac{e_i x^T + x e_i^T}{\|x\|}.$$

For any given choice of differentiable coefficients $(k^{\|}, k^{\perp})$, the expression (72) can be inserted into (71b), and Hamilton's equations may be solved numerically to yield the landmark trajectories. Once these are known, the time dependent velocity field (63) can be computed (with $x_a = q^a$ and $\alpha_a = p_a$); integrating numerically the differential equation $\partial_t \varphi_{0t}^v(t, x) = v(t, \varphi_{0t}^v(x))$, with initial condition $\varphi_{00}^v(x) = x$, yields the diffeomorphisms φ_{01}^v. This is shown in the next section for different choices of the TRI kernel $\mathbf{k}$.

6. Numerical results

Figure 9 shows the vector field $u(x) = \sum_{a=1}^{3} \mathbf{k}(x - q^a)\alpha_a$, which is of the type (63), with landmarks given by $q^1 = (-1, 0)$, $q^2 = (-1/2, 1)$, $q^3 = (1, 0)$ in $\mathcal{L}^3(\mathbb{R}^2)$, and momenta $\alpha_1 = (1/\sqrt{2}, 1/\sqrt{2})$, $\alpha_2 = (-2/\sqrt{5}, 1/\sqrt{5})$, $\alpha_3 =$

$(-2/\sqrt{5}, -1/\sqrt{5})$ (so in this case $\|\alpha_a\|_{\mathbb{R}^2} = 1$ for $a = 1, 2, 3$), but for three different choices of the TRI kernel $\mathbf{k}$. The first one is the Gaussian scalar kernel $\mathbf{k}(x) = k(\|x\|)\mathbb{I}_2$ where $k(r) = \exp(-\frac{1}{2}r^2/\sigma^2)$, with $2\sigma^2 = 1$, the second one is the kernel (31) from Example 2, and the third one is the kernel (30) from Example 1; in the last two cases, $a = 2$, $b = 1$, $2\sigma^2 = 1$ and $d = 2$, so that the two kernels are, respectively, curl-free and divegence-free. It should be apparent that given the completely different nature of the kernels and the corresponding vector fields, so should be the landmark trajectories obtained by integrating Hamilton's equation (71a) and (71b) as well as the corresponding diffeomorphisms of the ambient space. Here we shall give examples of such trajectories and diffeomorphisms for the above choices of the kernel.

6.1. Landmarks shooting

Figure 10 shows the landmark trajectories that result from shooting landmarks, i.e. from solving Hamilton's equations (71a) and (71b) as an initial value problem, namely with specified initial landmark sets in $\mathcal{L}^2(\mathbb{R}^2)$ and initial momenta. The grid in each graph represent the corresponding diffeomorphism φ_{01}^v, which is obtained from the landmark trajectories by integrating numerically the differential equation $\partial_t\varphi_{0t}^v(t,x) = v(t, \varphi_{0t}^v(x))$ with initial condition $\varphi_{00}^v(x) = x$, where v is given by equation (63). This is done for the above three choices of Gaussian kernels (scalar, curl-free, and divergence-free), with parameters $\sigma^2 = 1/32$, $b = \sigma^2$, $a = b\sigma^2$ and obviously $d = 2$ since we are working in two dimensions. Specifically, the first column shows the numerical solution of the above problem when the initial landmark set is $(q^1, q^2) = ((0,0), (0, 0.15))$ and the momenta are parallel and both oriented to the right, namely $(p_1, p_2) = ((15, 0), (15, 0))$.

The geodesics in the scalar case had already been examined in [42]; we note that their behavior in the curl-free case is qualitatively similar. By Liouville's Theorem [3] a time-dependent divergence-free vector field v generates a flow φ_{0t}^v that preserves volume; such preservation shows clearly for the diffeomorphism in the last figure on the right, which is distinctly different from the previous two in the shape of the landmark geodesics as well. On the other hand, the second column illustrates the geodesic curves resulting from shooting landmarks in parallel but opposite directions: that is, with initial configuration $(q^1, q^2) = ((-0.4, -0.125), (0.4, 0.125))$ and initial momenta $(p_1, p_2) = ((20, 0), (-20, 0))$. Once again, we notice that volume is preserved in the diffeomorphism induced by the divergence-free kernel; the path followed by the landmarks in this case is also characteristically different, in that in order to avoid compression the two landmarks do not spiral but tend to steer away from each other.
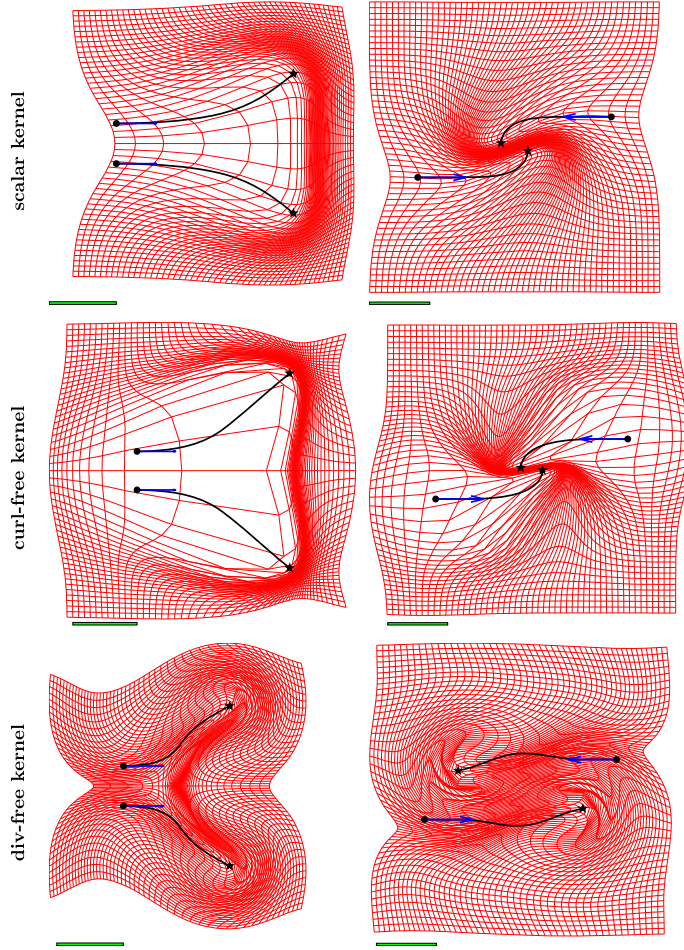
Figure 10: Landmark trajectories in $\mathcal{L}^2(\mathbb{R}^2)$ for three different kernels (scalar, curl-free, and div-free); bullets ($\bullet$) and stars ($\star$) are the initial and final configurations, respectively, and the arrows are the initial momenta. The grid illustrates the diffeomorphism φ_{01}^v of the ambient space. The length of the bar at the bottom left of each graph shows the width $\sqrt{2}\sigma$ of each kernel.

6.2. Exponential map for landmarks

Figure 11 shows exponential maps [36] $\exp_q(t\dot{q}(0))$ in $\mathcal{L}^2(\mathbb{R}^2)$ for the above choices of the Gaussian kernel (scalar, curl-free, and divergence-free); with the same choice of parameters as in the landmark shooting examples. In all three cases, the *initial* landmark positions are fixed at $q = (q^1, q^2) =$
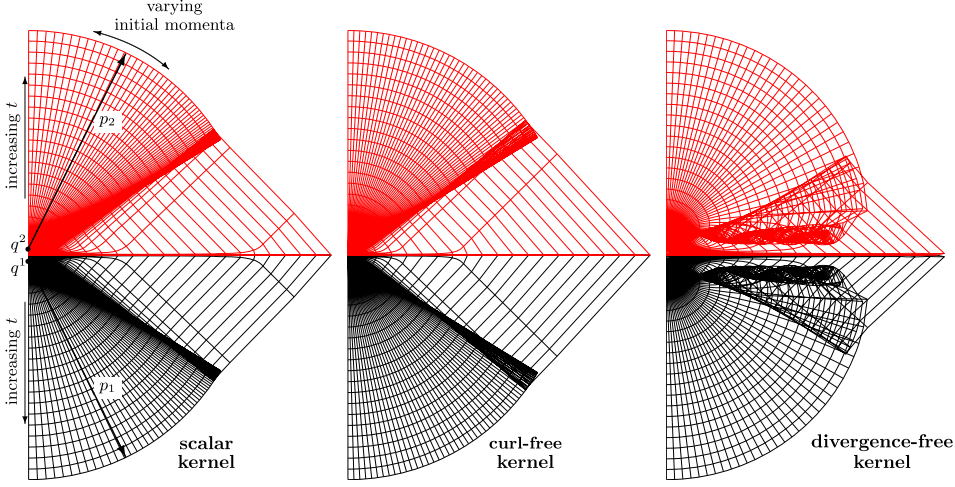
Figure 11: Exponential maps $\exp_q(t\dot{q}(0))$ in $\mathcal{L}^2(\mathbb{R}^2)$ for scalar, curl-free, and div-free Gaussian-type kernels; the initial landmark configuration $q = (q^1, q^2)$ is shown in the first graph. The black grids (bottom halves) and red grids (top halves) illustrate the geodesic trajectories initiating from q^1 and q^2 respectively, for different choices of initial momenta (p_1, p_2 are chosen symmetrically with respect to the horizontal axis), and sampled in $t \in [0, 1]$.

$((0, -0.125), (0, 0.125))$, whereas the *initial* momenta are of the type $p = (p_1, p_2) = (50(\cos\theta, \sin\theta), 50(\cos\theta, -\sin\theta))$ with $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ (the first of Hamilton's equations (71a) relates the momenta p and the landmark velocities), and the paramerer t is sampled in [0,1]. We observe that the exponential maps for the Gaussian scalar and curl-free kernels are qualitatively similar, in that they have one "folding": this illustrates the existence of *conjugate points*, i.e. landmark sets in $\mathcal{L}^2(\mathbb{R}^2)$ that are connected to the initial configuration $q = (q^1, q^2)$ by distinct geodesic paths. The situation becomes considerably more complex in the divergence-free kernel case, where we observe *multiple* foldings: the exponential map keeps adding extra folds as t increases, and the existence of several distinct trajectories connecting landmark sets to the initial configuration is apparent.

Figure 12 shows a detail of the same exponential map in the region where most foldings occur, and a single trajectory of landmarks in the divergence-free case (with initial horizontal momenta, $\theta = 0$), at different spatial scales; note that the latter trajectory exhibits an apparent "periodic" behavior. Our interpretation of this phenomenon, that only appears in the divergence-free case, is that for two landmarks to move in the same direction in an incompressible fluid, they have to "swim" in such fashion so that the fluid gets out
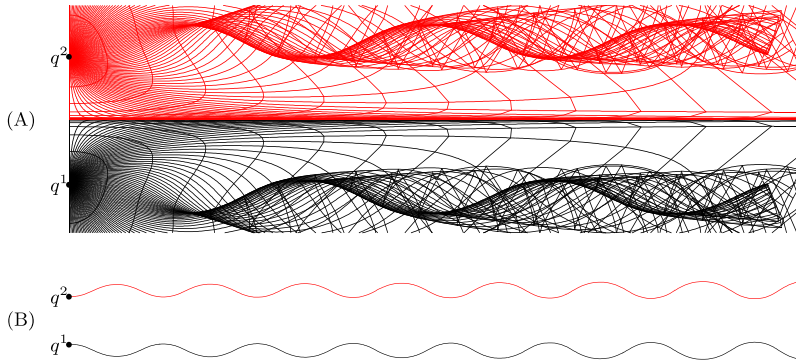
Figure 12: (A) Detail of the exponential map and (B) a landmark trajectory in the divergence-free case.

of the way as they proceed along their paths. This qualitative behavior and the related differential geometry of $\mathcal{L}^N(\mathbb{R}^d)$, equipped with the metric (67) for TRI kernels, should be investigated further. Regarding the computational complexity for both the shooting trajectories and the exponential maps, the running time for our code is just a few seconds on a Macbook Pro laptop computer equipped with a 2.66 GHz Intel Core i7 processor.

6.3. Application to real data

Finally, Figure 13 shows the high-resolution deformation of a template for *planum temporale* [55, 56, 57] to a target using the *currents* approach for surface matching [28, 66] and using three different deformation kernels. The first one is a *scalar* the Gaussian kernel $\mathbf{k}(x) = k(\|x\|)$, $x \in \mathbb{R}^3$, where k is given by (17); the second one and third one are, respectively, the Gaussian-type *curl-free* and *divergence-free* matrix-valued TRI kernels of Example 2 and Example 1, in three dimensions. The diameter of the *planum temporale* template is 44mm, while the length parameter of the Gaussian kernels is $\sigma = 7$mm. The deformation is sampled and shown at four different times: $t = 0$ (no deformation), $t = 1/3$, $t = 2/3$, and $t = 1$ (template matched to target). A cubic grid is also shown to illustrate, for each kernel and at each sampled time instant, the corresponding deformation of the ambient space.

The code was written in C++ and run on on a Macbook Pro laptop computer equipped with a 2.66 GHz Intel Core i7 processor: each of the three computations took approximately one and a half hours (it was about 25% faster in the scalar case). The template surface has 854 vertices and 1,532 triangles, while the target surface has 662 vertices and 1,160 triangles: there-
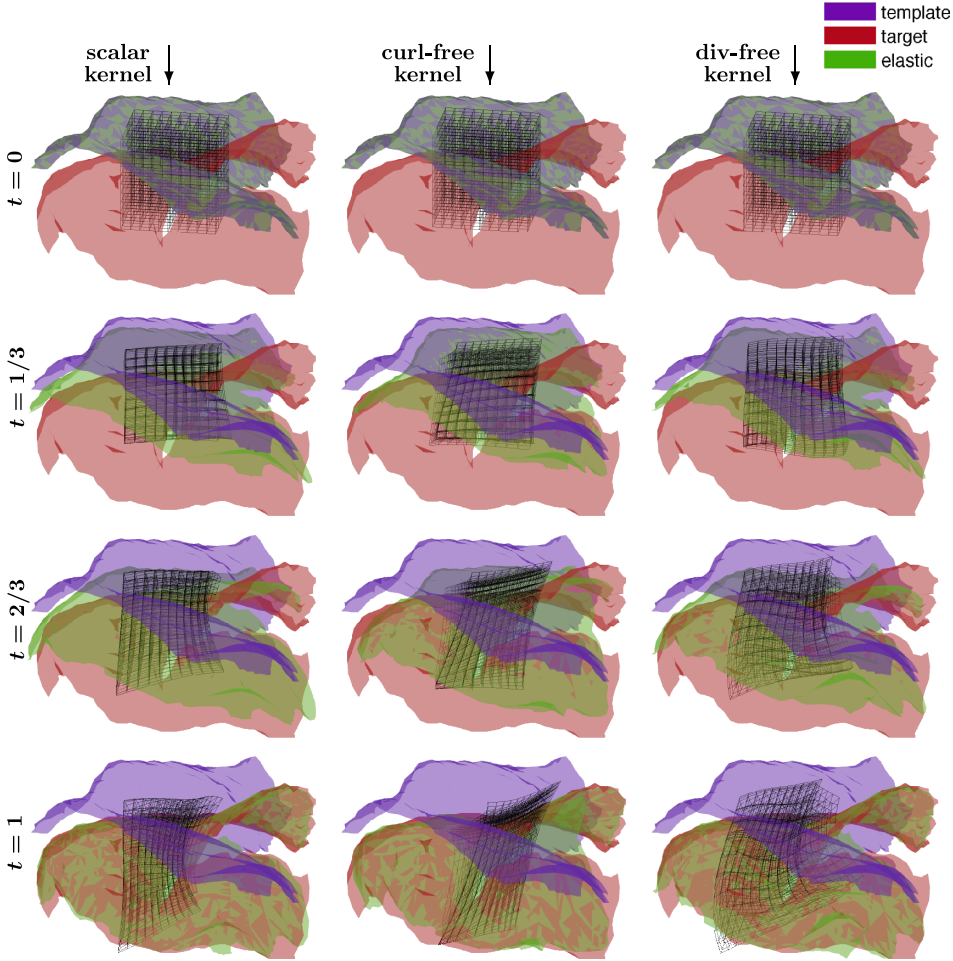
Figure 13: Deformation of a template (in purple) for *planum temporale* to a target (in red) using three different Gaussian kernels: scalar (first column), curl-free (second column), and divergence free (third column). The diameter of the *planum temporale* template is 44mm, while the length parameter of the Gaussian kernels is $\sigma = 7$mm. The deformation is sampled at four different times along the geodesic flow: $t = 0$, $t = 1/3$, $t = 2/3$, and $t = 1$, and each sample is shown in the first, second, third, and fourth rows, respectively. The deformed template is plotted in green in each figure. A cubic grid is also shown to illustrate, for each type of kernel and at each time instant, the corresponding deformation of the ambient space.

fore we have 854 moving points, which is also the number of momentum vectors. Thus, at each step of the energy minimization process we integrate a system of $854 \times 3 = 2{,}562$ differential equations in order to evaluate the functional and compute its gradient. Moreover, for the scalar kernel we coded the matching algorithm in the CUDA language (Compute Unified Device Architecture) for Nvidia graphics processing units, which allows developers to access the parallel throughput architecture of such devices: the resulting computational time for the matching in the scalar case was just around *one minute* (as opposed to one and a half hours!). While we have not yet followed the same procedure for the matrix-valued (curl-free and divergence-free) kernels, we are confident that the computing time in these two instances would be comparable to the scalar case.

Curl-free vs. divergence-free. Finally, we should note that we chose the *planum temporale* just to demonstrate that data of this nature can be handled in a computationally efficient manner with TRI kernels of the above type. However, for specific applications certain kernels are more appropriate than others. For example, as we recalled previously, by Liouville's Theorem [3] divergence-free vector fields produce volume-preserving diffeomorphisms, therefore divergence-free kernels are more suitable for applications where the deformations involved are *a priori* known to preserve volume; this is a property e.g. of the contractions of the muscular tissues of the heart [9, 40]. On the other hand, curl-free kernels may be used for the longitudinal studies of deformations that are "purely" caused by the growth or loss of matter; in neuroscience the former instance typically occurs during the development of the brain, while the latter is either caused by aging or by neurodegenerative diseases such as Alzheimer's. Regarding the latter case, we should mention a recent paper by Lorenzi *et al.* [39], where the authors use the irrotational (curl-free) component of the deformation and the corresponding potential to identify areas of maximal volume change in patients affected by the disease, and quantify longitudinal atrophy in such regions.

7. Conclusions

In this paper we have provided a thorough and systematic study of matrix-valued definite positive kernels that induce translation- and rotation-invariant metrics in the corresponding Reproducing Kernel Hilbert Spaces. We have conducted a novel analysis of such kernels in the Fourier domain: here the classification of spaces of curl-free and divergence free vector fields is very natural, in that each corresponds to one of the two coefficients of the Fourier transform of the kernel being equal to zero. This allows one to perform the Hodge decomposition in RKHS via the computation of Hankel transforms (which may be done analytically or numerically). We have also devised a

method for building curl-free and divergence-free kernels via an appropriate differentiation of scalar kernels; we have also proven that this constructive procedure is general, in that any TRI kernel (more precisely, the terms of its Hodge decomposition) may be obtained this way. Such a method may thus be used to design a large class of positive kernels. We have given a summary of the application of the LDDMM approach to the interpolation of vector fields and the matching of landmark points, together with the differential geometric interpretation and the derivation of the corresponding Hamiltonian. Finally, we have given numerical examples of the dynamics of landmark points and the corresponding flow of diffeomorphisms for different choices of TRI kernels, as well as the illustration of an application of such kernels to the surface matching of *planum temporale*.

Future work should focus on the design of kernels that are appropriate for different applications, especially in computational anatomy and related fields. As mentioned in the previous section, divergence-free kernels are suitable for volume-preserving transformations, therefore such kernels should be suitable for applications where shapes are known to deform in such manner (for example, the muscular tissues of the human heart have this property). On the other hand, curl-free kernels are more appropriate for the longitudinal studies of deformations that are mostly due to the growth or loss of matter. More in general, given a parameter-dependent family of kernels (where the parameters may determine e.g. the size of the kernel, or the weights in a combination of curl-free and divergence-free terms), for a specific application one has the interest of determining, or *learning*, the parameters that are optimal for a dataset of deformed anatomical shapes, i.e. those that best describe the deformations. Developing a comprehensive collection of TRI kernels with different properties, such as differing degrees of regularity at the origin and varying rates of convergence to zero at infinity, would also be useful.

The shape of geodesics should be studied in relation to the choice of the TRI kernel; in particular, note that all the kernels that we have considered in the examples have a length constant (e.g. the parameter $\sigma > 0$ for Gaussian- and Bessel-type kernels), so that the induced dynamics in the space of landmarks (and, for that matter, in the group of diffeomorphisms itself) is not scale invariant; furthermore, the regularity of the kernels at zero and the heaviness of their tails are significant in determining both the qualitative dynamics and the complexity of numerical implementations. Last, but not least, the visualization of the exponential map in the previous section (especially in the case of divergence-free TRI kernels) reveals the presence of several conjugate points, i.e. landmark configurations that are connected by distinct geodesics; the study of the dynamics of $\mathcal{L}^N(\mathbb{R}^d)$ induced by the action of diffeomorphism groups should be furthered, and the investigation of the differential geometry of the space of landmarks that was started in [42]

in the case of scalar kernels, which is useful for the goal of performing statistics on such manifold (e.g. for determining the uniqueness of the Karcher mean [54]), should be extended to the case of TRI kernels.

Acknowledgements

Appendix A. Relevant properties of Hankel transforms and Bessel functions

• **Fourier-Bessel integral** [20, Chapter 2] [52, Chapter 1]**:** this is also known as Hankel's integral formula, or the inversion formula for Hankel transforms:

$$(73) \qquad \int_0^\infty \Big\{ \int_0^\infty f(r)\, J_\mu(2\pi\varrho r)\, r\, dr \Big\} J_\mu(2\pi t\varrho)\, \varrho\, d\varrho = \frac{f(t)}{(2\pi)^2}, \quad t > 0.$$

• **Recurrence relations for Bessel functions of the 1st kind** [1, §9.1.7]**:**

$$(74) \qquad J_{\nu-1}(z) + J_{\nu+1}(z) = \frac{2\nu}{z} J_\nu(z)$$

$$(75) \qquad J_{\nu-1}(z) - \frac{\nu}{z} J_\nu(z) = J'_\nu(z)$$

$$(76) \qquad -J_{\nu+1}(z) + \frac{\nu}{z} J_\nu(z) = J'_\nu(z)$$

• **A result on modified Bessel functions of the 2nd kind:**

Lemma A.1. *For fixed $\sigma > 0$ and arbitrary ν the function $f_\nu(r) := (\frac{r}{\sigma})^\nu K_\nu(\frac{r}{\sigma})$, $r > 0$, is such that:*

$$(77) \qquad f'_\nu(r) = -\frac{1}{\sigma}\Big(\frac{r}{\sigma}\Big)^\nu K_{\nu-1}\Big(\frac{r}{\sigma}\Big), \qquad f''_\nu(r) = \frac{1}{\sigma^2} f_\nu(r) + \frac{2\nu-1}{r} f'_\nu(r)\,.$$

Proof. We have $f'_\nu(r) = \frac{\nu}{\sigma}(\frac{r}{\sigma})^{\nu-1} K_\nu(\frac{r}{\sigma}) + \frac{1}{\sigma}(\frac{r}{\sigma})^\nu K'_\nu(\frac{r}{\sigma})$; using the property $K'_\nu(z) = -K_{\nu-1}(z) - \frac{\nu}{z} K_\nu(z)$ of modified Bessel functions [1, §9.6.26] yields

the first of (77). Differentiating the latter gives

$$f_\nu''(r) = \frac{d}{dr}\Big\{-\frac{1}{\sigma}\Big(\frac{r}{\sigma}\Big)\cdot\Big(\frac{r}{\sigma}\Big)^{\nu-1}K_{\nu-1}\Big(\frac{r}{\sigma}\Big)\Big\}$$
$$\overset{(*)}{=} -\frac{1}{\sigma^2}\Big(\frac{r}{\sigma}\Big)^{\nu-1}K_{\nu-1}\Big(\frac{r}{\sigma}\Big) + \frac{1}{\sigma^2}\Big(\frac{r}{\sigma}\Big)^{\nu}K_{\nu-2}\Big(\frac{r}{\sigma}\Big),$$

where in $(*)$ we have used the first of equations (77) with $\nu-1$ instead of ν. At this point we use the property $K_{\nu-2}(z) = K_\nu(z) - 2\frac{\nu-1}{z}K_{\nu-1}(z)$ of modified Bessel functions [1, §9.6.26], and elementary manipulations lead to the second of equations (77). □

• **Asymptotic expansion for large arguments of the modified Bessel function** K_ν **[1, §9.7.2]:**
(78)
$$K_\nu(z) \sim \sqrt{\frac{\pi}{2z}}e^{-z}\Big[1 + \frac{\lambda-1}{8z} + \frac{(\lambda-1)(\lambda-9)}{2!(8z)^2} + \frac{(\lambda-1)(\lambda-9)(\lambda-25)}{3!(8z)^3} + \cdots\Big]$$

valid for fixed ν and $|\arg z| < \frac{3}{2}\pi$; in the above expression, $\lambda = 4\nu^2$.

Appendix B. Bochner's Theorem and consequences

We shall reformulate the generalized version of Bochner's theorem as follows:

Bochner's Theorem. *Consider a matrix-valued function* $\mathbf{k} \in L^1(\mathbb{R}^m, \mathbb{R}^{d\times d})$ *whose Fourier transform* $\widehat{\mathbf{k}}$ *is also* L^1, *and let* $K(x,y) := \mathbf{k}(x-y)$, $x, y \in \mathbb{R}^m$. *The following facts are equivalent:*

   i. *K is a positive definite function;*

  ii. *for all $f \in C_c^\infty(\mathbb{R}^m, \mathbb{C}^d)$ it is the case that $\int_{\mathbb{R}^m} f^*(\xi)\,\widehat{\mathbf{k}}(\xi)\,f(\xi)\,d\xi \geq 0$;*

 iii. *for all $\xi \in \mathbb{R}^m$, it is the case that $\widehat{\mathbf{k}}(\xi)$ is a self-adjoint positive definite matrix.*

Proof. We have that $\mathbf{k}(x) = \int_{\mathbb{R}^m} \widehat{\mathbf{k}}(\xi)e^{2\pi i\langle x,\xi\rangle}d\xi$, $x \in \mathbb{R}^m$ so we first note that for fixed $N \in \mathbb{N}$, $\boldsymbol{x} = (x_1, \ldots, x_N) \in (\mathbb{R}^m)^N$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N) \in (\mathbb{R}^d)^N$ we may write

$$(79) \qquad \sum_{a,b=1}^N \alpha_a \cdot \mathbf{k}(x_a - x_b)\alpha_b = \int_{\mathbb{R}^m} f_{\boldsymbol{x},\boldsymbol{\alpha}}^*(\xi)\,\widehat{\mathbf{k}}(\xi)f_{\boldsymbol{x},\boldsymbol{\alpha}}(\xi)\,d\xi,$$

where $f_{\boldsymbol{x},\boldsymbol{\alpha}}(\xi) := \sum_{a=1}^N e^{-2\pi i x_a\cdot\xi}\alpha_a$. We now proceed with the proof.

(i. $\Rightarrow$ ii.) Fix $f \in C_c^\infty(\mathbb{R}^m, \mathbb{C}^d)$ and $\varepsilon > 0$. By the continuity of integration, there exists $A > 0$ such that $C_A := [-A, A]^m$ verifies $\int_{\mathbb{R}^m \setminus C_A} \|\widehat{\mathbf{k}}(\xi)\| d\xi < \varepsilon$. We can choose A large enough such that $\operatorname{supp} f \subseteq C_A$. Consider now the *periodic extension* of f in $\mathbb{R}^m$ and its Fourier series expansion:

$$f_{\mathrm{per}}(\xi) := \sum_{p \in \mathbb{Z}^m} f(\xi - Ap) = \sum_{p \in \mathbb{Z}^m} c_p e^{-2\pi i \frac{p}{2A} \cdot \xi}, \ \xi \in \mathbb{R}^m,$$

$$\text{with } c_p := \int_{\mathbb{R}^m} f(\xi) \, e^{2\pi i \frac{p}{2A} \cdot \xi} \, d\xi, \ p \in \mathbb{Z}^m.$$

The above Fourier series converges uniformly, whence there exist $M \in \mathbb{N}$, points $p_1, \ldots, p_M \in \mathbb{Z}^m$ such that $| f(\xi) - \sum_{a=1}^N \exp(-2\pi i \frac{p_a}{2A} \cdot \xi) \alpha_a | < \varepsilon$ for all $\xi \in C_A$, where we have set $\alpha_a := c_{p_a}$, for $a = 1, \ldots, M$. In other words, there exist $\boldsymbol{x} = (x_1, \ldots, x_M)$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M)$ such that

$$\sup_{\xi \in C_A} \left\{ \left| f(\xi) - \sum_{a=1}^N \exp\left(-2\pi i x_a \cdot \xi\right) \alpha_a \right| \right\} < \varepsilon.$$

Since $\|f - f_{\boldsymbol{x},\boldsymbol{\alpha}}\|_{C_A,\infty} < \varepsilon$ we have $\|f_{\mathrm{per}} - f_{\boldsymbol{x},\boldsymbol{\alpha}}\|_\infty < \varepsilon$, whence $\|f_{\boldsymbol{x},\boldsymbol{\alpha}}\|_\infty < \|f_{\mathrm{per}}\|_\infty + \varepsilon = \|f\|_\infty + \varepsilon$. Now,

$$\left| \int_{\mathbb{R}^m} f_{\boldsymbol{x},\boldsymbol{\alpha}}^*(\xi) \widehat{\mathbf{k}}(\xi) f_{\boldsymbol{x},\boldsymbol{\alpha}}(\xi) \, d\xi - \int_{\mathbb{R}^m} f^*(\xi) \widehat{\mathbf{k}}(\xi) f(\xi) \, d\xi \right|$$

$$\leq \underbrace{\left| \int_{\mathbb{R}^m \setminus C_A} f_{\boldsymbol{x},\boldsymbol{\alpha}}^*(\xi) \widehat{\mathbf{k}}(\xi) f_{\boldsymbol{x},\boldsymbol{\alpha}}(\xi) \, d\xi \right|}_{I_1}$$

$$+ \underbrace{\left| \int_{C_A} f_{\boldsymbol{x},\boldsymbol{\alpha}}^*(\xi) \widehat{\mathbf{k}}(\xi) f_{\boldsymbol{x},\boldsymbol{\alpha}}(\xi) \, d\xi - \int_{C_A} f^*(\xi) \widehat{\mathbf{k}}(\xi) f(\xi) \, d\xi \right|}_{I_2},$$

with

$$I_1 \leq \int_{\mathbb{R}^m \setminus C_A} \|\widehat{\mathbf{k}}(\xi)\| \|f_{\boldsymbol{x},\boldsymbol{\alpha}}(\xi)\|^2 d\xi < \left(\|f\|_\infty + \varepsilon\right)^2 \int_{\mathbb{R}^m \setminus C_A} \|\widehat{\mathbf{k}}(\xi)\| d\xi < \left(\|f\|_\infty + \varepsilon\right)^2 \varepsilon$$

and

$$I_2 \leq \left| \int_{C_A} (f_{\boldsymbol{x},\boldsymbol{\alpha}}(\xi) - f(\xi))^* \widehat{\mathbf{k}}(\xi) f_{\boldsymbol{x},\boldsymbol{\alpha}}(\xi) \, d\xi \right| + \left| \int_{C_A} f^*(\xi) \widehat{\mathbf{k}}(\xi) (f_{\boldsymbol{x},\boldsymbol{\alpha}}(\xi) - f(\xi)) \, d\xi \right|$$

$$\leq \int_{C_A} \|f_{\boldsymbol{x},\boldsymbol{\alpha}}(\xi) - f(\xi)\| \|\widehat{\mathbf{k}}(\xi)\| \|f_{\boldsymbol{x},\boldsymbol{\alpha}}(\xi)\| \, d\xi + \int_{C_A} \|f(\xi)\| \|\widehat{\mathbf{k}}(\xi)\| \|f_{\boldsymbol{x},\boldsymbol{\alpha}}(\xi) - f(\xi)\| \, d\xi$$

$$< \varepsilon \|\widehat{\mathbf{k}}\|_{L^1} \left(\|f\|_\infty + \varepsilon\right) + \|f\|_\infty \|\widehat{\mathbf{k}}\|_{L^1} \varepsilon,$$

where $\|\widehat{\mathbf{k}}\|_{L_1} = \int_{\mathbb{R}^m} \|\widehat{\mathbf{k}}(\xi)\| d\xi$. Therefore it must be the case that

$$I_1 + I_2 \leq (\|f\|_\infty + \varepsilon)^2 \varepsilon + \|\widehat{\mathbf{k}}\|_{L^1}(\|f\|_\infty + \varepsilon)\varepsilon + \|\widehat{\mathbf{k}}\|_{L^1}\|f\|_\infty \varepsilon,$$

which concludes this part of the proof.

(ii. ⇒ iii.) The matrix $\widehat{\mathbf{k}}(\xi)$ is self-adjoint for all ξ by the arbitrariness of $f \in C_c^\infty(\mathbb{R}^m, \mathbb{C}^d)$. Assume that for a $\xi_0 \in \mathbb{R}^m$ the matrix $\widehat{\mathbf{k}}(\xi_0)$ has an eigenvalue $\lambda_0 < 0$, with eigenvector $\gamma_0 \in \mathbb{C}^m$. By the continuity of $\widehat{\mathbf{k}}$ there exists a neighborhood Ω_0 of ξ_0 such that $\gamma_0^* \widehat{\mathbf{k}}(\xi) \gamma_0 < 0$ in Ω_0; let $f(\xi) := \psi(\xi)\gamma_0$, $\xi \in \mathbb{R}^m$, where $\psi \in C_c^\infty(\mathbb{R}^m, \mathbb{R})$, with supp $\psi \subseteq \Omega_0$. Then we must have $\int_{\mathbb{R}^m} f^* \widehat{\mathbf{k}} f \, d\xi = \int_{\Omega_0} \psi^2 \gamma_0^* \widehat{\mathbf{k}} \gamma_0 \, d\xi < 0$, which is a contradiction.

(iii. ⇒ i.) If $\widehat{\mathbf{k}}(\xi)$ is self-adjoint with nonnegative eigenvalues for all $\xi \in \mathbb{R}^m$ then the right-hand side of (79) is nonnegative for all $\boldsymbol{x}$ and $\boldsymbol{\alpha}$; whence the function $K(x, y)$ is positive definite. □

We now proceed with the proof of Proposition 3.13, that states that translation- and rotation-invariant kernels $\mathbf{k}$ of the type (10) are *strictly* positive definite as long as at least one of the two coefficients $h^\|, h^\perp : \mathbb{R}^+ \to \mathbb{R}$ of the Fourier transform of $\mathbf{k}$ is positive somewhere.

Proof of Proposition 3.13. In the TRI kernel setting $m = d$. Note that for any $\boldsymbol{x} = (x_1, \ldots, x_N) \in (\mathbb{R}^d)^N$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N) \in (\mathbb{R}^d)^N$ (where at least one of the α_a's is nonzero), and for any $\xi \in \mathbb{R}^d$, the vector $f_{\boldsymbol{x},\boldsymbol{\alpha}}(\xi) \in \mathbb{C}^m$ introduced in the previous proof is *neither* parallel *nor* perpendicular to ξ.

First assume that there exists $r_0 \geq 0$ such that $h^\|(r_0) > 0$ or $h^\perp(r_0) > 0$. Choose $\xi_0 \in \mathbb{R}^d$ such that $\|\xi_0\| = r_0$, and arbitrary $\boldsymbol{x} = (x_1, \ldots, x_N)$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)$. We shall write:

$$f_{\boldsymbol{x},\boldsymbol{\alpha}}(\xi_0) = u_0^\| + u_0^\perp, \text{ with } u_0^\| := \mathrm{Pr}_{\xi_0}^\| f_{\boldsymbol{x},\boldsymbol{\alpha}}(\xi_0) \text{ and } u_0^\perp := \mathrm{Pr}_{\xi_0}^\perp f_{\boldsymbol{x},\boldsymbol{\alpha}}(\xi_0);$$

by the remark above, we have $u_0^\| \neq 0$, $u_0^\perp \neq 0$. It is immediate to verify that

$$f_{\boldsymbol{x},\boldsymbol{\alpha}}^*(\xi_0)\mathrm{Pr}_{\xi_0}^\| f_{\boldsymbol{x},\boldsymbol{\alpha}}(\xi_0) = \|u_0^\|\|^2 > 0 \text{ and } f_{\boldsymbol{x},\boldsymbol{\alpha}}^*(\xi_0)\mathrm{Pr}_{\xi_0}^\perp f_{\boldsymbol{x},\boldsymbol{\alpha}}(\xi_0) = \|u_0^\perp\|^2 > 0,$$

while $f_{\boldsymbol{x},\boldsymbol{\alpha}}^*(\xi)\mathrm{Pr}_\xi^\| f_{\boldsymbol{x},\boldsymbol{\alpha}}(\xi) \geq 0$ and $f_{\boldsymbol{x},\boldsymbol{\alpha}}^*(\xi)\mathrm{Pr}_\xi^\perp f_{\boldsymbol{x},\boldsymbol{\alpha}}(\xi) \geq 0$ for generic $\xi \in \mathbb{R}^d$. In conclusion

$$\int_{\mathbb{R}^d} f_{\boldsymbol{x},\boldsymbol{\alpha}}^*(\xi) \widehat{\mathbf{k}}(\xi) f_{\boldsymbol{x},\boldsymbol{\alpha}}(\xi) \, d\xi$$

$$= \int_{\mathbb{R}^d} \left\{ h^\|(\|\xi\|) f_{\boldsymbol{x},\boldsymbol{\alpha}}^*(\xi)\mathrm{Pr}_\xi^\| f_{\boldsymbol{x},\boldsymbol{\alpha}}(\xi) + h^\perp(\|\xi\|) f_{\boldsymbol{x},\boldsymbol{\alpha}}^*(\xi)\mathrm{Pr}_\xi^\perp f_{\boldsymbol{x},\boldsymbol{\alpha}}(\xi) \right\} d\xi > 0,$$

and by (79) the kernel is *strictly* definite positive. Vice versa, by (79) if $\mathbf{k}$ is strictly definite positive then the above strict inequality holds for all $\boldsymbol{x} =$

$(x_1, \ldots, x_N) \in (\mathbb{R}^d)^N$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N) \in (\mathbb{R}^d)^N$ where at least one of the α_a's is nonzero. Therefore either $h^{\|}$ or $h^{\perp}$ must be strictly positive somewhere in $(0, \infty)$. This completes the proof. $\qquad\square$

Appendix C. Proof of Proposition 3.9

Note that since f is symmetric so must be its Fourier transform, i.e. $\widehat{f}(\xi) = \widehat{f}(-\xi)$. We now prove that it is actually *radially* symmetric. Using polar coordinates in $\mathbb{R}^d$ [26] we can write, for any $\xi \in \mathbb{R}^d$:

$$(80) \quad \widehat{f}(\xi) = \widehat{f}(-\xi) := \int_{\mathbb{R}^d} f(x) e^{2\pi i x \cdot \xi} dx = \int_0^\infty r^{d-1} g(r) \int_{\mathbb{S}^{d-1}} e^{2\pi i r s \cdot \xi} \, ds \, dr,$$

where we use the symbol $\mathbb{S}^n$ for the unit n-sphere. Fix ξ, and let η be such that $\|\eta\| = \|\xi\|$; in other words $\eta = R\xi$ for some $R \in \mathrm{O}(d, \mathbb{R})$. We have that $\widehat{f}(\eta) = \int_0^\infty r^{d-1} g(r) \int_{\mathbb{S}^{d-1}} e^{2\pi i r s \cdot \eta} ds \, dr$, but

$$\int_{\mathbb{S}^{d-1}} e^{2\pi i r\, s \cdot \eta} \, ds = \int_{\mathbb{S}^{d-1}} e^{2\pi i r\, s \cdot R\xi} \, ds = \int_{\mathbb{S}^{d-1}} e^{2\pi i r (R^T s) \cdot \xi} \, ds = \int_{\mathbb{S}^{d-1}} e^{2\pi i r\, s' \cdot \xi} \, ds',$$

since $|\det R| = 1$. Whence $\widehat{f}(\eta) = \widehat{f}(\xi)$, i.e. $\widehat{f}(\xi) = G(\|\xi\|)$, for some function $G : \mathbb{R}^+ \to \mathbb{R}$. Without loss of generality in (80) we can choose ξ to be parallel to the standard basis vector e_1. Letting $\varrho = \|\xi\|$ and $\cos\varphi = s \cdot e_1$ we can write $\xi \cdot s = \varrho \cos\varphi$ (note that $\|s\| = 1$ on $\mathbb{S}^{d-1}$). This yields:

$$
\begin{aligned}
\int_{\mathbb{S}^{d-1}} e^{2\pi i r\, s \cdot \xi} \, ds &= \int_{\mathbb{S}^{d-1}} e^{2\pi i \varrho r \cos\varphi} \, ds \\
&= \sigma(\mathbb{S}^{d-2}) \int_0^\pi e^{2\pi i \varrho r \cos\varphi} (\sin\varphi)^{d-2} \, d\varphi \\
&\overset{(*)}{=} \sigma(\mathbb{S}^{d-2}) \int_{-1}^1 e^{2\pi i \varrho r u} (1 - u^2)^{\frac{d-3}{2}} \, du \\
(81) \qquad\qquad &= 2\sigma(\mathbb{S}^{d-2}) \int_0^1 (1 - u^2)^{\frac{d-3}{2}} \cos(2\pi \varrho r u) \, du,
\end{aligned}
$$

where $\sigma(\mathbb{S}^n) = 2\pi^{\frac{n+1}{2}}/\Gamma(\frac{n+1}{2})$ is the surface area of the unit n-sphere; in $(*)$ we have performed the substitution $u = \cos\varphi$. We now use the integral representation of Bessel functions [1, §9.1.20]

$$J_\nu(z) = \frac{2(\frac{1}{2}z)^\nu}{\sqrt{\pi}\,\Gamma(\nu + \frac{1}{2})} \int_0^1 (1 - u^2)^{\nu - \frac{1}{2}} \cos(zu) \, du, \quad \text{valid for } \Re\nu > -\tfrac{1}{2},$$

to get

$$\int_0^1 (1-u^2)^{\frac{d-3}{2}} \cos(2\pi\varrho r u)\, du = \frac{\sqrt{\pi}\,\Gamma(\frac{d-1}{2})}{2(\pi\varrho r)^{\frac{d}{2}-1}}\, J_{\frac{d}{2}-1}(2\pi\varrho r);$$

inserting the above into (81) and the latter into (80) completes the proof.

References

[1] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions.* Dover Publications, New York, 1964.

[2] V. I. Arnold. Sur la géometrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits. *Annales de l'Institut Fourier*, 16(1):319–361, 1966. MR0202082

[3] V. I. Arnold. *Mathematical Methods of Classical Mechanics*, volume 60 of *Graduate Texts in Mathematics.* Springer, New York, second edition, 1989. MR0997295

[4] N. Aronszajn. La théorie des noyaux reproduisants et ses applications, I. *Mathematical Proceedings of the Cambridge Philosophical Society*, 39(03):133–153, 1943. MR0008639

[5] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68(3):337–404, May 1950. MR0051437

[6] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. Journal on Computer Vision*, 61(2):139–157, 2005.

[7] S. Bergman. Functions satisfying certain partial differential equations of elliptic type and their representation. *Duke Mathematical Journal*, 14:349–366, 1947. MR0022262

[8] S. Bergman. *The Kernel Function and Conformal Mapping.* American Mathematical Society, Providence, Rhode Island, second edition, 1950. MR0038439

[9] A. Bistoquet and J. Oshinski. Myocardial deformation recovery from cine MRI using a nearly incompressible biventricular model. *Medical Image Analysis*, 12(1):69–85, 2008.

[10] S. Bochner. Monotone funktionen, stieltjessche integrale und harmonische analyse. *Mathematische Annalen*, 108(1):378–410, 1933. MR1512856

[11] L. Bonaventura, A. Iske, and E. Miglio. Kernel-based vector field reconstruction in computational fluid dynamic models. *International Journal for Numerical Methods in Fluids*, 66(6):714–729, 2011. MR2839797

[12] R. N. Bracewell. *The Fourier Transform and its Applications.* McGraw-Hill, New York, third edition, 1999. MR0924577

[13] P. Cachier and N. Ayache. Isotropic energies, filters and splines for vector field regularization. *Journal of Mathematical Imaging and Vision*, 20(3):251–265, 2004. MR2060147

[14] Y. Cao, M. I. Miller, S. Mori, R. L. Winslow, and L. Younes. Diffeomorphic matching of diffusion tensor images. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR '06)*, New York, June 2006.

[15] Y. Cao, M. I. Miller, R. L. Winslow, and L. Younes. Large deformation diffeomorphic metric mapping of vector fields. *IEEE Transactions on Medical Imaging*, 24(9):1216–1230, 2005.

[16] C. Carmeli, E. De Vito, A. Toigo, and V. Umanità. Vector-valued Reproducing Kernel Hilbert Spaces and Universality. *Analysis and Applications*, 08(01):19, 2010.

[17] A. J. Chorin and J. E. Marsden. *A Mathematical Introduction to Fluid Mechanics.* Springer-Verlag, New York, 1993. MR1218879

[18] G. Christensen, R. Rabbitt, and M. Miller. Deformable templates using large deformation kinematics. *Image Processing, IEEE Transactions on*, 5(10):1435–1447, 1996.

[19] E. De Vito, V. Umanità, and S. Villa. An extension of Mercer theorem to matrix-valued measurable kernels. *Applied and Computational Harmonic Analysis*, 34(3):339–351, May 2013. MR3027907

[20] V. A. Ditkin and A. P. Prudnikov. *Integral Transforms and Operational Calculus*, volume 78 of *International Series of Monographs in Pure and Applied Mathematics.* Pergamon Press, Oxford, 1965. MR0196422

[21] F. Dodu and C. Rabut. Irrotational or divergence-free interpolation. *Numerische Mathematik*, 98(3):477–498, 2004. MR2088924

[22] P. Dupuis, U. Grenander, and M. I. Miller. Variational problems on flows of diffeomorphisms for image matching. *Quarterly of Applied Mathematics*, 56(3):587–600, Sept. 1998. MR1632326

[23] A. Erdélyi. *Tables of Integral Transforms*, volume 2. McGraw Hill, Jan. 1954.

[24] L. C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, Rhode Island, 1998. MR1625845

[25] P. T. Fletcher, S. Venkatasubramanian, and S. Joshi. The geometric median on Riemannian manifolds with application to robust atlas estimation. *NeuroImage*, 45:S143–152, Mar. 2009.

[26] G. B. Folland. *Real Analysis: Modern Techniques and Their Applications*. John Wiley & Sons, second edition, 1999. MR1681462

[27] J. A. Glaunès. *Transport par difféomorphismes de points, de mesures et de courants pour la comparaison de formes et l'anatomie numérique*. PhD thesis, Université Paris 13, France, Sept. 2005.

[28] J. A. Glaunès and S. Joshi. Template estimation from unlabeled point set data and surfaces for computational anatomy. In X. Pennec and S. Joshi, editors, *Proc. of the International Workshop on the Mathematical Foundations of Computational Anatomy (MFCA)*, pages 29–39, October 2006.

[29] J. A. Glaunès, A. Qiu, M. I. Miller, and L. Younes. Large deformation diffeomorphic metric curve mapping. *International Journal of Computer Vision*, 80(3):317–336, Dec. 2008.

[30] J. A. Glaunès, M. Vaillant, and M. I. Miller. Landmark matching via large deformation diffeomorphisms on the sphere. *Journal of Mathematical Imaging and Vision*, 20:170–200, 2004. MR2049789

[31] U. Grenander and M. I. Miller. Computational anatomy: an emerging discipline. *Quarterly of Applied Mathematics*, 56(3):617–694, Sept. 1998. MR1668732

[32] M. Hein and O. Bousquet. *Kernels, associated structures and generalizations*. Technical Report 127, Max Planck Institute for Biological Cybernetics, 2004.

[33] K. Itô, editor. *Encyclopedic Dictionary of Mathematics*. MIT Press, Cambridge, Massachusetts, USA, second edition, 1993. Prepared by the Mathematical Society of Japan.

[34] K. Jänich. *Vector Analysis*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 2001. MR1811820

[35] S. C. Joshi and M. I. Miller. Landmark matching via large deformation diffeomorphisms. *IEEE Transactions on Image Processing*, 9(8):1357–1370, Aug. 2000. MR1808275

[36] J. Jost. *Riemannian Geometry and Geometric Analysis*. Springer-Verlag, New York, third edition, 2002. MR1871261

[37] J. M. Lee. *Riemannian Manifolds: an Introduction to Curvature*, volume 176 of *Graduate Texts in Mathematics*. Springer, New York, 1997. MR1468735

[38] J. M. Lee. *Introduction to Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer, New York, 2003. MR1930091

[39] M. Lorenzi, N. Ayache, and X. Pennec. Regional flux analysis of longitudinal atrophy in alzheimer's disease. In N. Ayache, H. Delingette, P. Golland, and K. Mori, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 7510 of *Lecture Notes in Computer Science*, pages 739–746. Springer, Jan. 2012.

[40] T. Mansi, X. Pennec, M. Sermesant, H. Delingette, and N. Ayache. iLogDemons: a Demons-based registration algorithm for tracking incompressible elastic biological tissues. *International Journal of Computer Vision*, 92(1):92–111, Mar. 2011.

[41] C. A. Micchelli and M. Pontil. On learning vector-valued functions. *Neural Computation*, 17(1):177–204, Jan. 2005. MR2175914

[42] M. Micheli, P. W. Michor, and D. Mumford. Sectional curvature in terms of the cometric, with applications to the Riemannian manifolds of landmarks. *SIAM Journal on Imaging Sciences*, 5(1):394–433, 2012. MR2902666

[43] M. Micheli, P. W. Michor, and D. Mumford. Sobolev metrics on diffeomorphism groups and the derived geometry on spaces of submanifolds. *Izvestiya of the Russian Academy of Sciences, Math. Series*, 77(3):109–138, 2013. MR3098790

[44] P. W. Michor and D. Mumford. Riemannian geometries on spaces of plane curves. *Journal of the European Mathematical Society*, 8:1–48, 2006. MR2201275

[45] P. W. Michor and D. Mumford. An overview of the Riemannian metrics on spaces of curves using the Hamiltonian approach. *Applied and Computational Harmonic Analysis*, 23:74–113, 2007. MR2333829

[46] M. I. Miller, A. Trouvé, and L. Younes. On the metrics and Euler-Lagrange equations of computational anatomy. *Annual Review of Biomedical Engineering*, 4:375–405, 2002.

[47] M. I. Miller and L. Younes. Group actions, homeomorphisms, and matching: a general framework. *International Journal of Computer Vision*, 41(1/2):61–84, 2001.

[48] D. Mumford. The geometry and curvature of shape spaces. In U. Zannier, editor, *Colloquium De Giorgi 2009*, volume 3 of *Colloquia of the Scuola Normale Superiore di Pisa*, pages 43–53. Springer, Jan. 2012.

[49] D. Mumford and A. Desolneux. *Pattern Theory: The Stochastic Analysis of Real-World Signals.* CRC Press, 2010. MR2723182

[50] D. Mumford and P. W. Michor. On Euler's equation and 'EPDiff'. *Journal of Geometric Mechanics*, 5(3):319–344, September 2013. MR3110104

[51] K. P. Murphy. *Machine Learning: A Probabilistic Perspective.* MIT Press, Cambridge, Mass., 2012.

[52] F. Oberhettinger. *Tables of Bessel Transforms.* Springer-Verlag, New York, 1972. MR0352888

[53] A. Papoulis. *Systems and Transforms with Applications in Optics.* McGraw-Hill, New York, 1968.

[54] X. Pennec. *Probabilities and Statistics on Riemannian Manifolds: a Geometric Approach*. Research Report 5093, INRIA, January 2004.

[55] A. Qiu, M. Vaillant, P. Barta, J. T. Ratnanather, and M. I. Miller. Region-of-interest-based analysis with application of cortical thickness variation of left planum temporale in schizophrenia and psychotic bipolar disorder. *Human Brain Mapping*, 29(8):973–985, Aug. 2008.

[56] J. T. Ratnanather, P. E. Barta, N. A. Honeycutt, N. Lee, H. M. Morris, A. C. Dziorny, M. K. Hurdal, G. D. Pearlson, and M. I. Miller. Dynamic programming generation of boundaries of local coordinatized submanifolds in the neocortex: application to the planum temporale. *NeuroImage*, 20:359–377, Sept. 2003.

[57] J. T. Ratnanather, C. B. Poynton, D. V. Pisano, B. Crocker, E. Postell, S. Cebron, E. Ceyhan, N. A. Honeycutt, P. B. Mahon, and P. E. Barta. Morphometry of superior temporal gyrus and planum temporale in schizophrenia and psychotic bipolar disorder. *Schizophrenia Research*, 150:476–483, Nov. 2013.

[58] W. Rudin. *Principles of Mathematical Analysis.* McGraw-Hill, 3rd edition, Jan. 1976. MR0385023

[59] I. J. Schoenberg. Metric spaces and completely monotone functions. *The Annals of Mathematics*, 39(4):811–841, Oct. 1938. MR1503439

[60] I. J. Schoenberg. Metric spaces and positive definite functions. *Transactions of the American Mathematical Society*, 44(3):522–536, Nov. 1938. MR1501980

[61] B. Scholkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA, 2001.

[62] L. Schwartz. Sous-espaces hilbertiens d'espaces vectoriels topologiques et noyaux associés (noyaux reproduisants). *Journal d'Analyse Mathématique*, 13(1):115–256, 1964. MR0179587

[63] I. N. Sneddon. *Fourier Transforms*. McGraw-Hill, New York, 1951. MR0041963

[64] E. M. Stein and G. Weiss. *Introduction to Fourier Analysis on Euclidean Spaces*. Princeton University Press, Nov. 1971. MR0304972

[65] A. Trouvé. *Infinite Dimensional Group Action and Pattern Recognition*. Technical report, DMI, École Nationale Supérieure, Paris, May 1995. MR1360567

[66] M. Vaillant and J. A. Glaunès. Surface matching via currents. In G. Christensen and M. Sonka, editors, *Information Processing in Medical Imaging (IPMI)*, volume 3565 of *Lecture Notes in Computer Science*, pages 1–5. Springer, 2005.

[67] G. Wahba. *Spline Models for Observational Data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia, PA, 1990. MR1045442

[68] L. Younes. Jacobi fields in groups of diffeomorphisms and applications. *Quarterly of Applied Mathematics*, 65(1):113–134, 2007. MR2313151

[69] L. Younes. *Shapes and Diffeomorphisms*, volume 171 of *Applied Mathematical Sciences*. Springer, 2010. MR2656312

[70] L. Younes, P. W. Michor, J. Shah, and D. Mumford. A metric on shape space with explicit geodesics. *Rendiconti Lincei – Matematica e Applicazioni*, 9:25–57, 2008. MR2383560

[71] S. Zhang, L. Younes, J. Zweck, and J. T. Ratnanather. Diffeomorphic surface flows: A novel method of surface evolution. *SIAM Journal on Applied Mathematics*, 68(3):806–824, Jan. 2008. MR2375296

Mario Micheli
Department of Mathematics
University of Washington
Box 354350
Seattle, WA 98195
USA
E-mail address: micheli@uw.edu

Joan A. Glaunès
Laboratoire MAP5
Université Paris Descartes and CNRS
Sorbonne Paris Cité
45 rue des Saints-Pères
75006 Paris
France
E-mail address: alexis.glaunes@mi.parisdescartes.fr