UNIVERSITÀ DEGLI STUDI DI PADOVA

SCUOLA GALILEIANA DI STUDI SUPERIORI

Università degli Studi di Padova

Scuola Galileiana di Studi Superiori

Classe di Scienze Naturali

Tesi di Diploma Galileiano

Sobolev reproducing kernels on compact Lie groups

Relatrice/Relatore
Prof. Ernesto De Vito

Diplomanda/o
Luca Wellmeier

Sobolev reproducing kernels on compact Lie groups

Luca Wellmeier

November 6, 2023

Abstract. In this paper we demonstrate how to compute the reproducing kernel of Sobolev spaces with real smoothness index on a compact Lie group. In Euclidean space these are called Matérn kernels. For compact manifolds one needs to know the eigenstructure of the Laplace-Beltrami operator, which is generally hard to find. However, on a compact Lie group the Laplacian can be identified with the quadratic Casimir operator, whose spectrum is explicitly described by the Peter-Weyl theorem. It requires knowing the irreducible unitary representations of the group, which are known in many low-dimensional cases. The approach also allows to easily incorporate different bi-invariant Riemannian metrics.

Contents

1 Introduction

Kernel methods are a set of powerful, non-parametric learning algorithms. They allow for non-linear, implicit transformation of the original data into a very high-dimensional feature space by means of a kernel function. The choice of a kernel function corresponds to setting up the notion of distance between data points, and is generally more intuitive than constructing a neural architecture. Furthermore, they can be applied to a wide variety of tasks as many fundamental linear algorithms can be kernelized. Examples include linear/ridge regression, support vector machines, principal component analysis, and Gaussian process regression.

While undisputably useful in practice, they also bear significant importance for understanding the theoretical foundations of machine learning. Recent studies have shown that kernel methods can serve as a proxy for understanding the myths of deep learning (see e.g. [4, 18, 15] as introductory reads). Fortunately, since they are linear methods at heart, theoretical studies about kernel methods can draw from a huge pool of highly developed tools from functional analysis, probability theory and statistics. This makes them much more favorable to work with compared to deep learning methods.

One example of a particularly popular kernel function is the Matérn kernel defined on Euclidean space. It possesses a continuous parameter controlling its smoothness, which links it intimitely to Sobolev spaces with real smoothness index (a.k.a. Bessel potential spaces). Its most popular member is probably the Laplacian kernel, which is "a bit less than" differentiable. In fact, choosing estimators with the right smoothness seems to be of fundamental importance for a successful application of a learning algorithm to a given problem. For an introduction to the Euclidean Matérn kernel, refer to [29] and for a recent survey on the ins and outs of Matérn kernels see [21].

Efforts have been made to generalize Matérn kernels to Riemannian manifolds [8, 11, 19]. As to be expected, the nice closed form expression is not what characterizes them. In particular, simply replacing the Euclidean norm by the Riemannian distance does not yield a well-behaved kernel, since the curvature is ignored [12]. Rather, one can characterize them as the reproducing kernel of a Sobolev space. Thus, instead of the term Matérn kernel, we use the term Sobolev kernel in this paper.

In this paper, we collect and review the theoretical tools to compute Sobolev kernels on compact Lie groups. Independently, they are all well-known and there is no new mathematics developed. Instead, the focus lies in laying out the foundations for further theoretical study and discussing some potential directions, since, to the best of the authors knowledge, they have not yet been considered in the literature.

In Section 2, we first fix some notation related to analysis on Riemannian manifolds and we review the basics about reproducing kernel Hilbert spaces. The main part consists of a study of the Laplace-Beltrami operator on Riemannian manifolds, how one can define Sobolev spaces with its help, and how to compute the reproducing kernels. The main result states that one needs to know the exact eigenstructure of the Laplace-Beltrami operator, to compute the kernel.

Section 3 is the core of this paper. In it, we carry out the spectral analysis of the Laplacian on compact Lie groups. We first introduce basic terminology and concepts about Lie groups, and review the notions of left/right/bi-invariant Riemannian metrics. Then, we discuss the algebra of left-invariant partial differential operators and show how the Laplacian can be identified with the quadratic Casimir operator. Finally, we introduce unitary representations of a Lie group, and prove the Peter-Weyl theorem: a deep result that decomposes L^2 into according to the action of the left/right regular representations. Finally, we conclude that this decomposition also determines the eigenstructure of the Laplacian.

1.1 Future directions

We indicate some directions for interesting future studies. For the sake of reading flow, it is probably best to read this section last.

- Finding all (or sufficiently many) irreducible representations of a Lie group is a hard problem. While the full eigenstructure of some concrete low-dimensional Lie groups are well-known (e.g. [22, 3, 7]), most works focus on eigenvalues and multiplicities and circumvent the eigenfunctions where possible. It would be quite helpful to survey what is already available: full descriptions for concrete groups, algorithms/implementations. A starting point for the latter could be [10] where the authors provide a constructive proof of the Peter-Weyl theorem (unlike ours which relies on Zorn's lemma).

- It seems much like the natural next step would be compact Riemannian symmetric spaces like spheres, projective space, Grassmannians, ... They are naturally linked to Lie groups, left-invariance can be replaced by G-invariance, and they admit a "substantial Fourier theory" (cf. [27, Section 7.6.5]). Moreover, much effort has already been put towards studying their Laplacians [11, 23, 9]. The question is to what extent the eigenstructures are known/feasible to compute.

- It would be very interesting to perform extensive experimentation with these Sobolev kernels. In view of [18], not only do they have the usual Sobolev smoothness control, but they also allow to easily integrate different (bi-invariant) Riemannian metrics. These are both very direct controls that could explain better how the geometry of data and the regularity properties of the estimator interplay and how they influences learning behavior with particular emphasis on benign overfitting [2] and double descent [5]. While the non-commutative case is of indepedent theoretic and applied (e.g. in robotics) interest, the n-torus has already proven to be a very promising proxy for Euclidean space (see [28, Chapter 4] for a discussion). It is especially feasible from a representation-theoretic perspective as it is abelian (all irreps are 1-dimensional) and we can easily write down the spectrum of the Laplacian in full detail.

2 Sobolev spaces on compact Riemannian manifolds

We review what is already known about Sobolev kernels on general compact Riemannian manifolds following the exposition in [11]. We introduce basic concepts about Riemannian manifolds, study the Laplace-Beltrami operator, review reproducing kernel Hilbert spaces and conclude with properties and an expression for the reproducing kernel of Sobolev spaces with smoothness index greather than half of the dimension of the space in Theorem 2.19.

2.1 Review of Riemannian manifolds

In this section we fix notation and recall some of the required concepts of Riemannian manifolds. Full details of everything that is discussed here is provided in [17, 16], although our notations slightly differ.

Throughout, let (M, g) denote a smooth connected n-dimensional Riemannian manifold with tangent bundle TM. We denote the set of vector fields by $\Gamma(TM)$. The space of differentiable k-forms, i.e. smooth sections of the bundle of alternating k-tensors, is denoted by $\Omega^k(M)$, where $\Omega^0(M) = C^\infty(M)$. The exterior derivative $d\colon \Omega^k(M) \to \Omega^{k+1}$ coincides with the usual differential in case $k = 0$.

Orientation and integration. Recall that M is orientable if and only if there is a non-vanishing top-dimensional form $\omega \in \Omega^n(M)$. If this is the case, we can always find a unique such ω_g, the Riemannian volume form, so that $\omega_g(E_1, \ldots, E_n) = 1$ for every oriented, orthonormal, local frame $(E_1, \ldots, E_n)$, $E_i \in \Gamma(TM)$. By setting $\mathrm{Vol}(U) = \mathrm{Vol}_g(U) \coloneqq \left|\int_U \omega_g\right|$ and extending to the Borel σ-algebra of M, we obtain a Radon measure $\mathrm{Vol} = \mathrm{Vol}_g$.

Flow and Lie brackets. The set of smooth vector fields (i.e. smooth sections of the tangent bundle) is denoted by $\Gamma(TM)$. For any $X \in \Gamma(TM)$ and any point $p \in M$ there exists a unique integral curve defined on a maximal open real interval, which passes through p and whose derivative coincides with the value of the vector field at all time. We denote by $\Phi_t^X\colon M \to M$ the diffeomorphism that assigns to each initial point the point of the corresponding integral curve at time t. This map is called the flow of the vector field X and possesses some nice properties: in any small enough open time frame and at any point $p \in M$, $\Phi_\cdot^X(p)$ solves the ODE defined by the vector field as a first-order differential operator:

$$\frac{d}{dt}\Phi_t^X(p) = X_{\phi_t^X(p)}.$$

Moreover, in adequate time frames the flow defines an action of $\mathbb{R}$ on M since $\Phi_t^X \circ \Phi_s^X = \Phi_{t+s}^X$. A vector field is called complete if the flow can be defined for all time. The classes of manifolds that we are interested in, namely compact ones and Lie groups, have the nice property that all vector fields are complete.

The Lie derivative of a vector field Y along another X is defined as the smooth vector field

$$(\mathcal{L}_X Y)_p \coloneqq \left.\frac{d}{dt}\right|_{t=0} (d\Phi_{-t}^X)_{\Phi_t^X(p)}(Y_{\Phi_t^X(p)}).$$

Then, the Lie brackets of $X, Y \in \Gamma(TM)$, $[X, Y] := \mathcal{L}_X Y$, turn the vector space $\Gamma(TM)$ into a Lie algebra: they are bilinear, alternating ($[X, X] = 0$) and satisfy the Jacobi identity $[X, [Y, Z]] + [Z, [X, Y]] + [Y, [Z, X]] = 0$. This structure has an alternative interpretation. Since every vector field defines a first-order differential operator on $C^\infty(M)$ (which satisfies the product rule $X(fg) = fXg + gXf$) we can define the product of two vector fields as their composition. Then, it turns out that $[X, Y] = XY - YX$: the Lie brackets measure the difference between flowing along Y then X versus the other way around. Two vector fields commute if their Lie bracket is zero.

Covariant derivative. The Riemannian metric g gives rise to a metric on M: Since any two points can be joined by a smooth curve $\gamma\colon [a, b] \to M$, the infimum of the length $\int_a^b \sqrt{g(\gamma'(t), \gamma'(y))_{\gamma(t)}}\, dt$ over all such curves can be shown to define a metric, which is called the Riemannian distance. Note that, even though the metric is well-defined, there is no guarantee that length-minimizing curves exist. We call M geodesically complete if they always do. The Hopf-Rinow theorem then states that M is geodesically complete if and only if it is complete as a metric space, and we simply call M complete in that case. Consequently, all compact Riemannian manifolds are complete.

The Riemannian structure also suggests that we should investigate how different tangent spaces are related to each other. The Lie derivative, already provided a notion of derivative along vector fields, but it has the big drawback that it is not "fully" linear: $\mathcal{L}_{fX} Y \neq f\mathcal{L}_X Y$ in general. The right notion is that of a connection: a map $\nabla\colon \Gamma(TM) \times \Gamma(TM) \to \Gamma(TM)$, denoted $(X, Y) \mapsto \nabla_X Y$, which is $C^\infty(M)$-linear in X, $\mathbb{R}$-linear in Y, and satisfies the product rule $\nabla_X(fY) = f\nabla_X Y + (Xf)Y$. In fact, there is a unique connection if we impose compatibility with the metric structure, $\nabla_X g(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z)$, and symmetry, $\nabla_X Y - \nabla_Y X = [X, Y]$. It is called the Levi-Civita connection. We will only ever consider the Levi-Civita connection in this paper.

Important differential operators. First, the gradient of $f \in C^\infty(M)$ is defined as the unique vector field $\operatorname{grad} f$ such that $g(\operatorname{grad} f, v) = df(v)$ for all $v \in TM$. Thus, the gradient is a map $\operatorname{grad}\colon C^\infty(M) \to \Gamma(TM)$. Next up, we can define the divergence of a vector field $X \in \Gamma(TM)$ as $\operatorname{div} X := \operatorname{trace}(Y \mapsto \nabla_Y X)$. Note that the map $Y \mapsto \nabla_Y X$ is a linear map at each tangent space, so that using the trace here is justified. The divergence is a map $\operatorname{div}\colon \Gamma(TM) \to C^\infty(M)$. The defintion is in analogy to the Euclidean case: Computing the connection in local coordinates allows thinking of $Y \mapsto \nabla_Y X$ as the Jacobian matrix, and the sum of the diagonal is exactly the divergence in $\mathbb{R}^n$. Finally, the Laplacian is the differential operator $\Delta\colon C^\infty(M) \to C^\infty(M)$, defined by $\Delta f := \operatorname{div}(\operatorname{grad} f)$. If we additionally assume that M is compact, then we have the following formulations of the divergence theorem and Green's identity

$$\int_M \operatorname{div} X \, dV = 0, \qquad \int_M f_1 \Delta f_2 \, dV + \int_M g(\operatorname{grad} f_1, \operatorname{grad} f_2)\, dV = 0. \tag{1}$$

2.2 The Laplace-Beltrami operator and Sobolev spaces

In this section we are going to review some facts about Sobolev spaces on compact Riemannian manifolds. We skip proofs in order to avoid a discussion of the necessary tools from functional analysis.

Throughout, let (M, g) be a compact, connected and oriented Riemannian manifold with volume element denoted by dV. As a first step, we need to extend the Laplacian to an operator on $L^2(M)$. We start with a heuristic argument by means of Green's identity. If $f \neq 0$ is a non-vanishing smooth function, then

$$\langle f, \Delta f \rangle_{L^2} = \int_M f \Delta f \, dV = -\int_M g(\operatorname{grad} f, \operatorname{grad} f) \, dV < 0.$$

This hints to negativity. Moreover, we also have symmetry. Indeed by the same identity,

$$\langle f_1, \Delta f_2 \rangle_{L^2} = -\int_M g(\operatorname{grad} f_1, \operatorname{grad} f_2) \, dV = -\int_M g(\operatorname{grad} f_2, \operatorname{grad} f_1) \, dV = \langle \Delta f_1, f_2 \rangle_{L^2} .$$

Now, trivially, $C^\infty(M) \subset L^2(M)$ since M is compact, and recall that the smooth functions form of dense subset of $L^2(M)$. In conclusion, we have a densely defined symmetric operator, sparking hope that we might be able to extend it nicely onto $L^2(M)$. This is indeed the case if we assume the Riemannian manifold to be complete. Moreover, in the compact case (which implies completeness) we can make statements about the spectrum:

Theorem 2.1 (Sturm-Liouville decomposition)**.** *Let M be a compact Riemannian manifold. Then the operator $\Delta \colon C^\infty(M) \to L^2(M)$ extends uniquely to an unbounded self-adjoint negative operator $\Delta \colon L^2(M) \to L^2(M)$. There is an orthonormal basis $\{f_k\}_{k\geq 1}$ of $L^2(M)$ consisting of smooth eigenfunctions together with a divergent sequence $\{\lambda_k\}_{k\geq 1}$ of negative eigenvalues such that $\Delta f_k = \lambda_k f_k$ and each eigenvalue appears only a finite number of times.*

The proof of the first statement can be found in [25, Section 2] and the second part is proven in [6, Theorem 18].

Definition 2.2. *Let M be a compact Riemannian manifold. Define the operator $J \coloneqq (I - \Delta)^{-1/2}$. The Bessel potential is defined as the symbol J^s for a real number $s \in \mathbb{R}$ and acts on $f \in L^2(M)$ as*

$$J^s f = \sum_k (1 - \lambda_k)^{-s/2} \langle f, f_k \rangle_{L^2(M)} f_k .$$

The Sobolev space $H^s(M)$ with real smoothness index $s > 0$ is defined as the domain of the Bessel potential, which then defines an operator $J^s \colon H^s(M) \to L^2(M)$.

Proposition 2.3. *The domain of the Bessel potential for $s > 0$ is the set*

$$\left\{ f \in L^2(M) \,\middle|\, \|J_s f\|_{L^2} = \sum_k (1 - \lambda_k)^{-s} \left| \langle f, f_k \rangle_{L^2} \right|^2 < \infty \right\}.$$

Moreover, $H^s(M)$ becomes a Hilbert space when equipped with the inner product $\langle f, g \rangle_{H^s} \coloneqq \langle J^{-s} f, J^{-s} g \rangle_{L^2}$.

The proof can be found in [11] and references therein. Finally, the main reason that we can speak about reproducing kernel Hilbert spaces, which are spaces of functions, is the following result.

Theorem 2.4 (Rellich-Kondrachov theorem)**.** *Let M be a compact Riemannian manifold. For $0 < s < t$ embedding $H^t(M) \hookrightarrow H^s(M)$ is compact. Futhermore, if $s > \dim(M)/2$, then we have another compact embedding $H^s(M) \hookrightarrow C(M)$.*

See [26, Propositions 3.3 and 4.4] or [1, Theorem 2.34] for a proof.

2.3 Reproducing kernels

We review the basic concepts in the theory of reproducing kernel Hilbert spaces (see [24, Chapters 4] for more details, or [28, Chapter 3] for a more applied perspective) and conclude with an explicit description of the Sobolev reproducing kernel on a compact Riemannian manifold. Throughout, let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ and X a set (we impose more structure as we go).

Definition 2.5. *A function $k \colon X \times X \to \mathbb{K}$ is called a kernel if there is a $\mathbb{K}$-Hilbert space F and a map $\Phi \colon X \to F$ such that*

$$k(x, x') = \langle \Phi(x'), \Phi(x) \rangle_F \tag{2}$$

for all $x, x' \in X$. In that case, Φ is called the feature map and F the feature space.

Given a concrete problem, feature maps should be constructued, so that the associated kernel acts as a measure of similarity. In fact, a kernel k naturally induces a pseudometric on the underlying space X: if $\Phi \colon X \to F$ is a feature map, then the kernel metric d_k is defined by

$$d_k(x, x') := \|\Phi(x) - \Phi(x')\|_H = \sqrt{k(x, x) + k(x', x') - 2k(x, x')}.$$

Note that this definition is independent of the feature map and that d_k is only a proper metric if Φ is injective but, nevertheless, generates a topology on X. Our first result characterizes continuity.

Lemma 2.6. *If $(X, \mathcal{T})$ is a topological space and k a kernel on X with feature map $\Phi \colon X \to F$, then the following are equivalent:*

1. *k is continuous w.r.t. the product space topology,*
2. *k is continuous separately in both arguments and $x \mapsto k(x, x)$ is continuous,*
3. *$\iota \colon (X, \mathcal{T}) \to (X, d_k)$ is continuous, and*
4. *$\Phi \colon (X, \mathcal{T}) \to F$ is continuous.*

Proof. (1) immediately implies (2).

If (2) holds, then $d_k(\cdot, x)\colon (X, \mathcal{T}) \to \mathbb{R}$ is continuous for each x. Therefore, the "open" ε-ball $\{x' \in X \mid d_k(x', x) < \varepsilon\}$ around x is truly open w.r.t. $\mathcal{T}$ and (3) follows.

Let (3) hold. It is easy to see that the map $\Phi'\colon (X, d_k) \to F$ is continuous by definition of the kernel metric. Therefore, so is $\Phi = \Phi' \circ \iota$ and (4) follows.

Finally, let (4) hold. In order to show (1) and close the equivalence, it suffices to observe that the product $\Phi \times \Phi\colon X \times X \to F \times F$ is continuous and $k = \langle \cdot, \cdot \rangle_F \circ (\Phi \times \Phi)$. □

Such kernels are in a one-to-one relationship with specific kind of feature spaces. In order to establish that result, let us first define a special class of kernel-space combinations.

Definition 2.7. *Let H be a $\mathbb{K}$-Hilbert space of functions $X \to \mathbb{K}$ where sum and scalar multiplication are defined as the usual pointwise ones. A function $k\colon X \times X \to \mathbb{K}$ is called a reproducing kernel of H if $k_x := k(\cdot, x) \in H$ for any x and it satisfies the reproducing property*

$$f(x) = \langle f, k_x \rangle \tag{3}$$

for any x and $f \in H$. In that case k_x is called canonical feature map. If a Hilbert function space H has a reproducing kernel, we call it a reproducing kernel Hilbert spaces (RKHS, pl. RKHSs).

Proposition 2.8. *A reproducing kernel k is a kernel in the previous sense and satisfies the following version of Cauchy-Schwarz:*

$$|k(x, x')|^2 \leq k(x, x)k(x', x'). \tag{4}$$

Proof. A reproducing kernel on a space H is easily seen to be a kernel by the reproducing property (Eq. (3)) with the canonical feature map $x \mapsto k_x$:

$$k(x, x') = k_{x'}(x) = \langle k_{x'}, k_x \rangle_H. \tag{5}$$

The estimate follows immediately:

$$|k(x, x')|^2 = |\langle k_{x'}, k_x \rangle|^2 \leq \|k_{x'}\|^2 \|k_x\|^2 = k(x, x)k(x', x'). \tag{6}$$

□

They have an important characterization, which is quite handy from a functional analysis point-of-view. It reveals that RKHSs are much easier to deal with than Lebesgue spaces for instance, convergence in norm already implies pointwise convergence.

Proposition 2.9. *Let H be a Hilbert space of functions on X and for each $x \in X$ let $L_x\colon H \to \mathbb{K}$ be the evaluation functional defined by $L_x(f) = f(x)$. Then, H is an rkhs iff L_x is bounded. In that case, norm convergence implies pointwise convergence.*

Proof. Let first H be an rkhs with reproducing kernel k. Then,

$$|L_x(f)| = |f(x)| = |\langle f, k_x \rangle_H| \leq \|k_x\|_H \|f\|_H, \tag{7}$$

so $\|L_x\| \leq \|k_x\|_H < \infty$. In this case, if $f_n \to f$ converges in H-norm, then

$$|f_n(x) - f(x)| = |L_x(f_n - f)| \leq \|L_x\|_H \|f_n - f\|_H \to 0, \tag{8}$$

and the function f_n converges pointwise to f.

Conversely, let H be a Hilbert function space with bounded evaluation functionals L_x. Consider the function $k(x, x') = \langle L_x, L'_x \rangle_{H^*}$ defined as the inner product on the Hilbert space dual. Let $R\colon H^* \to H$ be the isometric, (anti-)linear Riesz isomorphism. Then

$$k_{x'}(x) = k(x, x') = \langle L_x, L_{x'} \rangle_{H^*} = \langle RL_{x'}, RL_x \rangle_H = L_x(RL_{x'}) = (RL_{x'})(x) \tag{9}$$

or simply $k(\cdot, x') = RL_{x'}$. The reproducing property follows:

$$f(x') = L_{x'} f = \langle f, RL_{x'} \rangle_H = \langle f, k_{x'} \rangle_H.$$

Thus, k is a reproducing kernel and H a rkhs. □

In fact, such kernels are unique to the space in the following sense.

Proposition 2.10. *Let H be a rkhs with reproducing kernel k and pick an onb $(e_i)_i$ of H. Then,*

$$k(x, x') = \sum_i \overline{e_i(x')} e_i(x). \tag{10}$$

In particular, if a Hilbert space of functions has a reproducing kernel, it must be unique.

Proof. A straightforward application of Parseval's identity and reproducing property:

$$k_{x'} = \sum_i (e_i \otimes e_i)(k'_x) = \sum_i \langle k_{x'}, e_i \rangle_H e_i = \sum_i \overline{e_i(x')} e_i. \tag{11}$$

This expression converges in H-norm, thus, by the previous result also pointwise and the claim follows. □

The following result describes RKHSs quite explicitly.

Proposition 2.11. *Let H be a rkhs with reproducing kernel k. Define the vector subspace*

$$H_{pre} := \left\{ f = \sum_{i=1}^{\infty} c_i k_{x_i} \colon x_i \in X, c_i \in \mathbb{K}, \#\{c_i \neq 0\} < \infty \right\} \subseteq H, \tag{12}$$

of all finite linear combinations of the canonical feature map. If we equip it with the inner product

$$\langle f, g \rangle_{pre} = \langle \sum_i c_i k_{x_i}, \sum_j d_j k_{y_j} \rangle_{pre} := \sum_{i,j} c_i \overline{d_j} k(x_i, y_j), \tag{13}$$

then H is the completion of the pre-Hilbert space H_{pre}.

Proof. Observe that $\langle\cdot,\cdot\rangle_{\text{pre}}$ is nothing but the restriction of $\langle\cdot,\cdot\rangle_H$. Therefore, it suffices to prove density. If this wasn't the case, H_{pre} has a non-trivial orthogonal complement in H. This would imply that there are $f \in H_{\text{pre}}^{\perp}$ and $x \in X$ such that $f(x) \neq 0$. But since $k_x \in H_{\text{pre}}$

$$0 = \langle f, k_x \rangle = f(x) \neq 0. \tag{14}$$

□

We have already shown that a reproducing kernels are unique, if they exist. The following result establishes a reverse statement.

Theorem 2.12 (Rkhss and kernels are 1-to-1)**.** *Each rkhs H has the unique reproducing kernel k given by*

$$k(x,x') = \sum_i \overline{e_i(x')} e_i(x), \tag{15}$$

where $(e_i)_i$ is any onb of H. Conversely, every kernel k with feature map Φ and feature space F gives rise to a unique (up to isomorphism) rkhs H for which it is a reproducing kernel. This space is given by

$$H = \{V(g) \coloneqq (x \mapsto \langle g, \Phi(x)\rangle_F) \colon g \in g\} \tag{16}$$

equipped with the norm

$$\|f\|_H \coloneqq \inf_{g \in V^{-1}(f)} \|g\|_F. \tag{17}$$

Proof. The first part of the theorem has already been proven in Proposition 2.10.

We now prove that the space H as defined above is indeed a Hilbert space by establishing an isometry with a subspace of F via the operator $V \colon F \to H$. First, it is easily seen that $\ker V$ is closed (take a convergent sequence and notice that $V(\cdot)(x)$ is continuous) so that we have an orthogonal sum $H = \ker V \oplus F'$. The restriction $V|_{F'}$ is, therefore, injective and also surjective: if $f = V(g)$, decompose $g = g_0 + g'$, thus, $f = V|_{F'} g'$. Moreover, we find (with the same technique) that $\|f\|_H^2 = \left\|(V|_{F'})^{-1} f\right\|_{F'}^2$, which implies that $V|_{F'}$ is an isometric isomorphism and H is a Hilbert space.

Next, we show that k is a reproducing kernel of H. Since Φ is its feature map we have

$$k_x = \langle \Phi(x), \Phi(\cdot)\rangle_F = V\Phi(x) \in H. \tag{18}$$

Moreover, $\langle g, \Phi(x)\rangle = 0$ for any $g \in \ker V$ so that $\Phi(x) \in (\ker V)^{\perp} = F'$. The reproducing property follows by using the isometry of $V|_{F'}$

$$f(x) = \langle (V|_{F'})^{-1} f, \Phi(x)\rangle_F = \langle f, V|_{H'}\Phi(x)\rangle_H = \langle f, k_x\rangle_H. \tag{19}$$

Finally, we prove uniqueness. Observe first that the inner product on the dense subspace H_{pre} of H (from Proposition 2.11) is only dependent on the kernel k so that it is a dense for all rkhss with this reproducing kernel. Now let H_1 and H_2 be two rkhss with reproducing kernel k. It clearly is enough to show that $H_1 \subseteq H_2$ is an isometric inclusion as the other inclusion will work the same way. Both contain the dense subset

H_{pre}. We fix $f \in H_1$ together with a sequence $f_n \in H_{\text{pre}}$ such that $f_n \to f$ in H_1. This sequence is also contained in H_2 and is Cauchy also there, so that $f_n \to g$ in H_2 for some g. But norm convergence already implies point-wise convergence which implies $f = g \in H_2$. Finally, since norms coincide on the dense subset

$$\|f\|_{H_1} = \lim \|f_n\|_{H_{\text{pre}}} = \|f\|_{H_2}.$$

□

Now we add more structure to our space. In view of the previous section on Sobolev spaces, the goal here is to study RKHSs of square-integrable functions. For this, we prove two easy results on separability and measurability.

Lemma 2.13. *If X is a separable space and k a continuous kernel then the associated rkhs is separable.*

Proof. By the previous lemma we know that the feature map Φ associated to k is continuous. This implies that $\Phi(X)$ is separable and, therefore, also H_{pre} as the span. The former space is dense and the claim follows. □

Lemma 2.14. *Let k be a kernel on a measurable space X and let H be the associated rkhs. Then k_x is measurable for all x iff all $f \in H$ are.*

Proof. Since $k_x \in H$, the sufficiency is immediate. If, conversely, all k_x are measurable, then so are all functions in H_{pre} (as in the proof of Theorem 2.12). Fix now $f \in H$ and a sequence $(f_n)_n \subset H_{\text{pre}}$ such that $f_n \to f$ in H. Since this implies point-wise convergence, and all f_n are measurable, so is f. □

From now on, we let X be a topological space with a finite Borel measure μ so that $L^2(X, \mu)$ is separable. Moreover, H will be a separable RKHS with continous kernel k.

Proposition 2.15 (Inclusion operator)**.** *Assuming that*

$$\|k\|_{L^2(\mu)} := \left(\int_X k(x,x)\, d\mu(x) \right)^{1/2} < \infty$$

is finite, all members of H are square-integrable and the canonical embedding $H \to L^2(\mu)$ sending functions to their equivalence classes is continuous with operator norm bounded by $\|k\|_{L^2(\mu)}$.

Proof. Fix an $f \in H$. Recall that $\|k_x\|_H = \sqrt{k(x,x)}$. Both claims follow from a simple application of Hoelder's inequality:

$$\begin{aligned}
\|f\|^2_{L^2(\mu)} &= \int_X |f(x)|^2\, d\mu(x) = \int_X |\langle f, k_x \rangle_H|^2\, d\mu(x) \\
&\leq \|f\|^2_H \int_X k(x,x)\, d\mu(x) = \|f\|^2_H \, \|k\|_{L^2(\mu)}.
\end{aligned}$$

□

We denote the canonical embedding by I_k.

Definition 2.16. *The kernel operator is defined as the operator* $K := I_k I_k^* \colon L^2(\mu) \to L^2(\mu)$.

Proposition 2.17 (Compactness)**.** *We assume that* $\|k\|_{L^2(\mu)} < \infty$. *Then, both the inclusion* I_k *and its adjoint* I_k^* *are Hilbert-Schmidt operators with* $\|I_k\|_{HS} = \|k\|_{L^2(\mu)} = \|I_k^*\|_{HS}$. *Moreover, the kernel operator is positive, self-adjoint and trace-class with* $\operatorname{trace} K \leq \|k\|^2_{L^2(\mu)}$.

Proof. We start with the inclusion I_k. Fix an orthonormal basis $(e_i)_{i\in J}$ (J countable by assumption) of H. Then,

$$\begin{aligned}\|I_k\|^2_{\mathrm{HS}} &= \sum_i \|I_k e_i\|^2_{L^2(\mu)} \\ &= \int_X \sum_i |I_k e_i(x)|^2 \, d\mu(x) = \int_X \sum_i |e_i(x)|^2 \, d\mu(x) = \|k\|^2_{L^2(\mu)},\end{aligned}$$

where the last equation follows from Eq. (10). Thus, S_μ is Hilbert-Schmidt. Recalling that a compact operator is Hilbert-Schmidt iff its adjoint is, the same holds for I_k^* with equal norm.

Since K is a compositions of Hilbert-Schmidt operators, they are trace-class themselves and, in particular, compact. Note that they are clearly positive and self-adjoint as the compositions of a compact operator with its adjoint. By the previous results $\operatorname{trace} K$ is bounded by $\|I_k\|\|I_k^*\| = \|k\|^2_{L^2(\mu)}$. □

Finally, the compactness of the kernel operator motivates the following key result.

Theorem 2.18 (Mercer's theorem)**.** *Let* X *be a compact metric space with a Borel measure* μ *with full support and* k *a continuous kernel on* X. *Let* $(e_i, \lambda_i)_{i\in I}$ *be the eigensystem of the kernel operator* $K_\mu \colon L^2(\mu) \to L^2(\mu)$ *Then, the functions* $\widetilde{e}_i := \lambda_i^{-1} I_k^* e_i$ *make up an ONB* $\{\sqrt{\lambda_i}\widetilde{e}_i\}_{i\in I}$ *of* H. *Then,*

$$k(x, x') = \sum_i \lambda_i e_i(x) e_i(x'), \quad x, x' \in X.$$

We come back to the case of Sobolev spaces. The following result is the main one of this section: it explicitly describes the kernel operator and gives an expression for the kernel using Mercer's theorem.

Theorem 2.19. *Let* M *be a compact Riemannian manifold. For* $s > \dim(M)/2$, *the Sobolev space* $H^s(M)$ *is a reproducing kernel Hilbert space with reproducing kernel* $k^s \colon M \times M \to \mathbb{C}$. *In that case the kernel integral operator* $K^s \colon L^2(M) \to L^2(M)$ *satisfies*

$$K^s = I_s I_s^* = J^{2s},$$

where $I_s \colon H^s(M) \hookrightarrow L^2(M)$ *and the kernel, evaluated at* $p, p' \in M$, *is explicitly expressed as*

$$k^s(p, p') = \sum_k (1 - \lambda_k)^{-s} f_k(p) \overline{f_k(p')}.$$

Here, we make use of the eigenstructure of the Laplace-Beltrami operator from Theorem 2.1.

Proof. By the Rellich-Kondrachov theorem (Theorem 2.4), $H^s(M)$ can be identified as a space of continuous functions provided $s > \dim(M)/2$. In particular, all members are bounded since M is compact and, thus, all evaluation functionals are bounded. We compute the kernel operator. Let $h \in L^2(M)$ and $g \in H^s(M)$. Then,

$$\langle h, I_s g\rangle_{L^2} = \left\langle J^{-2s} J^{2s} h, I_s g\right\rangle_{L^2} = \left\langle J^{-s} J^{2s} h, J^{-s} I_s g\right\rangle_{L^2}.$$

Now, $J^{2s}h$ and g are both part of the domain of J^{-s}, It follows that there is $f \in H^s(M)$ such that $I_s f = J^{2s}h$ and $\langle h, I_s g\rangle_{L^2} = \langle f, g\rangle_{H^s}$. Consequently, $I_s^* h = f$ and $I_s I_s^* h = J^{2s}h$. Finally, using the volume element of (M, g) as a measure (with full support), Mercer's theorem (Theorem 2.18) implies the kernel expression and the proof is complete. □

3 Analysis on Lie groups

We now specialize our theory to compact Lie groups. First, we review the necessary concepts of Lie groups and explore the notion of left/right/bi-invariant Riemannian metrics. Afterwards, we introduce the universal enveloping algebra; an important tool in representation theory which can be identified with an algebra of differential operators on the group. In particular, we show that Laplacian under this identification corresponds to the quadratic Casimir element. Afterwards, we review the theory of unitary representations, prove the Peter-Weyl theorem and relate the L^2-decomposition that it produces to the eigenstructure of the Laplacian with Theorem 3.38.

3.1 Basic concepts

We start with our review of Lie groups. Everything presented is classical and can be found for instance in [17].

Let $M_n(\mathbb{C}) \cong \mathbb{R}^{4n^2}$ be the (real) manifold of all $n \times n$-matrices with complex coefficients. We will switch often between real and "artificial" complex structure in this sense. Note that $\det\colon M_n(\mathbb{C}) \to \mathbb{C}$ is smooth, so that $\mathrm{GL}_n(\mathbb{C}) = \{A \in M_n(\mathbb{C}) \mid \det(A) \neq 0\}$ is an open submanifold with equal dimension. Clearly, matrix multiplication is smooth and so is inversion (look at Cramer's rule).

Definition 3.1. *A (real) Lie group is a group and equipped with the structure of a smooth manifold such that both multiplication and inversion are smooth. A matrix Lie group is a closed subgroup of* $\mathrm{GL}_n(\mathbb{C})$.

We will also always assume that the Lie group is connected. One can show that this is actually not a restriction in the case of a compact groups: they can always be embedded into $U(n)$, the compact group of unitary matrices with complex coefficients (with the orthogonal matrix group $O(n)$ as a special case for purely real groups). Due to the applied nature of this paper we will mostly be concerned with matrix Lie groups (which, actually, is not a restriction for compact groups). We will be using multiplicative notations (xy for multiplication, x^{-1} for inversion and 1 for the identity element). If G is a Lie group and $y \in G$, then the left and right multplication maps

$$\begin{aligned} L_y &\colon G \to G, \quad x \mapsto yx \\ R_y &\colon G \to G, \quad x \mapsto xy^{-1} \end{aligned}$$

are both diffeomorphisms.

Definition 3.2. *A vector field $X \in \Gamma(TG)$ is called left-invariant if $(L_y)_* X = X$ (or more explicitly $dL_y \circ X = X \circ L_y$) for all $y \in G$. The vector subspace of all left-invariant vector fields is denoted $\Gamma^L(TG)$.*

Recall that a smooth vector field X with flow $\Phi_t^X \colon G \to G$ acts as a differential operator on $C^\infty(G)$ via the Lie derivative:

$$Xf(x) \coloneqq \mathcal{L}_X f \, \frac{d}{dt} \left(f \circ \Phi_t^X \right) \Big|_{t=0} (x).$$

This operation satisfies the product rule $X(fg) = fXg + gXf$, and, reversely, any linear operator on $C^\infty(G)$ that satisfies it, must be coming from a vector field (see e.g. [17, Proposition 8.15]). As already discussed earlier, the Lie brackets

$$[X, Y] = XY - YX, \quad \text{i.e. } [X, Y](f) = X(Yf) - Y(Xf)$$

make $\Gamma(TG)$ a Lie algebra. For Lie groups, we have the following key result for the subspace of left-invariant vector fields. (or right-invariant, the theory works the same)

Proposition 3.3. *The left-invariant vector fields $\Gamma^L(TG)$ of a Lie group together with the Lie bracket form a Lie subalgebra of $\Gamma(TG)$, which we explicitly denote* $\mathrm{Lie}(G)$*. The two vector spaces* $\mathrm{Lie}(G)$ *and $T_1 G$ are canonically isomorphic.*

Proof. First, we need to show that $[\Gamma^L(TG), \Gamma^L(TG)] \subset \Gamma^L(TG)$. If X and Y are left-invariant, then

$$(L_x)_*[X, Y] = [(L_x)_* X, (L_x)_* Y] = [X, Y],$$

where the first equality follows from the fact that L_x is a diffeomorphism and [17, Corollary 8.31]. Thus, $\Gamma^L(TG)$ is a Lie subalgebra of $\Gamma(TG)$.

Now, define a map $\mathrm{Lie}(G) \to T_1 G$ by $X \mapsto X_1$. This is clearly linear and injective: if $X_1 = 0 \in T_1 G$, then for any $x \in G$, $X_g = (X \circ L_g)_1 = (dL_g \circ X)_1 = 0$ by left-invariance. To prove surjectivity, let $v \in T_1 G$ be arbitrary. We claim that $(v^L)_x \coloneqq dL_x(v) \in T_x G$ is a smooth left-invariant vector field, which, clearly, satisfies $(v^L)_1 = v$. We skip the smoothness check (see [17, proof of Theorem 8.37] for that) and we conclude with left-invariance. Indeed,

$$(v^L \circ L_x)_y = (v^L)_{xy} = dL_{xy}(v) = (dL_x \circ dL_y)(v) = (dL_x \circ v^L)_y,$$

where we used that $d(f \circ g) = df \circ dg$ for smooth maps f and g. $\square$

This result has several consequences, that clarify how much structure Lie groups really have in comparison to general Riemannian manifolds:

- the Lie algebra $L = \mathrm{Lie}(G)$ of G has the same dimensionality as the group itself: $\dim L = \dim T_1 G = \dim G$,
- any left-invariant vector field X on G is already determined by its value at 1: $X_x = ((X_1)^L)_x$,
- left-invariant vector fields are complete,
- any vector space basis of L corresponds to a global frame on TG, so that G is parallelizable (i.e. TG is diffeomorphic to $G \times T_1 G$; see [17, Chapter 10]), and
- as a consquence G is orientable.

It is straightforward to compute the Lie algebra in many important cases. For instance, $Lie(\mathrm{GL}_n, \mathbb{C}) = M_n(\mathbb{C}) \cong M_{2n}(\mathbb{R})$. In an abstract setting with a vector space V, we might also write $\mathrm{Lie}(\mathrm{GL}(V)) = \mathrm{End}(V)$.

We give another characterization for the tangent space.

Definition 3.4. *A one-parameter subgroup γ of G is a Lie group homomorphism (i.e. a smooth group homomorphism) $\gamma\colon \mathbb{R} \to G$.*

Proposition 3.5 (Characterization of one-parameter subgroups)**.** *The one-parameter subgroups of G are exactly the maximal integral curves of left-invariant vector fields starting at the identity.*

Proof. First, let γ be an integral curve for some $X \in \mathrm{Lie}(G)$, starting at 1. Since every left-invariant vector field is complete γ is defined on the whole of $\mathbb{R}$. The left-invariance of X implies that $t \mapsto L_y(\gamma(t))$ for any $\gamma(s) = y$ is another integral curve of X starting at y. But $t \mapsto \gamma(s+t)$ is also an integral curve starting at y, which means that they are equal, and that $\gamma(s)\gamma(t) = \gamma(s+t)$. Thus, γ is a one-parameter subgroup.

On the other hand, let γ be a one-parameter subgroup and d/dt the (left-invariant) vector field of $\mathbb{R}$. We define the left-invariant vector field $X = d\gamma(d/dt) \in \mathrm{Lie}(G)$. Then, we have for any point in time $t_0 \in \mathbb{R}$

$$\gamma'(t_0) = d_{t_0}\gamma \left(\left. \frac{d}{dt} \right|_{t_0} \right) = X_{\gamma(t_0)},$$

so γ is an integral curve for X. □

Note that one-parameter subgroups are in bijection with left-invariant vector fields, and therefore also with $T_1 G$. What we get in addition, is a very explicit expression for such curves via the exponential map. The following properties can all be found in [17, Proposition 20.5 and Proposition 20.8].

Proposition 3.6. *Let the map* $\exp\colon \mathrm{Lie}(G) \to G$ *be defined by* $\exp(X) = \gamma_X(1)$*, where* γ_X *is the unique one-parameter subgroup associated to* X *established by the previous result. It has the following properties:*

1. *the curve* $t \mapsto \exp(tX)$ *is exactly the one-parameter subgroup generated by* X*,*
2. $\exp((s+t)X) = \exp(sX)\exp(tX)$*,*
3. $\exp(X)^{-1} = \exp(-X)$*,*
4. $\exp(X)^n = \exp(nX)$*,*
5. $d\exp|_0\colon \mathrm{Lie}(G) \to \mathrm{Lie}(G)$ *is the identity map,*
6. $\exp$ *restricts to diffeomorphism from some neighborhood of* $0 \in \mathrm{Lie}(G)$ *to a neighborhood of* $1 \in G$*,*
7. *the flow of a left-invariant vector field* X *is given by* $\Phi_t^X = R_{-\exp(tX)}$*.*

In the case of a matrix Lie group $G \subseteq \mathrm{GL}_n(\mathbb{C})$ with Lie algebra $L = \mathrm{Lie}(G) \subseteq M_n(\mathbb{C})$, this simplies to the matrix exponential: Let $A \in L$ be a matrix and denote by A^L the corresponding left-invariant vector field. The one-parameter subgroup γ that A generates must satisfy the initial-value problem

$$\gamma'(t) = A^L\big|_{\gamma(t)}, \quad \gamma(0) = 1_n.$$

It is well-known that the solution to this ODE is given by

$$\gamma(t) = \exp(tA) := \sum_{k=0}^{\infty} \frac{t^k A^k}{k!},$$

which converges to an invertible matrix.

Next, we prove that a Lie group homomorphism gives rise to a compatible homomorphism on the Lie algebras, sometimes called the differential homomorphism.

Proposition 3.7. *If* $\varphi\colon G \to H$ *is a homomorphism of Lie groups, then* $d\varphi := d_1\varphi\colon \mathrm{Lie}(G) \to \mathrm{Lie}(H)$ *is a Lie algebra homomorphism and the following diagram commutes:*

$$\begin{tikzcd} G \arrow[r, "\varphi"] & H \\ \mathrm{Lie}(G) \arrow[u, "\exp"] \arrow[r, "d\varphi"] & \mathrm{Lie}(H). \arrow[u, "\exp"'] \end{tikzcd}$$

Note the slight abuse of notation here: we identify Lie algebra and tangent space without comment.

Proof. Let $X^L \in \mathrm{Lie}(G)$ denote the vector field, and $X = X^L_1$ the tangent vector, that defines it uniquely. We first need to assert that the vector field Y^L determined by $Y \coloneqq d_1\varphi(X^L_1) = d_1\varphi(X)$ is left-invariant. Note that, since φ is a group homomorphism, $\varphi(L_y(x)) = L_{varphi(x)}(\varphi(y))$ for all $x, y \in G$. In other words, $\varphi \circ L_y = L_{\varphi(y)} \circ \varphi$. Therefore,

$$d\varphi(X^L_y) = (d\varphi \circ dL_y)(X^L_e) = (dL_{\varphi(y)} \circ d\varphi)(X^L_e) = dL_{\varphi(y)}(Y^L_e) = Y_{\varphi(y)}.$$

Then, since $d\varphi([X^L, Y^L]) = [d\varphi \circ X^L, d\varphi \circ Y^L] \in \mathrm{Lie}(H)$, we know that $d\varphi$ is a homomorphism of Lie algebras.

Next, we prove that the diagram commutes, i.e. $\exp(d\varphi(X)) = \varphi(\exp(X))$. Knowing that $t \mapsto \exp(td\varphi(X))$ is the one-parameter subgroup generated by $d\varphi(X)$, we only need to show that on the right-hand side, $\psi(t) \coloneqq \varphi(\exp(tX))$ has initial velocity $\psi'(0) = d\varphi(X)_e$:

$$\psi'(0) = \left.\frac{d}{dt}\right|_{t=0} \varphi(\exp(tX)) = d\varphi_0 \underbrace{\left(\left.\frac{d}{dt}\right|_{t=0} \exp(tX) \right)}_{X_1} = d\varphi(X)_e$$

and the proof is concluded. ☐

Definition 3.8. *Let G be a Lie group with Lie algebra $L = \mathrm{Lie}(G)$. For $x \in G$ denote by $C_x \colon G \to G$ the conjugation map $C_x = L_x \circ R_{x^{-1}} \colon y \mapsto xyx^{-1}$. The adjoint representation of G is the map*

$$\mathrm{Ad}\colon G \to \mathrm{GL}(L), \quad x \mapsto \mathrm{Ad}(x) = X \mapsto dC_x(X).$$

Its derivatve $\mathrm{ad} = d\,\mathrm{Ad}\colon L \to \mathrm{End}(L)$ *will also be referred to as adjoint representation.*

Lemma 3.9. *The adjoint (Lie algebra) representation is given by* $\mathrm{ad}(X) = [X, \cdot]$.

Proof. The linear map $d\,\mathrm{Ad}(X) \in \mathrm{End}(L)$ is determined by its value at the identity. Thus, we can compute the action of $d\,\mathrm{Ad}(X)$ on Y using the exponential map: by Proposition 3.6 $\exp(tX)$ is exactly the one-parameter subgroup generated by X. Therefore,

$$d\,\mathrm{Ad}(X)Y = \left(\left.\frac{d}{dt}\right|_{t=0} \mathrm{Ad}(\exp(tX)) \right) Y = \left.\frac{d}{dt}\right|_{t=0} \left(\mathrm{Ad}(\exp(tX))Y\right).$$

Now, $\mathrm{Ad}(\exp(tX))Y$ is left-invariant, and thus also determined by its value at the identity. Therefore, using the fact that $\mathrm{Ad}(x) = dR_{x^{-1}} \circ dL_x$ and the the expression for the flow of left-invariant vector fields (Proposition 3.6),

$$(\mathrm{Ad}(\exp(tX))Y)_1 = (dR_{\exp(-tX)} \circ dL_{\exp(tX)})(Y_1) = d\Phi^X_{-t}\left(Y_{\Phi^X_t(1)}\right).$$

Plugging this back in and taking derivative in t completes the proof. ☐

3.2 Invariant Riemannian metrics and measures

Now we equip our Lie groups with Riemannian metrics that are compatible with the group structure. More extensive treatments can be found in [16, 20].

Definition 3.10. *A Riemannian metric g on a Lie group G called left-invariant if $(L_y)^* g = g$ for all $y \in G$, i.e.*

$$g(X \circ L_y, Y \circ L_y) = g(X, Y)$$

holds in each tangent space. If $(R_y)^* g = g$ for all y, it is called right-invariant, and bi-invariant if it is both left- and right-invariant.

Due to expressiveness of the Lie algebra such metrics are easily characterizes with bilinear forms on the tangent space.

Proposition 3.11. *There is a bijection between the two sets*

$$\{\textit{left-invariant Riemannian metrics } \langle \cdot, \cdot \rangle \textit{ on } G\},$$

and

$$\{\textit{inner products } B \textit{ on } \mathrm{Lie}(G)\}.$$

This correspondence restricts to a bijection between bi-invariant Riemannian metrics and Ad-invariant inner products B on $\mathrm{Lie}(G)$, i.e. $B(\mathrm{Ad}(x)X, \mathrm{Ad}(x)Y) = B(X, Y)$.

Proof. The evaluation of any Riemannian metric at the identity gives an inner product on the tangent space. Reversely, if an inner product on $\mathrm{Lie}(G)$ is given, we obtain a left-invariant Riemannian metric via the push-forward with L_x on any tangent space $T_x G$. Finally, a left-invariant Riemannian metric is bi-invariant if and only if it is right-invariant, i.e. if and only if $R_y^* L_y^* g = g$. But $R_y \circ L_y = \mathrm{Ad}(y^{-1})$ and the claim follows. $\square$

In particular, left-invariant measures always exist. Just pick any inner product on the underlying vector space of the Lie algebra.

Invariant metrics possess some nice properties. First note that if $X, Y \in \mathrm{Lie}(G)$ are left-invariant and g a left-invariant Riemannian metric, then $g(X, Y) = g_1(X_1, Y_1)$ is a constant function (the two initial vectors are pushed forward by the same diffeomorphism, so their relative orientation doesn't change). In that case, $\nabla_X Y$ is also left-invariant. Moreover, By the compatibility with the metric structure of the Levi-Civita connection, one obtains $g(\nabla_X Y, Z) + g(Y, \nabla_X Z) = 0$. Combining this with the symmetry we have

$$g(\nabla_X Y, Z) = \frac{1}{2}\big(g([X, Y], Z) - g([Y, Z], X) + g([Z, X], Y)\big). \tag{20}$$

To study what happens if we even require bi-invariance we first prove a simple characterization.

Lemma 3.12. *If $G \subseteq U_n$ is a closed, connected subgroup of the unitary group, then any left-invariant Riemannian metric is bi-invariant if and only if $\mathrm{ad}(X) \colon \mathrm{Lie}(G) \to \mathrm{Lie}(G)$ is skew-Hermitian for every $X \in \mathrm{Lie}(G)$.*

Proof. Let $x \in G$ be close enough to the identity so that there is a unique $X \in \mathrm{Lie}(G)$ with $x = \exp(X)$. We have by Proposition 3.7 that $\mathrm{Ad}(x) = \mathrm{Ad}(\exp(X)) = \exp(\mathrm{ad}(x))$. Since G is a group of unitary matrices, $\mathrm{Ad}(x)^{-1} = \overline{\mathrm{Ad}(x)}^T$ if and only if $\exp(-\mathrm{ad}(X)) = \exp(\overline{\mathrm{ad}(X)}^T)$ if and only $-\mathrm{ad}(X) = \overline{\mathrm{ad}(X)}^T$. To conclude we use the fact that any topological group G can be generated by a neighborhood of the identity (e.g. [13]). $\square$

Therefore, the last terms in Eq. (20) cancel and we obtain a very simple expression for the Levi-Civita connection:

$$\nabla_X Y = \frac{1}{2}[X, Y]. \tag{21}$$

The natural question to answer is: are there always bi-invariant metrics? While the answer is a no in general, the case for compact Lie groups is positive. The following characterization was proven in [20, Lemma 7.5].

Theorem 3.13 (Milnor (1976))**.** *A Lie group (connected!) admits a bi-invariant Riemannian metric if and only if it is isomorphic to $K \times \mathbb{R}^k$, for some compact Lie group K and some integer $k \geq 0$.*

Next up, we study uniqueness. We cannot expect much: abelian groups have zero restrictions and *any* inner product on the Lie algebra corresponds to a bi-invariant metrics. Recall that a Lie algebra is called simple if it has no non-trivial ideals.

Proposition 3.14. *If G is compact and $\mathrm{Lie}(G)$ is simple, then the bi-invariant Riemannian metric is unique up to positive scalar factor.*

Proof. If B is an Ad-invariant inner product, then any other inner product is of the form $B'(X, Y) = B(SX, Y)$ for some Hermitian matrix S. If the new metric is also bi-invariant, then $\mathrm{ad}(Z)$ is skew-Hermitian in both inner product spaces. As a consequence,

$$B(\mathrm{ad}(Z)X, SY) = -B(X, \mathrm{ad}(Z)SY)$$

and

$$B'(\mathrm{ad}(Z)X, Y) = -B'(X, \mathrm{ad}(Z)Y) = B(SX, \mathrm{ad}(Z)Y) = B(X, S\,\mathrm{ad}(Z)Y)$$

so S and $\mathrm{ad}(Z)$ commute. Therefore, if V is an eigenspace of S, then $[Z, V] = \mathrm{ad}(Z)V \subseteq V$: eigenspaces of S are ideals of the Lie algebra, which, in turn, was assumed to be simple. Thus, $B' = \lambda B$ for the unique eigenvalue. $\square$

Example 3.15. Let G be a connected Lie group with Lie algebra $L = \mathrm{Lie}(G)$. Define the symmetric bilinear form

$$\kappa(X, Y) := \mathrm{trace}(\mathrm{ad}(X) \circ \mathrm{ad}(Y)).$$

Note that the symmetry follows from the corresponding property of the trace. In the representation theory of Lie algebras This form is called the Killing form. In view of Proposition 3.11, this might be a somewhat canonical candidate for a Riemannian metric.

In fact, the non-degeneracy of this form is equivalent to the Lie algebra being semi-simple, i.e. L has no non-zero abelian ideals (see e.g. [14] for a standard reference). The form is clearly Ad-invariant. Moreover, one can prove that the Killing form is negative definite if the Lie algebra is even simple and not-abelian. In that case, by the previous uniqueness result, we have found the only bi-invariant Riemannian metric. But one can also prove, for instance, that $-\kappa$ is an inner product on L already if G has a compact universal covering group. See [20, Section 7] for more details.

Finally, we briefly discuss integration on Lie groups. Let $dV = dV_g$ be the volume element of G induced by a Riemannian volume form.

Theorem 3.16 (Specalized Haar's theorem)**.** *Let G be a Lie group with a left-invariant Riemannian metric g. Then, the volume element dV defines a left-invariant, non-zero Radon measure and is, as such, unique up to positive scalar factor.*

Note that the left-invariance ($\mathrm{Vol}(xE) = \mathrm{Vol}(E)$ for all Borel sets E) follows from the left-invariance of the Riemannian volume form. The uniqueness will not be proven here and the reader is referred to [13, Section 2.4]. In the context of topological groups this measure (left-invariant, non-zero Radon) is called the Haar measure. One can already prove existence and uniqueness on locally compact Hausdorff groups. The mentioned reference contains a detailed treatment of the subject. The last thing we are going to discuss is unimodularity, which is a property (of a group by uniqueness up to factor) asserting that any left-invariant Haar measure is also right-invariant. Abelian groups trivially possess this property.

Proposition 3.17. *If G is a compact Lie group, any left-invariant non-zero Radon measure is also right-invariant.*

Proof. Fix one Haar measure, here denoted by μ. Let $M\colon G \to \mathbb{R}_{>0}$ be the function defined as follows: Define another measure by $\mu_y(E) := \mu(Ey)$. Note that it is left-invariant, too, by associativity:

$$\mu_y(xE) = \mu((xE)y) = \mu(x(Ey)) = \mu_y(E).$$

Then, we let $M(y)$ be the unique positive number λ such that $\mu_y = \lambda\mu$. Clearly, G is unimodular if and only if $M \equiv 1$.

By construction, M is a homomorphism of groups: indeed, for any Borel set $E \subseteq G$

$$M(xy)\mu(E) = \mu(Exy) = M(y)\mu(Ex) = M(y)M(x)\mu(E),$$

so $M(xy) = M(x)M(y)$. Moreover, M can be seen to be continuous (see [13, Proposition 2.24]). Thus, the image of the compact group G under M must be a compact subgroup of $(\mathbb{R}, \cdot)$. The only candidate is the trivial subgroup (easy to see also if we consider the group isomorphism $\log\colon (\mathbb{R}_{>0}, \cdot) \to (\mathbb{R}, +)$). $\square$

3.3 Partial differential operators

In this section we study the universal enveloping algebra, which is a classical tool in representation theory. There, it is a convenient environment to work in as its representations coincide with those of the Lie algebra, but the enveloping algebra is associative, which is not the case in general for the Lie algebra (cf [14]). For our purposes, however, we will interpret it as the algebra of left-invariant partial differential operators, extending the first-order ones (i.e. vector fields). We will find the Laplacian as a very well-behaved second-order member.

Definition 3.18. *Let G be a Lie group and $L = \mathrm{Lie}(G)$ its Lie algebra. The universal enveloping algebra* $\mathrm{PDO}(G)$ *is the unital associative algebra obtained as the quotient of the tensor algebra*

$$\bigoplus_{m=0}^{\infty} (\underbrace{L \otimes \cdots \otimes L}_{m \text{ times}})$$

by the two-sided ideal generated by elements of the form

$$(X \otimes Y - Y \otimes X) - [X, Y].$$

The algebra operation in the quotient is denoted using concatenation.

The symbol PDO stems from the fact that its elements can be considered left-invariant partial differential operators. Indeed, by only imposing the algebraic rule $\mathrm{PDO}(G) \ni XY - YX = [X, Y] \in \mathrm{Lie}(G)$, we can seamlessly embed the left-invariant vector fields of G. The action of some monomial $D = X_1 \ldots X_k \in \mathrm{PDO}(G)$ on $f \in C^\infty$ is given by

$$Df(x) = \frac{d}{dt}\bigg|_{t=0} f\big(x \exp(tX_k) \ldots \exp(tX_1)\big). \tag{22}$$

For a general operator the action extends by linearity.

Theorem 3.19 (Universal property of $\mathrm{PDO}(G)$)**.** *Let $\iota\colon \mathrm{Lie}(G) \to \mathrm{PDO}(G)$ denote the canonical map. Then, any algebra homomorphism $\varphi\colon \mathrm{Lie}(G) \to A$ into some unital, associative, real algebra A factors over* $\mathrm{PDO}(G)$ *as $\varphi = \overline{\varphi} \circ \iota$ for a unique unital algebra homomorphism $\overline{\varphi}\colon \mathrm{PDO}(G) \to A$. This means that we have the following commutative diagram:*

$$\begin{tikzcd} \mathrm{Lie}(G) \arrow[rr, "\varphi"] \arrow[drr, "\iota"'] & & A \\ & & \mathrm{PDO}(G). \arrow[u, "\overline{\varphi}"'] \end{tikzcd}$$

As a unital associative $\mathbb{R}$-algebra, $\mathrm{PDO}(G)$ *is unique with this property.*

We skip the proof of the universal property and refer the reader to [22, Theorem 8.3.34].

Proposition 3.20. *If G is a matrix Lie group, then the canonical map $\iota\colon \mathrm{Lie}(G) \to \mathrm{PDO}(G)$ is injective. As a consequence, given a basis $\{X_j\}_j$ of $\mathrm{Lie}(G)$, the monomial operators $X_1 \dots X_k$ make up a basis of the underlying $\mathbb{R}$-vector space.*

Proof. Let $L = \mathrm{Lie}(G) \subseteq M_n(\mathbb{C}) = \mathrm{Lie}(\mathrm{GL}_n(\mathbb{C}))$ and $\varphi\colon L \hookrightarrow M_n(\mathbb{C})$ be the inclusion map. By the universal property, φ (which is trivially a homomorphism of algebras) factors over $\mathrm{PDO}(G)$: there exists another algebra homomorphism $\overline{\varphi}\colon \mathrm{PDO}(G) \to M_n(\mathbb{C})$ such that $\varphi = \overline{\varphi} \circ \iota$. Since φ is injective, the same must hold for ι. $\square$

The following element of $\mathrm{PDO}(G)$ is the protagonist of this paper.

Definition 3.21. *Let G be a matrix Lie group and B an* Ad*-invariant inner product on $L = \mathrm{Lie}(G)$. and let $X_1, \dots, X_n$ be a basis of L which is orthonormal w.r.t. B, and set $\hat{B}_{ij} \coloneqq B(X_i, X_j)$. Define another basis by*

$$X^i \coloneqq \sum_{j=1}^{n} \left(\hat{B}^{-1}\right)_{ij} X_j.$$

Then, the quadratic Casimir operator of G with respect to the Riemannian metric is defined as the element

$$\Omega(B) \coloneqq \sum_i X_i X^i \in \mathrm{PDO}(G).$$

Proposition 3.22. *The quadratic Casimir operator $\Omega(B)$ is independent of the choice of basis and and commutes with every $D \in \mathrm{PDO}(G)$. As a consequence, $\Omega(B)$ is both left- and right-invariant.*

Proof. Let $\{X_i\}_i$ be the basis of L that the quadratic Casimir operator was defined with and pick another basis $\{Y_i\}_i$. Then, there exists an invertible matrix T such that $Y_i = \sum_j T_{ij} X_j$. We compute the Gram matrix $\hat{C}$ of the bilinear form w.r.t. to that new basis:

$$\hat{C}_{ij} = B(Y_i, Y_j) = B\left(\sum_k T_{ik} X_k, \sum_l T_{jl} X_l\right) = \sum_{k,l} T_{ik} B(X_k, X_l) T_{jl}.$$

Thus, $\hat{C} = T\hat{B}T^T$. Let $\overline{\Omega}(B)$ denote the quadratic Casimir operator w.r.t. the new basis. We prove that it is equal to the $\Omega(B)$:

$$\begin{aligned}
\overline{\Omega}(B) &= \sum_{i,j} \left(\hat{C}^{-1}\right)_{ij} Y_i Y_j = \sum_{i,j} \left(\hat{C}^{-1}\right)_{ij} \sum_k T_{ik} X_k \sum_l T_{jl} X_l \\
&= \sum_{k,l} X_k X_l \sum_{i,j} T_{ik} \left(\hat{C}^{-1}\right)_{ij} T_{jl} \\
&= \sum_{k,l} X_k X_l \sum_{i,j} (T^T)_{ki} \left((T^T)^{-1} \hat{B}^{-1} T^{-1}\right)_{ij} T_{jl} = \Omega(B).
\end{aligned}$$

Next, let $D \in \mathrm{PDO}(G)$. We need to show $D\Omega(B) = \Omega(B)D$. By construction, $B(X^i, X_j) = \delta_{ij}$ and we can consider B an inner product on L via $\langle X_i, X_j \rangle_L \coloneqq B(X^i, X_j)$

for which $\{X_i\}_i$ is an ONB. For differential operators $D_1, D_2 \in \mathrm{PDO}(G)$, we set $[D_1, D_2] \coloneqq D_1D_2 - D_2D_1$, which satisfies the Jacobi identity. If for now we take $D \in L$,

$$[D, \Omega(B)] = \left[D, \sum_i X_i X^i\right] = \sum_i \left([D, X_i]X^i, X_i[D, X^i]\right).$$

We examine the terms one by one. We can pick constants c_{ij}, d_{ij} such that $[D, X_i] = \sum_j c_{ij}X_j$ and $[D, X^i] = \sum_j d_{ij}X^j$. Due to the fact that the bilinear form B is Ad-invariant, we can relate them. Indeed,

$$c_{ij} = \langle X_j, [D, X_i]\rangle_L = B(X^j, [D, X_i]) = B([X^j, D], X_i) = -B([D, X^j], X_i).$$

Therefore,

$$c_{ij} = -B\left(\sum_k d_{jk}X^k, X_i\right) = -\sum_k d_{jk}\langle X_k, X_i\rangle_L = -d_{ij}.$$

But then

$$[D, \Omega(B)] = \sum_{i,j}(c_{ij}X_jX^i + d_{ij}X_iX^j) = \sum_{i,j}(c_{ij} + d_{ij})X_jX^i = 0$$

and Ω commutes with first-order D. For higher-order monomials $D = X_1 \dots X_k$ observe that $[D, \Omega(B)] = X_1[X_2 \dots X_k, \Omega(B)] + [X_1, \Omega(B)]X_2 \dots X_k$ and proceed by induction. The general case follows from linearity.

Finally, we prove that actually any left-invariant differential operator Ω is right-invariant, if it commutes with every left-invariant $D \in \mathrm{PDO}(G)$. To prove that $R_y\Omega = \Omega R_y$, $y \in G$, as operators on $C^\infty(G)$. In view of Eq. (22), let $\Omega = X_1 \dots X_k \in \mathrm{PDO}(G)$ be monomial let $\exp(t\Omega)$ be short for $\exp(tX_k)\dots\exp(tX_1)$. we have that for $f \in C^\infty(G)$ and $x \in G$

$$f(xy^{-1}\exp(t\Omega)) = f(xy^{-1}y\exp(t\Omega)y^{-1}) = f(x\exp(ty\Omega y^{-1})y^{-1}),$$

where we inserted $1 = y^{-1}y$ before each $\exp(tX_i)$ and where we used the corresponding property of the exponential map in the last equality. After taking the derivative it follows that $\Omega R_y = R_y \mathrm{Ad}(y)\Omega$ with $\mathrm{Ad}(y)\Omega \coloneqq \mathrm{Ad}(y)(X_1)\dots\mathrm{Ad}(y)(X_k)$. However, since D commutes with all differential operators, it commutes with Y, chosen such that $\exp(Y) = y$, which means that $\mathrm{Ad}(y)\Omega = \Omega$. □

Proposition 3.23. *Let G be a matrix Lie group with bi-invariant metric g and associated Ad-invariant inner product B on* $\mathrm{Lie}(G)$*. Then, $\Omega(B)$ is identified with the Laplacian Δ w.r.t. the metric g.*

Proof. Recall from Section 2.1 that $\Delta f = \mathrm{div}(\mathrm{grad}\, f) = \mathrm{trace}(Y \mapsto \nabla_Y(\mathrm{grad}\, f))$ in analogy to the Euclidean case. Let X_i be an orthonormal basis of $\mathrm{Lie}(G)$ (that is, an orthonormal global frame) and let $f \in C^\infty(G)$. Now, by the discussions in Section 3.2

and the simple expression for the Levi-Civita connection in Eq. (21), we have

$$\begin{aligned}\Delta f &= \sum_i g(X_i, \nabla_{X_i}(\operatorname{grad} f)) \\ &= \sum_i X_i\, g(X_i, \operatorname{grad} f) - \sum_i g(\underbrace{\nabla_{X_i} X_i}_{[X_i, X_i]=0}, \operatorname{grad} f) \\ &= \sum_i X_i(X_i f),\end{aligned}$$

where we used the compatibility with the metric in the first line (note that $\operatorname{grad} f$ has no reason to be left-invariant, so that we cannot use the special cases). The claim follows when using the inclusion map. □

3.4 Unitary representations

Since we proof the main result, the Peter-Weyl theorem, in the next section, we need to review the basics of the theory of Lie (topological, really) group represention theory. The main concept will be that of a unitary representation. This exposition is based on [13, Chapter 3].

Definition 3.24. *A unitary representation of a Lie group G is a complex Hilbert space V together with a continuous group homomorphism $\xi\colon G \to U(V)$ into the unitary group of V equipped with the strong operator topology (i.e. all maps $x \mapsto \xi(x)v$ are continuous). The dimension of a representation is that of V.*

If the context allows, we may sometimes simply say that ξ is a representation without mentioning the space or that V is a representation without mentioning the map. While the above definition might seem not fit for the category of Lie groups, we will now see that it is the right one for the kinds of representations that will matter to us. In fact, one can already develop the theory on compact topological groups, and non of the results of this section is Lie group-specific expect this one:

Proposition 3.25. *If (V, ξ) is a finite-dimensional unitary representation, then ξ is smooth.*

Proof. The vector space V is equipped with the smooth structure of $V \cong \mathbb{C}^n \cong \mathbb{R}^{2n}$. Then, by Proposition 3.7 we have a commutative diagram

$$\begin{array}{ccc} G & \xrightarrow{\quad\xi\quad} & \mathrm{GL}(V) \\ \uparrow{\scriptstyle\exp} & & \uparrow{\scriptstyle\exp} \\ \mathrm{Lie}(G) \cong \mathbb{R}^{\dim G} & \xrightarrow{d\varphi} & \mathrm{End}(V) \cong M_{2n}(\mathbb{R}). \end{array}$$

It suffices to show differentiability locally around 1, since the left multiplication map is a diffeomorphism. Indeed, Proposition 3.6 stated that exp is a local diffeomorphism sending a neighborhood of 0 in $\mathrm{Lie}(G)$ to a neighborhood of 1 in G. Thus, we can reverse the left arrow (locally) and obtain a composed map of the form smooth-linear-smooth, and the claim follows. $\square$

Definition 3.26. *Let (V, ξ) and (W, η) be two unitary representations. A bounded linear operator $A\colon V \to W$ is called an intertwining operator if the following diagram commutes for all $x \in G$:*

$$\begin{tikzcd} V \arrow[r, "\xi(x)"] \arrow[d, "A"'] & V \arrow[d, "A"] \\ W \arrow[r, "\eta(x)"'] & W. \end{tikzcd}$$

The set of all such operators is denoted $\mathrm{Hom}_G(\xi, \eta)$. If there is a unitary operator in it, so that $\eta(x) = U\xi(x)U^*$, then ξ and η are called unitarily equivalent.

Definition 3.27. *Let (V, ξ) be a unitary representation of G. A closed Hilbert subspace $W \subseteq V$, is called ξ-invariant if $\xi(x)W \subseteq W$ for all $x \in G$. A unitary representation is called irreducible if it has no non-trivial invariant subspaces.*

If W is ξ-invariant the restriction $\xi|_W$ defines a representation on W. The subspaces $\{0\}$ and W make up the two trivial examples of invariant subspaces.

Definition 3.28. *Let $\{V_{\xi_i}\}_i$ be a family of unitary representations. Their direct sum $\bigoplus_i V_{\xi_i}$ is the unitary representations defined on the direct sum as Hilbert spaces as $\xi(x)(\sum_i v_i) \coloneqq \sum_i \xi_i(x)v_i$.*

Proposition 3.29. *Let $W \subseteq V$ be a ξ-invariant subspace. Then $W^\perp$ is invariant, too, and $V = W \oplus W^\perp$ as representations.*

Proof. Let $w \in W$, $w' \in W^\perp$. Then $\langle w, \xi(x)w' \rangle = \left\langle \xi(x^{-1})w, w' \right\rangle = 0$, since W is already invariant. Therefore, $\pi(x)w' \in W^\perp$. $\square$

As a direct consequence, every finite-dimensional unitary representation is the direct sum of irreducible representations. The following result is a simple characterization of invariant subspaces using the orthogonal projection.

Proposition 3.30. *Let (V, ξ) be a unitary representation. If $W \subseteq V$ is a closed Hilbert subspace, then it is ξ-invariant if and only if the orthogonal projection P_W is an intertwining operator $P_W \in \mathrm{Hom}_G(V, V)$.*

Proof. First, let P_W be intertwining. Pick $w \in W$. By the intertwining property, $\xi(x)w = \xi(x)P_W w = P_W \pi(x)w \in W$, and W is invariant. Conversely, if W is invariant and $w \in W$, then $\xi(x)P_W w = \xi(x)w = P\xi(x)w$. If $v \in W^\perp$, $\xi(x)P_W v = 0 = P_W \xi(x)v$. Since $V = W \oplus W^\perp$ the claim follows. $\square$

The following result is a fundamental tool in representation theory.

Theorem 3.31 (Schur's lemma)**.** *Let ξ be a unitary representation. It is irreducible if and only* $\mathrm{Hom}_G(\xi, \xi)$ *contains only scalar multiples of the identity. In that case if η is another irreducible unitary representation, then* $\mathrm{Hom}_G(\xi, \eta)$ *is one-dimensional if they are equivalent and trivial otherwise.*

Proof. First assume that ξ is not irreducible. Then, there exists a non-trivial ξ-invariant subspace and a non trivial projection in $\mathrm{Hom}_G(\xi, \xi)$, which cannot be a scalar multiple of the identity. Converselt, let $U \in \mathrm{Hom}_G(\xi, \xi)$ such that T is not a multiple of the identity. Then, both of the operators $\frac{1}{2}(T + T^*)$ and $\frac{1}{2i}(T - T^*)$ are intertwining and at least one of them is not a scalar multiple of I; call it T'. In any case, T' is a self-adjoint operator, so every operator that commutes with T' also commutes with all spectral indicator functions $\chi_E(T')$, $E \subset \mathbb{R}$ (by [13, Theorem 1.51 (c)]). In particular, they all commute with $\pi(x)$. Therefore, there must exist projection operators.

If now $T \in \mathrm{Hom}_G(\xi, \eta)$ then $T^* \in \mathrm{Hom}_G(\eta, \xi)$, since

$$T^*\eta(x) = (\eta(x)(x^{-1})T)^* = (T\xi(x)(x^{-1})T)^*$$

, Therfore, $TT^* \in \mathrm{Hom}_G(\xi, \xi)$ and $T^*T \in \mathrm{Hom}_G(\eta, \eta)$, and $TT^* = cI$, $T^*T = cI$. This implies either $T = 0$ or $c^{-1/2}T$ is unitary, and thus $\mathrm{Hom}(\xi, \eta) = \{0\}$ and that it consists of scalar multiples of unitary operators. If we take $T_1, T_2 \in \mathrm{Hom}_G(\xi, \xi)$, then $T_2^{-1}T_1 = T_2^*T_1 \in \mathrm{Hom}_G(\xi, \xi)$. Thus, $T_2^{-1}T_1 = cI$, and finally $T_1 = cT_2$ and $\dim(\mathrm{Hom}(\xi, \xi)) = 1$. □

The fact that a finite-dimensional representation decomposes as a direct sum of irreducible ones generalizes to infinite dimensions if we consider this paper's target class of compact groups.

Theorem 3.32 (Decomposition theorem)**.** *On a compact (Lie) group G every irreducible unitary representation is finite-dimensional (and in particular smooth). Every unitary representation is the direct sum of irreducible ones.*

Proof. Let (V, ξ) be a unitary representation. Fix a unit vector $u \in V$ and define the operator the averaged projection operator

$$Tv \coloneqq \int \langle v, \xi(x)u \rangle \, \xi(u) \, dx.$$

Then, T can be shown to be a positive, compact operator and $T \in \mathrm{Hom}_G(\xi, \xi)$ [13, Lemma 5.1]. Immediately, one sees that if ξ is irreducible, Schur's lemma (Theorem 3.31) implies that $T = cI$. But then the identity operator is compact and and the representation must be finite-dimensional.

Otherwise, we let ξ be an arbitrary unitary representation. Then, by the spectral theorem, T has a non-zero eigenvalue, whose eigenspace must of finite dimension. But T is intertwining and the eigenspace ξ-invariant. In conclusion, we have asserted existence of a finite-dimensional, ξ-invariant subspace which can be decomposed into irreducible ones iteratively using Proposition 3.29.

We conclude by Zorn's lemma, which makes sure that there is a maximal family $\mathcal{V}_i$ of of mutually orthogonal, irreducible, ξ-invariant subspaces. Now, the orthogonal complement of the direct sum $\bigoplus_i \mathcal{V}_i$ must be ξ-invariant itself. If it was non-trivial, it would nessecarily contain an irreducible subspace, which, in turn, would contradict the maximality of $\mathcal{V}_i$. Therefore, $V = \bigoplus_i \mathcal{V}_i$. □

3.5 Peter-Weyl theorem and spectrum of the Laplacian

Throughout, let G be a compact connected Lie group equipped with a bi-invariant Riemannian metric g, and let dy denote the associated normalized Haar measure established in Theorem 3.16 coming from the volume element $dV(y)$. Moreover, $L^2(G)$ is the Lebesgue space w.r.t. that measure. The following is based on [22, Chapters 7 and 8].

Definition 3.33. *The unitary dual $\widehat{G}$ of a compact (Lie) group G is the set of all equivalence classes of irreducible unitary representations under unitary equivalence.*

Elements of $\widehat{G}$ are denoted $[\xi]$ where ξ is an irreducible unitary representations acting as the representative.

Note that a finite-dimensional unitary representation (V, ξ), can be expressed in terms of matrices: Let $m = \dim(\xi)$ be the dimension of the representation and denote by $U_m, (\mathbb{C})$ the group of unitary matrices with complex coefficients. Choose an orthonormal basis of V. Then, $(\mathbb{C}^m, \widetilde{\xi})$ with $\widetilde{\xi}\colon G \to U_m(\mathbb{C})$, $\widetilde{\xi}_{ij}(x) := \langle e_i, \xi(x)e_j \rangle$, clearly defines a unitarily equivalent representation.

Definition 3.34. *A representation $\xi\colon G \to U_m(\mathbb{C})$, for some integer $m > 0$, is called a matrix representation. For brevity, we may say that an element $[\xi] \in \widehat{G}$ is matrix-represented if ξ is a matrix representative. The smooth functions $\xi_{ij}\colon G \to \mathbb{C}$ are called matrix coefficients.*

Lemma 3.35. *Let $[\xi], [\eta] \in \widehat{G}$ be both matrix-representated with $\dim \xi = m$ and $\dim \eta = n$. Then,*

$$\langle \xi_{ij}, \eta_{kl} \rangle_{L^2(G)} = \begin{cases} 0, & \text{if } [\xi] \neq [\eta], \\ \frac{1}{m}\delta_{ik}\delta_{jl}, & \text{if } \xi = \eta. \end{cases}$$

Proof. Fix $j = 1, \ldots, m$ and $l = 1, \ldots, n$ and let E be the matrix that is zero everywhere except for a one at (j, l), i.e. $E_{pq} := \delta_{pj}\delta_{lq}$. For the representation spaces V_ξ and V_η of ξ and η respectively, we define the operator $A\colon V_\xi \to V_\eta$ by the element-wise integral

$$A := \int_G \xi(y)E\eta(y^{-1})\,dy.$$

By writing out the matrix multiplication and using the fact that representations are unitary we find

$$A_{ik} = \int_G \sum_{p=1}^{m}\sum_{q=1}^{n} \xi_{ip}(y)E_{pq}\zeta_{qk}(y^{-1})\,dy = \int_G \xi_{ij}(y)\overline{\zeta_{kl}(y)}\,dy = \langle \xi_{ij}, \zeta_{kl} \rangle_{L^2(G)}\,.$$

This operator is intertwining: indeed, using a substitution $z = xy$

$$\xi(x)A = \int_G \xi(xy)E\eta(y^{-1})\, dy = \int_G \xi(z)E\eta(z^{-1}x)\, dz = A\eta(x).$$

Thus, $A \in \mathrm{Hom}_G(\xi, \eta)$. But the representations are both irreducible, and Schur's lemma (Theorem 3.31) implies in this case that $A = \lambda I$, $\lambda \in \mathbb{C}$, if $[\xi] = [\eta]$ and $A = 0$ otherwise.

The only thing left to do is to compute the inner product in case $[\xi] = [\eta]$. Clearly $m = n$ and there is a unitary intertwining operator $U\colon V_\xi \to V_\eta$ so that $\eta(x) = U\xi(x)U^*$. Then,

$$\begin{aligned}
\langle \xi_{kj}, \eta_{kl} \rangle_{L^2(G)} &= A_{kk} = \lambda = \frac{1}{m}\,\mathrm{trace}(A) \\
&= \frac{1}{m}\int_G \mathrm{trace}(\xi(y)\,E\,\eta(y^{-1}))\, dy \\
&= \frac{1}{m}\int_G \mathrm{trace}(\xi(y)\,E\,U\xi(y^{-1})U^*)\, dy \\
&= \frac{1}{m}\int_G \mathrm{trace}(\xi(y)\,E\,\xi(y^{-1}))\, dy \\
&= \frac{1}{m}\int_G \mathrm{trace}(E)\, dy = \frac{1}{m}\delta jl,
\end{aligned}$$

where we made use of the commutativity of matrices when computing the trace of their product. ☐

Definition 3.36. *The left regular representation π_L and the right regular representation π_R are the unitary representations $\pi_L, \pi_R\colon G \to U(L^2(G))$ defined by*

$$\begin{aligned}
(\pi_L(y)f)(x) &:= (L_y f)(x) = f(yx), \\
(\pi_R(y)f)(x) &:= (R_y f)(x) = f(xy^{-1}).
\end{aligned}$$

They are in fact unitary representations: observe that

$$\langle f, \pi_R(y)g \rangle_{L^2(G)} = \int f(x)\overline{g(xy^{-1})}\, dx = \int f(zy)\overline{g(z)}\, dz = \left\langle \pi_R(y^{-1})f, g \right\rangle_{L^2(G)},$$

where we used the change of variables $z = xy^{-1}$. Therefore, $\pi_R(y)^* = \pi_R(y^{-1}) = \pi_R(y)^{-1}$. A very important result for us is the decomposition of these two representations:

Theorem 3.37 (Peter-Weyl theorem)**.** *Let G be a compact Lie group. The set*

$$\left\{ \sqrt{\dim(\xi)}\,\xi_{ij} \,\middle|\, [\xi] \in \widehat{G} \text{ matrix-represented}, 1 \le i, j \le \dim(\xi) \right\} \subset C^\infty(G) \tag{23}$$

forms an orthonormal basis of $L^2(G)$. For a given matrix-represented $[\xi] \in \widehat{G}$, define the subspaces

$$\mathcal{H}^\xi_{\cdot,j} := \mathrm{span}\{\xi_{ij} \mid 1 \le i \le \dim(\xi)\} \quad \text{and} \quad \mathcal{H}^\xi_{i,\cdot} := \mathrm{span}\{\xi_{ij} \mid 1 \le j \le \dim(\xi)\}.$$

Then, each "row space" $\mathcal{H}^{\xi}_{\cdot,j}$ is π_L-invariant and each "column space" $\mathcal{H}^{\xi}_{i,\cdot}$ is π_R-invariant and ξ is equivalent to the restricition of π_L onto $\mathcal{H}^{\xi}_{\cdot,j}$ and π_R onto $\mathcal{H}^{\xi}_{i,\cdot}$, respectively. As a consequence, $L^2(G)$ decomposes both as the left regular and as the right regular representation

$$\bigoplus_{[\xi]\in\widehat{G}} \bigoplus_{j=1}^{\dim(\xi)} \mathcal{H}^{\xi}_{\cdot,j} = L^2(G) = \bigoplus_{[\xi]\in\widehat{G}} \bigoplus_{i=1}^{\dim(\xi)} \mathcal{H}^{\xi}_{i,\cdot}.$$

Proof. It will be enough to prove the theorem for the left regular representation π_L. The right regular case is analogous.

First note that

$$\pi_L(y)\xi_{ij}(x) = \xi_{ij}(yx) = (\xi(y)\xi(x))_{ij} = \sum_{k=1}^{\dim(\xi)} \xi_{ik}(y)\xi_{kj}(x).$$

This proves that for each y, $\pi_L(y)\xi_{ij}$ is a linear combination of the $\{\xi_{kj}\}_k$ and therefore, $\mathcal{H}^{\xi}_{\cdot,j}$ is π_L-invariant. Moreover, letting $\{e_i\}_i$ be the standard basis of $\mathbb{C}^{\dim(\xi)}$, we can define an operator

$$A\left(\sum_i \lambda_i e_i\right) := \sum_i \lambda_i \xi_{ij},$$

which intertwines between the matrix representation ξ and the restricted left regular representation $\pi_L|_{\mathcal{H}^{\xi}_{\cdot,j}}$. Indeed, for every $x, y \in G$ and $k = 1, \dots, \dim(\xi)$

$$\begin{aligned}((A\xi(x))e_k)(y) &= A\left(\sum_i \xi_{ki}(x)e_i\right)(y) = \sum_i \xi_{ki}(x)\xi_{ij}(y) \\ &= (\xi(x)\xi(y))_{kj} = \xi_{kj}(xy) = ((\pi_L(x)A)e_k)(y).\end{aligned}$$

We prove that the matrix coefficients form an orthonormal basis of $L^2(G)$. By Lemma 3.35 we already know that they are orthogonal. Let us assume that the closed span $\mathcal{H}$ of the orthonormal set in line 23 is not $L^2(G)$. As we have already shown, $\mathcal{H}$ must be π_L-invariant, and so is $\mathcal{H}^\perp$ by Proposition 3.29. Since the latter is a direct sum of irreducible unitary representations by the Decomposition theorem (Theorem 3.32), there exists a non-trivial π_L-invariant closed subspace $E \subset \mathcal{H}^\perp$ with a matrix representation $\xi \sim \pi_L|_E$. Let $\{f_i\}_{i=1}^{\dim(\xi)} \subset E$ be an orthonormal basis such that for any $y \in G$

$$\pi_L(y)f_i = \sum_j \xi_{ij}(y)f_j(x).$$

Since these live in a Lebesgue space we can have point-wise equality

$$f_i(yx) = \sum_j \xi_{ij}(y)f_j(x)$$

only for almost every $x \in G$. Thus, we define three sets

- $N(y)$ is the set of all $x \in G$ such that Eq. (23) does not hold,
- $M(x)$ is the set of all $y \in G$ such that Eq. (23) does not hold,
- and K is the set of all $(x, y) \in G \times G$ such that Eq. (23) does not hold.

Since these sets are clearly measurable (f_i and ξ_{ij} are all measurable), and the domain of integration is compact so that the measure of K is finite, we may apply Fubini's theorem:

$$\int_G \int_{M(x)} 1\, dy\, dx = \int_K 1\, dy \otimes dx = \int_G \underbrace{\int_{N(y)} 1\, dx}_{=0}\, dy = 0.$$

We have proven that for almost every $x \in G$, the set $M(x)$ has measure zero. Pick one such x_0 so that Section 3.5 holds for x_0 and almost every $y \in G$. Using $z := yx_0$ we can rewrite

$$f_i(z) = \sum_j \xi_{ij}(zx_0^{-1}) f_j(x_0) = \sum_j \sum_k \xi_{ik}(z)\xi_{kj}(x_0^{-1}) f_j(x_0) = \sum_k \xi_{ik}(z) \sum_j \xi_{kj}(x_0^{-1}) f_j(x_0).$$

Therefore, $f_i \in \operatorname{span}\{\xi_i k\}_k \subset \mathcal{H}$, but since by assumption $f_i \in \mathcal{H}^\perp$, it must be that $f_i = 0$. The index was arbitrary, so that $E = 0$. A contradiction. □

The following theorem breaks down the question of finding the eigenstructure of the Laplace-Beltrami operator in the sense of the Sturm-Liouville decomposition (Theorem 2.1) to finding all (or sufficiently many in applied contexts) irreducible representations of the group. Since these are known in many cases it also allows to explicitly write down the the reproducing kernel of Sobolev spaces (see Theorem 2.19).

Theorem 3.38. *Let G be a compact Lie group and let $[\xi] \in \widehat{G}$ be matrix-represented. The spaces*

$$\mathcal{H}^\xi := \bigoplus_{j=1}^{\dim(\xi)} \mathcal{H}^\xi_{\cdot,j} = \bigoplus_{i=1}^{\dim(\xi)} \mathcal{H}^\xi_{i,\cdot}$$

from the Peter-Weyl theorem (Theorem 3.37) are exactly the eigenspaces of the Laplace-Beltrami operator Δ and $\Delta|_{\mathcal{H}^\xi} = \lambda_\xi I$ for some $\lambda_\xi < 0$.

Note that the eigenspaces do not depend on the chosen Riemannian metric. The dependence is only expressed through the eigenvalues.

Proof. By Proposition 3.22 the Laplace-Beltrami operator is bi-invariant, so that it commutes with both $\pi_L(x)$ and $\pi_R(x)$ for any $x \in G$. Thus, it commutes with every irreducible unitary matrix representation ξ of $L^2(G)$ by the Peter-Weyl theorem (Theorem 3.37). In particular, it sends all $\mathcal{H}^\xi_{\cdot,j}$ and $\mathcal{H}^\xi_{i,\cdot}$ into themselves. Therefore, $\Delta\xi_{ij}$ lies in the intersection of row and column space which is simply the linear span of ξ_{ij}: $\Delta\xi_{ij} = \lambda_{ij}\xi_{ij}$.

We claim that all λ_{ij} are equal. We have

$$(\Delta\pi_L(y)\xi_{ij})(x) = \Delta(x \mapsto \xi_{ij}(yx))(x) = \Delta\left(x \mapsto \sum_k \xi_{ik}(y)\xi_{kj}(x)\right)(x) = \sum_k \lambda_{ik}\xi_{ik}(y)\xi_{kj}(x)$$

on the one hand and

$$(\pi_L(y)\Delta\xi_{ij})(x) = \lambda_{ij}\xi_{ik}(yx) = \sum_k \lambda_{ij}\xi_{ik}(y)\xi_{kj}(x)$$

on the other. By linear independence of the matrix coefficients, it follows that $\lambda_{ik}\xi_{kj}(y) = \lambda_{ij}\xi_{kj}(y)$ and thus $\lambda_{ik} = \lambda_{ij}$ for all i, j, k, i.e. any row of the matrix λ_{ij} is filled with only one value. A similar argument with the left regular representation shows that $\lambda_{kj} = \lambda_{ij}$, i.e. columns are all filled with the same value. The claim is proven.

Finally, $\lambda := \lambda_{ij}$ must be negative since the Laplace-Beltrami is negative definite by the Sturm-Liouville decomposition (Theorem 2.1). $\square$

References

[1] Thierry Aubin. *Some Nonlinear Problems in Riemannian Geometry.* Springer Monographs in Mathematics. Springer-Verlag, Berlin, 1998, pp. xviii+395. ISBN: 3-540-60752-8. DOI: 10.1007/978-3-662-13006-3. URL: https://doi.org/10.1007/978-3-662-13006-3.

[2] Peter L. Bartlett et al. "Benign Overfitting in Linear Regression". In: *Proceedings of the National Academy of Sciences* 117.48 (Dec. 2020), pp. 30063–30070. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1907378117. arXiv: 1906.11300 [cs, math, stat]. URL: http://arxiv.org/abs/1906.11300 (visited on 11/02/2023).

[3] Brian Lee Beers and Richard S. Millman. "The Spectra of the Laplace-Beltrami Operator on Compact, Semisimple Lie Groups". In: *American Journal of Mathematics* 99.4 (1977), pp. 801–807. ISSN: 0002-9327. DOI: 10.2307/2373866. JSTOR: 2373866. URL: https://www.jstor.org/stable/2373866 (visited on 11/02/2023).

[4] Mikhail Belkin, Siyuan Ma, and Soumik Mandal. *To Understand Deep Learning We Need to Understand Kernel Learning.* June 14, 2018. DOI: 10.48550/arXiv.1802.01396. arXiv: 1802.01396 [cs, stat]. URL: http://arxiv.org/abs/1802.01396 (visited on 11/02/2023). preprint.

[5] Mikhail Belkin et al. "Reconciling Modern Machine Learning Practice and the Bias-Variance Trade-Off". In: *Proceedings of the National Academy of Sciences* 116.32 (Aug. 6, 2019), pp. 15849–15854. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1903070116. arXiv: 1812.11118 [cs, stat]. URL: http://arxiv.org/abs/1812.11118 (visited on 11/02/2023).

[6] Pierre H. Bérard. *Spectral Geometry: Direct and Inverse Problems.* Vol. 1207. Lecture Notes in Mathematics. Springer-Verlag, Berlin, 1986, pp. xiv+272. ISBN: 3-540-16788-9. DOI: 10.1007/BFb0076330. URL: https://doi.org/10.1007/BFb0076330.

[7] Petre Birtea, Ioan Casu, and Dan Comanescu. *Laplace-Beltrami Operator on the Orthogonal Group in Cartesian Coordinates*. Oct. 20, 2023. DOI: 10.48550/arXiv.2206.06790. arXiv: 2206.06790 [math-ph]. URL: http://arxiv.org/abs/2206.06790 (visited on 11/02/2023). preprint.

[8] Viacheslav Borovitskiy et al. *Matérn Gaussian Processes on Riemannian Manifolds*. Apr. 17, 2023. DOI: 10.48550/arXiv.2006.10160. arXiv: 2006.10160 [cs, stat]. URL: http://arxiv.org/abs/2006.10160 (visited on 09/22/2023). preprint.

[9] Fida El Chami. "Spectre du laplacien sur les formes versus spectre des volumes : le cas des grassmanniennes". PhD thesis. Université Paris-Sud, Oct. 17, 2000. URL: https://theses.hal.science/tel-04200996 (visited on 11/02/2023).

[10] Thierry Coquand and Bas Spitters. "A Constructive Proof of the Peter-Weyl Theorem". In: *Mathematical Logic Quarterly* 51.4 (2005), pp. 351–359. ISSN: 1521-3870. DOI: 10.1002/malq.200410037. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/malq.200410037 (visited on 11/02/2023).

[11] Ernesto De Vito, Nicole Mücke, and Lorenzo Rosasco. *Reproducing Kernel Hilbert Spaces on Manifolds: Sobolev and Diffusion Spaces*. May 26, 2019. DOI: 10.48550/arXiv.1905.10913. arXiv: 1905.10913 [cs, math, stat]. URL: http://arxiv.org/abs/1905.10913 (visited on 09/22/2023). preprint.

[12] Aasa Feragen, Francois Lauze, and Søren Hauberg. *Geodesic Exponential Kernels: When Curvature and Linearity Conflict*. Nov. 17, 2014. DOI: 10.48550/arXiv.1411.0296. arXiv: 1411.0296 [cs]. URL: http://arxiv.org/abs/1411.0296 (visited on 09/22/2023). preprint.

[13] Gerald B. Folland. *A Course in Abstract Harmonic Analysis*. 2nd ed. Textbooks in Mathematics. CRC Press, Boca Raton, FL, 2016, xiii+305 pp.+loose errata. ISBN: 978-1-4987-2713-6.

[14] James E. Humphreys. *Introduction to Lie Algebras and Representation Theory*. Vol. 9. Graduate Texts in Mathematics. Springer-Verlag, New York-Berlin, 1972, pp. xii+169.

[15] Arthur Jacot, Franck Gabriel, and Clément Hongler. *Neural Tangent Kernel: Convergence and Generalization in Neural Networks*. Feb. 10, 2020. DOI: 10.48550/arXiv.1806.07572. arXiv: 1806.07572 [cs, math, stat]. URL: http://arxiv.org/abs/1806.07572 (visited on 11/02/2023). preprint.

[16] John M. Lee. *Introduction to Riemannian Manifolds*. Vol. 176. Graduate Texts in Mathematics. Springer, Cham, 2018, pp. xiii+437. ISBN: 978-3-319-91754-2 978-3-319-91755-9.

[17] John M. Lee. *Introduction to Smooth Manifolds*. 2nd ed. Vol. 218. Graduate Texts in Mathematics. Springer, New York, 2013, pp. xvi+708. ISBN: 978-1-4419-9981-8.

[18] Tengyuan Liang and Alexander Rakhlin. "Just Interpolate: Kernel "Ridgeless" Regression Can Generalize". In: *The Annals of Statistics* 48.3 (June 1, 2020). ISSN: 0090-5364. DOI: 10.1214/19-AOS1849. arXiv: 1808.00387 [cs, math, stat]. URL: http://arxiv.org/abs/1808.00387 (visited on 11/02/2023).

[19] Anton Mallasto and Aasa Feragen. “Wrapped Gaussian Process Regression on Riemannian Manifolds”. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. June 2018, pp. 5580–5588. DOI: 10.1109/CVPR.2018.00585.

[20] John Milnor. “Curvatures of Left Invariant Metrics on Lie Groups”. In: *Advances in Math.* 21.3 (1976), pp. 293–329. ISSN: 0001-8708. DOI: 10.1016/S0001-8708(76)80002-3. URL: https://doi.org/10.1016/S0001-8708(76)80002-3.

[21] Emilio Porcu et al. *The Matérn Model: A Journey through Statistics, Numerical Analysis and Machine Learning*. Mar. 5, 2023. DOI: 10.48550/arXiv.2303.02759. arXiv: 2303.02759 [math, stat]. URL: http://arxiv.org/abs/2303.02759 (visited on 09/22/2023). preprint.

[22] Michael Ruzhansky and Ville Turunen. *Pseudo-Differential Operators and Symmetries*. Vol. 2. Pseudo-Differential Operators. Theory and Applications. Birkhäuser Verlag, Basel, 2010, pp. xiv+709. ISBN: 978-3-7643-8513-2. DOI: 10.1007/978-3-7643-8514-9. URL: https://doi.org/10.1007/978-3-7643-8514-9.

[23] Henrik Schlichtkrull, Peter Trapa, and Jr Vogan. *Laplacians on Spheres*. July 20, 2018. DOI: 10.48550/arXiv.1803.01267. arXiv: 1803.01267 [math]. URL: http://arxiv.org/abs/1803.01267 (visited on 11/02/2023). preprint.

[24] Ingo Steinwart and Andreas Christmann. *Support Vector Machines*. Information Science and Statistics. Springer, New York, 2008, pp. xvi+601. ISBN: 978-0-387-77241-7.

[25] Robert S Strichartz. “Analysis of the Laplacian on the Complete Riemannian Manifold”. In: *Journal of Functional Analysis* 52.1 (June 1, 1983), pp. 48–79. ISSN: 0022-1236. DOI: 10.1016/0022-1236(83)90090-3. URL: https://www.sciencedirect.com/science/article/pii/0022123683900903 (visited on 09/22/2023).

[26] Michael E. Taylor. *Partial Differential Equations I. Basic Theory*. 2nd ed. Vol. 115. Applied Mathematical Sciences. Springer, New York, 2011, pp. xxii+654. ISBN: 978-1-4419-7054-1. DOI: 10.1007/978-1-4419-7055-8. URL: https://doi.org/10.1007/978-1-4419-7055-8.

[27] Hans Triebel. *Theory of Function Spaces. II*. Vol. 84. Monographs in Mathematics. Birkhäuser Verlag, Basel, 1992, pp. viii+370. ISBN: 3-7643-2639-5. DOI: 10.1007/978-3-0346-0419-2. URL: https://doi.org/10.1007/978-3-0346-0419-2.

[28] Luca Wellmeier. “High-Capacity Hypothesis Spaces in Modern Statistical Learning”. MA thesis. University of Padua, Apr. 2023.

[29] Holger Wendland. *Scattered Data Approximation*. Vol. 17. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, Cambridge, 2005, pp. x+336. ISBN: 978-0-521-84335-5 0-521-84335-9.