University of Amsterdam

MSc Artificial Intelligence

Master Thesis

A Wigner-Eckart Theorem for Steerable Kernels of General Compact Groups

by

Leon Lang

12383201

July, 2020

48 Credit Points
Research carried out from November 2019 to July 2020

Supervisor/Examiner:
Maurice Weiler
Dr. Patrick Forré

Assessor:
Dr.ir. Erik J. Bekkers

Acknowledgments

First and foremost, I thank my supervisor Maurice Weiler for providing me with the ideas that sparked my investigations into the connections between deep learning, physics, and representation theory that underlie this thesis. I was glad to have his constant encouragement and feedback and also large freedom in the pursuit of this research. I also thank my brother Lucas for his patience in explaining to me the Wigner-Eckart Theorem from the perspective of a quantum chemist and for discussions related to the meaning of fields and spherical tensor operators in a physical context. Additionally, I thank Patrick Forré for useful discussions on the link between steerable kernels and representation operators and Gabriele Cesa for discussions on the connection between real and complex representations of compact groups. My thanks also go to Tom Lieberum who gave feedback on the introduction and on a talk about early results of this thesis. Additionally, I thank Stefan Dawydiak and Terrence Tao for online discussions on aspects surrounding a real version of the Peter-Weyl Theorem and Rupert McCallum for discussions on different aspects of the mathematical ideas behind this thesis.

Summary

Equivariant neural networks recently emerged as a principled way to do deep learning when symmetries of the prediction task are known in advance. An important class of equivariant networks are steerable CNNs. Convolutions in this setting use a steerable kernel which guarantees that the output features transform predictably when the input features transform under the symmetry group. These steerable kernels show a remarkable similarity to representation operators — generalizations of spherical tensor operators — which are central to quantum mechanics. Such representation operators can be described concisely using the Wigner-Eckart Theorem. By extending the kernel linearly to the space of square-integrable functions on a homogeneous space, we get a precise link between steerable kernels and representation operators. This allows us to prove a Wigner-Eckart Theorem for steerable kernels of general compact groups which also completely covers the kernel theory of gauge equivariant CNNs whenever their so-called structure group is compact. Consequently, in the compact case, we obtain a general description of how to parameterize steerable and gauge equivariant CNNs. In our result, steerable kernel bases are expressed using endomorphisms of irreducible representations, Clebsch-Gordan coefficients, and harmonic basis functions on a homogeneous space. We discuss the symmetry groups $\mathrm{U}(1)$, $\mathrm{SO}(2)$, $\mathbb{Z}_2$, $\mathrm{SO}(3)$ and $\mathrm{O}(3)$ and derive concrete steerable kernel bases between arbitrary irreducible input and output fields. By a thorough investigation, we show that the kernel bases are consistent with prior results obtained for these symmetry groups, in cases where they have been described before. While we only derive concrete kernel bases for groups that are relevant in image processing, we note that our work applies just as well to groups like $\mathrm{SU}(2)$ or $\mathrm{SU}(3)$ that appear in physics. We hope that this new link between the theory of equivariant deep learning and quantum mechanics will lead to fruitful collaborations between physicists and chemists on the one hand and deep learning researchers on the other hand.

Contents

List of Symbols

General Set Theory and Functions

$A \cap B$	intersection of sets A and B
$A \cup B$	union of sets A and B
$\bigcap_{i \in I} A_i$	intersection of sets A_i
$\bigcup_{i \in I} A_i$	union of sets A_i
$\bigsqcup_{i \in I} A_i$	union of sets A_i which are disjoint from each other
$A \subseteq B$	A is a subset of B
$A \subsetneq B$	A is a strict subset of B
$A \setminus B$	set of all elements in A which are not in B
$A \times B$	Cartesian product of sets or structures (e.g. groups) A, B
$\emptyset$	empty set
$X \coloneqq Y$	X is defined as Y
$\sim$	often an equivalence relation
$[x]$	equivalence class with respect to an equivalence relation
$\mathbf{1}_A$	indicator function of set A
$f \circ g$	composition of two composable functions f and g
f^{-1}	either the inverse of function f or the preimage function
$f\|_A$	restriction of a function f to a subset A

Numbers and Collections of Numbers

$\mathbb{N}$	natural numbers including 0
$\mathbb{Z}$	integers
$\mathbb{R}$	field of real numbers
$\mathbb{C}$	field of complex numbers
$\mathbb{H}$	skew-field of quaternions
$\mathbb{K}$	one of the two fields $\mathbb{R}$ and $\mathbb{C}$
$\mathbb{K}^n$	n-dimensional canonical vector space over $\mathbb{K}$
$\overline{x}$	complex conjugate of x

Groups

G	a compact topological group
$1, e$	neutral element of a group with multiplication as operation
0	neutral element of an additive group
$G \rtimes H$	semidirect product of two groups G and H
C_N	group of planar rotations of a regular N-gon
D_N	group of planar rotations and reflections of a regular N-gon
$\mathrm{SO}(n)$	special orthogonal group in n real dimensions
$\mathrm{O}(n)$	orthogonal group in n real dimensions
$\mathrm{O}(V)$	orthogonal group of a *real* Hilbert space V
$\mathrm{SU}(n)$	special unitary group in n complex dimensions
$\mathrm{U}(n)$	unitary group in n complex dimensions
$\mathrm{U}(V)$	unitary group of a *complex* Hilbert space V
$\mathrm{E}(n)$	Euclidean motion group in n dimensions

Basic Representation Theory

ρ	a linear representation of a group
ρ^v	The function $G \to V$, $g \mapsto \rho(g)(v)$
ρ^{uv}	matrix coefficient of the unitary representation ρ
$\rho^{\mathrm{in}}, \rho^{\mathrm{out}}$	representations of the in-field and out-field, respectively
ρ_{Hom}	Hom-representation on $\mathrm{Hom}_{\mathbb{K}}(V, V')$ of representations ρ and ρ'
$\rho \otimes \rho'$	tensor product representation on $V \otimes V'$ of representations ρ and ρ'
$\mathrm{Ind}_G^H \rho$	induced representation on H or a representation ρ on G
$\hat{G}$	set of isomorphism classes of unitary representations on G
l	an isomorphism class of unitary representations
ρ_l	a representative of isomorphism class l
V_l	vector space on which ρ_l acts
v_l^i or Y_l^n	fixed chosen orthonormal basis vector of V_l

Vector Spaces and Hilbert Spaces

$\dim(V)$	dimension of $\mathbb{K}$-vectorspace V
$V \perp W$	V and W are perpendicular
$V \cong W$	V and W are isomorphic with respect to their structures

$V \ncong W$	V and W are not isomorphic with respect to their structures
$\langle f \vert g \rangle$	bra-ket notation of a scalar product on a Hilbert space
$\langle y \vert f \vert x \rangle$	equivalent to $\langle y \vert f(x) \rangle$ for a function f
$\mathrm{null}(f)$	null space of f
$\mathrm{im}(f)$	image of f
f^*	adjoint of the operator f
id_V	identity function on V

(Hilbert) Space Constructions from Other Spaces

$\mathrm{Hom}_{\mathbb{K}}(V, W)$	space of $\mathbb{K}$-linear functions from V to W
$\mathrm{Aut}_{\mathbb{K}}(V)$	space of invertible $\mathbb{K}$-linear functions from V to itself, sometimes written $\mathrm{GL}(V, \mathbb{K})$ in the literature
$\mathrm{Hom}_{G,\mathbb{K}}(V, W)$	space of intertwiners from V to W
$\mathrm{Hom}_G(X, W)$	space of G-equivariant continuous maps from X to W, for a homogeneous space X
$\mathrm{End}_{G,\mathbb{K}}(V)$	space of endomorphisms of V, i.e. intertwiners from V to V
$V \otimes W$	tensor product of two vector spaces over their common field. Also denotes the tensor product of pre-Hilbert spaces
$\bigoplus_{i \in I} V_i$	(orthogonal) direct sum of all V_i
$\widehat{\bigoplus}_{i \in I} V_i$	topological closure of the (orthogonal) direct sum of all V_i
$\mathrm{span}_{\mathbb{K}}(M)$	vector subspace of a $\mathbb{K}$-vector space spanned by M
$V^\perp$	orthogonal complement of V
$E_\lambda(\varphi)$	eigenspace of φ for eigenvalue λ

Topological Spaces, Metric Spaces, Normed Spaces

$\mathcal{T}$	topology
U_x	open neighborhood of $x \in X$
$\mathcal{U}_x$	set of all open neighborhoods of $x \in X$
$\lim_{U \in \mathcal{U}_x}$	limit over the directed set of open neighborhoods of x
$\lim_{k \to \infty} x_k$	limit of the sequence $(x_k)_k$
$\overline{A}$	topological closure of $A \subseteq X$
$\lVert x \rVert$	norm of x
$\lvert x \rvert$	absolute value of x
$d(x, x')$	distance of x, x' according to metric d
$\mathrm{B}_\epsilon(x)$	ϵ-ball around x according to some metric d

Homogeneous Spaces and the Peter-Weyl Theorem

X	a homogeneous space of G
$x^* \in X$	arbitrary point
S^n	n-dimensional sphere in $(n+1)$-dimensional space
μ	a measure on a compact group G or its Homogeneous Space X
$\int_X$	integral on a space X with respect to its measure
$L^2_{\mathbb{K}}(X), L^2_{\mathbb{K}}(G)$	Hilbert space of square-integrable functions on X and G with values in $\mathbb{K}$
λ	unitary representation on $L^2_{\mathbb{K}}(X)$ or $L^2_{\mathbb{K}}(G)$
$g(x)$	arbitrary lift of x with respect to projection $\pi : G \to X,\ g \mapsto gx^*$
$\mathrm{av}(f)$	average of $f : G \to \mathbb{K}$ along cosets
π^*	lift of functions $L^2_{\mathbb{K}}(X) \to L^2_{\mathbb{K}}(G)$
δ_x	Dirac delta function at point x
δ_U	approximated Dirac delta function for nonempty open set U
ρ_l^{ij}	abbreviation for $\rho_l^{v^i v^j}$ for orthonormal basis vectors $v^i, v^j \in V_l$
$\mathcal{E}$	linear span of all matrix coefficients of irreducible unitary representations
$\mathcal{E}_l$	linear span of all matrix coefficients of ρ_l
$\mathcal{E}_l^j$	linear span of all matrix coefficients ρ_l^{ij} with varying i but fixed j
n_l, m_l	multiplicity of l in orthogonal decomposition of $L^2_{\mathbb{K}}(G)$ and $L^2_{\mathbb{K}}(X)$, respectively
V_{li}	copy of V_l appearing in the Peter-Weyl decomposition of $L^2_{\mathbb{K}}(X)$
p_{li}	canonical projection $p_{li} : L^2_{\mathbb{K}}(X) \to V_{li}$ and $p_{li} : \bigoplus_{l'i'} V_{l'i'} \to V_{li}$
$\sin_m, \cos_m$	the functions $x \mapsto \sin(mx)$ and $x \mapsto \cos(mx)$
$Y_l^n, {}^rY_l^n$	complex- and real-valued version of a spherical harmonic
$D_l, {}^rD_l$	complex- and real-valued version of Wigner D-matrix

Kernels and Representation Operators

K	kernel $K : X \to \mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}})$
$K \star f$	convolution of kernel K with input f
$\mathcal{K}$	kernel operator or (more generally) representation operator $\mathcal{K} : T \to \mathrm{Hom}_{\mathbb{K}}(U, V)$

$\hat{K}$	kernel operator $\hat{K} : L^2_{\mathbb{K}}(X) \to \mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}})$ corresponding to a kernel K
$\mathcal{K}\|_X$	kernel $\mathcal{K}\|_X : X \to \mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}})$ corresponding to a kernel operator $\mathcal{K}$
$\tilde{\mathcal{K}}$	for a representation operator $\mathcal{K} : T \to \mathrm{Hom}_{\mathbb{K}}(U, V)$, this denotes the corresponding map $\tilde{\mathcal{K}} : T \otimes U \to V$ under the hom-tensor adjunction

The Wigner-Eckart Theorem

ρ_l, ρ_J	input- and output representations on the spaces V_l and V_J
Y_j^m, Y_l^n, Y_J^M	fixed chosen orthonormal basis vectors of the abstract irreducible representations V_j, V_l, V_J
$\langle JM\|K(x)\|ln\rangle$	matrix element of $K(x)$ for a kernel K and $x \in X$
$[l]$	dimension of l'th irrep V_l as $\mathbb{K}$-vector space
m_j	number of times V_j is in the Peter-Weyl decomposition of $L^2_{\mathbb{K}}(X)$
$[J(jl)]$	number of times V_J is in the direct sum decomposition of $V_j \otimes V_l$
c, c_{jis}	endomorphisms, mostly on V_J. c_{jis} are endomorphisms appearing in the Wigner-Eckart Theorem for steerable kernels
c_r	basis endomorphism, indexed with index set $r \in R$
$\langle JM\|c\|JM'\rangle$	matrix element at indices M, M' for endomorphism c
l_s, l_{jis}	linear equivariant isometric embeddings $l_s : V_J \to V_j \otimes V_l$ and $l_{jis} : V_J \to V_{ji} \otimes V_l$
p_{jis}	projection $p_{jis} : V_{ji} \otimes V_l \to V_J$ corresponding to (i.e.: adjoint to) the embedding l_{jis}
$\langle s, JM\|jmln\rangle$	Clebsch-Gordan coefficient corresponding to l_s
$\mathrm{CG}_{J(jl)s}$	3-dimensional matrix of Clebsch-Gordan coefficients
Y_{ji}^m	harmonic basis function, for example spherical harmonic. Element of $V_{ji} \subseteq L^2_{\mathbb{K}}(X)$
$\langle Y_{ji}^m \| x\rangle$	shorthand notation for $\lim_{U \in \mathcal{U}_x} \langle Y_{ji}^m \| \delta_U \rangle$. Equal to $\overline{Y_{ji}^m(x)}$
$\langle Y_{ji}\|x\rangle$	row vector with entries $\langle Y_{ji}^m \| x\rangle$
Rep	isomorphism between tuples of endomorphisms and kernel operators
Ker	isomorphism between tuples of endomorphisms and steerable kernels
K_{jisr}	basis kernel
w_{jisr}	learnable parameter

1. Introduction

1.1. Steerable and Gauge Equivariant Kernels and their Symmetry Properties

Deep learning is the workhorse of much of modern research in machine learning. Especially convolutional neural networks (CNNs) are ubiquitous and led to some of the great successes in previous years: AlexNet [1] was a landmark success in the classification of images by machine learning systems, and is thought of having led to the deep learning revolution.
CNNs are neural networks that distinguish themselves from fully connected neural networks by two main properties: local connectivity and weight sharing. Local connectivity leads to a processing of the network that hierarchically builds abstract features. Thus, for example, an eye is recognized by the presence of certain characteristic parts in the correct relative configuration, for example eye lids, pupils, and the iris. These parts themselves are assembled and recognized from lower-level features like specific color patterns and edges.
The weight sharing plays another role: by copying filters and placing them on all positions at the image, the idea is formalized that local features "mean the same everywhere", and that therefore the network should process the image in the same way everywhere. In classification, we see that convolutional neural networks preserve the invariance of meaning under certain symmetries: if the image is translated, e.g. moved to the right, then its meaning does not change. Since the processing of the network is exactly the same at the new position compared to the old one, the network also *assigns* the same meaning as before, and so the invariance of meaning under translation is preserved. More precisely, the output of a translated image under the CNN layer is precisely a translation of the output of the original image. In more diagrammatic fashion, we can express this as follows: we denote by I the image and by t a translation operator that, say, translates the image a little to the right. K is the filter, or kernel, that is convolved with the image in order to produce local features. Then both paths in the following diagram lead to the same result:

$$\begin{tikzcd} I \arrow[r, mapsto, "t"] \arrow[d, mapsto, "K\star"'] & t(I) \arrow[d, mapsto, "K\star"] \\ K \star I \arrow[r, mapsto, "t"'] & \begin{array}{c} K \star t(I) = \\ t(K \star I) \end{array} \end{tikzcd} \tag{1.1}$$

In recent years, this symmetry property formed the starting point for investigations into generalizations of this so-called *translation equivariance*. The main motivation is as follows: in many cases, there are symmetries besides translations that also preserve meaning. We therefore want our networks to preserve these symmetries. The most obvious examples for this appear in medical image analysis and the analysis of satellite images. When analyzing small-scale structure like patterns on skin patches, there is no relation between the orientation of a certain pattern and its medical meaning. For example, a skin anomaly should be classified as cancer irrespective of whether this pattern is upside down or not. In fact, there even is no ground truth orientation at all that would make it sensible to talk about "upside-down", and all orientations are equally likely. Usual CNNs have the problem that they do not preserve the patterns under rotation and reflection, and so they have to relearn the patterns in all appearing orientations. What we want, however, is the analog of Diagram 1.1. That is, additionally to the translation equivariance which we still want to fulfill, we want the following: assume r is a rotation or reflection operator that takes an image and outputs its rotation or reflection. Then for all such operators, the two paths in the following diagram should lead to the same result:

$$\begin{tikzcd} I \arrow[r, mapsto, "r"] \arrow[d, mapsto, "K\star"'] & r(I) \arrow[d, mapsto, "K\star"] \\ K \star I \arrow[r, mapsto, "r"'] & K \star r(I) = r(K \star I) \end{tikzcd} \tag{1.2}$$

Recent years have seen great success in formalizing this idea in different settings and came up with remarkable solutions to the underlying technical problems [2–4].
Recently, it became clear that this requirement for equivariance with respect to symmetry transformations is also related to physics [5–7]. Especially gauge equivariant CNNs [7] provide an interesting new perspective. What they address is the problem of applying neural networks to data on curved and topologically nontrivial shapes, for example the sphere. The problem then is that there is no preferred orientation for applying the kernel, and so the outcome of the convolution becomes ambiguous. The crucial idea is to view the outcome of the convolution as a field of features *expressed in a certain gauge*, which is a choice of a local reference frame. The desired property is that first convolving and then changing the gauge leads to the same outcome as first changing the gauge and then convolving the result. These changes in gauges, or reference frames, are no active transformations of the input but just passive changes in the viewpoint. However, when interpreting them as active changes in the *measurements* of the quantities involved, then the requirement to respect gauge transformations leads to a similar picture as Diagram 1.2. This is intimately connected to physics, where there is in the same way no preferred coordinate system in which to apply our physical theories. A change in the coordinate system will change the re-

sulting physical quantities and predictions, but only *relative to the chosen coordinates*, whereas the absolute predicted quantity remains the same.
What all of this suggests is that there is one theory for equivariant kernels that incorporates all the previously discussed examples. For gauge equivariant kernels this might not be obvious since they operate on curved shapes, so-called *Riemannian manifolds*. However, since the kernels actually live in the tangent space of the manifold, the flat story applies. A kernel is then formalized as a general function

$$K : \mathbb{R}^n \to \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$$

that locally maps between spaces of feature vectors $\mathbb{R}^{c_{\text{in}}}$ and $\mathbb{R}^{c_{\text{out}}}$. It is useful to give the feature vectors themselves an orientation, for example in order to detect edges in different angles. This means that there is a transformation group G like $\mathrm{O}(n)$ that can manipulate the input- and output features by transformation rules, called *linear representations*, ρ_{in} and ρ_{out}. For a representation ρ and each g in G, $\rho(g)$ is then a matrix, and group multiplication corresponds to multiplication of matrices:

$$\rho(gg') = \rho(g) \cdot \rho(g').$$

Additionally, the transformation group G is assumed to act naturally on $\mathbb{R}^n$ itself, for example if it is $\mathrm{O}(n)$. The kernel then needs to fulfill the following transformation rule for all $x \in \mathbb{R}^n$ and $g \in G$ in order to obey Diagram 1.2:

$$K(g \cdot x) = \rho_{\text{out}}(g) \cdot K(x) \cdot \rho_{\text{in}}(g)^{-1}. \tag{1.3}$$

This is the kernel constraint that first appeared in Cohen and Welling [3] and later, in a refined version, in Weiler et al. [8] and Weiler and Cesa [9]. It is the same kernel constraint that later reappeared in gauge equivariant CNNs [7]. A kernel that fulfills this constraint is called a *steerable kernel*.

1.2. An Analogy between Steerable Kernels and Spherical Tensor Operators

Symmetries play a large role in physics, as we already hinted at above when discussing gauge equivariant CNNs. Additionally, whenever we *actively* rotate all the "actors" in a physical interaction in the same manner, we expect that the physical behavior will essentially not change — or, more precisely, rotate in the same way as the physical actors we started with.
An important example of this is state transitions of electrons in a hydrogen atom. Basis states are in this context described by quantum numbers, including the so-called orbital angular momentum quantum number l and the magnetic quantum number n. Hereby, we separate off the radial part and ignore it. The state of the electron is then described by the so-called *ket* $|ln\rangle$. How can state transitions to another basis state $|JM\rangle$ emerge? One possibility for this is the absorption of a photon. The oscillating

electromagnetic field of this photon induces an operator $\mathcal{K}_j^m$ that can, via certain selection rules, induce changes between different states the electron might be in. Hereby, j and m are themselves quantum numbers, but this does not matter for now.
Mathematically speaking, possible transitions are described as follows: the starting state $|ln\rangle$ and the goal state $|JM\rangle$ both live in some large space $\mathcal{H}$, called *Hilbert space*, and $\mathcal{K}_j^m$ is an operator $\mathcal{K}_j^m : \mathcal{H} \to \mathcal{H}$. By definition, Hilbert spaces come equipped with a scalar product. The amplitude a of the state transition, which is closely related to the probability of this transition, is then given by the expression

$$a = \left\langle JM \middle| \mathcal{K}_j^m \middle| ln \right\rangle . \tag{1.4}$$

This can roughly be imagined as follows: when expressed in bases, one can associate to $\mathcal{K}_j^m$ a matrix, and to the *bra* $\langle JM|$ a row vector, whereas to the *ket*, $|ln\rangle$, there corresponds a column vector. Actually, the matrix of $\mathcal{K}_j^m$ might have infinitely many rows and columns. However, in the following, we will only consider the "suboperator" that maps from quantum states of quantum number l to those of quantum number J. This is basically a restriction and "corestriction" of the full operator, where the corestriction happens by an orthogonal projection on the component of quantum number J[1]. From here on, we mean with $\mathcal{K}_j^m$ only this smaller operator. Note that the result of multiplying a row vector first with a matrix and then with a column vector is just a scalar, and since the operations of the vectors and matrices correspond to those of the operator acting on the bras and kets, we obtain $a \in \mathbb{C}$.
Where does symmetry enter the picture? Imagine we rotate both the starting state $|ln\rangle$ and the goal state $|JM\rangle$ of the electron, as well as the oscillating electromagnetic field of the photon, all with the same rotation. What we then expect is that the amplitude of this state transition will not change since the physical laws are symmetric. That is, assume that $\left(\mathcal{K}_j^m\right)^g$ is the operator corresponding to the rotation g of the electromagnetic field, $|ln\rangle^g$ the rotation of the starting state and $|JM\rangle^g$ the rotation of the goal state. Then we expect an invariant amplitude

$$a = \left\langle JM \right|^g \left(\mathcal{K}_j^m\right)^g \left| ln \right\rangle^g .$$

Now, what does it mathematically mean to "rotate a basis state" or to "rotate an operator"? The quantum states of the electron live in representations with orbital angular momentum quantum numbers l and J, which really means — precisely as in the case of steerable kernels — that they come equipped with maps D_J and D_l that take a rotation $g \in \mathrm{SO}(3)$ and map it to an operator that can rotate states, $D_l(g)$ and $D_J(g)$. Expressed in bases, these correspond to the so-called *Wigner D-matrices*. The rotation of the ket, $|ln\rangle$, is then given by

$$|ln\rangle^g = D_l(g)|ln\rangle$$

[1] For a subspace $U \subseteq \mathcal{H}$, the restriction of $\mathcal{K}_j^m$ is defined as $\mathcal{K}_j^m|_U \coloneqq \mathcal{K}_j^m \circ i_U$ where $i_U : U \to \mathcal{H}$ is the inclusion. The corestriction to a subspace $V \subseteq \mathcal{H}$ is defined as $P_V \circ \mathcal{K}_j^m$, where $P_V : \mathcal{H} \to V$ is the orthogonal projection operator. We perform both a restriction and a corestriction.

for a fixed rotation g. The rotation of the bra $\langle JM|$ is given by

$$\langle JM|^g = \langle JM|D_J(g)^*$$

where $D_J(g)^*$ is the *adjoint* of $D_J(g)$. The adjoint of an operator hereby corresponds to the conjugate transpose of the corresponding matrix. From this, we can figure out analytically what the rotation of the operator $\mathcal{K}_j^m$ must be. Namely, we obtain the following relation from all the previous equations:

$$\begin{aligned}\big\langle JM\big|\mathcal{K}_j^m\big|ln\big\rangle &= a \\ &= \big\langle JM\big|^g \left(\mathcal{K}_j^m\right)^g \big|ln\big\rangle^g \\ &= \big\langle JM\big|D_J(g)^* \cdot \left(\mathcal{K}_j^m\right)^g \cdot D_l(g)\big|ln\big\rangle.\end{aligned}$$

Note that this equality must hold for all basis states $|ln\rangle$ and $\langle JM|$, which really means that the middle parts of these terms are forced to be equal. A comparison and a reordering — using that D_J is *unitary* and therefore inverted by building the adjoint — gives us the following definition for the rotation of the operator $\mathcal{K}_j^m$:

$$\left(\mathcal{K}_j^m\right)^g = D_J(g) \cdot \mathcal{K}_j^m \cdot D_l(g)^{-1}. \tag{1.5}$$

With some delight we see that this equation is relatively similar to Equation 1.3. The former equation for kernels expresses that a steerable kernel in rotated coordinates is given by the kernel in original coordinates, only conjugated by the representations corresponding to the input- and output field. The new equation says that a rotated operator in physics is given by conjugating the original operator with the representations of the input- and output states.

We now work on making this relation between operators in physics and kernels even stronger. For this, remember that j and m, the indices that define the operator $\mathcal{K}_j^m$, are also quantum numbers: j is an orbital angular momentum quantum number and m a magnetic quantum number. Actually, one then has one operator $\mathcal{K}_j^{m'}$ for each magnetic quantum number m'. From physics, it is well-known that the operators $\mathcal{K}_j^m$ transform under rotation in the same way as the basis kets in the linear representation D_j. Let $D_j^{m'm}$ be the matrix elements of the corresponding Wigner D-matrices, where m' is the row index and m the column index. That the $\mathcal{K}_j^m$ transform as the basis kets in this representation means the following:

$$\left(\mathcal{K}_j^m\right)^g = \sum_{m'} D_j^{m'm}(g)\mathcal{K}_j^{m'}. \tag{1.6}$$

Comparing with Equation 1.5 we obtain:

$$\sum_{m'} D_j^{m'm}(g)\mathcal{K}_j^{m'} = D_J(g) \cdot \mathcal{K}_j^m \cdot D_l(g)^{-1}. \tag{1.7}$$

A collection of operators $\mathcal{K}_j^{m'}$ transforming with this rule is called a *spherical tensor operator* in physics. If $j = 0$, then there is only one operator with the trivial transformation law, which is called a *scalar operator*. For the case $j = 1$, there are three

operators that transform in the same way as vectors in $\mathbb{R}^3$ under the standard matrix representation of $\mathrm{SO}(3)$. This case is then called a *vector operator*. Tensor operators are the generalization to arbitrary $j \in \mathbb{N}_{\geq 0}$.
In order to make the analogy to steerable kernels stronger, we would like to interpret a spherical tensor operator as *one object* $\mathcal{K}$, in the same way as a kernel K is one single object and not just a disjoint collection of matrices in $\mathbb{R}^{c_{\mathrm{out}} \times c_{\mathrm{in}}}$. For this, we interpret $\mathcal{K}$ as a function that assigns to arbitrary kets of quantum number j an operator. Namely, the kets $|jm'\rangle$ are the basis of the space on which the representation D_j acts. We then define $\mathcal{K}$ as the *unique linear map* which is given on basis kets as follows:

$$\mathcal{K} : |jm\rangle \mapsto \mathcal{K}_j^m.$$

We can then deduce from Equation 1.7 the following, where we insert the identity in the first step, use the definition of the matrix elements of D_j and a swap in order in the second step and the linearity of $\mathcal{K}$ in the third step:

$$\begin{aligned}
\mathcal{K}\big(D_j(g)|jm\rangle\big) &= \mathcal{K}\Big(\sum\nolimits_{m'} |jm'\rangle\langle jm'|D_j(g)|jm\rangle\Big) \\
&= \mathcal{K}\Big(\sum\nolimits_{m'} D_j^{m'm}(g)|jm'\rangle\Big) \\
&= \sum_{m'} D_j^{m'm}(g)\mathcal{K}\left(|jm'\rangle\right) \\
&= \sum_{m'} D_j^{m'm}(g)\mathcal{K}_j^{m'} \\
&= D_J(g) \cdot \mathcal{K}_j^m \cdot D_l(g)^{-1} \\
&= D_J(g) \cdot \mathcal{K}(|jm\rangle) \cdot D_l(g)^{-1}.
\end{aligned} \tag{1.8}$$

If now $|v\rangle = \sum_{m'} \langle jm'|v\rangle \cdot |jm'\rangle$ is any ket of quantum number j, not necessarily a basis ket, then from the linearity of $\mathcal{K}$ and Equation 1.8 we obtain

$$\mathcal{K}\big(D_j(g) \cdot |v\rangle\big) = D_J(g) \cdot \mathcal{K}(|v\rangle) \cdot D_l(g)^{-1}. \tag{1.9}$$

This equation is essentially the starting point for the definition of a *representation operator* as a generalization of spherical tensor operators that can be found in Jeevanjee [10]. This, finally, really looks like Equation 1.3. In this comparison, the action of the group G on $\mathbb{R}^n$ in deep learning is replaced by the action of $\mathrm{SO}(3)$ via D_j on the kets of quantum number j.
Thus, we see the following analogies:

1. Input features in deep learning correspond to starting states, given as kets $|ln\rangle$, in quantum mechanics.

2. Output features in deep learning correspond to goal states, or bras $\langle JM|$, in quantum mechanics.

3. Steerable kernels in deep learning correspond to spherical tensor operators in quantum mechanics.

Note that we stick from now on to the view — somewhat unfamiliar in the physics literature — that spherical tensor operators are linear functions that map kets of quantum number j to operators from states of quantum number l to states of quantum number J. This is more abstract than the view that it is a collection of finitely many operators $\mathcal{K}_j^{m'}$ with certain transformation properties, however more suitable for our aims due to the analogy with steerable kernels.

1.3. The Wigner-Eckart Theorem and Research Questions

An important question in physics is how to describe such spherical tensor operators. Crucially, spherical tensor operators are *linear* functions of the kets with orbital angular momentum quantum number j, and so, as all linear functions, they are completely determined by their matrix elements with respect to bases of the involved spaces. If $[j]$, $[l]$, and $[J]$ are the dimensions of the three involved spaces, then the spherical tensor operator is described by a $([J] \times [l]) \times [j]$-tensor. The reason is that for each ket of quantum number j, the result is a whole operator from states of quantum number l to states of quantum number J. The number of matrix elements that need to be determined then seems quite large and it might be a hassle to figure it all out. However, this concern neglects Equation 1.9 which tells us how such an operator changes under rotation of the ket of quantum number j. This imposes strong relations on different matrix elements. In fact, these relations are so strong that one can show that there is a *single complex number* that is able to completely characterize the spherical tensor operator. This is the content of the famous Wigner-Eckart Theorem [10]:

Theorem 1.3.1. *Assume $\mathcal{K}$ is a spherical tensor operator that maps kets of quantum number j to operators from quantum states of quantum number l to quantum states of quantum number J. Then there is a unique complex number, called* reduced matrix element *and denoted by $\langle J \| \mathcal{K} \| l \rangle$, that completely determines $\mathcal{K}$. More precisely, there are coupling coefficients $\langle JM | jmln \rangle$, the so-called* Clebsch-Gordan coefficients, *which are completely independent of the spherical tensor operator $\mathcal{K}$, such that the matrix elements of $\mathcal{K}$ are given as follows:*

$$\left\langle JM | \mathcal{K}_j^m | ln \right\rangle = \langle J \| \mathcal{K} \| l \rangle \cdot \langle JM | jmln \rangle .$$

This makes us wonder: can this result be transported over into the realm of deep learning in order to get a description of all possible steerable kernels? At first sight, this seems difficult: while we noted that spherical tensor operators are linear functions mapping kets to operators, steerable kernels are certainly not linear in their input in $\mathbb{R}^n$ in any meaningful sense. This leads to the following set of research questions:

1. *Is it possible to "linearize" a steerable kernel K to a map $\hat{K}$ that is linear in its input?*

2. *Does the linear version $\hat{K}$ then share enough properties with spherical tensor operators from physics such that a generalized Wigner-Eckart Theorem can be proven for it?*

3. *Is it possible to transfer this result to get a description of the original kernel K?*

4. *Does this result help us in parameterizing equivariant neural networks?*

5. *In what generality is all of this possible?*

In the next section, we sketch the answers to these questions that we describe in detail in the rest of this thesis.

1.4. A Wigner-Eckart Theorem for Steerable Kernels of General Compact Groups

From now on, we write "order" instead of quantum number, since this is the more common term in deep learning.
The answer to all the first four of the research questions is an unambiguous "yes". The last question does not have a definitive answer yet, however, we are able to *completely cover the theory of steerable CNNs on $\mathbb{R}^n$ and gauge equivariant CNNs for compact structure groups* with our investigations. Note that the following is just a sketch of the final result. We will define all the terminology in more detail and clarity in the chapters to come.
We work in the following general setting: G is an *arbitrary compact transformation group* that can, for example, act as a transformation group that fixes the origin in $\mathbb{R}^n$, including groups like $\mathrm{O}(n)$ or finite groups as examples[2]. X is any orbit under that action, i.e. a set of points in $\mathbb{R}^n$ that can be interchanged by the action of G[3]. One can show that a theory of steerable kernels restricted to such orbits is enough in order to recover steerable kernels that are described on the whole of $\mathbb{R}^n$. Furthermore, ρ_l and ρ_J are *irreducible* representations of G corresponding to the input field and the output field of the steerable kernel. More general finite-dimensional input- and output representations can be assembled from such irreducible ones, and so one does not lose generality by restricting to irreducible representations. These representations thereby either act on a real or a complex vector space, and in order to cover both, we write $\mathbb{K}$ instead of $\mathbb{R}$ or $\mathbb{C}$. Let $[l]$ and $[J]$ be the dimensions of the input- and output representation. Then the kernel is a function

$$K : X \to \mathbb{K}^{[J]\times[l]}.$$

[2]Note that we assume all topological spaces in this work to be Hausdorff. The reader should not worry at this point, if she or he does not know this term.

[3]In the most general formulation we find, X is an arbitrary homogeneous space of G and thus need not be thought of as being embedded in $\mathbb{R}^n$. For mitigating confusion, we mention that this homogeneous space does not have the same meaning as in the general theory of equivariant CNNs on homogeneous spaces [11], which we discuss in more detail in the chapter on related work 5 and our conclusion 7.

We further need the following ingredients:

1. $\hat{G}$ is the set of isomorphism classes of so-called *irreducible unitary* representations of the compact group G.

2. For $j \in \hat{G}$, m_j is the number of times that the j'th representation appears as a "direct summand" in the space of square-integrable functions on X, $L^2_{\mathbb{K}}(X)$. For this to make sense, this space of square-integrable functions also carries a representation of the compact group G in a suitable way. In the examples we describe in Chapter 6, the number m_j is always 0 or 1, but other possibilities exist in theory.

3. For each $j \in \hat{G}$, let $[j]$ be the dimension of the irreducible representation of order j. Then for each $i = 1, \dots, m_j$ and $m = 1, \dots, [j]$, we let $Y^m_{ji} : X \to \mathbb{K}$ be a square-integrable function such that the collection of all these functions for fixed j and i is *steerable*. That is, it transforms under the transformation of the space X via G in the same way as the basis vectors of the irreducible representation of order j. We also call them *harmonic basis functions* in analogy with the spherical harmonics. The collection of all these functions for all j, i, and m forms an orthonormal basis of $L^2_{\mathbb{K}}(X)$. These functions exist according to the *Peter-Weyl Theorem* 2.1.22.

4. For all j, let $[J(jl)]$ be the number of times that the representation ρ_J appears as a direct summand in the tensor product of the representations of order j and l. This number can be different from 0 or 1 as we will see in the Example of $\mathrm{SO}(2)$-equivariant CNNs in Section 6.2.

5. For each $s = 1, \dots, [J(jl)]$, there exist so-called *Clebsch-Gordan coefficients* $\langle s, JM | jmln \rangle$ as above. These are coupling coefficients between basis vectors of the irreducible representation of order J appearing in the tensor product of representations of order j and order l. Different from the situation in physics, they now contain an additional index s, which can be imagined as an additional quantum number.

We then prove the following Wigner-Eckart Theorem for steerable kernels, which we state in more detail in Theorem 4.1.13:

Theorem 1.4.1 (Wigner-Eckart Theorem for steerable kernels)**.** *A steerable kernel $K : X \to \mathbb{K}^{[J] \times [l]}$ is completely and uniquely determined by an arbitrary collection $\{c_{jis}\}$ of maps $c_{jis} : \mathbb{K}^{[J]} \to \mathbb{K}^{[J]}$, called* endomorphisms*, with the following indices and properties:*

1. *The indices are $j \in \hat{G}$, $i = 1, \dots, m_j$ and $s = 1, \dots, [J(jl)]$.*

2. *$c_{jis} : \mathbb{K}^{[J]} \to \mathbb{K}^{[J]}$ is linear.*

3. *c_{jis} commutes with the J'th representation, that is: for all $g \in G$ we have $\rho_J(g) \circ c_{jis} = c_{jis} \circ \rho_J(g)$.*

The matrix elements of $K(x) \in \mathbb{K}^{[J]\times[l]}$ are then for each $x \in X$ given as follows:

$$\langle JM|K(x)|ln\rangle = \sum_{j\in\hat{G}} \sum_{i=1}^{m_j} \sum_{s=1}^{[J(jl)]} \sum_{m=1}^{[j]} \sum_{M'=1}^{[J]} \langle JM|c_{jis}|JM'\rangle \cdot \langle s, JM'|jmln\rangle \cdot \left\langle Y_{ji}^m|x\right\rangle .$$

Here, the $\langle JM|c_{jis}|JM'\rangle$ are the matrix elements of the function c_{jis}, replacing the reduced matrix element in the original Wigner-Eckart Theorem. It is the only part of the right-hand side of the formula that depends on the kernel K[4]*. $\left\langle Y_{ji}^m|x\right\rangle$ is "physics notation" for $\overline{Y_{ji}^m(x)}$, where the overline denotes complex conjugation. This is the only part in the right-hand side that depends on the input x. The Clebsch-Gordan coefficients, $\langle s, JM'|jmln\rangle$, do neither depend on the kernel, nor on the input.*

An important remark is the following: As you may notice, the formerly "reduced matrix element" in the original Wigner-Eckart Theorem 1.3.1 is now replaced by matrix elements of the endomorphisms c_{jis} that depend on the indices M and M'. In the physics context, one works over the complex numbers $\mathbb{C}$ and this dependence disappears. The general concept that applies to both the complex and the real case is the notion of an endomorphism as defined above. We call the $\langle JM|c_{jis}|JM'\rangle$ the *generalized reduced matrix elements* of the kernel K.

1.5. What is this Theorem Good for?

Now that we have described this theorem, we would like to know what we hope to get from it. After all, proving such a theorem in such generality is considerable work, and we might wonder if it is worth the effort. The following reasons make us confident that it is:

1. When a basis $\{c_r \mid r \in R\}$ of the space of endomorphisms of $\mathbb{R}^{[J]}$ is known, then it can be shown, see Theorem 4.1.15, that this leads to a description of a basis for the space of steerable kernels $K : X \to \mathbb{K}^{[J]\times[l]}$. Namely, for each $j \in \hat{G}$, $i = 1, \ldots, m_j$, $s = 1, \ldots, [J(jl)]$ and $r \in R$, one then obtains a basis kernel $K_{jisr} : X \to \mathbb{K}^{[J]\times[l]}$ with the following matrix elements:

$$\langle JM|K_{jisr}(x)|ln\rangle = \sum_{m=1}^{[j]} \sum_{M'=1}^{[J]} \langle JM|c_r|JM'\rangle \cdot \langle s, JM'|jmln\rangle \cdot \left\langle Y_{ji}^m|x\right\rangle .$$

   The collection of all K_{jisr} then forms a basis for steerable kernels $X \to \mathbb{K}^{[J]\times[l]}$. This, in turn, tells us how to parameterize our equivariant neural network layer:

[4]The kernel depends on the c_{jis}, since they determine it. But in the other direction, the kernel actually also determines the c_{jis} uniquely and we can therefore say that these endomorphisms and their matrix elements "depend on the kernel".

We need to learn coefficients $w_{jisr} \in \mathbb{K}$, and an arbitrary steerable kernel K is then given by the linear combination

$$K = \sum_{j \in \hat{G}} \sum_{i=1}^{m_j} \sum_{s=1}^{[J(jl)]} \sum_{r \in R} w_{jisr} \cdot K_{jisr}.$$

2. For this to work, we need to be able to determine irreducible representations for each $j \in \hat{G}$, endomorphisms, Clebsch-Gordan coefficients, and harmonic basis functions Y_{ji}^m in practice. We will see in examples in Chapter 6 that this is generally a doable task.

3. The level of generality of this theorem means that we are relieved from thinking the same arguments over and over again in specific use cases. Our theorem clearly separates the *general structure* of steerable kernels, which is always the same, from the specifics of the situation at hand. These specifics are best thought of as being independent of the theory of steerable CNNs, since they are representation-theoretic in nature.

4. What we will see is that the search for a steerable kernel basis in specific use cases never *just* provides us with this kernel basis. Arguably, along the way we understand a great deal about the representation theory of the specific group in question. This always has to happen anyway, but with our method, it is very explicit and separated from the concerns of equivariant deep learning.

5. In the past, there was work *deriving* kernel constraints for steerable and gauge equivariant CNNs in full generality, but no general *solution* strategy for this constraint. This is the first general solution of how to parameterize steerable and gauge equivariant kernels for compact transformation groups. Thus, it will provide practitioners working with specific groups with a guide and a tool for validating their findings. Since solutions in the past were often heuristic and did in many cases not *prove* the completeness of the resulting kernel space, nor that it was linearly independent, it seems reassuring to have this general result.

6. What seems especially useful is the identification of *endomorphisms* as a central building block of steerable kernels, which was probably not observed before. This helps in a better understanding of the differences between kernel spaces of different transformation groups. For example, as we will see in examples that we derive in Chapter 6, it helps to explain why $\mathrm{SO}(2)$-equivariant networks over the real numbers have more steerable kernels than over the complex numbers, whereas for the group $\mathrm{SO}(3)$ this distinction is not present.

7. Since we emphasize the core abstract structure of the problem and solution throughout this work, it may help for generalizing the results even further in the future. This may provide general solutions for equivariant kernels that are defined on non-flat spaces, like spherical CNNs or general CNNs on homogeneous

spaces [11, 12]. While the underlying topological spaces in these cases are assumed to be homogeneous spaces, which is less general than the situation of gauge equivariant CNNs, the *kernel theory* for general CNNs on homogeneous spaces is actually a further generalization that we do not fully cover in this work.

8. Finally, the strong analogy between steerable kernels and spherical tensor operators, and the presence of a Wigner-Eckart Theorem in both settings, makes us hope for fruitful collaborations between physicists, chemists, and deep learning researchers. This might lead to further insights into the nature of learning in the presence of symmetries and ultimately a greater understanding of inductive biases in general. For physicists and chemists, it might help in creating machine learning systems that make useful predictions for physical experiments.

1.6. A Tour through the Thesis

We now outline the structure of this thesis. In Chapter 2, we start by reviewing the foundations of the representation theory of compact groups, including Haar measures on the group and their homogeneous spaces. We will formulate the Peter-Weyl Theorem 2.1.22 which we use in crucial steps in this work. It tells us how to view the space of square-integrable functions on a homogeneous space itself as a representation of the compact group, and how it splits into irreducible representations. In the second half of this chapter, we go over some central steps of the proof of this theorem. We do this since the theorem is usually found in the literature only for *complex* representations. We, however, also need it for *real* representations. This also leads to a slight change in the formulation of the theorem itself concerning the multiplicities of the irreducible representations in the regular representation.

Equipped with a clear understanding of the representation theory of compact groups, we first engage with steerable CNNs in Chapter 3. We start with a description of steerable CNNs as they appear in the literature. Then, we reformulate steerable kernels in more abstract terms as certain maps on general homogeneous spaces of a compact group. We argue that this abstract formulation is all we need in order to determine steerable kernels in practice. Once we have this abstract formulation, we will see that this *almost* looks like spherical tensor operators — or their representation-theoretic generalizations, representations operators — from physics. However, the linearity will still be missing. In Theorem 3.1.7 we will then make a precise connection to representation operators which are defined on the space of square-integrable functions on the homogeneous space. We call these *kernel operators*. This link is the bridge that we need in order to be able to transport physical results over into the realm of deep learning. In the second half of the chapter, we do a thorough proof of Theorem 3.1.7.

In Chapter 4 we will then formulate and prove the Wigner-Eckart Theorem for steerable kernels of general compact groups 4.1.13 which we outlined before. This Theorem is split into several parts: Firstly, a basis-independent Wigner-Eckart Theorem for kernel operators, which is essentially a generalization of the usual Wigner-Eckart Theorem from physics. It makes in essential parts use of the Peter-Weyl Theorem outlined in Chapter 2. Secondly, a basis-independent Wigner-Eckart Theorem for steerable kernels, which uses the version for kernel operators and the correspondence between steerable kernels and kernel operators outlined before in Chapter 3. And thirdly, a basis-dependent version for steerable kernels that is identical to the formula outlined in Theorem 1.4.1 in this introduction. We discuss some practical considerations for how to apply this theorem in practice. The second half of the chapter contains a proof of the first part of this Wigner-Eckart Theorem, which we leave out before.

In Chapter 5 we discuss related work. We first compare with prior work on steerable CNNs and gauge equivariant CNNs, since this most obviously falls within the scope of the theory outlined in this work. We then also discuss other work on the realm of equivariant deep learning that is inspired by representation theory and physics. We conclude by discussing purely theoretical work that was published before.

In Chapter 6, we then look at specific example applications of our theory. In these examples, we look at specific compact transformation groups G, specific, relevant homogeneous spaces X of the group (basically just the orbit of the chosen action of G on $\mathbb{R}^n$) and one of the fields $\mathbb{R}$ or $\mathbb{C}$. For this combination we derive a basis for the space of steerable kernels between arbitrary irreducible input- and output representations of the group. Specifically, we look at harmonic networks [4], $\mathrm{SO}(2)$-equivariant networks for real representations [9], $\mathbb{Z}_2$-equivariant networks for real representations, $\mathrm{SO}(3)$-equivariant networks for both real and complex representations [6, 8], and $\mathrm{O}(3)$-equivariant networks for both real and complex representations. In these results, we show that our theory is consistent with already implemented networks, but also show how to parameterize steerable CNNs for cases that did to the best of our knowledge not appear in published work yet. The investigation of $\mathbb{Z}_2$-equivariant CNNs will additionally show that our result is consistent with group convolutional CNNs for the regular representation [2]. By using the same guideline for all examples, we see that applying our theorem is a doable task that can be accomplished for all compact groups that one wishes to investigate in practice.

Finally, in Chapter 7 we come to our conclusions and discuss future work, especially centered around the question of how to further generalize the results in this work.

In Appendix A, we summarize some of the main notions from the theory of topological spaces, metric spaces, normed vector spaces and (pre-)Hilbert spaces that we use throughout this work.

Chapters 2, 3, and 4 contain the bulk of the theoretical work. We recommend the reader to first only read the first halves of these chapters, Sections 2.1, 3.1 and 4.1, since they contain the formulation of the most important results and the main intuitions, whereas the second halves of these chapters mainly contain proofs that can be skipped when going over the material for the first time. We only very occasionally make use of any concepts defined in these second halves.

1.7. Prerequisites

While we try to make this thesis accessible, it is clear at the same time that prior knowledge in some areas of mathematics and deep learning is useful for appreciating this work. In the realm of deep learning, we expect the reader to be familiar with convolutional neural networks (CNNs). Additionally, it is useful if the reader has engaged before with the literature in equivariant deep learning, with the most important prior sources to consult being Cohen and Welling [2], Weiler et al. [8] and Weiler and Cesa [9].

With respect to mathematical prerequisites, it is clearly useful if the reader has prior knowledge and intuitions in representation theory. We define all of the notions that we use, but it is impossible to give a thorough introduction to the way of thinking in this area in such a short space. If the reader has prior knowledge only in the representation theory of finite groups, or maybe in the representation theory of algebras instead of that of groups, then we expect that the intuitions carry over well enough in order to read this thesis.

Additionally, the reader clearly needs a good foundation in linear algebra and calculus at the level of a first-year undergraduate in mathematics or related fields. Additionally, the reader should have some prior knowledge in measure theory in order to understand and appreciate the definition and properties of the so-called Haar measure. However, different from many texts in the realm of artificial intelligence and machine learning, this work does never make use of any techniques from statistics or probability theory, so this is not required as prior knowledge.

In the appendix, we collect some results on topology, the theory of metric and normed spaces, and (pre-)Hilbert spaces that we use throughout the text. The recommendation is similar as with the prerequisites in representation theory: It is useful to have prior encounters with these areas of mathematics since we cannot give a thorough introduction into the way of thinking of these subjects.

2. Representation Theory of Compact Groups

In this chapter, we outline the main ingredients of the representation theory of compact groups that we need for our applications to steerable CNNs. Usually, this theory is only developed for representations over the complex numbers. However, since we want to apply it also to steerable CNNs using real representations, we need to be a bit more careful. In particular, we need to make sure that the Peter-Weyl Theorem is correctly stated and proven.

The outline is as follows: In Section 2.1, we start by stating all the important definitions and concepts from group theory and representation theory of (unitary) representations that are needed for formulating the Peter-Weyl Theorem. After defining Haar measures both for compact groups and their homogeneous spaces and shortly discussing their square-integrable functions, we formulate the Peter-Weyl Theorem 2.1.22. In Section 2.2, then, we give a proof of this version of the Peter-Weyl Theorem, carefully making sure to not use properties that are only true over $\mathbb{C}$. In some essential steps, mainly the density of the matrix coefficients in the regular representation, we refer to the literature, since the proof clearly does not make use of $\mathbb{C}$ per se. While we initially only give the proof for the regular representation, i.e. the space of square-integrable functions on the group itself, we end this section with a discussion of general unitary representations and, in particular, the space of square-integrable functions for an arbitrary homogeneous space.

In the whole chapter, let $\mathbb{K}$ be the field of real or complex numbers.

2.1. Foundations of Representation Theory and the Peter-Weyl Theorem

2.1.1. Preliminaries of Topological Groups and their Actions

In this section, we define preliminary concepts from topological groups and their actions. This material can, for example, be found in detail in Arhangel'skii and Tkachenko [13]. For the topological concepts that we use, we refer to Appendix A.1.

Definition 2.1.1 (Group, Abelian Group)**.** A *group* $G = (G, \cdot, (\cdot)^{-1}, e)$, most often simply written G, consists of the following data:

1. A set G of group elements $g \in G$.

2. A multiplication $\cdot : G \times G \to G$, $(g, h) \mapsto g \cdot h$.

3. An inversion $(\cdot)^{-1} : G \to G$, $g \mapsto g^{-1}$.

4. A distinguished unit element $e \in G$. It is also called neutral element.

They are assumed to have the following properties for all $g, h, k \in G$:

1. The multiplication is associative: $g \cdot (h \cdot k) = (g \cdot h) \cdot k$.

2. The unit element is neutral with respect to multiplication: $e \cdot g = g = g \cdot e$.

3. The inversion of an element multiplied with itself is the neutral element: $g \cdot g^{-1} = g^{-1} \cdot g = e$.

A group is called *abelian* if, additionally, the multiplication is commutative: $g \cdot h = h \cdot g$ for all $g, h \in G$. If this is the case, a group is often written as $G = (G, +, -(\cdot), 0)$.

If we consider several groups at once, say G and H, then we often do not distinguish their multiplication, inversion, and neutral elements in notation. It will be clear from the context which group the operation belongs to.

Definition 2.1.2 (Subgroup)**.** Let G be a group and $H \subseteq G$ a subset. H is called a *subgroup* if:

1. For all $h, h' \in H$ we have $h \cdot h' \in H$.

2. For all $h \in H$ we have $h^{-1} \in H$.

3. The neutral element $e \in G$ is in H.

Consequently, H is also a group with the restrictions of the multiplication and inversion of G to H.

Definition 2.1.3 (Group homomorphism)**.** Let G and H be groups. A function $f : G \to H$ is called a *group homomorphism* if it respects the multiplication, inversion, and neutral element, i.e. for all $g, h \in G$:

1. $f(g \cdot h) = f(g) \cdot f(h)$.

2. $f(g^{-1}) = f(g)^{-1}$.

3. $f(e) = e$.

The second and third properties automatically follow from the first and so do not need to be verified in order to prove that a certain function is a group homomorphism.

Definition 2.1.4 (Topological Group, Compact Group)**.** Let G be a group and $\mathcal{T}$ be a topology of the underlying set of G. Then $G = (G, \mathcal{T})$ is called a *topological group* [13] if both multiplication $G \times G \to G$, $(x, y) \mapsto x \cdot y$ and inversion $G \to G$, $x \mapsto x^{-1}$ are continuous maps. Additionally, we always assume the topology to be Hausdorff.
A topological group is called *compact* if the underlying topological space is compact.

From now on, all groups considered are compact topological groups. Furthermore, whenever G is a finite group, we assume that it is a topological group with the *discrete topology*, i.e. the topology with respect to which all subsets of G are open.
We will need the following definition in order to define homogeneous spaces:

Definition 2.1.5 (Group Action)**.** Let G be a compact group and X a topological space. Then a *group action* of G on X is a continuous function $\cdot : G \times X \to X$ with the following properties:

1. $(g \cdot h) \cdot x = g \cdot (h \cdot x)$ for all $g, h, \in G$ and $x \in X$.

2. $e \cdot x = x$ for all $x \in X$.

We will often simply write gx instead of $g \cdot x$. Also, note that the multiplication within G is denoted by the same symbol as the group action on the space X.

Definition 2.1.6 (Orbit)**.** Let $\cdot : G \times X \to X$ be a group action. Let $x \in X$. Then it's *orbit*, denoted $G \cdot x$, is given by the set

$$G \cdot x \coloneqq \{g \cdot x \mid g \in G\} \subseteq X.$$

Definition 2.1.7 (Transitive Action, Homogeneous Space)**.** Let $\cdot : G \times X \to X$ be a group action. This action is called *transitive* if for all $x, y \in X$ there exists $g \in G$ such that $gx = y$. Equivalently, each orbit is equal to X, that is: For all $x \in X$ we have $G \cdot x = X$.
X is called a *homogeneous space* (with respect to the action) if the action is transitive, X is Hausdorff and $X \neq \emptyset$.

The Hausdorff condition and non-emptiness in the definition of homogeneous spaces is needed for Lemma 2.1.21, which is necessary to even define a normalized Haar measure on a homogeneous space. Some texts in the literature may define homogeneous spaces without these conditions.

Definition 2.1.8 (Stabilizer Subgroup)**.** Let $\cdot : G \times X \to X$ be a group action. Let $x \in X$. The *stabilizer subgroup* G_x is the subgroup of G given by

$$G_x \coloneqq \{g \in G \mid gx = x\} \subseteq G.$$

Example 2.1.9. The multiplication of the group G is a group action of G on itself. G is a homogeneous space with this action. Furthermore, for each $g \in G$ the stabilizers G_g are the trivial subgroup e.
In general, homogeneous spaces with the property that all stabilizers are trivial are called *torsors* or *principal homogeneous spaces*. Principal homogeneous spaces are topologically indistinguishable from the group itself.

2.1.2. Linear and Unitary Representations

In this section, we define many of the foundational concepts about linear and unitary representations [14, 15].
Whenever we will consider linear or unitary representations of compact groups, we want those representations to be *continuous*. This requires that the vector spaces on which our groups act carry themselves a topology. Prototypical examples of such vector spaces are (pre-)Hilbert spaces. They are the main examples of vector spaces considered in this work. Foundational concepts about (pre-)Hilbert spaces can be found in Appendix A.2. The most important difference between how we view pre-Hilbert spaces and how it can often be found in the literature is that in this work, scalar products are antilinear in the first component and linear in the second. This is the convention usually chosen in physics.
For a vector space V over $\mathbb{K}$ let $\mathrm{Aut}_{\mathbb{K}}(V)$ be the group of invertible linear functions from V to V. Sometimes in the literature, this is also written $\mathrm{GL}(V, \mathbb{K})$. The multiplication is given by function composition and the neutral element by the identity function id_V on V.

Definition 2.1.10 (Linear Representation)**.** Let G be a compact group and V be a $\mathbb{K}$-vector space carrying a topology, for example a (pre)-Hilbert space. Then a *linear representation* of G on V is a group homomorphism $\rho : G \to \mathrm{Aut}_{\mathbb{K}}(V)$ which is continuous in the following sense: for all $v \in V$, the function

$$\rho^v : G \to V, \ \ g \mapsto \rho^v(g) := \rho(g)(v)$$

is continuous. From the definition we obtain $\rho(e) = \mathrm{id}_V$, $\rho(g \cdot h) = \rho(g) \circ \rho(h)$ and $\rho(g^{-1}) = \rho(g)^{-1}$ for all $g, h \in G$. For simplicity, we also just say *representation* or G-representation instead of *linear representation*. Instead of denoting the representation by ρ, we often denote it by V if the function ρ is clear from the context.

Note that in this definition, V can be any abstract topological $\mathbb{K}$-vector space with a topology and does not need to be a space $\mathbb{K}^n$ or something similar. Consequently, we usually do not view the functions $\rho(g)$ as matrices, but as abstract linear automorphisms from V to V.

Definition 2.1.11 (Intertwiner)**.** Let $\rho : G \to \mathrm{Aut}_{\mathbb{K}}(V)$ and $\rho' : G \to \mathrm{Aut}_{\mathbb{K}}(V')$ be two representations over the same group G. An *intertwiner* between them is a linear function $f : V \to V'$ that is additionally equivariant with respect to ρ and ρ' and continuous. Equivariance means that for all $g \in G$ one has $f \circ \rho(g) = \rho'(g) \circ f$, which means the following diagram commutes:

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "\rho(g)"'] & V' \arrow[d, "\rho'(g)"] \\ V \arrow[r, "f"'] & V' \end{tikzcd}$$

Definition 2.1.12 (Equivalent representations)**.** Let $\rho : G \to \mathrm{Aut}_{\mathbb{K}}(V)$ and $\rho' : G \to \mathrm{Aut}_{\mathbb{K}}(V')$ be two representations. They are called *equivalent* if there is an intertwiner $f : V \to V'$ that has an inverse. That is, there exists an intertwiner $g : V' \to V$ such that $g \circ f = \mathrm{id}_V$ and $f \circ g = \mathrm{id}_{V'}$.

In categorical terms, equivalent representations are isomorphic in the category of linear representations. The reason we do not call them isomorphic is that there is a stronger notion of isomorphism between representations which we will later use, namely isomorphisms of *unitary* representations.

Definition 2.1.13 (Invariant Subspace, Subrepresentation, Closed Subrepresentation)**.** Let $\rho : G \to \mathrm{Aut}_{\mathbb{K}}(V)$ be a representation. An *invariant subspace* $W \subseteq V$ is a linear subspace of V such that $\rho(g)(w) \in W$ for all $g \in G$ and $w \in W$. Consequently, the restriction $\rho|_W : G \to \mathrm{Aut}_{\mathbb{K}}(W)$, $g \mapsto \rho(g)|_W : W \to W$ is a representation as well, called *subrepresentation* of ρ.
A subrepresentation is called *closed* if W is closed in the topology of V.

Definition 2.1.14 (Irreducible Representation)**.** A representation $\rho : G \to \mathrm{Aut}_{\mathbb{K}}(V)$ is called *irreducible* if $V \neq 0$ and if the only closed subrepresentations of V are 0 and V itself. An irreducible representation is also shortly called *irrep*.

Definition 2.1.15 (Unitary group)**.** Let V be a pre-Hilbert space. The *unitary group* $\mathrm{U}(V)$ of V is defined as the group of all linear invertible maps $f : V \to V$ that respect the inner product, i.e. $\langle f(x)|f(y)\rangle = \langle x|y\rangle$ for all $x, y \in V$. It is a group with respect to the usual composition and inversion of invertible linear maps.

Note that if the field $\mathbb{K}$ is the real numbers, then what we call "unitary" is actually called *orthogonal*, and the group would be denoted $\mathrm{O}(V)$. However, the mathematical properties are essentially the same, and since the term "unitary" is more widely used (as normally, representations over the complex numbers are considered) we stick with "unitary".
More generally, we have the following:

Definition 2.1.16 (Unitary Transformation)**.** Let V, V' be two pre-Hilbert spaces. A *unitary transformation* $f : V \to V'$ is a bijective linear function such that $\langle f(x)|f(y)\rangle = \langle x|y\rangle$ for all $x, y \in V$. These can be regarded as isomorphisms between pre-Hilbert spaces.

Note that unitary transformations are in particular isometries, i.e. they keep the distances of vectors with respect to the metric defined by the scalar product. For the definition of this metric, see the discussion before and after Definition A.1.14.

Definition 2.1.17 (Unitary representation)**.** Let V be a pre-Hilbert space and G a group. Then a representation $\rho : G \to \mathrm{Aut}_{\mathbb{K}}(V)$ is called a *unitary representation* if $\rho(g) \in \mathrm{U}(V)$ for all $g \in G$. We then write $\rho : G \to \mathrm{U}(V)$.

In this whole chapter, the space V of a unitary representation is supposed to be a *Hilbert space*, instead of just a pre-Hilbert space. Only in chapter 4 will we consider unitary representations on pre-Hilbert spaces. Note that all finite-dimensional pre-Hilbert spaces are already complete by Proposition A.2.16, so in these cases, there is no difference. The same proposition also shows that for finite-dimensional unitary representations, we can ignore the topological closedness condition in order to check whether it is irreducible. It will later turn out that all irreducible representations of a compact group are automatically finite-dimensional anyway, see Proposition 2.2.8, so this further simplifies our considerations.

As before with the unitary group, a unitary representation is actually called "orthogonal representation" when the field is the real numbers $\mathbb{R}$. $\mathrm{U}(V)$ is then replaced by $\mathrm{O}(V)$. We again stick with $\mathrm{U}(V)$ whenever the field $\mathbb{K}$ is not specified.

Definition 2.1.18 (Isomorphism of Unitary Representations)**.** Let $\rho : G \to \mathrm{U}(V)$, $\rho' : G \to \mathrm{U}(V')$ be unitary representations and $f : V \to V'$ an intertwiner. f is called an *isomorphism* (of unitary representations) if, additionally, f is a unitary transformation. The representations are then called *isomorphic*. For this, we write $\rho \cong \rho'$ or $V \cong V'$ depending on whether we want to emphasize the representations or the underlying vector spaces.

We note the following, which we will frequently use: due to the unitarity of $\rho(g)$ for a unitary representation ρ, we have $\rho(g)^* = \rho(g)^{-1}$, i.e. the adjoint is the inverse. Adjoints are defined in Definition A.2.11 and this statement is proven more generally in Proposition A.2.13. Overall, this means that $\langle \rho(g)(v) | w \rangle = \langle v | \rho(g)^{-1}(w) \rangle$ for all v, w and g.

In the end, it will turn out that the Peter-Weyl Theorem which we aim at is exclusively a statement about unitary representations. One may then wonder whether this is too restrictive. After all, the representations that we consider for steerable CNNs (with precise definitions given in Section 3.1) are not necessarily unitary, and so it is not immediately obvious how the Peter-Weyl Theorem will be able to help for those. However, as it turns out, all linear representations on finite-dimensional spaces *can* be considered as unitary, and so the theory applies. We will discuss this in Proposition 2.1.20 once we understand Haar measures on compact groups.

2.1.3. The Haar Measure, the Regular Representation and the Peter-Weyl Theorem

Now that we have introduced many notions in the representation theory of compact groups, we can formulate the most important result, the *Peter-Weyl Theorem* that we will use throughout this work. In the next section, we will then go through a step-by-step proof of this theorem. The material in this section is based on Kowalski [15], Nachbin and Bechtolsheim [16] and Knapp [14]. We thank Stefan Dawydiak for a discussion about the Peter-Weyl Theorem over the real numbers [17].

We assume that the reader knows what a measure is [18]. Let G be a compact group. A standard result is that there exists a measure μ on G, called a *Haar measure* that, among other properties, fulfills the following:

1. $\mu(S)$ can be evaluated for all Borel sets $S \subseteq G$. Here, the Borel sets form the smallest so-called σ-algebra that contains all the open sets.

2. In particular, we can evaluate $\mu(S)$ for all open or closed sets $S \subseteq G$.

3. The Haar measure is normalized: $\mu(G) = 1$.

4. μ is left and right invariant: $\mu(gS) = \mu(S) = \mu(Sg)$ for all $g \in G$ and S measurable.

5. μ is inversion invariant: $\mu(S^{-1}) = \mu(S)$ for all S measurable.

These properties then translate into properties of the associated *Haar integral*: let $f : G \to \mathbb{K}$ be integrable with respect to μ, then we obtain:

1. $\int_G 1 dg = 1$ for the constant function with value 1.
2. $\int_G f(hg)dg = \int_G f(g)dg = \int_G f(gh)dg$ for all $h \in G$.
3. $\int_G f(g^{-1})dg = \int_G f(g)dg$.

Example 2.1.19 (Finite Groups)**.** If G is a finite group with n elements, then the Haar measure is just the normalized counting measure which assigns $\mu(g) = \frac{1}{n}$ for all $g \in G$. Each function $f : G \to \mathbb{K}$ is then integrable, and its integral is just given by

$$\int_G f(g)dg = \frac{1}{n} \sum_{g \in G} f(g).$$

In this special case, one can easily verify all properties of Haar measures and Haar integrals stated above.

With this measure defined, we can already understand why all linear representations on finite-dimensional spaces can be considered as unitary:

Proposition 2.1.20. *Let $\rho : G \to \mathrm{Aut}_{\mathbb{K}}(V)$ be a linear representation on a finite-dimensional space V. Then there exists a scalar product $\langle \cdot | \cdot \rangle_\rho : V \times V \to \mathbb{K}$ that makes $(V, \langle \cdot | \cdot \rangle)$ a Hilbert space and such that ρ becomes a unitary representation with respect to this scalar product.*

Proof. Since V is finite-dimensional, there is an isomorphism of vector spaces to some $\mathbb{K}^n$. Consequently, there is *some* scalar product $\langle \cdot | \cdot \rangle : V \times V \to \mathbb{K}$ that makes V a Hilbert space. However, this scalar product does not necessarily make ρ a unitary representation. However, we can define $\langle \cdot | \cdot \rangle_\rho : V \times V \to \mathbb{K}$ by

$$\langle v | w \rangle_\rho \coloneqq \int_G \langle \rho(g)(v) | \rho(g)(w) \rangle \, dg.$$

That this integral exists is due to the continuity of linear representations and since also the scalar product is continuous by Proposition A.2.7. It can easily be checked that this construction makes V a Hilbert space. And due to the right invariance of the Haar measure, we can check that ρ is a unitary representation with respect to this scalar product. Namely, for arbitrary $g' \in G$ we have:

$$\begin{aligned}\big\langle \rho(g')(v) \big| \rho(g')(w) \big\rangle_\rho &= \int_G \big\langle \rho(g)\rho(g')v \big| \rho(g)\rho(g')w \big\rangle dg \\ &= \int_G \big\langle \rho(gg')(v) \big| \rho(gg')(w) \big\rangle dg \\ &= \int_G \big\langle \rho(g)(v) \big| \rho(g)(w) \big\rangle dg \\ &= \langle v | w \rangle_\rho .\end{aligned}$$

□

Now, for a measure space Y with corresponding measure μ, we can consider the space of square-integrable functions on Y with values in $\mathbb{K}$, denoted $L^2_{\mathbb{K}}(Y)$ (the measure is omitted in the notation since there is usually no ambiguity). In these spaces, functions are identified if they coincide on a set with measure 0. $L^2_{\mathbb{K}}(Y)$ is clearly a vector space over $\mathbb{K}$, but it turns out that it can even be considered to be a Hilbert space as follows:

$$\langle f | g \rangle := \int_Y \overline{f(y)} g(y) dy.$$

Here, the overline means complex conjugation. The Hilbert space properties are easily verified.
In particular, one can consider the space $L^2_{\mathbb{K}}(G)$ of square-integrable functions on the group G itself. Now the claim is that $L^2_{\mathbb{K}}(G)$ can actually be equipped with a prototypical structure as a unitary representation over G which makes this space, in some sense, "universal among unitary representations". This works with the following canonical representation, called the *regular representation*:

$$\lambda : G \to \mathrm{U}(L^2_{\mathbb{K}}(G)), \; [\lambda(g)(f)]\,(g') := f(g^{-1}g').$$

continuity of this map is non-trivial and is for example shown in Knapp [14]. However, the more algebraic properties of being a unitary representation are quite easy to appreciate. First of all, we clearly see that λ is a group homomorphism mapping each group element to a linear automorphism. And finally, the unitarity of this representation can be understood as a direct consequence of the properties of the Haar measure, where we notably make only use of the left-invariance:

$$\begin{aligned}\langle \lambda(g)(f) | \lambda(g)(h) \rangle &= \int_G \overline{[\lambda(g)(f)]\,(g')} \cdot [\lambda(g)(h)]\,(g') dg' \\ &= \int_G \overline{f(g^{-1}g')} \cdot h(g^{-1}g') dg'\end{aligned}$$

$$
\begin{aligned}
&= \int_G \overline{f(g')} h(g') dg' \\
&= \langle f | h \rangle .
\end{aligned}
$$

We saw in Example 2.1.9 that G is a homogeneous space with respect to the action on itself. We can now ask whether these constructions can also work if X is an arbitrary homogeneous space of G. This requires us to define a suitable measure on X. This is indeed possible. For a fixed element $x^* \in X$, denote the stabilizer subgroup by $H = G_{x^*} \subseteq G$. Then the Hausdorff property of X allows to write down a homeomorphism between X and G/H, which in turn will allow us to use a canonical measure on G/H that we study below. We denote cosets $gH \in G/H$ by $[g]$.

Lemma 2.1.21. *Let X be a homogeneous space of the compact group G and H the stabilizer subgroup of a fixed element $x^* \in X$. Then the map*

$$\varphi : G/H \to X, \; [g] \mapsto gx^*$$

is a homeomorphism. Furthermore, H is topologically closed.

Proof. Let $\tilde{\varphi} : G \to X$, $g \mapsto gx^*$. This map is equal to the composition of the maps $G \to G \times X$, $g \mapsto (g, x^*)$ and $G \times X \to X$, $(g, x) \mapsto gx$. Both these are continuous, and thus $\tilde{\varphi}$ is continuous as well. Furthermore, note that if $g^{-1}g' \in H$, then there is $h \in H$ such that $g' = gh$, and thus

$$\tilde{\varphi}(g') = \tilde{\varphi}(gh) = (gh)x^* = g(hx^*) = gx^* = \tilde{\varphi}(g)$$

which means that by Proposition A.1.12, the map $\varphi : G/H \to X$, $[g] \mapsto gx^*$ is a well-defined continuous map. It is surjective since the action is transitive by definition of a homogeneous space. Furthermore, it is injective since if $gx^* = g'x^*$ then $x^* = (g^{-1}g')x^*$ and thus $g^{-1}g' \in H$, which means $[g] = [g']$.
Overall, φ is a continuous bijective map from G/H to X. Furthermore, G/H is compact since it is the continuous image of the compact group G under the projection $G \to G/H$, see Proposition A.1.8. Since X is Hausdorff by definition of homogeneous spaces, φ is a homeomorphism according to Proposition A.1.9.
Now, since X is Hausdorff and φ is a homeomorphism, it follows that G/H is Hausdorff as well. Then, necessarily, H is a topologically closed subgroup of G, see Bourbaki [19], Chapter III, Section 2.5, Proposition 13. □

Every space G/H where H is topologically closed allows a measure μ with very similar properties to those of G [16]. Since the stabilizer H is closed and $X \cong G/H$ by Lemma 2.1.21, we can do these constructions for X as well, as we outline now. The only properties that we now miss are the right-invariance and inversion-invariance: We simply can't ask for them since G does not naturally act on X from the right and since we cannot invert elements in X. But left-invariance does hold and this means that

$$\lambda : G \to L^2_{\mathbb{K}}(X), \; [\lambda(g)(f)](x) := f(g^{-1}x)$$

makes $L^2_{\mathbb{K}}(X)$ a unitary representation over G, as can be shown in the exact same way as for $L^2_{\mathbb{K}}(G)$.
Let $\hat{G}$ be the set of isomorphism classes of irreducible unitary representations over G. Furthermore, let $\rho_l : G \to V_l$ be a fixed representative of such an isomorphism class $l \in \hat{G}$. We write isomorphism classes as "l" (and later also j and J) in order to bring to mind quantum numbers used in quantum mechanics. Recall from linear algebra that a countable sum of subspaces of a vector space is called *direct* if no nontrivial subspace of any of the considered spaces is contained in the sum of all the other considered spaces[1]. Furthermore, recall that two subspaces $U, W \subseteq V$ of a Hilbert space V are called perpendicular or orthogonal if $\langle u|w\rangle = 0$ for all $u \in U$ and $w \in W$. We then write $U \perp W$. We can now formulate the Peter-Weyl Theorem. Intuitively, it says that $L^2_{\mathbb{K}}(X)$ splits into an orthogonal direct sum of the irreducible unitary representations, where each irreducible unitary representation appears maximally as often as its own dimension (and may not appear at all):

Theorem 2.1.22 (Peter-Weyl Theorem). *Let G be a compact group. Let X be a homogeneous space. There are numbers $m_l \in \mathbb{N}_{\geq 0}$ for all $l \in \hat{G}$ and closed-invariant subspaces $V_{li} \subseteq L^2_{\mathbb{K}}(X)$ for all $l \in \hat{G}$ and $i \in \{1, \dots, m_l\}$ such that the following hold:*

1. *$V_{li} \cong V_l$ as unitary representations for all i and l.*
2. *$m_l \leq \dim(V_l) < \infty$ for all l.*
3. *$V_{li} \perp V_{l'j}$ whenever $l \neq l'$ or $i \neq j$.*
4. *$\bigoplus_{l \in \hat{G}} \bigoplus_{i=1}^{m_l} V_{li}$ is topologically dense in $L^2_{\mathbb{K}}(X)$, written $L^2_{\mathbb{K}}(X) = \widehat{\bigoplus}_{l \in \hat{G}} \bigoplus_{i=1}^{m_l} V_{li}$.*

Now additionally consider G as a homogeneous space of itself. Then the same holds for $L^2_{\mathbb{K}}(G)$ as well, with numbers $n_l \leq \dim(V_l) < \infty$. We additionally have the following:

1. *$m_l \leq n_l$.*
2. *If $\mathbb{K} = \mathbb{C}$, then $n_l = \dim(V_l)$.*

Note that the representative V_l is not assumed to be embedded in $L^2_{\mathbb{K}}(X)$. It is just isomorphic, as a unitary representation, to each of the $V_{li} \subseteq L^2_{\mathbb{K}}(X)$.

Example 2.1.23. For $G = \mathrm{U}(1)$ and $\mathbb{K} = \mathbb{C}$ we have $L^2_{\mathbb{C}}(\mathrm{U}(1)) = \widehat{\bigoplus}_{l \in \mathbb{Z}} V_{l1}$ and all irreducible representations V_l are 1-dimensional.
For $G = \mathrm{SO}(2)$ (which is as a group isomorphic to $\mathrm{U}(1)$) and $\mathbb{K} = \mathbb{R}$, we obtain $L^2_{\mathbb{R}}(\mathrm{SO}(2)) = \widehat{\bigoplus}_{l \geq 0} V_{l1}$, and all irreducible representations V_l with $l \geq 1$ are two-dimensional, whereas V_0 is one-dimensional. Thus, here we see an example where the multiplicity of most irreducible representations in the regular representation is 1 and therefore smaller than their dimension, which cannot happen for representations over the complex numbers.

[1]For a vector space V and subspaces $(U_i)_{i \in I}$, their sum $\sum_{i \in I} U_i$ is the set of sums $\sum_{i \in J} u_i$ with $J \subseteq I$ finite and $u_i \in U_i$ for all i. It is itself a subspace of V.

Both of these results are standard results in Fourier analysis. These examples are discussed in more detail, especially with respect to their applications in deep learning, in Section 6.1 and 6.2.

2.2. A Proof of the Peter-Weyl Theorem

This section presents a proof of the Peter-Weyl Theorem, as formulated in Theorem 2.1.22. We mostly skip the *analytical* parts of the proof[2], since they are well-presented in the literature and clearly work over both the real and complex numbers. However, the more *algebraic* parts of the proof usually make use of the property of the complex numbers to be algebraically closed, which does not hold for the real numbers. This is invoked usually both in the proof of a version of Schur's Lemma, as well as in proving Schur's orthogonality. We therefore carefully adapt the proof of the Peter-Weyl Theorem in the literature so that it also works over the real numbers, and formulate and prove versions of Schur's Lemma 2.2.6 and Schur's orthogonality 2.2.7 that work in general.

This section can be skipped completely if the interest is mainly in the applications of the Peter-Weyl Theorem. In this case, the reader is advised to directly move on to Chapter 3.

We note the following convention that applies to this section: for all unitary representations $\rho : G \to \mathrm{U}(V)$ that we consider here, V is a Hilbert space (instead of just a pre-Hilbert space).

2.2.1. Density of Matrix Coefficients

An important ingredient in the construction of the spaces V_{li} that appear in the formulation of the Peter-Weyl Theorem 2.1.22 are matrix coefficients, which together generate those spaces in case that one considers the regular representation on $L^2_{\mathbb{K}}(G)$.

Definition 2.2.1 (Matrix coefficients)**.** Let $\rho : G \to \mathrm{U}(V)$ be a unitary representation. A *matrix coefficient* is any function of the form

$$\rho^{uv} : G \to \mathbb{K}, \; g \mapsto \overline{\langle u|\rho(g)(v)\rangle}$$

for arbitrary $u, v \in V$.

The term "matrix coefficient" comes from the analogy to matrix elements of linear maps between pre-Hilbert spaces of which orthonormal bases are fixed. Later, in Definition 4.1.9 we will also define the notion of "matrix elements" separately. The term "matrix coefficient" only applies to unitary representations.

Remark 2.2.2. By definition of linear representations, the function $g \mapsto \rho(g)(v)$ is continuous. Thus, since scalar products of Hilbert spaces are also continuous as functions on $V \times V$, see Proposition A.2.7, every matrix coefficient $\rho^{uv} : G \to \mathbb{K}$ is continuous.

[2] I.e., those parts that deal with approximations of square-integrable functions by matrix elements.

As a continuous function on a compact space, it is of course also square-integrable, i.e. $\rho^{uv} \in L^2_{\mathbb{K}}(G)$. The Peter-Weyl Theorem basically asserts that these matrix coefficients can be considered as the building blocks of all square-integrable functions.
Furthermore, one may wonder why there is a complex conjugation in the definition. The reason for this is that, otherwise, the isomorphism that we will construct in Proposition 2.2.12 is not linear but conjugate linear. The reason why this can nevertheless be called a matrix coefficient is that this actually *is* the matrix coefficient (without complex conjugation) on a conjugate Hilbert space, as explained in the next Proposition, which we took from Williams [20].

Proposition 2.2.3. *Let $\rho : G \to \mathrm{U}(V)$ be a unitary representation on a Hilbert space V with scalar multiplication $\cdot_V$ and scalar product $\langle\cdot|\cdot\rangle_V$. We have the following:*

1. *$\tilde{V} := V$ (equality as abelian groups) with $\alpha \cdot_{\tilde{V}} v := \overline{\alpha} \cdot_V v$ and $\langle u|v\rangle_{\tilde{V}} := \overline{\langle u|v\rangle}$ is again a Hilbert space, the so-called* conjugate Hilbert space *of V.*

2. *$\tilde{\rho} : G \to \mathrm{U}(\tilde{V})$ with $\tilde{\rho}(g) := \rho(g)$ is again a unitary representation.*

3. *For the matrix coefficients, we have $\tilde{\rho}^{uv}(g) = \overline{\rho^{uv}(g)}$.*

Proof. All these assertions are easy to check. As a demonstration, we do 3:

$$\tilde{\rho}^{uv}(g) = \overline{\langle u|\tilde{\rho}(g)(v)\rangle_{\tilde{V}}} = \overline{\overline{\langle u|\rho(g)(v)\rangle}_V} = \overline{\rho^{uv}(g)}.$$

That's what we wanted to show. □

As a consequence of this proposition, the matrix coefficient $\rho^{uv}(g)$ is equal to $\overline{\tilde{\rho}^{uv}(g)}$, thus being a "matrix coefficient without complex conjugation above the scalar product" of the conjugate unitary representation.

Theorem 2.2.4. *The linear span of the matrix-coefficients of finite-dimensional, unitary, irreducible representations of G are dense in $L^2_{\mathbb{K}}(G)$ for all compact groups G.*

Proof. For $\mathbb{K} = \mathbb{C}$, this is shown in Knapp [14]. The same proof, without adaptions, also works for $\mathbb{K} = \mathbb{R}$. Note that the cited proof uses a definition of matrix coefficients without the complex conjugation. However, Proposition 2.2.3 shows those span the same space, and thus we can apply it to our situation. □

2.2.2. Schur's Lemma, Schur's Orthogonality and Consequences

In this section, we state and prove versions of Schur's Lemma and Schur's Orthogonality [14] that are valid for both $\mathbb{K} = \mathbb{R}$ and $\mathbb{K} = \mathbb{C}$.

Lemma 2.2.5. *Let $\rho : G \to \mathrm{U}(V)$ and $\rho' : G \to \mathrm{U}(V')$ be unitary representations. Furthermore, let $f : V \to V'$ be an intertwiner. Then the adjoint $f^* : V' \to V$ is also an intertwiner.*

Proof. The adjoint $f^* : V' \to V$ is the unique continuous linear function from V' to V such that, for all $v \in V$ and $v' \in V'$, we have

$$\langle f(v)|v'\rangle = \langle v|f^*(v')\rangle \,.$$

This always exists according to Definition A.2.11. Note that with f being an intertwiner and using the unitarity of the representations, we obtain for all $g \in G, v \in V$ and $v' \in V'$:

$$\begin{aligned}\langle v|\rho(g)f^*(v')\rangle &= \left\langle \rho(g^{-1})(v)\middle|f^*(v')\right\rangle \\ &= \left\langle f\rho(g^{-1})(v)\middle|v'\right\rangle \\ &= \left\langle \rho'(g^{-1})f(v)\middle|v'\right\rangle \\ &= \langle f(v)|\rho'(g)(v')\rangle \\ &= \langle v|f^*\rho'(g)(v')\rangle\end{aligned}$$

from which we deduce $\rho(g)f^* = f^*\rho'(g)$ from Proposition A.2.14 for all $g \in G$, i.e. f^* is an intertwiner. □

Lemma 2.2.6 (Schur's Lemma for unitary Representations)**.** *Assume $\rho : G \to \mathrm{U}(V)$ and $\rho' : G \to \mathrm{U}(V')$ are irreducible unitary representations with V finite-dimensional. Also assume that $f : V \to V'$ is an intertwiner. Then either $f = 0$ or there is $\mu \in \mathbb{R}_{>0}$ such that μf is an isomorphism.*

Proof. For this proof, we follow the exposition of Tao [21]. We thank Terrence Tao for confirming in the discussion below his blogpost that this lemma can also be proven over the real numbers.
Let $f^* : V' \to V$ be the adjoint of f, which is also an intertwiner by Lemma 2.2.5. Now, set $\varphi := f^* \circ f : V \to V$. As a composition of intertwiners, φ is also an intertwiner. Furthermore, for arbitrary composable continuous linear functions between Hilbert spaces one always has $(g \circ h)^* = h^* \circ g^*$ and $(g^*)^* = g$, which easily follows from the definition and uniqueness of adjoints. Consequently, we have

$$\varphi^* = (f^* \circ f)^* = f^* \circ (f^*)^* = f^* \circ f = \varphi,$$

and so φ is self-adjoint. Thus, $\langle \varphi(v)|w\rangle = \langle v|\varphi(w)\rangle$ for all $v, w \in V$, from which we conclude that the matrix of φ corresponding to any orthonormal basis of V is Hermitian or, if $\mathbb{K} = \mathbb{R}$, even symmetric. Such an orthonormal basis exists by Proposition A.2.10. From the Spectral Theorem for Hermitian or symmetric matrices [22] we conclude that φ is unitarily (or for real matrices: orthogonally) diagonalizable with only real eigenvalues. Thus, there is an orthogonal decomposition of V into eigenspaces: $V = \bigoplus_{\lambda \text{ eigenvalue}} E_\lambda(\varphi)$.
Let $E_\lambda(\varphi)$ be any eigenspace. We now claim that it is an invariant subspace of ρ. Indeed, for all $g \in G$ and $v \in E_\lambda(\varphi)$ we have since φ is an intertwiner:

$$\varphi(\rho(g)(v)) = \rho(g)(\varphi(v)) = \rho(g)(\lambda v) = \lambda\rho(g)(v).$$

Since V is finite-dimensional, $E_\lambda(\varphi)$ is topologically closed by Proposition A.2.16, and since V is irreducible, we necessarily have $E_\lambda(\varphi) = 0$ or $E_\lambda(\varphi) = V$. Since not all eigenspaces can be zero, we conclude that there is an eigenvalue λ with $E_\lambda(\varphi) = V$, meaning $\varphi = \lambda\, \mathrm{id}_V$.
Assume $f \neq 0$. We now claim that $\lambda > 0$. Indeed, note that for all $v \in V$ we have

$$\begin{aligned}\lambda \|v\|^2 &= \langle \varphi(v)|v\rangle \\ &= \langle f^* \circ f(v)|v\rangle \\ &= \langle f(v)|f(v)\rangle \\ &= \|f(v)\|^2.\end{aligned}$$

Thus, if $v \in V$ is any vector with $f(v) \neq 0$, then we obtain $\lambda = \left(\frac{\|f(v)\|}{\|v\|}\right)^2 > 0$.
Now define $g : V \to V'$ as $g = \lambda^{-\frac{1}{2}} f$. g is clearly still an intertwiner. We can also show it is an isometry:

$$\begin{aligned}\langle g(v)|g(w)\rangle &= \lambda^{-1} \langle f(v)|f(w)\rangle \\ &= \lambda^{-1} \langle \varphi(v)|w\rangle \\ &= \lambda^{-1}\lambda \langle v|w\rangle \\ &= \langle v|w\rangle .\end{aligned}$$

Note that since V' is irreducible and $f(V) \subseteq V'$ topologically closed due to V being finite-dimensional, we necessarily have that f is surjective. Thus, we have shown that μf with $\mu := \lambda^{-\frac{1}{2}}$ is an isomorphism of unitary representations. $\square$

Proposition 2.2.7 (Schur's Orthogonality). *Let $\rho : G \to \mathrm{U}(V)$ and $\rho' : G \to \mathrm{U}(V')$ be nonisomorphic irreducible unitary representations of the compact group G, of which at least one is finite-dimensional. Let ρ^{uv} and $\rho'^{u'v'}$ be matrix coefficients of them, which are functions in $L^2_{\mathbb{K}}(G)$ due to their continuity. Then they are orthogonal, i.e. $\left\langle \rho^{uv} \middle| \rho'^{u'v'} \right\rangle = 0$.*

Proof. Without loss of generality, we can assume V' to be finite-dimensional. Assume that $l : V' \to V$ is *any* linear function. We can associate to it the function $f : V' \to V$ given by

$$f(w') := \int_G \rho(g) l \rho'(g)^{-1} w' dg.$$

For all $h \in G$ we have

$$\begin{aligned}\rho(h) f \rho'(h)^{-1} &= \int_G \rho(h)\rho(g) l \rho'(g)^{-1} \rho'(h)^{-1} dg \\ &= \int_G \rho(hg) l \rho'(hg)^{-1} dg \\ &= \int_G \rho(g) l \rho'(g)^{-1} dg \\ &= f,\end{aligned}$$

and thus $\rho(h)f = f\rho'(h)$, which means that f is an intertwiner. In this derivation, $\rho(h)$ could be put insight the integral since $\rho(h)$ is continuous and an integral is a limit over finite sums, which commutes with the continuous $\rho(h)$. By Schur's Lemma 2.2.6, we necessarily have $f = 0$. Now look at the specific linear function $l : V' \to V$ given by $l(w') := \langle v'|w'\rangle\, v$ with the fixed vectors v, v' corresponding to the matrix coefficients. We obtain $f = 0$, for f defined as before, and thus:

$$\begin{aligned}
0 = \langle u|f(u')\rangle &= \left\langle u\middle| \int_G \rho(g) l \rho'(g)^{-1}(u') dg \right\rangle \\
&= \int_G \left\langle u\middle|\rho(g) l \rho'(g)^{-1}(u')\right\rangle dg \\
&= \int_G \left\langle u\middle|\rho(g)\left[\left\langle v'\middle|\rho'(g)^{-1}(u')\right\rangle v\right]\right\rangle dg \\
&= \int_G \langle u|\rho(g)(v)\rangle \cdot \left\langle v'\middle|\rho'(g)^{-1}(u')\right\rangle dg \\
&= \int_G \overline{\overline{\langle u|\rho(g)(v)\rangle}} \cdot \overline{\langle u'|\rho'(g)(v')\rangle} dg \\
&= \int_G \overline{\rho^{uv}(g)} \rho'^{u'v'}(g) dg \\
&= \left\langle \rho^{uv}\middle|\rho'^{u'v'}\right\rangle
\end{aligned}$$

In this derivation, the integral could be put out of the scalar product since the scalar product is continuous, see Proposition A.2.7, and since integrals are certain limits over finite sums, with which the scalar product commutes. □

Note that there are more general Schur's orthogonality relations in the case that $\mathbb{K} = \mathbb{C}$, see Knapp [14], Corollary 4.10. These then engage with the matrix coefficients of one and the same representation. This, together with a version of Schur's Lemma that only holds over $\mathbb{C}$ leads to the strengthening of the Peter-Weyl Theorem that shows that the multiplicities n_l are given by $\dim(V_l)$.

Proposition 2.2.8. *All irreducible unitary representations of a compact group G are finite-dimensional.*

Proof. Assume $\rho : G \to \mathrm{U}(V)$ was an irreducible unitary representation on an infinite-dimensional space V. Let ρ^{uv} be any of its matrix coefficients. By Proposition 2.2.7, and since an infinite-dimensional representation can never be isomorphic to a finite-dimensional representation, ρ^{uv} is perpendicular to all matrix coefficients of finite-dimensional irreducible unitary representations. Due to the linearity of the scalar product, ρ^{uv} is perpendicular to the whole linear span of these matrix coefficients and thus to the topological closure of this span. The last step follows from the continuity of the scalar product, see Proposition A.2.7. By Theorem 2.2.4 this closure is the whole space $L^2_{\mathbb{K}}(G)$. Therefore, ρ^{uv} is even perpendicular to itself, and thus $\rho^{uv} = 0$.
Overall, for arbitrary $u, v \in V$ and $g \in G$ we obtain $0 = \rho^{uv}(g) = \overline{\langle u|\rho(g)(v)\rangle}$ and thus (by setting $u = \rho(g)(v)$) $\rho(g)(v) = 0$ and consequently $\rho(g) = 0$. We obtain

$\rho = 0$, a contradiction. Thus infinite-dimensional irreducible unitary representations cannot exist. $\square$

As a consequence, we mention that the finiteness conditions in Schur's Lemma and Schur's Orthogonality were not necessary to state since all irreducible unitary representations are finite-dimensional anyway. We obtain from this and from Schur's Lemma 2.2.6 that isomorphism classes and equivalence classes of irreducible unitary representations are one and the same.

2.2.3. A Proof of the Peter-Weyl Theorem for the Regular Representation

In this section, we engage with the Peter-Weyl Theorem for the regular representation on $L^2_{\mathbb{K}}(G)$. The case of $L^2_{\mathbb{K}}(X)$ for a homogeneous space X will be dealt with in Section 2.2.4. The core arguments in the proofs of this section are adapted from Williams [20]. As before, let $\hat{G}$ be the set of isomorphism classes of irreducible representations of G. For $l \in \hat{G}$ let ρ_l be a representative for the isomorphism class l. Furthermore, for each $\rho_l : G \to \mathrm{U}(V_l)$, let $v_l^1, \ldots, v_l^{\dim(V_l)}$ be an arbitrary orthonormal basis, which exists due to Proposition A.2.10 (mostly written without the superscript, i.e. as $v^1, v^2, \ldots,$ if the corresponding isomorphism class is clear). Denote $\rho_l^{ij} := \rho_l^{v^i v^j}$. Remember that matrix coefficients of unitary representations are continuous by Remark 2.2.2, and thus functions in $L^2_{\mathbb{K}}(G)$. Then, let $\mathcal{E} \subseteq L^2_{\mathbb{K}}(G)$ be the linear span of the matrix coefficients of all irreducible unitary representations. In the next Lemma, we want to show that $\mathcal{E}$ is already spanned by the matrix coefficients corresponding to representatives of isomorphism classes and their orthonormal bases:

Lemma 2.2.9. *We have*

$$\mathcal{E} = \mathrm{span}_{\mathbb{K}} \left\{ \rho_l^{ij} \mid l \in \hat{G}, i, j \in \{1, \ldots, \dim(V_l)\} \right\}.$$

Proof. First, we show that isomorphic representations don't add distinct matrix coefficients. Thus, let $\rho \cong \rho_l$ and let $f : V \to V_l$ be the corresponding isomorphism. Then we have $\rho_l(g) \circ f = f \circ \rho(g)$ and thus, since f is a unitary transformation, $\rho(g) = f^* \circ \rho_l(g) \circ f$, for all $g \in G$, see Proposition A.2.13. Now let $u, v \in V$ be arbitrary. We obtain

$$\begin{aligned}
\rho^{uv}(g) &= \overline{\langle u | \rho(g)(v) \rangle} \\
&= \overline{\langle u | f^* \rho_l(g) f(v) \rangle} \\
&= \overline{\langle f(u) | \rho_l(g)(f(v)) \rangle} \\
&= \rho_l^{f(u) f(v)}(g),
\end{aligned}$$

which proves the first claim. Now we want to show that we only need to consider the ρ_l^{ij}. Thus, let $u, v \in V_l$ be arbitrary. They allow for linear combinations

$$u = \sum_i \lambda^i v^i, \; v = \sum_i \mu^i v^i$$

with coefficients $\lambda^i, \mu^i \in \mathbb{K}$. We obtain:

$$\begin{aligned}
\rho_l^{uv}(g) &= \overline{\langle u|\rho_l(g)(v)\rangle} \\
&= \sum\nolimits_i \sum\nolimits_j \lambda^i \overline{\mu^j} \cdot \overline{\langle v^i|\rho_l(g)(v^j)\rangle} \\
&= \left(\sum\nolimits_i \sum\nolimits_j \lambda^i \overline{\mu^j} \rho_l^{ij}\right)(g),
\end{aligned}$$

thus showing that ρ_l^{uv} is in the linear span of the matrix coefficients corresponding to the orthonormal basis. This concludes the proof. □

For an isomorphism class $l \in \hat{G}$, let $\mathcal{E}_l := \operatorname{span}\left\{\rho_l^{ij} \mid i,j \in \{1,\dots,\dim(V_l)\}\right\} \subseteq L^2_{\mathbb{K}}(G)$ be the linear subspace of $\mathcal{E}$ generated by matrix coefficients corresponding to l. Let furthermore for all j the space $\mathcal{E}_l^j \subseteq \mathcal{E}_l$ be the subspace generated by all ρ_l^{ij} for $i \in \{1,\dots,\dim(V_l)\}$. In the next lemma, we prove that these are actually closed subrepresentations of the regular representation.

Lemma 2.2.10. *For $j \in \{1,\dots,\dim(V_l)\}$, $\mathcal{E}_l^j$ is a closed invariant subspace of $L^2_{\mathbb{K}}(G)$. In particular, $\mathcal{E}_l$ is a closed invariant subspace of $L^2_{\mathbb{K}}(G)$.*

Proof. Closedness follows immediately since this space is finite-dimensional and thus complete, see Proposition A.2.16. We need to show that $\lambda(g)\rho_l^{ij} \in \mathcal{E}_l^j$ for all $g \in G$ and all i,j. We can compute this directly:

$$\begin{aligned}
\left(\lambda(g)\rho_l^{ij}\right)(g') &= \rho_l^{ij}(g^{-1}g') \\
&= \overline{\langle v^i|\rho_l(g^{-1}g')(v^j)\rangle} \\
&= \overline{\langle \rho_l(g)(v^i)|\rho_l(g')(v^j)\rangle} \\
&= \overline{\left\langle \sum\nolimits_{i'} \langle v^{i'}|\rho_l(g)(v^i)\rangle\, v^{i'} \middle| \rho_l(g')(v^j)\right\rangle} \\
&= \sum\nolimits_{i'} \langle v^{i'}|\rho_l(g)(v^i)\rangle \cdot \overline{\langle v^{i'}|\rho_l(g')(v^j)\rangle} \\
&= \sum\nolimits_{i'} \overline{\langle v^i|\rho_l(g^{-1})(v^{i'})\rangle} \rho_l^{i'j}(g') \\
&= \left(\sum\nolimits_{i'} \rho_l^{ii'}(g^{-1})\rho_l^{i'j}\right)(g')
\end{aligned}$$

where the coefficients $\rho_l^{ii'}(g^{-1})$ do not depend on g'. Consequently, $\lambda(g)\rho_l^{ij} \in \mathcal{E}_l^j$. □

Lemma 2.2.11. *Let $\rho : G \to \mathrm{U}(V)$ and $\rho' : G \to \mathrm{U}(V')$ be unitary representations, ρ being irreducible and $V' \neq 0$. Furthermore, assume that $f : V \to V'$ is a surjective intertwiner. Then V' is also irreducible and f an equivalence.*

Proof. Assume by contradiction that V' is reducible. Thus, there is a nontrivial closed invariant subspace $0 \subsetneq W \subsetneq V'$. Now the following can easily be checked:

1. $0 \subsetneq f^{-1}(W) \subsetneq V$.

2. $f^{-1}(W)$ is an invariant subspace of V.

3. $f^{-1}(W)$ is a closed subset of V.

Once we have this, we have a contradiction to the fact that V is irreducible.
1 and 2 can be checked by the reader, and 3 follows since V is, as an irreducible representation, finite-dimensional by Proposition 2.2.8 and thus every subspace is closed by Proposition A.2.16.
Therefore, we know that V' is irreducible. Now use Schur's Lemma 2.2.6 to conclude that f, being nonzero, necessarily is an equivalence. □

Proposition 2.2.12. *There is an equivalence of representations $f_l^j : V_l \to \mathcal{E}_l^j$ given on the orthonormal basis by $f_l^j(v^i) = \rho_l^{ij}$. Consequently, there is an isomorphism $V_l \cong \mathcal{E}_l^j$ of unitary representations.*

Proof. We need to show that f_l^j is equivariant. Using the result of the derivation of Lemma 2.2.10, we compute

$$\begin{aligned} f_l^j(\rho_l(g)(v^i)) &= f_l^j\left(\sum\nolimits_{i'} \langle v^{i'} | \rho_l(g)(v^i)\rangle v^{i'}\right) \\ &= \sum\nolimits_{i'} \left\langle v^{i'} \middle| \rho_l(g)(v^i)\right\rangle f_l^j(v^{i'}) \\ &= \sum\nolimits_{i'} \overline{\langle v^i | \rho_l(g^{-1})(v^{i'})\rangle} \rho_l^{i'j} \\ &= \sum\nolimits_{i'} \rho_l^{ii'}(g^{-1}) \rho_l^{i'j} \\ &= \lambda(g)\rho_l^{ij} \\ &= \lambda(g)\left(f_l^j(v^i)\right), \end{aligned}$$

so $f_l^j \circ \rho_l(g) = \lambda(g) \circ f_l^j$ for all $g \in G$, which is what we wanted to show. That f is an intertwiner also requires it to be continuous: this follows since V_l is finite-dimensional, and so all linear functions on it are continuous.
Now, that f_l^j is even an equivalence follows from Lemma 2.2.11 by noting that $\mathcal{E}_l^j \neq 0$. Indeed, if it was zero then we would have $\rho_l^{ij}(g) = 0$ for all i, and thus $\rho(g)$ would not be invertible, in contrast that it is a unitary automorphism.
Thus, there is even an isomorphism $V_l \cong \mathcal{E}_l^j$ by Schur's Lemma 2.2.6. □

Lemma 2.2.13. *Let $\rho : G \to \mathrm{U}(V)$ be a unitary representation. Let $V_1 \subseteq V$ be a subrepresentation. Then the orthogonal complement $V_1^{\perp}$ is a subrepresentation as well.*

Proof. We have $\langle v | v_1 \rangle = 0$ for all $v \in V_1^{\perp}$ and all $v_1 \in V_1$. Now, let $g \in G$ be arbitrary. From the unitarity of ρ we obtain

$$\langle \rho(g)(v) | v_1 \rangle = \left\langle v \middle| \rho(g^{-1})(v_1)\right\rangle = 0.$$

The last step follows from $\rho(g^{-1})(v_1) \in V_1$, which holds since V_1 is a subrepresentation. Overall, this shows $\rho(g)(v) \in V_1^{\perp}$ as well, and so this is a subrepresentation. □

Lemma 2.2.14. *Let $\rho : G \to \mathrm{U}(V)$ be a finite-dimensional unitary representation. Furthermore, assume that W_1, W_2 are irreducible subrepresentations. If they are not isomorphic, then they are perpendicular, i.e. $\langle w_1 | w_2 \rangle = 0$ for all $w_1 \in W_1$ and $w_2 \in W_2$.*

Proof. Let $P : V \to W_1$ be the orthogonal projection from V to W_1, defined as the adjoint of the canonical inclusion $i : W_1 \to V$, i.e. defined by the property

$$\langle w_1 | P(v) \rangle = \langle i(w_1) | v \rangle = \langle w_1 | v \rangle$$

for all $v \in V$ and $w_1 \in W_1$, see also Proposition A.2.15. We now show that P is equivariant. For all $g \in G$, $v \in V$ and $w_1 \in W_1$ we have:

$$\begin{aligned} \big\langle w_1 \big| P(\rho(g)(v)) \big\rangle &= \langle w_1 | \rho(g)(v) \rangle \\ &= \big\langle \rho(g^{-1})(w_1) \big| v \big\rangle \\ &= \big\langle \rho(g^{-1})(w_1) \big| P(v) \big\rangle \\ &= \big\langle w_1 \big| \rho(g)(P(v)) \big\rangle, \end{aligned}$$

where we used in the third step that W_1 is a subrepresentation. Since this holds for all $w_1 \in W_1$, we obtain $P(\rho(g)(v)) = \rho(g)(P(v))$ by Proposition A.2.14 and overall that P is equivariant.
In particular, also the restriction $P|_{W_2} : W_2 \to W_1$ is equivariant. Since W_1 and W_2 are not isomorphic, we obtain by Schur's Lemma 2.2.6 that $P|_{W_2} = 0$, i.e. for all $w_1 \in W_1$ and $w_2 \in W_2$ we have $\langle w_1 | w_2 \rangle = \langle w_1 | P|_{W_2}(w_2) \rangle = \langle w_1 | 0 \rangle = 0$. Thus, W_1 and W_2 are perpendicular as claimed. □

Proposition 2.2.15. *Let $\rho : G \to \mathrm{U}(V)$ be any finite-dimensional unitary representation. Then V decomposes into an orthogonal direct sum*

$$V = \bigoplus_{i=1}^{n} V_i$$

such that $V_i \subseteq V$ are irreducible subrepresentations of ρ.

Proof. Let V_1 be any irreducible subrepresentation of V: This can be obtained by noting that if V is not already irreducible (in which case $V_1 = V$), then we find a nontrivial subrepresentation $0 \subsetneq W \subsetneq V$. By iteratively proceeding with W, we eventually need to reach an irreducible representation since V is finite-dimensional.
Now, let $V_1^{\perp}$ be the orthogonal complement of V_1. From Lemma 2.2.13 we know that this is a subrepresentation of V. By induction on the dimension of V, and since $V_1^{\perp}$ has strictly smaller dimension, we can assume that $V_1^{\perp}$ already splits into an orthogonal direct sum of irreducible subrepresentations $V_1^{\perp} = \bigoplus_{i=2}^{n} V_i$, and overall, $V = \bigoplus_{i=1}^{n} V_i$ is the decomposition we were looking for. □

The following proposition will not be used now, but we make use of it later when showing that there are only finitely many basis kernels in a steerable CNN for a compact group:

Proposition 2.2.16 (Krull-Remak-Schmidt Theorem)**.** *In the situation of Proposition 2.2.15, the orthogonal direct sum decomposition is essentially unique. That is, the type and multiplicities of the irreducible direct summands is always the same.*

Proof. If one has one decomposition of V in which an irreducible representation U does not appear, then it cannot appear in *any* decomposition since U would be perpendicular to all the irreps in the decomposition of V by Lemma 2.2.14 and thus zero. Therefore, the types of irreducible representations is always the same. That the multiplicities are always the same follows by the same argument and for dimension-reasons. □

We can now finally prove The Peter-Weyl Theorem 2.1.22 for the case that $X = G$:

Proof. By Proposition 2.2.15 and Lemma 2.2.10 there is some orthogonal decomposition $\mathcal{E}_l = \bigoplus_{i=1}^{n_l} V_{li}$ into irreducible invariant subspaces. Now assume that there is an i such that $V_{li} \ncong V_l$. By Proposition 2.2.12 this means that $V_{li} \ncong \mathcal{E}_l^j$ for all $j = 1, \dots, \dim(V_l)$. By Lemma 2.2.14 we obtain $V_{li} \perp \mathcal{E}_l^j$ for all j and thus, since $\sum_j \mathcal{E}_l^j = \mathcal{E}_l$, we obtain $V_{li} \perp \mathcal{E}_l$ and overall $V_{li} = 0$, a contradiction.
Thus, the assumption was wrong and all V_{li} in the orthogonal direct sum are isomorphic to V_l.
Now let $l \neq l'$ and i, j be arbitrary. We have $\mathcal{E}_l \perp \mathcal{E}_{l'}$ by Proposition 2.2.7, and thus in particular $V_{li} \perp V_{l'j}$. Furthermore, we have $n_l \leq \dim(V_l)$ since $\mathcal{E}_l = \sum_{j=1}^{\dim(V_l)} \mathcal{E}_l^j = \bigoplus_{i=1}^{n_l} V_{li}$, and $\dim(V_l) < \infty$ by Proposition 2.2.8.
Moreover, we have $\bigoplus_{l \in \hat{G}} \bigoplus_{i=1}^{n_l} V_{li} = \bigoplus_{l \in \hat{G}} \mathcal{E}_l = \mathcal{E}$, which is topologically dense in $L^2_{\mathbb{K}}(G)$ by Theorem 2.2.4.
Finally, that $n_l = \dim(V_l)$ if $\mathbb{K} = \mathbb{C}$ follows by invoking a stronger version of Schur's orthogonality than we have developed, and which works only over the complex numbers [14]. □

2.2.4. A Proof of the Peter-Weyl Theorem for General $L^2_{\mathbb{K}}(X)$

Now let X be a homogeneous space of G. Then, as mentioned in Section 2.1.3, there is a measure μ on X which is left-G-invariant [16] in the sense that we have for all $g \in G$ and all square-integrable functions $f \in L^2_{\mathbb{K}}(X)$:

$$\int_X f(g \cdot x) dx = \int_X f(x) dx.$$

Furthermore, let $\pi : G \to X$ be the projection given by $g \mapsto gx^*$ for a fixed element $x^* \in X$. One important result is that there is a Fubini-like theorem for evaluation of integrals on G using the invariant measure on X. Namely, for arbitrary $x \in X$, let $g(x) \in G$ be any lift, i.e. any element in G with $\pi(g(x)) = x$. This exists since the action is transitive. Let $H := G_{x^*} \subseteq G$ be the stabilizer subgroup. For a square-integrable function $f : G \to \mathbb{K}$, we can then construct the average $\mathrm{av}(f) : X \to \mathbb{K}$ by

$$\mathrm{av}(f)(x) := \int_H f(g(x)h) dh,$$

where we integrate using the Haar-measure on H[3]. If it is hard to understand why this is called an average, note that $X \cong G/H$, i.e. points in X can be interpreted as cosets of G, and then the average just averages over cosets[4].
This construction is well-defined, i.e. does not depend on the specific choice of the lift $g(x)$. Indeed, let $g(x)'$ be another lift of x. Then $g(x)' = g(x)h'$ for some $h' \in H$, since H is the stabilizer subgroup. Consequently, using the invariance of the Haar measure, we see:

$$\int_H f(g(x)'h)dh = \int_H f(g(x)h'h)dh = \int_H f(g(x)h)dh,$$

and thus the well-definedness of the average $\mathrm{av}(f) : X \to \mathbb{K}$. Integration of f on the whole of G is a "complete" average, and thus we can hope that averaging $\mathrm{av}(f)$ leads to this complete integral. This is indeed the case, i.e. $\mathrm{av}(f)$ is square-integrable on X and one has [16]

$$\int_G f(g)dg = \int_X \mathrm{av}(f)(x)dx. \tag{2.1}$$

We will use this important result later in order to see that $L^2_{\mathbb{K}}(X)$ embeds with good properties into $L^2_{\mathbb{K}}(G)$.
We now want to prove the Peter-Weyl Theorem for $L^2_{\mathbb{K}}(X)$. We first present a general argument showing an orthogonal decomposition of $L^2_{\mathbb{K}}(X)$ into irreducible subspaces, and then use a specific argument to deduce that the multiplicities of irreducible subrepresentations are necessarily bounded by the multiplicities in $L^2_{\mathbb{K}}(G)$.

Proposition 2.2.17. *Let $\rho : G \to \mathrm{U}(V)$ be any unitary representation. Then there is a dense subrepresentation which splits as an orthogonal direct sum of irreducible subrepresentations.*

Proof. We sketch the proof in Kowalski [15], Corollary 5.4.2. In this book, the proof is done only for the complex numbers $\mathbb{C}$, but it is obvious that each step carries over without any changes to arbitrary $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. The rough steps are as follows:

1. From ρ one builds a function $\overline{\rho} : L^2_{\mathbb{K}}(G) \to \mathrm{Hom}_{\mathbb{K}}(V, V)$, given by $\overline{\rho}(\varphi)(v) = \int_G \varphi(g)\rho(g)(v)dg$. This is analogous to our construction of kernel operators from kernels, which we will handle in the next chapter, See Theorem 3.1.7.

2. Given $v \in V$ fixed, one obtains the function $\overline{\rho}^v : L^2_{\mathbb{K}}(G) \to V$, $\varphi \mapsto \overline{\rho}(\varphi)(v)$. One can check easily that this is an intertwiner.

3. For each finite-dimensional subrepresentation $E \subseteq L^2_{\mathbb{K}}(G)$, the image $\overline{\rho}^v(E) \subseteq V$ is a finite-dimensional subrepresentation of V.

[3]Such a Haar measure exists since $H \subseteq G$ is a topologically closed subgroup of a compact group by Proposition 2.1.21 and thus compact itself by standard topological results [23]. Note that this measure fulfills $\mu(H) = 1$ and is thus *not* the same as the restriction of the measure on G to H.

[4]Here, G/H is the set of equivalence classes in G with respect to the equivalence relation $g \sim g'$ if $g^{-1}g' \in H$, which has a quotient topology as explained in Definition A.1.11. The equivalence classes are given by the cosets gH for $g \in G$.

4. For $v \neq 0$, using analytical arguments and the Peter-Weyl Theorem for $L^2_{\mathbb{K}}(G)$, one can prove that there is an E such that $\overline{\rho}^v(E) \subseteq V$ is not zero.

Having that, one can use Proposition 2.2.15 in order to deduce that $\overline{\rho}^v(E)$ contains an irreducible subrepresentation, and so does V.
With this at hand, one can proceed inductively as follows: Given an irreducible subrepresentation $V_1 \subseteq V$, one can consider the orthogonal complement $V_1^\perp$, which is by Lemma 2.2.13 again a subrepresentation of V. Thus, this also has, by the same argument as above, an irreducible subrepresentation V_2 and so on. By induction (or better: using Zorn's Lemma), one can then "fill up" V with orthogonal irreducible subrepresentations, deducing the result. □

Consequently, since $L^2_{\mathbb{K}}(X)$ carries a unitary representation of G by $[\lambda(g)(\varphi)](x) := \varphi(g^{-1}x)$, we can deduce that it contains a dense subrepresentation which splits as an orthogonal direct sum of irreducible subrepresentations. But we would like to know more details about this, in particular the multiplicities of the irreps. For this to work, we want to embed $L^2_{\mathbb{K}}(X)$ into $L^2_{\mathbb{K}}(G)$ and thus deduce a more specific result from the decomposition of $L^2_{\mathbb{K}}(G)$.
Let as before $x^* \in X$ be an arbitrary point and let $\pi : G \to X$ be the projection given by $\pi(g) := gx^*$. Consider the function $\pi^* : L^2_{\mathbb{K}}(X) \to L^2_{\mathbb{K}}(G)$ given by $\pi^*(\varphi) := \varphi \circ \pi$. It is unclear a priori whether this is well-defined: For example, it might be that an $f : X \to \mathbb{K}$ which is zero outside a measure 0 set gets lifted to $\pi^*(f) : G \to \mathbb{K}$ which does not have this property, and thus π^* would not be an actual function[5]. Thus, we need some lemmas:

Lemma 2.2.18. *Let $f : X \to \mathbb{K}$ be square-integrable. Then we have* $\mathrm{av}(\pi^*(f)) = f$.

Proof. Using Equation 2.1 and that H is the stabilizer subgroup we compute:

$$\begin{aligned}
\mathrm{av}(\pi^*(f))(x) &= \int_H \pi^*(f)(g(x)h)dh \\
&= \int_H f(\pi(g(x)h))dh \\
&= \int_H f(\pi(g(x)))dh \\
&= \int_H f(x)dh \\
&= f(x)\int_H 1dh \\
&= f(x)\mu(H) \\
&= f(x).
\end{aligned}$$

□

[5] Remember that functions in $L^2_{\mathbb{K}}(X)$ for any measurable space X are identified if they agree outside a set of measure 0.

Lemma 2.2.19. *Let $A \subseteq X$ be any measurable set. Let $\mathbf{1}_A : X \to \{0, 1\} \subseteq \mathbb{K}$ be its indicator function. Then $\pi^*(\mathbf{1}_A) = \mathbf{1}_{\pi^{-1}(A)}$.*

Proof. This can easily be checked. □

Lemma 2.2.20. *Let $\varphi : X \to \mathbb{K}$ be zero outside a measure zero set A. Then $\pi^*(\varphi)$ is zero outside $\pi^{-1}(A)$ which is also a measure zero set.*

Proof. If $g \notin \pi^{-1}(A)$ then $\pi(g) \notin A$ and thus:

$$0 = \varphi(\pi(g)) = \pi^*(\varphi)(g)$$

which proves the first statement. The second is shown as follows using both Lemmas 2.2.18 and 2.2.19 and Equation 2.1:

$$\begin{aligned}
\mu(\pi^{-1}(A)) &= \int_G \mathbf{1}_{\pi^{-1}(A)}(g)dg \\
&= \int_G \pi^*(\mathbf{1}_A)(g)dg \\
&= \int_X \mathrm{av}(\pi^*(\mathbf{1}_A))(x)dx \\
&= \int_X \mathbf{1}_A(x)dx \\
&= \mu(A) \\
&= 0,
\end{aligned}$$

thus showing what was claimed. □

Thus, our concern about well-definedness as a function is invalid and we can now prove an embedding result:

Proposition 2.2.21. *$\pi^* : L^2_{\mathbb{K}}(X) \to L^2_{\mathbb{K}}(G)$ is a well-defined intertwiner and a unitary transformation, i.e. for all $\varphi, \psi \in L^2_{\mathbb{K}}(X)$ we have $\langle \pi^*(\varphi) | \pi^*(\psi) \rangle_{L^2_{\mathbb{K}}(G)} = \langle \varphi | \psi \rangle_{L^2_{\mathbb{K}}(X)}$.*

Proof. For well-definedness, we still need to show that $\pi^*(\varphi)$ is again square-integrable for square-integrable $\varphi : X \to \mathbb{K}$. This is indeed the case due to Equation 2.1. Namely, let $|\pi^*(\varphi)|^2 : G \to \mathbb{K}$ and consider its average $\mathrm{av}(|\pi^*(\varphi)|^2)$. Clearly, we have $|\pi^*(\varphi)|^2 = \pi^*(|\varphi|^2)$ and thus, using Lemma 2.2.18, $\mathrm{av}(|\pi^*(\varphi)|^2) = |\varphi|^2$. We obtain:

$$\begin{aligned}
\int_G |\pi^*(\varphi)|^2(g)dg &= \int_X \mathrm{av}(|\pi^*(\varphi)|^2)(x)dx \\
&= \int_X |\varphi(x)|^2 dx \\
&< \infty.
\end{aligned}$$

Thus, π^* is not only well-defined but even fulfills $\|\pi^*(\varphi)\|_{L^2_{\mathbb{K}}(G)} = \|\varphi\|_{L^2_{\mathbb{K}}(X)}$, which also shows the continuity of π^*. With similar arguments, we show that π^* respects the whole scalar product, i.e. is a uniform transformation:

$$\begin{aligned}\langle \pi^*(\varphi)|\pi^*(\psi)\rangle_{L^2_{\mathbb{K}}(G)} &= \int_G \left(\overline{\pi^*(\varphi)} \cdot \pi^*(\psi)\right)(g)dg \\ &= \int_X \mathrm{av}(\overline{\pi^*(\varphi)} \cdot \pi^*(\psi))(x)dx \\ &= \int_X \overline{\varphi(x)}\psi(x)dx \\ &= \langle \varphi|\psi\rangle_{L^2_{\mathbb{K}}(X)} .\end{aligned}$$

The step from the second to the third line follows as before by noting that $\overline{\pi^*(\varphi)} \cdot \pi^*(\psi) = \pi^*(\overline{\varphi} \cdot \psi)$ and invoking Lemma 2.2.18 again.
The linearity of π^* is obvious, and the equivariance is done as follows: note that for arbitrary $g, g' \in G$ we have $\pi(g^{-1}g') = (g^{-1}g')x^* = g^{-1}(g'x^*) = g^{-1}\pi(g')$ and therefore:

$$\begin{aligned}\left[\pi^*(\lambda(g)\varphi)\right](g') &= (\lambda(g)\varphi)(\pi(g')) \\ &= \varphi(g^{-1}\pi(g')) \\ &= \varphi(\pi(g^{-1}g')) \\ &= \pi^*(\varphi)(g^{-1}g') \\ &= \left[\lambda(g)\pi^*(\varphi)\right](g').\end{aligned}$$

Thus, we shown everything which was to show. □

Thus, $\pi^* : L^2_{\mathbb{K}}(X) \to L^2_{\mathbb{K}}(G)$ is an embedding which even preserves the scalar product. We can therefore view $L^2_{\mathbb{K}}(X)$ as a subspace: $L^2_{\mathbb{K}}(X) \subseteq L^2_{\mathbb{K}}(G)$[6].
We can finally complete the proof of the Peter-Weyl Theorem 2.1.22:

Proof of Theorem 2.1.22. Assume that

$$\bigoplus_{l\in\hat{G}} \bigoplus_{i=1}^{m_l} V_{li} \subseteq L^2_{\mathbb{K}}(X) \subseteq L^2_{\mathbb{K}}(G)$$

is a dense subspace such that the direct sum is orthogonal, where $V_{li} \cong V_l$ for all l, i. This exists by Proposition 2.2.17.
Remember that n_l denotes the multiplicity of V_l as a subrepresentation in $L^2_{\mathbb{K}}(G)$. We now want to show that $m_l \leq n_l$. Since V_{li} is perpendicular to all $\mathcal{E}_{l'}$ with $l' \neq l$ by Lemma 2.2.14, V_{li} must be contained in the orthogonal complement of $\bigoplus_{l'\neq l} \mathcal{E}_{l'}$. This is exactly $\mathcal{E}_l$, which we show in a final lemma after this proof. So $V_{li} \subseteq \mathcal{E}_l$ for all i. Thus, we obtain the result $m_l \leq n_l$ by dimension reasons. This was all there was left to show. □

[6] In this notation, we suppress that this embedding depends on the specific base point x^* which was chosen. For another base point, the embedding differs by a unitary automorphism on $L^2_{\mathbb{K}}(G)$ as the reader may want to check.

Lemma 2.2.22. *We have* $\mathcal{E}_l = \left(\bigoplus_{l\neq l'\in\hat{G}} \mathcal{E}_{l'}\right)^{\perp}$

Proof. We already know $\mathcal{E}_l \subseteq \left(\bigoplus_{l\neq l'\in\hat{G}} \mathcal{E}_{l'}\right)^{\perp}$ from Proposition 2.2.7. Now, assume this inclusion is not an equality. Then there is $v \notin \mathcal{E}_l$ such that $v \in \left(\bigoplus_{l\neq l'\in\hat{G}} \mathcal{E}_{l'}\right)^{\perp}$. The space $\operatorname{span}_{\mathbb{K}}(v, \mathcal{E}_l)$ does contain an orthonormal basis by Proposition A.2.10, where the procedure of Gram-Schmidt orthonormalization allows starting with an orthonormal basis of $\mathcal{E}_l$ and to fill it up to one of the whole space $\operatorname{span}_{\mathbb{K}}(v, \mathcal{E}_l)$. Thus, we can assume $v \in \mathcal{E}_l^{\perp}$ as well. Overall, $v \in \left(\bigoplus_{l'\in\hat{G}} \mathcal{E}_{l'}\right)^{\perp}$, and by taking topological closure and using that the scalar product is continuous by Proposition A.2.7, obtain $v \in \left(\widehat{\bigoplus}_{l'\in\hat{G}}\mathcal{E}_{l'}\right)^{\perp} = (L^2_{\mathbb{K}}(G))^{\perp}$ by the Peter-Weyl Theorem for the regular representation. This means $v = 0 \in \mathcal{E}_l$, a contradiction to $v \notin \mathcal{E}_l$.
Thus, our assumption is wrong and such a vector v cannot exist. We obtain the equality as desired. □

3. The Correspondence between Steerable Kernels and Kernel Operators

In this chapter, we formulate and prove Theorem 3.1.7, which gives a precise one-to-one correspondence between steerable kernels on the one hand, and certain representation operators which we call *kernel operators* on the other hand.

This correspondence will be the main bridge between the theory of steerable CNNs and the mathematical foundations of quantum mechanics that underlies this work. It will allow us in Chapter 4 to crucially exploit an insight from quantum mechanics, namely the Wigner-Eckart Theorem, in order to better understand steerable kernels and to ultimately get a complete description of steerable kernel bases. In this sense, the correspondence established in this chapter may be seen as the main theoretical insight of this work.

The structure is as follows: In Section 3.1, we formulate the correspondence between steerable kernels and kernel operators. We do this by first studying steerable CNNs and the kernel constraint, which progressively leads us to consider steerable kernels on homogeneous spaces of general compact groups. This abstract formulation of steerable kernels will have apparent similarities to the concept of representation operators from physics, which we study next. However, they importantly differ in the fact that steerable kernels are not linear, whereas representation operators are, a difference that we need to bridge. Finally, after defining kernel operators as special representation operators, we give the formulation of the correspondence in Theorem 3.1.7 and shortly give some intuitions about why it is true. Then, in Section 3.2, we give a detailed and rigorous proof of this correspondence.

As in Chapter 2, $\mathbb{K}$ is either of the two fields $\mathbb{R}$ and $\mathbb{C}$.

3.1. Fundamentals of the Correspondence

3.1.1. Steerable Kernels and the Restriction to Homogeneous Spaces

In this section, we shortly explain steerable CNNs and reduce them conceptually to a form that is most useful for the correspondence between steerable kernels and kernel operators that we want to develop in this chapter. The concept of steerable CNNs

outlined here essentially follows Weiler and Cesa [9]. In a nutshell, steerable CNNs work as follows:
The network is supposed to process data given as functions $\mathbb{R}^n \to \mathbb{K}^c$, where n is the dimension of the space where the input-features are defined. Such functions are also called feature fields. c is the dimension of the features themselves, i.e. the number of channels. For example, planar RGB-images correspond to the case $n = 2$ and $c = 3$. Furthermore, a compact group G, see Definition 2.1.4, is considered that acts on $\mathbb{R}^n$ by rotations or reflections or both, for example the special orthogonal group $\mathrm{SO}(n)$, the orthogonal group $\mathrm{O}(n)$ or the finite groups C_N or D_N[1]. Then for each layer, the input and output has a certain *type*, i.e. representation, which may differ from layer to layer. That is, the input (and output as well) consists of a function $f : \mathbb{R}^n \to \mathbb{K}^c$, and G acts on $\mathbb{K}^c$ with a linear representation ρ, see Definition 2.1.10. This action induces an action of the semi-direct product $(\mathbb{R}^n, +) \rtimes G$ on the space of all signals[2], where $t \in (\mathbb{R}^n, +)$ and $g \in G$:

$$\Big(\left[\mathrm{Ind}_G^{\mathbb{R}^n \rtimes G} \rho\right] (tg) \cdot f\Big)(x) := \rho(g) \cdot f(g^{-1}(x - t)).$$

Let the kernel that "maps" between the layers by convolution[3] be given by a function

$$K : \mathbb{R}^n \to \mathbb{K}^{c_{\mathrm{out}} \times c_{\mathrm{in}}}.$$

That is, for an input $f_{\mathrm{in}} : \mathbb{R}^n \to \mathbb{K}^{c_{\mathrm{in}}}$, the output $f_{\mathrm{out}} : \mathbb{R}^n \to \mathbb{K}^{c_{\mathrm{out}}}$ is given by

$$f_{\mathrm{out}}(x) = \left[K \star f_{\mathrm{in}}\right](x) = \int_{\mathbb{R}^n} K(y - x) f_{\mathrm{in}}(y) dy,$$

where $K(y - x) \in \mathbb{K}^{c_{\mathrm{out}} \times c_{\mathrm{in}}}$ is viewed as a linear transformation from $\mathbb{K}^{c_{\mathrm{in}}}$ to $\mathbb{K}^{c_{\mathrm{out}}}$. The goal is now to find kernels K such that convolution with these kernels commutes with the induced actions on the input and output fields. That is, for all input fields f_{in} and for all $t \in \mathbb{R}^n$ and $g \in G$ we want the following property:

$$K \star \Big(\left[\mathrm{Ind}_G^{\mathbb{R}^n \rtimes G} \rho_{\mathrm{in}}\right] (tg) \cdot f_{\mathrm{in}} \Big) = \left[\mathrm{Ind}_G^{\mathbb{R}^n \rtimes G} \rho_{\mathrm{out}}\right] (tg) \cdot (K \star f_{\mathrm{in}})$$

It was shown in Weiler et al. [8] that a kernel K has this equivariance property if and only if, for all $g \in G$ and $x \in \mathbb{R}^n$, it holds:

$$K(gx) = \rho_{\mathrm{out}}(g) \circ K(x) \circ \rho_{\mathrm{in}}(g)^{-1}. \tag{3.1}$$

This thesis will create a general theory for how to *solve* this kernel constraint, which means to find a parameterization for the space of all kernels that fulfill this constraint.

[1] We will study some of these groups in the Examples in Chapter 6.

[2] The semidirect product $\mathbb{R}^n \rtimes G$ can be imagined as the smallest subgroup of the group of all isometries of $\mathbb{R}^n$ that contains both the translations $\mathbb{R}^n$ and the transformations G. It is not important to know the abstract definition of a semidirect product in our context.

[3] The operation is actually a so-called "correlation", but the term "convolution" is more widespread in the deep learning context and we follow this convention.

We now explain how to make this problem more tractable: formally, the action of G on $\mathbb{R}^n$ is a group action as in Definition 2.1.5. However, it cannot be transitive as in Definition 2.1.7 since G is compact and $\mathbb{R}^n$ is not. Thus $\mathbb{R}^n$ splits into a disjoint union of orbits, see Definition 2.1.6, of the action:

$$\mathbb{R}^n = \bigsqcup_{k \in K} X_k.$$

That this is a *disjoint* union can be explained as follows: define the relation $\sim$ on $\mathbb{R}^n$ by $x \sim x'$ if $gx = x'$ for some $g \in G$. This is then an equivalence relation, and so $\mathbb{R}^n$ splits into a disjoint union of equivalence classes. One then can show that these equivalence classes are precisely the orbits of the group action. For example, such orbits take the form of spheres S^{n-1} if $G = \mathrm{SO}(n)$ or $G = \mathrm{O}(n)$ and the form of a finite set of points if $G = \mathrm{C}_N$ or $G = \mathrm{D}_N$.
The idea is now that the kernel constraint 3.1 only constrains the behavior of the kernel at each orbit individually, and thus a solution on each orbit can be "patched together" to a solution on the whole of $\mathbb{R}^n$. Indeed, assume that $K_k : X_k \to \mathbb{K}^{c_{\text{out}} \times c_{\text{in}}}$ individually fulfill the kernel constraint, which means that for all $x_k \in X_k$ and $g \in G$ we have

$$K_k(gx_k) = \rho_{\text{out}}(g) \circ K_k(x_k) \circ \rho_{\text{in}}(g)^{-1}.$$

Then, define the patch of these orbit-kernels by $K : \mathbb{R}^n \to \mathbb{K}^{c_{\text{out}} \times c_{\text{in}}}$ as $K(x) = K_k(x)$ if $x \in X_k$. This is well-defined since each x is in precisely one orbit. Then clearly, K fulfills the kernel constraint 3.1. Moreover, each kernel K which fulfills the kernel constraint emerges from such a construction, since we can simply set $K_k := K|_{X_k}$. Overall, we see that we can restrict our attention to orbits. In Weiler et al. [24] and later Weiler et al. [8], a discretized implementation is done where the kernel is discretized into finitely many orbits with a smooth Gaussian radial profile. We will come back to these practical questions of parameterization in Remark 4.1.18, once we have fully developed the theory of steerable CNNs.

3.1.2. An Abstract Definition of Steerable Kernels

Motivated by the discussion in the last section, we now define steerable kernels in precise terms and will stick to that definition throughout this work. The definition will be more abstract than usual in the deep learning community, but we are rewarded since such an abstract definition makes it easier to apply theoretical results from mathematics and physics.
Without loss of generality, we will in the rest of this work only consider kernels on orbits. Thus, let $X := G \cdot x$ be an arbitrary orbit. We consider steerable kernels $K : X \to \mathbb{K}^{c_{\text{out}} \times c_{\text{in}}}$. Note that the restriction of the action $G \times \mathbb{R}^n \to \mathbb{R}^n$ to X, written $G \times X \to X$, makes X to a homogeneous space of G, see Definition 2.1.7. Thus, instead of viewing X as a subset of $\mathbb{R}^n$, we view X as an arbitrary homogeneous space of an arbitrary compact group G. Notably, this framework is more general than usually studied in the context of steerable CNNs on $\mathbb{R}^n$, since we allow also groups

that are not Lie groups and homogeneous spaces which are not naturally embedded in an $\mathbb{R}^n$, as well as finite homogeneous spaces of finite groups all at the same time.
Furthermore, note that $\mathbb{K}^{c_{\text{out}} \times c_{\text{in}}}$ can be viewed as the space of linear functions from $\mathbb{K}^{c_{\text{in}}}$ to $\mathbb{K}^{c_{\text{out}}}$, written $\text{Hom}_{\mathbb{K}}(\mathbb{K}^{c_{\text{in}}}, \mathbb{K}^{c_{\text{out}}})$. This is useful since it allows us to view $\mathbb{K}^{c_{\text{in}}}$ and $\mathbb{K}^{c_{\text{out}}}$ in more abstract terms. Namely, we replace these spaces by arbitrary finite-dimensional $\mathbb{K}$-vector spaces V_{in} and V_{out} together with linear representations $\rho_{\text{in}} : G \to \text{Aut}_{\mathbb{K}}(V_{\text{in}})$ and $\rho_{\text{out}} : G \to \text{Aut}_{\mathbb{K}}(V_{\text{out}})$.
Overall, this means that steerable kernels are certain (not linear in a meaningful sense) maps $K : X \to \text{Hom}_{\mathbb{K}}(V_{\text{in}}, V_{\text{out}})$. The only property they need to fulfill is the kernel constraint $K(gx) = \rho_{\text{out}}(g) \circ K(x) \circ \rho_{\text{in}}(g)^{-1}$ for all $g \in G$ and $x \in X$. This looks a bit like an equivariance property: we plug the "rotation" gx into K and express that this is the same as a certain transformation of $K(x)$ in the output-space, which consists of linear functions from V_{in} to V_{out}. This can be made more precise by defining the Hom-representation on $\text{Hom}_{\mathbb{K}}(V_{\text{in}}, V_{\text{out}})$:

Definition 3.1.1 (Hom-Representation)**.** Let $\rho_{\text{in}} : G \to \text{Aut}_{\mathbb{K}}(V_{\text{in}})$ and $\rho_{\text{out}} : G \to \text{Aut}_{\mathbb{K}}(V_{\text{out}})$ be two finite-dimensional G-representations over the field $\mathbb{K}$. The space $\text{Hom}_{\mathbb{K}}(V_{\text{in}}, V_{\text{out}})$ of $\mathbb{K}$-linear (not necessarily G-equivariant) functions from V_{in} to V_{out} also carries a G-representation, with action

$$[\rho_{\text{Hom}}(g)]\,(f) \coloneqq \rho_{\text{out}}(g) \circ f \circ \rho_{\text{in}}(g)^{-1}.$$

We call this the *Hom-representation.*

Remark 3.1.2. Of course, one needs to check that this is indeed a linear representation. Continuity follows from the continuity of ρ_{in} and ρ_{out} as follows: the topology on $\text{Hom}_{\mathbb{K}}(V_{\text{in}}, V_{\text{out}})$ is just the Euclidean topology of $\mathbb{K}^{c_{\text{out}} \times c_{\text{in}}}$ coming from a basis of V_{in} and V_{out}. In these bases, $\rho_{\text{in}}(g)$ and $\rho_{\text{out}}(g)$ are given by matrices. All matrix coefficients are continuous by Remark 2.2.2. Now, in order to show that ρ_{Hom} is continuous, pick a fixed element $f \in \mathbb{K}^{c_{\text{in}} \times c_{\text{out}}}$. One needs to show that the map

$$\rho_{\text{Hom}}^f : G \to \mathbb{K}^{c_{\text{in}} \times c_{\text{out}}}, \ g \mapsto \rho_{\text{out}}(g) \circ f \circ \rho_{\text{in}}(g^{-1})$$

is continuous. Since all matrix coefficients are continuous and since also the inversion $G \to G, \ g \mapsto g^{-1}$ is continuous by the definition of a topological group, the map ρ_{Hom}^f is basically just a stacked linear combination of continuous functions and thus continuous itself.
The linearity of each $\rho_{\text{Hom}}(g)$ is also clear. So what needs to be checked is that ρ_{Hom} is a group homomorphism. And indeed, it is, exploiting the corresponding property of ρ_{in} and ρ_{out}:

$$\begin{aligned}
[\rho_{\text{Hom}}(gg')]\,(f) &= \rho_{\text{out}}(gg') \circ f \circ \rho_{\text{in}}(gg')^{-1} \\
&= \rho_{\text{out}}(g) \circ \left(\rho_{\text{out}}(g') \circ f \circ \rho_{\text{in}}(g')^{-1}\right) \circ \rho_{\text{in}}(g)^{-1} \\
&= [\rho_{\text{Hom}}(g)]\left([\rho_{\text{Hom}}(g')]\,(f)\right) \\
&= [\rho_{\text{Hom}}(g) \circ \rho_{\text{Hom}}(g')]\,(f),
\end{aligned}$$

and so the claim follows.

With this definition in mind, steerable kernels $K : X \to \mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}})$ are just functions with the property $K(gx) = [\rho_{\mathrm{Hom}}(g)]\,(K(x))$. This is the equivariance-view we were looking for. Summarizing, we have the following abstract definition of steerable kernels:

Definition 3.1.3 (Steerable Kernel)**.** Let G be any compact group and X be any homogeneous space of G. Furthermore, let $\rho_{\mathrm{in}} : G \to \mathrm{Aut}_{\mathbb{K}}(V_{\mathrm{in}})$ and $\rho_{\mathrm{out}} : G \to \mathrm{Aut}_{\mathbb{K}}(V_{\mathrm{out}})$ be finite-dimensional representations of G. We assume that $\mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}})$ is equipped with the Hom-representation ρ_{Hom}.
A *steerable kernel* is an equivariant function $K : X \to \mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}})$, i.e. a function such that

$$K(gx) = [\rho_{\mathrm{Hom}}(g)]\,(K(x)) \tag{3.2}$$

for all $g \in G$ and $x \in X$. We denote the vector-space of all these kernels by

$$\mathrm{Hom}_G(X, \mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}})) = \{K : X \to \mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}}) \mid K \text{ is a steerable }\}\,.$$

Notably, steerable kernels are *not linear* in a meaningful sense with respect to their input.

That the space of steerable kernels forms a vector space, as claimed in this definition, can easily be checked. In the next section, we will see that Definition 3.1.3 looks suspiciously like *representation operators* considered in physics.

3.1.3. Representation Operators and Kernel Operators

Now that we have a clear abstract idea of what steerable kernels are, we can begin to establish analogies to physics. In this section, we therefore formulate what so-called representation operators are, which play a central role in quantum physics [10] and will then formulate the main theorem of this chapter, Theorem 3.1.7. This establishes the bridge between the realms of deep learning and quantum physics that we need in order to exploit physical insights.
Without spending too much time on the physical intuitions behind this mathematical formalism, for which we refer back to the introduction, we right away come to the main definition. It is basically a mathematical formalization and generalization of the concept of a spherical tensor operator. We thereby restrict to finite-dimensional input- and output representations due to our specific applications:

Definition 3.1.4 (Representation Operator)**.** Let $\rho_{\mathrm{in}} : G \to \mathrm{Aut}_{\mathbb{K}}(V_{\mathrm{in}})$ and $\rho_{\mathrm{out}} : G \to \mathrm{Aut}_{\mathbb{K}}(V_{\mathrm{out}})$ be finite-dimensional G-representations. Let $\lambda : G \to \mathrm{Aut}_{\mathbb{K}}(T)$ be a third G-representation, not necessarily finite-dimensional. Then a *representation operator* is an intertwiner $\mathcal{K} : T \to \mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}})$, where the right space is equipped with the Hom-representation as in Definition 3.1.1. We denote the vector space of all these representation operators by

$$\mathrm{Hom}_{G,\mathbb{K}}(T, \mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}})) = \{\mathcal{K} : T \to \mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}}) \mid \mathcal{K} \text{ is an intertwiner}\}\,.$$

Note that representation operators are by definition *linear*, which is a requirement that needs to be fulfilled for the standard Wigner-Eckart Theorem. We clearly see strong similarities between this definition and the formalization of steerable kernels in Definition 3.1.3. The main difference is that we assume representation operators to be linear. This is in notation captured by the subscript $\mathbb{K}$ that we put in the corresponding Hom-space. One may think that there is another difference, namely coming from the fact that intertwiners are by definition *continuous* with respect to the topologies involved. Two things need to be said about this:

1. First of all, one may wonder what continuity for representation operators actually means. This can be clarified as follows: By assumption, G-representations are always on vector spaces with topologies, and thus T has a topology. Furthermore, in Remark 3.1.2 we clarified the topology on $\mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}})$. Then, being continuous just means, as always, to be continuous with respect to the topologies of these two spaces.

2. The second remark is that this apparent difference in the requirement of continuity for steerable kernels and representation operators is actually non-existent. This is explained by the following Proposition

Proposition 3.1.5. *Let $K : X \to \mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}})$ be a steerable kernel. Then K is continuous.*

Proof. For brevity, denote $V := \mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}})$ and $\rho := \rho_{\mathrm{Hom}}$. Let $x^* \in X$ be any point and G_{x^*} the stabilizer corresponding to the action of G on X. Remember the homeomorphism $\varphi : G/H \to X$, $[g] \mapsto gx^*$ from Lemma 2.1.21. Since this is a homeomorphism, the kernel K is continuous if and only if the composition $K \circ \varphi$ is continuous, since then $K = (K \circ \varphi) \circ \varphi^{-1}$ is a composition of continuous functions. Thus, we evaluate $K \circ \varphi$:

$$(K \circ \varphi)([g]) = K(\varphi([g])) = K(gx^*) = \rho(g)(K(x^*)),$$

where in the last step we have used the equivariance of K. Thus, if we set $v^* := K(x^*) \in V$, then we obtain the simple relation $(K \circ \varphi)([g]) = \rho(g)(v^*)$. This is by definition just the unique map on the quotient, $G/H \to V$, coming from $\rho^{v^*} : G \to V$, $g \mapsto \rho(g)(v^*)$. This last map is continuous by definition of a linear representation. The universal property of quotients Proposition A.1.12 then shows that $K \circ \varphi$ is continuous as well, and so we are done. All of this is visualized in the following commutative diagram, where $q : G \to G/H$, $g \mapsto [g]$ is the canonical projection:

$$\begin{tikzcd}
 & G \arrow[dl, "q"'] \arrow[d, "(\cdot)\cdot x^*"] \arrow[dr, "\rho^{v^*}"] & \\
G/H \arrow[r, "\sim", "\varphi"'] \arrow[rr, bend right, "K\circ\varphi"'] & X \arrow[r, "K"'] & V
\end{tikzcd}$$

□

Thus, the only difference between steerable kernels and representation operators is indeed the linearity. We now look at special representation operators that play the main role in this work:

Definition 3.1.6 (Kernel Operator). Let $\rho_{\text{in}} : G \to \text{Aut}_{\mathbb{K}}(V_{\text{in}})$ and $\rho_{\text{out}} : G \to \text{Aut}_{\mathbb{K}}(V_{\text{out}})$ be finite-dimensional G-representations. Let $\lambda : G \to \text{U}(L^2_{\mathbb{K}}(X))$ be the standard unitary representation on the space of square-integrable functions of a homogeneous space X, given, as in Section 2.1.3, by

$$[\lambda(g)(\varphi)](g') = \varphi(g^{-1}g').$$

A *kernel operator* is a representation operator $\mathcal{K} : L^2_{\mathbb{K}}(X) \to \text{Hom}_{\mathbb{K}}(V_{\text{in}}, V_{\text{out}})$. We denote the space of these by

$$\begin{aligned}&\text{Hom}_{G,\mathbb{K}}(L^2_{\mathbb{K}}(X), \text{Hom}_{\mathbb{K}}(V_{\text{in}}, V_{\text{out}}))\\ &\qquad\qquad = \left\{ \mathcal{K} : L^2_{\mathbb{K}}(X) \to \text{Hom}_{\mathbb{K}}(V_{\text{in}}, V_{\text{out}}) \mid \mathcal{K} \text{ is an intertwiner} \right\}.\end{aligned}$$

Notably, kernel operators are $\mathbb{K}$-linear in their input.

3.1.4. Formulation of the Correspondence between Steerable Kernels and Kernel Operators

The following Theorem lies at the heart of our investigations and establishes that steerable kernels can be considered as kernel operators. More precisely, we will give an explicit isomorphism between the space of steerable kernels and the space of kernel operators.

We shortly explain why the theorem is useful. First of all, using a Wigner-Eckart theorem for kernel operators that we prove in Theorem 4.1.13, one can explicitly describe a basis B of the space of kernel operators $\text{Hom}_{G,\mathbb{K}}(L^2_{\mathbb{K}}(X), \text{Hom}_{\mathbb{K}}(V_{\text{in}}, V_{\text{out}}))$. Then, since we have an isomorphism of vector spaces to the space of steerable kernels, one can "carry over" this basis to a basis for the space of steerable kernels, namely $\text{Hom}_G(X, \text{Hom}_{\mathbb{K}}(V_{\text{in}}, V_{\text{out}}))$. This basis will then have a convenient explicit form that we establish in Theorem 4.1.15 and is exactly what we need in order to parameterize an equivariant neural network layer. We now come to a precise formulation of the theorem:

Theorem 3.1.7 (Kernel-Operator-Correspondence). *Let $\rho_{\text{in}} : G \to \text{Aut}_{\mathbb{K}}(V_{\text{in}})$ and $\rho_{\text{out}} : G \to \text{Aut}_{\mathbb{K}}(V_{\text{out}})$ be finite-dimensional G-representations and X be a homogeneous space of G. Then there is an isomorphism*

$$\begin{tikzcd}\text{Hom}_G(X, \text{Hom}_{\mathbb{K}}(V_{\text{in}}, V_{\text{out}})) \arrow[r, bend left, "\widehat{(\cdot)}"] & \text{Hom}_{G,\mathbb{K}}(L^2_{\mathbb{K}}(X), \text{Hom}_{\mathbb{K}}(V_{\text{in}}, V_{\text{out}})) \arrow[l, bend left, "(\cdot)|_X"]\end{tikzcd}$$

between the space of steerable kernels on the left and the space of kernel operators on the right. For a steerable kernel $K : X \to \mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}})$ *and a kernel operator* $\mathcal{K} : L^2_{\mathbb{K}}(X) \to \mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}})$*, these inverse maps are given by* $\hat{K}(f) := \int_X f(x)K(x)dx$ *and* $\mathcal{K}|_X(x) := \mathcal{K}(\delta_x)$*. Here,* δ_x *is the Dirac delta function of* $x \in X$*.*

This theorem requires some explanation. First of all, $\hat{K}$ is supposed to be a kernel operator, i.e. a map $L^2_{\mathbb{K}}(X) \to \mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}})$. Thus, $\hat{K}(f)$ should be a linear function $V_{\mathrm{in}} \to V_{\mathrm{out}}$. The formal expression of it can indeed be considered as such:

$$\hat{K}(f) = \int_X f(x)K(x)dx : v_{\mathrm{in}} \mapsto \int_X f(x)\left[K(x)\right](v_{\mathrm{in}})dx \in V_{\mathrm{out}}. \tag{3.3}$$

Due to the continuity of K proven in Proposition 3.1.5[4] and the integrability of f, the function $X \to V_{\mathrm{out}}$, $x \mapsto f(x)\left[K(x)\right](v_{\mathrm{in}})$ is also integrable, meaning the expression in Equation 3.3 can be evaluated. This explains the meaning of the map $\widehat{(\cdot)}$ in Theorem 3.1.7.

For the map $(\cdot)|_X$ in the other direction, we need to shortly explain what we mean by the Dirac delta function. A formal description will be given in Section 3.2.3, whereas here we focus on the intuitions. Such a "function" $\delta_x : X \to \mathbb{K}$ at a point $x \in X$ can be imagined as a function taking value infinity at x and zero elsewhere. It is characterized by the property that $\int_X \delta_x(x')f(x')dx' = f(x)$ for any function $f \in L^2_{\mathbb{K}}(X)$. We think of δ_x as being a function in $L^2_{\mathbb{K}}(X)$, even though technically, it is not in this space. This is since $\infty \notin \mathbb{K}$.

Now, $\mathcal{K}|_X(x) = \mathcal{K}(\delta_x)$ is defined as the value that $\mathcal{K}$ takes at the Dirac delta function δ_x. However, this is, at first sight, not well-defined since $\delta_x \notin L^2_{\mathbb{K}}(X)$. Formally, we can *approximate* the Dirac delta by scaled indicator functions δ_U with integral 1 on subsets U of X around x that get increasingly smaller. Then we can define $\mathcal{K}|_X(x)$ as the limit of $\mathcal{K}(\delta_U)$ as U tends to x.

Now that we have understood the formulation of the theorem, we might wonder, why should such a theorem be true? A first intuition comes from an analogy with linear algebra: Namely, assume B is a basis of a $\mathbb{K}$-vector space V and W any other vector space. Then linear maps $\hat{f} : V \to W$ are in one-to-one correspondence with (not assumed to be linear) functions $f : B \to W$, and this isomorphism is given by restriction and linear extension:

$$\begin{tikzcd} \mathrm{Hom}(B, W) \arrow[r, bend left, "\widehat{(\cdot)}"] & \mathrm{Hom}_{\mathbb{K}}(V, W). \arrow[l, bend left, "(\cdot)|_B"] \end{tikzcd}$$

Thus, we can think of the homogeneous space X as a "smooth basis" of the space of square-integrable functions. Sums are then replaced by integrals.

For the actual proof, one direction seems pretty clear from the properties of the Dirac delta:

$$\hat{K}\big|_X(x) = \hat{K}(\delta_x) = \int_X \delta_x(x')K(x')dx' = K(x).$$

[4]This means that all matrix elements of $K(x)$ for chosen bases of V_{in} and V_{out} are continuous.

But the other direction is less obvious: it seems like the space of kernel operators is considerably larger than the space of steerable kernels, since kernel operators are defined on a larger space. Therefore it is hard to believe that the construction is also inverse in the other direction. However, it pays off to ponder a bit more over what the Dirac delta construction does: Basically, we "embed" X into $L^2_{\mathbb{K}}(X)$ by means of the Dirac delta functions, i.e. $x \mapsto \delta_x$ and, as such, view X as a subset of $L^2_{\mathbb{K}}(X)$ (albeit a subset that is only in approximation in that space). Steerable kernels are then "partial" kernel operators in the sense that they are only defined on this subset $X \subseteq L^2_{\mathbb{K}}(X)$. What then needs to be understood is why there is only a unique extension of each steerable kernel K to a kernel operator $\mathcal{K}$ on the whole of $L^2_{\mathbb{K}}(X)$: if this is understood, then the space of kernel operators cannot be larger than the space of steerable kernels. And indeed, if there is an extension of K to $\mathcal{K}$ on $L^2_{\mathbb{K}}(X)$, it *has to* be unique: each $f \in L^2_{\mathbb{K}}(X)$ can be approximated by finite linear combinations of scaled indicator functions. Then by *linearity* of the kernel operator $\mathcal{K}$, we can evaluate $\mathcal{K}(f)$ by knowing $\mathcal{K}(\delta_U)$ for scaled indicator functions δ_U on small measurable sets U. And these approximate $K(x) = \mathcal{K}(\delta_x)$ for $x \in U$ arbitrarily well by construction. This determines the behavior of $\mathcal{K}$. The details of all of this can be found in the next section.

3.2. A Formal Proof of the Correspondence between Steerable Kernels and Kernel Operators

Here, we give a step-by-step proof of Theorem 3.1.7. The details of this investigation will not be needed later, and so a reader who is mainly interested in the applications to steerable CNNs can safely skip reading this section and go on reading Chapter 4.

3.2.1. A Reduction to Unitary Irreducible Representations

In this section, we make the proof more manageable by reducing $\mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}})$ to an irreducible representation. First, remember that Proposition 2.1.20 shows that there is a scalar product on $\mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}})$ such that it's Hom-representation becomes unitary. Since all norms on finite-dimensional spaces are equivalent, as is well known, this will not change the topology. Then, we can decompose $\mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}})$ into an orthogonal direct sum of irreducible unitary representations by Proposition 2.2.15. Let $\mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}}) \cong \bigoplus_{i=1}^n V_i$ be such a decomposition. We get canonical[5] isomorphisms

$$\mathrm{Hom}_G(X, \mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}})) \cong \bigoplus_{i=1}^{n} \mathrm{Hom}_G(X, V_i)$$

[5] "Canonical" once the decompositions into irreducible representations is already chosen.

and
$$\mathrm{Hom}_{G,\mathbb{K}}(L^2_{\mathbb{K}}(X), \mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}})) \cong \bigoplus_{i=1}^{n} \mathrm{Hom}_{G,\mathbb{K}}(L^2_{\mathbb{K}}(X), V_i).$$

Thus, we can show Theorem 3.1.7 by showing it for irreducible unitary representations instead of $\mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}})$. Overall, we have reduced our Theorem to the following, simpler statement:

Theorem 3.2.1. *Let $\rho : G \to \mathrm{U}(V)$ be an irreducible unitary representation and X a homogeneous space of G. Then there is an isomorphism*

$$\begin{tikzcd} \mathrm{Hom}_G(X, V) \arrow[r, bend left, "\widehat{(\cdot)}"] & \mathrm{Hom}_{G,\mathbb{K}}(L^2_{\mathbb{K}}(X), V) \arrow[l, bend left, "(\cdot)|_X"] \end{tikzcd}$$

which is given as follows: for $K \in \mathrm{Hom}_G(X, V)$ we set $\hat{K}(f) = \int_X f(x)K(x)dx$ and for $\mathcal{K} \in \mathrm{Hom}_{G,\mathbb{K}}(L^2_{\mathbb{K}}(X), V)$ we set $\mathcal{K}|_X(x) = \mathcal{K}(\delta_x)$, with δ_x being the Dirac delta function at point x.

From now on, we assume that X and $\rho : G \to \mathrm{U}(V)$ is fixed as in the formulation of Theorem 3.2.1.

3.2.2. Well-Definedness of $\widehat{(\cdot)}$

Lemma 3.2.2. *The function $\widehat{(\cdot)} : \mathrm{Hom}_G(X, V) \to \mathrm{Hom}_{G,\mathbb{K}}(L^2_{\mathbb{K}}(X), V)$ is well-defined, i.e.: For an equivariant function $K : X \to V$, the function $\hat{K} : L^2_{\mathbb{K}}(X) \to V$ is linear, equivariant and continuous.*

Proof. Linearity of $\hat{K}$ is clear. Equivariance can be proven using the equivariance of K and the left invariance of the Haar measure on the homogeneous space X:

$$\begin{aligned}
\hat{K}(\lambda(g)f) &= \int_X (\lambda(g)f)(x)K(x)dx \\
&= \int_X f(g^{-1} \cdot x)K(x)dx \\
&= \int_X f(x)K(g \cdot x)dx \\
&= \int_X f(x) \left[\rho(g) \left(K(x)\right)\right] dx \\
&= \rho(g) \left[\int_X f(x)K(x)dx\right] \\
&= \rho(g) \left[\hat{K}(f)\right].
\end{aligned}$$

The action by $\rho(g)$ could be put out of the integral since $\rho(g)$ it is linear and continuous, and since integrals can be approximated by finite sums.
Now about continuity: By Proposition A.1.18, we only need to show continuity in 0. Thus, let $(f_k)_k$ be a sequence of functions $f_k \in L^2_{\mathbb{K}}(X)$ with $\lim_{k\to\infty} \|f_k\|_{L^2} = 0$. Then we obtain

$$\begin{aligned}
\|\hat{K}(f_k)\|_V &= \left\| \int_X f_k(x) K(x) dx \right\|_V \\
&\le \int_X |f_k(x)| \cdot \|K(x)\|_V dx \\
&\le \max_{x'} \|K(x')\|_V \cdot \int_X |f_k(x)| dx,
\end{aligned}$$

where the continuity of K proven in Proposition 3.1.5 was used[6]. For the right expression, using the Cauchy-Schwarz inequality Proposition A.2.3 we obtain

$$\begin{aligned}
\int_X |f_k(x)| dx &= \int_X |f_k(x)| \cdot 1 dx \\
&= |\langle |f_k| \mid 1 \rangle| \\
&\le \|f_k\|_{L^2} \cdot \|1\|_{L^2} \\
&= \|f_k\|_{L^2}.
\end{aligned}$$

So, overall, if $\lim_{k\to\infty} \|f_k\|_{L^2} = 0$, then $\lim_{k\to\infty} \|\hat{K}(f_k)\|_V = 0$ as well, which proves continuity. □

3.2.3. Construction and Well-Definedness of $(\cdot)|_X$

In order to show the well-definedness of the function $\mathcal{K} \mapsto \mathcal{K}|_X$, we first need to clarify the definition of this function. In the limiting process of its construction, we unfortunately cannot rely on the usual definitions using sequences indexed by natural numbers since there may be "too many open neighborhoods of points". Thus we first need to discuss the more general directed sets and sequences indexed by them, i.e. so-called *nets* [23].

Definition 3.2.3 (Partially Ordered Set, Directed Set)**.** Let I be any index set and $\le$ a relation on it. $I = (I, \le)$ is a *partially ordered set* if:

1. $\le$ is *reflexive*, i.e. $i \le i$ for all $i \in I$.
2. $\le$ is *antisymmetric*, that is: $i \le j$ and $j \le i$ together imply $i = j$.
3. $\le$ is *transitive*, that is: $i \le j$ and $j \le k$ together imply $i \le k$.

[6] since $\|K\|$ is continuous on X, which is compact by Proposition A.1.8 as an image of the compact group G, it has a maximum by Corollary A.1.25.

A partially ordered set I is called *directed* if for all $i, j \in I$ there exists $k \in I$ such that $i \leq k$ and $j \leq k$.

Example 3.2.4. Clearly, the natural numbers together with the standard order relation form a directed set.
An important example for our purposes is the following: let Z be any topological space (for example our homogeneous space X) and $x \in Z$ be any point. Furthermore, define $\mathcal{U}_x$ as the set of open neighborhoods of x, i.e. open sets $U \subseteq Z$ such that $x \in U$. On this set, we define $U \leq V$ if $U \supseteq V$, i.e. by *reversed* inclusion. Then $(\mathcal{U}_x, \leq)$ is a directed set:

1. Reflexivity is clear since $V \supseteq V$ for all V.

2. Antisymmetry is clear since $U \supseteq V$ and $V \supseteq U$ together clearly imply $U = V$.

3. Transitivity is clear since $U \supseteq V$ and $V \supseteq W$ together clearly imply $U \supseteq W$.

4. For directedness, let $U, V \in \mathcal{U}_x$. Define $W = U \cap V$. Then $W \in \mathcal{U}_x$ and clearly $U \supseteq W$ and $V \supseteq W$, which is what was to show.

Note that $\mathcal{U}_x$ is usually *not totally ordered*, i.e. there are usually $U, V \in \mathcal{U}_x$ such that neither $U \supseteq V$ nor $V \supseteq U$.

Definition 3.2.5 (Net)**.** Let Z be any topological space and I a directed set. Then a *net* in Z is a function $x : I \to Z$. We write a net as $(x_i)_{i \in I}$, in analogy to sequences.

Definition 3.2.6 (Convergence of Nets)**.** Let $(x_i)_{i \in I}$ be a net in a topological space Z. Let $x \in Z$. We say that $(x_i)_{i \in I}$ *converges to* x, written $\lim_{i \in I} x_i = x$, if the following holds: For all open neighborhoods U of x there is an $i_0 \in I$ such that for all $i \geq i_0$ we have $x_i \in U$.

Definition 3.2.7 (Approximated Dirac Delta)**.** For $\emptyset \neq U \subseteq X$ open, we define the approximated Dirac delta by

$$\delta_U(x) = \frac{1}{\mu(U)} \cdot \mathbf{1}_U = \begin{cases} \frac{1}{\mu(U)}, & x \in U \\ 0, & \text{else.} \end{cases}$$

A priori, it is unclear that open sets have positive measure, which is needed for the well-definedness of this construction. Thus, we need the following lemma:

Lemma 3.2.8. *Let $\emptyset \neq U \subseteq X$ be an open set. Then $\mu(U) > 0$.*

Proof. Consider the family of open sets $(gU)_{g \in G}$. That all of these sets are necessarily open follows since the action $G \times X \to X$ is continuous, and thus by the definition of a group action, each $g \in G$ induces a homeomorphism $X \to X, x \mapsto gx$. Now, since the action is transitive, $(gU)_{g \in G}$ is an open cover of X, and since X is compact, see Definition A.1.7, it has an open subcover $(g_iU)_{i=1}^n$ with $g_i \in G$. Note that $\mu(g_iU) =$

$\mu(U)$ for all i since the measure μ on X is by definition left invariant under the action of G. Overall, we obtain

$$1 = \mu(X) = \mu\left(\bigcup_{i=1}^{n} g_i U\right) \leq \sum_{i=1}^{n} \mu(g_i U) = \sum_{i=1}^{n} \mu(U) = n \cdot \mu(U)$$

and thus $\mu(U) \geq \frac{1}{n} > 0$. □

Now we have established all concepts needed for defining the map $\mathcal{K}|_X$. What we call "intertwiners" in what follows are basically kernel operators, only that the space V does not have the structure of a Hom-space.

Definition 3.2.9 (Restriction of intertwiners)**.** Let $\mathcal{K} : L^2_{\mathbb{K}}(X) \to V$ be an intertwiner. Let $x \in X$. Let $\mathcal{U}_x$ be the net of open neighborhoods of x from Example 3.2.4. Then we define

$$\mathcal{K}|_X(x) := \mathcal{K}(\delta_x) := \lim_{U \in \mathcal{U}_x} \mathcal{K}(\delta_U),$$

where δ_U is the approximated Dirac delta from Definition 3.2.7 and where the limit is a limit of nets as in Definition 3.2.6.

While it is clear that this limit is unique if it exists [23], it is somewhat unclear why it exists in the first place. For this, we need to better understand the properties of the (approximated) Dirac delta. The most important one is the following, which we hinted at already in the intuitions we gave before this section: basically, Dirac deltas help for evaluating continuous functions at specific points:

Lemma 3.2.10. *For each $x \in X$ and $Y : X \to \mathbb{K}$ continuous we have* $\lim_{U \in \mathcal{U}_x} \langle \delta_U | Y \rangle = Y(x)$.

Proof. We have

$$\begin{aligned}
\left| \langle \delta_U | Y \rangle - Y(x) \right| &= \left| \int_X \delta_U(x') Y(x') dx' - \mu(U) \cdot \frac{1}{\mu(U)} Y(x) \right| \\
&= \left| \int_U \frac{1}{\mu(U)} Y(x') dx' - \int_U \frac{1}{\mu(U)} Y(x) dx' \right| \\
&= \left| \int_U \frac{1}{\mu(U)} (Y(x') - Y(x)) dx' \right| \\
&\leq \int_U \frac{1}{\mu(U)} \left| Y(x') - Y(x) \right| dx'.
\end{aligned}$$

Let $\epsilon > 0$. Since Y is continuous in x, there is $U_\epsilon \in \mathcal{U}_x$ such that $Y(x') \in \mathrm{B}_\epsilon(Y(x))$ for all $x' \in U_\epsilon$ or, equivalently, $|Y(x') - Y(x)| < \epsilon$. Thus, for all $U_\epsilon \supseteq U$, i.e. all $U_\epsilon \leq U$ in $\mathcal{U}_x$ we obtain

$$\left| \langle \delta_U | Y \rangle - Y(x) \right| \leq \int_U \frac{1}{\mu(U)} |Y(x') - Y(x)| dx'$$

$$\leq \int_U \frac{1}{\mu(U)} \epsilon dx'$$
$$= \epsilon \cdot \mu(U) \cdot \frac{1}{\mu(U)}$$
$$= \epsilon$$

and consequently $\lim_{U \in \mathcal{U}_x} \langle \delta_U | Y \rangle = Y(x)$. □

Before we can show the well-definedness of $\mathcal{K}|_X$, we first want to get a better description of $\mathcal{K}$. For that, recall from the Peter-Weyl Theorem that $L^2_{\mathbb{K}}(X) = \widehat{\bigoplus}_{l \in \hat{G}} \bigoplus_{i=1}^{m_l} V_{li}$. With this at our disposal, we can formulate the following Lemma on the form of intertwiners on $L^2_{\mathbb{K}}(X)$:

Lemma 3.2.11. *Let $\mathcal{K} : L^2_{\mathbb{K}}(X) \to V$ be an intertwiner. Let $l \in \hat{G}$ be the unique index such that $V \cong V_{li}$ for all $i = 1, \dots, m_l$. Let Y_{li}^n, $n = 1, \dots, [l]$ be an orthonormal basis of V_{li} where $[l] = \dim(V_l)$. Then*

$$\mathcal{K}(f) = \sum\nolimits_{i=1}^{m_l} \sum\nolimits_{n=1}^{[l]} \langle Y_{li}^n | f \rangle \, \mathcal{K}(Y_{li}^n)$$

for all $f \in L^2_{\mathbb{K}}(X)$.

Proof. We can write $f \in L^2_{\mathbb{K}}(X)$ according to the discussion after Definition A.2.9 as

$$f = \sum\nolimits_{l' \in \hat{G}} \sum\nolimits_{i=1}^{m_{l'}} \sum\nolimits_{n=1}^{[l']} \langle Y_{l'i}^n | f \rangle \, Y_{l'i}^n.$$

Note that $\mathcal{K}|_{V_{l'i}} : V_{l'i} \to V$ is an intertwiner as well, and so by Schur's Lemma 2.2.6 it is necessarily zero unless $l' = l$ is the unique index such that $V_{li} \cong V$. Due to its continuity and linearity, $\mathcal{K}$ commutes with infinite sums and we obtain

$$\begin{aligned}
\mathcal{K}(f) &= \sum\nolimits_{l' \in \hat{G}} \sum\nolimits_{i=1}^{m_{l'}} \sum\nolimits_{n=1}^{[l']} \langle Y_{l'i}^n | f \rangle \, \mathcal{K}\left(Y_{l'i}^n\right) \\
&= \sum\nolimits_{l' \in \hat{G}} \sum\nolimits_{i=1}^{m_{l'}} \sum\nolimits_{n=1}^{[l']} \langle Y_{l'i}^n | f \rangle \, \mathcal{K}|_{V_{l'i}}\left(Y_{l'i}^n\right) \\
&= \sum\nolimits_{i=1}^{m_l} \sum\nolimits_{n=1}^{[l]} \langle Y_{li}^n | f \rangle \, \mathcal{K}(Y_{li}^n).
\end{aligned}$$

□

Corollary 3.2.12. *We have $\mathcal{K}|_X(x) = \sum_{i=1}^{m_l} \sum_{n=1}^{[l]} \overline{Y_{li}^n(x)} \mathcal{K}(Y_{li}^n)$. In particular, the defining limit exists.*

Proof. Since the Y_{li}^n are by the proof of the Peter-Weyl Theorem in the finite-dimensional space $\mathcal{E}_l$ spanned by matrix coefficients of the irreducible representation $\rho_l : G \to \mathrm{U}(V_l)$ and since these matrix coefficients are continuous by Remark 2.2.2, the Y_{li}^n are as finite linear combinations of them also continuous functions. Thus, from Lemma 3.2.10 and 3.2.11 together we obtain:

$$\mathcal{K}|_X(x) = \lim_{U \in \mathcal{U}_x} \mathcal{K}\left(\delta_U\right)$$

$$
\begin{aligned}
&= \lim_{U \in \mathcal{U}_x} \sum\nolimits_{i=1}^{m_l} \sum\nolimits_{n=1}^{[l]} \langle Y_{li}^n | \delta_U \rangle \, \mathcal{K}(Y_{li}^n) \\
&= \sum\nolimits_{i=1}^{m_l} \sum\nolimits_{n=1}^{[l]} \left[\lim_{U \in \mathcal{U}_x} \langle Y_{li}^n | \delta_U \rangle \right] \mathcal{K}(Y_{li}^n) \\
&= \sum\nolimits_{i=1}^{m_l} \sum\nolimits_{n=1}^{[l]} \overline{Y_{li}^n(x)} \mathcal{K}(Y_{li}^n).
\end{aligned}
$$

The complex conjugation came into play since the order in the scalar product is swapped compared to Lemma 3.2.10. $\square$

Thus, since we now know that $\mathcal{K}|_X$ *as a function* makes sense, we can finally prove the well-definedness of $\mathcal{K} \mapsto \mathcal{K}|_X$,

Lemma 3.2.13. *The function $(\cdot)|_X : \mathrm{Hom}_{G,\mathbb{K}}(L^2_{\mathbb{K}}(X), V) \to \mathrm{Hom}_G(X, V)$ is well-defined, that is: for a linear, equivariant and continuous function $\mathcal{K} : L^2_{\mathbb{K}}(X) \to V$, the restriction $\mathcal{K}|_X : X \to V$ is equivariant.*

Proof. We have

$$
\begin{aligned}
\mathcal{K}|_X(g \cdot x) &= \lim_{U \in \mathcal{U}_{gx}} \mathcal{K}(\delta_U) \\
&= \lim_{U \in \mathcal{U}_x} \mathcal{K}(\delta_{gU}) \\
&= \lim_{U \in \mathcal{U}_x} \mathcal{K}(\lambda(g)\delta_U) \\
&= \lim_{U \in \mathcal{U}_x} \rho(g) \left[\mathcal{K}(\delta_U)\right] \\
&= \rho(g) \Big[\lim_{U \in \mathcal{U}_x} \mathcal{K}(\delta_U) \Big] \\
&= \rho(g) \left[\mathcal{K}|_X(x)\right],
\end{aligned}
$$

where the steps are justified as follows: The first step is just the definition of $\mathcal{K}|_X$. The second step uses that the open neighborhood of gx are precisely the g-translated open neighborhoods of x since $g : X \to X$ is a homeomorphism. The third step is easy to check. The fourth step uses the equivariance of $\mathcal{K}$. The fifth step uses the continuity of $\rho(g)$, which follows since $\rho(g)$ is a unitary transformation. The last step is again the definition of $\mathcal{K}|_X$. $\square$

3.2.4. $\widehat{(\cdot)}$ and $(\cdot)|_X$ Are Inverse to Each Other

With all this preparation, we can finish the proof of Theorem 3.2.1 and consequently of Theorem 3.1.7:

Proof of Theorem 3.2.1. After all the preparation, we only need to still show that the maps $\widehat{(\cdot)}$ and $(\cdot)|_X$ are inverse to each other. For $\hat{K}\big|_X = K$, i.e. the injectivity of the function $K \mapsto \hat{K}$ and surjectivity of the function $\mathcal{K} \mapsto \mathcal{K}|_X$, we compute:

$$
\hat{K}\big|_X(x) = \lim_{U \in \mathcal{U}_x} \hat{K}(\delta_U)
$$

$$
\begin{aligned}
&= \lim_{U \in \mathcal{U}_x} \int_X \delta_U(x') K(x') dx' \\
&= K(x).
\end{aligned}
$$

The last step follows from Lemma 3.2.10 by identifying $V = V_l$ with $\mathbb{K}^{[l]}$ and viewing K as consisting of continuous component functions $K^n : X \to \mathbb{K}$, $n \in \{1, \dots, [l]\}$. The continuity of K was shown in Proposition 3.1.5.
For showing $\widehat{\mathcal{K}|_X} = \mathcal{K}$ we do a computation using the description of $\mathcal{K}$ from Lemma 3.2.11 and the description of $\mathcal{K}|_X$ from Corollary 3.2.12:

$$
\begin{aligned}
\widehat{\mathcal{K}|_X}(f) &= \int_X f(x) \mathcal{K}|_X(x) dx \\
&= \int_X f(x) \Big(\sum\nolimits_{i=1}^{m_l} \sum\nolimits_{n=1}^{[l]} \overline{Y_{li}^n(x)} \mathcal{K}(Y_{li}^n) \Big) dx \\
&= \sum\nolimits_{i=1}^{m_l} \sum\nolimits_{n=1}^{[l]} \Big(\int_X f(x) \overline{Y_{li}^n(x)} dx \Big) \mathcal{K}(Y_{li}^n) \\
&= \sum\nolimits_{i=1}^{m_l} \sum\nolimits_{n=1}^{[l]} \langle Y_{li}^n | f \rangle \, \mathcal{K}(Y_{li}^n) \\
&= \mathcal{K}(f).
\end{aligned}
$$

This finally finishes the proof. $\square$

4. A Wigner-Eckart Theorem for Steerable Kernels of General Compact Groups

In Chapter 3 we have seen the most important theoretical insight of this work: steerable kernels on a homogeneous space X correspond one-to-one to kernel operators (certain representation operators) on the space of square-integrable functions $L^2_{\mathbb{K}}(X)$. In this chapter, we will develop the most important consequence of this correspondence: a Wigner-Eckart Theorem for steerable kernels and consequently a description of a basis for steerable kernels. This works for both fields $\mathbb{R}$ and $\mathbb{C}$, for an arbitrary compact group G, an arbitrary homogeneous space X and arbitrary finite-dimensional input- and output fields.

In Section 4.1 we will work towards formulating the most important theorems. Since these will involve tensor products, we will start with defining and studying tensor products of pre-Hilbert spaces and (unitary) representations. Afterward, we will define the Clebsch-Gordan coefficients, which relate a tensor product of irreducible representations to the irreducible subrepresentations of this tensor product. This will lead to a formulation of the original Wigner-Eckart Theorem similar as it appears in quantum mechanics, including a proof. The original Wigner-Eckart Theorem is a statement about representation operators on irreducible representations. However, we consider kernel operators on $L^2_{\mathbb{K}}(X)$ which is not irreducible. Also, different from the original Theorem, we also consider representations over the real numbers, which leads to a replacement of reduced matrix elements by *endomorphisms*. Therefore we then formulate a generalization of the original theorem. Then, using the correspondence between kernel operators and steerable kernels from Theorem 3.1.7, we can transform this into a Wigner-Eckart Theorem for steerable kernels and ultimately a statement about a basis of the space of steerable kernels. We conclude with some remarks about how to use the basis kernels in practice.

Afterward, in Section 4.2, we give the remaining proof of the Wigner-Eckart Theorem for kernel operators, which we omit in the section before. First, we reduce the statement to the dense subspace of $L^2_{\mathbb{K}}(X)$ which is a direct sum of all irreducible subrepresentations. We then describe a correspondence between representation operators and intertwiners on a certain tensor product, the so-called hom-tensor adjunction. Finally, we finish with the full proof of the Wigner-Eckart Theorem.

As always, let $\mathbb{K}$ be either of the two fields $\mathbb{R}$ and $\mathbb{C}$ and G be a compact topological group. X is any homogeneous space of G.

4.1. A Wigner-Eckart Theorem for Steerable Kernels and their Kernel Bases

4.1.1. Tensor Products of pre-Hilbert Spaces and Unitary Representations

In order to state the Wigner-Eckart Theorem, we need the notion of representations on tensor products. This is defined similarly to Hom-representations, see Definition 3.1.1. For this, we first need to discuss the notion of a tensor product of vector spaces:

Definition 4.1.1 (Tensor Product)**.** Let V and V' be two vector spaces over $\mathbb{K}$. Then $V \otimes V'$, the *tensor product* of V and V', is a vector space over $\mathbb{K}$ with the following properties:

1. There is a bilinear function $\otimes : V \times V' \to V \otimes V'$, $(v, v') \mapsto v \otimes v'$. $V \otimes V'$ is generated by elements of the form $v \otimes v'$.

2. It has the following universal property: for any *bilinear* function $\beta : V \times V' \to P$ into a vector space P, there is a *unique linear* function $\overline{\beta} : V \otimes V' \to P$ given on elements of the form $v \otimes v'$ by $\overline{\beta}(v \otimes v') = \beta(v, v')$. In other words, the following diagram commutes:

$$\begin{tikzcd} V \times V' \arrow[r, "\beta"] \arrow[d, "\otimes"'] & P \\ V \otimes V' \arrow[ur, "\overline{\beta}"'] & \end{tikzcd}$$

3. If V and V' are finite-dimensional with bases $\{v_1, \dots, v_n\} \subseteq V$ and $\{v'_1, \dots, v'_m\} \subseteq V'$, then $\{v_i \otimes v'_j\}_{i,j} \subseteq V \otimes V'$ is a basis of $V \otimes V'$. In particular, the dimension of $V \otimes V'$ is $n \cdot m$.

Property 3 follows from 1 and 2 and would therefore not necessarily be needed in the definition. The explicit construction of tensor products shall not matter for our purposes since the properties above characterize it up to isomorphism. The second property stated in the definition is of large importance since it tells us how we can define linear functions on $V \otimes V'$: if we have a guess for such a function $\varphi : V \otimes V' \to P$ (of which we don't yet know whether its "assignment rule" is well-defined), then we just need to test whether the function $\tilde{\varphi} : V \times V' \to P$ given by $\tilde{\varphi}(v, v') := \varphi(v \otimes v')$ is bilinear. If it is, then φ is a well-defined linear function. We will use this soon in the following context: Assume $f : V \to V$ and $g : V' \to V'$ are linear functions. Then we would like to define a function $f \otimes g : V \otimes V' \to V \otimes V'$ by $(f \otimes g)(v \otimes v') = f(v) \otimes g(v')$. For this to work, we need to test whether the assignment $(v, v') \mapsto f(v) \otimes g(v')$ is a bilinear function $V \times V' \to V \otimes V'$. Clearly, it is, and so $f \otimes g$ is a well-defined linear function! We use this in Definition 4.1.3 in order to define the tensor product of representations.

Since we actually deal with Hilbert spaces most of the time, we would like to build tensor products of Hilbert spaces. However, their definition is not completely straightforward since one cannot just take the tensor product of the underlying vector spaces but needs to additionally build the *completion* of the resulting space [25]. Since this complicates the considerations related to a correspondence we later formulate in Proposition 4.2.4, we go a slightly different route. Instead of describing the tensor product of Hilbert spaces, we describe the tensor product of *pre-Hilbert spaces*, which does not require a completion step. Recall from Definition A.2 that a pre-Hilbert space is basically a Hilbert space that is not necessarily complete.

Definition 4.1.2 (Tensor product of pre-Hilbert spaces)**.** Let V, V' be two pre-Hilbert spaces with scalar products $\langle \cdot | \cdot \rangle$ and $\langle \cdot | \cdot \rangle'$. Then the tensor product of vector spaces $V \otimes V'$ can be made into a pre-Hilbert space using the scalar product which is given on generators by

$$\langle v \otimes v' | w \otimes w' \rangle_{\otimes} \coloneqq \langle v | w \rangle \cdot \langle v' | w' \rangle' .$$

This is then anti-linearly extended in the first (i.e. "Bra"), and linearly extended in the second (i.e. "Ket") component.

One can show that this makes $V \otimes V'$ a pre-Hilbert space. For simplicity, we will from now on not notationally distinguish the different scalar products involved. With this preparation, we can come to the notion of tensor product representations:

Definition 4.1.3 (Tensor Product Representation)**.** Let $\rho : G \to \mathrm{Aut}_{\mathbb{K}}(V)$ and $\rho' : G \to \mathrm{Aut}_{\mathbb{K}}(V')$ be two linear representations, where V and V' are pre-Hilbert spaces. Then on the tensor product $V \otimes V'$ of pre-Hilbert spaces, we can define the *tensor product representation* $\rho \otimes \rho'$ by

$$\rho \otimes \rho' : G \to \mathrm{Aut}_{\mathbb{K}}(V \otimes V'), \ g \mapsto \rho(g) \otimes \rho'(g),$$

where $\rho(g) \otimes \rho'(g) : V \otimes V' \to V \otimes V'$ is given on generators by

$$(\rho(g) \otimes \rho'(g)) \, (v \otimes v') \coloneqq \rho(g)(v) \otimes \rho'(g)(v').$$

Lemma 4.1.4. *The map $\rho \otimes \rho' : G \to \mathrm{Aut}_{\mathbb{K}}(V \otimes V')$ defined above is a linear representation.*

Proof. Clearly, each $(\rho \otimes \rho')(g)$ is linear and we have $(\rho \otimes \rho')(gg') = (\rho \otimes \rho')(g) \circ (\rho \otimes \rho')(g')$. Thus, for showing that it is a linear representation, we need to show it is continuous. Assume we already knew continuity of all maps $(\rho \otimes \rho')^{v \otimes v'} : G \to V \otimes V'$, $g \mapsto \big[(\rho \otimes \rho')(g)\big](v \otimes v')$. Then for linear combinations $\xi = \sum_{i=1}^{n} \lambda_i \, (v_i \otimes v_i')$ we obtain using the linearity of $(\rho \otimes \rho')(g)$:

$$\begin{aligned}
(\rho \otimes \rho')^{\xi}(g) &= \big[(\rho \otimes \rho')(g)\big](\xi) \\
&= \big[(\rho \otimes \rho')(g)\big] \left(\sum\nolimits_{i=1}^{n} \lambda_i \, (v_i \otimes v_i') \right) \\
&= \sum\nolimits_{i=1}^{n} \lambda_i \big[(\rho \otimes \rho')(g)\big](v_i \otimes v_i')
\end{aligned}$$

$$= \left(\sum_{i=1}^{n} \lambda_i (\rho \otimes \rho')^{v_i \otimes v_i'}\right)(g).$$

Now, since scalar multiplication and addition in topological vector spaces is continuous, and since pre-Hilbert spaces are special topological vector spaces, the continuity of $(\rho \otimes \rho')^{\xi}$ follows from that of all $(\rho \otimes \rho')^{v \otimes v'}$.
What's left is proving the continuity of functions of the form $(\rho \otimes \rho')^{v \otimes v'}$. For notational simplicity, write $f = \rho^{v} : G \to V$ and $f' : \rho'^{v'}$, which are both continuous since ρ and ρ' are linear representations. We want to show that also $f \otimes f' : G \to V \otimes V'$ is continuous. We can test continuity in each point $g_0 \in G$ separately by Definition A.1.6. For each $g \in G$ we then obtain, with Re being the real part of a complex number:

$$\begin{aligned}
&\|(f \otimes f')(g) - (f \otimes f')(g_0)\|^2 \\
&\qquad = \Big\| \left[f(g) \otimes f'(g) - f(g) \otimes f'(g_0)\right] + \left[f(g) \otimes f'(g_0) - f(g_0) \otimes f'(g_0)\right] \Big\|^2 \\
&\qquad = \Big\| f(g) \otimes \left[f'(g) - f'(g_0)\right] + \left[f(g) - f(g_0)\right] \otimes f'(g_0) \Big\|^2 \\
&\qquad = \Big\| f(g) \otimes \left[f'(g) - f'(g_0)\right] \Big\|^2 + \Big\| \left[f(g) - f(g_0)\right] \otimes f'(g_0) \Big\|^2 \\
&\qquad\quad + 2\,\mathrm{Re} \left\langle f(g) \otimes \left[f'(g) - f'(g_0)\right] \,\middle|\, \left[f(g) - f(g_0)\right] \otimes f'(g_0) \right\rangle \\
&\qquad = \|f(g)\|^2 \cdot \|f'(g) - f'(g_0)\|^2 + \|f(g) - f(g_0)\|^2 \cdot \|f'(g_0)\|^2 \\
&\qquad\quad + 2\,\mathrm{Re} \Big(\langle f(g) | f(g) - f(g_0) \rangle \cdot \langle f'(g) - f'(g_0) | f'(g_0) \rangle \Big).
\end{aligned}$$

All in all we see the following: If g is sufficiently close to g_0, then due to the continuity of f, f', the scalar product, multiplication in $\mathbb{K}$ and the real part, $\|(f \otimes f')(g) - (f \otimes f')(g_0)\|^2$ gets arbitrarily close to 0. This shows the continuity of $f \otimes f'$ and we are done. □

Lemma 4.1.5. *Let $\rho : G \to \mathrm{U}(V)$ and $\rho' : G \to \mathrm{U}(V')$ be* unitary *representations on pre-Hilbert spaces. Then also $\rho \otimes \rho' : G \to \mathrm{U}(V \otimes V')$ is a well-defined unitary representation.*

Proof. According to Lemma 4.1.4 we only need to check whether all $\rho(g) \otimes \rho'(g)$ are unitary transformations. This follows immediately from the unitarity of $\rho(g)$ and $\rho'(g)$. □

4.1.2. The Clebsch-Gordan Coefficients and the Original Wigner-Eckart Theorem

In this section, we describe the Clebsch-Gordan coefficients and the original Wigner-Eckart Theorem. Except for the proof, we roughly follow Jeevanjee [10]. For the proof, we follow the more general treatment in Agrawala [26][1].

[1] It is more general in that it considers arbitrary groups and the situation that the considered irreducible representation appears several times in a tensor product representation instead of just once.

For our aims, let $\rho_j : G \to \mathrm{U}(V_j)$ and $\rho_l : G \to \mathrm{U}(V_l)$ be representatives of isomorphism classes of irreducible unitary representations[2]. Then consider their tensor product representation

$$\rho_j \otimes \rho_l : G \to \mathrm{U}(V_j \otimes V_l)$$

which is again a unitary representation according to Lemma 4.1.5. If V_j and V_l are of dimension $[j]$ and $[l]$, respectively, then $V_j \otimes V_l$ is of dimension $[j] \cdot [l]$. Since it is a finite-dimensional unitary representation, it is itself an orthogonal direct sum of finitely many irreducible unitary representations by Proposition 2.2.15:

$$V_j \otimes V_l \cong \bigoplus_{J \in \hat{G}} \bigoplus_{s=1}^{[J(jl)]} V_J.$$

Here $\hat{G}$ is, as before, the set of isomorphism classes of irreducible unitary representations and $[J(jl)]$ is the number of times that $\rho_J : G \to \mathrm{U}(V_J)$ appears in the direct sum decomposition of $V_j \otimes V_l$. Note that for most J we have $[J(jl)] = 0$, and for some J we may have $[J(jl)] > 1$, see Section 6.2, where it turns out that ρ_0 is contained twice in $\rho_m \otimes \rho_m$.
Now, choose — once and for all — orthonormal bases of all involved irreps, which exists according to Proposition A.2.10:

$$\begin{aligned} \left\{Y_j^m \mid m = 1, \dots, [j]\right\} &\subseteq V_j, \\ \left\{Y_l^n \mid n = 1, \dots, [l]\right\} &\subseteq V_l, \\ \left\{Y_J^M \mid M = 1, \dots, [J]\right\} &\subseteq V_J. \end{aligned}$$

This notation is supposed to remind about spherical harmonics since they form a basis for irreducible representations of the group $\mathrm{SO}(3)$. But as mentioned in the footnote, we do not consider these basis elements to be functions here.
Furthermore, let $l_s : V_J \to V_j \otimes V_l$ be the linear, equivariant and isometric (i.e. scalar product preserving) embeddings that correspond to the direct sum decomposition of $V_j \otimes V_l$ into irreps, where s ranges in $\{1, \dots, [J(jl)]\}$. With this in mind, we can define the Clebsch-Gordan coefficients:

Definition 4.1.6 (Clebsch-Gordan Coefficients)**.** The *Clebsch-Gordan Coefficients* are given by

$$\langle s, JM | jmln \rangle := \left\langle l_s(Y_J^M) \middle| Y_j^m \otimes Y_l^n \right\rangle.$$

Note that in the literature, people usually only consider Clebsch-Gordan coefficients of the specific groups $\mathrm{SO}(3)$, $\mathrm{SU}(2)$, $\mathrm{SU}(3)$ or similar groups appearing in physics. Also note that in the physics context, there is only one linear, equivariant, isometric embedding l_s, which follows directly from Schur's Lemma 4.1.8. Therefore, it is sensible that the embedding is usually not part of the notation of these coefficients. In

[2] Those are a priori *not* assumed to be embedded in a space of square-integrable functions. For such embedded representations, we write V_{ji} instead.

our case, however, when considering real representations, there can be several such embeddings l_s. This happens if the endomorphism space of V_J is nontrivial. An example is given by the two-dimensional irreducible representations of $\mathrm{SO}(2)$ over the real numbers which we discuss in Section 6.2. Since, however, we do not want to depart too much from the notation usually considered in physics, we also omit the embedding from the notation. The index s however needs to be present in order to index the possibly different appearances of V_J in $V_j \otimes V_l$.

With this preparation, we can explain the Wigner-Eckart Theorem the way it is usually considered in physics, as a prelude for the generalization that we consider in the next section.

In this (and only this!) section, we assume that our field is $\mathbb{C}$, since this is the case considered in physics. The Wigner-Eckart Theorem aims to obtain a description for all possible representation operators $\mathcal{K} : V_j \to \mathrm{Hom}_{\mathbb{C}}(V_l, V_J)$. As discussed in the introduction, this is, for example, useful for describing state transitions in the electrons of hydrogen atoms. To motivate the generalization in the next section, we shortly explain the derivation: we can consider the equivalent function $\mathcal{K} : V_j \otimes V_l \to V_J$ given by $\tilde{\mathcal{K}}(v_j \otimes v_l) := [\mathcal{K}(v_j)]\,(v_l)$ on the tensor product. As one can compute, and as we will see in more generality in Proposition 4.2.4, $\tilde{\mathcal{K}} : V_j \otimes V_l \to V_J$ is an intertwiner, where on the left we consider the tensor product representation. We assume, as is the case for $G = \mathrm{SO}(3)$ or $G = \mathrm{SU}(2)$ for usual applications in physics, that V_J is exactly once a direct summand of $V_j \otimes V_l$. Then, since by Schur's Lemma 2.2.6 there cannot be nontrivial equivariant linear maps between nonisomorphic irreps, $\tilde{\mathcal{K}}$ restricted to each direct summand of $V_j \otimes V_l$ vanishes, except the one isomorphic to V_J. More precisely, assume that

$$V_j \otimes V_l \cong V_J \oplus \bigoplus_{l'} V_{l'}$$

is a decomposition of $V_j \otimes V_l$ into copies of irreducible representations, where each $V_{l'}$ is nonisomorphic to V_J. Then the information contained in $\tilde{\mathcal{K}}$ is essentially equal to the information contained in the restriction $\tilde{\mathcal{K}}|_{V_J} : V_J \to V_J$. Since it is an intertwiner from a representation to itself, it deserves a special name. We state the following definition for arbitrary $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, since it will be of crucial importance in our generalization of the Wigner-Eckart Theorem:

Definition 4.1.7 (Endomorphism)**.** Let $\rho : G \to \mathrm{Aut}_{\mathbb{K}}(V)$ be a linear representation. An intertwiner from V to V is called *endomorphism*. The vector space of endomorphisms is written as

$$\mathrm{End}_{G,\mathbb{K}}(V) := \mathrm{Hom}_{G,\mathbb{K}}(V, V).$$

A version of Schur's Lemma gives a very simple description for endomorphisms of irreducible representations in the case that the underlying field is the complex numbers $\mathbb{C}$. It makes use of the property of the complex numbers to be algebraically closed:

Lemma 4.1.8 (Schur's Lemma)**.** *Let $\rho : G \to \mathrm{Aut}_{\mathbb{K}}(V)$ be an irreducible representation. If the underlying field is the complex numbers $\mathbb{C}$, then the set of endomorphisms, i.e.*

intertwiners from V to V, only consists of the complex multiples of the identity:

$$\mathrm{End}_{G,\mathbb{C}}(V) = \{c \cdot \mathrm{id}_V \mid c \in \mathbb{C}\} \cong \mathbb{C}.$$

Proof. See Jeevanjee [10]. □

This means that $\tilde{\mathcal{K}}|_{V_J} = c \cdot \mathrm{id}_{V_J}$ for some complex number $c \in \mathbb{C}$. Now if we let $p : V_j \otimes V_l \to V_J$ be the projection corresponding to the direct sum decomposition of $V_j \otimes V_l$, then we obtain

$$\tilde{\mathcal{K}} = \tilde{\mathcal{K}}|_{V_J} \circ p = (c \cdot \mathrm{id}_{V_J}) \circ p = c \cdot p.$$

That is, we have just found out that one complex number, c, is able to completely characterize $\tilde{\mathcal{K}}$ and consequently $\mathcal{K}$! This is *basically* already the Wigner-Eckart Theorem. However, it is useful to find a formulation that describes $\mathcal{K}$ with respect to bases of the different irreducible representations. For this, we define matrix elements of representation operators. Before we come to the definition, we introduce some notation: If $f : V \to V'$ is a linear continuous map between Hilbert spaces, we set

$$\langle y|f|x\rangle \coloneqq \langle y|f(x)\rangle$$

for each $x \in V$ and $y \in V'$. The symmetry in this notation is supposed to remind about the fact that f has an adjoint, see Definition A.2.11, and thus can be applied to y just as well as to x, but we will not make use of this fact.

Definition 4.1.9 (Matrix element)**.** Let T, V_l and V_J be unitary representations with orthonormal bases $\{Y_j^m\} \subseteq T$ (with j possibly also varying), $\{Y_l^n\} \subseteq V_l$ and $\{Y_J^M\} \subseteq V_J$, respectively. Let $\mathcal{K} : T \to \mathrm{Hom}_{\mathbb{K}}(V_l, V_J)$ be a representation operator. Then it's *matrix elements* are given by the scalars

$$\left\langle JM\middle|\mathcal{K}_j^m\middle|ln\right\rangle \coloneqq \left\langle Y_J^M\middle|\mathcal{K}(Y_j^m)\middle|Y_l^n\right\rangle.$$

In the same way, if $f : V_l \to V_J$ is any linear (not necessarily equivariant) map, then its matrix elements are given by the scalars

$$\langle JM|f|ln\rangle \coloneqq \left\langle Y_J^M\middle|f\middle|Y_l^n\right\rangle.$$

Remark 4.1.10. We shortly explain this term. Usually, in linear algebra, one has to do with linear functions $f : V \to V'$ between vector spaces carrying bases $\{v_j\} \subseteq V$ and $\{v_i'\} \subseteq V'$. For each basis element $v_j \in V$ one can then find coefficients $A_{ij} \in \mathbb{K}$ such that

$$f(v_j) = \sum\nolimits_i A_{ij} v_i'.$$

The A_{ij} are called the *matrix elements* of f and characterize f completely. Now if the bases are orthonormal bases as in Definition A.2.9, then the coefficients are given by

$$A_{ij} = \langle v_i'|f(v_j)\rangle = \langle v_i'|f|v_j\rangle.$$

In a similar way we can understand the matrix elements of a representation operator, only that the linear function itself depends on a chosen basis vector of V_j. As for linear functions, the matrix elements of a representation operator completely characterize it.

One last remark: since in this section, V_J appears only once as a direct summand in $V_j \otimes V_l$, we omit the additional "quantum number" s in the notation for the Clebsch-Gordan coefficients. With this preparation, we can formulate and prove the original version of the Wigner-Eckart Theorem. Remember that there is a unique complex number c such that $\tilde{\mathcal{K}}$ is given by $\tilde{\mathcal{K}} = c \cdot p$ for a projection $p : V_j \otimes V_l \to V_J$. We now denote this by $\langle J \| \mathcal{K} \| l \rangle := c$.

Theorem 4.1.11 (Wigner-Eckart Theorem)**.** *The matrix elements of the representation operator $\mathcal{K} : V_j \to \mathrm{Hom}_{\mathbb{C}}(V_l, V_J)$ are given by*

$$\big\langle JM \big| \mathcal{K}_j^m \big| ln \big\rangle = \langle J \| \mathcal{K} \| l \rangle \cdot \big\langle JM \big| jmln \big\rangle,$$

with the $\big\langle JM \big| jmln \big\rangle$ being the Clebsch-Gordan coefficients (which are independent from the representation operator $\mathcal{K}$).

Proof. Let $i : V_J \to V_j \otimes V_l$ be the embedding corresponding to the direct sum decomposition of $V_j \otimes V_l$. It is an adjoint of the projection $p : V_j \otimes V_l \to V_J$ according to the proof of Proposition A.2.15. By what we've argued above, there exists some $c \in \mathbb{C}$ such that:

$$\begin{aligned}
\big\langle JM \big| \mathcal{K}_j^m \big| ln \big\rangle &= \big\langle Y_J^M \big| \mathcal{K}(Y_j^m) \big| Y_l^n \big\rangle \\
&= \big\langle Y_J^M \big| \tilde{\mathcal{K}}(Y_j^m \otimes Y_l^n) \big\rangle \\
&= \big\langle Y_J^M \big| c \cdot p(Y_j^m \otimes Y_l^n) \big\rangle \\
&= c \cdot \big\langle Y_J^M \big| p(Y_j^m \otimes Y_l^n) \big\rangle \\
&= c \cdot \big\langle i(Y_J^M) \big| Y_j^m \otimes Y_l^n \big\rangle \\
&= \langle J \| \mathcal{K} \| l \rangle \cdot \big\langle JM \big| jmln \big\rangle.
\end{aligned}$$

As a short explanation: in the fifth step it was used that i and p are adjoint to each other, and consequently, we move from considering the tensor product in V_J to that one in $V_j \otimes V_l$. In the last step, the definition of the Clebsch-Gordan coefficients was used, and additionally, the notation $\langle J \| \mathcal{K} \| l \rangle := c$ that we mentioned before the theorem. The index s is everywhere missing since V_J appears only once in $V_j \otimes V_l$. This finishes the proof. □

Definition 4.1.12 (Reduced Matrix Element)**.** The unique number $c = \langle J \| \mathcal{K} \| l \rangle \in \mathbb{C}$ in this theorem is called the *reduced matrix element*. To reiterate, it characterizes the representation operator completely.

4.1.3. The Wigner-Eckart Theorem for Steerable Kernels

Now that we have seen the Wigner-Eckart Theorem in a version similar to how it usually appears in physics, it is time to state the version which we will need in this work for applications in deep learning. The treatment is similar to the formulation in Agrawala [26], which presents a generalization of the Wigner-Eckart Theorem to the case that V_J may appear several times as a direct summand in the direct sum decomposition of the tensor product. However, this paper still only considers the Wigner-Eckart Theorem for the case of the complex numbers $\mathbb{C}$ only. If we allow the real numbers as well, we cannot be sure that endomorphisms of irreducible representations are just given by one number. This is a complication we will deal with below by allowing matrix elements of general endomorphisms. Furthermore, we will deal with topological considerations that did not play a role in Agrawala [26]. And lastly, we transport the theorem over into the realm of steerable kernels.

We only consider the case that the input- and output representations are irreducible. Weiler and Cesa [9] then show how to extend the description of the basis kernels to basis kernels of general finite-dimensional input- and output representations. Furthermore, we can assume the input- and output representations to be unitary by Proposition 2.1.20. And finally, we can assume them to even be representatives of the isomorphism classes, since the basis kernels transform predictably under isomorphisms. Thus, assume the input-representation to be $\rho_l : G \to \mathrm{U}(V_l)$ and the output-representation to be $\rho_J : G \to \mathrm{U}(V_J)$. The idea is now that kernel operators $\mathcal{K} : L^2_{\mathbb{K}}(X) \to \mathrm{Hom}_{\mathbb{K}}(V_l, V_J)$ can be described on each direct summand of the domain individually, and that on each of these summands, arguments similar to those for the original Wigner-Eckart Theorem apply.

As a reminder, G is any compact group and X any homogeneous space of G. Furthermore, according to the Peter-Weyl Theorem 2.1.22 the space $L^2_{\mathbb{K}}(X)$ has a dense subset which is a direct sum of irreducible unitary representations:

$$L^2_{\mathbb{K}}(X) = \widehat{\bigoplus_{j \in \hat{G}}} \bigoplus_{i=1}^{m_j} V_{ji}.$$

Each V_{ji} is, as a subrepresentation of $L^2_{\mathbb{K}}(X)$, isomorphic to V_j. V_j is itself not assumed to be embedded in $L^2_{\mathbb{K}}(X)$.

For arbitrary $j \in \hat{G}$, fix once and for all orthonormal bases $\{Y_{ji}^m\} \subseteq V_{ji}$ corresponding to the basis $\{Y_j^m\}$ of V_j[3]. Furthermore, assume that for all $s = 1, \dots, [J(jl)]$, $p_{jis} : V_{ji} \otimes V_l \to V_J$ is a projection which is an adjoint of the linear equivariant isometric embedding $l_{jis} : V_J \to V_{ji} \otimes V_l$. This is assumed to be aligned with the embeddings $V_J \to V_j \otimes V_l$ with respect to the isomorphisms $V_j \cong V_{ji}$ that underlie the correspondence of basis elements $Y_j^m \sim Y_{ji}^m$. What this means is that the Clebsch-Gordan coefficients with respect to all of these embeddings, for all i, are equal:

$$\left\langle l_{jis}(Y_J^M) \middle| Y_{ji}^m \otimes Y_l^n \right\rangle = \langle s, JM | jmln \rangle .$$

Now we state and prove the Wigner-Eckart Theorem, which gives an explicit description of representation operators $\mathcal{K} : L^2_{\mathbb{K}}(X) \to \mathrm{Hom}_{\mathbb{K}}(V_l, V_J)$ in terms of endomorphisms of V_J and then transfers this statement over to a statement about steerable kernels $K : X \to \mathrm{Hom}_{\mathbb{K}}(V_l, V_J)$. Before we state the theorem, we want to shortly

[3] i is like an additional quantum number in physics.

explain what to *expect*: in the derivation of the original Wigner-Eckart Theorem in Section 4.1.2, we saw that a kernel operator could be expressed as $\tilde{\mathcal{K}} : V_j \otimes V_l \to V_J$. This was in turn equal to $\tilde{\mathcal{K}} = c \circ p$ for an endomorphism $c : V_J \to V_J$ and the projection p corresponding to the appearance of V_J in the direct sum decomposition of $V_j \otimes V_l$. This time, however, V_J can be found very often in $L^2_{\mathbb{K}}(X) \otimes V_l$, namely:

1. For each isomorphism class of irreps $j \in \hat{G}$,
2. For each appearance $i = 1, \dots, m_j$ of the irrep V_j in $L^2_{\mathbb{K}}(X)$ and
3. For each appearance $s = 1, \dots, [J(jl)]$ of the irrep V_J in the tensor product representation $V_j \otimes V_l$. $[J(jl)]$ can be zero, which means that j does not contribute.

We therefore expect $\tilde{\mathcal{K}}$ to be a whole sum of compositions of endomorphisms with projections, for each combination of valid j, i and s. Furthermore, the specific structure of $L^2_{\mathbb{K}}(X)$ will be exploited as well by using orthogonal projections from $L^2_{\mathbb{K}}(X)$ to summands V_{ji}. Overall, we hope this sufficiently motivates the theorem:

Theorem 4.1.13 (Wigner-Eckart Theorem for Steerable Kernels). *We state the theorem in three parts:*

1. *(Basis-independent Wigner-Eckart for Kernel Operators) There is an isomorphism of vector spaces*

$$\mathrm{Rep} : \bigoplus_{j \in \hat{G}} \bigoplus_{i=1}^{m_j} \bigoplus_{s=1}^{[J(jl)]} \mathrm{End}_{G,\mathbb{K}}(V_J) \to \mathrm{Hom}_{G,\mathbb{K}}(L^2_{\mathbb{K}}(X), \mathrm{Hom}_{\mathbb{K}}(V_l, V_J))$$

which is given by

$$[\mathrm{Rep}((c_{jis})_{jis})(\varphi)]\,(v_l) := \sum_{j \in \hat{G}} \sum_{i=1}^{m_j} \sum_{s=1}^{[J(jl)]} \sum_{m=1}^{[j]} \langle Y_{ji}^m | \varphi \rangle \cdot c_{jis}\left(p_{jis}(Y_{ji}^m \otimes v_l)\right) \tag{4.1}$$

where $(c_{jis})_{jis}$ is a tuple of endomorphisms, $\varphi : X \to \mathbb{K}$ is any square-integrable function and $v_l \in V_l$ is any element.

2. *(Basis-independent Wigner-Eckart for Steerable Kernels) There is an isomorphism of vector spaces*

$$\mathrm{Ker} : \bigoplus_{j \in \hat{G}} \bigoplus_{i=1}^{m_j} \bigoplus_{s=1}^{[J(jl)]} \mathrm{End}_{G,\mathbb{K}}(V_J) \to \mathrm{Hom}_{G}(X, \mathrm{Hom}_{\mathbb{K}}(V_l, V_J))$$

which is given by

$$[\mathrm{Ker}((c_{jis})_{jis})(x)]\,(v_l) := \sum_{j \in \hat{G}} \sum_{i=1}^{m_j} \sum_{s=1}^{[J(jl)]} \sum_{m=1}^{[j]} \langle Y_{ji}^m | x \rangle \cdot c_{jis}\left(p_{jis}(Y_{ji}^m \otimes v_l)\right)$$

where $(c_{jis})_{jis}$ *is a tuple of endomorphisms,* $x \in X$ *is any point and* $v_l \in V_l$ *is any element. Here,* $\overline{\langle Y_{ji}^m | x \rangle} := \lim_{U \in \mathcal{U}_x} \langle Y_{ji}^m | \delta_U \rangle$*, which is according to Proposition 3.2.10 equal to* $\overline{Y_{ji}^m(x)}$*.*

3. *(Basis-dependent Wigner-Eckart for Steerable Kernels) Let* $K = \mathrm{Ker}((c_{jis})_{jis})$ *be the steerable kernel corresponding to the tuple of endomorphisms* $(c_{jis})_{jis}$ *according to the isomorphism above. Then the matrix elements of* $K(x) \in \mathrm{Hom}_{\mathbb{K}}(V_l, V_J)$ *are explicitly given by*

$$\langle JM|K(x)|ln\rangle = \sum_{j\in\hat{G}}\sum_{i=1}^{m_j}\sum_{s=1}^{[J(jl)]}\sum_{m=1}^{[j]}\sum_{M'=1}^{[J]} \langle JM|c_{jis}|JM'\rangle \cdot \langle s, JM'|jmln\rangle \cdot \langle Y_{ji}^m|x\rangle. \tag{4.2}$$

Remark 4.1.14. Before we come to the proof, we have some remarks to make about this theorem:

1. In line with the usual convention, we call the $\langle JM|c_{jis}|JM'\rangle$ the *generalized reduced matrix elements* of the representation operator $\mathcal{K}$. Different from the situation in physics, these can depend nontrivially on the specific basis indices M and M'. If the space of endomorphisms is 1-dimensional, as is the case when considering representations over $\mathbb{C}$, then each c_{jis} is a diagonal matrix, meaning that it is characterized by only one complex number, for simplicity with the same name c_{jis}. Then one has $\langle JM|c_{jis}|JM'\rangle = \delta_{MM'} \cdot c_{jis}$ and the sum over M' disappears. What this means for the matrix form of basis kernels of steerable CNNs will be discussed in Corollary 4.1.16.

2. The coefficients $\langle s, JM'|jmln\rangle$ are as before the *Clebsch-Gordan coefficients*. Note that the input x of K appears only in $\langle Y_{ji}^m|x\rangle$. Those two parts of the right-hand side of the formula are always the same, independent of the kernel K.

3. The Clebsch-Gordan coefficients are traditionally defined with respect to isometric embeddings $l_{jis} : V_J \to V_j \otimes V_l$ since this makes them less ambiguous. However, we mention that the property of being isometric is no requirement for the construction of Clebsch-Gordan coefficients or the proof of the Wigner-Eckart Theorem, being equivariant and linear is sufficient. This then means that the copies $l_s(Y_J^M)$ do not anymore form an orthonormal basis. We will use this relaxation in the example in Section 6.2, where we do not want to be bothered with obtaining *isometric* embeddings.

4. The names for the isomorphisms in the theorem are meant as follows: Rep is the map that maps a tuple of endomorphisms to a kernel operator, which is a special **rep**resentation operator. Ker maps a tuple of endomorphisms to a steerable **ker**nel. It is *not* meant as a notation for a kernel in the sense of a nullspace in linear algebra.

5. Furthermore, a reader with a background in abstract algebra may wonder why we build the direct sum of spaces of endomorphisms instead of the direct product. The reason is that a posteriori, it turns out that only finitely many j contribute nontrivially, and so the direct sum is equal to the direct product. For a proof of the finiteness, see Remark 4.1.17 below.

6. As a last remark, we want to mention that part 1 of the theorem is not the most general version we could do. We chose to formulate the Wigner-Eckart Theorem for $L^2_{\mathbb{K}}(X)$ specifically since this is the space we use it for. However, an appropriate isomorphism can probably be formulated for any unitary representation instead of $L^2_{\mathbb{K}}(X)$, only that we then need to take care that we replace direct sums by direct products if the index sets on the left side are infinite. Additionally, V_l and V_J could be replaced by arbitrary finite-dimensional representations, and an appropriate adaptation of the theorem would apply. Whether V_l and V_J could also be replaced by infinite-dimensional unitary representations would need to be explored, but an extension to such a case seems possible.

Proof of Theorem 4.1.13. The proof of 1 will be done in Section 4.2 since it requires some work. However, the proofs of 2 and 3 are relatively straightforward once we believe 1 and so we do them here:
From 1 we know that Rep is an isomorphism. Furthermore, from Theorem 3.1.7 we know that

$$(\cdot)|_X : \mathrm{Hom}_{G,\mathbb{K}}(L^2_{\mathbb{K}}(X), \mathrm{Hom}_{\mathbb{K}}(V_l, V_J)) \to \mathrm{Hom}_{G,\mathbb{K}}(X, \mathrm{Hom}_{\mathbb{K}}(V_l, V_J))$$

is an isomorphism as well, and this is given by $\mathcal{K}|_X(x) := \lim_{U\in\mathcal{U}_x} \mathcal{K}(\delta_U)$, where we take the limit over the directed set of open neighborhoods of x. We define the isomorphism Ker now simply as the composition, i.e. $\mathrm{Ker} := (\cdot)|_X \circ \mathrm{Rep}$. This isomorphism is then explicitly given by:

$$\begin{aligned}
[\mathrm{Ker}((c_{jis})_{jis})(x)]\,(v_l) &= [\mathrm{Rep}((c_{jis})_{jis})|_X(x)]\,(v_l) \\
&= \lim_{U\in\mathcal{U}_x} [\mathrm{Rep}((c_{jis})_{jis})(\delta_U)]\,(v_l) \\
&= \lim_{U\in\mathcal{U}_x} \sum_{j\in\hat{G}} \sum_{i=1}^{m_j} \sum_{s=1}^{[J(jl)]} \sum_{m=1}^{[j]} \langle Y_{ji}^m | \delta_U \rangle \cdot c_{jis}\left(p_{jis}(Y_{ji}^m \otimes v_l)\right) \\
&= \sum_{j\in\hat{G}} \sum_{i=1}^{m_j} \sum_{s=1}^{[J(jl)]} \sum_{m=1}^{[j]} \left[\lim_{U\in\mathcal{U}_x} \langle Y_{ji}^m | \delta_U \rangle\right] \cdot c_{jis}\left(p_{jis}(Y_{ji}^m \otimes v_l)\right) \\
&= \sum_{j\in\hat{G}} \sum_{i=1}^{m_j} \sum_{s=1}^{[J(jl)]} \sum_{m=1}^{[j]} \langle Y_{ji}^m | x \rangle \cdot c_{jis}\left(p_{jis}(Y_{ji}^m \otimes v_l)\right).
\end{aligned}$$

This already proves 2. Now, in the following computation, we will use that $c_{jis} \circ p_{jis} = c_{jis} \circ \mathrm{id}_{V_J} \circ p_{jis}$ and that, inspired by notation in physics, we can write the identity on

V_J as $\mathrm{id}_{V_J} = \sum_{M'=1}^{[J]} \left|Y_J^{M'}\right\rangle \cdot \left\langle Y_J^{M'}\right|$. For 3, we then compute

$$
\begin{aligned}
&\langle JM|K(x)|ln\rangle \\
&= \left\langle Y_J^M \middle| K(x) \middle| Y_l^n \right\rangle \\
&= \left\langle Y_J^M \middle| \left[\mathrm{Ker}((c_{jis})_{jis})(x)\right] (Y_l^n) \right\rangle \\
&= \sum_{j\in\hat{G}} \sum_{i=1}^{m_j} \sum_{s=1}^{[J(jl)]} \sum_{m=1}^{[j]} \left\langle Y_{ji}^m \middle| x \right\rangle \cdot \left\langle Y_J^M \middle| c_{jis} \circ p_{jis} \middle| Y_{ji}^m \otimes Y_l^n \right\rangle \\
&= \sum_{j\in\hat{G}} \sum_{i=1}^{m_j} \sum_{s=1}^{[J(jl)]} \sum_{m=1}^{[j]} \sum_{M'=1}^{[J]} \left\langle Y_{ji}^m \middle| x \right\rangle \cdot \left\langle Y_J^M \middle| c_{jis} \middle| Y_J^{M'} \right\rangle \cdot \left\langle Y_J^{M'} \middle| p_{jis} \middle| Y_{ji}^m \otimes Y_l^n \right\rangle \\
&= \sum_{j\in\hat{G}} \sum_{i=1}^{m_j} \sum_{s=1}^{[J(jl)]} \sum_{m=1}^{[j]} \sum_{M'=1}^{[J]} \left\langle JM \middle| c_{jis} \middle| JM' \right\rangle \cdot \left\langle s, JM' \middle| jmln \right\rangle \cdot \left\langle Y_{ji}^m \middle| x \right\rangle.
\end{aligned}
$$

In the last step, we used the Clebsch-Gordan coefficients, see Definition 4.1.6 and, as mentioned before, that p_{jis} is adjoint to the embedding $l_{jis} : V_J \to V_{ji} \otimes V_l$. □

4.1.4. General Steerable Kernel Bases

Now that we have a Wigner-Eckart Theorem for steerable kernels, which gives a one-to-one correspondence between steerable kernels and tuples of endomorphisms, we can finally describe what a *basis* of the space of steerable kernels looks like. For this, additionally to the notation in the last section, we assume that $\{c_r \mid r \in R\}$ is a basis of $\mathrm{End}_{G,\mathbb{K}}(V_J)$.

Theorem 4.1.15 (Steerable kernel bases). *A basis of the space of steerable kernels $\mathrm{Hom}_G(X, \mathrm{Hom}_{\mathbb{K}}(V_l, V_J))$ is given by*

$$\{K_{jisr} : X \to \mathrm{Hom}_{\mathbb{K}}(V_l, V_J) \mid j \in \hat{G}, i \in \{1, \dots, m_j\}, s \in \{1, \dots, [J(jl)]\}, r \in R\},$$

where the basis kernels K_{jisr} have matrix elements

$$\langle JM|K_{jisr}(x)|ln\rangle = \sum_{m=1}^{[j]} \sum_{M'=1}^{[J]} \left\langle JM \middle| c_r \middle| JM' \right\rangle \cdot \left\langle s, JM' \middle| jmln \right\rangle \cdot \left\langle Y_{ji}^m \middle| x \right\rangle. \tag{4.3}$$

Now, for each $M' \in \{1, \dots, [J]\}$, let $\mathrm{CG}_{J(jl)s}^{M'}$ be the $[j] \times [l]$-matrix of Clebsch-Gordan coefficients $\langle s, JM'|jmln\rangle$, with only m and n varying. Furthermore, let $\langle Y_{ji}|x\rangle$ be the row vector with entries $\left\langle Y_{ji}^m \middle| x\right\rangle$ for $m = 1, \dots, [j]$. In matrix-notation with respect to the bases $\{Y_J^M\} \subseteq V_J$ and $\{Y_l^n\} \subseteq V_l$, we can then express the basis kernel $K_{jisr}(x) : V_l \to V_J$ as follows:

$$K_{jisr}(x) = c_r \cdot \begin{pmatrix} \langle Y_{ji}|x\rangle \cdot \mathrm{CG}_{J(jl)s}^{1} \\ \vdots \\ \langle Y_{ji}|x\rangle \cdot \mathrm{CG}_{J(jl)s}^{[J]} \end{pmatrix}. \tag{4.4}$$

In this formula, all "dots" mean conventional matrix multiplication and c_r is by abuse of notation the matrix of the endomorphism c_r.

Proof. For the first statement, note that a basis for $\bigoplus_{j\in\hat{G}}\bigoplus_{i=1}^{m_j}\bigoplus_{s=1}^{[J(jl)]}\mathrm{End}_{G,\mathbb{K}}(V_J)$ is given by all the tuples $t_{jisr} := (0,\dots,c_r,\dots,0)$ that have c_r at position jis, for all combinations of j, i, s and r. Thus, from the isomorphism Ker in the second part of Theorem 4.1.13 we obtain that all $K_{jisr} := \mathrm{Ker}(t_{jisr})$ together form a basis for the space of steerable kernels $\mathrm{Hom}_G(X, \mathrm{Hom}_{\mathbb{K}}(V_l, V_J))$. When applying the basis-dependent form in part 3 of that theorem to K_{jisr}, the first three sums in Equation 4.2 just disappear since t_{jisr} is zero almost everywhere. Furthermore, c_{jis} is replaced by the basis endomorphism c_r. We obtain the claimed result.
For the final statement on the matrix representation, note that

$$\begin{aligned}\langle JM|K_{jisr}(x)|ln\rangle &= \sum_{m=1}^{[j]}\sum_{M'=1}^{[J]}\big\langle JM\big|c_r\big|JM'\big\rangle\cdot\big\langle s, JM'\big|jmln\big\rangle\cdot\big\langle Y_{ji}^m\big|x\big\rangle\\ &= \sum_{M'=1}^{[J]}\big\langle JM\big|c_r\big|JM'\big\rangle\sum_{m=1}^{[j]}\big\langle Y_{ji}^m\big|x\big\rangle\cdot\big\langle s, JM'\big|jmln\big\rangle\\ &= c_r^M\cdot\left(\sum_{m=1}^{[j]}\langle Y_{ji}^m|x\rangle\cdot\langle s, JM'|jmln\rangle\right)_{M'=1}^{[J]}\\ &= c_r^M\cdot\left(\big\langle Y_{ji}\big|x\big\rangle\cdot\mathrm{CG}_{J(jl)s}^{M'-n}\right)_{M'=1}^{[J]}.\end{aligned}$$

Here, c_r^M is the M'th row of the matrix c_r. The result follows by dropping the indices M and n. □

The next corollary means that endomorphisms can be ignored if the space of endomorphisms is 1-dimensional, which is in particular the case if $\mathbb{K}=\mathbb{C}$.

Corollary 4.1.16. *Assume that* $\dim(\mathrm{End}_{G,\mathbb{K}}(V_J)) = 1$*. Then a basis of steerable kernels* $K : X \to \mathrm{Hom}_{\mathbb{K}}(V_l, V_J)$ *is given by all* K_{jis} *with matrices*

$$K_{jis}(x) = \begin{pmatrix}\langle Y_{ji}|x\rangle\cdot\mathrm{CG}_{J(jl)s}^{1}\\ \vdots\\ \langle Y_{ji}|x\rangle\cdot\mathrm{CG}_{J(jl)s}^{[J]}\end{pmatrix}. \tag{4.5}$$

In particular, this is the case if $\mathbb{K}=\mathbb{C}$*.*

Proof. In this case, a basis for the space of endomorphisms is given by the single endomorphism $c = \mathrm{id}_{V_J}$. Postcomposition with the identity does not change the matrix, and so the result follows.
For $\mathbb{K}=\mathbb{C}$ we have $\dim(\mathrm{End}_{G,\mathbb{C}}(V_J)) = 1$ by Schur's Lemma 4.1.8, and thus the result follows. □

We end with two remarks regarding the parameterization of steerable CNNs. The first remark considers the case of steerable CNNs of the form $K : X \to \mathrm{Hom}_{\mathbb{K}}(V_l, V_J)$ on a homogeneous space X. The second remark connects this back to the case that X is an orbit embedded in $\mathbb{R}^n$.

Remark 4.1.17 (Parameterization in the abstract). First of all, we want to understand that there are only finitely many basis kernels K_{jisr}. To this end, note that the index sets for i, s, and r are necessarily finite for all j, and thus we need to understand the finite range of j. A priori, j can run over the whole set $\hat{G}$, which can be infinite. But, as we argue now, for only finitely many $j \in \hat{G}$ we can have V_J in a direct sum decomposition of $V_j \otimes V_l$, which rescues the finiteness:
Namely, V_J is in the direct sum decomposition of $V_j \otimes V_l$ if and only if the vector space $\mathrm{Hom}_{G,\mathbb{K}}(V_j \otimes V_l, V_J)$ is nonzero by Schur's Lemma 2.2.6. By the hom-tensor adjunction that we will show in Proposition 4.2.4 in more generality, this is the case if an only if $\mathrm{Hom}_{G,\mathbb{K}}(V_j, \mathrm{Hom}_{\mathbb{K}}(V_l, V_J))$ is nonzero. And finally, this is the case if and only if V_j is in a direct sum decomposition of the representation $\mathrm{Hom}_{\mathbb{K}}(V_l, V_J)$, again by Schur's Lemma. Now, since $\mathrm{Hom}_{\mathbb{K}}(V_l, V_J)$ is finite-dimensional, this can only be the case for finitely many j, and so we are done[4].
Overall, this means the following: To parameterize an equivariant neural network, one needs arbitrary parameters $w_{jisr} \in \mathbb{K}$ for all combinations of $j \in \hat{G}$, $i \in \{1, \ldots, m_j\}$, $s \in \{1, \ldots, [J(jl)]\}$ and $r \in R$. A general steerable Kernel $K : X \to \mathrm{Hom}_{\mathbb{K}}(V_l, V_J)$ then takes the form

$$K = \sum_{j \in \hat{G}} \sum_{i=1}^{m_j} \sum_{s=1}^{[J(jl)]} \sum_{r \in R} w_{jisr} K_{jisr},$$

with the basis kernels K_{jisr} as in Theorem 4.1.15.

Remark 4.1.18 (Parameterization in practice). Remember that our original motivation for the use of homogeneous spaces in Section 3.1.1 was that $\mathbb{R}^n$ splits as a disjoint union of homogeneous spaces, on which the kernel constraint acts completely separately. For simplicity, we assume that the compact group acting on $\mathbb{R}^n$ is either $G = \mathrm{SO}(n)$ or $G = \mathrm{O}(n)$, but the general ideas hold also for the finite transformation groups in $\mathbb{R}^n$ — the only difference is that in these finite cases, the set of representatives of orbits becomes larger.
Thus, $\mathbb{R}^n$ splits into orbits $\mathbb{R}^n = \bigsqcup_{r \geq 0} S^{n-1}(r)$, where $S^{n-1}(r)$ is the sphere of radius r (with $S(0) = \{0\}$ being a single point).
We'll discuss the orbit $X_0 = \{0\}$, the origin, separately below. But note that all other orbits are necessarily homeomorphic to each other and thus can be treated on equal footing. Therefore, let S^{n-1} be the standard sphere with radius 1 and $K_{jisr} : S^{n-1} \to \mathrm{Hom}_{\mathbb{K}}(V_l, V_J)$ be basis kernels for this choice. Then for a general steerable kernel $K : \mathbb{R}^n \to \mathrm{Hom}_{\mathbb{K}}(V_l, V_J)$ there are *arbitrary* functions $w_{jisr} : \mathbb{R}_{>0} \to \mathbb{K}$ such that, for all $x \in \mathbb{R}^n \setminus \{0\}$, we have:

$$K(x) = \sum_{j \in \hat{G}} \sum_{i=1}^{m_j} \sum_{s=1}^{[J(jl)]} \sum_{r \in R} w_{jisr}(\|x\|) \cdot K_{jisr}\left(\frac{x}{\|x\|}\right).$$

For $x = 0$, we might use our heavy theory to solve the kernel constraint, but it is more illuminating to do it from scratch since this case is so simple: we have $K(0) : V_l \to V_J$,

[4] Of course, for this argument, we need the uniqueness of direct sum decompositions. But this follows if we assume the Hom-representation to be unitary, which works by Proposition 2.1.20 and then using the Krull-Remak-Schmidt Theorem, Proposition 2.2.16.

and the kernel constraint takes the form

$$K(0) = K(g \cdot 0) = \rho_J(g) \circ K(0) \circ \rho_l(g)^{-1}$$

for all $g \in G$, which is equivalent to $K(0) \circ \rho_l(g) = \rho_J(g) \circ K(0)$ for all $g \in G$. This just means that $K(0) : V_l \to V_J$ is an intertwiner, and by Schur's Lemma 2.2.6 it is either 0 if $l \neq J$ or an arbitrary endomorphism $V_J \to V_J$ if $l = J$. Thus, assuming $l = J$ and choosing basis-endomorphisms $c_r : V_J \to V_J$, there are coefficients $w_r \in \mathbb{K}$ such that

$$K(0) = \sum\nolimits_{r \in R} w_r \cdot c_r.$$

The reader may find it interesting to check that this solution is precisely what is also predicted by our theory using that $L^2_{\mathbb{K}}(\{0\}) \cong \mathbb{K}$ is just isomorphic to the trivial representation of G.

All in all, we now know what the most general steerable kernels look like. In practice, one needs to choose the functions $w_{jisr} : \mathbb{R}_{>0} \to \mathbb{K}$. For representations over the real numbers, i.e. with $\mathbb{K} = \mathbb{R}$, one choice is to only consider finitely many radii and Gaussian radial profiles around them. Then instead of learning the whole function w_{jisr}, one learns finitely many real parameters that choose "how activated" a basis kernel K_{jisr} is for a certain radius. This is for example the route taken in Weiler et al. [8], Weiler and Cesa [9], Weiler et al. [24]. If one deals with complex representations, one usually goes the same route, only that the parameters that choose how "activated" the basis kernels are will then be *complex numbers*. One can either parameterize them as $a + ib$ with a real part a and a complex part b. This intuitively means that a activates the standard version of the kernel K_{jisr}, whereas b activates the kernel iK_{jisr}, which can be imagined as a version of the kernel turned by $90°$. One other possibility is to parameterize a complex number as $\alpha \cdot e^{i\beta}$ with a scaling factor $\alpha > 0$ and a phase shift β. This is the route chosen in Worrall et al. [4].

In Chapter 6 we will look at examples of determining the basis kernels K_{jisr}, which will hopefully further illuminate the theorem. In the next section, we go back to the theory and prove the remaining parts of the Wigner-Eckart Theorem.

4.2. Proof of the Wigner-Eckart Theorem for Kernel Operators

In this section, we prove the first part of Theorem 4.1.13, the Wigner-Eckart Theorem for Kernel Operators, since we have skipped this in the last section. It is not necessary to read this section and the reader may wish to directly go to the chapter on related work 5 or the chapter on examples 6. We will make frequent use of topological concepts from Appendix A.1 in this section.

The strategy is the following: in Section 4.2.1, we show that

$$\mathrm{Hom}_{G,\mathbb{K}}(L^2_{\mathbb{K}}(X), \mathrm{Hom}_{\mathbb{K}}(V_l, V_J)) \cong \mathrm{Hom}_{G,\mathbb{K}}\left(\bigoplus_{j \in \hat{G}} \bigoplus_{i=1}^{m_j} V_{ji}, \mathrm{Hom}_{\mathbb{K}}(V_l, V_J) \right),$$

which basically means that we can ignore the "topological closure" of the direct sum which is dense in $L^2_{\mathbb{K}}(X)$. This works, intuitively, since kernel operators are continuous, and so they are completely determined by what they do on a dense subset. Then, in section 4.2.2, we show that

$$\mathrm{Hom}_{G,\mathbb{K}}\left(\bigoplus_{j\in\hat{G}}\bigoplus_{i=1}^{m_j} V_{ji}, \mathrm{Hom}_{\mathbb{K}}(V_l, V_J)\right) \cong \mathrm{Hom}_{G,\mathbb{K}}\left(\bigoplus_{j\in\hat{G}}\bigoplus_{i=1}^{m_j} V_{ji}\otimes V_l, V_J\right),$$

which is the main step that we need in order to be able to make use of the Clebsch-Gordan coefficients, namely when we decompose the tensor product. Finally, in Section 4.2.3, we finish the proof of Theorem 4.1.13.

4.2.1. Reduction to a Dense Subspace of $L^2_{\mathbb{K}}(X)$

In this section, we reduce the statement to representation operators on $\bigoplus_{j\in\hat{G}}\bigoplus_{i=1}^{m_j} V_{ji}$. For simplicity, we write the double direct sum from now on as $\bigoplus_{ji}$.
Furthermore, remember that V_l and V_J are finite-dimensional, and thus $\mathrm{Hom}_{\mathbb{K}}(V_l, V_J)$ can be identified with matrices in $\mathbb{K}^{[J]\times[l]}$. This space is a Euclidean space and thus has a scalar product and consequently also a norm, see Appendix A.1. Consequently, each kernel operator is a continuous map between *normed* vector spaces, which we'll use in the following.
A short terminological note: kernel operators are just representation operators on $L^2_{\mathbb{K}}(X)$ and only have their name due to the relation to steerable kernels. Thus, the terminological difference to representation operators in the following reduction result has no further meaning:

Lemma 4.2.1. *The restriction map*

$$\mathrm{Hom}_{G,\mathbb{K}}(L^2_{\mathbb{K}}(X), \mathrm{Hom}_{\mathbb{K}}(V_l, V_J)) \to \mathrm{Hom}_{G,\mathbb{K}}\left(\bigoplus\nolimits_{ji} V_{ji}, \mathrm{Hom}_{\mathbb{K}}(V_l, V_J)\right)$$

given by $\mathcal{K} \mapsto \mathcal{K}|_{\bigoplus_{ji} V_{ji}}$, between kernel operators on the left and representation operators on the right is an isomorphism.

Proof. First of all, the kernel operators on the left are actually uniformly continuous by Proposition A.1.18. Thus, by Lemma A.1.22, the restriction map is an injection into *uniformly continuous* representation operators on $\bigoplus_{ji} V_{ji}$. The set of all these maps is equal to the set of all representation operators by Proposition A.1.18 again.
Thus, in order to be finished, we only need to see that the unique extension of a representation operator $\mathcal{K} : \bigoplus_{ji} V_{ji} \to \mathrm{Hom}_{\mathbb{K}}(V_l, V_J)$ to a continuous function $\overline{\mathcal{K}} : L^2_{\mathbb{K}}(X) \to \mathrm{Hom}_K(V_l, V_J)$ is a kernel operator, which means it is linear and equivariant.
For linearity, let $a \in \mathbb{K}$ and $f \in L^2_{\mathbb{K}}(X)$. Let $(f_k)_k$ be a sequence in $\bigoplus_{ji} V_{ji}$ that converges to f. Using the continuity of $\overline{\mathcal{K}}$ and the linearity of $\mathcal{K}$ we obtain:

$$\overline{\mathcal{K}}(a\cdot f) = \overline{\mathcal{K}}\big(\lim_{k\to\infty}(a\cdot f_k)\big)$$

$$\begin{aligned}
&= \lim_{k\to\infty} \overline{\mathcal{K}}(a \cdot f_k) \\
&= \lim_{k\to\infty} \mathcal{K}(a \cdot f_k) \\
&= \lim_{k\to\infty} a \cdot \mathcal{K}(f_k) \\
&= a \cdot \lim_{k\to\infty} \overline{\mathcal{K}}(f_k) \\
&= a \cdot \overline{\mathcal{K}}\big(\lim_{k\to\infty} f_k\big) \\
&= a \cdot \overline{\mathcal{K}}(f).
\end{aligned}$$

Linearity with respect to addition can be shown similarly. For the equivariance we can essentially argue in the same way, only that we additionally need to use the continuity of the representations $\lambda : G \to \mathrm{U}(L^2_{\mathbb{K}}(X))$ and $\rho_{\mathrm{Hom}} : G \to \mathrm{Aut}_{\mathbb{K}}(\mathrm{Hom}_{\mathbb{K}}(V_l, V_J))$. □

4.2.2. The Hom-Tensor Adjunction

Lemma 4.2.2. *Let $\mathcal{K} : \bigoplus_{li} V_{li} \to V$ be linear and equivariant, where V is an irrep. Then $\mathcal{K}$ is continuous.*

Proof. By Schur's Lemma 4.1.8[5], we know that $\mathcal{K}$ factors through the irreducible representations that are isomorphic to V. That is, let V_j be that irrep and $p_{ji} : \bigoplus_{li} V_{li} \to V_{ji}$ be the canonical projections. Then there are intertwiners $c_i : V_{ji} \to V$ such that $\mathcal{K} = \sum_i c_i \circ p_{ji}$. Each c_i is continuous since it is a linear function between *finite-dimensional* normed vector spaces. Since also summation on normed vector spaces is continuous, we only need to show that the projections p_{ji} are continuous.
This follows from the following fact on how the norm on $\bigoplus_{li} V_{li}$ is composed from the norms on each V_{li}: For an element $f = \sum_{li} f_{li} \in \bigoplus_{li} V_{li}$ with $f_{li} \in V_{li}$, we have:

$$\|f\|^2 = \sum_{li} \|f_{li}\|^2.$$

The reason for this is that the V_{li} are perpendicular to each other. Consequently, if $(f^k)_k$ with $f^k \in \bigoplus_{li} V_{li}$ converges to 0, then also $(p_{ji}(f^k))_k = (f^k_{ji})_k$ converges to 0, which shows the continuity of p_{ji} in 0 and thus general continuity by Proposition A.1.18. □

Remark 4.2.3. Note the curious fact that we cannot get rid of the equivariance condition in the preceding Lemma. I.e., if we have a linear function $\mathcal{K} : \bigoplus_l V_l \to V$, then we cannot deduce that $\mathcal{K}$ is continuous. We omit the index i for simplicity. If equivariance is no requirement, then we only deal with vector spaces, which are in

[5] Schur's Lemma applies since it is a statement about irreducible representations which are necessarily finite-dimensional. This means that the continuity condition in the definition of intertwiners is vacuous and thus we don't need to worry about $\mathcal{K}$ not being continuous *a priori*.

general isomorphic to spaces of (maybe infinite) tuples of elements in $\mathbb{K}$. Thus, let the function $\mathcal{K} : \bigoplus_{l \in \mathbb{N}} \mathbb{K} \to \mathbb{K}$ given by

$$(a_l)_l \mapsto \sum\nolimits_l l \cdot a_l.$$

This is linear but not continuous in 0. The latter can be seen by considering the sequence $(a^k)_k$ with $a^k = (0, \dots, 0, \frac{1}{k}, 0, \dots)$ that has value $\frac{1}{k}$ on position k and otherwise only zeros. This sequence converges to the 0-sequence in norm. However, we have $\mathcal{K}(a^k) = 1$ for all k, thus the images do not converge to $0 = \mathcal{K}(0)$. □

From the preceding lemma, we are able to obtain the following alternative description of representation operators:

Proposition 4.2.4 (Hom-tensor Adjunction)**.** *The map*

$$\tilde{(\cdot)} : \mathrm{Hom}_{G,\mathbb{K}}\left(\bigoplus\nolimits_{ji} V_{ji}, \mathrm{Hom}_{\mathbb{K}}(V_l, V_J)\right) \to \mathrm{Hom}_{G,\mathbb{K}}\left(\left(\bigoplus\nolimits_{ji} V_{ji}\right) \otimes V_l, V_J\right)$$

given by

$$\tilde{\mathcal{K}}(v_j \otimes v_l) \coloneqq [\mathcal{K}(v_j)]\,(v_l)$$

is an isomorphism.

Proof. For continuity, note the following: by straightforward extensions of Lemma 4.2.2, all linear and equivariant maps $\bigoplus_{ji} V_{ji} \to \mathrm{Hom}_{\mathbb{K}}(V_l, V_J)$ and $\left(\bigoplus_{ji} V_{ji}\right) \otimes V_l \to V_J$ are necessarily continuous, and thus we can ignore continuity altogether. The rest of the proof can be done as in Agrawala [26]. For illustrating the most important part, we show that $\tilde{\mathcal{K}}$ is actually equivariant:

$$\begin{aligned}
\tilde{\mathcal{K}}\big(\,[(\rho_j \otimes \rho_l)(g)]\,(v_j \otimes v_l)\big) &= \tilde{\mathcal{K}}\big(\,[\rho_j(g)]\,(v_j) \otimes [\rho_l(g)]\,(v_l)\big) \\
&= [\mathcal{K}(\rho_j(g)(v_j))]\,(\rho_l(g)(v_l)) \\
&= \big[\rho_{\mathrm{Hom}}(g)(\mathcal{K}(v_j))\big]\,(\rho_l(g)(v_l)) \\
&= \big(\rho_J(g) \circ \mathcal{K}(v_j) \circ \rho_l(g)^{-1}\big)\,(\rho_l(g)(v_l)) \\
&= \rho_J(g)(\mathcal{K}(v_j)(v_l)) \\
&= \rho_J(g)(\tilde{\mathcal{K}}(v_j \otimes v_l)).
\end{aligned}$$

□

Remark 4.2.5. Some readers may wonder why this is called an *adjunction*. With removing some of the notation in the Proposition, one has

$$\mathrm{Hom}_{G,\mathbb{K}}(T, \mathrm{Hom}_{\mathbb{K}}(U, V)) \cong \mathrm{Hom}_{G,\mathbb{K}}(T \otimes U, V).$$

Now, for notational clarity, set $F \coloneqq \mathrm{Hom}_{\mathbb{K}}(U, \cdot)$ and $H \coloneqq (\cdot) \otimes U$ and remove the subscripts. Then the formula can be written as

$$\mathrm{Hom}(T, F(V)) \cong \mathrm{Hom}(H(T), V).$$

With replacing the notation if the Hom-spaces with a scalar product, and the isomorphism sign with equality, this reads as follows:

$$\langle T | F(V) \rangle = \langle H(T) | V \rangle \, .$$

Similar to adjoints in Hilbert spaces, we can then view F and H as adjoint to each other. In categorical terms, they are a pair of adjoint functors, see Lane et al. [27].

4.2.3. Proof of Theorem 4.1.13

After the work done in the prior sections, we are ready to complete the proof of Theorem 4.1.13!

Proof of Theorem 4.1.13. Only the first part of that theorem still needs to be proven. We have the following string of isomorphisms, which we will explain below:

$$\begin{aligned}
\mathrm{Hom}_{G,\mathbb{K}}(L^2_{\mathbb{K}}(X), \mathrm{Hom}_{\mathbb{K}}(V_l, V_J)) &\cong \mathrm{Hom}_{G,\mathbb{K}}\left(\bigoplus\nolimits_{ji} V_{ji}, \mathrm{Hom}_{\mathbb{K}}(V_l, V_J)\right) \\
&\cong \mathrm{Hom}_{G,\mathbb{K}}\left(\left(\bigoplus\nolimits_{ji} V_{ji}\right) \otimes V_l, V_J\right) \\
&\cong \mathrm{Hom}_{G,\mathbb{K}}\left(\bigoplus\nolimits_{ji} (V_{ji} \otimes V_l), V_J\right) \\
&\cong \bigoplus_{ji} \mathrm{Hom}_{G,\mathbb{K}}(V_{ji} \otimes V_l, V_J) \\
&\cong \bigoplus_{ji} \bigoplus_{s=1}^{[J(jl)]} \mathrm{Hom}_{G,\mathbb{K}}(V_J, V_J) \\
&= \bigoplus_{j \in \hat{G}} \bigoplus_{i=1}^{m_j} \bigoplus_{s=1}^{[J(jl)]} \mathrm{End}_{G,\mathbb{K}}(V_J).
\end{aligned}$$

The steps are justified as follows:

1. For the first step, use Lemma 4.2.1.
2. For the second step, use Proposition 4.2.4.
3. For the third step, use that there is a natural isomorphism $\left(\bigoplus_{ji} V_{ji}\right) \otimes V_l \cong \bigoplus_{ji}(V_{ji} \otimes V_l)$.
4. For the fourth step, use that linear equivariant maps can be described on each direct summand individually (and that we do not need to worry about continuity due to Lemma 4.2.2).
5. For the fifth step, precompose with the linear equivariant isometric embeddings $l_{jis} : V_J \to V_{ji} \otimes V_l$ and use, again, that linear equivariant maps can be described on each direct summand individually. Furthermore, use Schur's Lemma 2.2.6 in order to see that the other summands disappear.

6. The last step is just a reformulation.

Now, we call the string of isomorphisms from right to left

$$\mathrm{Rep} : \bigoplus_{j\in\hat{G}} \bigoplus_{i=1}^{m_j} \bigoplus_{s=1}^{[J(jl)]} \mathrm{End}_{G,\mathbb{K}}(V_J) \to \mathrm{Hom}_{G,\mathbb{K}}(L^2_{\mathbb{K}}(X), \mathrm{Hom}_{\mathbb{K}}(V_l, V_J))$$

and are only left with understanding that it is actually given by Equation 4.1. For this, we take a tuple $(c_{jis})_{jis}$ of endomorphisms and explicitly trace back "where it comes from". As in Lemma 4.2.2, let $p_{ji} : \bigoplus_{j'i'} V_{j'i'} \to V_{ji}$ be the canonical projection, which is by Proposition A.2.15 explicitly given by $p_{ji}(\varphi) = \sum_{m=1}^{[j]} \langle Y_{ji}^m | \varphi \rangle Y_{ji}^m$. Furthermore, let $p_{jis} : V_{ji} \otimes V_l \to V_J$ be the projections corresponding to the embeddings l_{jis}. Then from bottom to top, $(c_{jis})_{jis}$ gets transformed as follows:

$$\begin{aligned}
(c_{jis})_{jis} &\mapsto \left(\sum_{s=1}^{[J(jl)]} c_{jis} \circ p_{jis} \right)_{ji} \\
&\mapsto \sum_{j\in\hat{G}} \sum_{i=1}^{m_j} \sum_{s=1}^{[J(jl)]} c_{jis} \circ p_{jis} \circ (p_{ji} \otimes \mathrm{id}_{V_l}) \\
&\mapsto \mathrm{Rep}((c_{jis})_{jis})
\end{aligned}$$

In the very last step, the hom-tensor adjunction Proposition 4.2.4 is used, but in the other direction. As an illustration, the composition of functions over which we sum can be shown in the following commutative diagram:

$$\begin{tikzcd}
\bigoplus_{i'j'} V_{j'i'} \otimes V_l \arrow[r, "p_{jis}\otimes \mathrm{id}_{V_l}"] \arrow[rrr, bend right, "c_{jis}\circ p_{jis}\circ(p_{ji}\otimes \mathrm{id}_{V_l})"'] & V_{ji} \otimes V_l \arrow[r, "p_{ji}"] & V_J \arrow[r, "c_{jis}"] & V_J
\end{tikzcd}$$

We obtain:

$$\begin{aligned}
[\mathrm{Rep}((c_{jis})_{jis})(\varphi)]\,(v_l) &= \sum_{j\in\hat{G}} \sum_{i=1}^{m_j} \sum_{s=1}^{[J(jl)]} \left[c_{jis} \circ p_{jis} \circ (p_{ji} \otimes \mathrm{id}_{V_l})\right](\varphi \otimes v_l) \\
&= \sum_{j\in\hat{G}} \sum_{i=1}^{m_j} \sum_{s=1}^{[J(jl)]} (c_{jis} \circ p_{jis})(p_{ji}(\varphi) \otimes v_l) \\
&= \sum_{j\in\hat{G}} \sum_{i=1}^{m_j} \sum_{s=1}^{[J(jl)]} (c_{jis} \circ p_{jis}) \left(\sum_{m=1}^{[j]} \langle Y_{ji}^m | \varphi \rangle Y_{ji}^m \otimes v_l \right) \\
&= \sum_{j\in\hat{G}} \sum_{i=1}^{m_j} \sum_{s=1}^{[J(jl)]} \sum_{m=1}^{[j]} \langle Y_{ji}^m | \varphi \rangle \cdot c_{jis} \left(p_{jis} \left(Y_{ji}^m \otimes v_l \right) \right).
\end{aligned}$$

That, finally, finishes the proof. □

5. Related Work

Now that we are finally finished with developing our theory, we are ready to compare what we did with much of the prior work that has emerged in the preceding years. After that, we will discuss many example applications of our theory in Chapter 6.
We structure this chapter as follows: In Section 5.1 we look at prior work on $\mathrm{E}(2)$-equivariant CNNs, since their solution of the kernel constraint ultimately shows strong similarities to our work. Afterward, in Section 5.2, we look at other work that falls within the framework of steerable CNNs on $\mathbb{R}^2$ or $\mathbb{R}^3$. This also includes work on finite transformation groups, namely group convolutional CNNs, and harmonic networks, even though this research has been framed without the use of the term "steerable".
Afterward, in Section 5.3, we briefly look at gauge equivariant CNNs, which are generalizations of steerable CNNs that fulfill the same kernel constraint and are thus covered by our work. They are physics-inspired, as we are, and so in Section 5.4 we discuss more work that is inspired by physics and representation theory. This includes the use of Clebsch-Gordan coefficients, but also higher-dimensional symmetry groups like the Lorentz group appearing in special relativity.
Finally, in Section 5.5, we look at prior purely theoretical work. This prior work differs from us, in that it does not *solve* the kernel constraints that it describes.

5.1. General $\mathrm{E}(2)$-Equivariant Steerable CNNs

Similar to our work is Weiler and Cesa [9]. In this paper, the authors look at planar CNNs that can, for example, process image data. Instead of only looking at one transformation group, they actually look at all topologically closed subgroups of $\mathrm{O}(2)$ and derive and implement steerable CNNs for all of them. Those subgroups are $\mathrm{O}(2)$, $\mathrm{SO}(2)$, and C_N and D_N for natural numbers $N \in \mathbb{N}$.
It pays off to look at their strategy for solving the kernel constraint and to compare it with our method. For example, in their derivation for $\mathrm{SO}(2)$ and for irreducible, 2-dimensional input- and output representations, they look at kernels $K : S^1 \to \mathbb{R}^{2\times 2}$, where S^1 is homeomorphic to the orbit under the action of $\mathrm{SO}(2)$ of any point $0 \neq x \in \mathbb{R}^2$. Their ansatz is to expand each matrix element of a steerable kernel K, namely $K_{jk} : S^1 \to \mathbb{R}$ for $j, k \in \{1, 2\}$, in a Fourier basis:

$$K_{jk} = \sum_{l=0}^{\infty} A_{jk,l} \cos_l + B_{jk,l} \sin_l .$$

Here, we mean $\cos_l(\phi) = \sqrt{2}\cos(l \cdot \phi)$ and $\sin_l(\phi) = \sqrt{2}\sin(l \cdot \phi)$, even though in their work the normalization by $\sqrt{2}$ was not present. For $l = 0$, $\sin_l = 0$, so it can also be omitted. What they then essentially did was to use the kernel constraint in order to determine the Fourier coefficients $A_{jk,l}$ and $B_{jk,l}$. This then leads to the general form of a steerable kernel. Remember from Fourier theory that the Fourier coefficients can be determined as[1]

$$A_{jk,l} = \int_{S^1} \cos_l(x) K_{jk}(x) dx, \quad B_{jk,l} = \int_{S^1} \sin_l(x) K_{jk}(x) dx.$$

Now we wonder: What would their strategy have looked like, would they have used our method? In this case, in oder to determine the general form of K, one looks at the corresponding kernel operator $\hat{K} : L^2_{\mathbb{R}}(S^1) \to \mathbb{R}^{2\times 2}$ which is given by

$$\hat{K}(f) = \int_{S^1} f(x) K(x) dx.$$

Then one would use the Wigner-Eckart Theorem 4.1.13 in order to determine the matrix elements of $\hat{K}$. What this ultimately means is to determine $\hat{K}(f)$ for all orthonormal basis functions f, which are $\sin_l$ and $\cos_l$. Exemplary for $\cos_l$, we obtain:

$$\hat{K}(\cos_l) = \int_{S^1} \cos_l(x) K(x) dx = \left(\int_{S^1} \cos_l(x) K_{jk}(x) \right)_{j,k} = \left(A_{jk,l}\right)_{j,k}.$$

This is exactly the same data as before! So, probably without knowing, Weiler and Cesa [9] determined the matrix elements of a kernel operator in order to figure out a solution for these kernels. Our work can then interpreted as follows: instead of viewing these matrix elements as separate, we bundle them together to one object which we call a kernel operator and then use general ideas from representation theory and physics in order to determine their form. So, without intending to do so, we ultimately generalized and clarified the method in their work. In Section 6.2, we rederive the kernel constraint for $\mathrm{SO}(2)$-equivariant kernels from scratch with our method and demonstrate that the results coincide.

5.2. Other Work on Steerable CNNs

Another interesting paper to compare with is 3D steerable CNNs [8]. In this paper, the authors create steerable CNNs for the group $\mathrm{SO}(3)$ with the underlying field being the real numbers $\mathbb{R}$. Notably, their strategy is a bit different from ours. They consider kernels

$$K : S^2 \to \mathrm{Hom}_{\mathbb{R}}(V_l, V_J)$$

for irreducible representations $\rho_l : \mathrm{SO}(3) \to \mathrm{O}(V_l)$ and $\rho_J : \mathrm{SO}(3) \to \mathrm{O}(V_J)$[2]. Using our method, they would consider square-integrable functions on S^2 and view the

[1] In this formula, the measure on S^1 is considered to be normalized so that $\mu(S^1) = 1$.

[2] Note that since we work with real representations, what we usually call "unitary" representations are now orthogonal representations, and the unitary group is replaced by the orthogonal group.

kernel as a map $\hat{K} : L^2_{\mathbb{R}}(S^2) \otimes V_l \to V_J$, after which they would extract irreducible representations out of $L^2_{\mathbb{R}}(S^2) \otimes V_l$ using the Clebsch-Gordan decomposition. Instead, they decompose the right Hom-space into irreducible representations. This works, formally, by first constructing an isomorphism of representations

$$\mathrm{Hom}_{\mathbb{R}}(V_l, V_J) \cong V_l^* \otimes V_J \tag{5.1}$$

where V_l^* is the dual vector space of V_l, consisting of linear functions from V_l to $\mathbb{R}$. To make this precise, one considers the dual representation $\rho_l^* : \mathrm{SO}(3) \to \mathrm{O}(V_l^*)$ given by

$$\left[\rho_l^*(g)(\varphi)\right](v_l) \coloneqq \varphi\left(\rho_l(g)^{-1}(v_l)\right),$$

or more succinctly: $\rho_l^*(g)(\varphi) = \varphi \circ \rho_l(g)^{-1}$. Then, on the tensor product, one can consider the tensor product representation $\rho_l^* \otimes \rho_J$. Finally, one can build the isomorphism in Equation 5.1 from right to left by mapping $\varphi \otimes v_J$ to $\varphi_{v_J} : V_l \to V_J, v_l \mapsto \varphi(v_l) \cdot v_J$. Going this route, one can then decompose $V_l^* \otimes V_J$ into irreducible representations, which leaves one with Clebsch-Gordan coefficients which are different from ours. The reason is that they decompose the *specific* space $V_l^* \otimes V_J$ and search for *arbitrary* irreducible representations in them, whereas we decompose $V_j \otimes V_l$ for *arbitrary* j and always search for the *specific* representation ρ_J in there.
An issue is the following: While one has an isomorphism $\rho_l^* \cong \rho_l$ via the Bra-Ket convention for *real representations*, which they exploit, one in general does *not* have this isomorphism for complex representations. The reason is that the scalar product is then only conjugate linear in the first component, and so the map $V_l \to V_l^*, v_l \mapsto \langle v_l|$ is not linear and thus no isomorphism (being an isomorphism requires to be linear). In general, it can happen that there is no isomorphism between ρ_l and ρ_l^* at all. Overall, this leads us to believe that their method is less easy to generalize compared with our route using the Wigner-Eckart Theorem, which provides one unified story for all compact groups and both fields $\mathbb{R}$ and $\mathbb{C}$. In Section 6.5 the reader can find a derivation of steerable kernel bases for $\mathrm{SO}(3)$ over the real numbers using our method.
Similar to 3D Steerable CNNs are tensor field networks [6]. They are also equivariant with respect to the group $\mathrm{SE}(3)$ (meaning their compact transformation group is $\mathrm{SO}(3)$). Instead of operating with convolutions over an input field on a grid or, more smoothly, on a smooth block in $\mathbb{R}^3$, they perform discrete convolution of finitely many input tensors at points with arbitrary positions. Thus, they can operate on point clouds. Additionally, they have a self-interaction step in order to further process the feature vectors individually at each point.
Their work is very similar to ours and it seems that their theoretical justification is essentially a special case of the theory developed in our work. Namely, they consider input fields of order l and process this with harmonic basis functions of order j. Afterward, they linearly combine the result using Clebsch-Gordan coefficients in order to obtain an output field of order J. Comparing this with Equation 6.5 this seems very similar to what we do. The main difference is that they do not mention endomorphisms, different from our general treatment in Equation 4.3: the reason is

that the space of endomorphisms for each irreducible representation of $\mathrm{SO}(3)$ is 1-dimensional, meaning that the identity can be chosen as the basis endomorphism. This removes the endomorphisms from the equation.

Other foundational work in the realm of steerable CNNs are group convolutional CNNs by Cohen and Welling [2]. This paper mainly considers the groups C_N and D_N for $N = 4$. Group convolutional CNNs are basically steerable CNNs where only the regular representation is considered. In the more general framework of steerable CNNs, arbitrary finite-dimensional representations are considered. For the regular representations, solutions of the kernel constraint look as follows: one is allowed to construct arbitrary filters which then subsequently need to be copied and applied in different rotations and reflections, one for each group element. This is different from our method, in which the input- and output representations are first decomposed into irreducible representations, after which kernel solutions are computed for each pair of irreducible input- and output representations individually. More contemporary work on steerable CNNs like Weiler and Cesa [9] generally goes the route of decomposing representations into irreducible parts, even when they consider regular representations of finite groups. In Section 6.3, we discuss the example of the finite group $\mathbb{Z}_2$ and show that the route of decomposing the regular representation into irreducible subrepresentations leads to the same solution as in original group convolutional networks.

More recent work on group convolutional CNNs is Weiler et al. [24]. They deal with the following problem: if one considers discrete rotation groups with more than four rotation angles, then one must also rotate the filters by all such angles. If the filters are sampled in a pixel-basis, then this leads to approximation errors since the grid of the rotated filter does not align anymore with the grid of the sample space. Therefore, they go the route of choosing circular harmonics as their basis filters and then only learn the basis coefficients of these. Instead of rotating the grid-sampled filter, one can then rotate the linear combination of harmonic basis filters by doing a phase-shift in the coefficients. This alleviates the abovementioned problem. Additionally, they find a generalized version of He's weight initialization scheme [28] for an improved initialization of the neural network.

After group convolutional networks, steerable CNNs were proposed for the first time in a relatively modern formulation in Cohen and Welling [3]. They still worked with finite groups, but moved beyond the regular representation and already foreshadowed that the framework will work for continuous groups as well. The representations studied mostly in this work are so-called quotient representations of the group, of which the regular representation is a special case, and which have the nice advantage that they can be represented by permutation matrices. This allows to use regular ReLUs as nonlinearities.

However, they also went beyond this and studied irreducible representations. These cannot in general be represented by permutation matrices. However, they can, for the finite groups considered in the paper, be represented by so-called signed permutation matrices, which allows the use of concatenated ReLUs introduced in Shang et al. [29]. They mention that more general representations are possible as well and that one can assume them to be *orthogonal* (which we call *unitary* if the field $\mathbb{K}$ is not specified) since this allows to use any nonlinearity which only acts on the length of vectors. A detailed theoretical investigation of different nonlinearities can be found in Weiler and Cesa [9].

Harmonic Networks [4] are steerable CNNs in which the compact transformation group is $\mathrm{U}(1)$ and the field is the complex numbers $\mathbb{C}$. Mathematically speaking, this is the simplest case of a smooth transformation group: since the field is the complex numbers, each endomorphism space is 1-dimensional by Schur's Lemma 4.1.8, and so the endomorphisms can be ignored. Also, it turns out that each tensor product of irreducible representations is itself irreducible, and thus the Clebsch-Gordan coefficients can be ignored as well. Furthermore, the regular representation $L^2_{\mathbb{K}}(\mathrm{U}(1))$ contains each irreducible representation exactly once and the harmonic basis functions are just the characters since all irreducible representations are 1-dimensional. Notably, when I began developing the theory outlined in this work, I first focused solely on the case of harmonic networks, since they seemed doable and still provided enough structure to see the main parts of this theoretical work at play. We rederive harmonic networks in Section 6.1.

5.3. Gauge Equivariant CNNs

Recently, so-called gauge equivariant neural networks were proposed as generalizations of steerable CNNs [7]. These are neural networks that operate on feature fields on Riemannian manifolds, with the aim to be applicable to curved and topologically nontrivial surfaces like spheres for applications in domains such as medicine and climate science. Topologically nontrivial spaces pose a problem since it is a priori not clear how to apply the kernel: there is no orientation (or reference frame) that is ultimately preferred. For this reason, gauge equivariant CNNs go the route of expressing feature fields *relative to gauges*. This means that with respect to a certain local coordinatization of the manifold, the feature field is expressed as a field of feature vector coefficients. Then, the kernel is also applied with respect to the coordinatization. For two different coordinate systems, the result of the convolution will then differ. However, the goal is that they only differ up to a *gauge transformation*. That is, the results should describe the same quantity, just expressed with respect to the two different coordinate systems. When viewing coordinate changes as transformations of feature coefficient vectors (described by a representation of the so-called *structure group*), this becomes a requirement of equivariance. If instead one takes the global view and considers feature vectors as absolute quantities, this is in fact a requirement of *invariance* with respect to the chosen gauges. Different from steerable CNNs, the basic theory of gauge equivariant CNNs does not involve global active transformations of the feature fields themselves, which is the main difference between those theories.

Crucially, Cohen et al. [7] showed that the constraint on the kernel which needs to be fulfilled for the coordinate invariance is equal to the kernel constraint for steerable CNNs. Thus, our theory also fully covers the question of how to parameterize kernels

for gauge equivariant CNNs whenever the structure group is compact.
Similar networks were also considered by other authors. For example, Ruben Wiersma [30] is very similar to gauge equivariant CNNs since it also makes use of their parallel transport. The filters themselves are precisely the filters also used in harmonic networks [4].
Other work in that area is de Haan et al. [31]: here, gauge equivariant networks are discussed that operate on general discrete meshes.

5.4. Other Networks Inspired by Representation Theory and Physics

Gauge equivariant networks are physics-inspired, in the sense that the gauge symmetries that they preserve are the symmetries appearing in modern field theories. The underlying mathematical foundations for these physical theories are representation-theoretic. There is other work inspired by representation theory and physics as well. We first compare with Clebsch-Gordan nets [32] in order to understand the differences in how we and how they use the Clebsch-Gordan coefficients. Clebsch-Gordan nets operate on signals on the sphere and, after the first layer, operate fully in the Fourier space of signals on the whole rotation group $\mathrm{SO}(3)$. The feature spaces are then arbitrary complex finite-dimensional $\mathrm{SO}(3)$-representations which decompose into irreducible representations. In the Fourier space, usual nonlinearities like ReLUs cannot be meaningfully used for applications in image processing. One could transform the signals back onto the group $\mathrm{SO}(3)$ for using those, however, this would cause considerable computational cost as in spherical CNNs [12]. So instead, the authors search for an appropriate non-linearity in the Fourier space itself. What they choose is the following: For two irreducible representations $\rho_l : \mathrm{SO}(3) \to \mathrm{U}(V_l)$ and $\rho_{l'} : \mathrm{SO}(3) \to \mathrm{U}(V_{l'})$ which are in the feature space of a certain layer, they transform the data using the canonical map into the tensor product:

$$\otimes : V_l \times V_{l'} \to V_l \otimes V_{l'}.$$

This map is *bilinear*, and *not linear*, and thus they can use it as their nonlinearity. Then they decompose $V_l \otimes V_{l'}$ into irreducible subrepresentations. The explicit way of doing so are the Clebsch-Gordan coefficients. This gives then a linear isomorphism

$$C : V_l \otimes V_{l'} \to \bigoplus_{j=|l-l'|}^{l+l'} V_j$$

after which they can go on to process signals living in irreducible representations. In the linear, parameterized part of their processing, they just build arbitrary linear combinations of features that live in irreducible representations of the same order l. This is the most general linear function between features by a generalization of Schur's Lemma which is for example discussed in more detail in Bogatskiy et al. [5]. Note that

if one considers real representations instead of complex ones, this generalized Schur's Lemma is not quite true: the "coefficients" in the linear combinations of feature vectors are then allowed to be arbitrary *endomorphisms* of the irreducible representation.
We note that Clebsch-Gordan nets differ substantially from our work in how they use the Clebsch-Gordan coefficients: We use them in the description of our steerable kernels, and as such they are in our work part of the *linear* transformation steps, whereas in Clebsch-Gordan nets, these coefficients are part of the *nonlinear* step in the transformation.
Another very recent paper is Shutty and Wierzynski [33]. Their vision is to generalize equivariant neural networks to noncompact groups for which the theory is more complicated. Therefore, they consider the Lorentz group $\mathrm{SO}(3,1)$ as their point group and want to build neural networks which are then equivariant to the Poincaré group $\mathcal{P}_3 = (\mathbb{R}^4, +) \rtimes \mathrm{SO}(3,1)$. This group plays a central role in special relativity since it allows to formalize *changes of inertial reference frames* in Minkowski spacetime: not only are rotations and translations incorporated in this symmetry group, but also so-called *Lorentz boosts*. The authors mention that this might have applications in particle- and plasma physics. This is due to the high speeds involved in these application areas which make relativistic methods necessary. They write, as we also do in our work, that building neural networks that are equivariant with respect to such a group requires one to know the following:

1. Explicit irreducible representations.

2. Clebsch-Gordan coefficients.

3. Equivariant filter functions, which we call harmonic basis functions in our work[3].

They do not mention endomorphisms in this listing, probably because implicitly they only have groups in mind that have trivial endomorphism spaces. Furthermore, they do not engage with the question of how to parameterize *general* equivariant kernels, i.e. how to ensure that one has found all equivariant filter functions, and they do not prove their kernel decomposition. They mention that Clebsch-Gordan coefficients and equivariant filter functions can often be obtained analytically or numerically from the irreducible representations once these are known, and are therefore concerned mainly with how to find irreducible representations for defining their feature spaces. Instead of deriving those analytically, they *optimize* for irreducible representations by first optimizing for irreps of the corresponding Lie algebra, and then lifting this to irreps of the Lie group by exponentiation.
It is important to note that the Lorentz group is not covered by our theory since it is not compact, and thus the Peter-Weyl Theorem does not hold for such a group. However, the fact that Shutty and Wierzynski [33] decompose their kernels in a very similar way as we describe it suggests that a generalization of our results to such cases might be possible. We discuss this in more detail in Chapter 7.

[3]These are not the complete kernels but only the functions that we notate with Y_{ji}^m.

Very similar and published almost at the same time is Bogatskiy et al. [5]. They also consider Lorentz group equivariant networks, however operating completely in the Fourier space. This means, as in the case of Clebsch-Gordan networks [32], that not *fields* of vectors living in irreps are transformed, but just vectors in irreps themselves. In contrast to Shutty and Wierzynski [33], they *derive* instead of learn the irreducible representations. Furthermore, only irreps over the complex numbers $\mathbb{C}$ are considered. What this means is that the linear part of the network just consists of arbitrary linear mixings of components of the same irrep. In contrast, a mixing between different irreps, as in Clebsch-Gordan nets, happens only in the tensor product nonlinearity in terms of a Clebsch-Gordan decomposition. They derive the Clebsch-Gordan coefficients of the Lorentz group using the more well-known coefficients for $\mathrm{SU}(2)$. Additionally, a fundamental approximation result is proven, showing that they are able to approximate arbitrary continuous equivariant maps between finite-dimensional representations by their feed-forward neural network architecture. Additionally, they provide much useful information about the representation theory of the groups involved. Summarizing, we mention that this work differs from ours both in their use of the Clebsch-Gordan coefficients — which are a tool for decomposing tensor product representations into irreducible subrepresentations in their case — and in the fact that they do not process fields of activations.

5.5. Prior Theoretical Work

Since our work is purely theoretical, it is interesting to compare with earlier purely theoretical work. One recent paper is Esteves [34]. This paper is essentially a collection of well-known facts about the theory of equivariant CNNs. While it does not produce new results, it is a very concise introduction into the most important concepts. As in our work as well, the text starts with an introduction of representation theory, however only dealing with complex representations. Integration is, different from what we describe, treated for general locally compact groups instead of only compact groups. Fourier analysis and the Peter-Weyl Theorem is treated as well, including a description of the matrix coefficients of the irreducible representations of the groups $\mathrm{SU}(2)$ and $\mathrm{SO}(3)$.

In the last chapter, there is a collection and derivation of applications of these preliminary theories to the theory of equivariant CNNs. CNNs on finite groups [35], spherical CNNs [12], Clebsch-Gordan nets [32], 3D Steerable CNNs [8] (including a description of the solution of the kernel constraint), and other theoretical papers [36] [11] are discussed. We discuss these last two now:

Kondor and Trivedi [36] discuss scalar feature fields that are globally defined on a homogeneous space of a compact transformation group. They then derive general linear equivariant maps between such feature spaces and show that they are always given by convolution. In Cohen et al. [11], this is further generalized to feature fields that are possibly only locally described as functions to a space of feature vectors, however with the advantage that non-scalar fields like vector fields and tensor fields become

possible. It is important to study this in order to understand that it differs from our work. Namely, they look at groups H which they assume to be locally compact and unimodular. However, their H is *not*, as in our case, just the local structure group[4], but actually the global motion group, and thus also *not* assumed to be compact. Then, they let H act on the homogeneous space H/G with respect to the stabilizer subgroup G, which is also *not* assumed to be compact. For example, if one considers steerable CNNs on $\mathbb{R}^n$ with structure group $G = \mathrm{O}(n)$, then H is the Euclidean motion group $\mathrm{E}(n)$ and H/G would be $\mathbb{R}^n$.
Additionally, for two consecutive layers, the homogeneous space and stabilizer subgroup can change, so that there are two groups G_{in} and G_{out}. the groups have representations $\rho_{\mathrm{in}} : G_{\mathrm{in}} \to \mathrm{Aut}_{\mathbb{K}}(V_{\mathrm{in}})$ and $\rho_{\mathrm{out}} : G_{\mathrm{out}} \to \mathrm{Aut}_{\mathbb{K}}(V_{\mathrm{out}})$, which corresponds in the example above to the usual representations of $\mathrm{O}(n)$. For the general setting, feature fields then cannot globally be described as maps $H/G \to V$. The reason is that H/G may be twisted and thus its so-called feature bundle cannot necessarily be "trivialized"[5]. Nevertheless, feature fields can locally be described as maps $H/G \to V$, though the authors give other characterizations as well.
Now, what the authors show is that, again, linear equivariant maps between feature bundles are necessarily given by convolution with steerable kernels of a certain kind. One characterization of the space of steerable kernels is the space of bi-G_{in}-G_{out}-equivariant kernels of the following form:

$$\begin{aligned}&\mathrm{Hom}_{G_{\mathrm{in}} \times G_{\mathrm{out}}}(H, \mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}}))\\ &:= \Big\{K : H \to \mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}}) \mid K(g_{\mathrm{out}} h g_{\mathrm{in}}) = \rho_{\mathrm{out}}(g_{\mathrm{out}}) \circ K(h) \circ \rho_{\mathrm{in}}(g_{\mathrm{in}})\Big\}.\end{aligned}$$

We consider it an interesting adventure to test whether the techniques developed in our work would also allow one to come up with a *solution* for this kernel constraint. We discuss this in Chapter 7.

[4] With "structure group" we mean what we call G in the body of this work.

[5] This seems, *at first sight*, to also be a problem for us, since our homogeneous space X can also be topologically nontrivial, for example a sphere. The reason this is not an issue is that X is in applications of steerable CNNs naturally embedded in $\mathbb{R}^n$, and the feature bundle over this space *is* trivial.

6. Example Applications

In this chapter, we develop some relevant examples of the theory outlined in prior chapters. All of these examples are applications of Theorem 4.1.15 and Corollary 4.1.16. These examples are concerned with the following question: Given a *specific* field $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, compact transformation group G and homogeneous space X of G, how can a basis of steerable kernels $K : X \to \mathrm{Hom}_{\mathbb{K}}(V_l, V_J)$ for given irreducible representations $\rho_l : G \to \mathrm{U}(V_l)$ and $\rho_J : G \to \mathrm{U}(V_J)$ be determined? The theorems give an outline for what needs to be done in order to succeed in this task, and the steps are always as follows:

1. For each $l \in \hat{G}$, a representative for the isomorphism class of irreducible representations l needs to be determined. That is, one needs to determine $\rho_l : G \to \mathrm{U}(V_l)$ and an orthonormal basis $\{Y_l^n \mid n \in \{1, \dots, [l]\}\}$. We omit the index n if there is only one basis element. Usually, we have $V_l = \mathbb{K}^{[l]}$ and the orthonormal basis is just the standard basis.

2. The Peter-Weyl Theorem 2.1.22 gives the existence-statement for a decomposition of $L^2_{\mathbb{K}}(X)$ into irreducible subrepresentations. We need an *explicit* such decomposition, i.e.: we need to find multiplicities m_j, irreducible subrepresentations $V_{ji} \cong V_j$ for $i \in \{1, \dots, m_j\}$ and basis functions $Y_{ji}^m \in V_{ji} \subseteq L^2_{\mathbb{K}}(X)$ corresponding to the Y_j^m such that $L^2_{\mathbb{K}}(X) = \widehat{\bigoplus}_{j \in \hat{G}} \bigoplus_{i=1}^{m_j} V_{ji}$.

3. For each combination of j, l and J in $\hat{G}$, one needs to find the number of times $[J(jl)]$ that V_J appears in a direct sum decomposition of $V_j \otimes V_l$. Then, for each $s \in \{1, \dots, [J(jl)]\}$, and for all basis-indices M, m and n, one needs to determine the Clebsch-Gordan coefficients $\langle s, JM | jmln \rangle$. We omit the index s if V_J appears only once in the direct sum decomposition of $V_j \otimes V_l$.

4. For each J one needs to determine a basis $\{c_r \mid r \in R\}$ of the space of endomorphisms of V_J, namely $\mathrm{End}_{G,\mathbb{K}}(V_J)$.

Once all of this is done, one can then simply write down the basis kernels according to Equation 4.4 or, in case that the space of endomorphisms is 1-dimensional, Equation 4.5. The ingredients determined above are purely representation-theoretic information about the situation at hand, which hopefully makes the reader appreciate the results even more: We do not simply determine basis kernels. We understand in detail, along the way, the representation theory of the group and homogeneous space.
Note that we are not concerned with practical considerations related to how fine-grained to do this in practice (for example if the space on which the kernels operate splits into infinitely many orbits). For such questions, we refer back to Remark 4.1.18.

In the following sections, we discuss harmonic networks, $\mathrm{SO}(2)$-equivariant CNNs with real representations, reflection-equivariant networks, $\mathrm{SO}(3)$-equivariant CNNs with both complex and real representations, and $\mathrm{O}(3)$-equivariant CNNs with both complex and real representations. For each of these examples, we go through the four steps outlined above. We recommend looking at the first example in detail: we conduct it in the greatest detail and it is the easiest to understand and thus serves as a nice introduction.

6.1. Harmonic Networks

Here, we explain how the kernel constraint for harmonic networks [4] can be solved using our theory. Let $\mathrm{U}(1)$ be the group of rotations of $\mathbb{C} = \mathbb{R}^2$, i.e. the group of elements in $\mathbb{C}$ with length 1. This is also called the *circle group* since the group elements lie on a circle. In the case of harmonic networks, we have $\mathbb{K} = \mathbb{C}$, $G = \mathrm{U}(1)$, $X = S^1$. As in most examples that follow, we ignore the solution of the kernel constraint in the origin, since it is usually easy to solve.
We now go through the four steps outlined above. Our statements about the representation theory of the circle group can be found in Kowalski [15], chapter 5.

6.1.1. Construction of the Irreducible Representations of $\mathrm{U}(1)$

We have $\widehat{\mathrm{U}(1)} = \mathbb{Z}$, and for $l \in \mathbb{Z}$ we can construct a representative $\rho_l : \mathrm{U}(1) \to \mathrm{U}(V_l)$ as follows: $V_l = \mathbb{C}$ is just the canonical 1-dimensional $\mathbb{C}$-vector space, and ρ_l is given by

$$[\rho_l(g)]\,(z) := g^l \cdot z,$$

where g is regarded as an element in $\mathbb{C}$. One can easily check that this is an irreducible representation. The orthonormal basis element for each such representation is just given by $1 \in \mathbb{C} = V_l$. This already answers step 1 of the outline above.

6.1.2. The Peter-Weyl Theorem for $L^2_{\mathbb{C}}(S^1)$

For step 2, we need to determine the Peter-Weyl decomposition of $L^2_{\mathbb{C}}(S^1)$, where we regard S^1 as a subset of $\mathbb{C}$. Let $Y_{l1} : S^1 \to \mathbb{C}$ be given by $Y_{l1}(z) = z^{-l}$. Let $V_{l1} \subseteq L^2_{\mathbb{C}}(S^1)$ just be given by its span: $V_{l1} = \mathrm{span}_{\mathbb{C}}(Y_{l1})$. We want to see that this is a subrepresentation of $L^2_{\mathbb{C}}(S^1)$. To see this, remember that the unitary representation on $L^2_{\mathbb{C}}(X)$ is given by $\lambda : \mathrm{U}(1) \to \mathrm{U}(L^2_{\mathbb{C}}(S^1))$ with $[\lambda(g)\varphi]\,(z) = \varphi(g^{-1}z)$. We have

$$[\lambda(g)Y_{l1}]\,(z) = Y_{l1}(g^{-1}z) = (g^{-1}z)^{-l} = g^l \cdot z^{-l} = \left(g^l \cdot Y_{l1}\right)(z) \tag{6.1}$$

and thus $\lambda(g)Y_{l1} = g^l Y_{l1} \in V_{l1}$, which is what we claimed. Since the V_{l1} are 1-dimensional, they are necessarily irreducible for dimension reasons. Now, an important result from Fourier analysis is that the Y_{l1} for $l \in \mathbb{Z}$ actually form an orthonormal basis of $L^2_{\mathbb{C}}(S^1)$ and that, consequently, the Peter-Weyl decomposition of $L^2_{\mathbb{C}}(S^1)$ looks as follows:

$$L^2_{\mathbb{C}}(S^1) = \widehat{\bigoplus_{l\in\mathbb{Z}}} V_{l1}.$$

From this we see that the multiplicities m_l are all given by 1. What is missing is the connection to the irreps $\rho_l : \mathrm{U}(1) \to \mathrm{U}(V_l)$, but we have already indicated this in the notation. Namely, the map $f_l : V_l \to V_{l1}$ given by $z \mapsto z \cdot Y_{l1}$ is clearly an isomorphism of vector spaces, and due to Equation 6.1 even an isomorphism of representations:

$$\begin{aligned} f_l\big(\rho_l(g)(z)\big) &= f_l\big(g^l \cdot z\big) \\ &= (g^l \cdot z) \cdot Y_{l1} \\ &= z \cdot (g^l \cdot Y_{l1}) \\ &= z \cdot (\lambda(g)(Y_{l1})) \\ &= \lambda(g)\big(z \cdot Y_{l1}\big) \\ &= \lambda(g)\big(f_l(z)\big). \end{aligned}$$

Thus, $f_l \circ \rho_l(g) = \lambda(g) \circ f_l$ for all $g \in \mathrm{U}(1)$ and, as claimed, f_l turns out to be an isomorphism. This completely finishes step 2 of the outline above.

6.1.3. The Clebsch-Gordan Decomposition

For step 3, we proceed as follows: The map

$$f : V_j \otimes V_l \to V_{j+l}, \ z_j \otimes z_l \mapsto z_j \cdot z_l$$

is clearly well-defined and linear by the universal property of tensor products, see Definition 4.1.1. Furthermore, it is an isometry: namely, since the scalar product in $\mathbb{C}$ is just the usual multiplication (with the left entry being complex conjugated), we obtain

$$\begin{aligned} \big\langle f(z_j \otimes z_l) \big| f(z'_j \otimes z'_l) \big\rangle &= \big\langle z_j z_l \big| z'_j z'_l \big\rangle \\ &= \overline{z_j z_l} \cdot z'_j z'_l \\ &= \overline{z_j} z'_j \cdot \overline{z_l} z'_l \\ &= \big\langle z_j \big| z'_j \big\rangle \cdot \langle z_l | z'_l \rangle \\ &= \big\langle z_j \otimes z_l \big| z'_j \otimes z'_l \big\rangle. \end{aligned}$$

In the last step, we have used the definition of the scalar product on the tensor product, Definition 4.1.2. Thus, f is an isomorphism of Hilbert spaces. Finally, it also respects the representations since

$$\begin{aligned} f\Big(\left[(\rho_j \otimes \rho_l)(g)\right](z_j \otimes z_l)\Big) &= f\Big(\left[\rho_j(g)\right](z_j) \otimes \left[\rho_l(g)\right](z_l)\Big) \\ &= f\big(g^j z_j \otimes g^l z_l\big) \end{aligned}$$

$$
\begin{aligned}
&= g^j z_j \cdot g^l z_l \\
&= g^{j+l} \cdot (z_j z_l) \\
&= [\rho_{j+l}(g)] \, (f(z_j \otimes z_l))
\end{aligned}
$$

and thus $f \circ (\rho_j \otimes \rho_l)(g) = \rho_{j+l}(g) \circ f$ for all $g \in \mathrm{U}(1)$. Finally, the basis vectors correspond in the simplest possible way since $f(1 \otimes 1) = 1$.
Overall, what we've shown is the following: V_J is a direct summand of $V_j \otimes V_l$ if and only if $J = j + l$. If this is the case, we have $[J(jl)] = 1$ and can thus omit the index s. The only Clebsch-Gordan coefficient is then given by $\langle J1|j1l1\rangle = 1$ since the basis elements directly correspond.

6.1.4. Endomorphisms of V_J

This is the simplest part: Since we are considering representations over $\mathbb{C}$, Schur's Lemma 4.1.8 tells us that $\mathrm{End}_{\mathrm{U}(1),\mathbb{C}}(V_J)$ is 1-dimensional for each irrep J, and thus we can ignore the endomorphisms altogether.

6.1.5. Bringing Everything Together

We now show that a basis of steerable kernels $K : S^1 \to \mathrm{Hom}_{\mathbb{C}}(V_l, V_J)$ the group $\mathrm{U}(1)$ is given, when expressed as 1×1-matrix parameterized by S^1, by the basis function $Y_{l-J} : S^1 \to \mathbb{C}$. We remove the index "1" at the basis function to remove clutter. How can we see this result, using Equation 4.5?
Note that V_J can only appear as a direct summand of $V_j \otimes V_l$ if $j = J - l$ by what we've shown above. The "matrix" of Clebsch-Gordan coefficients $\mathrm{CG}_{J((J-l)l)}$ is then just the number 1. We can omit the vacuous indices i and s and obtain that the only basis kernel is given by

$$
\begin{aligned}
K_{J-l}(x) = \langle Y_{J-l}|x\rangle &= \overline{Y_{J-l}(x)} \\
&= \overline{x^{-(J-l)}} \\
&= x^{-(l-J)} \\
&= Y_{l-J}(x).
\end{aligned}
$$

This result is precisely equal to the one obtained in the original paper [4]. This concludes our investigations of harmonic networks.

6.2. $\mathrm{SO}(2)$-Equivariant Kernels for Real Representations

In this section, we look at the case $\mathbb{K} = \mathbb{R}$, $G = \mathrm{SO}(2)$ and $X = S^1$. In the following sections, we again step by step determine the representation-theoretic ingredients that we need for the application of our theorem. We remark that the resulting kernels are

not new, since Weiler and Cesa [9] have solved for this kernel basis already. However, we want to emphasize again that with our method, we learn more about the representation theory of $\mathrm{SO}(2)$ and thus get an overall better conceptual understanding of how the kernels arise.
Since it will help the presentation of our results, we set $\mathrm{SO}(2) = \mathbb{R}/2\pi\mathbb{Z}$, i.e. we view $\mathrm{SO}(2)$ as a group of angles. We also set $S^1 = \mathbb{R}/2\pi\mathbb{Z}$, that is we take the interval $[0, 2\pi]$ as the space where our functions are defined. Consequently, since we want our Haar measure to be normalized, we have to put the fraction $\frac{1}{2\pi}$ before all of our integrals, different from what we did in our treatment of $\mathrm{SO}(2)$ over $\mathbb{C}$.
Note that since we now consider representations over the real numbers, unitary representations become *orthogonal* and we write $\mathrm{O}(V)$ instead of $\mathrm{U}(V)$.

6.2.1. Construction of the Irreducible Representations of $\mathrm{SO}(2)$

The irreps of $\mathrm{SO}(2)$ over $\mathbb{R}$ are given by $\rho_l : \mathrm{SO}(2) \to \mathrm{O}(V_l)$, $l \in \mathbb{N}_{\geq 0}$. For $l = 0$, we have $V_0 = \mathbb{R}$ and the action is trivial. For $l \geq 1$, $V_l = \mathbb{R}^2$ as a vector space. The action is given by

$$\big[\rho_l(\phi)\big](v) = \begin{pmatrix} \cos(l\phi) & -\sin(l\phi) \\ \sin(l\phi) & \cos(l\phi) \end{pmatrix} \cdot v$$

for $\phi \in \mathrm{SO}(2) = \mathbb{R}/2\pi\mathbb{Z}$. The orthonormal basis is in both cases just given by standard basis vectors.

6.2.2. The Peter-Weyl Theorem for $L^2_{\mathbb{R}}(S^1)$

Now we look at square-integrable functions $L^2_{\mathbb{R}}(S^1)$ that we now assume to take *real values.* As before, $\mathrm{SO}(2)$ acts on this space by $(\lambda(\phi)f)(x) = f(x - \phi)$[1]. For notational simplicity, we write $\cos_l$ for the function that maps x to $\cos(lx)$, and analogously for $\sin_l$. One then can show the following, which is a standard result in Fourier analysis:

Proposition 6.2.1. *The functions* $\cos_l, \sin_l$, $l \geq 1$ *span an irreducible invariant subspace of* $L^2_{\mathbb{R}}(S^1)$ *of dimension* 2*, explicitly given by*

$$\mathrm{span}_{\mathbb{R}}(\cos_l, \sin_l) = \big\{\alpha \cos_l + \beta \sin_l \mid \alpha, \beta \in \mathbb{R}\big\}$$

which is isomorphic as an orthogonal representation to V_l *by* $\sqrt{2}\cos_l \mapsto \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ *and* $\sqrt{2}\sin_l \mapsto \begin{pmatrix} 0 \\ 1 \end{pmatrix}$[2]*. Furthermore,* $\sin_0 = 0$ *and* $\cos_0 = 1$ *are constant functions and their span is* 1*-dimensional and equivariantly isomorphic to* V_0 *by* $\cos_0 \mapsto 1$*.*
Finally, the functions $\sqrt{2} \cdot \cos_l$, $\sqrt{2} \cdot \sin_l$ *form an orthonormal basis of* $L^2_{\mathbb{R}}(S^1)$*, i.e. every function can be written uniquely as a (possibly infinite) linear combination of these basis functions.*

[1] Note that we have a subtraction now instead of a multiplicative inversion. This is because we view our group as additive.
[2] $\sqrt{2}$ acts as a normalization.

When setting $V_{l1} = \mathrm{span}_{\mathbb{R}}(\cos_l, \sin_l)$, we thus obtain a decomposition

$$L^2_{\mathbb{R}}(S^1) = \widehat{\bigoplus_{l \geq 0}} V_{l1}.$$

Thus, we have $m_l = 1$ for all $l \in \mathbb{N}$. All in all, we know everything there is to know about the Peter-Weyl Theorem in our situation.

6.2.3. The Clebsch-Gordan Decomposition

We now do the explicit decomposition of $V_j \otimes V_l$ into irreps, which will give us the Clebsch-Gordan coefficients that we need. Instead of doing the decomposition in terms of V_j and V_l themselves, in the proofs we actually use the isomorphic images V_{j1} and V_{l1} in $L^2_{\mathbb{R}}(S^1)$. For doing so, we first need some trigonometric formulas in our disposal:

Lemma 6.2.2. *The sine and cosine functions fulfill the following rules:*

1. $\sin_{j+l} = \sin_j \cos_l + \cos_j \sin_l$.
2. $\cos_{j+l} = \cos_j \cos_l - \sin_j \sin_l$.
3. $\cos_j \cos_l = \frac{1}{2}\left[\cos_{j+l} + \cos_{j-l}\right]$.
4. $\sin_j \cos_l = \frac{1}{2}\left[\sin_{j+l} + \sin_{j-l}\right]$.
5. $\cos_j \sin_l = \frac{1}{2}\left[\sin_{j+l} - \sin_{j-l}\right]$.
6. $\sin_j \sin_l = \frac{1}{2}\left[\cos_{j-l} - \cos_{j+l}\right]$.

Proof. The first two are well-known and the last four follow directly from the first two using $\sin_{-j} = -\sin_j$ and $\cos_{-j} = \cos_j$ where needed. $\square$

We will need the following general lemma:

Lemma 6.2.3. *Let $f : V \to V'$ be an intertwiner between representations $\rho : G \to \mathrm{Aut}_{\mathbb{K}}(V)$ and $\rho' : G \to \mathrm{Aut}_{\mathbb{K}}(V')$. Then $\mathrm{null}(f) = \{v \in V \mid f(v) = 0\}$ is an invariant linear subspace of V.*

Proof. This can very easily be checked by the reader. $\square$

As a remark on notation for the following proposition: We write the Clebsch-Gordan coefficients $\mathrm{CG}_{J(jl)s}$ of irreps V_J, V_j and V_l with dimensions $[J]$, $[j]$ and $[l]$ as a $[J] \times ([j] \times [l])$-tensor. That is, it consists of $[J]$ "rows", each of which is a $[j] \times [l]$-matrix. If V_J appears only once in the tensor product, we omit the index s as before.

Proposition 6.2.4. *We have the following decomposition results:*

1. *For $j = l = 0$ we have $V_0 \otimes V_0 \cong V_0$ and Clebsch-Gordan coefficients $\mathrm{CG}_{0(00)} = \Big([1]\Big)$.*

2. *For* $j = 0, l > 0$ *we have* $V_0 \otimes V_l \cong V_l$ *and Clebsch-Gordan coefficients* $\mathrm{CG}_{l(0l)} = \left(\begin{bmatrix} 1 & 0 \end{bmatrix} \atop \begin{bmatrix} 0 & 1 \end{bmatrix} \right)$.

3. *For* $j > 0, l = 0$*, we get* $V_j \otimes V_0 \cong V_j$ *and Clebsch-Gordan coefficients* $\mathrm{CG}_{j(j0)} = \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} \atop \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right)$.

4. *For* $j, l \geq 0$ *and* $j \neq l$ *we get* $V_j \otimes V_l \cong V_{|j-l|} \oplus V_{j+l}$*. The Clebsch-Gordan coefficients are given by* $\mathrm{CG}_{|j-l|(jl)} = \left(\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \atop \begin{bmatrix} 0 & -\frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix} \right)$ *and* $\mathrm{CG}_{j+l,(jl)} = \left(\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{bmatrix} \atop \begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix} \right)$.

5. *For* $j = l > 0$*, we get an isomorphism* $V_l \otimes V_l \cong V_{2l} \oplus V_0^2$*. We obtain the Clebsch-Gordan coefficients* $\mathrm{CG}_{0(ll)1} = \left(\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \right)$, $\mathrm{CG}_{0(ll)2} = \left(\begin{bmatrix} 0 & -\frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix} \right)$ *and* $\mathrm{CG}_{2l,(ll)} = \left(\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{bmatrix} \atop \begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix} \right)$*, the last one corresponding to the Clebsch-Gordan coefficients* $\mathrm{CG}_{j+l,(jl)}$ *from above. In* $\mathrm{CG}_{0(ll)1}$ *and* $\mathrm{CG}_{0(ll)2}$*, a fourth index is present, namely* 1 *and* 2*, respectively. This is the index "s" that was missing in all the prior examples, since this is the first time an irrep appears more than once in a tensor product decomposition.*

Proof. In the proof, instead of directly with the irreps $\rho_j : \mathrm{SO}(2) \to \mathrm{O}(V_j)$, we use the isomorphic copies V_{j1} in $L^2_{\mathbb{R}}(S^1)$ given in Proposition 6.2.1. Since we think that it does not help understanding to carry the index "1" in all computations, we omit this index.

The proof of 1, 2 and 3 is clear.

For 4, consider the (unnormalized) basis $\{b_{cc}, b_{cs}, b_{sc}, b_{ss}\}$ of $V_j \otimes V_l$, where for example $b_{cc} = \cos_j \otimes \cos_l$. Our goal is to express these basis elements with respect to basis elements of invariant subspaces. We do this by explicitly constructing an isomorphism to a decomposition of irreps. To that end, let $p : V_j \otimes V_l \to L^2_{\mathbb{R}}(S^1)$ be given by $f \otimes g \mapsto f \cdot g$, which is clearly a well-defined intertwiner. Let $b'_{cc} = p(b_{cc})$, and the same for the other basis vectors. We get as image of p the set

$$\begin{aligned} \mathrm{im}(p) &= \mathrm{span}_{\mathbb{R}} \left(b'_{cc}, b'_{cs}, b'_{sc}, b'_{ss} \right) \\ &= \mathrm{span}_{\mathbb{R}} \left(\cos_j \cdot \cos_l,\ \cos_j \cdot \sin_l,\ \sin_j \cdot \cos_l,\ \sin_j \cdot \sin_l \right), \end{aligned}$$

From Lemma 6.2.2 we obtain:

$$b'_{cc} - b'_{ss} = \cos_{j+l},\ b'_{cs} + b'_{sc} = \sin_{j+l},\ b'_{cc} + b'_{ss} = \cos_{j-l},\ b'_{sc} - b'_{cs} = \sin_{j-l}\,.$$

Since these are linearly independent basis functions, we obtain:

$$\operatorname{im}(p) = \operatorname{span}_{\mathbb{R}}\left(\cos_{j+l}, \sin_{j+l}, \cos_{j-l}, \sin_{j-l}\right) = V_{j+l} \oplus V_{|j-l|}.$$

Note for the last step that due to symmetry, $\cos_{j-l} = \cos_{l-j}$ and $\sin_{j-l} = -\sin_{l-j}$. Now we define the following second set of (not necessarily orthonormal) basis-elements in $V_j \otimes V_l$, corresponding to the basis-elements $V_{|j-l|} \oplus V_{l+j}$ by means of the isomorphism p:

$$c_1 = b_{cc} - b_{ss},\ c_2 = b_{cs} + b_{sc},\ c_3 = b_{cc} + b_{ss},\ c_4 = b_{sc} - b_{cs}.$$

We obtain $V_j \otimes V_l = \operatorname{span}_{\mathbb{R}}(c_1, c_2) \oplus \operatorname{span}_{\mathbb{R}}(c_3, c_4) \cong V_{l+j} \oplus V_{|j-l|}$. From the following equations we can read off the Clebsch-Gordan coefficients:

$$b_{cc} = \frac{1}{2}\left[c_1 + c_3\right],\ b_{cs} = \frac{1}{2}\left[c_2 - c_4\right],\ b_{sc} = \frac{1}{2}\left[c_2 + c_4\right],\ b_{ss} = \frac{1}{2}\left[c_3 - c_1\right].$$

The result follows.
Now, we prove 5. We have $j = l$ and still consider the same function p and overall notation. Note that $b'_{sc} - b'_{cs} = 0$ and $b'_{cc} + b'_{ss} = 1$ are constant function. Then by what was proven above, $\{c_1, c_2\}$ spans a space isomorphic to V_{2l}, c_3 a space isomorphic to the span of $\cos_0$, i.e. V_0, and c_4 spans the kernel, which is one-dimensional and also an invariant subspace due to Lemma 6.2.3, and therefore it spans a space isomorphic to V_0 as well. Overall, we obtain

$$V_l \otimes V_l = \operatorname{span}_{\mathbb{R}}(c_1, c_2) \oplus \operatorname{span}_{\mathbb{R}}(c_3) \oplus \operatorname{span}_{\mathbb{R}}(c_4) \cong V_{2l} \oplus V_0^2.$$

From this, we can as before read off the Clebsch-Gordan coefficients and obtain the claimed result. □

6.2.4. Endomorphisms of V_J

We now describe the endomorphisms of the irreducible representations, our last ingredient:

Proposition 6.2.5. *We have* $\operatorname{End}_{\mathrm{SO}(2),\mathbb{R}}(V_0) \cong \mathbb{R}$*, i.e. multiplications with all real numbers are valid endomorphisms of* V_0*. For* $l \geq 1$*, we get*

$$\operatorname{End}_{\mathrm{SO}(2),\mathbb{R}}(V_l) = \left\{ \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \middle| a, b \in \mathbb{R} \right\},$$

which is the set of all scaled rotations of $\mathbb{R}^2$*. When identifying* $\mathbb{R}^2 \cong \mathbb{C}$*, we can also view these transformations as arbitrary multiplications with a complex number.*
As a consequence, $\mathrm{id}_{\mathbb{R}}$ *is a basis for* $\operatorname{End}_{\mathrm{SO}(2),\mathbb{R}}(V_0)$ *and* $\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \right\}$ *a basis for* $\operatorname{End}_{\mathrm{SO}(2),\mathbb{R}}(V_l)$ *for* $l \geq 1$*.*

Proof Sketch. For $l \geq 1$ and an arbitrary matrix $E = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ that commutes with all rotation matrices $\rho_l(\phi)$, i.e. $E \circ \rho_l(\phi) = \rho_l(\phi) \circ E$, one can easily show the constraints $a = d$ and $b = -c$, from which the result follows. □

6.2.5. Bringing Everything Together

Now we have done all needed preparation and can solve the kernel constraint explicitly, using the matrix-form of the Wigner-Eckart Theorem for steerable kernels, Theorem 4.1.15. This is, as mentioned before, a new derivation of the results in Weiler and Cesa [9]. One can compare with table 8 in their appendix which only differs by (irrelevant) constants.

Proposition 6.2.6. *We consider steerable kernels $K : S^1 \to \mathrm{Hom}_{\mathbb{R}}(V_l, V_J)$, where V_l and V_J are irreducible representations of $\mathrm{SO}(2)$. Then the following holds:*

1. *For $l = J = 0$, we get $K(x) = a \cdot \begin{pmatrix} 1 \end{pmatrix}$ for every $x \in S^1$ and an arbitrary real number $a \in \mathbb{R}$ independent of x.*
2. *For $l = 0$, $J > 0$, a basis for steerable kernels is given by $\begin{pmatrix} \cos_J \\ \sin_J \end{pmatrix}$ and $\begin{pmatrix} -\sin_J \\ \cos_J \end{pmatrix}$.*
3. *For $l > 0$ and $J = 0$, a basis for steerable kernels is given by $\begin{pmatrix} \cos_l & \sin_l \end{pmatrix}$, $\begin{pmatrix} \sin_l & -\cos_l \end{pmatrix}$.*
4. *For $l, J > 0$, a basis for steerable kernels is given by the results stated in the proof.*

Proof. The proof of 1 is clear.
For 2, note that V_J can only appear in $V_j \otimes V_0$ if $j = J$. The relevant Clebsch-Gordan coefficients are by Proposition 6.2.4 therefore $\mathrm{CG}_{J(J0)} = \begin{pmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{pmatrix}$. Furthermore, the orthonormal basis of $V_{j1} = V_{J1}$ is given by Proposition 6.2.1 up to constants by $\{\cos_J, \sin_J\}$, which we have to write as a row-vector according to Theorem 4.1.15. Thereby, we can ignore the complex conjugation since we work over the real numbers. Our final ingredient is the endomorphism basis of V_J, which is by Proposition 6.2.5 given by $c_1 = \mathrm{id}_{\mathbb{R}^2}$ and $c_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Overall, the basis kernels are given by

$$c_i \cdot \begin{pmatrix} \begin{bmatrix} \cos_J & \sin_J \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ \begin{bmatrix} \cos_J & \sin_J \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{pmatrix} = c_i \cdot \begin{pmatrix} \cos_J \\ \sin_J \end{pmatrix}.$$

The result follows.
For 3, we find V_0 only in $V_j \otimes V_l$ if $j = l$, and even twice so. The relevant Clebsch-Gordan coefficients are therefore by Proposition 6.2.4 given by $\mathrm{CG}_{0(ll)1} = \left(\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\right)$ and $\mathrm{CG}_{0(ll)2} = \left(\begin{bmatrix} 0 & -\frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix}\right)$. The basis-functions in $V_{j1} = V_{l1}$ are by Proposition 6.2.1 up to constants $\{\cos_l, \sin_l\}$, again written as a row-vector. Finally, $V_J = V_0$ has only $\mathrm{id}_{\mathbb{R}}$ as a basis-endomorphism by Proposition 6.2.5, so this can be ignored altogether by Corollary 4.1.16. We obtain the following basis for steerable kernels:

$$\left(\begin{bmatrix} \cos_l & \sin_l \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\right) = \begin{pmatrix} \frac{1}{2}\cos_l & \frac{1}{2}\sin_l \end{pmatrix}$$
$$\left(\begin{bmatrix} \cos_l & \sin_l \end{bmatrix} \begin{bmatrix} 0 & -\frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix}\right) = \begin{pmatrix} \frac{1}{2}\sin_l & -\frac{1}{2}\cos_l \end{pmatrix}.$$

For 4, we consider only the case $l > J$. The case $l = J$ and $l < 0$ can be considered analogously and leads by trigonometric formulae to the same result. By Proposition 6.2.4 we have

$$V_{l-J} \otimes V_l \cong V_J \oplus V_{2l-J}, \;\; V_{l+J} \otimes V_l \cong V_J \oplus V_{2l+J},$$

i.e. $j = l - J$ and $j = l + J$ leads to a tensor product decomposition containing V_J, but no other j does. Thus, the relevant Clebsch-Gordan coefficients are by Proposition 6.2.4 the matrices $\mathrm{CG}_{J(l-J,l)}$ and $\mathrm{CG}_{J(l+J,l)}$, which are both equal and given by $\left(\begin{matrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \\ \begin{bmatrix} 0 & -\frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix} \end{matrix}\right)$. For $j = l - J$ and $j = l + J$, the basis functions of $V_{(l-J)1}$ and $V_{(l+J)1}$ are by Proposition 6.2.1 furthermore given by $\{\cos_{J-l}, \sin_{J-l}\}$ and $\{\cos_{l+J}, \sin_{l+J}\}$ respectively. Finally, V_J has again the two basis endomorphisms $c_1 = \mathrm{id}_{\mathbb{R}^1}$ and c_2 from above. Now, we do the computation for $j = l - J$, since for $j = l + J$ it is exactly the same and obtain the following two basis kernels:

$$c_i \cdot \left(\begin{matrix} \begin{bmatrix} \cos_{l-J} & \sin_{l-J} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \\ \begin{bmatrix} \cos_{l-J} & \sin_{l-J} \end{bmatrix} \cdot \begin{bmatrix} 0 & -\frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix} \end{matrix}\right) = c_i \cdot \begin{pmatrix} \frac{1}{2}\cos_{l-J} & \frac{1}{2}\sin_{l-J} \\ \frac{1}{2}\sin_{l-J} & -\frac{1}{2}\cos_{l-J} \end{pmatrix}.$$

Together with $j = l + J$, we get four basis kernels. This finishes the derivation. $\square$

6.3. $\mathbb{Z}_2$-Equivariant Kernels for Real Representations

In this section, we discuss steerable CNNs that use the finite group $\mathbb{Z}_2$, which we identify with $(\{-1, +1\}, \cdot)$, for their symmetries. We let this group act on the plane

$\mathbb{R}^2$ by vertical reflections, though other choices are possible as well:

$$x \cdot \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} xa \\ b \end{pmatrix}.$$

This example is very simple and one may see it as contrived to apply our relatively heavy theory to it. We include it mainly as a demonstration that our results can also be applied to non-smooth finite groups as instances of compact groups. Furthermore, we will fully recover the relationship to the original group convolutional CNNs from Cohen and Welling [2] and thereby demonstrate that all the different developed theories are consistent with each other.

6.3.1. The Irreducible Representations of $\mathbb{Z}_2$ over the Real Numbers

Let $\rho : \mathbb{Z}_2 \to \mathrm{Aut}_{\mathbb{R}}(V)$ be an irreducible real representation. Note that

$$\rho(-1) \circ \rho(-1) = \rho((-1) \cdot (-1)) = \rho(1) = \mathrm{id}_V,$$

and thus $\rho(-1)$ is an involution satisfying the equation $\rho(-1)^2 - \mathrm{id}_V = 0$. It is well-known from linear algebra that involutions are diagonalizable, and thus $\rho(-1)$ leaves 1-dimensional subspaces invariant. By irreducibility of ρ this means that V itself needs to be 1-dimensional. Consequently, we can assume $V = \mathbb{R}$ without loss of generality. Note that the computations above mean that we have

$$\big(\rho(-1) - \mathrm{id}_{\mathbb{R}}\big) \circ \big(\rho(-1) + \mathrm{id}_{\mathbb{R}}\big) = 0$$

and thus we need to have $\rho(-1) - \mathrm{id}_{\mathbb{R}} = 0$ or $\rho(-1) + \mathrm{id}_{\mathbb{R}} = 0$. It follows $\rho(-1) = \mathrm{id}_{\mathbb{R}}$ or $\rho(-1) = -\mathrm{id}_{\mathbb{R}}$. Overall, all these investigations mean that we have precisely two irreducible representations of $\mathbb{Z}_2$ up to equivalence. We call them $\rho_+ : \mathbb{Z}_2 \to \mathrm{O}(V_+)$ and $\rho_- : \mathbb{Z}_2 \to \mathrm{O}(V_-)$, where $\rho_+(-1) = \mathrm{id}_{\mathbb{R}}$ and $\rho_-(-1) = -\mathrm{id}_{\mathbb{R}}$ and $V_+ = V_- = \mathbb{R}$.

6.3.2. The Peter-Weyl Theorem for $L^2_{\mathbb{R}}(X)$

Here we do the Peter-Weyl decomposition for $L^2_{\mathbb{R}}(X)$, where X is one of the two homogeneous spaces $X = \{-1, 1\}$ and $X = \{0\}$ with the obvious actions coming from the groups $\mathbb{Z}_2$. This time, we also discuss orbits with only one point since we later want to get a description of kernels on the whole of $\mathbb{R}^2$ for comparisons with group convolutional CNNs.

We start with $X = \{-1, 1\}$. Note that the measure on X is just the normalized counting measure, and thus *all* functions $f : X \to \mathbb{R}$ are square-integrable. We define the two functions

$$f_+ : X \to \mathbb{R},\ f_+(x) = 1 \text{ for all } x \in X = \{-1, 1\},$$
$$f_- : X \to \mathbb{R},\ f_-(x) = x \text{ for all } x \in X = \{-1, 1\}.$$

We then define $V_{+1} = \mathrm{span}_{\mathbb{R}}(f_+)$ and $V_{-1} = \mathrm{span}_{\mathbb{R}}(f_-)$. This gives a decomposition

$$L^2_{\mathbb{R}}(X) = V_{+1} \oplus V_{-1}$$

since we have for all $f \in L^2_{\mathbb{R}}(X)$

$$f = \frac{f(1) + f(-1)}{2} \cdot f_+ + \frac{f(1) - f(-1)}{2} \cdot f_-.$$

Furthermore, the maps $1 \mapsto f_+$ and $1 \mapsto f_-$ give isomorphisms of representations $V_+ \cong V_{+1}$ and $V_- \cong V_{-1}$, respectively.
Now, assume that $X = \{0\}$ with the trivial action coming from $\mathbb{Z}_2$. Then $L^2_{\mathbb{R}}(X) = V_{+1}$ generated from the function $f_+ : X \to \mathbb{R}$, $f_+(0) = 1$. As before, $1 \mapsto f_+$ gives an isomorphism $V_+ \cong V_{+1}$. This concludes the investigations of the Peter-Weyl Theorem.

6.3.3. The Clebsch-Gordan Decomposition

We have the following four isomorphisms of representations:

$$\begin{aligned} V_+ \otimes V_+ &\cong V_+, \qquad & V_+ \otimes V_- &\cong V_-, \\ V_- \otimes V_+ &\cong V_-, \qquad & V_- \otimes V_- &\cong V_+, \end{aligned}$$

each time simply given by $a \otimes b \mapsto ab$. It can easily be checked that these are isomorphisms. In Section 6.6.3 the reader can find a proof for similar, sign-dependent isomorphisms for the case that the group is $\mathrm{O}(3)$. For each such isomorphism, there is precisely one Clebsch-Gordan coefficient and it is just given by 1. Thus, as in the case of harmonic networks in Section 6.1.5, we can just ignore the Clebsch-Gordan coefficients altogether in the final formulas for our basis kernels.

6.3.4. Endomorphisms of V_+ and V_-

Since V_+ and V_- are themselves only 1-dimensional, the endomorphism spaces are necessarily 1-dimensional as well and just given by arbitrary 1×1-matrices, i.e. arbitrary stretchings. As in the example of harmonic networks, we can therefore ignore the endomorphisms as well.

6.3.5. Bringing Everything Together

Different from the other examples, we will in this section not only engage with the final steerable kernels on homogeneous spaces but also discuss how these assemble to kernels defined on the whole plane $\mathbb{R}^2$. In the end, we will then also discuss how kernels for the regular representation would look like.
But first, we engage with the homogeneous spaces. We start with $X = \{-1, 1\}$ and consider steerable kernels $K : X \to \mathrm{Hom}_{\mathbb{R}}(V_{\mathrm{in}}, V_{\mathrm{out}})$ for irreducible V_{in} and V_{out}. There are four possibilities for the input and output representations:

Steerable Kernels $K : X \to \mathrm{Hom}_{\mathbb{R}}(V_+, V_+)$:

V_+ can only be in a tensor product $V \otimes V_+$ if the sign of V is positive as well. Such a space appears precisely once in $L^2_{\mathbb{R}}(X)$ according to Section 6.3.2. Since endomorphisms and Clebsch-Gordan coefficients do not appear by what we've shown before, and since complex conjugation doesn't do anything over the real numbers, a basis for steerable kernels is just given by the one kernel $K_+ = f_+$ itself. Here, we identify $\mathrm{Hom}_{\mathbb{R}}(V_+, V_+)$ with $\mathbb{R}$ since it only consists of 1×1-matrices.

Steerable Kernels $K : X \to \mathrm{Hom}_{\mathbb{R}}(V_+, V_-)$:

By the same arguments, a basis is given by the one kernel $K_- = f_-$.

Steerable Kernels $K : X \to \mathrm{Hom}_{\mathbb{R}}(V_-, V_+)$:

Again, a basis for steerable kernels is given by $K_- = f_-$.

Steerable Kernels $K : X \to \mathrm{Hom}_{\mathbb{R}}(V_-, V_-)$:

A basis is given by $K_+ = f_+$.
Finally, we also need to engage with the case that $X = \{0\}$ consists only of a single point. Similarly to above, in the "even" case that the signs of input- and output representations agree, a basis is given by $K_+ = f_+$ with $f_+(0) = 1$. If, however, the signs do not agree, then only $K = 0$ fulfills the constrained and the basis is empty.
Now, we assemble this to kernels on the whole of $\mathbb{R}^2$. We saw above that we only need to distinguish two cases, namely (a) the case that the signs of input and output representation agree and (b) that they do not.
For case (a), let $K : \mathbb{R}^2 \to \mathbb{R}$ be a steerable kernel, where $\mathbb{R}$ is isomorphic to the Hom-space between equal-sign representations. $\mathbb{R}^2$ splits disjointly into orbits, namely $\left\{ \begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} -a \\ b \end{pmatrix} \right\}$ for all $a \in \mathbb{R}_{\geq 0}$ and $b \in \mathbb{R}$. If $a = 0$, then the orbit is just a single point, which means that we have a vertical line of single-point orbits. The solution above showed that on each orbit, the kernel needs to be constant (since f_+ is constant) and overall this just translates to

$$K \begin{pmatrix} a \\ b \end{pmatrix} = K \begin{pmatrix} -a \\ b \end{pmatrix}$$

for all $a \geq 0$ and $b \in \mathbb{R}$. Consequently, K is just an arbitrary left-right symmetric kernel.
In the case that the input- and output representations do not share their sign, by the same arguments we see that $K : \mathbb{R}^2 \to \mathbb{R}$ is an arbitrary left-right *anti*-symmetric kernel which is zero on the vertical line $\begin{pmatrix} 0 \\ b \end{pmatrix}$ for arbitrary $b \in \mathbb{R}$.
Other than these left-right restrictions, the kernel can be freely learned. Overall, this means that we learn one "half" of the kernel and can recover the other half by the symmetry property derived above.

6.3.6. Group Convolutional CNNs for $\mathbb{Z}_2$

We now investigate what all this means if we consider regular representations instead of irreducible representations, thus corresponding to group convolutional kernels as in [2]. In this case, we will see an interesting "twist" in the kernel, which makes this example more interesting than one might initially think. The twist emerges as follows: For regular representations, we consider steerable kernels

$$K : \mathbb{R}^2 \to \mathrm{Hom}_{\mathbb{R}}(L^2_{\mathbb{R}}(\mathbb{Z}_2), L^2_{\mathbb{R}}(\mathbb{Z}_2))$$

Now, there are two relatively canonical bases we can choose in the left and the right space. We already know from above that $\{f_+, f_-\}$ is the basis to choose if we want to express steerable kernels corresponding to irreducible representations. However, for vanilla group convolutional CNNs, the basis usually chosen is $\{e_{+1}, e_{-1}\}$ where $e_{+1}(x) = \delta_{+1,x}$ and $e_{-1}(x) = \delta_{-1,x}$. We then obtain the following four base change relations:

$$\begin{aligned} f_+ &= e_{+1} + e_{-1}, \qquad & f_- &= e_{+1} - e_{-1}, \\ e_{+1} &= \frac{1}{2}f_+ + \frac{1}{2}f_-, \qquad & e_{-1} &= \frac{1}{2}f_+ - \frac{1}{2}f_-. \end{aligned}$$

Thus, the base change matrices are given by

$$B = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \qquad B^{-1} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}.$$

Now, assume that $K : \mathbb{R}^2 \to \mathrm{Hom}_{\mathbb{R}}(L^2_{\mathbb{R}}(\mathbb{Z}_2), L^2_{\mathbb{R}}(\mathbb{Z}_2)) \cong \mathbb{R}^{2\times 2}$ is expressed with respect to the basis $\{f_+, f_-\}$. If we write K as a matrix

$$K = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix}$$

then we know that K_{11} and K_{22} map between equal-sign representations and K_{12} and K_{21} between unequal-sign representations. Consequently, from what we've found above, K_{11} and K_{22} are symmetric, whereas K_{12} and K_{21} are antisymmetric. What we now want to figure out is how exactly this translates to a property of the kernel expressed in the basis $\{e_+, e_-\}$.

Thus, let K' be this corresponding kernel. Then using the base change matrices above we obtain

$$\begin{aligned} \begin{pmatrix} K'_{11} & K'_{12} \\ K'_{21} & K'_{22} \end{pmatrix} &= K' \\ &= B \cdot K \cdot B^{-1} \\ &= \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2}\left[K_{11} + K_{12} + K_{21} + K_{22}\right] & \frac{1}{2}\left[K_{11} - K_{12} + K_{21} - K_{22}\right] \\ \frac{1}{2}\left[K_{11} + K_{12} - K_{21} - K_{22}\right] & \frac{1}{2}\left[K_{11} - K_{12} - K_{21} + K_{22}\right] \end{pmatrix}. \end{aligned}$$

What symmetry properties does this kernel obey? In order to understand this, we use the following convention: for $y \in \mathbb{R}^2$ we set $-y = \begin{pmatrix} -y_1 \\ y_2 \end{pmatrix}$, i.e. the vertically flipped image of y. Then we have, using the symmetry and anti-symmetry of the entries of the original kernel K:

$$\begin{aligned}
K'_{22}(-y) &= \frac{1}{2}\big[K_{11}(-y) - K_{12}(-y) - K_{21}(-y) + K_{22}(-y)\big] \\
&= \frac{1}{2}\big[K_{11}(y) + K_{12}(y) + K_{21}(y) + K_{22}(y)\big] \\
&= K'_{11}(y), \\
K'_{21}(-y) &= \frac{1}{2}\big[K_{11}(-y) + K_{12}(-y) - K_{21}(-y) - K_{22}(-y)\big] \\
&= \frac{1}{2}\big[K_{11}(y) - K_{12}(y) + K_{21}(y) - K_{22}(y)\big] \\
&= K'_{12}(y).
\end{aligned}$$

Thus the second row of K' is basically the same as the first, only that the kernels swap with each other and are internally flipped. This is a special case of the outcome in Cohen and Welling [2], which is also described rather clearly in Weiler et al. [24]: in group convolutional kernels which are steerable with respect to finite groups, the kernels get copied and applied in all orientations demanded by the group.

What we would still like to understand is if we can also reverse the direction: That is, assume that we start with a group convolutional kernel K' of which we know that $K_{22'}(-y) = K'_{11}(y)$ and $K'_{21}(-y) = K'_{12}(y)$ for all $y \in \mathbb{R}^2$. If we then do a base change, we would like to know if the resulting kernel consists of symmetric and antisymmetric entries. Namely, set

$$\begin{aligned}
\begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} &= K \\
&= B^{-1} \cdot K' \cdot B \\
&= \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \cdot \begin{pmatrix} K'_{11} & K'_{12} \\ K'_{21} & K'_{22} \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{2}\left[K'_{11} + K'_{12} + K'_{21} + K'_{22}\right] & \frac{1}{2}\left[K'_{11} - K'_{12} + K'_{21} - K'_{22}\right] \\ \frac{1}{2}\left[K'_{11} + K'_{12} - K'_{21} - K'_{22}\right] & \frac{1}{2}\left[K'_{11} - K'_{12} - K'_{21} + K'_{22}\right] \end{pmatrix}.
\end{aligned}$$

The reader can easily check that we can deduce that K_{11} and K_{22} are symmetric and that K_{12} and K_{21} are anti-symmetric. We have thus fully shown the equivalence of the kernel solutions in the setting of steerable CNNs compared to the setting of group convolutional CNNs for the specific group $\mathbb{Z}_2$.

6.4. $\mathrm{SO}(3)$-Equivariant Kernels for Complex Representations.

In the first two sections, we have discussed $\mathrm{U}(1) \cong \mathrm{SO}(2)$-equivariant kernels (i.e., $\mathrm{SE}(2)$-equivariant neural networks) both over $\mathbb{C}$ and $\mathbb{R}$. The situation over $\mathbb{R}$ was considerably more complicated and required new arguments. In this section, we will discuss $\mathrm{SO}(3)$-equivariant kernels (i.e. $\mathrm{SE}(3)$-equivariant neural networks) for complex representations. In Section 6.5 we will then look at the real case, which will essentially give the exact same results, thus differing somewhat from the considerations about $\mathrm{U}(1) \cong \mathrm{SO}(2)$. Different from the earlier sections, we will from now on be less explicit and care more about the general properties of the different functions and coefficients we consider. $\mathrm{SO}(3)$-equivariant networks with real coefficients have before been implemented in Weiler et al. [8] and Thomas et al. [6], among others.

6.4.1. The Irreducible Representations of $\mathrm{SO}(3)$ over the Complex Numbers

In this section, we state the complex irreducible representations of $\mathrm{SO}(3)$. We will not state the matrices explicitly since the matrix elements are considerably more complicated than in the earlier examples that we saw. For each $l \in \mathbb{N}_{\geq 0}$, there is one irreducible unitary representation

$$D_l : \mathrm{SO}(3) \to \mathrm{U}(V_l), \text{ where } V_l = \mathbb{C}^{2l+1}.$$

The matrices $D_l(g)$ for $g \in \mathrm{SO}(3)$ are called the *Wigner D-matrices*[3]. There are, up to equivalence, no other irreducible representations of $\mathrm{SO}(3)$ over $\mathbb{C}$. A reference for all this is the original work Wigner [37].

We note that the indices for the dimensions in $\mathbb{C}^{2l+1}$ are $-l, -l+1, \ldots, l-1, l$ by general convention.

6.4.2. The Peter-Weyl Theorem for $L^2_{\mathbb{C}}(S^2)$ as a Representation of $\mathrm{SO}(3)$

Here, we describe how $L^2_{\mathbb{C}}(S^2)$, considered as a unitary representation via $\lambda : \mathrm{SO}(3) \to \mathrm{U}(L^2_{\mathbb{C}}(S^2))$, with $[\lambda(g)\varphi](x) = \varphi(g^{-1}x)$, contains densely a direct sum of irreducible representations. For doing so, we proceed by first describing spherical harmonics without formulas and stating their orthonormality properties, and then stating how they transform under rotation. This will then yield the result. Note that we do not need to describe explicit formulas for the spherical harmonics, which are again somewhat complicated since we are more interested in their properties in relation to Hilbert space theory and representation theory. A reference for all this is MacRobert [38].

The spherical harmonics are continuous functions $Y_l^n : S^2 \to \mathbb{C}$ for $l \in \mathbb{N}_{\geq 0}$ and $n = -l, \ldots, l$. Thus, they are elements of $L^2_{\mathbb{C}}(S^2)$. They have the following properties:

[3] Here, the letter "D" stands for "Darstellung" which is the German term for "representation".

1. $\langle Y_l^n | Y_{l'}^{n'} \rangle = \delta_{ll'}\delta_{nn'}$ for all l, l', n, n'.

2. The linear span of the spherical harmonics is dense in $L^2_{\mathbb{C}}(S^2)$.

3. They transform as follows under rotation: $\lambda(g)(Y_l^n) = \sum_{n'=-l}^{l} D_l^{n'n}(g) Y_l^{n'}$, where $D_l^{n'n}(g)$ are the matrix elements of the Wigner D-matrices defined in Section 6.4.1.

Properties 1 and 2 together imply that the spherical harmonics form an orthonormal basis of $L^2_{\mathbb{C}}(S^2)$, see Definition A.2.9. Let

$$V_{l1} \coloneqq \operatorname{span}_{\mathbb{C}}(Y_l^n \mid n = -l, \ldots, l).$$

Then we already obtain $L^2_{\mathbb{C}}(S^2) = \widehat{\bigoplus}_{l \geq 0} V_{l1}$. Now, let $e^n \in \mathbb{C}^{2l+1}$ be the n'th standard basis vector, for $n = -l, \ldots, l$. Then property 3 means that the linear map given on basis vectors by

$$f : V_l \to V_{l1},\ e^n \mapsto Y_l^n$$

is an isomorphism of unitary representations. More precisely, f is clearly a unitary transformation and a linear isomorphism, and it is furthermore equivariant on basis vectors since

$$\begin{aligned}
f\left(D_l(g)(e^n)\right) &= f\left(\sum\nolimits_{n'=-l}^{l} D_l^{n'n}(g) e^{n'}\right) \\
&= \sum\nolimits_{n'=-l}^{l} D_l^{n'n}(g) f(e^{n'}) \\
&= \sum\nolimits_{n'=-l}^{l} D_l^{n'n}(g) Y_l^{n'} \\
&= \lambda(g)(Y_l^n) \\
&= \lambda(g)(f(e^n)).
\end{aligned} \tag{6.2}$$

General equivariance then follows from equivariance on basis vectors. This concludes this section.

6.4.3. The Clebsch-Gordan Decomposition

Explicit formulas for the Clebsch-Gordan coefficients of $\mathrm{SO}(3)$ are given in Bohm and Loewe [39]. The most important fact is the following: There is a decomposition

$$V_j \otimes V_l \cong \bigoplus_{J=|l-j|}^{l+j} V_J$$

of representations. Furthermore, the Clebsch-Gordan coefficients $\langle JM | jmln \rangle$ are all real numbers, a fact that we will use in Section 6.5.

6.4.4. Endomorphisms of V_J

As in the case of harmonic networks, this is again very simple: we are considering representations over $\mathbb{C}$, and so Schur's Lemma 4.1.8 tells us that $\mathrm{End}_{\mathrm{SO}(3)}(V_J)$ is 1-dimensional for each irrep J. We can therefore ignore the endomorphisms once again.

6.4.5. Bringing Everything Together

Now, with all this prior work, let us determine the equivariant kernels $K : S^2 \to \mathrm{Hom}_{\mathbb{C}}(V_l, V_J)$ for the irreducible representations $D_l : \mathrm{SO}(3) \to \mathrm{U}(V_l)$ and $D_J : \mathrm{SO}(3) \to \mathrm{U}(V_J)$. For this, we use Equation 4.5. Since each V_j appears only once in the direct sum decomposition of $L^2_{\mathbb{C}}(S^2)$ according to Section 6.4.2 and since V_J can only appear once in the direct sum decomposition of a tensor product $V_j \otimes V_l$ according to Section 6.4.3 , we do not need the indices i and s. Furthermore, as mentioned in the last section, the endomorphisms are trivial, which is why we also do not need the index r. Overall, we see that we simply have basis kernels $K_j : S^2 \to \mathrm{Hom}_{\mathbb{C}}(V_l, V_J)$ for all j with $|l - J| \leq j \leq l + J$[4]. They are explicitly given by

$$K_j(x) = \begin{pmatrix} \langle Y_j | x \rangle \cdot \mathrm{CG}^1_{J(jl)} \\ \vdots \\ \langle Y_j | x \rangle \cdot \mathrm{CG}^{[J]}_{J(jl)} \end{pmatrix}$$

for all $x \in S^2$. Remembering that $\langle Y_j^m | x \rangle = \overline{Y_j^m(x)}$, the individual matrix elements of $K_j(x)$ are then given by

$$\langle JM | K_j(x) | ln \rangle = \sum_{m=-j}^{j} \langle JM | jmln \rangle \cdot \overline{Y_j^m}(x).$$

This ends the discussion.

6.5. $\mathrm{SO}(3)$-Equivariant Kernels for Real Representations

In this section, we want to argue why the results in the last section transfer over to the real case as well. Most of the investigations in this section are probably well-known. However, we were not able to find sources that explicitly explain the representation theory of $\mathrm{SO}(3)$ over the real numbers, and so we develop lots of it here from scratch. We thereby make use of the theory over $\mathbb{C}$, some results about real spherical harmonics, and the general theory of real and quaternionic representations outlined in Bröcker and Dieck [40]. We need to somewhat turn the order around in this section in order

[4] We saw that V_J is a direct summand of $V_j \otimes V_l$ if and only if $|l - j| \leq J \leq l + j$. By doing case distinctions, one can show that this is the case if and only if $|l - J| \leq j \leq l + J$.

to develop the results. Therefore we first investigate the Peter-Weyl Theorem, then look at the endomorphism spaces of the appearing irreducible representations and afterward, as a consequence, show that the representations appearing in the decomposition of $L^2_{\mathbb{R}}(S^2)$ are already exhaustive.

6.5.1. The Peter-Weyl Theorem for $L^2_{\mathbb{R}}(S^2)$ as a Representation of $\mathrm{SO}(3)$

The most important finding is the following, which is taken from Gallier and Quaintance [41]: One can do a base change for the spherical harmonics as follows to obtain real versions of them. Namely, let

$$ {}^rY_l^n = \begin{cases} \frac{i}{\sqrt{2}}\left(Y_l^n - (-1)^n\, Y_l^{-n}\right) & \text{if } n < 0, \\ Y_l^0 & \text{if } n = 0, \\ \frac{1}{\sqrt{2}}\left(Y_l^{-n} + (-1)^n\, Y_l^n\right) & \text{if } n > 0. \end{cases} \tag{6.3} $$

One can then show that these functions are real-valued continuous functions and therefore ${}^rY_l^n \in L^2_{\mathbb{R}}(S^2)$. Furthermore, they are an orthonormal basis of this space. We can then, as before, set ${}^rV_{l1}$ as the span of the ${}^rY_l^n \in L^2_{\mathbb{R}}(S^2)$ and obtain a decomposition

$$ L^2_{\mathbb{R}}(S^2) = \widehat{\bigoplus_{l\geq 0}} {}^rV_{l1}. $$

We need to understand the transformation properties of these real-valued spherical harmonics under rotation. To understand this explicitly, we set $B_l \in \mathbb{C}^{(2l+1)\times(2l+1)}$ as the (complex) base change matrix between the complex and real spherical harmonics. Its entries are given according to Equation 6.3 such that the following relation holds for all $n = -l, \dots, l$:

$$ {}^rY_l^n = \sum_{n'=-l}^{l} B_l^{n'n} \cdot Y_l^{n'}. $$

Since for a given l, both the complex and real spherical harmonics are linearly independent, the matrix B_l is invertible. Let B_l^{-1} be its inverse. Then it is generally known from linear algebra that we also obtain the inverse relation:

$$ Y_l^n = \sum_{n'=-l}^{l} (B_l^{-1})^{n'n} \cdot {}^rY_l^{n'}. $$

Using both these relations and the rotation properties of the complex spherical harmonics from Section 6.4.2 we obtain the following rotation property for the real spherical harmonics:

$$ \lambda(g)({}^rY_l^n) = \sum_{n_1=-l}^{l} B_l^{n_1 n} \cdot \lambda(g)(Y_l^{n_1}) $$

$$
\begin{aligned}
&= \sum_{n_1=-l}^{l} B_l^{n_1 n} \sum_{n_2=-l}^{l} D_l^{n_2 n_1}(g) \cdot Y_l^{n_2} \\
&= \sum_{n_1=-l}^{l} B_l^{n_1 n} \sum_{n_2=-l}^{l} D_l^{n_2 n_1}(g) \cdot \sum_{n'=-l}^{l} (B_l^{-1})^{n' n_2} \cdot {}^r Y_l^{n'} \\
&= \sum_{n'=-l}^{l} \left(\sum_{n_1=-l}^{l} \sum_{n_2=-l}^{l} (B_l^{-1})^{n' n_2} \cdot D_l^{n_2 n_1}(g) \cdot B_l^{n_1 n} \right) {}^r Y_l^{n'} \\
&= \sum_{n'=-l}^{l} \left(B_l^{-1} \cdot D_l(g) \cdot B_l \right)^{n' n} \cdot {}^r Y_l^{n'}.
\end{aligned}
$$

Now if we set ${}^r D_l(g) := B_l^{-1} \cdot D_l(g) \cdot B_l$, then we obtain the transformation property

$$
\lambda(g)({}^r Y_l^n) = \sum_{n'=-l}^{l} {}^r D_l(g)^{n' n} \cdot {}^r Y_l^{n'} \tag{6.4}
$$

which is analogous to the one in Section 6.4.2.

Lemma 6.5.1. ${}^r D_l(g)^{n' n} \in \mathbb{R}$ *for all* $l \geq 0$, $n', n = -l, \dots, l$ *and* $g \in \mathrm{SO}(3)$.

Proof. Note that since ${}^r Y_l^n$ is a real-valued function, the rotation $\lambda(g)({}^r Y_l^n)$ is real-valued as well. Thus, it is in the space $L^2_{\mathbb{R}}(S^2)$. The real spherical harmonics are a basis of this space, which means that the coefficients when expanding $\lambda(g)({}^r Y_l^n)$ in this basis are necessarily real as well. These coefficients are precisely given by the ${}^r D_l(g)^{n' n}$ according to Equation 6.4. □

Now, we have the choice to view ${}^r D_l$ as either a real or a complex representation, but first we take the complex viewpoint and see it as a function ${}^r D_l : \mathrm{SO}(3) \to \mathrm{Aut}_{\mathbb{C}}(\mathbb{C}^{2l+1})$. Notationwise, the following is important: the "r" in ${}^r D_l$ indicates that the elements in this matrix are real but does not tell us on which space it acts. This will always be clarified by the context. We have the following:

Lemma 6.5.2. ${}^r D_l : \mathrm{SO}(3) \to \mathrm{U}(\mathbb{C}^{2l+1})$ *is an irreducible unitary representation and isomorphic to* D_l.

Proof. First of all, it is an actual linear representation since

$$
{}^r D_l(gg') = B_l^{-1} D_l(gg') B_l = B_l^{-1} D_l(g) B_l B_l^{-1} D_l(g') B_l = {}^r D_l(g) \cdot {}^r D_l(g')
$$

where we used that D_l is a linear representation. Now since Y_l^n and ${}^r Y_l^n$ are both orthonormal bases of $L^2_{\mathbb{C}}(S^2)$, the base change matrix B_l needs to be a unitary matrix. Consequently, ${}^r D_l(g) = B_l^{-1} D_l(g) B_l$ is as a product of unitary transformations itself unitary, which means that ${}^r D_l$ is a unitary representation. Furthermore, we obtain $B_l \cdot {}^r D_l(g) = D_l(g) \cdot B_l$, which means that B_l gives an isomorphism ${}^r D_l \cong D_l$ of unitary representations. From the fact that D_l is irreducible, we obtain that ${}^r D_l$ is irreducible as well. □

Now we take the real viewpoint. Let ${}^rV_l = \mathbb{R}^{2l+1}$.

Lemma 6.5.3. ${}^rD_l : \mathrm{SO}(3) \to \mathrm{O}({}^rV_l)$ *is an irreducible orthogonal representation.*

Proof. ${}^rD_l(g)$ is a unitary matrix for each $g \in \mathrm{SO}(3)$ by Lemma 6.5.2, and since its matrix elements are real by Lemma 6.5.1, it automatically is an orthogonal matrix. If it was reducible, then there would be a *real* base change matrix that brings ${}^rD_l$ in a nontrivial blog-diagonal shape. However, this base change would in particular be complex, meaning that we would conclude that the complex version of the representation ${}^rD_l$ is reducible. But it is not, due to Lemma 6.5.2. □

Now, remember that $L^2_{\mathbb{R}}(S^2) = \widehat{\bigoplus}_{l\geq 0} {}^rV_{l1}$ and that ${}^rV_{l1}$ is generated from the real spherical harmonics. Also, remember that the real spherical harmonics transform as in Equation 6.4. Thus, with the same arguments as in Equation 6.2 we obtain ${}^rV_{l1} \cong {}^rV_l$, which is from the preceding lemmas an irreducible orthogonal representation. Thus, we have found the Peter-Weyl decomposition of $L^2_{\mathbb{R}}(S^2)$.

6.5.2. Endomorphisms of ${}^rV_J$

In the next section, we will show that the ${}^rD_J : \mathrm{SO}(3) \to \mathrm{O}({}^rV_J)$ already given an exhaustive list of the irreducible representations of $\mathrm{SO}(3)$ over the real numbers. In this section, we first describe their endomorphism spaces since this will help in showing that there cannot be any other irreducible representations. Fortunately, the situation is again very simple:

Proposition 6.5.4. $\mathrm{End}_{\mathrm{SO}(3),\mathbb{R}}({}^rV_J)$ *is one-dimensional for each* $J \geq 0$.

Proof. Let $f : {}^rV_J \to {}^rV_J$ be an endomorphism. Since ${}^rV_J = \mathbb{R}^{2J+1}$ we can view f as a matrix in $\mathbb{R}^{(2J+1)\times(2J+1)}$. That f is an endomorphism then means

$$f \cdot {}^rD_J(g) = {}^rD_J(g) \cdot f$$

for all $g \in \mathrm{SO}(3)$. Now note that as a real matrix, f is in particular a complex matrix, i.e. $f \in \mathbb{C}^{(2J+1)\times(2J+1)}$. Also, remember that we can view ${}^rD_J$ also as a complex irreducible representation ${}^rD_J : \mathrm{SO}(3) \to \mathrm{U}(\mathbb{C}^{2J+1})$ by Lemma 6.5.2. What this means is that $f \in \mathrm{End}_{\mathrm{SO}(3),\mathbb{C}}(\mathbb{C}^{2J+1})$, which is isomorphic to $\mathbb{C}$ by Schur's Lemma 4.1.8. Thus, f is a complex multiple of the identity. Since f is a real matrix, it is thus a real multiple of the identity. The result follows. □

6.5.3. General Notes on the Relation between Real and Complex Representations

In the next section we show that there can, up to isomorphism, not be *other* irreducible representations than the ${}^rD_l : \mathrm{SO}(3) \to \mathrm{O}({}^rV_l)$. In order to do so, we first need to better understand the relationship between real and complex representations of

compact groups. These investigations will carry over to the investigations for $\mathrm{O}(3)$ that we do in Section 6.6 as well.
The following definition of a classification of real irreducible representations of a compact group G can be found in Bröcker and Dieck [40], Theorem II.6.7. In this book, it is a theorem, since the authors give an independent but equivalent definition of these notions.

Definition 6.5.5 (Real type, complex type and quaternionic type irreducible representations)**.** Let $\rho : G \to \mathrm{O}(V)$ be a *real* irreducible representation of a compact group G. Then ρ is said to be of

1. *real type* if $\mathrm{End}_{G,\mathbb{R}}(V) \cong \mathbb{R}$,
2. *complex type* if $\mathrm{End}_{G,\mathbb{R}}(V) \cong \mathbb{C}$ and
3. *quaternionic type* if $\mathrm{End}_{G,\mathbb{R}}(V) \cong \mathbb{H}$, where $\mathbb{H}$ are the quaternions.

Here, these isomorphisms respect both addition and multiplication. The multiplication in the endomorphism spaces is thereby given by composition of functions.

Furthermore, Bröcker and Dieck [40] shows in Theorem II.6.3 that there is no other possibility for an irreducible real representation, i.e. they can be completely categorized by being of real, complex or quaternionic type. Additionally, since $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{H}$ already differ in their $\mathbb{R}$-dimension, it is enough to check whether the $\mathbb{R}$-dimension of an endomorphism space is 1, 2 or 4 in order to do the classification.
In order to compare real and complex representations we need to define two functors between those[5]:

Definition 6.5.6 (Restriction and Extension)**.** Let ${}^c\rho : G \to \mathrm{Aut}_{\mathbb{C}}({}^cV)$ be a complex representation. Furthermore, let ${}^r\rho : G \to \mathrm{Aut}_{\mathbb{R}}({}^rV)$ be a real representation. Then we define their *restriction* and *extension* as follows:

1. Set $r({}^cV)$ as the $\mathbb{R}$-vector space that has the same underlying abelian group as ${}^cV$ and the scalar multiplication from $\mathbb{R}$ which is the restriction of the multiplication from $\mathbb{C}$. The *restriction* $r({}^c\rho) : G \to \mathrm{Aut}_{\mathbb{R}}(r({}^cV))$ is defined as the exact same map as ${}^c\rho$, only that $r({}^c\rho)(g) : r({}^cV) \to r({}^cV)$ is now viewed as an automorphism of *real* vector spaces.
2. We define the *extension* by $e({}^rV) := \mathbb{C} \otimes_{\mathbb{R}} {}^rV$, where $\mathbb{C}$ is regarded as an $\mathbb{R}$-vector space. This construction becomes a $\mathbb{C}$-vector space by scalar multiplication $z \cdot (z' \otimes v) := (zz') \otimes v$. We can then define $e({}^r\rho) : G \to \mathrm{Aut}_{\mathbb{C}}(e({}^rV))$ by setting $e({}^r\rho)(g) := \mathrm{id}_{\mathbb{C}} \otimes ({}^r\rho(g))$.

[5] We only define these functors on objects and not on morphisms. The reason is that we will never explicitly use their definitions on morphisms. More details on this can be found in Bröcker and Dieck [40], including other functors which are needed in the general theory. The reader should not worry if he or she does not know what a functor is.

Note that the extension operation doubles the $\mathbb{R}$-dimension, whereas for the restriction it stays equal. Therefore, we can not hope that these operations are inverse to each other. However, we have the following, almost as nice statement:

Proposition 6.5.7. *For each real representation $\rho : G \to \mathrm{Aut}_{\mathbb{R}}(V)$ there is a natural isomorphism $r(e(V)) \cong V \oplus V$ of $\mathbb{R}$-representations.*

Proof. This is the first statement in Bröcker and Dieck [40], Proposition II.6.1. □

The following definition is actually not the definition that Bröcker and Dieck [40] formulate. However, it is an equivalent characterization that follows from their Proposition II.6.6 (vii), (viii) and (ix) and is more convenient for our needs:

Definition 6.5.8 (Real type *complex* representation)**.** Let $\rho : G \to \mathrm{Aut}_{\mathbb{C}}(V)$ be a *complex* irreducible representation. Then ρ is called *of real type* if there is an isomorphism of real representations $r(V) \cong U \oplus U$ where

1. $\rho_U : G \to \mathrm{Aut}_{\mathbb{R}}(U)$ is an irreducible real representation and
2. $r(\rho) : G \to \mathrm{Aut}_{\mathbb{R}}(r(V))$ is the restriction of ρ, as defined in Definition 6.5.6.

Proposition 6.5.9. *Assume G is a compact group such that all complex irreducible representations are of real type. Then also all real irreducible representations are of real type.*

Proof. This follows from Bröcker and Dieck [40], Proposition II.6.6 (ii) and (iii). □

Proposition 6.5.10. *Let $\rho : G \to \mathrm{Aut}_{\mathbb{R}}(V)$ be an irreducible real representation of real type. Then its extension $e(\rho) : G \to \mathrm{Aut}_{\mathbb{C}}(e(V))$ given as in Definition 6.5.6 is an irreducible complex representation (also of real type).*

Proof. This is precisely Bröcker and Dieck [40], Proposition II.6.6(i). □

6.5.4. The Irreducible Representations of $\mathrm{SO}(3)$ over the Real Numbers

The rough strategy is to use the fact that the ${}^{r}D_l$, viewed as complex irreducible representations, are an exhaustive list of all the complex irreps. Then, using the restriction and extension operators r and e between real and complex representations, we can show that in the specific case of $\mathrm{SO}(3)$, there can not be any other real irreducible representations than the ${}^{r}D_l$, viewed as *real* representations.

Lemma 6.5.11. *All complex irreducible representations of $\mathrm{SO}(3)$ are of real type.*

Proof. From Section 6.4.1 and Lemma 6.5.2 we know that the ${}^{r}D_l : \mathrm{SO}(3) \to \mathrm{U}(\mathbb{C}^{2l+1})$ give us, up to equivalence, all the complex irreducible representations of $\mathrm{SO}(3)$. According to Definition 6.5.8 we now need to understand that its restriction splits into the direct sum of twice the same irreducible *real* representation. We do this as follows:

We can write $r(\mathbb{C}^{2l+1}) = \mathbb{R}^{2l+1} \oplus (i\mathbb{R})^{2l+1} = {}^rV_l \oplus i\,{}^rV_l$, which is a decomposition of $\mathbb{C}^{2l+1}$ when viewed as an $\mathbb{R}$-vector space. Then, we can note that both

$$\begin{aligned} {}^rD_l &: \mathrm{SO}(3) \to \mathrm{O}({}^rV_l) \text{ and} \\ {}^rD_l &: \mathrm{SO}(3) \to \mathrm{O}(i\,{}^rV_l) \end{aligned}$$

are well-defined $\mathbb{R}$-representations, which follows from the fact that the matrix elements are all real. Furthermore, the first map is actually an irreducible real representation by Lemma 6.5.3. The second one is isomorphic to the first since one can show that

$$i : {}^rV_l \to i\,{}^rV_l, \; a \mapsto i \cdot a$$

is an isomorphism of real $\mathrm{SO}(3)$-representations. This gives us precisely the splitting of $r(\mathbb{C}^{2l+1})$ as a representation that we were looking for. ☐

Corollary 6.5.12. *All irreducible real representations of $\mathrm{SO}(3)$ are of real type.*

Proof. This follows directly from Lemma 6.5.11 and Proposition 6.5.9. ☐

Proposition 6.5.13. *The ${}^rD_l : \mathrm{SO}(3) \to \mathrm{O}({}^rV_l)$ are, up to equivalence, all real irreducible representations of $\mathrm{SO}(3)$.*

Proof. Assume that $\rho : \mathrm{SO}(3) \to \mathrm{Aut}_{\mathbb{R}}(V)$ is an irreducible real representation of $\mathrm{SO}(3)$. It is of real type by Corollary 6.5.12. By Proposition 6.5.10, the extension $e(\rho) : G \to \mathrm{Aut}_{\mathbb{C}}(e(V))$ is an irreducible complex representation. Since the ${}^rD_l$ give us all complex irreducible representations up to equivalence by Section 6.4.1 and Lemma 6.5.2, there is an equivalence of complex $\mathrm{SO}(3)$-representations $e(V) \cong \mathbb{C}^{2l+1}$ for some l. Since functors respect isomorphisms (and equivalences are isomorphisms in the categories of G-representations) and the restriction operation is a functor[6], and using Proposition 6.5.7 as well as the proof of Lemma 6.5.11 we obtain:

$$V \oplus V \cong r(e(V)) \cong r(\mathbb{C}^{2l+1}) \cong {}^rV_l \oplus i\,{}^rV_l = {}^rV_l \oplus {}^rV_l.$$

Using the Krull-Remak-Schmidt Theorem 2.2.16, we see that there is an isomorphism of $\mathrm{SO}(3)$-representations $V \cong {}^rV_l$. This finishes the proof. ☐

6.5.5. The Clebsch-Gordan Decomposition

We are almost there. The only thing left to understand is the Clebsch-Gordan decomposition. Remember the following from Section 6.4.3: For the *complex* irreducible representations there are decompositions

$$V_j \otimes V_l \cong \bigoplus_{J=|l-j|}^{l+j} V_J$$

[6] The reader does not need to know what a functor is if he or she believes these statements.

where on each space, the representations D_j, D_l and D_J are given by the Wigner D-matrices. Furthermore, the Clebsch-Gordan coefficients are all real. Now, we know that ${}^rD_l$ is, as a complex representation, isomorphic to D_l by Lemma 6.5.2, and such a representation then acts on $\mathbb{C}^{2l+1}$ as well. Consequently, we also get the decomposition

$$\mathbb{C}^{2j+1} \otimes \mathbb{C}^{2l+1} \cong \bigoplus_{J=|l-j|}^{l+j} \mathbb{C}^{2J+1}$$

of the complex representations ${}^rD_j$ and ${}^rD_l$. Obviously, the Clebsch-Gordan coefficients can be chosen to be exactly the same as before, and thus they are again real. Let the above isomorphism be called f. Now, we can view all involved vector spaces as $\mathbb{R}$-vector spaces as well. Furthermore, we have subspaces ${}^rV_j = \mathbb{R}^{2j+1}$, ${}^rV_l = \mathbb{R}^{2l+1}$ and ${}^rV_J = \mathbb{R}^{2J+1}$ which are also invariant under the representations ${}^rD_j$, ${}^rD_l$ and ${}^rD_J$. Consequently, we can just restrict the isomorphism above to a map

$$f| : {}^rV_j \otimes {}^rV_l \to \bigoplus_{J=|l-j|}^{l+j} {}^rV_J.$$

which is well-defined since the Clebsch-Gordan coefficients are real. It needs to be injective, since it is a restriction of an isomorphism. For dimension reasons, the restriction then needs to be an isomorphism, and obviously, it has the exact same Clebsch-Gordan coefficients as the original map f[7].

6.5.6. Bringing Everything Together

By what we've shown in the last sections, we see that the situation is basically the same as in Section 6.4.5. The only thing that changes is that we now use the *real* spherical harmonics, and therefore the complex conjugation disappears. What this overall means is the following: let ${}^rD_l : \mathrm{SO}(3) \to \mathrm{O}({}^rV_l)$ and ${}^rD_J : \mathrm{SO}(3) \to \mathrm{O}({}^rV_J)$ be the representations determining the input and output fields. Then a basis for steerable kernels $K : S^2 \to \mathrm{Hom}_{\mathbb{R}}({}^rV_l, {}^rV_J)$ is given by kernels $K_j : S^2 \to \mathrm{Hom}_{\mathbb{R}}({}^rV_l, {}^rV_J)$ for all $|l-J| \leq j \leq l+J$. The matrix elements are given by

$$\langle JM|K_j(x)|ln\rangle = \sum_{m=-j}^{j} \langle JM|jmln\rangle \cdot {}^rY_j^m(x). \tag{6.5}$$

6.6. $\mathrm{O}(3)$-Equivariant Kernels for Complex Representations

In this section, we deal with $\mathrm{O}(3)$-equivariant kernels for complex representations and then, in the next section, will transport the results over to real representations.

[7]The reason for this is that the standard basis vectors in $\mathbb{C}^k$ which are used for the Clebsch-Gordan coefficients are exactly the standard basis vectors in $\mathbb{R}^k \subseteq \mathbb{C}^k$ by definition of this embedding.

In the earlier examples, we saw that the Peter-Weyl decomposition of $L^2_{\mathbb{K}}(X)$ always contained each irreducible representation of the symmetry group exactly once. The example of $\mathrm{O}(3)$ is the first in which this is not the case: *parity* will play a role in determining which irreducible representations make their way in the space of square-integrable functions and which do not. Overall, we hope that the example of $\mathrm{O}(3)$ is a sufficient justification for our use of the multiplicities m_j of irreducible representations that we considered in all our theorems. $\mathrm{O}(3)$-equivariant networks are to the best of our knowledge not described in any published work yet.

6.6.1. The Irreducible Representations of $\mathrm{O}(3)$

The most important observation is the following, after which we can deduce the irreducible representations of $\mathrm{O}(3)$ from those of $\mathrm{SO}(3)$:

Lemma 6.6.1. *Let $\mathbb{Z}_2 := (\{-1, +1\}, \cdot)$ be the group with two elements. Then the map*

$$\cdot : \mathbb{Z}_2 \times \mathrm{SO}(3) \to \mathrm{O}(3), \quad (s, g) \mapsto sg$$

is an isomorphism of groups.

Proof. It is a group homomorphism since $s \in \{-1, +1\}$ can be represented by a multiple of the identity matrix, and as such it commutes with every matrix g. That $\cdot$ is an isomorphism follows since all matrices in $\mathrm{O}(3)$ either have determinant 1 or -1. The matrices with determinant 1 form $\mathrm{SO}(3)$ and are the image of $\{+1\} \times \mathrm{SO}(3)$. The matrices with determinant -1 are the image of $\{-1\} \times \mathrm{SO}(3)$. □

Note the fact that for $g \in \mathrm{SO}(3)$, $-g$ has determinant -1, which we used in the proof. This does only hold for $g \in \mathrm{SO}(n)$ with n being *odd*. Therefore, the above lemma is not true for n even. In the even case, we obtain a *semidirect* product and the story complicates somewhat.
Earlier, we already considered tensor product representations of one and the same group. A related notion is that of tensor product representations of two different groups[8]:

Definition 6.6.2 (Tensor product representation)**.** Let G and H be two compact groups. Let $\rho_G : G \to \mathrm{Aut}_{\mathbb{K}}(V_G)$ and $\rho_H : H \to \mathrm{Aut}_{\mathbb{K}}(V_H)$ be representations of the two groups G and H. Then the *tensor product representation* is given by

$$\begin{gathered}\rho_G \otimes \rho_H : G \times H \to \mathrm{Aut}_{\mathbb{K}}(V_G \otimes V_H), \\ \big[\,(\rho_G \otimes \rho_H)\,(g,h)\big](v_G \otimes v_H) := \rho_G(g)(v_G) \otimes \rho_H(h)(v_H).\end{gathered}$$

This is again a linear representation.

Proposition 6.6.3. *Representatives of isomorphism classes of irreducible representations of $G \times H$ are given precisely by all the $\rho_G \otimes \rho_H$, where ρ_G and ρ_H run through representatives of isomorphism classes of irreducible representations of G and H, respectively.*

[8] It is not a direct generalization due to the presence of two different group elements being applied.

Proof. This is proven in chapter II, Proposition 4.14 and 4.15 of Bröcker and Dieck [40]. □

It is important to note that the proof of the above proposition uses the property of the complex numbers to be algebraically closed in crucial steps, and therefore it is unclear how exactly a generalization to representations over the real numbers looks like. Therefore, we will not use the above proposition in our later considerations for real representations of $\mathrm{O}(3)$.
However, in our current situation, we can apply it without problems. This proposition, together with Lemma 6.6.1, suggests that we should understand the irreducible representations of $\mathbb{Z}_2$. We already saw this for real representations before and essentially obtain the same result:

Lemma 6.6.4. *The irreducible representations of $\mathbb{Z}_2$ are up to equivalence precisely the following two, which we state for simplicity only on the generator:*

$$\begin{aligned}\rho_+ &: \mathbb{Z}_2 \to \mathrm{Aut}_{\mathbb{C}}(\mathbb{C}), \quad \rho_+(-1) = \mathrm{id}_{\mathbb{C}} \\ \rho_- &: \mathbb{Z}_2 \to \mathrm{Aut}_{\mathbb{C}}(\mathbb{C}), \quad \rho_-(-1) = -\,\mathrm{id}_{\mathbb{C}}\,.\end{aligned}$$

Proof. This can be shown in exactly the same way as in Section 6.3.1. □

Thus we are ready to state our result about the irreducible representations of $\mathrm{O}(3)$:

Proposition 6.6.5. *The irreducible representations of $\mathrm{O}(3)$ are up to equivalence given as follows: for each $l \in \mathbb{N}_{\geq 0}$ there are precisely two representations $D_{l+} : \mathrm{O}(3) \to \mathrm{U}(V_{l+})$ and $D_{l-} : \mathrm{O}(3) \to \mathrm{U}(V_{l-})$ with $V_{l+} = \mathbb{C}^{2l+1} = V_{l-}$, given as follows:*

$$\begin{aligned}D_{l+}(sg) &= D_l(g) \quad \textit{for all } s \in \mathbb{Z}_2,\ g \in \mathrm{SO}(3). \\ D_{l-}(sg) &= sD_l(g) \ \textit{for all } s \in \mathbb{Z}_2,\ g \in \mathrm{SO}(3).\end{aligned}$$

Proof. Remember from Section 6.4.1 that the irreducible representations of $\mathrm{SO}(3)$ are given by the Wigner D-matrices D_l. From Lemma 6.6.4 we know that the irreducible representations of $\mathbb{Z}_2$ are given by ρ_+ and ρ_-. From the isomorphism $\mathrm{O}(3) \cong \mathbb{Z}_2 \times \mathrm{SO}(3)$ from Lemma 6.6.1 and from Proposition 6.6.3 we thus obtain that the irreducible representations of $\mathrm{O}(3)$ are precisely given by all $\rho_+ \otimes D_l$ and $\rho_- \otimes D_l$. We now show that $\rho_- \otimes D_l$ is equivalent to D_{l-}: We have

$$\rho_- \otimes D_l : \mathrm{O}(3) \to \mathrm{Aut}_{\mathbb{C}}(\mathbb{C} \otimes V_l), \ \big[(\rho_- \otimes D_l)(sg)\big](z \otimes v) = sz \otimes \left[D_l(g)\right](v).$$

Now, consider the linear isomorphism $f : \mathbb{C} \otimes V_l \to V_{l+}$, $z \otimes v \mapsto zv$. We only need to check that it is equivariant and are then done:

$$\begin{aligned}f\big(\left[(\rho_- \otimes D_l)(sg)\right](z \otimes v)\big) &= f\big(sz \otimes [D_l(g)](v)\big) \\ &= sz \cdot \left[D_l(g)\right](v) \\ &= [sD_l(g)](zv) \\ &= [D_{l-}(sg)](f(z \otimes v)).\end{aligned}$$

The statement about D_{l+} can be shown using the exact same map f. □

6.6.2. The Peter-Weyl Theorem for $L^2_{\mathbb{C}}(S^2)$ as Representation of O(3)

The considerations in this section follow almost entirely from Section 6.4.2. There we saw that, as a representation over SO(3), we have a decomposition

$$L^2_{\mathbb{C}}(S^2) = \widehat{\bigoplus_{l \geq 0}} V_{l1}$$

with the spaces V_{l1} being spanned by the spherical harmonics Y_l^n, $n = -l, \dots, l$. We immediately see that in $L^2_{\mathbb{C}}(S^2)$, viewed as a representation over O(3), there is *not enough space* for all the irreducible representations, since they appear in pairs as shown in Proposition 6.6.5[9]. Thus, we need to figure out which irreducible representations are present and which are not. The core of this question is answered by the following proposition:

Lemma 6.6.6 (Parity in spherical harmonics)**.** *The spherical harmonics obey the following parity rules:*

$$Y_l^n(sx) = s^l \cdot Y_l^n(x)$$

for all $l \geq 0$, $n = -l, \dots, l$, $s \in \mathbb{Z}_2$ *and* $x \in S^2$.

Proof. This is a well-known property of the spherical harmonics. $\square$

Thus, together with Section 6.4.2 we get the following transformation behavior of spherical harmonics under the group O(3), where $s \in \mathbb{Z}_2$ and $g \in \mathrm{SO}(3)$:

$$\begin{aligned}
\lambda(sg)(Y_l^n) &= s^l \lambda(g)(Y_l^n) \\
&= s^l \sum_{n'=-l}^{l} D_l^{n'n}(g) Y_l^{n'} \\
&= \sum_{n'=-l}^{l} \left(s^l D_l^{n'n}(g) \right) Y_l^{n'} \\
&= \begin{cases} \sum_{n'=-l}^{l} D_{l+}^{n'n}(sg) Y_l^{n'}, & l \text{ even} \\ \sum_{n'=-l}^{l} D_{l-}^{n'n}(sg) Y_l^{n'}, & l \text{ odd.} \end{cases}
\end{aligned}$$

Thus, we obtain the following decomposition of $L^2_{\mathbb{C}}(S^2)$:

$$L^2_{\mathbb{C}}(S^2) = \widehat{\bigoplus_{\substack{l \geq 0 \\ l \text{ even}}}} V_{l1+} \oplus \widehat{\bigoplus_{\substack{l \geq 0 \\ l \text{ odd}}}} V_{l1-}.$$

Here, V_{l1+} and V_{l1-} are generated from the spherical harmonics of order l and we have $V_{l1+} \cong V_{l+}$ and $V_{l1-} \cong V_{l-}$ as representations according to the transformation behavior we saw above.

[9] With this, we mean the following: the irreducible representations of SO(3) already cover $L^2_{\mathbb{C}}(S^2)$. O(3) has *even more* irreducible representations than SO(3), so it is a priori clear that they cannot all fit into $L^2_{\mathbb{C}}(S^2)$.

6.6.3. The Clebsch-Gordan Decomposition

Remember from Section 6.4.3 that we have a decomposition of $\mathrm{SO}(3)$-representations

$$V_j \otimes V_l \cong \bigoplus_{J=|l-j|}^{l+j} V_J$$

given by real Clebsch-Gordan coefficients. Now for $\mathrm{O}(3)$, remember that as vector spaces we have for all j (and equally for l and J) equalities $V_j = V_{j-} = V_{j+}$, and so we guess that in the isomorphism above, we just need to figure out the correct signs in order to be compatible with the corresponding representations. The idea is that "multiplying the signs at the left" should lead to the "sign at the right", and this paradigm leads us to believe that there are the following isomorphisms:

$$\begin{aligned} V_{j+} \otimes V_{l+} &\cong \bigoplus_{J=|l-j|}^{l+j} V_{J+}, \quad & V_{j+} \otimes V_{l-} &\cong \bigoplus_{J=|l-j|}^{l+j} V_{J-}, \\ V_{j-} \otimes V_{l+} &\cong \bigoplus_{J=|l-j|}^{l+j} V_{J-}, \quad & V_{j-} \otimes V_{l-} &\cong \bigoplus_{J=|l-j|}^{l+j} V_{J+}. \end{aligned}$$

We just show the lower-left isomorphism since the arguments are always the same. So, assume that $f : V_j \otimes V_l \to \bigoplus_{J=|l-j|}^{l+j} V_J$ is an isomorphism and thus in particular intertwines the given representations. Now, we take the exact same map $f : V_{j-} \otimes V_{l+} \to \bigoplus_{J=|l-j|}^{l+j} V_{J-}$ and only need to figure out that it is equivariant with respect to the given representations, using the same property for the original isomorphism we started with:

$$\begin{aligned} f \circ \left[D_{j-}(sg) \otimes D_{l+}(sg)\right] &= f \circ \left[s(D_j(g) \otimes D_l(g))\right] \\ &= s \bigoplus_{J=|l-j|}^{l+j} D_J(g) \circ f \\ &= \bigoplus_{J=|l-j|}^{l+j} D_{J-}(sg) \circ f. \end{aligned}$$

This shows the claim. From these considerations, it also follows that the Clebsch-Gordan coefficients do not in any way depend on the signs of the spaces V_j, V_l, V_J. Thus, we write them generically as $\langle JM|jmln\rangle$.

6.6.4. Endomorphisms of V_J

As always over $\mathbb{C}$, Schur's Lemma 4.1.8 shows that the endomorphism spaces are 1-dimensional, and thus we can ignore endomorphisms.

6.6.5. Bringing Everything Together

Now we can finally compute the basis for steerable kernels. The section on the Clebsch-Gordan decomposition suggests that we need to do a case distinction for this. Namely, the possible kernels depend on the signs of V_l and V_J. The results basically follow analogously to the results in Section 6.4.5.

Steerable Kernels $K : S^2 \to \mathrm{Hom}_{\mathbb{C}}(V_{l+}, V_{J+})$:

V_{J+} can only be in a tensor product $V_j \otimes V_{l+}$ if the sign of j is positive. Spaces V_{j1+} appear in the tensor product decomposition of $L^2_{\mathbb{C}}(S^2)$ precisely for even j, according to Section 6.6.2. Thus, a basis for steerable kernels is given by all K_j with *even* $j \in \big\{|l-J|, \dots, l+J\big\}$. It has matrix elements

$$\langle JM|K_j(x)|ln\rangle = \sum_{m=-j}^{j} \langle JM|jmln\rangle \cdot \overline{Y_j^m}(x),$$

exactly as in Section 6.4.5.

Steerable Kernels $K : S^2 \to \mathrm{Hom}_{\mathbb{C}}(V_{l+}, V_{J-})$:

Analogously, a basis for steerable kernels is given by all K_j, with *odd* $j \in \big\{|l - J|, \dots, l+J\big\}$.

Steerable Kernels $K : S^2 \to \mathrm{Hom}_{\mathbb{C}}(V_{l-}, V_{J+})$:

Again, a basis for steerable kernels is given by all K_j with *odd* $j \in \big\{|l-J|, \dots, l+J\big\}$.

Steerable Kernels $K : S^2 \to \mathrm{Hom}_{\mathbb{C}}(V_{l-}, V_{J-})$:

As in the first case, a basis for steerable kernels is given by all K_j with *even* $j \in \big\{|l-J|, \dots, l+J\big\}$.
Thus, we have determined all kernel bases for the group $\mathrm{O}(3)$ over the complex numbers. Compared to $\mathrm{SO}(3)$, we see that the kernel spaces get roughly halved. The reason for this is that with a bigger symmetry group, the kernel needs to obey more rules, which means that the kernel constraint has fewer solutions.

6.7. $\mathrm{O}(3)$-Equivariant Kernels for Real Representations

Basically, we can argue exactly as in Section 6.5.4 in order to transport the results for complex representations over to the real world. We shortly sketch the procedure and outcome. As we know from Section 6.3.1, $\rho_- : \mathbb{Z}_2 \to \mathrm{O}(\mathbb{R})$ and $\rho_+ : \mathbb{Z}_2 \to \mathrm{O}(\mathbb{R})$ are the only irreducible real representations of $\mathbb{Z}_2$. Thus, for each $l \geq 0$ we obtain two

irreducible real representations ${}^rD_{l+} : \mathrm{O}(3) \to \mathrm{O}({}^rV_{l+})$ and ${}^rD_{l-} : \mathrm{O}(3) \to \mathrm{O}({}^rV_{l-})$. As before, they also act on complex vector spaces and are as such isomorphic to the complex irreducible representations of $\mathrm{O}(3)$. One can then show as in Lemma 6.5.11 that all complex irreducible representations are of real type since they split into two copies of the real version of this representation. Thus, by Corollary 6.5.12, all real irreducible representations are of real type, and this means that we can proceed exactly as in Proposition 6.5.13 in order to show that the ${}^rD_{l+}$ and ${}^rD_{l-}$ are already all the irreducible real representations of $\mathrm{O}(3)$ up to equivalence.
For the Peter-Weyl decomposition of $L^2_{\mathbb{R}}(S^2)$, we only need to note that the real spherical harmonics emerge with a base change from the complex ones, as seen in Equation 6.3, and thus fulfill the same parity rules as the complex spherical harmonics. This gives us a decomposition

$$L^2_{\mathbb{R}}(S^2) = \widehat{\bigoplus_{\substack{l \geq 0 \\ l \text{ even}}}} ({}^rV_{l1+}) \oplus \widehat{\bigoplus_{\substack{l \geq 0 \\ l \text{ odd}}}} ({}^rV_{l1-}).$$

For the Clebsch-Gordan coefficients, we again get decompositions

$${}^rV_j \otimes {}^rV_l \cong \bigoplus_{J=|l-j|}^{l+j} {}^rV_J$$

where the signs on the left must “multiply to” the signs on the right, as in Section 6.6.3. Finally, the endomorphism spaces must be 1-dimensional since the endomorphism spaces of the complex versions are 1-dimensional.
Overall, we obtain the same kernels as in Section 6.6.5, only that we need to use the real spherical harmonics as our steerable filters and can get rid of the complex conjugation.

7. Conclusion and Future Work

In this work, we have stated and proven a Wigner-Eckart Theorem for steerable kernels of general compact groups, which leads to a general description of how to parameterize steerable CNNs and gauge equivariant neural networks for compact transformation groups. We have done this by *linearizing* steerable kernels in a suitable way such that they can be understood similar to spherical tensor operators in physics. We have shown how the basis kernels can always be succinctly described using endomorphisms of irreducible representations, Clebsch-Gordan coefficients, and harmonic basis functions. In the examples, we have demonstrated that it is a structured and doable process to figure out these ingredients in specific use cases. This answers the first four of the research questions that we stated in Section 1.3.

Several open questions remain that we want to discuss here. We start by discussing how practitioners might go about to solve for steerable kernel bases in specific situations that they face. Afterward, we discuss how the results in this thesis might be further generalized, which is the fifth research question that we have formulated in the beginning.

7.1. Recommendations for Applying our Result to Find Steerable Kernel Bases of New Groups

We think in the following direction: we have stated in Theorem 2.1.22 a version of the Peter-Weyl Theorem that works for arbitrary homogeneous spaces of compact groups and for both the fields $\mathbb{R}$ and $\mathbb{C}$. However, this was only an *existence statement* about a decomposition of the space of square-integrable functions and does not directly answer the question of how one might obtain this decomposition in practice. As one could see in the demonstration of our examples in Chapter 6, this can usually be done. Nevertheless, we note that we made heavy use of well-known results in harmonic analysis along the way that would have been hard to obtain from scratch. The same certainly holds for some of the results on Clebsch-Gordan decompositions that we have taken from the literature, especially for the group $\mathrm{SO}(3)$. While it might not be worthwhile to try to prove general theorems in this area, especially since this is probably considered to be a subfield of harmonic analysis and representation theory and would take one further away from deep learning, we assume that the following pointers might help the interested practitioner in finding steerable kernel bases in practice:

1. Even if one is interested in steerable kernels for real representations, it is of-

ten useful to first try to solve for them in the complex case. This also has the advantage that one does not need to worry about endomorphisms.

2. Furthermore, in the complex, case one has access to the Peter-Weyl Theorem as it is stated in the literature. If the homogeneous space is the full group G then this means that normalizations of the matrix coefficients of irreducible representations form the harmonic basis functions of $L^2_{\mathbb{C}}(G)$. Since $L^2_{\mathbb{C}}(X)$ is a subspace of $L^2_{\mathbb{C}}(G)$, one can sometimes hope that the matrix coefficients also help for finding the basis functions of $L^2_{\mathbb{C}}(X)$, especially if $L^2_{\mathbb{C}}(X)$ is a subspace in a "non-twisted way".

3. Sometimes, one might have access to collections of square-integrable functions that transform as irreducible representations, but one is maybe not completely sure whether this already *exhausts* the space of square-integrable functions, i.e. whether the basis is complete. For knowing this, it would be useful to obtain a priori some knowledge about the multiplicities of the different irreducible representations in $L^2_{\mathbb{K}}(X)$. Some results in this direction can be found in Gallier and Quaintance [41], Chapter 6.10.

4. If one tries to transport the result over into the real realm, then one might be lucky, as in our examples of $\mathrm{SO}(3)$ and $\mathrm{O}(3)$. Here, one could show that all complex irreducible representations are of *real type*, which resulted in all real irreducible representations to be of real type as well. Additionally, they could all be found as direct summands of the complex irreducible representations with restricted scalars. If all irreducible real representations are of real type, then one can ignore endomorphisms also in the real case.

5. The most peculiar example we saw was that of $\mathrm{SO}(2)$-equivariant kernels over the real numbers. In this case, the result was not completely analogous to the one in the complex domain, i.e. the case of harmonic networks [4]. The reason was that the irreducible representations were mostly of complex type, instead of real type, which meant that the endomorphism spaces were 2-dimensional (i.e. isomorphic to $\mathbb{C}$). This complicates the considerations. It might be interesting to look into the underlying structure of this example once more in order to infer generalizable results. It is probably helpful to consult Bröcker and Dieck [40], Chapter II.6, for this.

6. Furthermore, let us mention that in practice, in very difficult cases, one may have no need to even solve all of these problems analytically, as was already observed by Shutty and Wierzynski [33]. This paper observed that Clebsch-Gordan coefficients can be found algorithmically by solving a linear program and that one may be able to *learn* irreducible representations. We note that additionally, endomorphisms can probably be found in a similar way. For endomorphism spaces, it may help to first find out the dimensionality of the endomorphism space, which can for example be achieved using the Schur indicator function, see Bröcker and Dieck [40], Proposition II.6.8.

7.2. A Possible Generalization to Equivariant CNNs on Homogeneous Spaces

We are overall most excited about the idea to further generalize the results in this work. As mentioned, our kernel solution completely covers that of steerable CNNs on $\mathbb{R}^n$ and of gauge equivariant CNNs operating on arbitrary Riemannian manifolds, at least for the case of compact transformation groups. This is a quite general setting, however, it assumes the kernel to always be defined on a flat space or, as in the case of gauge equivariant CNNs, the tangent spaces.
We are especially interested in generalizing the kernel solution to the case of equivariant CNNs on homogeneous spaces [11]. As mentioned in Section 5.5, one then deals with the case of a unimodular locally compact group H and two subgroups G_{in} and G_{out} that act via input- and output representations $\rho_{\mathrm{in}} : G_{\mathrm{in}} \to \mathrm{Aut}_{\mathbb{K}}(V_{\mathrm{in}})$ and $\rho_{\mathrm{out}} : G_{\mathrm{out}} \to \mathrm{Aut}_{\mathbb{K}}(V_{\mathrm{out}})$. One characterization for the space of steerable kernels in this setting is the following:

$$\begin{aligned} &\mathrm{Hom}_{G_{\mathrm{in}} \times G_{\mathrm{out}}}(H, \mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}})) \\ &= \big\{K : H \to \mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}}) \mid K(g_{\mathrm{out}} h g_{\mathrm{in}}) = \rho_{\mathrm{out}}(g_{\mathrm{out}}) \circ K(h) \circ \rho_{\mathrm{in}}(g_{\mathrm{in}})\big\}. \end{aligned} \tag{7.1}$$

How can one interpret this in light of representation theory? The most useful way seems the following, as we have already indicated in the notation with the subscript $G_{\mathrm{in}} \times G_{\mathrm{out}}$[1]. Namely, the group $G_{\mathrm{in}} \times G_{\mathrm{out}}$ acts from the left on H by

$$(g_{\mathrm{in}}, g_{\mathrm{out}}) \cdot h := g_{\mathrm{out}} h g_{\mathrm{in}}^{-1}.$$

Furthermore, $G_{\mathrm{in}} \times G_{\mathrm{out}}$ has a Hom-representation on $\mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}})$ that is constructed similarly to the Hom-representation of a single group, only that now two group elements take part in the action:

$$\big[\rho_{\mathrm{Hom}}(g_{\mathrm{in}}, g_{\mathrm{out}})\big](f) := \rho_{\mathrm{out}}(g_{\mathrm{out}}) \circ f \circ \rho_{\mathrm{in}}(g_{\mathrm{in}})^{-1}. \tag{7.2}$$

Then, one can see that the set of steerable kernels is just the set of $G_{\mathrm{in}} \times G_{\mathrm{out}}$-equivariant maps $H \to \mathrm{Hom}_{\mathbb{K}}(V_{\mathrm{in}}, V_{\mathrm{out}})$. More precisely, the kernel constraint can be written as

$$K\big((g_{\mathrm{in}}, g_{\mathrm{out}}) \cdot h\big) = \rho_{\mathrm{Hom}}(g_{\mathrm{in}}, g_{\mathrm{out}})\big(K(h)\big)$$

for all $h \in H$, $g_{\mathrm{in}} \in G_{\mathrm{in}}$ and $g_{\mathrm{out}} \in G_{\mathrm{out}}$. Note that there is now a sign flip compared to the original formulation in 7.1. As in our setting in Theorem 3.1.7, one can hope that there is a correspondence between such steerable kernels and kernel operators on the space of square-integrable functions on H, $L^2_{\mathbb{K}}(H)$. For this to make sense, one can define the following representation of $G_{\mathrm{in}} \times G_{\mathrm{out}}$ on $L^2_{\mathbb{K}}(H)$, which works completely analogously to what we have seen before:

$$\lambda : G_{\mathrm{in}} \times G_{\mathrm{out}} \to \mathrm{U}(L^2_{\mathbb{K}}(H)), \ \big[\lambda(g_{\mathrm{in}}, g_{\mathrm{out}})(\varphi)\big](h) := \varphi\big((g_{\mathrm{in}}, g_{\mathrm{out}})^{-1} \cdot h\big) = \varphi\big(g_{\mathrm{out}}^{-1} h g_{\mathrm{in}}\big).$$

[1] This subscript is not present in the original work.

It is not entirely clear whether a correspondence between kernels and kernel operators is possible in this level of generality. One important reason is that in the proof of this theorem, especially in Lemma 3.2.11, we have actually made use of the compactness of G in order to get a description of kernel operators. Namely, the Peter-Weyl Theorem was used, which is a statement about compact groups and not valid for locally compact groups. But we note that the *intuitions* as to why this theorem is true which we gave in Section 3.1.4 seem more general. Therefore, there is hope for a generalization.
However, even if this correspondence can be established, one cannot just ignore the missing compactness altogether. Namely, in order to establish a Wigner-Eckart Theorem for steerable kernels in Theorem 4.1.13, we have first established a Wigner-Eckart Theorem for kernel operators, which crucially depended on the ability to decompose $L^2_{\mathbb{K}}(X)$. Nevertheless, we mention that there might be a way out: in case that a group G is is of so-called type I, second-countable and locally compact, the space of square-integrable functions on G has a *direct integral decomposition*

$$L^2_{\mathbb{C}}(G) \cong \int_{\pi \in \hat{G}} E_\pi d\pi,$$

where E_π is the space of so-called Hilbert-Schmidt endomorphisms of the underlying Hilbert space of the irreducible representation with index π, and $d\pi$ is the so-called Plancherel measure. This is a generalization of the Peter-Weyl Theorem that can be found in Segal [42] and Mautner [43]. We have neither yet looked into the precise meaning of this formula, nor have we tried to investigate its applicability for proving a Wigner-Eckart Theorem for kernel operators on $L^2_{\mathbb{K}}(H)$ from above, but it is certainly a result that is worth to be further explored. Also, note that $L^2_{\mathbb{K}}(H)$ is not considered as a representation of H itself in our setting, but of the group $G_{\text{in}} \times G_{\text{out}}$, so this setting is a bit different from the one considered in the direct integral decomposition above.
We also mention that 7.1 is not the only characterization of the space of equivariant kernels given in Cohen et al. [11], and other ones might lead to an easier generalization of our work. We also have no clear sense yet what role the group structure of H plays, since the representation theory that enters is mostly the one of G_{in} and G_{out}. Furthermore, one might also decompose H into orbits of the left action of $G_{\text{in}} \times G_{\text{out}}$, which gives compact orbits if G_{in} and G_{out} are compact. It seems worthwhile to start generalizing our results by first looking at this compact case of the theory of CNNs on homogeneous spaces, and then to proceed from there.
Overall we hope that this discussion of further generalizations is a fruitful food for thought for any theoreticians who would like to tackle interesting problems on the intersection of deep learning, representation theory, harmonic analysis, and physics.

A. Mathematical Preliminaries

In this appendix, we state mathematical preliminaries that we use throughout the text. In this whole chapter, $\mathbb{K}$ is one of the two fields $\mathbb{R}$ or $\mathbb{C}$.

A.1. Concepts from Topology, Normed Spaces, and Metric Spaces

Since in this work, we want to develop the theory of representations over *compact* groups, and since this is a topological property, we need to formulate some topological concepts [23]. Additionally, the vector spaces on which our compact groups act also carry a topology, mostly coming from their Hilbert space structure.

Definition A.1.1 (Topological Space, Open Sets, Closed Sets)**.** A *topological space* $(X, \mathcal{T})$ consists of a set X and a set $\mathcal{T}$ of subsets of X, called the *open* sets, such that arbitrary unions and finite intersections of open sets are open. In particular, X and the empty set $\emptyset$ are open. *Closed* sets are the complements of open sets and fulfill dual axioms: arbitrary intersections and finite unions of closed sets are closed.

Let in the following X and Y be topological spaces.

Definition A.1.2 (Open Neighborhood)**.** Let $x \in X$. An open set $U \subseteq X$ is called *open neighborhood* of x if $x \in U$.

Definition A.1.3 (Hausdorff Space)**.** X is called a *Hausdorff space* if two distinct points can always be separated by open sets, i.e. for all $x, y \in X$ there exist U_x, U_y open such that $x \in U_x, y \in U_y$, and $U_x \cap U_y = \emptyset$.

In this work, all topological spaces are Hausdorff.

Definition A.1.4 (Subspace)**.** Assume $A \subseteq X$ is a subset. Then the set $\mathcal{T}_A \coloneqq \{U \cap A \mid U \in \mathcal{T}\}$ is a topology for A and thus makes A a topological space as well. It is called a *subspace* of X.

Whenever we consider a subset of a topological space, it is viewed as a topological space with this construction.

Definition A.1.5 (Closure, density)**.** For $A \subseteq X$, its *closure* $\overline{A}$ is defined as the smallest closed subset of X that contains A. Equivalently, it is the intersection of all closed subsets of X containing A, which is closed by the axioms of a topology. A is called *dense in* X if $\overline{A} = X$.

Definition A.1.6 (Continuous Function, homeomorphism)**.** A function $f : X \to Y$ is called *continuous* if preimages of open sets are always open. Equivalently, for each point $x_0 \in X$ and each open neighborhood V of $f(x_0)$ there is an open neighborhood U of x_0 such that $f(U) \subseteq V$.
A *homeomorphism* is a continuous bijective function with a continuous inverse.

Note that compositions of continuous functions are continuous as well.

Definition A.1.7 (Open Cover, Compact Space)**.** An *open cover* of X is a family of open sets $\{U_i\}_{i \in I}$ that cover X, i.e. $X = \bigcup_{i \in I} U_i$. X is called *compact* if all open covers have a finite subcover, that is: For all open covers $\{U_i\}_{i \in I}$ there exists a finite subset $J \subseteq I$ such that $\{U_i\}_{i \in J}$ is still an open cover of X.

Proposition A.1.8. *If X is compact and $f : X \to Y$ is continuous, then $f(X) \subseteq Y$ is compact as well. In particular, if f surjective, then Y is compact.*

Proof. See Sutherland [44], Proposition 13.15. □

Proposition A.1.9. *Let $f : X \to Y$ be a continuous bijection and assume that X is compact and that Y is Hausdorff. Then the inverse f^{-1} is continuous as well and thus f is a homeomorphism.*

Proof. See Sutherland [44], Proposition 13.26. □

Definition A.1.10 (Product Topology)**.** The *product topology* on $X \times Y$ is the coarsest (i.e. smallest in terms of inclusion) topology that makes both projections $p_X : X \times Y \to X$ and $p_Y : X \times Y \to Y$ continuous.

If Z is a third topological space and we have two continuous functions $f_X : Z \to X$ and $f_Y : Z \to Y$, then the function $f_X \times f_Y : Z \to X \times Y$, $z \mapsto (f_X(z), f_Y(z))$ is continuous as well.

Definition A.1.11 (Quotient Map, Quotient Space)**.** A continuous function $f : X \to Y$ is called a *quotient map* if f is surjective and if $U \subseteq Y$ is open if and only if $f^{-1}(U) \subseteq X$ is open.
Let $\sim$ be any equivalence relation on X and $X/\sim$ be the quotient set formed by identifying equivalent elements. Let $q : X \to X/\sim$ be the canonical function sending each element to its equivalence class. We define $U \subseteq X/\sim$ to be open if $q^{-1}(U) \subseteq X$ is open. Then q is a quotient map and $X/\sim$ is called a *quotient space.*

Proposition A.1.12 (Universal property of Quotient Maps)**.** *Let $q : X \to X/\sim$ be a standard quotient map and $f : X \to Y$ be any continuous function such that $f(x) = f(x')$ whenever $x \sim x'$. Then there is a unique continuous function $\overline{f} : X/\sim \to Y$ such that the following diagram commutes:*

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "q"'] & Y \\ X/\sim \arrow[ur, "\overline{f}"'] & \end{tikzcd}$$

$\overline{f}$ is given on equivalence classes by $\overline{f}([x]) = f(x)$.

Proof. See Conway [23], Proposition 2.8.7. □

It can be shown that all quotient maps are equivalent to a construction of the form $q : X \to X/\sim$. Namely, for a quotient map $f : X \to Y$, define $\sim$ by $x \sim x'$ if $f(x) = f(x')$. Then the map $\overline{f} : X/\sim \to Y$, $[x] \mapsto f(x)$ is a well-defined continuous map by the universal property of quotient maps Proposition A.1.12. One can show that this is a homeomorphism. Thus for a quotient map $f : X \to Y$ we also call Y a quotient space.
Our route for defining concrete topologies is in most cases through the existence of inner products on Hilbert spaces, which will be defined in detail in Definition A.2.1. Namely, inner products define norms, which define metrics [45], which in turn define topologies. For this, we need some definitions:

Definition A.1.13 (Norm)**.** Let V be a $\mathbb{K}$-vector space, A *norm* on V is a map $\|\cdot\| : V \to \mathbb{R}_{\geq 0}$ with the following properties for all $\lambda \in \mathbb{K}$ and $v, w \in V$:

1. $\|v\| = 0$ if and only if $v = 0$.
2. $\|\lambda v\| = |\lambda| \cdot \|v\|$.
3. Triangle inequality: $\|v + w\| \leq \|v\| + \|w\|$.

If $\langle\cdot|\cdot\rangle : V \times V \to \mathbb{K}$ is an inner product on a Hilbert space, then it defines a norm $\|\cdot\| : V \to \mathbb{R}_{\geq 0}$ by $\|x\| := \sqrt{\langle x|x\rangle}$.

Definition A.1.14 (Metric)**.** Let Y be a set. A *metric* on Y is a function $d : Y \times Y \to \mathbb{R}_{\geq 0}$ with the following properties for all $x, y, z \in Y$:

1. $d(x, y) = 0$ if and only if $x = y$.
2. Symmetry: $d(x, y) = d(y, x)$.
3. Triangle inequality: $d(x, z) \leq d(x, y) + d(y, z)$.

A norm $\|\cdot\| : V \to \mathbb{R}_{\geq 0}$ defines a metric $d : V \times V \to \mathbb{R}$ by setting $d(x, y) := \|x - y\|$. In turn, a metric defines a topology as follows: open balls are given by all sets of the form $\mathrm{B}_\epsilon(x) := \{y \in V \mid d(x, y) < \epsilon\}$ for all $x \in V$ and $\epsilon > 0$. Open sets are then defined as arbitrary unions of arbitrary open balls.
Additionally, we need notions about convergence in this work. Since we will deal with them mostly in the context of metric spaces (with normed vector spaces and Hilbert spaces being special cases, as explained above), we focus on these notions for metric spaces.

Definition A.1.15 (Convergent Sequence)**.** Let Y be a metric space. Then a sequence $(y_k)_k$ in Y is said to *converge to* y if for all $\epsilon > 0$ there is a $k_\epsilon \in \mathbb{N}$ such that $y_k \in \mathrm{B}_\epsilon(y)$ for all $k \geq k_\epsilon$.

With this in mind, one can give an equivalent definition of continuity that applies to metric spaces:

Definition A.1.16 (Continuity in metric spaces)**.** A function $f : Y \to Z$ between metric spaces is *continuous in* $y \in Y$ if for each sequence $(y_k)_k$ of points $y_k \in Y$ converging to a point $y \in Y$, we also have that the sequence $f(y_k)$ converges to $f(y)$. This can be understood in terms of the function "commuting with limits":

$$\lim_{k \to \infty} f(y_k) = f\Big(\lim_{k \to \infty} y_k\Big).$$

Furthermore, $f : Y \to Z$ is called *continuous* if it is continuous in all points $y \in Y$.

Equivalently, the following holds: $f : Y \to Z$ is continuous in $y \in Y$ if and only of for all $\epsilon > 0$ there is a $\delta > 0$ such that $f\left(\mathrm{B}_\delta(y)\right) \subseteq \mathrm{B}_\epsilon(f(y))$.

Definition A.1.17 (Uniform continuity)**.** A function $f : Y \to Z$ between metric spaces is called *uniformly continuous* if for each $\epsilon > 0$ there is a $\delta > 0$ such that for all $y, y' \in Y$ with $d_Y(y, y') < \delta$ we obtain $d_Y(f(y), f(y')) < \epsilon$.

The following is a result we use several times in the main text:

Proposition A.1.18. *Let $f : V \to V'$ be a linear function between normed vector spaces. Then the following are equivalent:*

1. *f is uniformly continuous.*

2. *f is continuous.*

3. *f is continuous in* 0*.*

Proof. Trivially, 1 implies 2, which in turn implies 3. Now assume 3, i.e. f is continuous in 0. Let $\epsilon > 0$. Then by continuity in 0, there exists $\delta > 0$ such that for all $v \in V$ with $\|v\| = \|v - 0\| < \delta$ we obtain $\|f(v)\| = \|f(v) - f(0)\| < \epsilon$. Now let $v, v' \in V$ be arbitrary with $\|v - v'\| < \delta$. Then by the linearity of f we obtain:

$$\|f(v) - f(v')\| = \|f(v - v')\| < \epsilon,$$

which is exactly what we wanted to show. $\square$

Sometimes, sequences *look like* they converge since their elements get ever closer to each other. However, not all such sequences need to converge. Therefore, there is the following notion:

Definition A.1.19 (Cauchy Sequence)**.** Let Y be a metric space. A sequence $(y_k)_k$ in Y is a *Cauchy Sequence* if for all $\epsilon > 0$ there is $k_\epsilon \in \mathbb{N}$ such that for all $k, k' > k_\epsilon$ we have $d(y_k, y_{k'}) < \epsilon$.

For example, one can consider the metric space $\mathbb{R} \setminus \{0\}$ together with the usual metric. Then the sequence $\left(\frac{1}{k}\right)_k$ is a Cauchy sequence but does not converge since the limit (in $\mathbb{R}$!), which would be 0, is not in $\mathbb{R} \setminus \{0\}$. Thus, the following notion is useful:

Definition A.1.20 (Complete Metric Space)**.** A metric space Y is called *complete* if every Cauchy sequence converges.

Definition A.1.21 (Completion)**.** Let Y be a metric space. A *completion of* Y is a metric space Y' which contains Y as a dense subspace and such that Y' is complete.

Proposition A.1.22 (Universal Property of Completions)**.** *Assume that $Y \subseteq Y'$ is a pair of metric spaces, where Y' is a completion of Y. Then the following universal property holds:*
Let Z be any complete metric space and $f : Y \to Z$ be any uniformly continuous function. Then there is a unique continuous function $f' : Y' \to Z$ that extends f, i.e. such that $f'|_Y = f$. f' furthermore is also uniformly continuous. This can be expressed by the following commutative diagram, where $i : Y \to Y'$ is the canonical inclusion:

$$\begin{tikzcd} Y \arrow[r, "f"] \arrow[d, "i"] & Z \\ Y' \arrow[ur, "f'"'] & \end{tikzcd}$$

Proof. See for example Kaplansky [45]. □

Definition A.1.23 (Boundedness)**.** Let Y be a metric space. A subset $A \subseteq Y$ is called *bounded* if there is a constant $C > 0$ such that $d(a, b) \leq C$ for all $a, b \in A$.

Theorem A.1.24 (Heine-Borel Theorem)**.** *A subset $A \subseteq \mathbb{K}^d$ is compact if and only if it is closed and bounded.*

Proof. See Conway [23], Theorem 1.4.8. □

Corollary A.1.25 (Extreme Value Theorem)**.** *Let $f : X \to \mathbb{R}$ be continuous, where X is any nonempty compact topological space. Then f has a maximum and a minimum.*

Proof. By Proposition A.1.8, $f(X) \subseteq \mathbb{R}$ is compact. By Theorem A.1.24 this means that $f(X)$ is closed and bounded. Boundedness means that the supremum is finite and closedness means that the supremum must lie in $f(X)$, and consequently it is a maximum. For the minimum, the same arguments apply. □

A.2. Pre-Hilbert Spaces and Hilbert Spaces

Here, we state foundational concepts in the theory of Hilbert spaces [46].

Definition A.2.1 (pre-Hilbert Space, Hilbert space)**.** A *pre-Hilbert space* $V = (V, \langle\cdot|\cdot\rangle)$ consists of the following data:

1. A vector space V over $\mathbb{K}$.
2. An inner product $\langle\cdot|\cdot\rangle : V \times V \to \mathbb{K}$, $(x, y) \mapsto \langle x|y\rangle$.

It has the following properties that hold for all $x, x', y, y' \in V$, $\lambda \in \mathbb{K}$:

1. The inner product is conjugate linear in the first component: $\langle x + x' | y \rangle = \langle x|y \rangle + \langle x'|y \rangle$ and $\langle \lambda x|y \rangle = \overline{\lambda} \langle x|y \rangle$, where $\overline{\lambda}$ is the complex conjugate of λ.

2. The inner product is linear in the second component: $\langle x|y + y' \rangle = \langle x|y \rangle + \langle x|y' \rangle$ and $\langle x|\lambda y \rangle = \lambda \langle x|y \rangle$.

3. The inner product is conjugate symmetric: $\langle y|x \rangle = \overline{\langle x|y \rangle}$

4. The inner product is positive definite: $\langle x|x \rangle > 0$ unless $x = 0$.

If additionally, the following statement holds, then V is called a *Hilbert Space*:

5. V, together with the norm $\| \cdot \| : V \to V$ induced from the inner product by $\|x\| := \sqrt{\langle x|x \rangle}$, and consequently the metric defined by $d(x, y) := \|x - y\|$, is a complete metric space as in Definition A.1.20.

Remark A.2.2. Of course, all Hilbert Spaces are pre-Hilbert spaces, and so all Propositions about pre-Hilbert spaces in the following apply to Hilbert spaces just as well. Note that the first property follows from the second and third. We also mention that usually, inner products on Hilbert spaces are assumed to be linear in the first and conjugate linear in the second component, in contrast to how we view it. The reason for our choice is that our work is inspired by connections to physics where our convention is more common. It is basically the bra-ket convention. Furthermore, note that if $\mathbb{K} = \mathbb{R}$, then conjugate linear maps are linear and thus the inner product will be linear in both components. Additionally, it will be symmetric instead of only conjugate symmetric.

Proposition A.2.3 (Cauchy-Schwartz inequality)**.** *For any two elements v, w in a pre-Hilbert space V, we have*

$$|\langle v|w \rangle| \leq \|v\| \cdot \|w\|.$$

We have equality if and only if v and w are linearly dependent.

Proof. See Debnath and Mikusinski [46], Theorem 3.2.9. □

Definition A.2.4 (Orthogonality)**.** Two vectors v, w in a pre-Hilbert space V are called *orthogonal*, written $v \perp w$, if$\langle v|w \rangle = 0$.

Obviously, being orthogonal is a symmetric relation.

Definition A.2.5 (Orthogonal Complement)**.** Let V be a pre-Hilbert space and $W \subseteq V$ a subset. $v \in V$ is *orthogonal to* W if $\langle v \mid w \rangle = 0$ for all $w \in W$.
The *orthogonal complement* of W, denoted $W^{\perp}$, is the set of all vectors in V that are orthogonal to W.

Proposition A.2.6 (Closedness of Complements)**.** *Let $W \subseteq V$ be a subset of a pre-Hilbert space V. Then $W^{\perp}$ is a topologically closed linear subspace of V.*

Proof. See Debnath and Mikusinski [46], Theorem 3.6.2. □

Proposition A.2.7 (Continuity of Scalar Product)**.** *For any pre-Hilbert space V, the scalar product $\langle\cdot|\cdot\rangle : V \times V \to \mathbb{K}$ is continuous.*

Proof. See Debnath and Mikusinski [46], Theorem 3.3.12. □

Definition A.2.8 (Orthonormal System)**.** A family $(v_i)_{i\in I}$ of elements in a pre-Hilbert space is called *orthonormal system* if $\|v_i\| = 1$ for all $i \in I$ and $v_i \perp v_j$ for all $i \neq j$.

Definition A.2.9 (Orthonormal Basis)**.** An orthonormal system $(v_i)_{i\in I}$ in a Hilbert space V is called *orthonormal basis* if the linear span of all $\{v_i\}_{i\in I}$ is dense in V. If this is the case, then each $v \in V$ can be uniquely written as

$$v = \sum_{i\in I} \alpha_i v_i$$

with only countably many $\alpha_i \in \mathbb{K}$ being nonzero. The coefficients are given by $\alpha_i = \langle v_i|v\rangle$.

We stress that while the index set I can be uncountably infinite, the sequence expansions of each element in V only have countably many entries. It is obvious from the Peter-Weyl Theorem 2.1.22 and this definition that the functions

$$\left\{Y_{li}^m \mid l \in \hat{G}, i \in \{1,\ldots,m_l\}, m \in \{1,\ldots,[l]\}\right\}$$

form an orthonormal basis of $L^2_{\mathbb{K}}(X)$.

Proposition A.2.10 (Gram-Schmidt Orthonormalization)**.** *For every linearly independent sequence $(y_k)_k$ in a pre-Hilbert space V with $N \in \mathbb{N} \cup \{\infty\}$ elements, one can find an orthonormal sequence $(v_k)_k$ in V such that the following holds: for all $n \in \mathbb{N}$, $n \leq N$, the progressive linear span stays the same:*

$$\operatorname{span}_{\mathbb{K}}(v_1,\ldots,v_n) = \operatorname{span}_{\mathbb{K}}(y_1,\ldots,y_n).$$

In particular, since every finite-dimensional Hilbert space has a vector space basis, it necessarily also has an orthonormal basis.

Proof. See Debnath and Mikusinski [46], page 110. □

Definition A.2.11 (Adjoint of an operator)**.** Let $f : V \to V'$ be a continuous linear function between Hilbert spaces. Then there is a unique continuous linear function $f^* : V' \to V$ such that for all $v \in V$ and $v' \in V'$ one has:

$$\langle f(v)|v'\rangle_{V'} = \langle v|f^*(v')\rangle_V .$$

f^* is called the *adjoint* of f.

The existence of adjoints is for example discussed in Debnath and Mikusinski [46], page 158. This book only considers the case of operators on a Hilbert space to itself, but these considerations generalize to the setting with two different Hilbert spaces. One has the following:

Proposition A.2.12. *Let $f : V \to V'$ and $g : V' \to V''$ be continuous linear functions between Hilbert spaces. Then:*

1. $(f^*)^* = f$.
2. $\mathrm{id}_V^* = \mathrm{id}_V$.
3. $(g \circ f)^* = f^* \circ g^*$.

Proof. All of these properties follow directly from the uniqueness of adjoints. □

Proposition A.2.13. *Let $f : V \to V'$ be a unitary transformation between Hilbert spaces, i.e. an invertible linear function such that $\langle f(v) | f(w) \rangle = \langle v | w \rangle$ for all $v, w \in V$. Then the adjoint is the inverse, i.e. $f^* = f^{-1}$.*

Proof. First of all, the inverse f^{-1} is again continuous due to the unitarity of f. Furthermore, due to the unitarity, we obtain

$$\begin{aligned} \langle f(v) | v' \rangle &= \left\langle f(v) \middle| f(f^{-1}(v')) \right\rangle \\ &= \left\langle v \middle| f^{-1}(v') \right\rangle \end{aligned}$$

for all $v \in V$ and $v' \in V'$. Due to the uniqueness of adjoints, we obtain $f^{-1} = f^*$. □

The following proposition is sometimes used in the main text:

Proposition A.2.14. *Let $v, w \in V$ be two elements in a pre-Hilbert space such that $\langle v | u \rangle = \langle w | u \rangle$ for all $u \in V$. Then $v = w$.*

Proof. We have

$$\langle v - w | u \rangle = \langle v | u \rangle - \langle w | u \rangle = 0$$

for all $u \in V$. In particular, when setting $u = v - w$ we obtain

$$\langle v - w | v - w \rangle = 0$$

and thus $v - w = 0$, i.e. $v = w$. □

Proposition A.2.15 (Orthogonal Projection Operators)**.** *Let $W \subseteq V$ be a topologically closed subspace of a Hilbert space. Then there is a continuous linear function $P : V \to W$ such that for all $v \in V$ and $w \in W$ we have*

$$\langle P(v) | w \rangle = \langle v | w \rangle .$$

Furthermore, if W is finite-dimensional and $w_1, \dots, w_n$ and orthonormal basis, then P is given explicitly by

$$P(v) = \sum_{i=1}^{n} \langle w_i | v \rangle \, w_i.$$

Proof. That W is topologically closed means that W, with the scalar product inherited from V, is a complete metric space. Thus, W is a Hilbert space as well. Therefore, the continuous linear embedding $i : W \to V$ given by $w \mapsto w$ has an adjoint $i^* : V \to W$ by Definition A.2.11. Set $P := i^*$. For arbitrary $v \in V$ and $w \in W$ we obtain:

$$\begin{aligned} \langle P(v)|w\rangle &= \langle i^*(v)|w\rangle \\ &= \langle v|i(w)\rangle \\ &= \langle v|w\rangle\,. \end{aligned}$$

For the second statement, note that for all $j \in \{1, \dots, n\}$ we have, using that the w_i are orthonormal:

$$\begin{aligned} \left\langle \sum\nolimits_{i=1}^n \langle w_i|v\rangle\, w_i \middle| w_j \right\rangle &= \sum\nolimits_{i=1}^n \overline{\langle w_i|v\rangle}\, \langle w_i|w_j\rangle \\ &= \langle v|w_j\rangle \\ &= \langle P(v)|w_j\rangle\,. \end{aligned}$$

By Proposition A.2.14 and since the w_j generate W we obtain $\sum_{i=1}^n \langle w_i|v\rangle\, w_i = P(v)$ as claimed. □

Proposition A.2.16. *Let $(V, \langle\cdot|\cdot\rangle)$ be a finite-dimensional pre-Hilbert space. Then this space is already complete and thus a Hilbert space.*
In particular, all finite-dimensional subspaces of Hilbert spaces are topologically closed.

Proof. The proof of the Gram-Schmidt orthonormalization in Proposition A.2.10 does not make use of the completeness of the Hilbert space, and thus it holds for pre-Hilbert spaces as well. Consequently, V, being finite-dimensional, has an orthonormal basis. It is thus isomorphic to $\mathbb{K}^n$ together with the standard scalar product, which is well-known to be complete. Thus, V is a Hilbert space.
Now, let $W \subseteq V$ be a finite-dimensional subspace of a Hilbert space V which may be infinite-dimensional. Then W is a pre-Hilbert space and by what was just shown a Hilbert space. Consequently, all sequences in W which have a limit in V need, by completeness, to have that limit already in W. This shows that W is topologically closed. □

Bibliography

[1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton. ImageNet Classification with Deep Convolutional Neural Networks. *Neural Information Processing Systems*, 25, Jan 2012. doi: 10.1145/3065386.

[2] Taco Cohen and Max Welling. Group Equivariant Convolutional Networks. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 48, pages 2990–2999, New York, New York, USA, 20–22 Jun 2016. PMLR.

[3] Taco S. Cohen and Max Welling. Steerable CNNs, 2016.

[4] Daniel E. Worrall, Stephan J. Garbin, Daniyar Turmukhambetov, and Gabriel J. Brostow. Harmonic Networks: Deep Translation and Rotation Equivariance. *CoRR*, abs/1612.04642, 2016.

[5] Alexander Bogatskiy, Brandon Anderson, Jan T. Offermann, Marwah Roussi, David W. Miller, and Risi Kondor. Lorentz Group Equivariant Neural Network for Particle Physics, 2020.

[6] Nathaniel Thomas, Tess Smidt, Steven M. Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor Field Networks: Rotation- and Translation-Equivariant Neural Networks for 3D Point Clouds. *ArXiv*, abs/1802.08219, 2018.

[7] Taco Cohen, Maurice Weiler, Berkay Kicanaoglu, and Max Welling. Gauge Equivariant Convolutional Networks and the Icosahedral CNN. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 1321–1330, Long Beach, California, USA, 09–15 Jun 2019. PMLR.

[8] Maurice Weiler, Mario Geiger, Max Welling, Wouter Boomsma, and Taco Cohen. 3D Steerable CNNs: Learning Rotationally Equivariant Features in Volumetric Data, 2018.

[9] Maurice Weiler and Gabriele Cesa. General $E(2)$-Equivariant Steerable CNNs, 2019.

[10] Nadir Jeevanjee. *An introduction to tensors and group theory for physicists*. Birkhäuser, New York, NY, 2011. doi: 10.1007/978-0-8176-4715-5.

[11] Taco Cohen, Mario Geiger, and Maurice Weiler. A General Theory of Equivariant CNNs on Homogeneous Spaces. *CoRR*, abs/1811.02017, 2018.

[12] Taco S. Cohen, Mario Geiger, Jonas Köhler, and Max Welling. Spherical CNNs. In *International Conference on Learning Representations*, 2018.

[13] Alexander Arhangel'skii and Mikhail Tkachenko. *Topological Groups and Related Structures*. Jan 2008.

[14] Anthony Knapp. *Lie Groups Beyond an Introduction, Second edition*, volume 140. Jan 2002. doi: 10.1007/978-1-4757-2453-0.

[15] E. Kowalski. *An Introduction to the Representation Theory of Groups*. Graduate Studies in Mathematics. American Mathematical Society, 2014. ISBN 9781470409661.

[16] L. Nachbin and L. Bechtolsheim. *The Haar integral*. University series in higher mathematics. Van Nostrand, 1965.

[17] Stefan Dawydiak. Is there a Peter-Weyl-Theorem over the real numbers?, 2020.

[18] T. Tao. *An Introduction to Measure Theory*. Graduate studies in mathematics. American Mathematical Society, 2013. ISBN 9781470409227.

[19] N. Bourbaki. *General Topology: Chapters 1-4*. Elements of mathematics. Springer, 1998. ISBN 9783540642411.

[20] Dana P. Williams. The Peter-Weyl Theorem for Compact Groups, 1991.

[21] Terrence Tao. The Peter-Weyl Theorem, and non-abelian Fourier analysis on compact groups, 2011.

[22] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, USA, 2nd edition, 2012. ISBN 0521548233.

[23] John Conway. *A Course in Point Set Topology*. Jan 2014. doi: 10.1007/978-3-319-02368-7.

[24] Maurice Weiler, Fred Hamprecht, and Martin Storath. Learning Steerable Filters for Rotation Equivariant CNNs. pages 849–858, Jun 2018. doi: 10.1109/CVPR.2018.00095.

[25] R.V. Kadison and J.R. Ringrose. *Fundamentals of the Theory of Operator Algebras. Volume I*. Fundamentals of the Theory of Operator Algebras. American Mathematical Society, 1997. ISBN 9780821808191.

[26] Vishnu Agrawala. Wigner-Eckart theorem for an arbitrary group or Lie algebra. *Journal of Mathematical Physics*, 21, July 1980. doi: 10.1063/1.524639.

[27] S.M. Lane, S.J. Axler, Springer-Verlag (Nowy Jork)., F.W. Gehring, and P.R. Halmos. *Categories for the Working Mathematician*. Graduate Texts in Mathematics. Springer, 1998. ISBN 9780387984032.

[28] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *IEEE International Conference on Computer Vision (ICCV 2015)*, 1502, Feb 2015. doi: 10.1109/ICCV.2015.123.

[29] Wenling Shang, Kihyuk Sohn, Diogo Almeida, and Honglak Lee. Understanding and Improving Convolutional Neural Networks via Concatenated Rectified Linear Units. Mar 2016.

[30] Klaus Hildebrandt Ruben Wiersma, Elmar Eisemann. CNNs on Surfaces using Rotation-Equivariant Features. *Transactions on Graphics*, 39(4), July 2020. doi: 10.1145/3386569.3392437.

[31] Pim de Haan, MauricWeiler, Taco Cohen, and Max Welling. Gauge Equivariant Mesh CNNs: Anisotropic convolutions on geometric graphs, 2020.

[32] Risi Kondor, Zhen Lin, and Shubhendu Trivedi. Clebsch-Gordan Nets: a Fully Fourier Space Spherical Convolutional Neural Network. In *NeurIPS*, 2018.

[33] Noah Shutty and Casimir Wierzynski. Learning Irreducible Representations of Noncommutative Lie Groups, Jun 2020.

[34] Carlos Esteves. Theoretical Aspects of Group Equivariant Neural Networks, 2020.

[35] Carlos Esteves, Yinshuang Xu, Christine Allen-Blanchette, and Kostas Daniilidis. Equivariant Multi-View Networks. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1568–1577, 2019.

[36] Risi Kondor and Shubhendu Trivedi. On the Generalization of Equivariance and Convolution in Neural Networks to the Action of Compact Groups. Feb 2018.

[37] E.P. Wigner. *Gruppentheorie und ihre Anwendung auf die Quantenmechanik der Atomspektren*. Die Wissenschaft. J.W. Edwards, 1944.

[38] T.M. MacRobert. *Spherical Harmonics: An Elementary Treatise on Harmonic Functions, with Applications*. Methuen, 1947.

[39] A. Bohm and M. Loewe. *Quantum Mechanics: Foundations and Applications*. Springer study edtion. Springer New York, 1993. ISBN 9780387953304.

[40] T. Bröcker and T. Dieck. *Representations of Compact Lie Groups*. Graduate Texts in Mathematics. Springer Berlin Heidelberg, 2003. ISBN 9783540136781.

[41] J. Gallier and J. Quaintance. *Differential Geometry and Lie Groups: A Second Course*. Geometry and Computing. Springer International Publishing, 2020. ISBN 9783030460464.

[42] I. E. Segal. An Extension of Plancherel's Formula to Separable Unimodular Groups. *Annals of Mathematics*, 52(2):272–292, 1950. ISSN 0003486X.

[43] F. I. Mautner. Note on the Fourier inversion formula on groups. *Transactions of the American Mathematical Society*, 78:371–384, 1955.

[44] Wilson Sutherland. Introduction to Metric and Topological Spaces. 1975.

[45] I. Kaplansky. *Set Theory and Metric Spaces*. AMS Chelsea Publishing Series. AMS Chelsea Publishing, 2001. ISBN 9780821826942.

[46] L. Debnath and P. Mikusinski. *Introduction to Hilbert Spaces with Applications*. Elsevier Science, 2005. ISBN 9781493300358.