

Equivariant and Coordinate Independent Convolutional Networks

What is the appropriate geometric structure for neural networks that process spatial signals on Euclidean spaces or more general manifolds? This question takes us on a journey which leads to a gauge field theory of convolutional networks.

Feature vector fields: The spatial signals we are interested in are fields of feature vectors. Feature fields allow to describe data like images, audio, videos, point clouds, or tensor fields, such as fluid flows and electromagnetic fields.

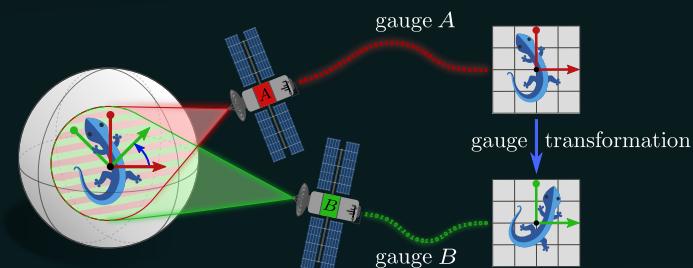
Equivariant networks commute with actions of some symmetry group on their feature spaces. The relevant group actions in this work are geometric transformations of feature fields, like translations, rotations, or reflections of images.

Equivariant models generalize everything they learn over the considered group of transformations. This property makes them significantly more data efficient, interpretable, and robust in comparison to non-equivariant models.

Convolutional Neural Networks (CNNs) are the most common network architecture for processing feature fields. Conventional CNNs operate on Euclidean spaces and are translation equivariant, i.e. position independent. This work explains how to extend CNNs to be equivariant under more general symmetries of space.

Coordinate independence: Manifolds are in general not equipped with a canonical choice of coordinates. Feature fields and neural network layers are hence required to be coordinate independent, that is, expressible relative to different frames of reference.

The ambiguity of local frames represents the gauge freedom of our neural field theory. We show that the demand for coordinate independence requires CNNs to be equivariant under local gauge transformations.



To offer an easy entry, the first part of this work focuses on the representation theory of equivariant convolutional networks on Euclidean spaces.

The insights gained in the Euclidean setting are subsequently leveraged to develop the full gauge theory of coordinate independent CNNs on Riemannian manifolds.

In the last part, we turn to a discussion of practical applications on specific manifolds. A comprehensive literature review demonstrates the generality of our theory by showing for more than 100 models from the literature how they can be understood as specific instantiations of “Equivariant and Coordinate Independent CNNs”.

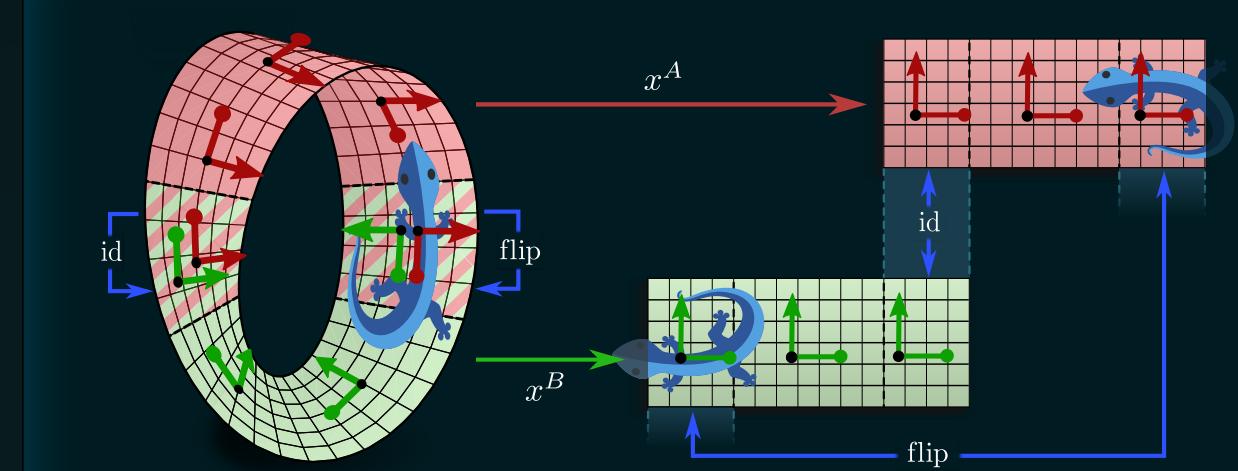
Equivariant and Coordinate Independent CNNs



Maurice
Weiler

EQUIVARIANT AND COORDINATE INDEPENDENT CONVOLUTIONAL NETWORKS

A GAUGE FIELD THEORY OF NEURAL NETWORKS



Maurice Weiler
Patrick Forré
Erik Verlinde
Max Welling

EQUIVARIANT AND COORDINATE INDEPENDENT CONVOLUTIONAL NETWORKS

A GAUGE FIELD THEORY OF NEURAL NETWORKS

The work described in this book has been carried out at the QUVA Deep Vision lab at the University of Amsterdam. Funding for this research was provided by Qualcomm Technologies, Inc.

Copyright ©2023 by M. Weiler, Amsterdam, The Netherlands.

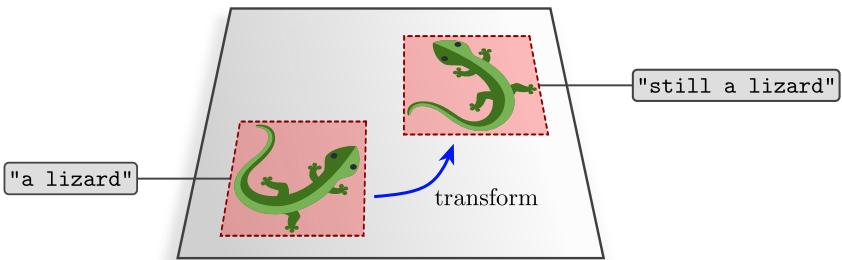
Summary

In this book, *Equivariant and Coordinate Independent Convolutional Networks*, we develop a gauge theory of artificial neural networks for processing spatially structured data like images, audio, or videos. The standard neural network architecture for such data are convolutional networks, which are characterized by their position-independent inference. Generalizing whatever they learn over spatial locations, convolutional networks are substantially more data efficient and robust in comparison to non-convolutional models. This characteristic is especially important in domains like medical imaging, where training data is scarce.

The independence from spatial locations is formally captured by the networks' *translation group equivariance*, i.e. their property to commute with translations of their input signals. We show that the convolutional network design is not only sufficient for translation equivariance but is actually a necessary condition – convolutions can therefore be derived by demanding the model's equivariance. The first part of this work leverages this insight to define *generalized convolutional networks* which are *equivariant under larger symmetry groups*. Such models generalize their inference over additional geometric transformations, for instance, rotations or reflections of patterns in images. We demonstrate empirically that they exhibit a significantly enhanced data efficiency, convergence rate, and final performance in comparison to conventional convolutional networks. Our publicly available implementation found wide use in the research community.

In the second part, we extend convolutional networks further to process signals on Riemannian manifolds. Beyond flat Euclidean images, this setting includes, e.g., spherical signals like global weather patterns on the earth's surface, or signals on general surfaces like artery walls or the cerebral cortex. We show that convolution kernels on manifolds are required to be *equivariant under local gauge transformations* if the networks' inference is demanded to be *coordinate independent*. The resulting coordinate independent networks are proven to be equivariant with respect to the manifolds' global symmetries (isometries).

Our objective is not to propose yet another equivariant network design for a narrow application domain, but to devise a *unifying mathematical framework* for convolutional networks. The last part of this book demonstrates the generality of our differential geometric formulation of convolutional networks by showing that is able to explain a vast number of equivariant network architectures from the literature.



Convolutional neural networks (CNNs) process data in a position-independent manner. Having learned to detect a pattern at one specific location, they will detect it at any other location as well. This work develops a theory of generalized equivariant CNNs which extend this property to 1) further geometric transformations and 2) arbitrary manifolds.

(Lizards adapted under the Creative Commons Attribution 4.0 International license by courtesy of Twitter.)

Acknowledgement

First and foremost, I would like to thank my doctoral advisor *Max Welling* for his exceptional support and supervision throughout my doctoral studies. Without Max’s scientific curiosity and enthusiasm for fundamental questions, this work would not have been possible in its current theoretical scope. I am deeply grateful for his continual advice and guidance, paired with the great degree of freedom I was given to pursue my own research agenda. Max and I met for the first time during a talk I gave when visiting AMLab. When I mentioned my intuition of “independent local transformations” of patterns in images, Max got very excited and proposed a gauge theoretical description of convolutional networks. Years later, we found this description, which is the main contribution of this book. This little anecdote highlights Max’s remarkable capability to see connections between different fields, which repeatedly pointed me in the right direction whenever I got stuck during my research.

Thanks to my co-supervisor *Erik Verlinde* for his valuable feedback on my research from a physicist’s perspective. Erik pointed out multiple connections between our coordinate independent neural networks and Einstein’s theory of relativity.

I would like to express my deepest gratitude to *Patrick Forré*, who taught me the mathematics of associated fiber bundles and helped me formalizing our differential geometric formulation of convolutional networks. We had countless insightful blackboard sessions where he scrutinized the mathematical constructions I presented to him, often pointing out deficiencies that needed to be fixed. During that process, I learned a lot about rigorous mathematical thinking, which will surely be of great utility in my future research. Patrick is AMLab’s “go-to guy” for tricky mathematical questions, and is in this capacity of immeasurable value to the lab.

Special thanks are due to *Gabriele Cesa* and *Leon Lang*, who I supervised and taught as master’s students, but from whom I soon learned a lot in return. They contributed substantially to the representation theory and implementation of steerable (gauge equivariant) convolution kernels, which are at the heart of equivariant CNNs. Both were furthermore supporting me with plenty of helpful discussions and gave valuable feedback on this manuscript.

I truly enjoyed being part of AMLab and QUVA lab and I am grateful to everyone who helped shape this unique and stimulating scientific environment. Many thanks specifically to *Pim de Haan*, *Erik Jenner*, *Andrii Skliar*, *Roberto Bondesan*, *Changyong Oh*, *Rob Romijnders*, *Thomas Kipf*, and *Erik Bekkers* for fruitful collaborations, discussions, and brainstorming sessions. Thanks also go to *Taco Cohen*, who invited me prior to my PhD studies to AMLab and with who I was happy to publish several impactful papers. I also want to point out the role of *Fred Hamprecht* and *Martin Storath*, who got me in contact with equivariant deep learning in the first place.

Getting through my dissertation would not have been possible without the support of good friends. A big thanks to *Nawid Sayed*, *Felix Weidner*, and *Tyll Wunram* for always having an open ear and providing emotional support in good times and bad. Thanks to *Berkay Kicanaoglu*, *Anil Baslamisli*, *Ismailcan Ersahin*, *Gjorgji Strezoski*, and *Mert Kilickaya* for welcoming me to Amsterdam and for all the fun coffee breaks and memorable nights. Sporting with *Victor Halfhide*, *Thomas*, *Gjorgji*, *Berkay*, *Tyll*, *Noureddien Hussein*, and all the people I met at USC allowed me to recharge my battery after a long day of research.

Finally, I would like to thank my family for their unconditional love, their advice, and the home and shelter they gave me. All the geometric constructions conceived and built with my father *Hermann*, the countless hours of studying mathematics with my mother *Marita*, and each single toy disassembled with my brother *Marcel* to reveal their inner workings certainly left their traces in my geometric way of thinking, and hence in this book.

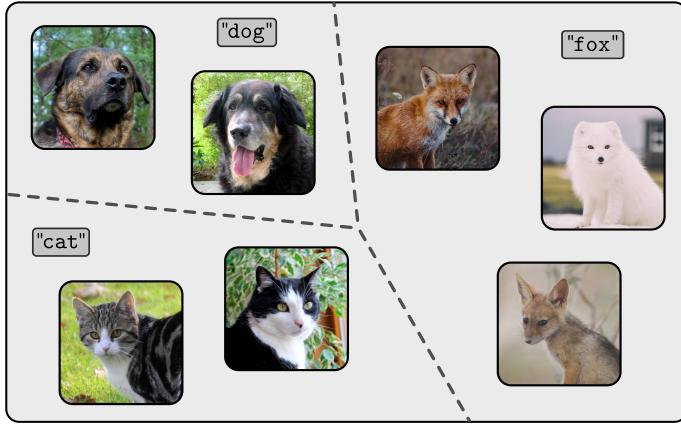
Why Equivariant & Coordinate Independent Convolutional Networks ?

A visual introduction

The gauge theory of “Equivariant and Coordinate Independent CNNs” describes neural networks for processing spatially structured data like audio, images, videos, or more general signals on more general spaces. Why should such networks be *equivariant*, *coordinate independent*, or *convolutional*, and why are they described by a *gauge theory*? The following paragraphs aim to give an informal and intuitive motivation for these properties and the content of this work. A more technical introduction and full overview of this book’s content is given in Chapter 1.

Why *equivariant* neural networks ?

To motivate the merits of an equivariant network design, consider the simple application of image classification. The goal is to partition the space of images into various classes:



(Animal photos adapted under the free license by courtesy of [Freepik](#).)

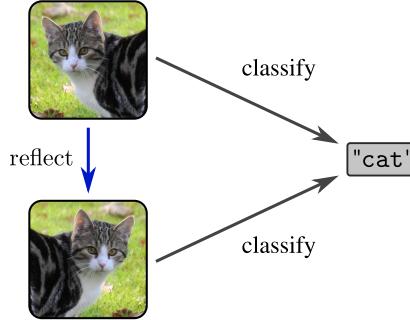
To solve this task, the network has to become *invariant* to the intra-class variability of each class. For instance, there are many different appearances of fox images, and all of them should be mapped to the same class label “fox”. The intra-class variabilities are partly due to images showing truly different instances, e.g. the red and the arctic fox. In addition, “one and the same” image can occur in different appearances:



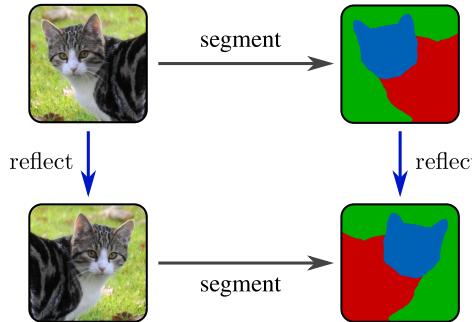
In this work, we are focusing specifically on *geometric transformations* of spatial signals, like, for instance, translations, rotations, reflections, scaling, or other affine transformations. Such transformations are mathematically described by *group actions* of the corresponding *symmetry group*.¹

¹An introduction to groups, actions, representations, and equivariant maps is found in Appendix B.

Being *invariant* to a geometric transformation means that the network's prediction does not change when its input is being acted on by the transformation group. This property is visualized by the following *commutative diagram*:



To motivate *equivariance*, consider an image segmentation task (pixel-wise classification) instead of image classification. The network's output should transform here in the same way as its input. The commutativity of the diagram below captures this requirement graphically:

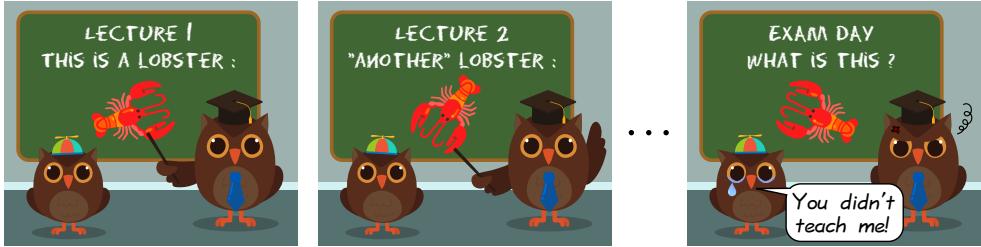


A function $f: X \rightarrow Y$ which commutes in such a way with group actions on its domain X and codomain Y is said to be *group equivariant*:

$$f \circ \text{action}_X = \text{action}_Y \circ f \quad \text{or, diagrammatically,} \quad \begin{array}{ccc} X & \xrightarrow{f} & Y \\ \text{action}_X \downarrow & & \downarrow \text{action}_Y \\ X & \xrightarrow{f} & Y \end{array}$$

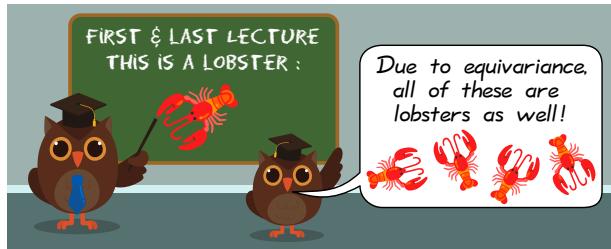
The visualized segmentation task is therefore reflection equivariant, in this specific case with the same reflection group action on its input and output. Probably somewhat surprisingly, the invariant classification task fits under the umbrella of equivariance as well. The difference here is that the label space is “transformed” by the trivial reflection group action ($\text{action}_Y = \text{identity}$), which acts by “doing nothing”, and may therefore be collapsed in the diagram. Invariant functions are hence a special case of equivariant functions, such that we can talk without loss of generality about equivariant functions only. For rigorous definitions of invariance and equivariance, and their mutual relation, we refer to Appendix B.4.

Why should we be interested in studying the equivariance properties of neural networks? In principle, a neural network would *learn* to be invariant or equivariant whenever this is desirable for the task it is being trained on. However, a naive network would have to learn this explicitly, that is, it would need to be shown samples in every possible geometric pose before it would understand their equivalence. This approach is clearly undesirable, as it leads to long training times and yields non-robust predictions.



(Vector graphics adapted under the Apache license 2.0 by courtesy of Google and the free license by courtesy of Freepik.)

A more sensible approach is to design the networks such that they are *by construction* constrained to be equivariant. Instead of having to learn over and over again how to process essentially the same image, such networks *automatically generalize their knowledge over all considered transformations*. Equivariance reduces the models' complexity and number of parameters, which frees learning capacity, accelerates the training process, and leads to an improved performance.



This work explains in a fairly general setting how to construct invariant and equivariant convolutional neural networks (equivariant CNNs).

Why do we need a *gauge theory* of neural networks ?

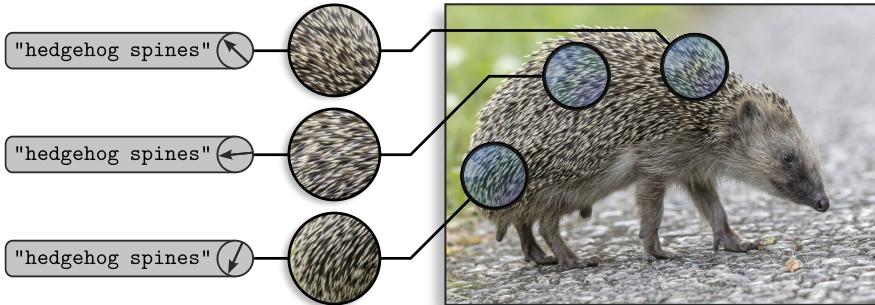
With respect to which specific group of transformations should a convolutional network be made equivariant? The answer to this question depends of course on the symmetries present in the distribution of data under consideration. Most obvious are *global transformations*, acting on signals as a whole.

For instance, aerial images do usually not exhibit a preferred origin or orientation. To process such data, it is hence reasonable to employ networks that are equivariant under the action of global isometries (i.e. translations, rotations and reflections of images).



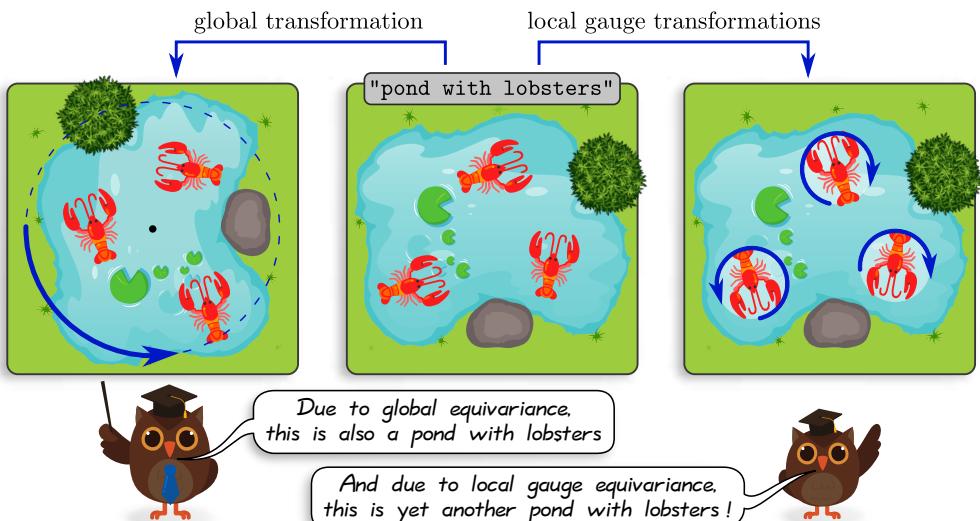
In sets of photos that are not taken from an aerial view, the gravitational field imposes a preferred directionality, breaking the rotational symmetry at the global scale. Networks for processing such images should hence only be equivariant under translations and reflections, but not under (global) rotations.

In both of these examples, we argued based on the signals' content at their *global scale*, however, the distribution of *local patterns* may differ from this analysis. To give an example, the image of the hedgehog below exhibits a preferred directionality at the global scale (hog, but hopefully not owl), while local patterns like the spines occur in arbitrarily rotated poses (all of may appear):



(Hedgehog adapted under the free license by courtesy of [Freepik](#).)

When processing such signals, the model should not only generalize everything it has learned over *global transformations* of the full signal (global equivariance), but also over *independent transformations of local patterns*, called “*gauge transformations*”.²



Our gauge theory of convolutional networks formalizes such local gauge transformations and explains how to construct gauge equivariant network layers. The neural activations (feature vectors) of gauge equivariant networks will thereby be guaranteed to encode the *same content* (e.g. “hedgehog spines”) in *different poses* (e.g.)³, while the features of non-equivariant networks would be entirely unrelated. We find empirically that local rotation and reflection equivariance is essentially always beneficial, since low level features like edges or corners usually appear in arbitrary orientations.

²Gauge transformations and gauge equivariant layers are more thoroughly introduced below.

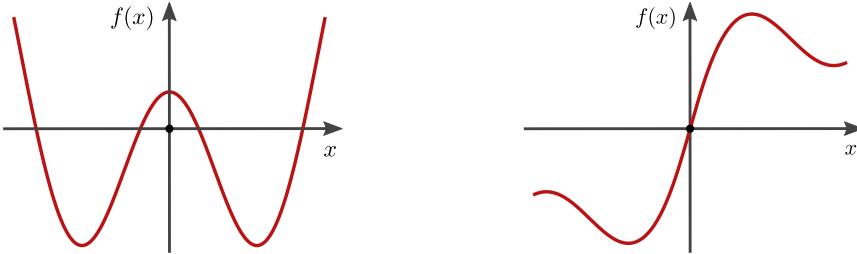
³Mathematically, such feature vectors with “poses” are elements of a *group representations space*.

How to construct equivariant models ?

How can we guarantee the equivariance of a machine learning model? As a conceptually simple example, let's consider a linear regression task (curve fitting), and assume that the ground truth $f : \mathbb{R} \rightarrow \mathbb{R}$ is either known to be *symmetric* or known to be *antisymmetric* (i.e. invariant or equivariant w.r.t. multiplications with -1)⁴:

$$\text{symmetric: } f(-x) = f(x)$$

$$\text{antisymmetric: } f(-x) = -f(x)$$



A naive (non-equivariant) approach would be to fit an ordinary polynomial

$$y(x) = \sum_{n=0}^N w_n x^n$$

of some degree $N \in \mathbb{N}$, where the $w_n \in \mathbb{R}$ are $N+1$ trainable parameters. However, this general polynomial model would ignore our prior knowledge on the functions' symmetry properties – they would have to be learned from the training data.

To take advantage of the symmetries, we can constrain the polynomial model to respect the symmetries a-priori. Doing so forces either the odd or even terms to zero, leaving us with *equivariant models*

$$y_{\text{symm}}(x) = \sum_{n \text{ even}}^N w_n x^n \quad (\text{since } (-x)^n = x^n \text{ for even } n \in \mathbb{N})$$

$$\text{and} \quad y_{\text{anti}}(x) = \sum_{n \text{ odd}}^N w_n x^n \quad (\text{since } (-x)^n = -(x^n) \text{ for odd } n \in \mathbb{N}),$$

which are *by design symmetric and antisymmetric*, respectively. Note that, in comparison to unconstrained polynomials, these equivariant models have approximately half the number of parameters. Furthermore, they generalize over reflections: having seen training data for $x > 0$ only, they are automatically fitted for all $x < 0$ as well.

On an abstract level, we started with some space of generic machine learning models, which is subsequently restricted to a *subspace of equivariant models*:

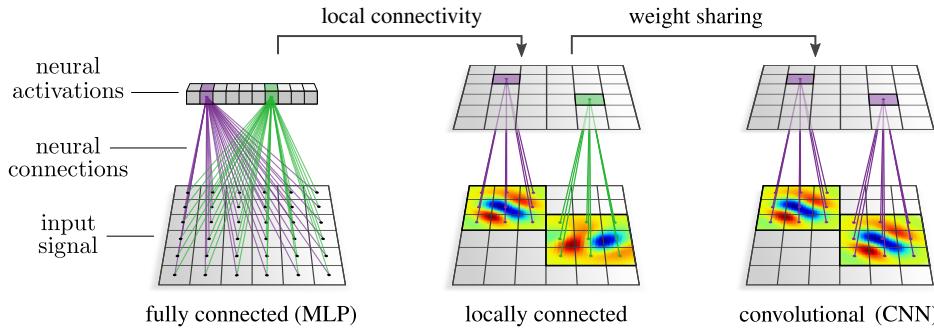
$$\begin{array}{ccc} \text{equivariant models} & \stackrel{\substack{\text{symmetry} \\ \text{constraint}}}{\subset} & \text{generic models} \end{array}$$

The layers of equivariant neural networks may similarly be defined as symmetry-constrained instances of common network operations (e.g. linear maps or bias summations). Specifically for the *translation group*, the subspaces of equivariant network operations correspond exactly to *convolutional network layers*.

⁴We encourage the reader to visualize the defining constraints $f(-x) = \pm f(x)$ of (anti)symmetric functions by drawing two commutative diagrams similar to those on page v. Hint: all nodes are \mathbb{R} , the blue transformation arrows are multiplications with -1 , and the gray function arrows are labeled by f .

Why convolutional neural networks ?

Convolutional neural networks (CNNs) are the standard network architecture for spatially structured signals. They differ from plain, fully connected networks in two respects: firstly, they usually have a *local neural connectivity*, and secondly, they *share synapse weights* (e.g. a convolution kernel) between different spatial locations:



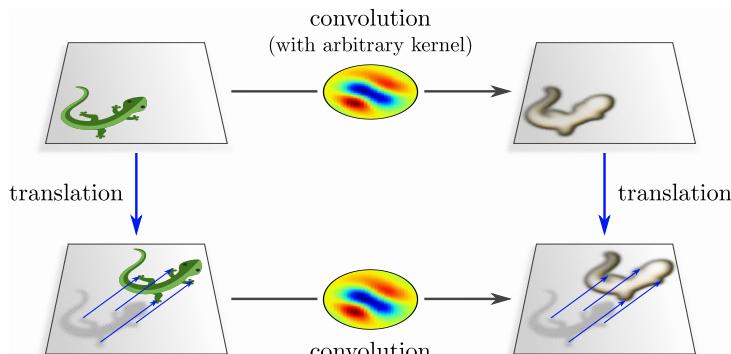
The weight sharing requirement applies to any type of operation employed in convolutional networks, enforcing, for instance, that one and the same bias vector or nonlinearity is to be used at every spatial position. CNNs owe their name to so-called convolution operations, which are exactly those linear maps that share weights. Intuitively, convolutions can be thought of as sliding a template pattern – the *convolution kernel* – across space, matching it at each single position with the signal to produce a response field.

Both the local connectivity and weight sharing reduce the number of model parameters (visualized by the number and color of the synapse weights in the graphics above), which makes CNNs less hungry for training data in comparison to fully connected networks. The local connectivity implies in addition that each neuron is associated with a specific spatial location (the center of its receptive field), such that they are naturally arranged in “feature maps”.

More important for us is, however, that the *spatial weight sharing implies the translation equivariance of convolutional networks*:

$$\text{spatial weight sharing} \implies \text{translation equivariance}$$

To see that this is indeed the case, note that any translation of a network’s input shifts patterns to other neurons’ receptive fields. Given that the neural connectivity is shared, these neurons are guaranteed to evoke the same responses as those at the previous location.



(Lizards adapted under the Creative Commons Attribution 4.0 International license by courtesy of Twitter.)

Instead of *defining* convolutional networks as sharing weights, and subsequently observing that they happen to be equivariant, we show in this work that one can reverse the implication arrow and *derive* weight sharing by demanding their equivariance:

$$\text{spatial weight sharing} \iff \text{translation equivariance}$$

We define conventional CNNs therefore equivalently as those neural networks that are translation equivariant.

Generalized equivariant CNNs & steerable kernels

The mutual implication “*weight sharing* \iff *translation equivariance*” suggests the generalization of convolutional networks by means of requiring their equivariance under extended groups of (global) transformations. A quite general family of symmetry groups, covering most practically relevant transformations of Euclidean space, are *affine groups* $\text{Aff}(G)$. They always contain *translations*, but augment them with additional transformations in the so-called *structure group* $G \leq \text{GL}(d)$.⁵ The additional transformations in G allow to model, for instance, rotations, reflections, or the scaling and shearing of signals.

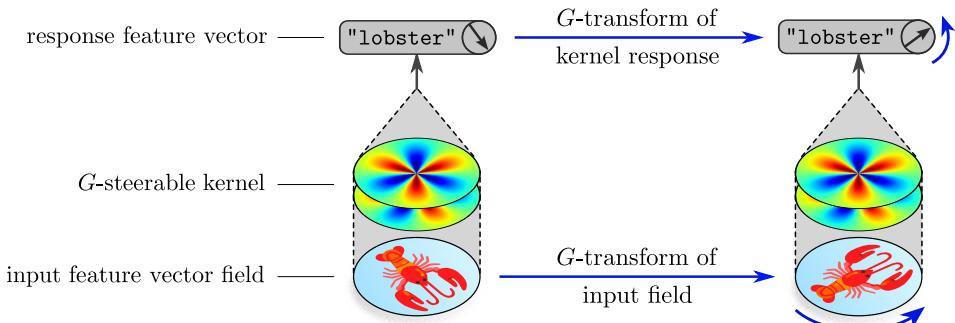
As for conventional CNNs, we prove a mutual implication between the networks’ equivariance and weight sharing, however, now over affine transformations:

$$\text{affine weight sharing} \iff \text{affine group equivariance}$$

Recall that affine groups consist of 1) translations and 2) additional G -transformations. The generalized affine weight sharing constraint splits accordingly into requirements for

1. spatial (or translational) weight sharing – just as for conventional CNNs – and
2. so-called “*G-steerability*” constraints on the shared neural connectivity.

The latter are additional G -equivariance requirements that any shared operation, e.g. any shared kernel, bias, or nonlinearity, needs to satisfy. For instance, a G -steerable kernel guarantees that any G -transformation of its input results in a corresponding G -transformation of its response feature vector.

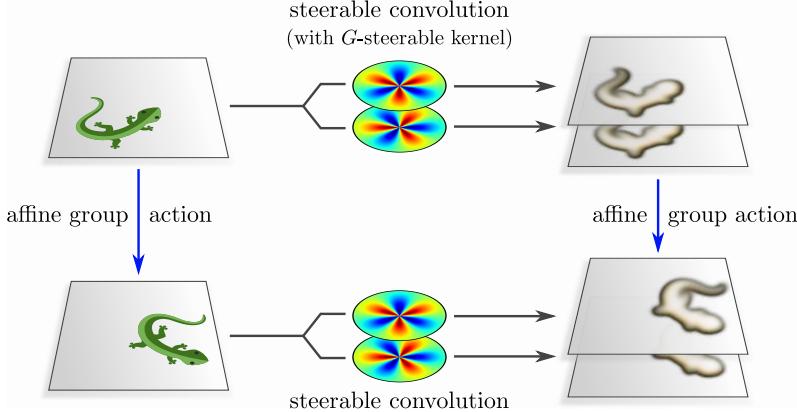


Since the responses are *vector valued features*, steerable kernels comprise in general multiple channels.⁶ Some clarifying examples follow shortly.

⁵Affine groups $\text{Aff}(G) := (\mathbb{R}^d, +) \rtimes G$ are a semidirect product \rtimes of translations $(\mathbb{R}^d, +)$ and some matrix subgroup $G \leq \text{GL}(d)$ of general linear transformations (e.g. rotations in $\text{SO}(d)$).

⁶Steerable kernels are actually matrix-valued, having $c_{\text{out}} \times c_{\text{in}}$ channels when their input is a feature vector field of c_{in} -dimensional feature vectors and their response is a c_{out} -dimensional feature vector.

Applied at one single location, a kernel produces a single response feature vector, as shown above. Convolutions apply the kernel at every point of space, and result therefore in a whole *feature vector field*. Since convolutions are translation equivariant, and steerable kernels are G -equivariant, any convolution with a steerable kernel is jointly translation and G -equivariant – they are hence $\text{Aff}(G)$ -equivariant, as desired.

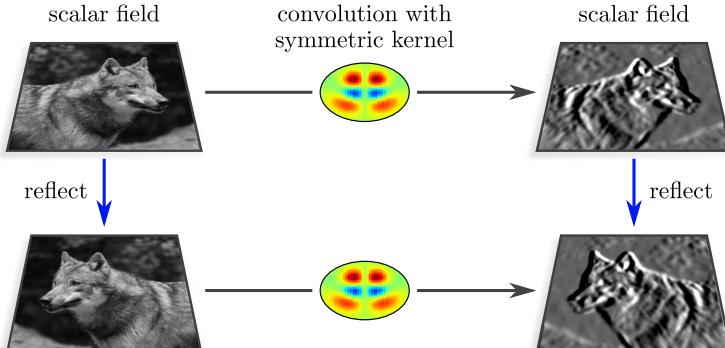


The main challenge in constructing $\text{Aff}(G)$ -equivariant CNNs is to solve for and to parameterize the subspaces of G -steerable convolution kernels. These subspaces are characterized by the specific G -symmetries that their constituent kernels have to satisfy. The particular details of the kernels' symmetries depend thereby on the choice of transformation laws (G -actions) according to which their input and output feature vectors are supposed to transform.

To build a more concrete intuition for steerable kernels, we turn to some specific examples.

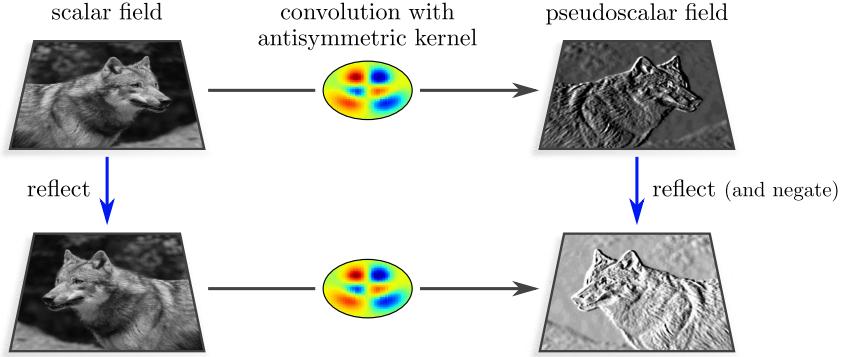
Reflection steerable kernels: The simplest non-trivial example is the one where G is the *reflection group*, such that $\text{Aff}(G)$ consists of translations and reflections.

Assume the convolution input to be a *scalar field*, modeling, for instance, a grayscale image. Let the convolution kernel be *symmetric*, i.e. invariant under reflections. When being applied to a reflected input such a kernel is guaranteed to produce exactly the same responses as for the original input, however, now located at spatially reflected positions. As the output field transforms in this example just like the input field, it is of scalar type as well.



Note that general, non-symmetric kernels would not satisfy this equivariance diagram. Their two response fields on the right-hand side would rather be mutually unrelated.

Next, let us consider *antisymmetric kernels*, which negate under reflections. Due to this property, the response field to a spatially reflected scalar input field will not only appear spatially reflected, but will *additionally change its sign*. This transformation behavior is a valid reflection group action, corresponding to so-called *pseudoscalar fields*.



In both examples, we assumed a specific input field type and reflection steerable kernels to be given, from which the output field type followed:

$$\text{input field type, steerable kernel} \implies \text{output field type}$$

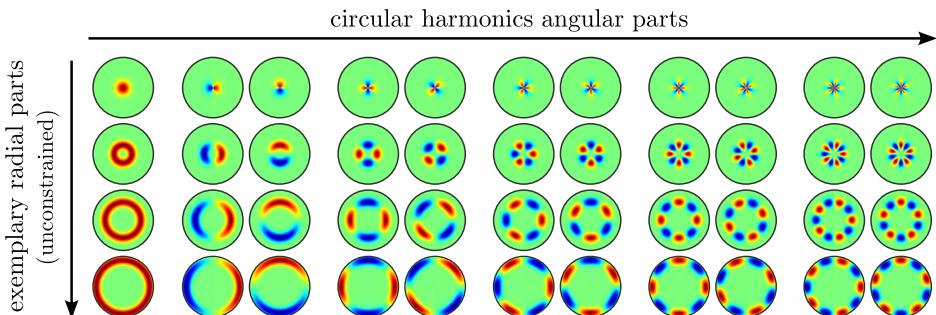
Our formulation of $\text{Aff}(G)$ -equivariant convolutions is, conversely, starting from the input and output field types, and derives subsequently the corresponding subspaces of steerable kernels that map equivariantly between them:

$$\text{input \& output field types} \implies \text{steerable kernels}$$

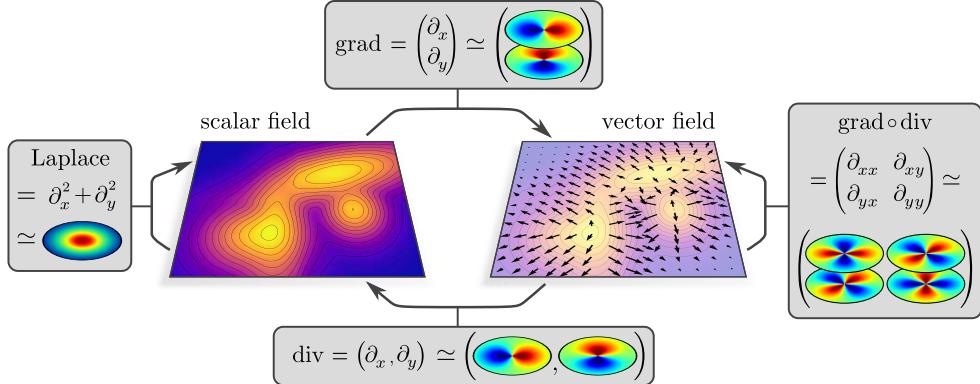
Specifically for scalar inputs and scalar or pseudoscalar outputs, these subspaces contain exactly symmetric and antisymmetric kernels, respectively.

Rotation steerable kernels: For our next example, consider the group $G = \text{SO}(2)$ of *rotations* in two dimensions. In general, $\text{SO}(2)$ -steerability imposes some rotational symmetry constraint on the kernels' angular part, but does not affect their radial part.

A natural basis to expand such kernels are *circular harmonics*, which are the Fourier basis of the kernels' angular parts in polar coordinates. This basis is *complete*, that is, any (including non-steerable) kernel can be constructed from it. $\text{SO}(2)$ -steerability constrains the basis to subsets of admissible angular frequencies, which depend on the specific types of feature fields considered.



As feature field types, the visualization below considers *scalar* and *vector fields*.⁷ In contrast to the previous examples, tangent vectors are not one-dimensional features, but comprise multiple channels. Steerable kernels for mapping between such field types are hence accordingly *matrix-valued*. They can be viewed as spatially extended counterparts of well-known differential operators, like the Laplacian (*scalar* \rightarrow *scalar*), gradient (*scalar* \rightarrow *vector*), divergence (*vector* \rightarrow *scalar*), or gradient of the divergence (*vector* \rightarrow *vector*).

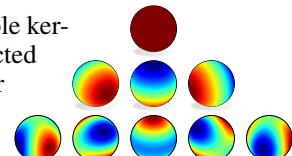


More general solutions than those visualized here exist: for instance, the mapping (*vector* \rightarrow *vector*) allows in addition for rotationally invariant kernels, and the radial parts are only exemplary and get optimized during the training process. More expressive steerable kernel spaces, which allow for higher circular harmonic frequencies and the mixing of different frequencies, require more complex field types, e.g. general tensor fields. A full overview of the complete solution spaces of $SO(2)$ -steerable kernels is given in Table 5.2.

Note that the resemblance to common differential operators in the example above is no coincidence – it is a direct consequence of the rotational symmetries of the laws of nature, which necessitate *steerable partial differential operators* [137].

General steerable kernels: Although they are often formulated differently, *any* equivariant convolutional network assumes some types of feature fields and applies corresponding G -steerable kernels. Our contribution is to present a *unified formulation*, which describes all of these models in a single coherent framework.

We furthermore found a complete characterization of G -steerable kernel spaces, which explains in general how they may be constructed from *harmonic basis functions* of G . For instance, the angular parts of $SO(3)$ -steerable kernels will always be assembled from (subsets of) spherical harmonics, shown on the right.



Gauge equivariance: As mentioned above, G -steerability ensures kernel responses to transform in a G -equivariant manner with their input. Recall that CNNs usually have a *local connectivity*, i.e. apply kernels of finite spatial extent. Convolutions with G -steerable kernels that have such narrow receptive fields are automatically *gauge equivariant* w.r.t. independent G -valued *gauge transformations* of the kernels' fields of view at different locations.

This claim is intuitively plausible, however, to put it on a formal basis, we need to introduce the gauge theoretic framework of *coordinate independent CNNs*, which we do next.

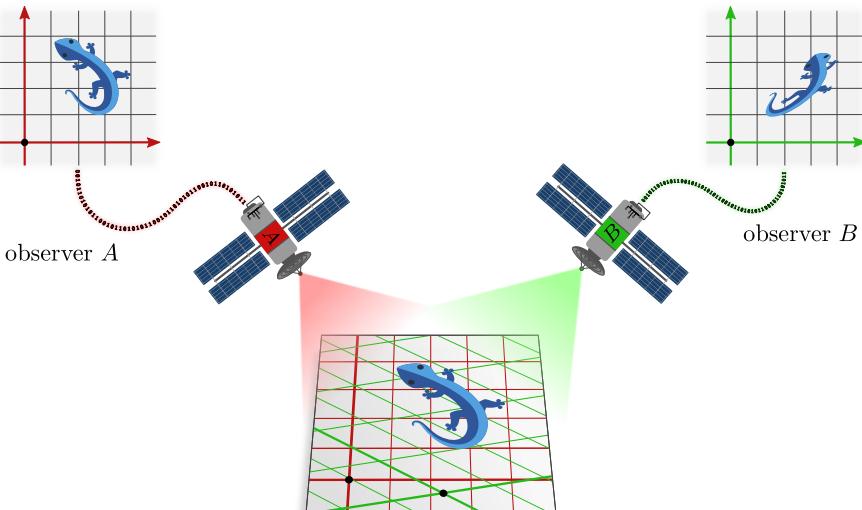
⁷Note that this diagram is in contrast to the previous ones *not commutative*. For instance, div is not the inverse of grad.

Why *coordinate independent* neural networks?

So far, we have only considered *active transformations* of signals. *Passive transformations* do not act on signals themselves, but rather on the *coordinate systems* used to describe them:



How is this relevant for equivariant networks? Equivariant CNNs are guaranteed to respond equivalently to *actively* transformed variants of the same signal. However, the perception of a network depends only on its *relative* alignment towards the signal. We can therefore equivalently think about keeping the signal fixed, and *passively* transforming the network's viewpoint. A network's viewpoint can be identified with its internal *frame of reference*, which justifies our interpretation as passive coordinate transformations.⁸



(Lizards adapted under the Creative Commons Attribution 4.0 International license by courtesy of Twitter.)

The fact that active and passive transformations are indistinguishable from the observer's viewpoint implies that the predictions of equivariant networks transform in a well defined manner when passively changing their viewpoint – equivariant networks and their features are hence *coordinate independent*, i.e. obey the *principle of covariance*. Conversely, a requirement for equivariance follows when demanding coordinate independence.⁹

The equivalence of the active and passive interpretation holds not only for *global coordinate charts*, but also for *local frames of reference* (gauges), which we leverage next to formalize the networks' local gauge equivariance. This local viewpoint becomes actually strictly necessary when extending CNNs to general manifolds, as they do in general neither come with, nor admit global coordinates.

⁸The network takes on the same role as an *observer* in special relativity.

⁹This statement assumes crucially that the network is co-moving with frames and their passive coordinate transformations. It would not be true if the network was held in fixed relation to the signal.

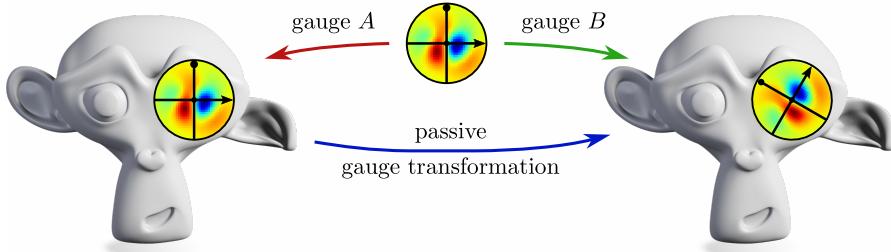
Manifolds, local frames & gauge symmetries: This work covers not only CNNs on Euclidean spaces, but generalizes them to arbitrary *Riemannian manifolds*, including, for instance, spheres or other curved surfaces, as shown below.

On Euclidean spaces, we defined convolutional networks by demanding their equivariance and leveraging the implication “*weight sharing \Leftarrow equivariance*”. However, manifolds are in general *asymmetric*, that is, there exist in general no (global) transformations w.r.t. which we could demand equivariance, such that this strategy does no longer apply. Instead, we define CNNs on manifolds immediately as coordinate independent networks with spatially shared synapse weights. It turns out that this requires, once again, equivariance constraints (G -steerability constraints), however, now in a local *gauge theoretic* sense.

To motivate gauges and local gauge transformations, assume you are given a convolution kernel, which should be applied at each spatial location of a manifold (weight sharing). In contrast to Euclidean spaces, it is unclear how to do this, since there exists in general *no preferred reference direction* (e.g. rotation or reflection) along which the kernel should be aligned. A specific choice of alignment, which can be identified with a choice of reference frame, is what we call a “*gauge*”:¹⁰

$$\text{geometric kernel alignment} \equiv \text{choice of local reference frame} \equiv \text{gauge}$$

Local gauge transformations are, accordingly, passive transformations between reference frames and kernel alignments.



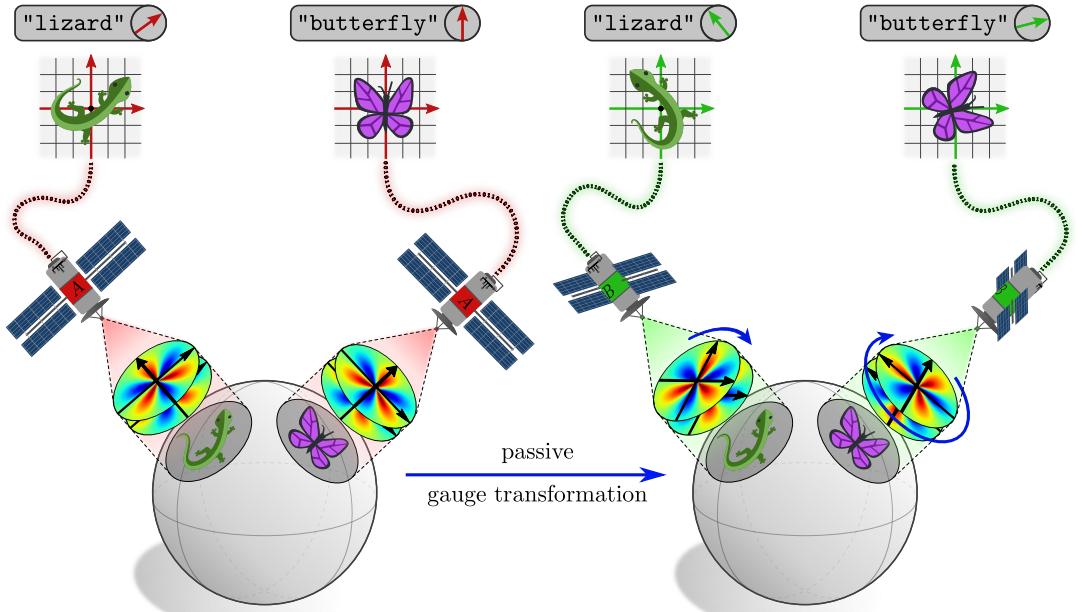
Above, we interpreted *neural networks* – as a whole – as *global observers*, whose geometric alignment relative to the signal was identified with a global coordinate chart. Their coordinate independence (viewpoint independence) was in one-to-one relation to their *equivariance under global coordinate transformations*.

In the gauge theoretic interpretation, we are instead viewing *every single neuron* – whose synapses (e.g. kernels) form a local sub-network – as an *independent local observer*. Their geometric alignment is described by a local reference frame (gauge) of the corresponding tangent space. Demanding the network’s (local) coordinate independence requires the shared kernel’s *equivariance under local gauge transformations*. This gauge equivariance constraint is exactly the G -steerability constraint discussed above.

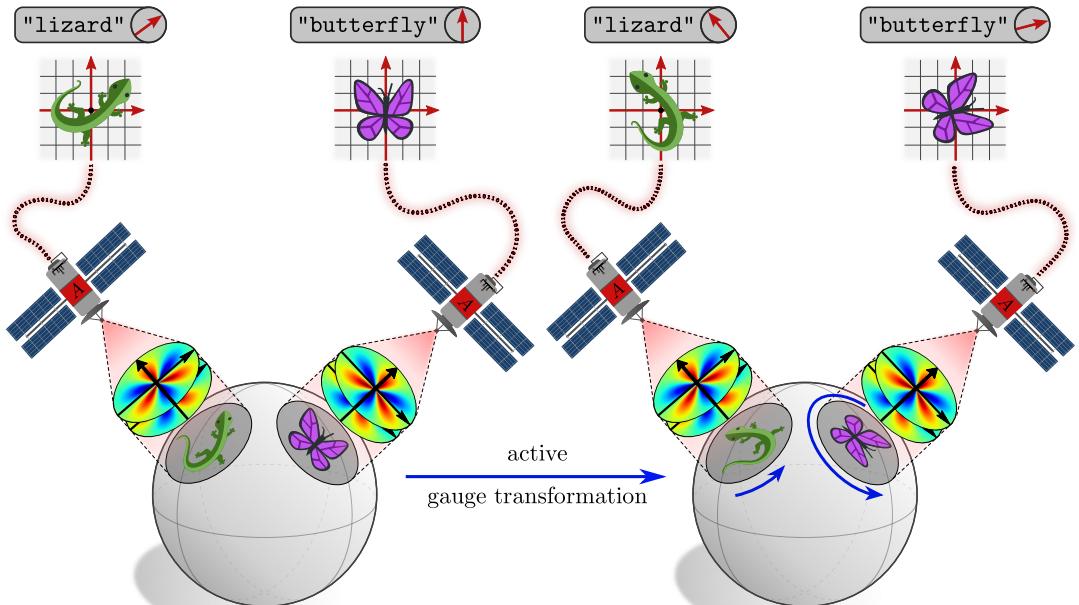
$$\text{local coordinate independence} \iff \text{gauge steerable kernels}$$

Intuitively, a kernel’s steerability regulates G -ambiguities of reference directions by guaranteeing that their responses in different alignments differ merely by some G -transformation. Steerable kernel responses therefore encode both the *gauge independent content* of the signal in their field of view, and its *geometric pose relative to the chosen gauge*.

¹⁰Mathematically, gauges are formalized as local trivializations of the tangent bundle.

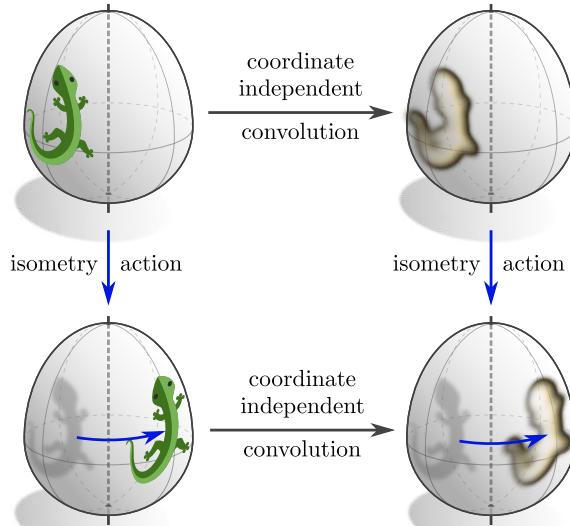


The active counterpart of passive local gauge transformations are *independent active transformations of signal patches* in each kernel’s local field of view (keeping the gauges fixed). As in the passive case, the kernel’s G -steerability ensures that their responses transform equivariantly, i.e. encode the same content in a different pose.



(Lizards and butterflies adapted under the Creative Commons Attribution 4.0 International license by courtesy of Twitter.)

As mentioned above, we can in general not assume Riemannian manifolds to have non-trivial symmetries (called isometries). However, if a manifold happens to have a non-trivial isometry group, it is natural to ask for the network's isometry equivariance, i.e. the property to commute with isometry group actions on signals. Our coordinate independent CNNs satisfy this property by construction (the details depend on the choice of structure group G).



(Lizards adapted under the Creative Commons Attribution 4.0 International license by courtesy of Twitter.)

This result closes the loop to our Euclidean steerable CNNs above, which were defined by requiring the networks' global transformation equivariance. In fact, the gauge theoretic formulation of coordinate independent CNNs reduces on Euclidean spaces to steerable CNNs.¹¹

Theoretical and practical significance

The primary theoretical contribution of the gauge theory of “Equivariant and Coordinate Independent CNNs” is that it represents a *unified theory of convolutional networks*, which is capable of explaining many independently proposed CNN architectures in one single framework. To substantiate this claim, we review more than 100 different models from the literature, showing for each single one how it can be viewed as a specific instantiation of our general formulation. An overview of these models – ranging from Euclidean CNNs over polar, spherical, icosahedral and Möbius CNNs to convolutional networks on general surfaces and meshes – is given in Table 14.1 on page 273.

The reformulation of these networks as coordinate independent CNNs does not only clarify their mutual relations, but implies in addition that *all of them are actually equivariant under local gauge transformations*. This insight is remarkable, since these models were usually conceived with solely global transformations in mind.

Applications: Above, we were mainly considering image processing tasks as exemplary applications, however, coordinate independent CNNs are applicable to arbitrary types of signals or fields (including e.g. vector or tensor fields) on any Riemannian manifold.

¹¹The networks are on Euclidean spaces not only isometry equivariant, but $\text{Aff}(G)$ -equivariant.

Classically, equivariant networks were proposed for applications with *global symmetries*, where they were shown to be up to $10\times$ more data efficient than non-equivariant models [329]. They are of particular importance for *biomedical imaging tasks*, where training data is typically scarce and accurate results are of greatest societal relevance.



A probably less expected result is that equivariant CNNs outperform non-equivariant models even in applications without global symmetries. For instance, we find that replacing conventional CNN layers with their rotation steerable counterparts reduces the error rate in image classification tasks with globally preferred directionality by up to 25%; see Section 6.4.



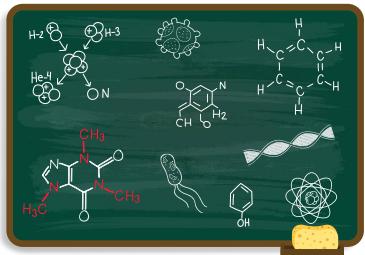
The filter bank on the left shows convolution kernels of a non-steerable CNN which was trained in a similar setting. It is apparent that the kernels learned to respond to oriented patterns in any possible direction – gauge equivariant layers incorporate such symmetries of the neural connectivity by design, alleviating the model from the burden of having to learn it explicitly.

Our theory explains CNNs on manifolds of *arbitrary dimensionality* d . For instance, audio signals are sampled on $d = 1$ time axis. Translation equivariance makes the networks generalize over different points in time, while scale equivariance makes them generalize over (local) pitch-shifts. Typical examples for $d = 3$ dimensions are videos, point clouds, or voxel images like fMRI scans of the brain. Specifically for the latter, one could alternatively think about processing signals directly on the $d = 2$ -dimensional curved surface of the cerebral cortex. Coordinate independent CNNs would describe all of these approaches in a common language.



Typical examples for $d = 3$ dimensions are videos, point clouds, or voxel images like fMRI scans of the brain. Specifically for the latter, one could alternatively think about processing signals directly on the $d = 2$ -dimensional curved surface of the cerebral cortex. Coordinate independent CNNs would describe all of these approaches in a common language.

A particularly exciting new application area is “*AI for science*”, which employs neural networks as computational tools in the physical and life sciences. The relevant laws of nature are usually characterized by a rich set of a-priori known symmetries (e.g. Lorentz invariance), which should be incorporated into the models. Once again, gauge symmetries play an important role: for instance, local patterns in molecules, like the highlighted $\cdots\text{N}-\text{CH}_3$, reappear multiple times in different poses.



Equivariant CNNs have also proven useful in *reinforcement learning*. For example, the game board and rules of “Go” are invariant under reflections and rotations by $\pi/2$. This implies a corresponding equivariance requirement on the learned policy in the sense that a transformed state should result in accordingly transformed actions. Our implementation of steerable CNNs has already been used for multiple reinforcement learning and robotic control tasks [216, 314, 360, 312, 129, 138, 188, 313, 315].

As these examples demonstrate, there is a wide variety of applications that rely on processing spatial signals. Symmetries, whether global or local, abound in such tasks. In all of these cases, “*Equivariant and Coordinate Independent CNNs*” explain how convolutional networks can be constructed, and how the tasks’ symmetries can be taken into consideration. The unified formulation enables thereby to transfer insights between different applications.

(Graphics on this page adapted from Kim et al. [149], Krizhevsky et al. [166], Winkler et al. [330], Freepik, and Rabich [237].)

Contents

1	Introduction	1
1.0	Introduction	1
1.1	Outline	15
I	Equivariant Convolutional Networks on Euclidean Spaces	21
2	Invariant and equivariant models	25
2.1	Equivariant machine learning and quotient hypothesis spaces	25
2.1.1	Invariant and equivariant target functions	25
2.1.2	Quotient hypothesis spaces	26
2.2	Equivariant neural networks	28
3	Translation equivariance & conventional Euclidean CNNs	33
3.1	Euclidean feature maps as translation group representations	33
3.2	Translation equivariant layers and convolutions	35
3.2.1	Convolutions as translation equivariant linear maps	36
3.2.2	Translation equivariant bias summation	38
3.2.3	Translation equivariant local nonlinearities	38
3.2.4	Translation equivariant local pooling operations	39
3.2.5	Translation invariant global pooling operations	41
4	Affine group equivariance & steerable Euclidean CNNs	43
4.1	Affine groups	44
4.2	Euclidean feature fields as induced affine group representations	46
4.3	Affine equivariant layers and steerable convolutions	49
4.3.1	Steerable convolutions as affine equivariant linear maps	49
4.3.2	Steerable bias summation	54
4.3.3	Steerable local nonlinearities	55
4.3.4	Steerable local pooling operations	57
4.3.5	Steerable global pooling operations	59
4.4	Local symmetries and equivariance group restriction	61
4.5	Affine group convolutions as regular steerable convolutions	63

5 G-steerable convolution kernels	67
5.1 The vector space of G -steerable kernels and steerable basis expansion	68
5.2 Simple example – reflection steerable kernels	70
5.2.1 Reflection steerable feature fields	71
5.2.2 Reflection steerable convolutions	72
5.3 A generalized Wigner-Eckart theorem for G -steerable kernels	76
5.3.4 SO(2) and O(2)-steerable kernels	80
5.3.5 Kernel sampling and anti-aliasing:	85
5.4 Alternative approaches to construct steerable kernels	86
6 Empirical evaluation of steerable CNNs	89
6.1 Generalization over group orbits	92
6.2 Data efficiency and convergence	94
6.3 Choice of equivariance group and group restriction	95
6.4 Drop-in replacement for conventional convolutions and natural image datasets	98
6.5 Field type, nonlinearity and symmetry group benchmarking	100
II An Introduction to Coordinate Independent CNNs	109
7 Gauges, gauge transformations and G-structures	113
7.1 Tangent spaces and reference frames	113
7.2 Coordinate independent functions on tangent spaces	118
7.3 Structure groups, G -structures and G -atlases	120
8 Coordinate independent feature vector fields	123
8.1 Gauge transformations of feature vectors	123
8.2 Parallel transport of feature vectors	127
8.2.1 Tangent vector transporters	127
8.2.2 Feature vector transporters	129
8.2.3 Compatibility of connections and G -structures	129
8.3 Isometry actions and induced gauge transformations	131
8.3.1 Pushforward of tangent vectors	131
8.3.2 Pushforward of reference frames and symmetries of the G -structure	132
8.3.3 Pushforward of feature vectors	133
9 Coordinate independent networks and GM-convolutions	135
9.1 Pointwise gauge equivariant operations	136
9.1.1 Gauge equivariant 1×1 -convolutions	136
9.1.2 Gauge equivariant bias summation	138
9.1.3 Gauge equivariant nonlinearities	140
9.2 Kernel field transforms and GM -convolutions	140
9.2.1 A local observer's view on feature fields	141
9.2.2 Coordinate independent kernels and kernel field transforms	143
9.2.3 GM -convolutions and G -steerable kernels	146
9.3 Isometry equivariance	151

10 Reflection steerable Möbius CNNs	155
10.1 Geometry of the Möbius strip	155
10.2 Orientation independent feature fields	158
10.3 Orientation independent convolutional networks	159
10.3.1 Orientation independent bias summation	160
10.3.2 Orientation independent nonlinearities	161
10.3.3 Orientation independent convolutions	161
10.3.4 General isometry equivariant kernel field transforms on the Möbius strip	162
10.4 Numerical implementation of Möbius convolutions	163
10.5 Empirical evaluation of Möbius convolutions	167
III Fiber Bundle Theory of Coordinate Independent CNNs	173
11 Associated bundles and coordinate free feature fields	177
11.1 A brief introduction to fiber bundles	177
11.1.1 Fiber bundles in general	178
11.1.2 Vector bundles	180
11.1.3 G -bundles	181
11.1.4 Associated G -bundles	182
11.1.5 Principal G -bundles	182
11.1.6 Sections and fields	183
11.1.7 Bundle morphisms	184
11.2 The tangent bundle TM and frame bundle FM	185
11.2.1 Tangent bundle TM	186
11.2.2 Frame bundle FM	187
11.2.3 TM as $GL(d)$ -associated vector bundle $(FM \times \mathbb{R}^d)/GL(d)$	187
11.3 G -structures GM and associated feature vector bundles \mathcal{A}	189
11.3.1 G -structures GM	189
11.3.2 TM as G -associated vector bundle $(GM \times \mathbb{R}^d)/G$	192
11.3.3 Associated feature vector bundles \mathcal{A}	192
11.3.4 Associated feature vector field and feature spaces	193
11.4 Local bundle trivializations of TM , FM , GM and \mathcal{A}	193
11.4.1 Trivializations of TM	194
11.4.2 Induced trivializations of FM and frame fields	194
11.4.3 G -atlas induced G -structure GM	198
11.4.4 Induced trivializations of associated bundles \mathcal{A}	199
11.4.5 Summarizing remarks	201
11.5 Parallel transporters on associated bundles	201
11.5.1 Transport on TM	202
11.5.2 Transport on FM	202
11.5.3 Compatibility of connections and G -structures	203
11.5.4 Transport on GM	204
11.5.5 Transport on \mathcal{A}	205
12 Coordinate free formulation of kernel field transforms and GM-convolutions	207
12.1 $1 \times 1 GM$ -convolutions	208
12.1.1 $1 \times 1 GM$ -convolutions as vector bundle M -morphisms	208
12.1.2 $1 \times 1 GM$ -convolutions as homomorphism bundle sections	210
12.2 Kernel field transforms and GM -convolutions	214

12.2.1 Kernel fields	214
12.2.2 Kernel field transforms and GM -convolutions	218
12.2.3 Kernel field transforms and GM -convolutions in local coordinates	220
13 Isometry equivariance	225
13.1 Isometries and their action on manifolds, bundles and fields	226
13.1.1 Isometry groups	226
13.1.2 Isometry action on fiber bundles	228
13.1.3 Isometry action in local coordinates	234
13.1.4 Commutativity of isometry actions with the exponential map and transporters	239
13.2 Isometry equivariance of kernel field transforms and GM -convolutions	241
13.2.1 Isometry equivariance of general kernel field transforms	242
13.2.2 Isometry equivariance of GM -convolutions	245
13.3 Quotient kernel fields	248
13.3.1 Isometry induced quotient spaces	249
13.3.2 Quotient representative kernel fields and stabilizer constraints	253
IV Applications & Literature Review	263
14 Design choices and overview	267
15 Euclidean coordinate independent CNNs	275
15.1 Affine geometry of Euclidean spaces \mathbb{E}_d	276
15.1.1 Affine charts and $\text{Aff}(G)$ -atlases	277
15.1.2 Induced G -atlases and G -structures	278
15.1.3 Coordinate free affine transformations	281
15.2 Affine group equivariant CNNs on Euclidean spaces \mathbb{E}_d	284
15.2.1 Recovering steerable convolutions on \mathbb{R}^d	284
15.2.2 Affine chart independence	286
15.2.3 Affine group equivariance	287
15.3 Euclidean CNNs in the literature	289
16 Rotation equivariant CNNs on punctured Euclidean spaces	295
16.1 Global rotation equivariance on $\mathbb{E}_2 \setminus \{0\}$	296
16.2 Global rotation and scale equivariance on $\mathbb{E}_2 \setminus \{0\}$ via log-polar coordinates	297
16.3 Global rotation equivariance on $\mathbb{E}_3 \setminus \{0\}$	301
17 Spherical coordinate independent CNNs	305
17.1 Geometry of the 2-sphere S^2	306
17.2 Fully rotation equivariant spherical CNNs	311
17.3 Azimuthal rotation equivariant spherical CNNs on cylindrical topologies	318
17.4 Icosahedral approximations of spherical CNNs	324
18 Coordinate independent CNNs on general surfaces	331
18.1 Geometry of embedded surfaces	333
18.1.1 Classical differential geometry of embedded surfaces	333
18.1.2 Discretized geometry of surface meshes	336
18.2 Rotation-steerable surface convolutions	343
18.3 $\{e\}$ -steerable surface convolutions	353

V Appendix	361
A List of theorems and definitions	363
B Groups, representations and equivariant maps	367
B.1 Symmetry groups – basic definitions	367
B.2 Subgroups and products of groups	370
B.3 Group actions, orbits and quotient spaces	372
B.4 Invariant and equivariant maps	376
B.5 Group representations and intertwiner maps	378
C Coordinate chart formalism of differential geometry	387
C.1 Tangent spaces, cotangent spaces and dual bases	388
C.2 Differentials, gradients and Jacobians	389
C.3 Chart induced coordinate bases	391
C.4 Coordinate bases as local bundle trivializations	396
C.5 G -structures and vielbein fields	399
D Integration over tangent spaces	403
E Equivariant MLPs	405
F Equivariant convolutions on homogeneous spaces	409
F.1 Homogeneous spaces, group convolutions and group correlations	411
F.2 Scalar field convolutions on homogeneous spaces	413
F.3 Steerable CNNs on homogeneous spaces	416
G Coordinate independent weight sharing	427
H An intuition for the Wigner-Eckart theorem for steerable kernels	429
H.1 Harmonic kernels	430
H.2 Irrep endomorphisms	430
H.3 General irrep steerable kernels and Clebsch Gordan coefficients	432
I Existence and smoothness of kernel field transforms	435
J Regular feature fields as scalar functions on G-structure	439
K Quotient representative kernel fields – proofs	443
L Spherical convolutions as GM-convolutions – proofs	451
M Research questions & conclusions	457
Bibliography	477

Introduction

Computational methods became in the recent decades more and more relevant for a wide range of applications. The underlying algorithms are classically hardcoded, that is, the programmer specifies explicitly how the algorithms process data. While this approach is suitable for tasks with tightly controlled input and output spaces, it becomes quickly infeasible for more complicated tasks like computer vision or speech recognition, where the data to process exhibits substantial variability. Machine learning algorithms aim to resolve this issue by replacing hardcoded algorithms with adaptive models that are fitted to data.

While the machine learning paradigm takes the burden of hardcoding an algorithm from the programmer, the issue with tasks of increasing complexity remains – it manifests here in an increased demand for training data, which quickly becomes infeasible as well. A large part of machine learning research is focused on easing this issue by incorporating prior knowledge about the learning task into the machine learning model. One of the arguably most successful approaches is that of *group equivariant models*. Equivariant learning algorithms hardcode *symmetry properties* or *invariances* of the learning task directly into the space of models to be optimized over, which greatly enhances their data efficiency.

A prototypical example of this design principle are *convolutional neural networks* (CNNs) [175, 166]. Conventional CNNs process signals on Euclidean spaces – for instance images – and exploit their spatial structure via a *local neural connectivity* with *spatially shared synapse weights*. Since the same convolution kernel (neural connectivity) is applied at each point in space, convolutions are *translation equivariant maps* – any translation of their input

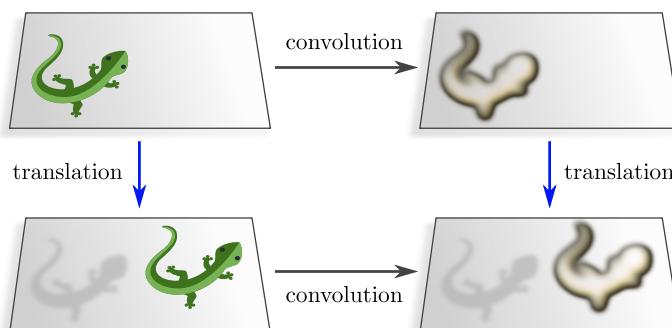


Figure 1.1: Convolutions are *translation equivariant maps*, i.e. they commute with translation group actions on their input and output feature maps. When a convolutional network learned how to process patterns at one specific location, it is guaranteed to generalize this knowledge to any other location.

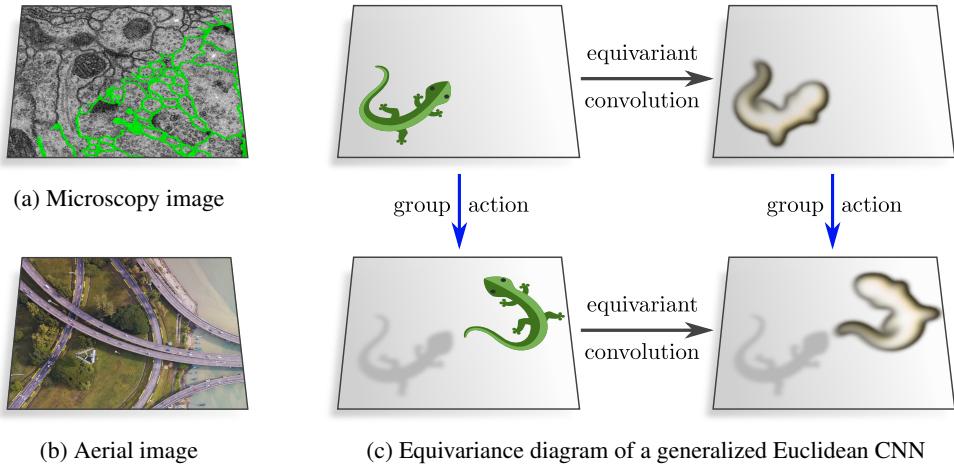


Figure 1.2: *Left:* Microscopy and aerial images are examples of signals that do not exhibit a preferred notion of directionality. Patterns like cell boundaries or streets occur therefore not only at different locations, but also in different rotations and reflections. *Right:* Neural networks processing such data should generalize their inference over these additional geometric transformations. Formally, this is captured by demanding that they are equivariant under (commute with) the extended group of symmetries. Conventional CNNs are merely translations equivariant, but disregard other transformations.

(Lizards adapted under the Creative Commons Attribution 4.0 International license by courtesy of Twitter.)

results in a corresponding translation of their output, as visualized by the commutative diagram in Fig. 1.1. As a consequence, *convolutional networks generalize their inference automatically over spatial positions*, that is, they do not explicitly need to relearn how to process a given pattern when it reappears at a different location. Due to their improved data efficiency and robustness, convolutional networks are nowadays de-facto the standard models for processing spatially structured data like audio, images, volumetric signals or videos.

Given the considerable empirical success of conventional Euclidean CNNs, there is a great interest in extending convolutional models to

- 1. be equivariant under *larger symmetry groups*, and ▷ Fig. 1.2
- 2. process signals on *more general domains*. ▷ Fig. 1.4

This work presents a *gauge theory* of *Equivariant and Coordinate Independent Convolutional Networks* on Riemannian manifolds which addresses both of these points. To offer an easy entry, and following the historic development, Part I begins by introducing equivariant convolutions on Euclidean spaces (“steerable CNNs”), before Parts II and III develop the full differential geometric formulation. Part IV demonstrates the generality of this formulation by explaining a vast array of convolutional networks from the literature as specific instantiations of coordinate independent CNNs; see Table 14.1 on page 273.

The remainder of this introductory section is accordingly split into a high level overview of the Euclidean and Riemannian formulations of coordinate independent CNNs, and applications thereof. A detailed outline of this work’s contents follows on page 15. The preface offers a less technical but more visual introduction and motivation.

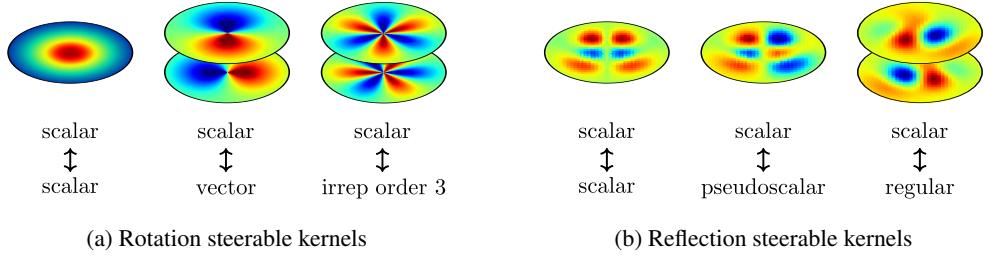


Figure 1.3: Simple examples of *steerable convolution kernels*, which guarantee an equivariant mapping between *feature fields* of different types (i.e. with different transformation laws). For simplicity, all examples assume scalar input fields, such that the multiplicity of kernel channels is determined by the output feature vector dimensionality. *Left:* SO(2)-steerable kernels obey a rotation equivariance constraint on their angular part; their radial part is unconstrained. A mapping from scalar to scalar fields requires isotropic kernels, while a mapping to vector fields requires a circular harmonic angular part of frequency one. Higher order circular harmonics result in feature vectors that transform according to higher order irreducible representations of SO(2). Note that the first two examples correspond in the infinitesimal limit to the Laplace and gradient operator [137]. *Right:* Reflection steerable kernels exhibit some kind of reflectional symmetry. Symmetric and antisymmetric kernels map scalar input fields to scalar or pseudoscalar output fields, respectively, the latter defined by a negation of their sign under reflections. A mapping to fields transforming according to the two-dimensional regular representation of the reflection group is performed by a convolution with a kernel whose two channels are reflected copies of each other. Regular representations explain the widely used group convolutions [52, 162] from a representation theoretic perspective; see Theorem 4.5.

Equivariant CNNs on Euclidean spaces

▷ Part I

In many applications, characteristic patterns in the signal appear not only at arbitrary locations, but also in *arbitrary directions or scale*; see for instance Figs. 1.2a and 1.2b. While generalizing over translations, conventional CNNs disregard such additional geometric factors of variation, that is, they need to explicitly learn over and over again how to process a given pattern in any single geometric pose it may appear in. Quite some effort has been made to alleviate this shortcoming by *extending the equivariance properties of convolutional networks to larger symmetry groups*, as visualized in Fig. 1.2c. Such generalized equivariant CNNs are guaranteed to share their inference over the extended group of symmetries, which makes them even more data efficient and robust as compared to conventional CNNs.

Motivation & overview

In recent years, the research community made major progress in developing equivariant Euclidean CNNs and demonstrated their superior empirical performance. However, while a plethora of equivariant model architectures has been proposed, most publications considered very specific settings and came up with their own nomenclature, notation and formulation, making it increasingly hard to keep an overview and understand how different models relate to each other. The first part of this work presents the theory of *Euclidean steerable CNNs* [53, 323, 322, 173, 137, 40], which unifies many of the proposed models in a common *representation theoretic* framework. Different architectures are shown to differ mainly in the considered *symmetry groups* and *feature field types*, i.e. the specific transformation laws (group representations) of their feature spaces. A geometrically consistent mapping between fields of different types is guaranteed by some form of equivariant convolution operation. We show that all of these operations can be abstractly described as conventional convolutions with symmetry-constrained “*group steerable*” kernels, some examples of which are shown in Fig. 1.3.

Affine groups	To cover a wide range of settings, we consider <i>affine groups</i> $\text{Aff}(G) := (\mathbb{R}^d, +) \rtimes G$, which consist of translations $(\mathbb{R}^d, +)$ of \mathbb{R}^d and transformations in some choice of <i>structure group</i> (or stabilizer subgroup) $G \leq \text{GL}(d)$. ¹ The latter describes for instance rotations, reflections, shearing or scaling, and controls the desired level of equivariance. Conventional CNNs, which are merely translation equivariant, are covered for the trivial group $G = \{e\}$.
Feature fields	The feature spaces of steerable CNNs consist of <i>feature vector fields</i> , i.e. fields of feature vectors that are attached to each point of Euclidean space. Feature fields differ from conventional feature maps in that they are equipped with an <i>affine group action</i> . The specifics of the action depend on their <i>field type</i> – common examples are scalar, vector or tensor fields, visualized in Fig. 4.3, or pseudoscalar and regular feature fields, shown in Fig. 5.1. Formally, field types are G -representations ρ (Def. B.5.1), and the feature spaces (spaces of feature fields) are $\text{Aff}(G)$ -representations $\text{Ind}_G^{\text{Aff}(G)} \rho$ that are induced from the field types (Def. 4.2.1).
Equivariant layers	Any layer (network operation) of a steerable CNN is required to be equivariant, which means that it needs to ensure that any transformation of its input field results in a corresponding transformation of its output field. A common approach to introduce novel equivariant network layers is to first specify their function definition, and subsequently prove their equivariance. The issue with this approach is that the operations are found heuristically instead of being <i>derived</i> from first principles, which makes it hard to assess their generality and does not cast light on how further equivariant operations could be found. We adopt therefore a different strategy, which proceeds instead by 1) assuming a flexible <i>ansatz</i> for the layer (e.g. linear maps), 2) fixing desired transformation laws (field types) of its input and output, and 3) solving the implied equivariance constraint on the ansatz. This approach applies to arbitrary groups and field types, and allows hence to solve for a whole class of equivariant layers simultaneously, resulting in a general theory of equivariant Euclidean convolutions.
Steerable kernels	A central result following from this approach is that $\text{Aff}(G)$ -equivariant layers rely generally on an $\text{Aff}(G)$ -invariant neural connectivity. In particular, the translational subgroup requires <i>spatial weight sharing</i> , while the structure group G constrains this shared connectivity to be G -steerable (i.e. G -equivariant). Specifically for <i>linear maps</i> as ansatz for the layer, spatial weight sharing implies that it is <i>necessarily a convolution</i> , while G -steerability imposes an equivariance constraint on the convolution kernel; see Fig. 1.3. As a consequence, all that is required to extend convolutions to be affine group equivariant is to ensure that the convolution kernel satisfies the steerability constraint.
Gauge transforms	Intuitively, G -steerable kernels summarize the features in their field of view such into an output feature vector that any G -transformation of their field of view results into a corresponding G -transformation of the summarizing feature vector. A convolution with a G -steerable kernel results thus in a whole field of output feature vectors with the correct transformation law. Steerable kernels are formally equivalent to <i>representation operators</i> from quantum mechanics (e.g. scalar, vector or tensor operators). A generalization of the <i>Wigner-Eckart theorem</i> describes their construction from <i>harmonic basis functions</i> like circular or spherical harmonics (Figs 5.2 and 5.3) and <i>Clebsch-Gordan coefficients</i> . The latter imply <i>transition rules</i> between feature field types, similar to the state transition rules in quantum mechanics.

¹ $\text{GL}(d)$ is the *general linear group* of \mathbb{R}^d , consisting of all invertible $d \times d$ matrices, and $G \leq \text{GL}(d)$ means that G is some *subgroup* of it (Def. B.2.1). \rtimes denotes *semidirect products* (Def. B.2.5).

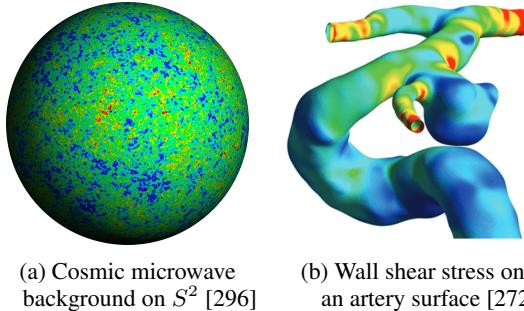


Figure 1.4: Characteristic patterns in signals on general Riemannian manifolds occur commonly at different locations and in arbitrary orientations. Convolutional networks address the former via their spatial weight sharing. We show that the latter requires convolution kernels to be equivariant under local gauge transformations. If the manifold has global symmetries (isometries), e.g. rotations of the sphere S^2 , the network can be designed to be equivariant w.r.t. these isometries.

Coordinate independent CNNs on Riemannian manifolds

▷ Parts II & III

A related line of research investigates the generalization of convolutional networks to *more general domains* like *Riemannian manifolds*. Figs. 1.4a and 1.4b shows the cosmic microwave background on the sphere and the wall shear stress of an artery’s surface as exemplary signals on non-Euclidean spaces that have been processed by convolutional models [228, 291]. As in the Euclidean case, distinctive patterns of features are usually appearing at different locations, implying that spatial weight sharing – i.e. a convolutional architecture – is still desirable.

Motivation & overview

A major complication in comparison to flat spaces is that *manifolds do not come with a preferred choice of reference direction*, along which a convolution kernel could be aligned to measure features; see Fig. 1.5. Since no reference direction is preferred, the kernel needs to be aligned *arbitrarily* on the manifold. The central theme of Parts II and III of this work is to regulate this arbitrariness by making the networks’ data processing independent from the specific alignment of convolution kernels. As we will show, this requires kernels again to be *steerable*, just as in the Euclidean setting. Since the response of a steerable kernel transforms predictably when its alignment is changed, the extracted information content is guaranteed to be the same for any (arbitrary) choice of alignment.

The use of steerable kernels makes the feature spaces of coordinate independent CNNs

1. *covariant under (passive) coordinate transformations* (gauge transformations), and
2. *equivariant under active transformations of the input signal*.

In contrast to the Euclidean case, we demand only the networks’ covariance (coordinate independence), from which equivariance follows automatically.

To make these statements more precise and explain the necessity for steerable kernels, we need to formalize the notion of “kernel alignment” on a manifold mathematically. We do so by identifying the alignment of a kernel at some point p of the manifold M with a *choice of local reference frame* – a *gauge* – of the corresponding tangent space $T_p M$. *Gauge transformations* are local passive transformations between choices of reference frames and, hence, kernel alignments. Fig. 1.7 visualizes the concept of aligning kernels along reference frames. Aligning a kernel relative to the canonical (uniquely preferred) *reference frame field* of the Euclidean plane \mathbb{R}^2 , shown in the top, results in the usual *kernel field* of Euclidean CNNs. A different frame field, shown in the bottom, implies an alternative kernel field and thus network. As stated above, the choice of frames is on most manifolds inherently ambiguous, such that no specific kernel alignment is preferred. Fig. 1.5 visualizes this issue for the sphere, where frames are only unique up to rotations.

Gauges

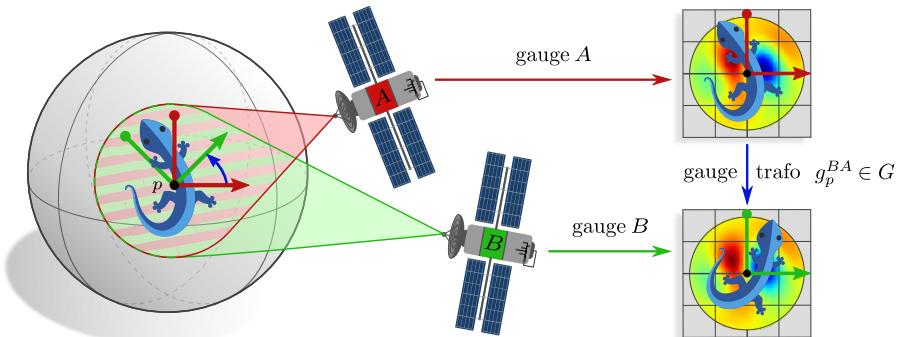


Figure 1.5: Different observers A and B may perceive a pattern of features from a different “viewpoint”. The satellites in our application are convolution kernels which summarize their local field of view around p into a feature vector at p . Their “viewpoint” is a choice of local reference frame (gauge) at p , along which the kernel is aligned. Since the observations from both viewpoints represent the same pattern, the kernel responses should contain equivalent information, that is, the inference should be *coordinate independent*. This constrains the convolution kernels to be *steerable*, i.e. *equivariant under local gauge transformations* (changes of reference frames). The level of gauge equivariance is determined by the *structure group* G , which depends both on the manifold and the application.

(Lizards adapted under the Creative Commons Attribution 4.0 International license by courtesy of Twitter.)

G-structures The level of ambiguity in the choice of frames depends on the *geometric structure* with which the manifold is equipped. Such structure often allows to *disambiguate reference frames up to certain symmetry transformations* (gauge transformations); see Fig. 1.6. This statement is best explained with a few examples:

- a naked *smooth manifold* M does not come with any preference in the choice of frames. Gauge transformations between general frames are arbitrary invertible linear maps, that is, they take values in the *general linear group* $G = \text{GL}(d)$, where $d = \dim(M)$.
- an *orientation* of the manifold allows to distinguish left-handed from right-handed frames. Gauge transformations between frames of either handedness are orientation preserving, i.e. they are elements of $G = \text{GL}^+(d)$ (invertible linear maps with positive determinant).
- a *volume form* allows to distinguish *unit volume frames*. Gauge transformations are then volume preserving, that is, they take values in the *special linear group* $G = \text{SL}(d)$.
- the *metric structure* of a Riemannian manifold allows to measure distances and angles in the tangent spaces and therefore allows to distinguish *orthonormal frames*. Gauge transformations between orthonormal frames are rotations and reflections in the *orthogonal group* $G = \text{O}(d)$.
- together, an *orientation and metric* imply *oriented orthonormal frames*. Gauge transformations are then only rotations in the *special orthogonal group* $G = \text{SO}(d)$.
- a *frame field* consists of a *unique frame* at every point of the manifold. Gauge transformations are in this case trivial, which is described by the *trivial group* $G = \{e\}$.

All of these geometric structures have in common that they define a preferred subset (subbundle) of frames such that gauge transformations take values in some *structure group* $G \leq \text{GL}(d)$. To emphasize the central role of the structure group G , such structures are denoted as *G-structures* GM . Visual examples of *G-structures* for different structure groups G and manifolds M are given in Fig. 1.6.

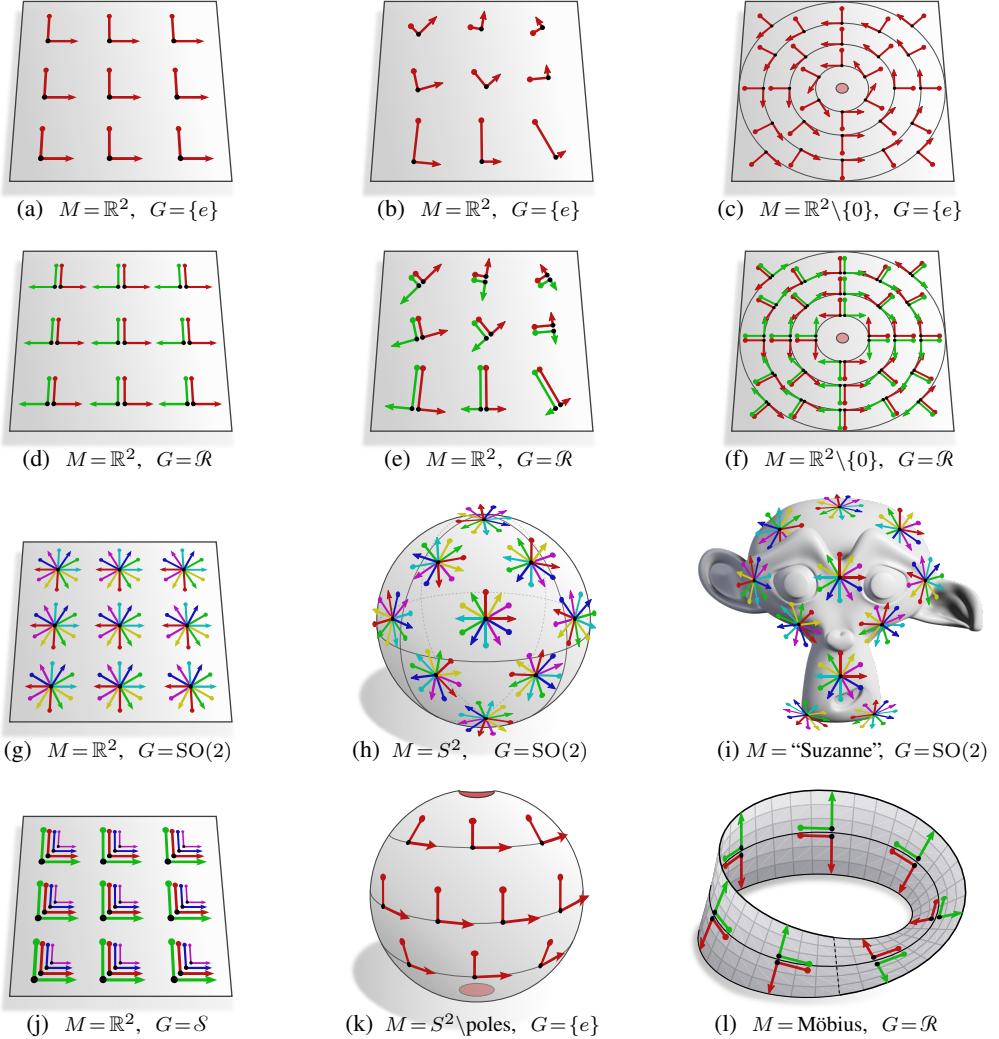
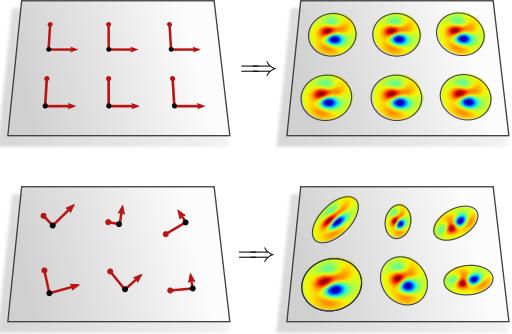


Figure 1.6: Exemplary G -structures GM for different structure groups G and manifolds M . The structure group G specifies which values gauge transformations can take, and therefore how “big” the subset of distinguished frames at each point is. Fig. 1.6a shows the canonical $\{e\}$ -structure (frame field) of \mathbb{R}^2 , corresponding to conventional Euclidean CNNs. The G -structures in Figs. 1.6d, 1.6g and 1.6j add reflected ($G = \mathcal{R}$), rotated ($G = \text{SO}(2)$) and scaled ($G = \mathcal{S}$) frames, respectively. They correspond to the $\text{Aff}(G)$ -steerable CNNs from Part I. G -structures are usually not unique: Figs. 1.6b and 1.6e show alternative G -structures on \mathbb{R}^2 (corresponding to an alternative metric w.r.t. which their frames are orthonormal). They might not be practically relevant but demonstrate the flexibility of our framework. The G -structures in Figs. 1.6c and 1.6f correspond to polar coordinates and model $\text{SO}(2)$ and $\text{O}(2)$ -equivariant (but not translation equivariant) CNNs. As G -structures are required to be continuous, the singularity at the origin 0 is removed. Fig. 1.6h shows the usual $\text{SO}(2)$ -structure on the 2-sphere S^2 , which is underlying $\text{SO}(3)$ -equivariant spherical CNNs. Spherical coordinates, which are singular at the (cut out) poles, imply the $\{e\}$ -structure in Fig. 1.6k. Topological obstructions may prevent continuous (non-singular) reductions of the structure group. For instance, topological spheres as in Fig. 1.6i require at least $G = \text{SO}(2)$, while non-orientable manifolds, like the Möbius strip in Fig. 1.6l, require at least $G = \mathcal{R}$ to admit the G -structures’ continuity. G -steerable kernels are hence strictly necessary for continuous convolution operations on topologically non-trivial manifolds.

Figure 1.7: A key property of convolutions is that they *share weights* over the manifold. The alignment of a convolution kernel is formalized by identifying it with a choice of reference frame. Different *frame fields* imply therefore different (convolutional) *kernel fields*.

The choice of frames, called *gauge*, is often not unique. The ambiguity in this choice is encoded in *G-structures*; shown in Fig. 1.6. To account for the arbitrariness of frames, the kernels are then required to be *G-steerable*; see Figs. 1.3 and 1.8.



Covariance Given a manifold with *G*-structure, we are confronted with an *inherent G-ambiguity in the choice of frames*. Facing the same issue during the development of his general theory of relativity, Albert Einstein proposed the *principle of (general) covariance* [80, 79]:

“Universal laws of nature are to be expressed by equations which hold good for all systems of coordinates, that is, are covariant with respect to any substitutions whatever.”

Along the same lines, we formulate a *principle of G-covariance* for deep learning:

“Convolutional neural networks are to be expressed by equations which hold good for arbitrary frames of the G-structure, that is, are covariant with respect to any G-valued gauge transformations.”

An important difference to Einstein’s *general covariance* is that we allow for arbitrary structure groups $G \leq \text{GL}(d)$, while general relativity is exclusively considering general linear gauge transformations in $G = \text{GL}(d)$. Our theory allows therefore to describe conventional non-covariant CNNs ($G = \{e\}$), generally covariant CNNs ($G = \text{GL}(d)$) and the whole spectrum of models in between.

Feature fields The principle of *G-covariance* demands in particular the coordinate independence of the networks’ feature spaces. Feature vectors are therefore *necessarily* associated with some *G-representation* ρ , which determines their transformation law under gauge transformations. The particular choice of group representation determines hereby the geometric type of a feature vector field. Note that such fields are the direct differential geometric generalization of the Euclidean feature fields of steerable CNNs, here defined with a focus on passive local gauge transformations of individual feature vectors instead of active $\text{Aff}(G)$ transformations of the global field. The latter is in the differential geometric setting described by *pushforward actions* on the feature field.

Steerable kernels Any network layer is required to respect the features’ (passive) transformation laws, that is, it needs to guarantee that its outputs transform as expected. Specifically for convolutions, *G-covariance* demands that applying a shared kernel relative to different *G*-frames should evoke the *same response up to a gauge transformation*, which requires the *G-steerability* (gauge equivariance) of convolution kernels. In this context, one may think of *G-steerable* kernels as measuring features *relative* to reference frames without introducing a dependence on their *absolute* alignment, which would break the *G-equivalence* of gauges. Fig. 1.8 visualizes the sharing of a reflection steerable kernel along different gauges, giving an intuition on how the responses’ *G-covariance* comes about.

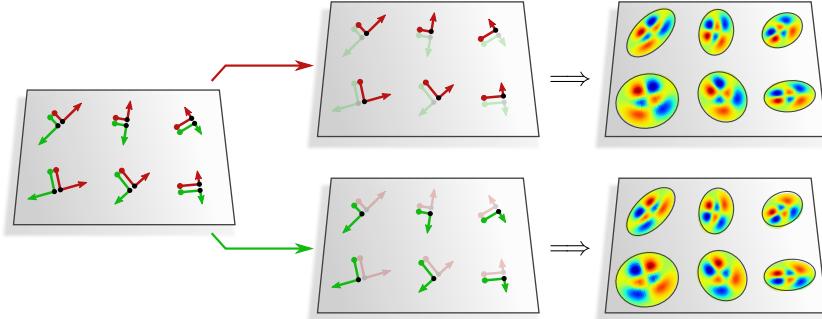


Figure 1.8: Sharing of a reflection steerable kernel along two different gauges (red and green) of a reflection G -structure. Due to the kernel’s reflection equivariance, the choice of gauge is in so far irrelevant that it will affect the responses only by a predictable gauge transformation, guaranteeing the features’ reflection covariance. Specifically, antisymmetric kernels map gauge invariant scalar fields to pseudoscalar fields, which flip their sign under gauge transformations. A symmetric kernel (Fig. 1.3b, left) would have resulted in a gauge invariant response, i.e. another scalar field.

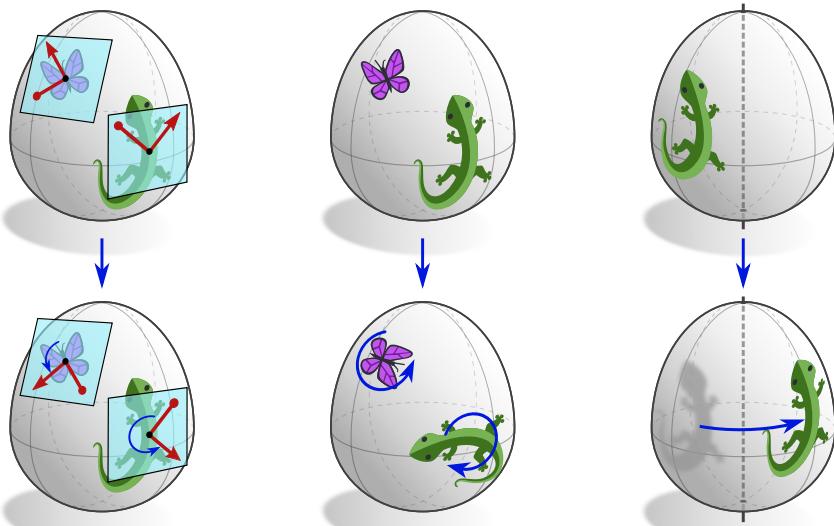


Figure 1.9: An overview of the types of transformations relevant for coordinate independent CNNs. *Left:* Passive local gauge transformations are transformations between local reference frames of the manifold’s tangent spaces. “Locality” refers to the fact that each tangent space has its own frame, and therefore independent gauge transformation. “Passive” means that only the coordinate representations of quantities like features change, while the actual geometric objects stay the same. *Middle:* The active variant of local gauge transformations transforms small patches of features independently from each other. A local observer (e.g. convolution kernel) can’t distinguish between active and passive gauge transformations; see Fig. 1.10. *Right:* Isometries are the (distance preserving) symmetries of Riemannian manifolds. They act via “pushforward” on feature fields, which can be thought as carrying the fields along with the group action. The isometry group of a manifold may be trivial. *All:* The features of coordinate independent CNNs are covariant under passive gauge transformations and equivariant under active gauge transformations and isometry group actions on signals.

(Lizards and butterflies adapted under the Creative Commons Attribution 4.0 International license by courtesy of Twitter.)

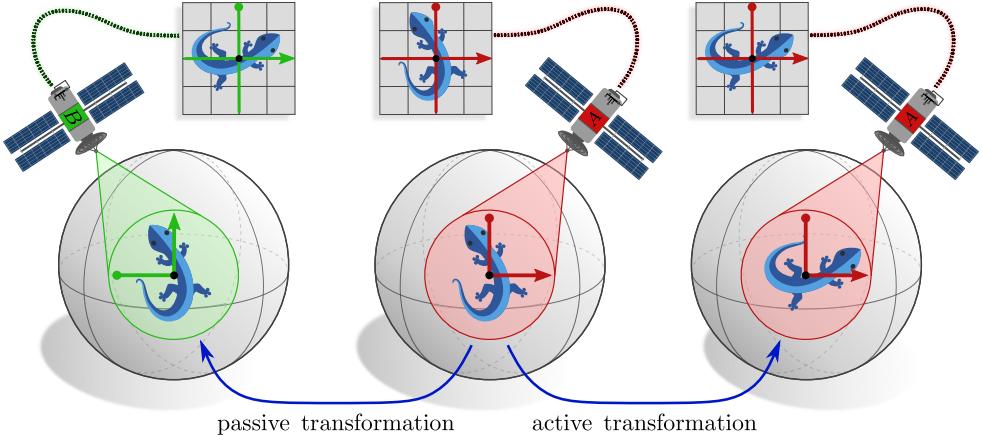


Figure 1.10: From the viewpoint of a local observer, passive gauge transformation of their own reference frame are equivalent to (inverse) active gauge transformation of the signal in their local field of view. In other words, observers (kernels) perceive features *relative* to their own frame of reference.

(Lizards adapted under the Creative Commons Attribution 4.0 International license by courtesy of Twitter.)

Besides being *covariant* under *passive* gauge transformations, which transform between reference frames but keep the actual signals on the manifold fixed (Fig. 1.9a), coordinate independent CNNs are *equivariant* under *active* transformations of the signals themselves.

Gauge equivariance Firstly, steerable kernels respond by definition equivariantly to active G -transformations of features in their field of view. The networks as a whole are therefore equivariant under independent “*active gauge transformations*” of local patterns, shown in Fig. 1.9b. Fig. 1.10 clarifies how this relates to the models’ G -covariance: a kernel’s response depends generally only on its relative alignment towards the signal, such that passive transformations of the alignment and (inverse) active transformations of the signal result in the same change of kernel response. If the kernel is in addition G -steerable, its responses change predictably, which is in the passive and active case referred to as G -covariance and G -equivariance, respectively.

Isometry equivariance Secondly, the manifold may have *global symmetries*, which, in the case of Riemannian manifolds, are distance preserving maps, called *isometries*. These isometries act via *pushforward* on feature fields, which can be thought of as “carrying features along” with the isometry action;² see Fig. 1.9c. We prove that a neural network is exactly then equivariant w.r.t. (a subgroup of) isometries if their neural connectivity – or *kernel field* – is invariant:

$$\text{isometry equivariant CNN} \iff \text{isometry invariant kernel field}$$

As visualized in Fig. 1.11, this requires that the neural connectivity is 1) shared spatially over the isometry orbits and 2) steerable under the isometry group’s stabilizer subgroups.

This result implies that the isometry equivariance of convolutions depends on the symmetries of their convolutional kernel field. Since we define convolutions as sharing kernels along frames of the G -structure, their *convolutional kernel fields inherit the symmetries of their underlying G-structure*; see Fig. 1.12 for examples.³ It follows as a corollary that *convolutions are equivariant w.r.t. that subgroup of isometries that are symmetries of the G-structure*. This result reduces the design of equivariant CNNs on manifolds to the design of invariant G -structures. The reader is encouraged to revisit the exemplary G -structures in Fig. 1.6 in regard to their symmetries and the implied equivariance properties.

²Technically, this requires isometries that are principal bundle automorphisms of the G -structure.

³The specific choice of gauge is by the kernel’s G -steerability irrelevant.

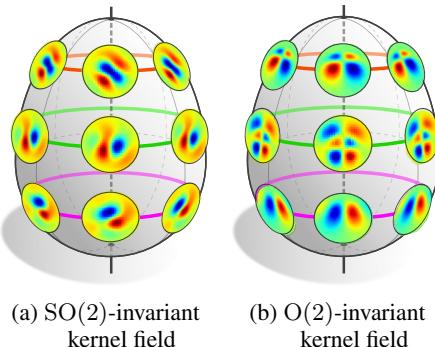
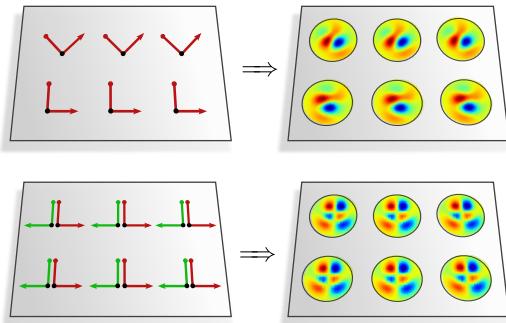


Figure 1.11: A neural network is *isometry equivariant* if and only if its kernel field (neural connectivity) is *isometry invariant*. This implies 1) spatial weight sharing over the *isometry orbits* (colored rings) and 2) a constraint on the kernels to be steerable w.r.t. their respective orbit's *stabilizer subgroup*. The $\text{SO}(2)$ and $\text{O}(2)$ actions on the egg have the same orbits, on which – but not between which – kernels are shared. While the stabilizers are for the $\text{SO}(2)$ action trivial (ignoring the poles), those for $\text{O}(2)$ are reflections, requiring an accordingly steerable connectivity. Note that the notion of “invariance” depends here on the field types, and might for instance imply antisymmetric kernels. The isometry equivariance of convolutions is discussed in Fig. 1.12.

Figure 1.12: *Convolutional kernel fields* are constructed by aligning a convolution kernel along (arbitrary) frames of a G -structure. They are hence invariant under the G -structure’s symmetries, and the implied *convolution is equivariant under the G -structure’s symmetries*; cf. Fig. 1.11. The top G -structure corresponds therefore to a convolution that is equivariant under horizontal (but not vertical) translations, while the bottom G -structure implies a fully translation and reflection equivariant convolution operation.



Until now, we did not comment on why we consider Riemannian manifolds and isometries instead of smooth manifolds and diffeomorphisms. The metric structure comes into play since we consider *spatially extended kernels*, whose matching with features is performed in *geodesic normal coordinates*. Specifically, the kernels are defined on flat space \mathbb{R}^d , and shared along gauges over the manifold’s tangent spaces. To match them with features, the feature fields are projected to the tangent spaces (or normal coordinates) by pulling (and parallel transporting) them back along *Riemannian exponential maps* as shown in Fig. 9.1. Since the exponential map depends on the manifold’s metric structure, this projection is deformed by non-isometric diffeomorphisms, which prevents the models’ full *diffeomorphism equivariance*. To remedy this shortcoming, and extend the models to be diffeomorphism equivariant, it would be necessary to replace the finite sized steerable kernels with *steerable partial differential operators*, which model local interactions [137]. Due to the wide use of spatially extended kernels in deep learning, we stick with this design choice throughout.

Diffeo-morphism equivariance

For Euclidean spaces $M = \mathbb{R}^d$, equipped with their canonical $\{e\}$ -structure (Fig. 1.6a), coordinate independent CNNs reduce to conventional Euclidean CNNs. Similarly, the G -structures in Figs. 1.6d, 1.6g and 1.6j recover the $\text{Aff}(G)$ -steerable Euclidean CNNs from Part I.⁴ This claim holds in fact for general affine groups, not only for isometries, i.e. $G \leq \text{O}(d)$, since the Riemannian exponential map commutes on Euclidean spaces with these more general symmetries. Note that coordinate independent CNNs allow to model more general convolutions on Euclidean space than steerable CNNs since they allow for alternative G -structures like those in Figs. 1.6b or 1.6e.

Euclidean CNNs

⁴There is a principal G -bundle isomorphism between these G -structures and $\text{Aff}(G)$.

Applications & literature review

▷ Chapters 6 & 10 and Part IV

The properties and practical utility of steerable and coordinate independent CNNs is demonstrated in multiple chapters with an applied focus.

Chapter 6 presents an *empirical evaluation of Euclidean steerable CNNs*, verifying their theoretically guaranteed generalization over affine transformations and showing their improved data efficiency and convergence rates. Since they generalize in particular over local gauge transformations, steerable CNNs are found to be highly useful even for natural images, which have a globally preferred directionality imposed by the gravitational field. A benchmark study, which uses datasets with different global symmetries, clarifies which choices of symmetry groups, field types and equivariant nonlinearities work best in practice.

Chapter 10 describes and evaluates an exemplary implementation of coordinate independent CNNs on the *Möbius strip*, equipped with the reflection group structure from Fig. 1.6l. The model is empirically shown to be isometry equivariant and outperform a naive non-covariant (and non-equivariant) implementation.

Part IV presents a comprehensive *literature review* on convolutional networks. It demonstrates the generality of our differential geometric formulation of convolutional networks by describing *more than 100 models* as specific instances of coordinate independent CNNs. As coordinate independent CNNs allow for arbitrary structure groups, they explain even non-covariant models ($G = \{e\}$) on manifolds, which fix some gauge heuristically, relative to which they apply non-steerable ($\{e\}$ -steerable) kernels. Chapter 15 focuses specifically on $\text{Aff}(G)$ -equivariant convolutions on Euclidean spaces, proving in particular how coordinate independent CNNs reduce in this setting to the classical formulation of steerable CNNs. Euclidean convolutions that are based on polar or more general hyperspherical G -structures, and are therefore rotation, but not translation equivariant, are discussed in Chapter 16. Log-polar coordinates result similarly in rotation and scale equivariant models. Chapter 17 describes $O(3)$, $SO(3)$, $O(2)$ and $SO(2)$ -equivariant spherical CNNs, and icosahedral approximations thereof. Implementations on general surfaces are covered in Chapter 18. Besides describing the models in the literature, all of these chapters start by explaining the differential geometry and constructions like exponential maps or transporters on their specific manifolds, which is helpful when implementing new models. An overview of the resulting taxonomy of convolutional networks is given in Table 14.1 on page 273.

An implementation of G -steerable kernels and Euclidean steerable CNNs for arbitrary field types of any structure group $G \leq \text{GL}(d)$ is available in our PyTorch extension `escnn` [39] (formerly `e2cnn` [38]). The library is designed to abstract away most of the complicated representation theory, such that the user only has to make basic choices like selecting the symmetry group or feature field types. It has been widely used for tasks like aerial imaging for object detection [116], deforestation segmentation [214], the processing of weather data [122] dendrite-core detection in material science [96], understanding the human ventral visual stream [147], fluid dynamics applications [317, 341, 225, 319], molecular energy prediction [40, 353], molecular recognition [14], fMRI imaging [146], morphologic profiling of cells and organelles [30], inverse problems [37], compressed sensing [168], reinforcement learning, robotics and planning [216, 314, 360, 312, 129, 138, 188, 313, 315, 350], pose regression [219], feature matching [227, 19, 286, 176, 7], oriented keypoint detection [177], object tracking [113], symmetry detection [266], symmetry learning [313], shape generation [305], 3d mesh generation from images [214], visual reasoning [208], differential privacy [123], and other applications [354, 260, 135, 43, 137, 94, 331, 145, 215, 104, 89, 103, 104, 111, 9, 124, 190, 169, 155, 239, 218].

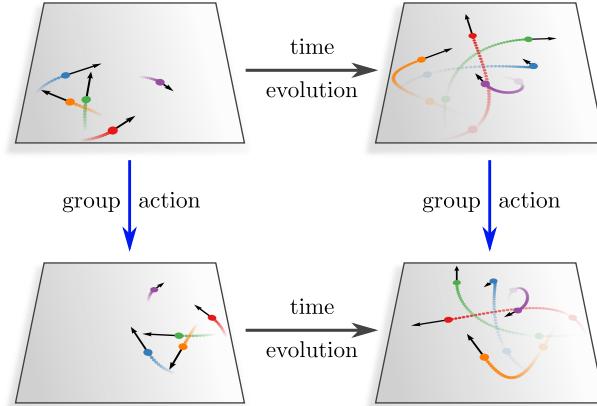


Figure 1.13: The *invariance* of the *laws of nature* under Poincaré (or Galilean) transformations implies, and is implied by, the *equivariance* of the system’s *time evolution*. This mutual implication is analogous to that between an *equivariant neural network* and its *invariant neural connectivity*; see Fig. 1.2c. Similar to an observer in physics, an equivariant network may only perform *relative* measurements. The deep learning analogue to the inertial frames in physics, related by Lorentz (or Poincaré) transformations, are frames of a G -structure, which are related by G -valued gauge transformations (or affine group actions). The laws of nature and the inference of convolutional networks are both governed by *steerable* (gauge equivariant) operators.

(Trajectories generated with code from Kipf et al. [153].)

Relation to physics

Our formulation of Equivariant and Coordinate Independent Convolutional Networks shows some striking similarities to theories in physics, which we briefly highlight here.

Field theories: First of all, the feature spaces of CNNs are spaces of *feature fields*, which are similar to the *tensor fields* occurring all throughout physics. Mathematically, both are formalized as *associated bundle sections*, and are hence described by *gauge field theories*.

Relativity: We find that *equivariant predictions* of convolutional networks are in one-to-one correspondence with an *invariant neural connectivity*. Equivalently, *invariant laws of nature* imply, and are implied by, an *equivariant system dynamics*,⁵ visualized in Fig. 1.13.

Such invariances of physical laws are captured by the *principle of relativity*. The word “relativity” refers hereby to the fact that invariant laws of physics allow an experimenter only to measure events *relative to their frame of reference*, but not in an absolute sense. Equivariant CNNs have exactly the same property: they allow only for *relative measurements of features*, but are constrained to be insensitive to their absolute pose. We can therefore claim to have found a theory of relativity of neural networks.

In physics, one usually starts by studying the special theory of relativity, which considers *flat Minkowski spacetime* and global *Poincaré transformations* between inertial frames. General relativity generalizes this setting to *curved spacetime* and distinguishes local tetrad frames, which are related by *local Lorentz transformations*. Similarly, we develop our theory of equivariant CNNs first on *flat Euclidean spaces*, equipped with *global affine group actions* that transition between affine charts (Def. 15.1.1). Coordinate independent CNNs generalize

⁵Eq. (B.28) in Appendix B.4 shows that this is a general property of equivariant maps.

this setting to *general Riemannian manifolds* and *local G -transformations* between frames of a G -structure. Note that all that is required to construct a relativistic neural network for fields on spacetime is to choose the Lorentz group $G = \mathrm{O}(1, 3)$ as structure group of a coordinate independent CNN.

From a philosophical viewpoint, it was argued that the assumption of invariant laws of nature is essential for our ability to discover them empirically. For instance, Eugene Wigner, a Nobel prize winner who greatly contributed to the application of group theory in modern physics, claimed in this context [328]:

“... if there were no phenomena which are independent of all but a manageable small set of conditions, physics would be impossible.”

In a universe in which the laws of nature were not Poincaré (or Galilean) invariant, experimental outcomes would depend on the experiment’s location, orientation and time, thus preventing any reproducibility and being at odds with the scientific method itself. The situation in machine learning is remarkably similar: if we would not assume an invariant neural connectivity (i.e. an equivariant network), the model would laboriously have to (re)learn how to process a given pattern in every possible pose and location. Following Wigner’s quotation, we retain:

If there was no inference which is independent of all but a manageable small set of conditions, machine learning would be infeasible.

Differential operators & equations: A seeming difference between the two research areas is that the field equations in physics involve *partial differential operators*, while convolutional networks are usually applying *spatially extended kernels*. We bridged this gap in our publication [137], which augments convolutional networks with (steerable) partial differential operators (PDOs). Interestingly, we found that common operators in physics, like *gradients*, *divergences*, *curl*, or the *Laplacian*, are all examples of *steerable PDOs* – this PDO steerability is actually necessary for the laws of nature to be invariant.⁶

Convolutions do furthermore play a great role in solving *inhomogeneous, linear, translation invariant partial differential equations* (PDEs): their solutions may be given by convolutions with *Green’s functions*, which are “impulse responses” of PDEs. Note the terms “linear” and “translation invariant”, which are exactly the conditions from which we derive convolutional network layers in Theorems 3.2.1 and 4.3.1. It would be interesting to investigate whether convolutional network layers can be viewed as solving some linear neural PDE, whose inhomogeneity is given by the layer’s input feature map.

Representation operators: Finally, we found that our *G -steerable kernels*, which *map between feature fields*, are mathematically equivalent to the *representation operators* from quantum mechanics, which *map between quantum states*. Both are therefore described by a Wigner-Eckart theorem [173], which essentially identifies the admissible irreducible G -representations that may be contained in the convolution kernel or quantum operator. The implied *state transition rules* in quantum mechanics are therefore equivalent to the *feature transition rules* in equivariant deep learning; see Fig. 5.4.

What is the reason for these similarities between physics and deep learning? The connection is simply a consequence of both being geometric theories – all of the similarities are fundamental results of the underlying representation theory and differential geometry.

⁶For these examples, invariant under Galilean instead of Poincaré transformations.

1.1 Outline

This work is organized into four main parts and an appendix. Part I develops a theory of *affine group equivariant CNNs on Euclidean spaces*. Feature spaces are formulated as *group representation spaces* and network layers are equivariant maps between them. Convolutions, or more general forms of weight sharing, are shown to follow from the requirement for group equivariance. Parts II and III generalize these models to Riemannian manifolds. Part II introduces this generalization in an easily accessible language, expressing feature fields and network layers relative to *local coordinates* (fiber bundle trivializations). The demanded *coordinate independence* (covariance) of the models requires features to be associated with some transformation law. Network layers are required to guarantee the correct transformation behavior of features. Part III formalizes these coordinate independent neural networks in terms of *associated fiber bundles*. This allows for a *global, coordinate free* formulation, which is particularly useful when investigating the networks' isometry equivariance. The definitions from Part II are recovered when expressing the coordinate free operations in local bundle trivializations (coordinates). The reader not familiar with fiber bundles may skip Part III at a first pass. Part IV turns to *applications on specific geometries*. It provides in particular a detailed review of convolutional networks from the literature and reformulates them as instantiations of coordinate independent CNNs for specific manifolds, G -structures and field types. The appendix covers mathematical background, discusses further details omitted in the main parts, and gives long proofs.

Detailed Overview

Part I: Chapter 2 gives a brief introduction to the general concept of *invariant and equivariant models* in machine learning. The basic design principle and advantages of such models are discussed in Section 2.1. Section 2.2 clarifies how equivariant *neural networks* may be constructed as sequences of equivariant layers.

Chapter 3 reviews *conventional Euclidean CNNs* [175] from a group theoretic perspective. The feature spaces of conventional CNNs are in Section 3.1 formalized as *regular translation group representations*, consisting of feature maps that are equipped with a translation group action. Section 3.2 derives typical CNN layers, like convolutions, bias summation, nonlinearities and pooling operations, from the requirement for the networks' *translation equivariance*. Such layers will generally exhibit some form of *spatial weight sharing*, which means that they act by applying some shared local operation at each point of the space.

Chapter 4 generalizes these constructions and results to *affine symmetry groups*, resulting in *Euclidean steerable CNNs* [53, 323, 322, 56, 137, 173, 40]. Affine groups $\text{Aff}(G)$ are briefly introduced in Section 4.1. The feature spaces of steerable CNNs are in Section 4.2 defined as *induced affine group representations*. They consist of feature vector fields, which differ from conventional feature maps in that they come with an affine group action instead of just a translation group action. Affine group equivariant layers are derived in Section 4.3. Just like conventional CNNs, these layers require *spatial weight sharing*, however, now with an additional *G -steerability constraint* on the shared neural connectivity. Section 4.4 explains how networks with a varying level of equivariance in different layers may be constructed and how this approach can be useful. Section 4.5 comments on *group convolution* based networks and proves that they are a special case of steerable CNNs for regular G -representations.

Steerable convolutions are ultimately just convolutions with *G -steerable kernels*, which are the subject of Chapter 5. Section 5.1 argues that steerable kernels form a *vector subspace* of

general convolution kernels and explains how these can be parameterized for learning. To build an intuition on such kernels, Section 5.2 derives some of them for the simple example of the *reflection group*. A general solution of the kernel constraint for arbitrary representations of compact groups G in terms of a *generalized Wigner-Eckart theorem* is presented in Section 5.3. Section 5.4 gives an overview of some alternative approaches to solving the steerability constraint and parameterizing steerable kernels.

Chapter 6 investigates the properties of steerable CNNs empirically and benchmarks different design choices like e.g. symmetry groups, field types or equivariant nonlinearities.

Part II: The second part of this work formulates a gauge theory of convolutional networks on Riemannian manifolds, expressed in local coordinates.

Chapter 7 introduces the underlying mathematical framework of *gauges*, *gauge transformations* and *G-structures*. Specifically, Section 7.1 defines gauges as choices of isomorphisms $T_p M \cong_{\text{vec}} \mathbb{R}^d$ which associate numerical coefficients in \mathbb{R}^d to tangent vectors in $T_p M$. A gauge determines not only how geometric quantities like tangent vectors are “measured”, but implies also coordinate expressions for *functions* that map between them. Section 7.2 explains such induced coordinate expressions of functions mapping between tangent spaces. As gauges are in one-to-one correspondence with choices of reference frames, a family of geometrically distinguished gauges corresponds to a bundle of preferred frames. These so-called *G-structures* are briefly discussed in Section 7.3.

The goal of Chapter 8 is to define coordinate independent feature spaces on Riemannian manifolds. Section 8.1 introduces coordinate independent *feature vector fields*, which generalize the Euclidean feature fields from Part I to the differential geometric setting. As in the case of tangent vectors, the numerical coefficients of feature vectors transform when transitioning between reference frames. The gauge transformation laws (“field types”) of feature vectors determine in particular their *parallel transport* and their pushforward when being acted on by *isometries*, which are described in Sections 8.2 and 8.3, respectively.

Chapter 9 develops *neural networks* that map between feature fields. *Pointwise operations*, like bias summation, 1×1 -convolutions and nonlinearities, are discussed in Section 9.1. Section 9.2 focuses on *coordinate independent convolutions* with spatially extended kernels. Each of these operations is initially introduced without the weight sharing assumption, that is, allowing for instance for a different kernel at each point of the manifold. These kernels (or biases or nonlinearities) are beyond the requirement for coordinate independence not constrained in any way. However, when requiring spatial weight sharing, they become constrained to be *gauge equivariant (G-steerable)* since only equivariant quantities can be shared in a coordinate independent manner. Section 9.3 gives a concise proof of the isometry equivariance of coordinate independent convolutions in terms of local coordinate expressions. The key idea here is that isometries can be viewed as inducing gauge transformations (passive interpretation), which are explained away by the kernels’ gauge equivariance.

Chapter 10 describes an implementation of orientation independent convolutions on the *Möbius strip*. After reviewing the geometry of the Möbius strip in Section 10.1, multiple types of feature fields are defined in Section 10.2. The following Sections 10.3 and 10.4 formulate orientation independent CNNs analytically and describe their implementation. Section 10.5 closes with an empirical evaluation of Möbius convolutions.

Part III: The third part formalizes and extends the content of Part II in the language of associated fiber bundles.

Chapter 11 defines the relevant associated bundles and their local trivializations. A general introduction to fiber bundles is given in Section 11.1. Sections 11.2 and 11.3 introduce

the tangent bundle TM , the frame bundle FM , G -structures GM and G -associated feature vector bundles \mathcal{A} . Feature fields are globally defined as sections of feature vector bundles. *Local bundle trivializations* (gauges), which are discussed in Section 11.4, express these bundles in coordinates, thereby recovering our definitions from Part II. We demonstrate in particular how local trivializations of the different bundles induce each other, such that their gauge transformations (transition maps) are synchronized. Section 11.5 discusses *parallel transporters* on G -bundles.

Chapter 12 reformulates the coordinate independent networks from Chapter 9 in terms of fiber bundles. 1×1 -convolutions are in Section 12.1 described as specific vector bundle M -morphisms. Alternatively, they may be viewed as sections of a homomorphism bundle. Section 12.2 introduces *coordinate free kernel fields* and *kernel field transforms*. These operations are similar to coordinate independent convolutions but are not required to share weights, i.e. may apply a different kernel at each spatial location. *Convolutional kernel fields* are constructed by sharing a G -steerable (gauge equivariant) kernel over the whole manifold. A *Coordinate free formulation of convolutions* is then defined as kernel field transforms with convolutional kernel fields. When expressing the coordinate free formulation of convolutions relative to local trivializations (gauges), we recover the coordinate expressions of convolutions from Section 9.2.

The *isometry equivariance* of convolutions is investigated in Chapter 13. After introducing isometries, Section 13.1 discusses their *pushforward action* on the fiber bundles. These action may again be expressed in local trivializations, resulting in the formulation from Section 8.3. Section 13.2 defines the action of isometries on kernel fields and proves that the *isometry equivariance of a kernel field transform* implies the *isometry invariance of its kernel field* and vice versa. Coordinate independent convolutions are proven to be equivariant under the action of those isometries which are bundle automorphisms (symmetries) of the G -structure GM . Section 13.3 investigates isometry invariant kernel fields in greater detail and proves that they are equivalent to *kernel fields on quotient spaces* of the isometry action – intuitively speaking, isometry invariant kernel fields are required to share kernels over the isometry orbits. This result implies in particular that isometry equivariant kernel field transforms on *homogeneous spaces* are necessarily coordinate independent convolutions.

Part IV: The fourth part of this work demonstrates that a vast number of convolutional networks from the literature can be interpreted as applying coordinate independent convolutions for some choice of G -structure and field types. It starts in Chapter 14 with a general discussion about the design choices of coordinate independent CNNs. Table 14.1 on page 273 gives an overview and classification of the models that are reviewed. The reader is invited to have a look at the G -structures that are visualized in Part IV as their symmetries give an intuitive idea about the properties of the corresponding convolutions.

Chapter 15 describes a family of coordinate independent CNNs on Euclidean spaces which corresponds exactly to the affine group equivariant *Euclidean steerable CNNs* from Part I. As a preparation, Section 15.1 describes the geometry of Euclidean spaces and constructs $\text{Aff}(G)$ -invariant G -structures from atlases of charts with $\text{Aff}(G)$ -valued transition maps. These G -structures are in Section 15.2 shown to result in $\text{Aff}(G)$ -equivariant convolutions, whose coordinate expressions are steerable CNNs. Section 15.3 comments briefly on such models found in the literature, which differ mainly in the assumed choices of structure groups and group representations.

Chapter 16 covers CNNs on *punctured Euclidean spaces* $\mathbb{E}_d \setminus \{0\}$, whose origin $\{0\}$ was removed. These models are rotation equivariant around the origin, however, they are not

translation equivariant. They are based on G -structures that correspond to polar coordinates, log-polar coordinates or spherical coordinates.

Spherical CNNs are covered in Chapter 17. Section 17.1 discusses the geometry of the (embedded) 2-sphere S^2 . Interpreting the tangent spaces as two-dimensional subspaces of an embedding space \mathbb{R}^3 , we derive closed form expressions of exponential and logarithm maps, frames, gauges, transporters and isometry actions. Section 17.2 reviews $\text{SO}(3)$ and $\text{O}(3)$ -equivariant spherical CNNs. We prove in particular that our theory includes the general formulation of spherical convolutions by Cohen et al. [56] as a special case. Spherical CNNs that are merely $\text{SO}(2)$ rotation equivariant around a fixed axis are described in Section 17.3. Section 17.4 reviews icosahedral CNNs. The icosahedron approximates the sphere but consists of locally flat faces which allow for an efficient implementation of convolution operations.

A survey of convolutional networks on *general two-dimensional surfaces* is found in Chapter 18. Section 18.1 provides a brief introduction to the classical differential geometry of embedded surfaces and their discretization in terms of triangle meshes. The surface convolutions in the literature are categorized in two classes: The first class, covered in Section 18.2, is based on $G = \text{SO}(2)$ -steerable kernels. These models are independent from the specific choice of right-handed, orthonormal frame. Section 18.3 reviews the second category of models, which are based on $\{e\}$ -steerable, i.e. non-equivariant kernels. These models rely explicitly on a choice of frame field. They differ therefore mainly in the heuristics that are used to determine reference frames. Note that such models are necessarily discontinuous on non-parallelizable manifolds like for instance topological spheres.

Appendix: The appendices cover mathematical background, additional information about the theory of equivariant CNNs and long proofs. A list of all proofs and definitions is given in Appendix A.

The *group and representation theory* that is required to define steerable and coordinate independent convolutions is discussed in Appendix B.

Appendix C introduces the *coordinate chart formalism* of differential geometry, and explains in detail how it relates to the fiber bundle formalism that we are using primarily. Charts are shown to induce specific bundle trivializations, known as coordinate bases (or holonomic bases). The gauge transformations between these trivializations are the well-known Jacobians of chart transition maps. Table C.1 gives an overview of the correspondences between the two formalisms.

Coordinate independent convolutions are computed by expressing feature fields in geodesic normal coordinates, where they are matched with G -steerable convolution kernels. This process involves an *integration over the tangent spaces* which is described in Appendix D.

Appendix E comments on *equivariant MLPs*, in particular how equivariant fully connected network layers may be constructed and parameterized.

Kondor and Trivedi [162], Cohen et al. [56] and Bekkers [10] proposed quite general theories of *convolutions on homogeneous spaces*. As these models share weights via the action of some symmetry group, they are very similar to our isometry equivariant kernel field transforms from Sections 13.2 and 13.3. Appendix F reviews these models and explains how they relate to our coordinate independent convolutions.

Appendix G comments on the *coordinate independence* of kernels and *weight sharing* along reference frames. A coordinate independent sharing of weights is only possible for G -steerable kernels.

The *Wigner-Eckart theorem* from Section 5.3 explains how steerable kernels are constructed from harmonic basis functions, irrep endomorphisms and Clebsch-Gordan coefficients. To give an intuition for the role of these ingredients, Appendix H introduces them step by step along with a succession of increasingly general cases of kernel constraints.

The next four appendices give long proofs that were deferred in the main text.

Specifically, Appendix I asserts the well-definedness of our kernel field transforms and convolutions given that the underlying kernel field is smooth and consists of compactly supported kernels. Well-definedness means here that the defining integrals exist and that the resulting feature fields are smooth.

Appendix J argues that feature fields which transform according to the regular representation of the structure group G are equivalent to scalar fields on the G -structure. This is relevant since some models, specifically group convolutions, take this viewpoint (cf. Section 4.5).

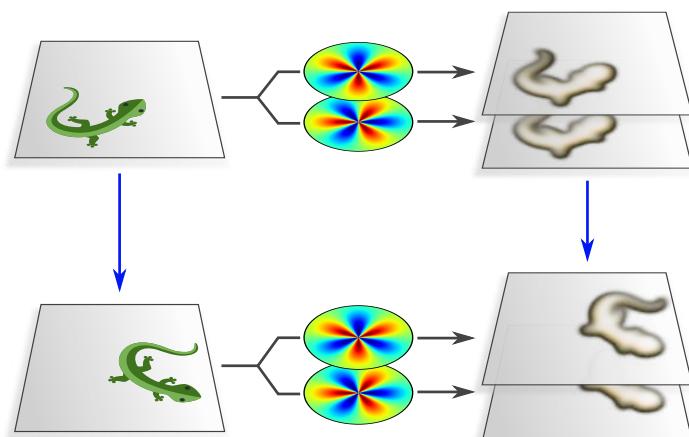
Appendix K.1 proves that isometry invariant kernel fields on the manifold are equivalent to kernel fields on quotient spaces of the isometry action. The special case of homogeneous spaces, on which isometry equivariant kernel field transforms are equivalent to coordinate independent convolutions, is covered in Appendix K.2.

The spherical convolutions of Cohen et al. [56] are in Appendix L proven to be a special case of our coordinate independent spherical convolutions – any spherical CNN that is covered by their theory is therefore explained by our theory as well.

This book is based on the main author’s doctoral dissertation. Appendix M lists the main research questions raised during the doctoral studies and summarizes brief answers. It furthermore discusses some conclusions and directions for future research. As this appendix provides an overview of the book’s content, it may be consulted as an alternative introduction.

PART I

EQUIVARIANT CONVOLUTIONAL NETWORKS ON EUCLIDEAN SPACES



Introduction & overview

Spatially structured signals, like images, videos or audio, are nowadays most commonly processed by convolutional neural networks (CNNs). A main characteristic of convolutional networks is their translation equivariance, i.e. their property to produce feature maps that move along when shifting the models' input (Fig. 1.1). As we will show, the convolutional network design is not just sufficient to produce an equivariant response, but is actually necessary – in other words, CNNs are fully characterized by and can be *derived* from the requirement for translation equivariance. Based on this insight, we generalize convolutional networks on Euclidean spaces by *defining* them as neural networks that are equivariant under an extended group of geometric symmetries.

Chapter 2 introduces the general idea and merits of *invariant and equivariant neural networks*. Chapter 3 formalizes *conventional Euclidean CNNs* in a representation theoretic language, defining their feature spaces as *translation group representations* and *deriving* their layers as equivariant maps between such representation spaces. This procedure is in Chapter 4 mirrored for more general *affine symmetry groups*, resulting in *Euclidean steerable CNNs*. Affine equivariant networks are found to rely on *steerable convolution kernels*, which are investigated further in Chapter 5. Chapter 6 presents an empirical study and benchmarking of steerable CNNs.

Euclidean steerable CNNs will be revisited in Part IV, Chapter 15, which explains how they are *recovered from their differential geometric generalization* to Riemannian manifolds from Parts II and III. This alternative formulation clarifies the models' independence from choices of *global affine coordinate charts*, occurring in implementations as choices of pixel grids. Appendix F is furthermore commenting on the generalization of steerable CNNs to homogeneous spaces, like, for instance, the sphere.

CHAPTER 2

Invariant and equivariant models

Viewed abstractly, feed forward neural networks are sequences of parameterized functions, denoted as layers. The most common approach to construct equivariant networks is, accordingly, to compose them as a sequence of equivariant layers, each satisfying an individual symmetry constraint.

An equivariant model design is twofold free lunch: On the one hand, it guarantees that the model respects the symmetries of the learning task by construction, and generalizes whatever it learned over group orbits, i.e. all inputs that are related by the symmetry group action. On the other hand, the symmetry constraint reduces the number of model parameters, which leads to a faster convergence and makes the models robust against overfitting.

Before discussing equivariant neural networks in Section 2.2, we review general equivariant models in machine learning and their reduced hypothesis spaces in Section 2.1.

2.1 Equivariant machine learning and quotient hypothesis spaces

The task in machine learning is typically to approximate some *target function* $\mathcal{T} : \mathcal{F}_{\text{in}} \rightarrow \mathcal{F}_{\text{out}}$ with a *model* $\mathcal{M} : \mathcal{F}_{\text{in}} \rightarrow \mathcal{F}_{\text{out}}$. The domain \mathcal{F}_{in} and codomain \mathcal{F}_{out} are the input and output *feature spaces*, respectively, for instance spaces of images and class labels. Denote the unconstrained *hypothesis space*, i.e. the space of models under consideration during the training process, by $\mathcal{H}_{\text{full}}$.

2.1.1 Invariant and equivariant target functions

Invariant and equivariant models are used in learning tasks where some symmetry group acts on the feature spaces and the target function commutes with these actions. Let G therefore be some symmetry group (Def. B.1.1) and consider some *group action* (Def B.3.1)

$$\triangleright_{\text{in}} : G \times \mathcal{F}_{\text{in}} \rightarrow \mathcal{F}_{\text{in}}, \quad (g, x) \mapsto g \triangleright_{\text{in}} x \quad (2.1)$$

on the input feature space \mathcal{F}_{in} . The target function \mathcal{T} is said to be *G-invariant* iff it satisfies

$$\mathcal{T}(g \triangleright_{\text{in}} x) = \mathcal{T}(x) \quad \forall g \in G, \quad x \in \mathcal{F}_{\text{in}}. \quad (2.2)$$

This is visualized by the *commutativity* of the following diagram:¹

$$\begin{array}{ccc}
 \mathcal{F}_{\text{in}} & & \mathcal{F}_{\text{out}} \\
 \downarrow g \triangleright_{\text{in}} & \nearrow L & \\
 \mathcal{F}_{\text{in}} & \nearrow L &
 \end{array} \tag{2.3}$$

Examples for invariant tasks are image classification, which is usually translation invariant, or the classification of a set (as a whole), where the order (permutation) of elements is irrelevant. These two examples are visualized in Figs. 2.1c and 2.1a, respectively.

While the output of an invariant function does not change when its input is transformed, the output of an equivariant function responds with a corresponding transformation. We consider therefore a second, potentially different group action

$$\triangleright_{\text{out}} : G \times \mathcal{F}_{\text{out}} \rightarrow \mathcal{F}_{\text{out}}, \quad (g, y) \mapsto g \triangleright_{\text{out}} y \tag{2.4}$$

on the output feature space \mathcal{F}_{out} . A *G-equivariant* target function w.r.t. these group actions satisfies a more general relation

$$\mathcal{T}(g \triangleright_{\text{in}} x) = g \triangleright_{\text{out}} \mathcal{T}(x) \quad \forall g \in G, x \in \mathcal{F}_{\text{in}}, \tag{2.5}$$

which corresponds to a slightly modified commutative diagram:

$$\begin{array}{ccc}
 \mathcal{F}_{\text{in}} & \xrightarrow{\mathcal{T}} & \mathcal{F}_{\text{out}} \\
 \downarrow g \triangleright_{\text{in}} & & \downarrow g \triangleright_{\text{out}} \\
 \mathcal{F}_{\text{in}} & \xrightarrow{\mathcal{T}} & \mathcal{F}_{\text{out}}
 \end{array} \tag{2.6}$$

Equivariant functions include invariant functions since the group action on \mathcal{F}_{out} may be chosen to be trivial (invariant), as visualized in the diagram in Eq. (B.27). Figs. 2.1d and 2.1b visualize examples of translation equivariant image segmentation and the permutation equivariant processing of set elements.

2.1.2 Quotient hypothesis spaces

Given that the target of a learning task is known to be invariant or equivariant, it is reasonable to guarantee this behavior in the considered class of machine learning models by design. Specifically, if \mathcal{T} is invariant, Eq. (2.2), one restricts to a constrained hypothesis space

$$\mathcal{H}_{\text{inv}} := \{m : \mathcal{F}_{\text{in}} \rightarrow \mathcal{F}_{\text{out}} \mid m(g \triangleright_{\text{in}} x) = m(x) \quad \forall g \in G, x \in \mathcal{F}_{\text{in}}\} \subseteq \mathcal{H}_{\text{full}} \tag{2.7}$$

consisting of invariant models only. Similarly, if \mathcal{T} is equivariant, Eq. (2.5), the hypothesis space should be constrained to equivariant models:

$$\mathcal{H}_{\text{equiv}} := \{m : \mathcal{F}_{\text{in}} \rightarrow \mathcal{F}_{\text{out}} \mid m(g \triangleright_{\text{in}} x) = g \triangleright_{\text{out}} m(x) \quad \forall g \in G, x \in \mathcal{F}_{\text{in}}\} \subseteq \mathcal{H}_{\text{full}} \tag{2.8}$$

¹Diagrams give a visual overview of functions and the spaces between which they map. For instance, the diagram on the left implies that there are functions $f : X \rightarrow Y$, $g : Y \rightarrow Z$ and $h : X \rightarrow Z$. If the compositions of functions along all paths with the same start and endpoint agree, the diagram is said to be *commutative*. Our example diagram is commutative if (and only if) $h = g \circ f$ holds.

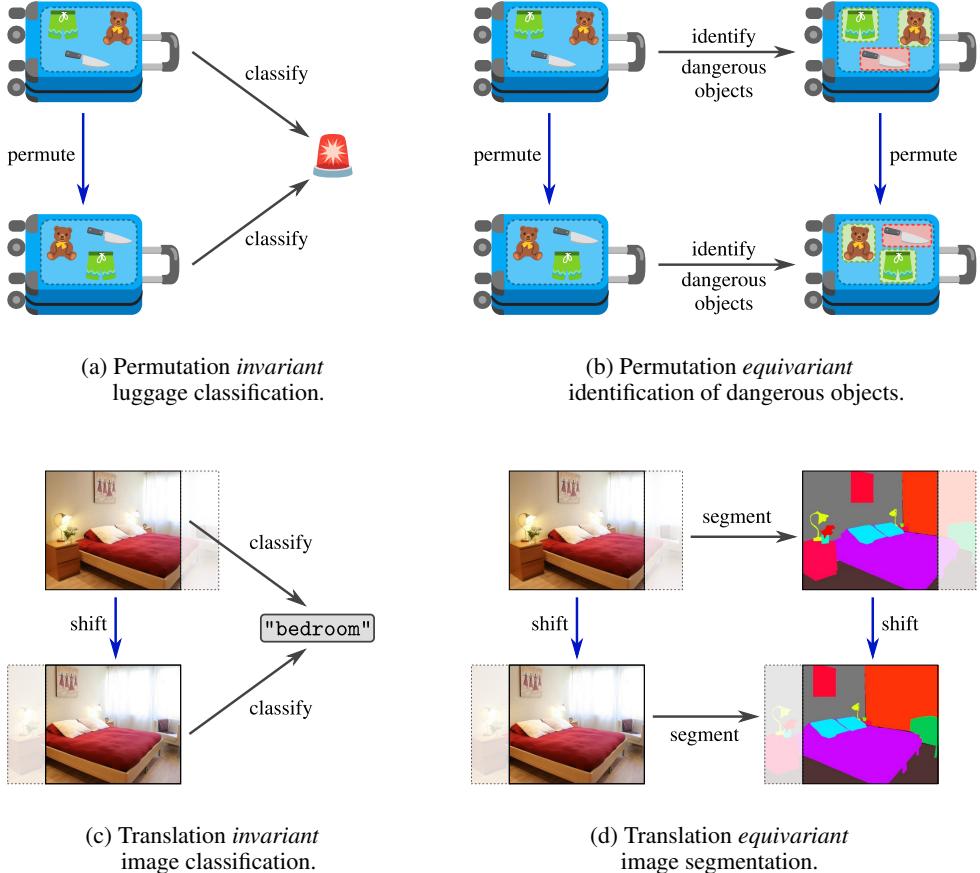


Figure 2.1: Invariant classification (left) and equivariant segmentation (right) of sets (top) and images (bottom). The elements of a set (or rather tuple) are acted on by the permutation group, while images are acted on by geometric transformations like translations, rotations or reflections. Invariant models are guaranteed to produce the same result when acting on their input. The outputs of equivariant models, on the other hand, transform in a predictable way with the input – the model commutes with the symmetry group actions on the input and output. Invariant maps are a special case of equivariant maps, where the group action on the model output is trivial. Both invariant and equivariant models generalize what they have learned over the symmetry group orbits, i.e. inputs that are related by the group action. Fig. 2.2 shows this in more detail for the case of the permutation invariant luggage classification in Fig. 2.1a.

(Vector graphics adapted under the Apache license 2.0 by courtesy of Google.)

Injecting this prior knowledge into the optimization problem can speed up training, regularize it against overfitting and improve the model accuracy [322].

The hypothesis space of invariant models can be understood as a *quotient hypothesis space*. To see this, note that invariant models assign the same response $m(x)$ to all elements on the *group orbit* (Def. B.3.3 and Fig. B.3)

$$G \triangleright_{\text{in}} x := \{g \triangleright_{\text{in}} x \mid g \in G\} \subseteq \mathcal{F}_{\text{in}} \quad (2.9)$$

of x in \mathcal{F}_{in} . Such models can therefore be viewed as assigning a response to orbits as a whole, instead of single elements. To make this precise, consider the *quotient space* (Def. B.3.4)

$$G \backslash \mathcal{F}_{\text{in}} := \{G \triangleright_{\text{in}} x \mid x \in \mathcal{F}_{\text{in}}\}, \quad (2.10)$$

consisting of all G -orbits in \mathcal{F}_{in} , and the corresponding *quotient map*

$$q_{\triangleright_{\text{in}}} : \mathcal{F}_{\text{in}} \rightarrow G \backslash \mathcal{F}_{\text{in}}, \quad x \mapsto G \triangleright_{\text{in}} x, \quad (2.11)$$

which sends elements in \mathcal{F}_{in} to their orbits. G -invariant models $m \in \mathcal{H}_{\text{inv}}$ are then in one-to-one correspondence to an unconstrained map $m^{\downarrow} : G \backslash \mathcal{F}_{\text{in}} \rightarrow \mathcal{F}_{\text{out}}$ on the quotient space, where $m = m^{\downarrow} \circ q_{\triangleright_{\text{in}}}$. This relation is visualized by the following commutative diagram:

$$\begin{array}{ccc} \mathcal{F}_{\text{in}} & \xrightarrow{m} & \mathcal{F}_{\text{out}} \\ q_{\triangleright_{\text{in}}} \downarrow & \nearrow m^{\downarrow} & \\ G \backslash \mathcal{F}_{\text{in}} & & \end{array} \quad (2.12)$$

The hypothesis space of invariance constrained maps can thus be viewed as the quotient hypothesis space

$$\mathcal{H}_{\text{inv}}^{\downarrow} := \{m^{\downarrow} : G \backslash \mathcal{F}_{\text{in}} \rightarrow \mathcal{F}_{\text{out}}\} \cong \mathcal{H}_{\text{inv}}. \quad (2.13)$$

Figs. 2.2 and 6.1 visualize this concept at concrete examples, which make the benefit of using symmetry constrained hypothesis spaces obvious.

A similar construction of quotient hypothesis spaces could be made for equivariant maps, however, it would be more technical since we would be required to work with quotient representatives (Def. B.3.5). The interested reader is pointed to Section 13.3, where we make this construction in a slightly different setting explicit.

2.2 Equivariant neural networks

A (feed forward) neural network is a sequence $L_N \circ L_{N-1} \circ \dots \circ L_3 \circ L_2 \circ L_1$ of parameterized layers $L_l : \mathcal{F}_{l-1} \rightarrow \mathcal{F}_l$, mapping between adjacent feature spaces:

$$\mathcal{F}_0 \xrightarrow{L_1} \mathcal{F}_1 \xrightarrow{L_2} \mathcal{F}_2 \xrightarrow{L_3} \dots \xrightarrow{L_{N-1}} \mathcal{F}_{N-1} \xrightarrow{L_N} \mathcal{F}_N \quad (2.14)$$

The feature spaces \mathcal{F}_l are usually defined to be vector spaces. The sequence of network layers is often composed of blocks which comprise a linear map, a bias summation and a nonlinearity.²

²Many other types of layers exist, including e.g. normalization layers or bilinear maps between features.

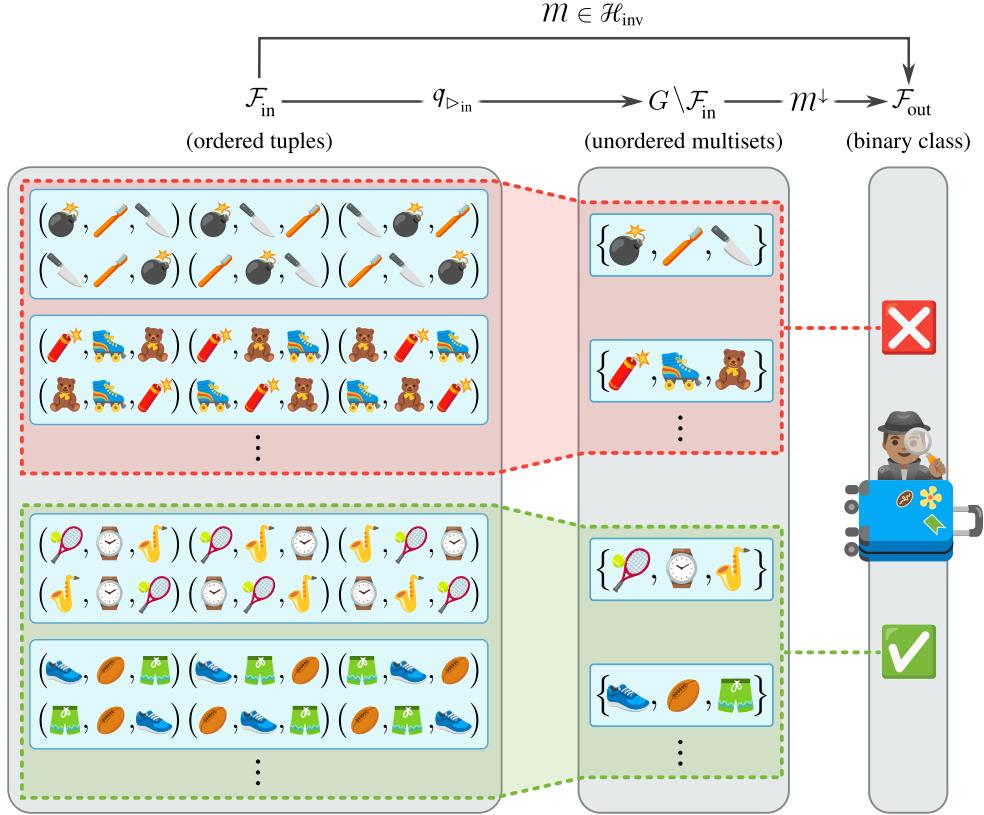


Figure 2.2: A detailed view on the permutation equivariant luggage inspection from Fig. 2.1a, visualizing the concepts of group orbits and quotient spaces. The elements in the luggage are *ordered 3-tuples* ($\text{red star}, \text{skateboard}, \text{teddy bear}$), acted on by the permutation group $G = S_3$. Whether a dangerous item is present in the tuple is independent from the specific ordering of the tuple elements – the classifier $m : \mathcal{F}_{\text{in}} \rightarrow \mathcal{F}_{\text{out}}$ should therefore be S_3 -*permutation invariant*, i.e. $m \in \mathcal{H}_{\text{inv}}$. This is equivalent to demanding that the model considers the luggage items as elements of an *unordered (multi)set* $\{\text{red star}, \text{skateboard}, \text{teddy bear}\}$. Formally, these multisets correspond to *group orbits* $S_3 \triangleright_{\text{in}} (\text{red star}, \text{skateboard}, \text{teddy bear}) \subset \mathcal{F}_{\text{in}}$ (blue boxes) of tuples, and can therefore be viewed as elements of the *quotient space* $S_3 \setminus \mathcal{F}_{\text{in}}$. The space of S_3 -invariant models $m \in \mathcal{H}_{\text{inv}}$ is equivalent to the quotient hypothesis space $\mathcal{H}_{\text{inv}}^{\downarrow}$ of unconstrained maps $m^{\downarrow} : S_3 \setminus \mathcal{F}_{\text{in}} \rightarrow \mathcal{F}_{\text{out}}$. Injecting such prior knowledge into the model by design greatly simplifies the learning problem and leads to an improved performance.

(Vector graphics adapted under the Apache license 2.0 by courtesy of Google.)

The most common way to construct equivariant networks is to design its layers such that they are *individually equivariant*.³ Each feature space \mathcal{F}_l is therefore to be equipped with its own group action

$$\triangleright_l : G \times \mathcal{F}_l \rightarrow \mathcal{F}_l, \quad (g, x) \mapsto g \triangleright_l x. \quad (2.15)$$

The actions \triangleright_0 and \triangleright_N on the network's input \mathcal{F}_0 and output \mathcal{F}_N are thereby specified by the learning task, while the intermediate actions \triangleright_l for $l = 1, \dots, N - 1$ are hyperparameters chosen by the user. Note that this is similar to the case of general neural networks, where the input and output feature spaces \mathcal{F}_0 and \mathcal{F}_N are given by the task, but the intermediate feature spaces \mathcal{F}_l for $l = 1, \dots, N - 1$ are user-chosen.

If each individual layer $L_l : \mathcal{F}_{l-1} \rightarrow \mathcal{F}_l$ is designed to be equivariant w.r.t. its input and output group actions, i.e. satisfies $L_l(g \triangleright_{l-1} x) = g \triangleright_l L_l(x)$ for any $g \in G$ and any $x \in \mathcal{F}_{l-1}$, the network is by induction equivariant as a whole. This is visualized by the following diagram, which commutes for any $g \in G$:

$$\begin{array}{ccccccccc} \mathcal{F}_0 & \xrightarrow{L_1} & \mathcal{F}_1 & \xrightarrow{L_2} & \mathcal{F}_2 & \xrightarrow{L_3} & \dots & \xrightarrow{L_{N-1}} & \mathcal{F}_{N-1} & \xrightarrow{L_N} & \mathcal{F}_N \\ \downarrow g \triangleright_0 & & \downarrow g \triangleright_1 & & \downarrow g \triangleright_2 & & & & \downarrow g \triangleright_{N-1} & & \downarrow g \triangleright_N \\ \mathcal{F}_0 & \xrightarrow{L_1} & \mathcal{F}_1 & \xrightarrow{L_2} & \mathcal{F}_2 & \xrightarrow{L_3} & \dots & \xrightarrow{L_{N-1}} & \mathcal{F}_{N-1} & \xrightarrow{L_N} & \mathcal{F}_N \end{array} \quad (2.16)$$

Invariant networks usually comprise 1) an equivariant subnetwork $L_N \circ \dots \circ L_1$, 2) an invariant map L_{inv} and 3) an unconstrained subnetwork $\tilde{L}_M \circ \dots \circ \tilde{L}_1$, operating on the resulting G -invariant features:⁴

$$\begin{array}{ccccccccc} \mathcal{F}_0 & \xrightarrow{L_1} & \dots & \xrightarrow{L_N} & \mathcal{F}_N & & & & \\ \downarrow g \triangleright_0 & & & & \downarrow g \triangleright_N & & & & \\ \mathcal{F}_0 & \xrightarrow{L_1} & \dots & \xrightarrow{L_N} & \mathcal{F}_N & & & & \\ & & & & & \searrow L_{\text{inv}} & & & \\ & & & & & \tilde{\mathcal{F}}_0 & \xrightarrow{\tilde{L}_1} & \dots & \xrightarrow{\tilde{L}_M} & \tilde{\mathcal{F}}_M \end{array} \quad (2.17)$$

An example are convolutional networks for translation invariant image classification, which usually composed of 1) a convolutional subnetwork, 2) a global pooling operation, and 3) an unconstrained MLP, operating on the resulting translation invariant features. A further variation are networks which become after some depth equivariant w.r.t. a subgroup of symmetries; see Section 4.4 and [322].

³An alternative approach is to use a non-equivariant network f , and feed *any* G -transformed input through it. The $|G|$ outputs $[f(gx)]_{g \in G}$ transform then according to the regular G -representation (Def. B.5.18), i.e. are permuted when acting on the input. Invariant responses could then be computed by taking the mean or the maximum over responses [174]. The disadvantage of this approach is that it enforces equivariance post-hoc, and does not reduce the number of model parameters.

⁴This design is again a special case of the fully equivariant formulation in Eq. (2.16), assuming trivial group actions for all layers \tilde{L}_l , $l = 0, \dots, M$. One may therefore w.l.o.g. consider “fully equivariant networks” as in Eq. (2.16).

We discuss the specific design of *equivariant MLPs* in Appendix E. The feature spaces are here *finite group representation spaces*, that is, finite vector spaces that are equipped with linear group actions (representations, Def. B.5.1). Linear equivariant layers between such representation spaces are intertwiners (Def. B.5.7). Appendix E characterizes these intertwiner spaces by decomposing them into their irreducible components.

For completeness, we need to mention that there are alternative approaches to construct equivariant networks. A simple approach is to use an unconstrained network and to train it to be approximately equivariant – this is usually achieved by *augmenting* the dataset with G -transformed samples. To give an example, one could use an MLP instead of a CNN to process images, and train them by feeding in translated versions of the images. This approach is clearly disadvantageous since the unconstrained network has a larger number of parameters and is therefore more prone to overfitting [106, 318]. An equivariant network, on the other hand, has a greatly reduced hypothesis space: instead of being required to learn the correct mapping for each G -related feature individually, it generalizes automatically over orbits in \mathcal{F}_{in} . Another approach to construct invariant networks was proposed by Laptev et al. [174]. The authors propose to use an unconstrained network, feed in all G -transformed versions of the input explicitly, and take the maximum response over the resulting responses. Since the G -action leads to a mere permutation of the responses, this design is G -invariant. The approach is, however computationally expensive and does not lead to a reduced number of parameters due to weight sharing.

Translation equivariance & conventional Euclidean CNNs

Spatially structured signals, like audio, images, voxel data, videos or physical fields on a flat spacetime, are with great success processed by Euclidean *convolutional neural networks* (CNNs) [175, 166]. As the name suggests, these models process the signals (or feature maps) by convolving them with some convolution kernel, the parameters of which are learned. The central difference in comparison to general linear maps is that the convolution operation is in addition *translation equivariant* – it commutes with translations of the feature maps. CNNs are therefore equivariant neural networks in the sense introduced in the previous chapter.

To make this precise, we introduce feature maps in the following Section 3.1 as *regular representations of the translation group*, i.e. equip them with a linear action of the translation group. Typical translation equivariant layers between such feature spaces are convolutions, bias summation, nonlinearities or local and global pooling layers. Instead of *defining* these layers and subsequently proving their equivariance, we show in Section 3.2 how they can be *derived* from the requirement for translation equivariance.

This formulation sets the stage for our definition of more general affine group equivariant (steerable) CNNs in Chapter 4. Fig. 3.2 gives an overview of the design choices that distinguish non-equivariant fully and locally connected networks, translation equivariant CNNs and affine group equivariant steerable CNNs. In a nutshell, translation equivariance requires translational weight sharing (e.g. convolutions), while affine group equivariance requires the shared neural connectivity additionally to be G -steerable (G -equivariant).

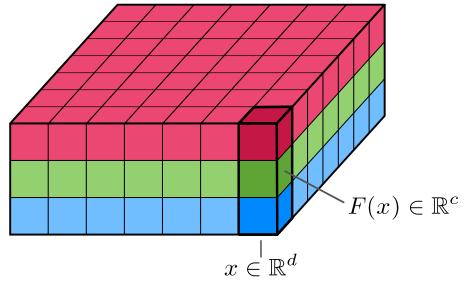
3.1 Euclidean feature maps as translation group representations

An *Euclidean feature map* in d dimensions with c *channels* is a function

$$F : \mathbb{R}^d \rightarrow \mathbb{R}^c \tag{3.1}$$

that assigns a c -dimensional feature vector $F(x)$ to each point $x \in \mathbb{R}^d$; see Fig. 3.1. An example are *audio signals* on $d = 1$ time dimension and with $c = 1$ or 2 channels for mono or stereo sources, respectively. *Images* on $d = 2$ spatial dimensions come usually with $c = 1, 3, 4$ or more channels, and are then denoted as grayscale images, RGB images, RGBA images or hyperspectral images, respectively. Volumetric data lives in $d = 3$ spatial dimensions, while videos are feature maps on $d = 2 + 1$ spacetime dimensions. The number of channels in intermediate layers of the network are user-chosen hyperparameters.

Figure 3.1: Euclidean feature maps are functions $F : \mathbb{R}^d \rightarrow \mathbb{R}^c$ that assign c -dimensional feature vectors $F(x) \in \mathbb{R}^c$ to each point $x \in \mathbb{R}^d$. The regular representation of the translation group $(\mathbb{R}^d, +)$ acts on feature maps by shifting them spatially; see Eq. (3.5). Shown here is a $c = 3$ channel RGB image in $d = 2$ spatial dimensions, discretized on a pixel grid. Feature fields, visualized in Figs. 4.3 and 4.4, generalize feature maps by equipping them with an affine group action.



The shift of an audio signal in time or of an image in space is modeled by an action of the *translation group* of the corresponding dimensionality. In d dimensions, the translation group is defined as the tuple $(\mathbb{R}^d, +)$, i.e. the set \mathbb{R}^d with addition $+ : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$, $(t, s) \mapsto t + s$ as binary operation. The inverse of $t \in \mathbb{R}^d$ is accordingly $-t$ and the identity element is $0 \in \mathbb{R}^d$.

The translation group acts naturally on the Euclidean space \mathbb{R}^d by shifting its points:

$$(\mathbb{R}^d, +) \times \mathbb{R}^d \rightarrow \mathbb{R}^d, \quad (t, x) \mapsto x + t \quad (3.2)$$

It acts furthermore on Euclidean feature maps $F : \mathbb{R}^d \rightarrow \mathbb{R}^c$ by sending them to shifted feature maps $[t \triangleright F] : \mathbb{R}^d \rightarrow \mathbb{R}^c$, defined by

$$[t \triangleright F](x) := F(x - t). \quad (3.3)$$

The feature spaces of translation equivariant Euclidean CNNs are spaces of such feature maps, which we formalize in the following definition. Since we want to apply convolutions, we require the feature maps in addition to be square integrable.

Definition 3.1.1 (Euclidean feature maps as regular translation group representations).

The feature spaces of translation equivariant Euclidean CNNs are vector spaces

$$L^2(\mathbb{R}^d, \mathbb{R}^c) := \left\{ F : \mathbb{R}^d \rightarrow \mathbb{R}^c \mid \int_{\mathbb{R}^d} dx \|F(x)\|^2 \leq \infty \right\} \quad (3.4)$$

of square integrable c -channel feature maps in d dimensions, equipped with the translation group action

$$\triangleright : (\mathbb{R}^d, +) \times L^2(\mathbb{R}^d, \mathbb{R}^c) \rightarrow L^2(\mathbb{R}^d, \mathbb{R}^c), \quad (t, F) \mapsto t \triangleright F \quad (3.5)$$

defined by

$$[t \triangleright F](x) := F(x - t). \quad (3.6)$$

This action corresponds, in fact, to a linear representation (Def. B.5.1) $(\mathbb{R}^d, +) \rightarrow \text{GL}(L^2(\mathbb{R}^d, \mathbb{R}^c))$ of the translation group, which is known as regular representation¹ (Def. B.5.18).

Note that such defined feature maps do not have a well defined behavior under other geometric transformations like rotations, reflections, scaling or shearing. This shortcoming will be alleviated in Chapter 4 where we define *Euclidean feature fields*, which are affine group representations.

¹More precisely, for c channels, this would be a direct sum of c regular representations.

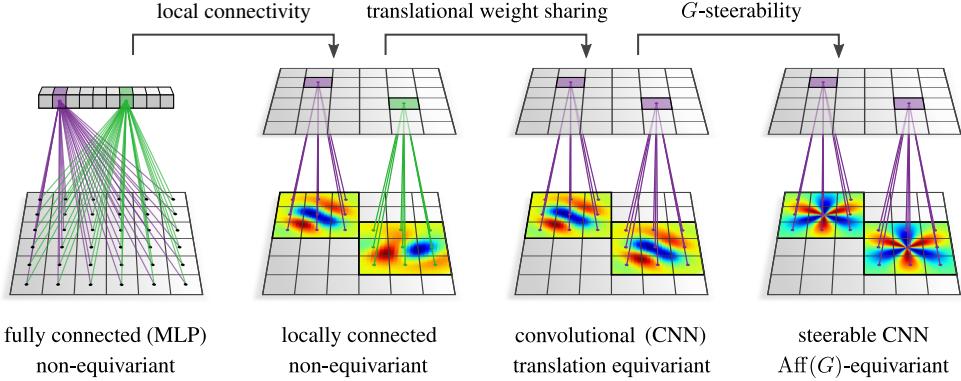


Figure 3.2: Architectural design choices that turn a fully connected network (MLP) into a convolutional network. *Left:* Each neuron of an unconstrained MLP connects with its own weights to each pixel of the input image. If there are M input pixels and N hidden neurons, this requires $\mathcal{O}(MN)$ parameters. Fully connected networks have no equivariance guarantees. *Middle left:* Restricting each neuron’s connectivity to a local receptive field reduces the number of parameters to $\mathcal{O}(M)$. Being associated to a specific region in the input image, the hidden neurons are naturally organized in a feature map. However, this network design still does not guarantee any equivariance properties. *Middle right:* Convolutional networks share synapse weights over spatial locations, which can be thought of as convolving the image with a convolution kernel. The translational weight sharing ensures the model’s translation equivariance. The parameter cost is independent of the image size, i.e. $\mathcal{O}(1)$ (it still scales with the number of channels and kernel size). *Right:* If the convolution kernel itself is additionally G -equivariant (steerable), where $G \leq \mathrm{GL}(d)$ could for instance model rotations, reflections, scaling or shearing, the resulting steerable CNN is equivariant under the affine group $\mathrm{Aff}(G) = (\mathbb{R}^d, +) \rtimes G$, i.e. simultaneous translations and G -transformations. The G -steerability constraint on the kernels can be viewed as a form of weight sharing over G -transformations and reduces the model’s parameter cost further. Chapter 3 derives the convolutional network design from the requirement for translation equivariance. Chapter 4 derives the additional G -steerability constraint on the kernels from the requirement for $\mathrm{Aff}(G)$ -equivariance. Note that the local connectivity (local kernel support) is not actually required for the model’s equivariance, but is rather an empirical design choice. We are furthermore modeling CNNs on continuous spaces \mathbb{R}^d , instead of discrete pixel grids \mathbb{Z}^d visualized above.

3.2 Translation equivariant layers and convolutions

Translation equivariant network layers between feature maps with c_{in} input channels and c_{out} output channels are functions $L : L^2(\mathbb{R}^d, \mathbb{R}^{c_{\text{in}}}) \rightarrow L^2(\mathbb{R}^d, \mathbb{R}^{c_{\text{out}}})$ such that the following diagram commutes for arbitrary translations $t \in (\mathbb{R}^d, +)$:

$$\begin{array}{ccc}
 L^2(\mathbb{R}^d, \mathbb{R}^{c_{\text{in}}}) & \xrightarrow{L} & L^2(\mathbb{R}^d, \mathbb{R}^{c_{\text{out}}}) \\
 \downarrow t \triangleright_{\text{in}} & & \downarrow t \triangleright_{\text{out}} \\
 L^2(\mathbb{R}^d, \mathbb{R}^{c_{\text{in}}}) & \xrightarrow{L} & L^2(\mathbb{R}^d, \mathbb{R}^{c_{\text{out}}})
 \end{array} \tag{3.7}$$

We derive in the following some of the most common CNN layers from the demand for translation equivariance.

3.2.1 Convolutions as translation equivariant linear maps

Linear translation equivariant functions between feature maps (regular translation group representations) are necessarily convolutions. In the most general case, we would have to consider *distributional convolutions*, including for instance the application of *partial differential operators* [137]. For simplicity, we will here focus on convolutions with kernels that are conventional functions (instead of general Schwartz distributions).

Our ansatz for an unconstrained linear map between feature maps is therefore given by an *integral transform*

$$\mathbb{I}_\kappa : L^2(\mathbb{R}^d, \mathbb{R}^{c_{\text{in}}}) \rightarrow L^2(\mathbb{R}^d, \mathbb{R}^{c_{\text{out}}}) \quad (3.8)$$

that is parameterized by a (square integrable) *matrix-valued two-argument kernel*

$$\kappa : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \quad (3.9)$$

and defined by²

$$\mathbb{I}_\kappa[F](x) := \int_{\mathbb{R}^d} dy \kappa(x, y) F(y); \quad (3.10)$$

see Fig. 3.3 for a visualization. The matrix-valued codomain of the kernel is thereby ensuring that the c_{in} input channels are mapped to c_{out} output channels. As proven in the following theorem, such linear maps are necessarily convolutions if they are required to be translation equivariant:

Theorem 3.2.1 (Regular translation intertwiners are convolutions). *The integral transform \mathbb{I}_κ in Eq. (3.8) is translation equivariant w.r.t. the regular translation group action on feature maps in Eq. (3.5), i.e. satisfies*

$$\mathbb{I}_\kappa[t \triangleright_{\text{in}} F] = t \triangleright_{\text{out}} \mathbb{I}_\kappa[F] \quad \forall t \in (\mathbb{R}^d, +), \quad (3.11)$$

if and only if the two-argument kernel satisfies $\kappa(x + t, y + t) = \kappa(x, y)$ for any $x, y, t \in \mathbb{R}^d$. Such kernels depend only on the relative distance between x and y and are therefore equivalent to a one-argument kernel

$$K : \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}, \quad \Delta x \mapsto K(\Delta x) := \kappa(\Delta x, 0). \quad (3.12)$$

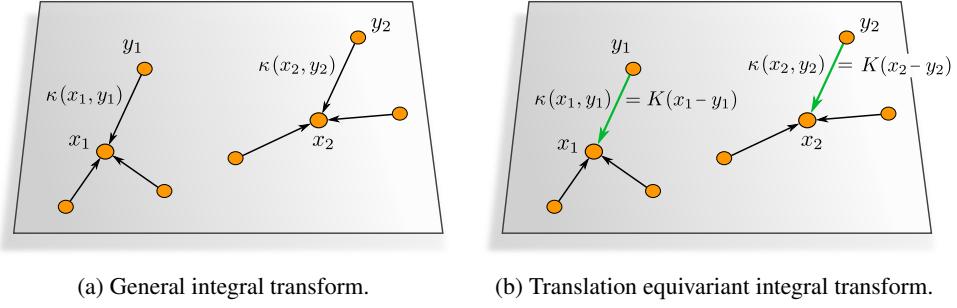
The integral transform reduces thus to a convolution integral

$$\mathbb{I}_\kappa[F](x) = [K * F](x) = \int_{\mathbb{R}^d} dy K(x - y) F(y). \quad (3.13)$$

Proof: Writing out the left-hand side of Eq. (3.11) yields

$$\begin{aligned} \mathbb{I}_\kappa[t \triangleright_{\text{in}} F](x) &= \int_{\mathbb{R}^d} dy \kappa(x, y) [t \triangleright_{\text{in}} F](y) \\ &= \int_{\mathbb{R}^d} dy \kappa(x, y) F(y - t) \\ &= \int_{\mathbb{R}^d} d\tilde{y} \kappa(x, \tilde{y} + t) F(\tilde{y}), \end{aligned} \quad (3.14)$$

²In components, $\mathbb{I}_\kappa[F]_i(x) = \sum_{j=1}^{c_{\text{in}}} \int_{\mathbb{R}^d} dy \kappa_{ij}(x, y) F_j(y)$, where $i = 1, \dots, c_{\text{out}}$ labels the output channels.



(a) General integral transform.

(b) Translation equivariant integral transform.

Figure 3.3: The general integral transform \mathbf{I}_κ in Eq. (3.8) maps a c_{in} -channel input feature map linearly to a c_{out} -channel output feature map. Output feature vectors $\mathbf{I}_\kappa[F](x) \in \mathbb{R}^{c_{\text{out}}}$ at $x \in \mathbb{R}^d$ are thereby computed by accumulating input feature vectors $F(y) \in \mathbb{R}^{c_{\text{in}}}$ via a matrix-valued two-argument kernel $\kappa : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$, integrating over all points $y \in \mathbb{R}^d$. A-priori, the kernel values are entirely unconstrained. However, when demanding *translation equivariance*, the kernel values $\kappa(x, y)$ may only depend on the *relative distance* $x - y$ of the arguments; see Fig. 3.3b. As proven in Theorem 3.2.1, this implies that the integral transform can be written as a *convolution integral* with a spatially shared single (relative) argument kernel $K : \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$.

while the right-hand side is given by

$$[t \triangleright_{\text{out}} \mathbf{I}_\kappa[F]](x) = \int_{\mathbb{R}^d} dy \kappa(x - t, y) F(y). \quad (3.15)$$

Demanding both sides to be equal for any translation $t \in (\mathbb{R}^d, +)$ and any feature map $F \in L^2(\mathbb{R}^d, \mathbb{R}^{c_{\text{in}}})$ implies the *translation invariance constraint*

$$\kappa(x + t, y + t) = \kappa(x, y) \quad \forall x, y, t \in \mathbb{R}^d \quad (3.16)$$

on the neural connectivity (spatial weight sharing) that was claimed in the theorem.

We can now w.l.o.g. choose $t = -y$ in Eq. (3.16), resulting in $\kappa(x, y) = \kappa(x - y, 0)$, and define the one-argument kernel $K(\Delta x) := \kappa(\Delta x, 0)$. This proves that the requirement for translation equivariance makes the integral transform a convolution. \square

The convolution with a kernel $K : \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ is essentially applying this kernel at each point to the image – it is therefore often thought of as a *translational weight sharing* of the neural connectivity. Convolution kernels are in practice often compactly supported, such that each neuron is only connected to a local receptive field; see Fig. 3.2. This design choice is not strictly necessary for the network’s translation equivariance, but reduces its parameter cost.

Theorem 4.3.1 in Section 4.3 generalizes these considerations to more general affine symmetry groups, resulting in an additional equivariance constraint on the convolution kernel itself.

3.2.2 Translation equivariant bias summation

After applying a linear operation, it is common to sum a learned bias to the network's features. While one could in principle sum an arbitrary *bias field* $\mathbf{b} : \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{in}}}$ to the feature maps, the requirement for translation equivariance demands that this bias field is translation invariant, i.e. spatially constant, as we will show in the following. Note that bias summation operations preserve the number of feature channels, i.e. $c_{\text{in}} = c_{\text{out}} =: c$, and therefore in particular the translation group action $\triangleright_{\text{in}} = \triangleright_{\text{out}} =: \triangleright$ on their input and output.

To derive the spatial bias weight sharing, consider an unconstrained bias summation operation

$$B_{\mathbf{b}} : L^2(\mathbb{R}^d, \mathbb{R}^c) \rightarrow L^2(\mathbb{R}^d, \mathbb{R}^c), \quad F \mapsto F + \mathbf{b} \quad (3.17)$$

that is parameterized by a general square integrable bias field

$$\mathbf{b} \in L^2(\mathbb{R}^d, \mathbb{R}^c). \quad (3.18)$$

Theorem 3.2.2 (Translation equivariant bias summation). *The bias summation layer in Eq. (3.17) is translation equivariant w.r.t. the regular translation group action on feature maps in Eq. (3.5), i.e. satisfies*

$$B_{\mathbf{b}}[t \triangleright F] = t \triangleright B_{\mathbf{b}}[F], \quad (3.19)$$

if and only if the bias field is constant:

$$\mathbf{b}(x) = b \quad \text{for some} \quad b \in \mathbb{R}^c \quad (3.20)$$

Proof: The left-hand side of Eq. (3.19) at $x \in \mathbb{R}^d$ becomes

$$B_{\mathbf{b}}[t \triangleright F](x) = [t \triangleright F](x) + \mathbf{b}(x) = F(x - t) + \mathbf{b}(x), \quad (3.21)$$

while the right-hand side is given by

$$[t \triangleright B_{\mathbf{b}}[F]](x) = [t \triangleright (F + \mathbf{b})](x) = F(x - t) + \mathbf{b}(x - t). \quad (3.22)$$

Equating the two expressions results in $\mathbf{b}(x) = \mathbf{b}(x - t)$ for arbitrary $x, t \in \mathbb{R}^d$ – the bias field is thus required to be translation invariant, i.e. constant. \square

3.2.3 Translation equivariant local nonlinearities

Since any sequence of linear operations and bias summation collapses to a single affine transformation, one alternates them with *nonlinearities*. The space of possible nonlinearities is vast, however, convolutional networks usually rely on such nonlinearities that *act locally*, i.e. individually on each feature vector $F(x) \in \mathbb{R}^c$ at $x \in \mathbb{R}^d$. Similar to the general bias fields above, these nonlinearities could again be position-dependent, but the requirement for translation equivariance demands them to be translationally invariant.

To prove this claim, consider a *spatially dependent* localized nonlinearity ansatz

$$\sigma : \mathbb{R}^d \times \mathbb{R}^{c_{\text{in}}} \rightarrow \mathbb{R}^{c_{\text{out}}}, \quad (x, f) \mapsto \sigma_x(f), \quad (3.23)$$

mapping a c_{in} -dimensional input feature vector f to a c_{out} -dimensional output feature vector. Let its action on feature fields be given by the operation

$$\mathfrak{S}_\sigma : L^2(\mathbb{R}^d, \mathbb{R}^{c_{\text{in}}}) \rightarrow L^2(\mathbb{R}^d, \mathbb{R}^{c_{\text{out}}}), \quad (3.24)$$

defined by

$$\mathfrak{S}_\sigma[F](x) := \sigma_x(F(x)) \quad (3.25)$$

Theorem 3.2.3 (Translation equivariant local nonlinearities). *The spatially dependent localized nonlinearity operation in Eq. (3.24) is translation equivariant w.r.t. the regular translation group action on feature maps in Eq. (3.5), i.e. satisfies*

$$\mathfrak{S}_\sigma[t \triangleright_{\text{in}} F] = t \triangleright_{\text{out}} \mathfrak{S}_\sigma[F], \quad (3.26)$$

if and only if the underlying field of localized nonlinearities is position-independent:

$$\sigma_x = \sigma : \mathbb{R}^{C_{\text{in}}} \rightarrow \mathbb{R}^{C_{\text{out}}} \quad \text{for any } x \in \mathbb{R}^d \quad (3.27)$$

Proof: Again, we write out the left-hand side of the constraint in Eq. (3.26),

$$\mathfrak{S}_\sigma[t \triangleright_{\text{in}} F](x) = \sigma_x([t \triangleright_{\text{in}} F](x)) = \sigma_x(F(x - t)), \quad (3.28)$$

and demand that it equals the right-hand side

$$[t \triangleright_{\text{out}} \mathfrak{S}_\sigma[F]](x) = \mathfrak{S}_\sigma[F](x - t) = \sigma_{x-t}(F(x - t)) \quad (3.29)$$

for any $x, t \in \mathbb{R}^d$. This results in the claimed position dependence $\sigma_x = \sigma_{x-t} := \sigma$ of the underlying pointwise nonlinearity field. \square

3.2.4 Translation equivariant local pooling operations

Besides convolutions, bias summation and nonlinearities, most CNN architectures rely on pooling operations. These pooling operations come in two flavors, namely *local pooling operations*, which are (faithfully) *translation equivariant*, and *global pooling operations*, which produce a *translation invariant* output, as required for instance for classification tasks. We start with local max pooling and average pooling and discuss the global operations in the next paragraph. Note that local pooling operations (with single pixel stride) are in discretized implementations on pixel grids often followed by a s pixel stride subsampling step, which is only partially equivariant as discussed below.

Local max pooling: *Local max pooling* is a nonlinear operation that produces a feature field whose value at $x \in \mathbb{R}^d$ is given by the *channel-wise maximum* of feature values in some *pooling region* $\mathcal{R}_x \subset \mathbb{R}^d$ around $x \in \mathbb{R}^d$:

$$\text{local_max_pool} : L^2(\mathbb{R}^d, \mathbb{R}^c) \rightarrow L^2(\mathbb{R}^d, \mathbb{R}^c), \quad F \mapsto \max_{y \in \mathcal{R}_x} F(y) \quad (3.30)$$

If the pooling regions at different points agree, this operation is translation equivariant:

Theorem 3.2.4 (Translation equivariance of local max pooling). *The local max pooling operation in Eq. (3.30) is translation equivariant, that is,*

$$\text{local_max_pool}[t \triangleright F](x) = (t \triangleright \text{local_max_pool}[F])(x) \quad (3.31)$$

for any $x \in \mathbb{R}^d$ and any $t \in (\mathbb{R}^d, +)$, if and only if the pooling windows are translation invariant (spatially shared), i.e. satisfy

$$t^{-1}\mathcal{R}_x = \mathcal{R}_{x-t} \quad \forall x \in \mathbb{R}^d, t \in (\mathbb{R}^d, +), \quad (3.32)$$

Proof: The statement follows by expanding the left-hand side

$$\begin{aligned} \text{local_max_pool}[t \triangleright F](x) &= \max_{y \in \mathcal{R}_x} F(y - t) \\ &= \max_{y \in t^{-1}\mathcal{R}_x} F(y) \end{aligned} \quad (3.33)$$

and the right-hand side

$$\begin{aligned}(t \triangleright \text{local_max_pool}[F])(x) &= \text{local_max_pool}[F](x - t) \\ &= \max_{y \in \mathcal{R}_{x-t}} F(y)\end{aligned}\quad (3.34)$$

and demanding that they agree for any feature map $F \in L^2(\mathbb{R}^d, \mathbb{R}^c)$ and arbitrary positions $x \in \mathbb{R}^d$ and translations $t \in (\mathbb{R}^d, +)$. \square

Local average pooling: Instead of taking the single maximal response in a pooling window, *local average pooling* computes their *channel-wise average*. If we allow for a weighted average, this operation is just a (channel-wise) convolution

$$\text{local_avg_pool}_k : L^2(\mathbb{R}^d, \mathbb{R}^c) \rightarrow L^2(\mathbb{R}^d, \mathbb{R}^c), \quad F \mapsto k * F \quad (3.35)$$

with a *scalar weighting kernel* $k : \mathbb{R}^d \rightarrow \mathbb{R}$.

Theorem 3.2.5 (Translation equivariance of local average pooling). *The local average pooling operation in Eq. (3.35) is by construction translation equivariant, that is,*

$$\text{local_avg_pool}[t \triangleright F](x) = (t \triangleright \text{local_avg_pool}[F])(x) \quad (3.36)$$

for any $x \in \mathbb{R}^d$ and any $t \in (\mathbb{R}^d, +)$.

Proof: This is a special case of Theorem 3.2.1, applied to each channel individually. \square

Subsampling: In practice, it is common to have *sampled feature maps* $F : \mathbb{Z}^d \rightarrow \mathbb{R}^d$ on a pixel grid \mathbb{Z}^d , equipped with the regular action of the discrete translation group $(\mathbb{Z}^d, +)$. The local pooling operations are then usually followed by a *subsampling* operation with a stride of $s \in \mathbb{N}$ pixels,

$$\text{subsample}_s : L^2(\mathbb{Z}^d, \mathbb{R}^c) \rightarrow L^2(s\mathbb{Z}^d, \mathbb{R}^c), \quad F \mapsto F|_{s\mathbb{Z}^d}, \quad (3.37)$$

given by a restriction of the domain and resulting in a feature map on the subgrid $s\mathbb{Z}^d$. This operation is *not* fully $(\mathbb{Z}^d, +)$ translation equivariant, but equivariant w.r.t. the subgroup $(s\mathbb{Z}^d, +)$ of translations by s pixels.

Theorem 3.2.6 (Translation subgroup equivariance of subsampling).

Let $\triangleright_{\mathbb{Z}^d}$ be the group action of the translation subgroup $(s\mathbb{Z}^d, +)$ on $L^2(\mathbb{Z}^d, \mathbb{R}^c)$ defined by $[t \triangleright_{\mathbb{Z}^d} F](x) := F(x - t)$ and let $\triangleright_{s\mathbb{Z}^d}$ be its restriction to an action on subsampled feature maps $L^2(s\mathbb{Z}^d, \mathbb{R}^c)$. The subsampling operation in Eq. (3.37) commutes with these actions, that is, for any $x \in s\mathbb{Z}^d$ and $t \in (s\mathbb{Z}^d, +)$ it holds that:

$$\text{subsample}_s[t \triangleright_{\mathbb{Z}^d} F](x) = (t \triangleright_{s\mathbb{Z}^d} \text{subsample}_s F)(x) \quad (3.38)$$

Proof: This claim is obvious since subsample_s is the restriction map and $\triangleright_{s\mathbb{Z}^d}$ is the correspondingly restricted group action. \square

Note that a repeated subsampling by strides of s_1, \dots, s_N pixels results in a network that is as a whole only $(\mathbb{Z}_{\prod_{i=1}^N s_i}, +)$ -equivariant. Zhang [348] analyzes this issue empirically and proposes to alleviate it by preceding the subsampling step with a low-pass filtering. The exactly $(\mathbb{Z}^d, +)$ -equivariant subsampling operation by Xu et al. [336] relies on choosing a shifted subsampling grid, corresponding to a choice of coset $t + s\mathbb{Z} \in \mathbb{Z}/s\mathbb{Z}$ (Def. B.2.2), in an equivariant (data-dependent) way.

3.2.5 Translation invariant global pooling operations

Global max pooling and global average pooling result in translation invariant features.

Global max pooling: The former is defined by

$$\text{global_max_pool} : L^2(\mathbb{R}^d, \mathbb{R}^c) \rightarrow \mathbb{R}^c, \quad F \mapsto \max_{x \in \mathbb{R}^d} F(x), \quad (3.39)$$

where the process of taking the maximum value is again understood to be taken independently for each channel.

Theorem 3.2.7 (Translation invariance of global max pooling). *The global max pooling operation in Eq. (3.39) is translation invariant, that is,*

$$\text{global_max_pool}[t \triangleright F](x) = \text{global_max_pool}[F](x) \quad (3.40)$$

for any $x \in \mathbb{R}^d$ and $t \in (\mathbb{R}^d, +)$.

Proof: The maximum value depends obviously not on its specific location ($\arg \max$). \square

Global average pooling: Global average pooling is a map

$$\text{global_avg_pool} : L^2(\mathbb{R}^d, \mathbb{R}^c) \rightarrow \mathbb{R}^c, \quad F \mapsto \int_{\mathbb{R}^d} dx F(x) \quad (3.41)$$

which averages the feature map over the whole Euclidean space.

Theorem 3.2.8 (Translation invariance of global average pooling). *The global average pooling operation in Eq. (3.41) is translation invariant, i.e. for any $x \in \mathbb{R}^d$ and $t \in (\mathbb{R}^d, +)$ one has*

$$\text{global_avg_pool}[t \triangleright F](x) = \text{global_avg_pool}[F](x). \quad (3.42)$$

Proof: The translation invariance follows from a simple substitution,

$$\begin{aligned} \text{global_avg_pool}[t \triangleright F] &= \int_{\mathbb{R}^d} dx F(x - t) = \int_{\mathbb{R}^d} d\tilde{x} F(\tilde{x}) \\ &= \text{global_avg_pool}(F), \end{aligned} \quad (3.43)$$

where $t \in (\mathbb{R}^d, +)$ is arbitrary. \square

Affine group equivariance & steerable Euclidean CNNs

There are many cases in which a signal processing task should not only be translation equivariant, but be equivariant under the action of *affine groups* $\text{Aff}(G)$. Besides translations, affine groups contain G -transformations, where the structure group $G \leq \text{GL}(d)$ could for instance model rotations, reflections, scaling or shearing of the signal. A prototypical use case are medical or satellite imaging tasks, where the rotation or reflection of the image is irrelevant; see Fig. 1.2.

In order to construct affine group equivariant network layers, we need to specify affine group actions on the feature spaces. A natural and quite general choice are *induced representations*, acting on *feature vector fields*. In addition to moving feature vectors spatially over \mathbb{R}^d , induced representations act on the feature vectors $F(x) \in \mathbb{R}^c$ themselves via some G -representation $\rho : G \rightarrow \text{GL}(c)$. Fig. 4.3 shows scalar fields and vector fields as specific examples of feature fields. Feature fields with a reflection group action are visualized in Fig. 5.1.

Steerable CNNs consist of $\text{Aff}(G)$ -equivariant maps between such feature fields. Most notably, linear equivariant maps between feature fields (induced representation intertwiners) are shown to be *convolutions with G -steerable kernels*, i.e. matrix-valued kernels satisfying a linear G -equivariance constraint; see Fig. 4.1. More generally, the *neural connectivity is required to be shared over $\text{Aff}(G)$ -transformations* – in addition to a spatial weight sharing, this requires a local G -equivariance of the shared operations.

After briefly introducing *affine groups* in Section 4.1, we define *Euclidean feature fields* in Section 4.2 as induced affine group representations. Section 4.3 derives steerable convolutions and other *affine equivariant network layers* like steerable bias summation, nonlinearities and pooling operations. While this section *derives* the G -steerability constraint on convolution kernels, a review of their construction and implementation is deferred to the following Chapter 5. Section 4.4 introduces an *equivariance group restriction* operation, which allows to build networks with different levels of equivariance in different layers. Most work on equivariant CNNs relies on *regular group convolutions* instead of general steerable convolutions. Section 4.5 clarifies the relation between the two, showing that group convolutions are a special case of steerable convolutions for regular inducing representations.

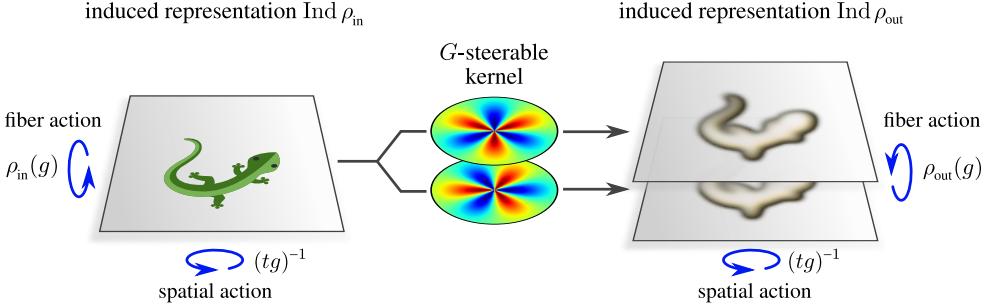


Figure 4.1: Schematic idea of convolutions with G -steerable kernels as induced representation intertwiners. The feature fields $F : \mathbb{R}^d \rightarrow \mathbb{R}^c$ of steerable CNNs are *affine group representations* which are induced from some G -representation $\rho : G \rightarrow \text{GL}(c)$. Given a group element $tg \in \text{Aff}(G)$, a feature field F transforms according to $[\text{Ind } \rho(tg)] F = \rho(g) F (tg)^{-1}$, that is, 1) a *spatial transformation* of the domain \mathbb{R}^d by $(tg)^{-1}$ and 2) a *fiber action* on the codomain \mathbb{R}^c by $\rho(g)$ (mixing channels). Theorem 4.3.1 proves that the most general linear $\text{Aff}(G)$ -equivariant integral transform (intertwiner) between induced representations $\text{Ind } \rho_{\text{in}}$ and $\text{Ind } \rho_{\text{out}}$ is given by *convolutions with G -steerable kernels* $K : \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$, i.e. matrix-valued kernels subject to the linear G -equivariance constraint in Eq. (4.18). For simplicity, we visualized the case of a $c_{\text{in}} = 1$ -dimensional scalar input field (trivial representation ρ_{in}) and a $c_{\text{out}} = 2$ -dimensional output field, transforming according the $\text{SO}(2)$ -irrep with frequency 3 – requiring a $\text{SO}(2)$ -steerable kernel of shape 2×1 and angular frequency 3; see Table 5.2. More generally, the input would have multiple channels and the matrix-valued kernel would comprise $c_{\text{out}} \times c_{\text{in}}$ elementary scalar kernels.

Lizards adapted under the Creative Commons Attribution 4.0 International license by courtesy of Twitter.)

An *empirical evaluation* of Euclidean steerable CNNs is presented in Chapter 6. For a *differential geometric formulation* of affine equivariant steerable CNNs, which discusses in particular their coordinate independence and local gauge equivariance, we point the reader to Chapter 15.

The theory of Euclidean steerable CNNs is largely understood, however, an implementation supporting general field types requires quite some experience and effort. To take this burden from the user, we provide the PyTorch extension `e2cnn` [38] and its successor `escnn` [39], which implement feature fields and equivariant layers for arbitrary representations ρ of compact structure groups $G \leq \text{O}(2)$ and $G \leq \text{O}(d)$, respectively. The frontend is thereby designed such that the user only has to select a symmetry group and its actions on feature spaces (field types), while all of the representation theory is hidden in the backend.

4.1 Affine groups

Affine groups $\text{Aff}(G) := (\mathbb{R}^d, +) \rtimes G$ are semidirect products (Def. B.2.5) of translations and some *structure group* $G \leq \text{GL}(d)$. This includes for instance the Euclidean groups $\text{E}(d)$, i.e. isometries of \mathbb{R}^d , for $G = \text{O}(d)$, or the special Euclidean groups $\text{SE}(d)$ for $G = \text{SO}(d)$. Pure translations are covered for the trivial structure group $G = \{e\}$.

As usual for semidirect products, any group element $tg \in \text{Aff}(G)$ can be uniquely split into factors, here a translation $t \in (\mathbb{R}^d, +)$ and an element $g \in G$. The affine group acts naturally

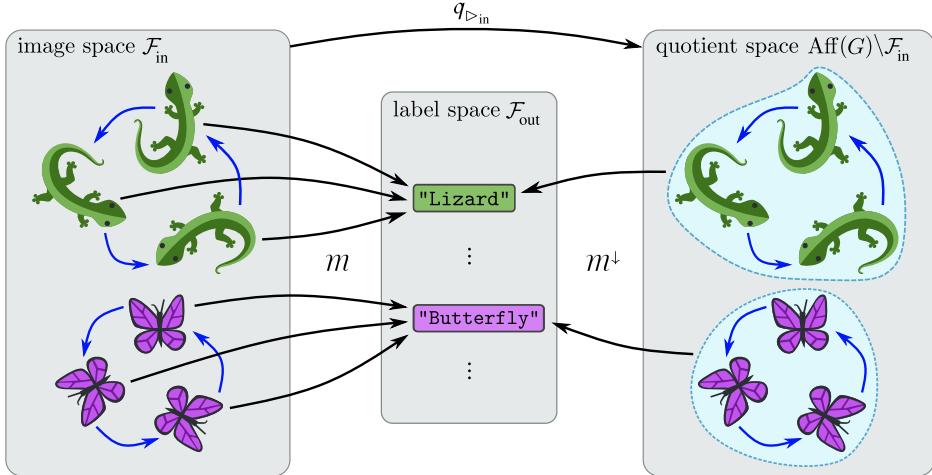


Figure 4.2: Affine equivariant CNNs generalize whatever they learn over $\text{Aff}(G)$ -orbits of images, i.e. subsets of images that are related by translations and G -transformations. For instance, an $\text{Aff}(\text{SO}(2))$ -*invariant* classification model is guaranteed to assign the same (invariant) class label to any translated or rotated version of the same image. It can therefore be viewed as effectively operating on the quotient image space $\text{Aff}(G) \backslash \mathcal{F}_{\text{in}}$, as explained in Section 2.1 and Fig. 2.2. More general $\text{Aff}(G)$ -*equivariant* CNNs respond with an affine group transformation of their output in \mathcal{F}_{out} if their input in \mathcal{F}_{in} is transformed. Section 6.1 investigates the generalization properties of $\text{Aff}(G)$ -steerable CNNs empirically. Lizards and butterflies adapted under the Creative Commons Attribution 4.0 International license by courtesy of Twitter.)

on Euclidean spaces \mathbb{R}^d :

$$\text{Aff}(G) \times \mathbb{R}^d \rightarrow \mathbb{R}^d, \quad (tg, x) \mapsto gx + t \quad (4.1)$$

Structure group elements act hereby by matrix multiplication, which is possible since the structure group is by definition a matrix group. For convenience, we note that the action of an inverse group element is given by $((tg)^{-1}, x) \mapsto g^{-1}(x - t)$, which is easily confirmed by composing the actions of tg and $(tg)^{-1}$.

Since affine groups contain translations $(\mathbb{R}^d, +) \leq \text{Aff}(G)$ as subgroups, their action on \mathbb{R}^d is *transitive* (Def. B.3.8). It is for $G \neq \{e\}$ not fixed point free (Def. B.3.10), but has stabilizer subgroups $\text{Stab}_x \leq \text{Aff}(G)$ (Def B.3.6) that are isomorphic to G . As we will see in Section 4.3 below, these local G -symmetries will result in a requirement for the operations' local G -steerability.

For completeness, we mention that affine groups may be viewed as a *principal G -bundle* $q : \text{Aff}(G) \rightarrow \text{Aff}(G)/G \cong \mathbb{R}^d$ over a Euclidean base space, where the quotient map $q : tg \mapsto tG$ serves as a bundle projection. The interpretation of G as *structure group* relies on this identification. More details on principal bundles in general can be found in Section 11.1 while the specific case of $\text{Aff}(G)$ as principal bundle and its local gauge transformations are discussed in Chapter 15. The reader may safely ignore the interpretation of $\text{Aff}(G)$ as principal bundle for now.

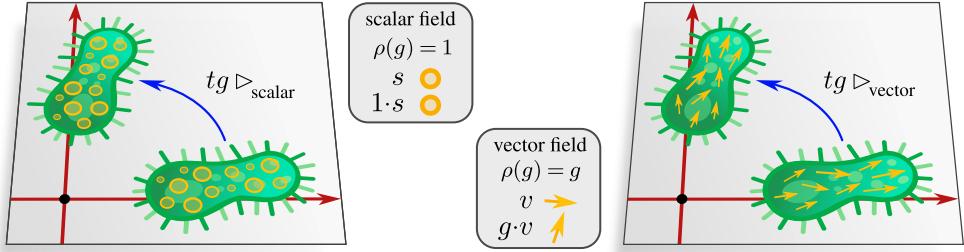


Figure 4.3: Transformation behavior of scalar and vector fields. A group element $tg \in \text{Aff}(G)$ acts on a scalar field $s : \mathbb{R}^d \rightarrow \mathbb{R}$ by moving its values spatially, that is, $[tg \triangleright_{\text{scalar}} s](x) := s((tg)^{-1}x)$. The individual vectors that make up a vector field $v : \mathbb{R}^d \rightarrow \mathbb{R}^d$ carry directional information. They are therefore additionally transformed by a multiplication with g , i.e., $[tg \triangleright_{\text{vector}} v](x) := gv((tg)^{-1}x)$. Feature fields generalize this transformation law, acting on feature vectors with some G -representation ρ ; see Eq. (4.3). Scalar, vector or tensor fields are recovered by choosing specific representations, which are the trivial representation $\rho(g) = 1$, the defining representation $\rho(g) = g$ and tensor product representations $\rho(g) = \otimes^s (g^{-1})^\top \otimes^r g$, respectively.

(Microbes adapted under the Apache license 2.0 by courtesy of Google.)

4.2 Euclidean feature fields as induced affine group representations

The laws of nature are formulated in terms of scalar, vector, or tensor-fields, which have a well defined transformation behavior under affine transformations (or even diffeomorphisms); see Fig. 4.3. Note that the transformation law $[(tg) \triangleright_{\text{vector}} v](x) = gv((tg)^{-1}x)$ of vector fields does not only involve a *spatial transformation* of the field on \mathbb{R}^d , but also a *G-action on each individual (feature) vector (fiber)* in \mathbb{R}^c . Feature vector fields generalize this transformation law by acting with some G -representation $\rho : G \rightarrow \text{GL}(c)$ (Def. B.5.1) on feature vectors. Specifically, a *feature vector field of geometric type* ρ is a map

$$F : \mathbb{R}^d \rightarrow \mathbb{R}^c \quad (4.2)$$

that is acted on by the ρ -induced $\text{Aff}(G)$ -action

$$[tg \triangleright_\rho F](x) = \rho(g) F((tg)^{-1}x). \quad (4.3)$$

↑ ↑
fiber action on \mathbb{R}^c spatial action on \mathbb{R}^d

More formally, we define Euclidean feature fields as follows:

Definition 4.2.1 (Euclidean feature fields). *The feature spaces of $\text{Aff}(G)$ -equivariant Euclidean steerable CNNs are vector spaces*

$$L^2(\mathbb{R}^d, \mathbb{R}^c) := \left\{ F : \mathbb{R}^d \rightarrow \mathbb{R}^c \mid \int_{\mathbb{R}^d} dx \|F(x)\|^2 \leq \infty \right\} \quad (4.4)$$

of square integrable c -channel feature fields in d spatial dimensions. They are associated to some G -representation (geometric type) $\rho : G \rightarrow \text{GL}(c)$ and transform according to the induced group action

$$\begin{aligned} \triangleright_\rho : \text{Aff}(G) \times L^2(\mathbb{R}^d, \mathbb{R}^c) &\rightarrow L^2(\mathbb{R}^d, \mathbb{R}^c), \\ (tg, F) &\mapsto tg \triangleright_\rho F := \rho(g) F(tg)^{-1}. \end{aligned} \quad (4.5)$$

As this action is linear, it corresponds to a group representation, known as induced representation^{1,2}

$$\text{Ind}_G^{\text{Aff}(G)}\rho : \text{Aff}(G) \rightarrow \text{GL}(L^2(\mathbb{R}^d, \mathbb{R}^c)) , \quad tg \mapsto tg \triangleright_\rho (\cdot) . \quad (4.6)$$

Euclidean feature fields are therefore just the elements of induced affine group representation spaces.

A trivial structure group $G = \{e\}$, and thus trivial inducing representation ρ , recovers Def. 3.1.1, i.e. Euclidean feature maps as regular translation group representations. A formal definition of induced representations on general homogeneous spaces (Def. B.3.11) is found in [35].

Different choices of representations ρ yield different *types* of feature fields. For instance, the *trivial representation*, defined by

$$\rho_{\text{triv}} : G \rightarrow \text{GL}(1), \quad g \mapsto 1 \quad (4.7)$$

describes fields of G -invariant *scalars*. Examples of scalar fields are grayscale images, temperature fields, pressure fields or probability distributions. *Tangent vector fields*, like optical flow or fluid velocity fields, correspond to the *defining representation*

$$\rho_{\text{vector}} : G \rightarrow \text{GL}(d), \quad g \mapsto g \quad (4.8)$$

of the matrix group. More general *tensor fields* of type (r, s) are described by *tensor product representations*

$$\rho_{\text{tensor}(r,s)} : G \rightarrow \text{GL}((\mathbb{R}^d)^{\otimes r} \otimes (\mathbb{R}^{d*})^{\otimes s}), \quad g \mapsto g^{\otimes r} \otimes (g^*)^{\otimes s}, \quad (4.9)$$

where $g^* = (g^{-1})^\top$ is the dual representation of the defining representation g . They model for instance diffusion tensor images, electromagnetic field tensors or stress tensors. Another common choice are *regular representations* (Def. B.5.18), which act by permuting feature vector entries. They are of great practical relevance in equivariant deep learning since they describe the transformation of the features of group convolutional networks [52]; see Section 4.5. Feature fields which transform under *irreducible representations* (irreps, Def. B.5.6) were investigated in [335, 323, 301, 161, 3, 322, 139]. A more detailed overview and an extensive benchmark of different field types or representations in deep learning is presented in Chapter 6.

Since finite dimensional unitary G -representations decompose generally into a *direct sum of irreducible representations* (Theorem B.5.16), one may view feature fields of such types as being a direct sum of the corresponding irrep fields. The original and decomposed representations and feature vectors are hereby related by a (linear) change of basis. Linear operations, like convolutions, may be derived in any choice of basis, implying that one may w.l.o.g. focus on irrep fields when studying them [322, 173, 40]. However, general nonlinear operations depend on the particular choice of basis, such that we will consider general G -representations, instead of only irreducible ones, in the following.

The coordinate free and coordinate independent description of feature fields on general manifolds is described in Sections 8.1, 11.3.3 and 11.4.4; see specifically the commutative diagram in Fig. 11.7. The induced representation action corresponds in this setting to the *pushforward of associated bundle sections* (pushforward of feature fields), discussed in Section 8.3 and Def. 13.1.2.

¹ $\text{Ind}_G^{\text{Aff}(G)}$ is a functor that turns G -representations into $\text{Aff}(G)$ -representations.

²There is generally a one-to-one correspondence between linear group actions and group representations; see the discussion around Eqs. (B.35) and (B.36) in Appendix B.5.1.

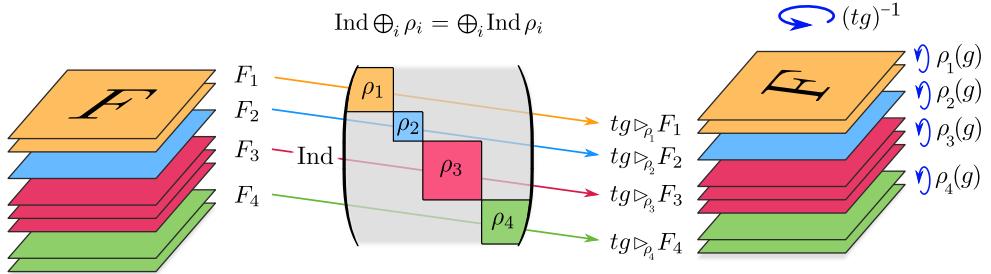


Figure 4.4: A full feature space of steerable CNNs comprises multiple individual feature fields $F_i : \mathbb{R}^d \rightarrow \mathbb{R}^{c_i}$ of potentially different types ρ_i and dimensionalities c_i . The composite field $F = \bigoplus_i F_i$ transforms according to the direct sum $\bigoplus_i \text{Ind}_G^{\text{Aff}(G)} \rho_i = \text{Ind}_G^{\text{Aff}(G)} \bigoplus_i \rho_i$, and can therefore be viewed as being of type $\bigoplus_i \rho_i$ (here $\rho_1 \oplus \rho_2 \oplus \rho_3 \oplus \rho_4$). The block structure of the direct sum representation guarantees hereby that the individual fields f_i transform independently from each other, that is, their channels do not mix under G -transformations. Compare this to Fig. 8.3, which emphasizes the passive transformation viewpoint, i.e. coordinate independence, of feature fields.

Stacked Euclidean feature fields: The feature maps of translation equivariant CNNs consist of multiple channels, which transform independently from each other under translations. The feature spaces of $\text{Aff}(G)$ -steerable CNNs consist in analogy of multiple independent feature fields $F_i : \mathbb{R}^d \rightarrow \mathbb{R}^{c_i}$, each coming with its own type ρ_i . Taken together, these individual fields make up a composite $(\sum_i c_i)$ -dimensional feature field, which is formally given by their *direct sum*³

$$F = \bigoplus_i F_i : \mathbb{R}^d \rightarrow \mathbb{R}^{\sum_i c_i} \quad (4.10)$$

and transforms thus according to the direct sum $\bigoplus_i \text{Ind}_G^{\text{Aff}(G)} \rho_i$ of the individual induced representations. The induction functor commutes with the direct sum [55],

$$\bigoplus_i \text{Ind}_G^{\text{Aff}(G)} \rho_i = \text{Ind}_G^{\text{Aff}(G)} \bigoplus_i \rho_i, \quad (4.11)$$

such that the composite feature field can be viewed as being of type

$$\rho = \bigoplus_i \rho_i. \quad (4.12)$$

A visual interpretation of the transformation law of stacked feature fields is given in Fig. 4.4.

As a practical example of a steerable feature space consisting of multiple fields consider an RGB image as depicted in Fig. 4.5. Similar to a grayscale image, the individual color channels encode intensity values which do not mix under G -transformations like rotations or reflections. The full RGB image is therefore to be identified with three scalar fields, each of which transforms independently under the trivial representation. Not all individual feature fields need to be of the same type ρ_i . For instance, in a weather forecasting application the input signal might consist of scalar

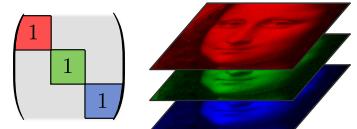


Figure 4.5: The three color channels of an RGB image are geometrically identified as scalar fields, the full feature space therefore transforms according to the direct sum $\rho(g) = (1) \oplus (1) \oplus (1)$.

³The direct sum \oplus of vectors $F_i(x)$ can be thought of as “stacking” these into a concatenated vector. Consistently with this, the direct sum of representations ρ_i can be thought of as building a block diagonal matrix containing the ρ_i as blocks; see Def. B.5.3 and Figs. 4.4 and 4.5.

fields encoding features like temperature or pressure and vector fields like wind flow velocities. A description as ρ_i fields of corresponding types ensures the geometrically correct processing of such data. While the field types ρ_i of a network's input and output are typically given by the learning task, the field types used in hidden layers are chosen by the user as a hyperparameter similar to the choice of channels for a conventional CNN.

Since stacked feature fields are defined as a direct sum, they transform independently from each other. This allows to investigate the equivariant network layers in the following section for each constituent field individually, that is, one can without loss of generality consider single (non-stacked) input and output feature fields. Layers that map between full feature spaces of stacked fields are then combined from these operations between individual fields.

4.3 Affine equivariant layers and steerable convolutions

$\text{Aff}(G)$ -equivariant (steerable) network layers between feature fields of types $\rho_{\text{in}} : G \rightarrow \text{GL}(c_{\text{in}})$ and $\rho_{\text{out}} : G \rightarrow \text{GL}(c_{\text{out}})$ are functions $L : L^2(\mathbb{R}^d, \mathbb{R}^{c_{\text{in}}}) \rightarrow L^2(\mathbb{R}^d, \mathbb{R}^{c_{\text{out}}})$ such that the following diagram commutes for arbitrary affine transformations $tg \in \text{Aff}(G)$:

$$\begin{array}{ccc} L^2(\mathbb{R}^d, \mathbb{R}^{c_{\text{in}}}) & \xrightarrow{L} & L^2(\mathbb{R}^d, \mathbb{R}^{c_{\text{out}}}) \\ tg \triangleright \rho_{\text{in}} \downarrow & & \downarrow tg \triangleright \rho_{\text{out}} \\ L^2(\mathbb{R}^d, \mathbb{R}^{c_{\text{in}}}) & \xrightarrow{L} & L^2(\mathbb{R}^d, \mathbb{R}^{c_{\text{out}}}) \end{array} \quad (4.13)$$

As done for translation equivariant layers in the previous Chapter 3, we derive here the most commonly used affine equivariant layers, i.e. steerable convolutions, biases, nonlinearities and pooling operations. Since translations $(\mathbb{R}^d, +) \leq \text{Aff}(G)$ form a subgroup of the affine groups, we will recover the constraints that we encountered for conventional CNNs, namely a requirement for *spatial weight sharing*. However, the spatially shared network connectivity (e.g. kernel or bias) is additionally constrained by the structure group $G < \text{Aff}(G)$, i.e. an additional *G-steerability constraint*. Together, these constraints can be viewed as a *weight sharing over $\text{Aff}(G)$ -transformations*.

4.3.1 Steerable convolutions as affine equivariant linear maps

Linear $\text{Aff}(G)$ -equivariant maps between feature fields (induced affine group representations) are necessarily convolutions with G -steerable (G -equivariant) kernels. As before, the convolution ensures translational equivariance. Fig. 4.6 visualizes the role of the kernel's G -steerability: it summarizes the ρ_{in} -features in its field of view such into a ρ_{out} -feature vector that any G -transformation of the field of view results in a corresponding G -transformation of the output feature vector. Applied in a convolutional manner, this ensures exactly that the resulting field of output feature vectors transforms according to their induced representation, Def. 4.2.1.

The statement that linear $\text{Aff}(G)$ -equivariant maps are convolutions with G -steerable kernels has been proven in different settings. Jenner and Weiler [137] considered general linear functionals, corresponding to steerable kernels in the distributional sense. This setting includes

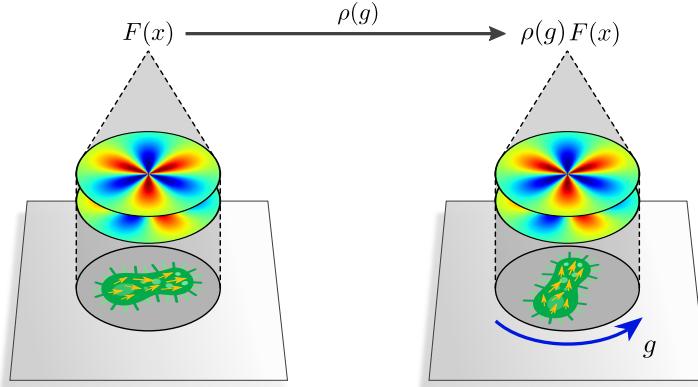


Figure 4.6: A convolution kernel at $x \in \mathbb{R}^d$ summarizes the field of input feature vectors around x into a single output feature vector $F(x) \in \mathbb{R}^{c_{\text{out}}}$. Since the kernel is applied in a convolutional manner, i.e. at each point $x \in \mathbb{R}^d$, it produces a field of output feature vectors. $\text{Aff}(G)$ -equivariant convolutions rely according to Theorem 4.3.1 on G -steerable kernels, satisfying the G -equivariance constraint in Eq. (4.18) and diagram (4.19). Intuitively, G -steerable kernels guarantee that a transformation of the input features in their field of view by $g \in G$ (or more specifically by $\text{Ind}_G^{\text{Aff}(G)} \rho_{\text{in}}(g)$) results in a transformation of the output feature vector by $\rho_{\text{out}}(g)$. Applied convolutionally, the resulting field of output feature vectors is guaranteed to transform according to the induced representation $\text{Ind}_G^{\text{Aff}(G)} \rho_{\text{out}}$. More details on G -steerable kernels are found in Chapter 5 below.

convolutions with classical kernel functions, but also with other Schwartz distributions, like e.g. partial differential operators. For simplicity, we will here follow the original derivation of the G -steerability constraint by Weiler et al. [323], considering specifically convolutions with classical kernel functions, derived from an integral transform ansatz. In contrast to their derivation, we allow here for general structure groups $G \leq \text{GL}(d)$ instead of *compact* structure groups $G \leq \text{O}(d)$ only. This will result in an additional volume scaling factor $|\det g|$ occurring in the G -steerability constraint on convolution kernels.

As simple ansatz for *unconstrained linear maps between feature fields*, we consider the same *integral transforms* as in Section 3.2, i.e.

$$\mathcal{I}_\kappa : L^2(\mathbb{R}^d, \mathbb{R}^{c_{\text{in}}}) \rightarrow L^2(\mathbb{R}^d, \mathbb{R}^{c_{\text{out}}}), \quad \text{defined by} \quad \mathcal{I}_\kappa[F](x) := \int_{\mathbb{R}^d} dy \kappa(x, y) F(y). \quad (4.14)$$

They are parameterized by square integrable *matrix-valued two-argument kernel* functions

$$\kappa : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \quad (4.15)$$

which specify how input features from any point $y \in \mathbb{R}^d$ are aggregated into an output feature vector at $x \in \mathbb{R}^d$; see Fig. 3.3.

The following theorem proves that the demand for affine group equivariance requires this integral transform to be a convolution with a G -steerable kernel:

Theorem 4.3.1 (Steerable convolutions). *The integral transform \mathbb{I}_κ in Eq. (4.14) is $\text{Aff}(G)$ equivariant w.r.t. induced affine group representations $\text{Ind}_G^{\text{Aff}(G)} \rho_{\text{in}}$ and $\text{Ind}_G^{\text{Aff}(G)} \rho_{\text{out}}$, that is, satisfies*

$$\mathbb{I}_\kappa [tg \triangleright_{\rho_{\text{in}}} F] = tg \triangleright_{\rho_{\text{out}}} \mathbb{I}_\kappa [F] \quad \forall tg \in \text{Aff}(G), \quad (4.16)$$

if and only if

1. it is a convolution integral

$$\mathbb{I}_\kappa [F](x) = \int_{\mathbb{R}^d} dy K(x-y) F(y) =: [K * F](x) \quad (4.17)$$

with a one-argument kernel $K : \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ satisfying the translation relativity condition $\kappa(x, y) =: K(x-y)$, and

2. the kernel K is G -steerable, that is,

$$K(gx) = \frac{1}{|\det g|} \rho_{\text{out}}(g) K(x) \rho_{\text{in}}(g)^{-1} \quad \forall x \in \mathbb{R}^d, g \in G. \quad (4.18)$$

This G -steerability constraint is captured by the commutativity of the diagram

$$\begin{array}{ccc} \mathbb{R}^d & \xrightarrow{K} & \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \\ g \downarrow & & \downarrow \frac{1}{|\det g|} \rho_{\text{out}}(g) [\cdot] \rho_{\text{in}}(g)^{-1} \\ \mathbb{R}^d & \xrightarrow{K} & \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \end{array} \quad (4.19)$$

for arbitrary $g \in G$. Chapter 5 below discusses G -steerable kernels in depth.

Proof: We write out the left-hand side of Eq. (4.16), which yields

$$\begin{aligned} \mathbb{I}_\kappa [tg \triangleright_{\rho_{\text{in}}} F](x) &= \int_{\mathbb{R}^d} dy \kappa(x, y) [tg \triangleright_{\rho_{\text{in}}} F](y) \\ &= \int_{\mathbb{R}^d} dy \kappa(x, y) \rho_{\text{in}}(g) F((tg)^{-1}y) \\ &= \int_{\mathbb{R}^d} d\tilde{y} |\det g| \kappa(x, (tg)\tilde{y}) \rho_{\text{in}}(g) F(\tilde{y}) \end{aligned} \quad (4.20)$$

after a substitution of $\tilde{y} := (tg)^{-1}y = g^{-1}(y-t)$ with $|\det(\frac{d\tilde{y}}{dy})| = |\det(g^{-1})| = |\det g|^{-1}$. The right-hand side is given by

$$\begin{aligned} [tg \triangleright_{\rho_{\text{out}}} \mathbb{I}_\kappa [F]](x) &= tg \triangleright_{\rho_{\text{out}}} \int_{\mathbb{R}^d} dy \kappa(x, y) F(y) \\ &= \int_{\mathbb{R}^d} dy \rho_{\text{out}}(g) \kappa((tg)^{-1}x, y) F(y). \end{aligned} \quad (4.21)$$

These expressions agree for any $tg \in \text{Aff}(G)$ and any feature map $F \in L^2(\mathbb{R}^d, \mathbb{R}^{c_{\text{in}}})$ if and only if

$$\kappa((tg)x, (tg)y) = \frac{1}{|\det g|} \rho_{\text{out}}(g) \kappa(x, y) \rho_{\text{in}}(g)^{-1} \quad (4.22)$$

holds for any $x, y \in \mathbb{R}^d$ and any $tg \in \text{Aff}(g)$.

This constraint implies:

1. As already seen in Theorem 3.2.1 on translation equivariant convolutions, the two argument kernel may only depend on the *relative translation* between its arguments, i.e. $\kappa(x + t, y + t) = \kappa(x, y)$ for any $t \in (\mathbb{R}^d, +) \leq \text{Aff}(G)$. This allows to express the two argument kernel in terms of a one (relative) argument kernel $K : \mathbb{R}^d \rightarrow \mathbb{R}^{C_{\text{out}} \times C_{\text{in}}}$, given by $K(\Delta x) := \kappa(\Delta x, 0)$. Equivalently, we have $\kappa(x, y) = K(x - y)$, implying that the integral transform is a convolution with kernel K .
2. Expressing the general constraint in Eq. (4.22) in terms of the one-argument kernel, we obtain

$$\begin{aligned}\kappa((tg)x, (tg)y) &= \kappa(gx + t, gy + t) \\ &= K(g(x - y)) \\ &= \frac{1}{|\det g|} \rho_{\text{out}}(g) K(x - y) \rho_{\text{in}}(g)^{-1},\end{aligned}\tag{4.23}$$

for any $x, y \in \mathbb{R}^d$ and $g \in G$, which, after inserting $\Delta x := x - y$, is the claimed G -steerability constraint. \square

After being published for Euclidean spaces in [323], this theorem was in [56] generalized to H -equivariant convolutions on *homogeneous spaces* H/G . The symmetry group H was hereby assumed to be locally compact and unimodular, such that no determinant factor occurred in the G -steerability constraint. Appendix F explains such convolutions on homogeneous spaces in more detail. Jenner and Weiler [137] generalized our result on Euclidean spaces furthermore from our integral transform ansatz to *arbitrary linear maps*, and proved that these can generally be expressed as convolutions with G -steerable *Schwartz distributions*. This framework covers in particular *equivariant partial differential operators*, including for instance *gradients*, *divergence*, *curl* and the *Laplace operator*.

The active $\text{Aff}(G)$ -equivariance of steerable convolutions is in Chapter 15 shown to be related to their (passive) independence from choices of coordinate charts or pixel grids with transition maps in $\text{Aff}(G)$; see Fig. 15.2. Note that the independence of the neural connectivity from choices of $\text{Aff}(G)$ -related charts is equivalent to the invariance of physical laws under choices of Poincaré-related inertial frames in the special theory of relativity. Steerable CNNs are therefore best thought of as performing $\text{Aff}(G)$ -*relative measurements*.

It is intuitively clear that compactly supported G -steerable kernels are equivariant w.r.t. independent “*local G -transformations*” of their field of view at different locations; see Fig. 4.7a. This property is not adequately captured by our derivation of $\text{Aff}(G)$ -steerable convolutions, which focused on *global* transformations of the image as a whole. We formalize this local G -equivariance in Parts II and III on coordinate independent CNNs, which generalize steerable CNNs to Riemannian manifolds and focus on passive gauge transformations, i.e. changes between local reference frames (Fig. 4.7b).

Many equivariant network architectures in the literature are described as *group convolutions* instead of steerable convolutions [52, 162, 324, 10]. While their theoretical formulation may look quite different, group convolutions are actually a special case of steerable convolutions, operating specifically on feature fields whose type is the *regular G -representation*. Steerable CNNs are more general in that they allow for arbitrary field types, which allows for instance to model vector or tensor fields. This equivalence of group convolutions and regular steerable convolutions is proven in Section 4.5.

Chapter 5 below expands on the parametrization and implementation of G -steerable kernels.

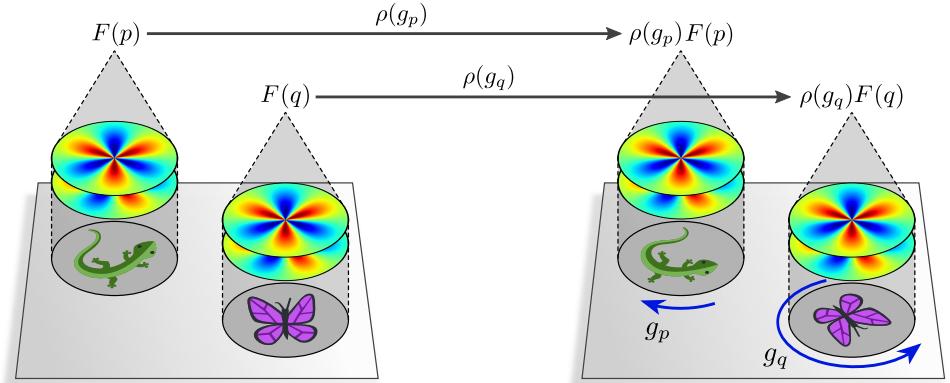
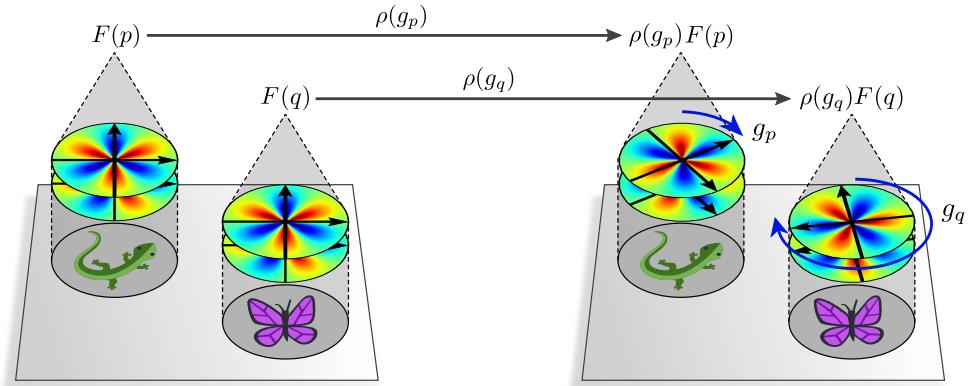
(a) *Active local gauge equivariance of steerable CNNs*(b) *Passive local coordinate independence of steerable CNNs*

Figure 4.7: *Top:* Steerable convolutions are not only globally $\text{Aff}(G)$ -equivariant, but more generally equivariant under independent local G -transformations g_p and g_q of the kernels' fields of view at different locations p and q in \mathbb{R}^d . This “local gauge equivariance” of steerable convolutions is formalized in Parts II and III. *Bottom:* As argued in Fig. 1.10, kernels only perceive their relative alignment towards features. The G -transformation of their response vectors when transforming their field of view is therefore equivalent to that when their own alignment is transformed. Parts II and III describe kernel alignments as choices of local reference frames and interpret steerable CNNs therefore as being coordinate independent.

(Lizards and butterflies adapted under the Creative Commons Attribution 4.0 International license by courtesy of Twitter.)

4.3.2 Steerable bias summation

In analogy to the case of translation equivariant bias summation, we investigate the summation

$$B_{\mathfrak{b}} : L^2(\mathbb{R}^d, \mathbb{R}^c) \rightarrow L^2(\mathbb{R}^d, \mathbb{R}^c), \quad F \mapsto F + \mathfrak{b} \quad (4.24)$$

of a square integrable field of bias vectors

$$\mathfrak{b} \in L^2(\mathbb{R}^d, \mathbb{R}^c), \quad (4.25)$$

however, now subject to an affine group equivariance constraint. Instead of only being required to be *spatially constant* (translation invariant), the bias field now has to be invariant under $\text{Aff}(G)$ -transformations – this implies in particular that *biases may only be summed to the trivial subrepresentations of ρ_{in}* . Note that we have again $c_{\text{in}} = c_{\text{out}} =: c$ and $\triangleright_{\text{in}} = \triangleright_{\text{out}} =: \triangleright$ since bias summation necessarily preserves the field type.

Theorem 4.3.2 (Affine equivariant bias summation). *The bias summation in Eq. (4.24) is $\text{Aff}(G)$ -equivariant w.r.t. the actions of the induced affine group representations $\text{Ind}_G^{\text{Aff}(G)}\rho_{\text{in}}$ and $\text{Ind}_G^{\text{Aff}(G)}\rho_{\text{out}}$, i.e. satisfies*

$$B_{\mathfrak{b}}[tg \triangleright_{\rho} F] = tg \triangleright_{\rho} B_{\mathfrak{b}}[F], \quad (4.26)$$

if and only if the bias field is $\text{Aff}(G)$ -invariant:

$$tg \triangleright_{\rho} \mathfrak{b} = \mathfrak{b} \quad \forall tg \in \text{Aff}(G) \quad (4.27)$$

This requires

1. *a spatially constant bias field, i.e. $\mathfrak{b}(x) = b$ for any $x \in \mathbb{R}^d$ and some shared bias vector $b \in \mathbb{R}^c$, and*
2. *this shared bias needs to be G -invariant, that is, $b = \rho(g)b$ for any $g \in G$.*

Proof: The left- and right-hand sides of Eq. (4.26) at $x \in \mathbb{R}^d$ become

$$\begin{aligned} B_{\mathfrak{b}}[tg \triangleright_{\rho} F](x) &= [tg \triangleright_{\rho} F](x) + \mathfrak{b}(x) \\ &= \rho(g)F((tg)^{-1}x) + \mathfrak{b}(x), \end{aligned} \quad (4.28)$$

and

$$\begin{aligned} [tg \triangleright_{\rho} B_{\mathfrak{b}}[F]](x) &= [tg \triangleright_{\rho} (F + \mathfrak{b})](x) \\ &= \rho(g)F((tg)^{-1}x) + \rho(g)\mathfrak{b}((tg)^{-1}x), \end{aligned} \quad (4.29)$$

respectively, which yields Eq. (4.27). Specifically, we have $\mathfrak{b}(x) = \mathfrak{b}(x - t)$ for any translation $t \in (\mathbb{R}^d, +) \leq \text{Aff}(G)$, implying the spatially constant bias field $\mathfrak{b}(x) = b$. Reinserting this into the constraint for general $tg \in \text{Aff}(G)$, we get the G -invariance $b = \rho(g)b$ for any $g \in G$. \square

To get an insight into the implications of the G -invariance constraint $b = \rho(g)b \quad \forall g \in G$, assume ρ to be reducible into a direct sum of irreps, which is by Theorems B.5.14 and B.5.16 w.l.o.g. the case for compact groups, including any $G \leq \text{O}(d)$. Let $Q \in \text{GL}(c)$ be a change of basis that decomposes ρ into irreps ρ_j , i.e. satisfies $Q\rho(g)Q^{-1} = \bigoplus_j \rho_j(g)$ for any $g \in G$, and let $Qb =: \tilde{b} = \bigoplus_j \tilde{b}_j$ be the corresponding decomposition of the bias vector into irreducible subspaces labeled by j . The constraint in this basis becomes

$\bigoplus_j \tilde{b}_j = [\bigoplus_j \rho_j(g)] [\bigoplus_j \tilde{b}_j]$, i.e. it splits into independent constraints $\tilde{b}_j = \rho_j(g)\tilde{b}_j$ on the subspaces. The latter is *always satisfied for the trivial irreducible subrepresentations of ρ* , satisfying $\rho_0(g) = 1$. For all non-trivial irreducible subrepresentations of ρ , the constraint is only satisfied for zero vectors, i.e. $\tilde{b}_{j \neq 0} = 0$. It follows that the vector space

$$\mathcal{B}_\rho^G := \{b \in \mathbb{R}^c \mid b = \rho(g)b \quad \forall g \in G\} \quad (4.30)$$

of G -invariant bias vectors coincides with the (subspaces of) trivial subrepresentations of ρ . Its dimensionality – and therefore the number of learnable parameters – equals the multiplicity of trivial subrepresentations contained in ρ .

Two simple examples of feature fields to which one might want to sum a spatially shared and G -invariant bias vector are scalar fields and tangent vector fields. By definition, scalars are G -invariant, that is, they transform according to the trivial representation $\rho(g) = 1 \quad \forall g \in G$. One can therefore sum any (scalar) bias $b \in \mathbb{R}$ to them. In contrast, tangent vectors transform according to the non-trivial, irreducible defining representation $\rho(g) = g$ of G . This representation does not contain any trivial subrepresentation, such that it is impossible to sum a shared bias vector to tangent vector fields while maintaining $\text{Aff}(G)$ -equivariance. As a third example, consider regular representations of compact groups, which describe for instance the feature fields of group convolutional networks. By the Peter-Weyl theorem B.5.22, it is known that regular representations contain exactly one trivial subrepresentation. The bias to be summed to regular feature fields is therefore seen to be described by a single parameter. Section 10.3.1 explicitly derives some more examples for representations of the reflection group.

4.3.3 Steerable local nonlinearities

As probably expected, locally acting $\text{Aff}(G)$ -equivariant nonlinearities are not only required to be spatially independent, but also equivariant w.r.t. field types ρ_{in} and ρ_{out} , between which they map. To show this, are again starting with a spatially dependent ansatz

$$\mathfrak{S}_\sigma : L^2(\mathbb{R}^d, \mathbb{R}^{c_{\text{in}}}) \rightarrow L^2(\mathbb{R}^d, \mathbb{R}^{c_{\text{out}}}), \quad (4.31)$$

for the local nonlinearities, given by

$$\sigma : \mathbb{R}^d \times \mathbb{R}^{c_{\text{in}}} \rightarrow \mathbb{R}^{c_{\text{out}}}, \quad (x, f) \rightarrow \sigma_x(f), \quad (4.32)$$

and

$$\mathfrak{S}_\sigma[F](x) := \sigma_x(F(x)), \quad (4.33)$$

and then demand affine equivariance.

Theorem 4.3.3 (Affine equivariant local nonlinearities). *The spatially dependent localized nonlinearity operation in Eq. (4.31) is $\text{Aff}(G)$ -equivariant w.r.t. the actions of the induced affine group representations $\text{Ind}_G^{\text{Aff}(G)}\rho_{\text{in}}$ and $\text{Ind}_G^{\text{Aff}(G)}\rho_{\text{out}}$, i.e. satisfies*

$$\mathfrak{S}_\sigma[tg \triangleright_{\rho_{\text{in}}} F] = tg \triangleright_{\rho_{\text{out}}} \mathfrak{S}_\sigma[F], \quad (4.34)$$

if and only if:

1. *the underlying field of localized nonlinearities is position-independent, that is, $\sigma_x = s$ for some shared nonlinearity $s : \mathbb{R}^{c_{\text{in}}} \rightarrow \mathbb{R}^{c_{\text{out}}}$ and any $x \in \mathbb{R}^d$, and*

2. this shared nonlinearity is G -equivariant, i.e. $\rho_{\text{out}}(g) \circ = \circ \rho_{\text{in}}(g) \quad \forall g \in G$.

Proof: The constraint in Eq. (4.34) demands that

$$\begin{aligned} \mathfrak{S}_\sigma [tg \triangleright_{\rho_{\text{in}}} F](x) &= \sigma_x([tg \triangleright_{\rho_{\text{in}}} F](x)) \\ &= \sigma_x(\rho_{\text{in}}(g) F((tg)^{-1}x)), \end{aligned} \quad (4.35)$$

and

$$\begin{aligned} [tg \triangleright_{\rho_{\text{out}}} \mathfrak{S}_\sigma[F]](x) &= \rho_{\text{out}}(g) \mathfrak{S}_\sigma[F]((tg)^{-1}x) \\ &= \rho_{\text{out}}(g) \sigma_{(tg)^{-1}x}(F((tg)^{-1}x)) \end{aligned} \quad (4.36)$$

agree for any $x \in \mathbb{R}^d$ and any $tg \in \text{Aff}(g)$, and therefore:

1. focusing specifically on translations $t \in (\mathbb{R}^d, +) \leq \text{Aff}(G)$ first, we obtain $\sigma_x(F(x-t)) = \sigma_{x-t}(F(x-t))$, implying the claimed position independence of the nonlinearity field
2. reinserting this partial result into the constraint yields, for any $tg \in \text{Aff}(G)$ and arbitrary feature fields, $\circ(\rho_{\text{in}}(g) F((tg)^{-1}x)) = \rho_{\text{out}}(g) \circ(F((tg)^{-1}x))$, implying the G -equivariance of \circ . \square

Due to the nonlinearity of the steerability constraint $\rho_{\text{out}}(g) \circ = \circ \rho_{\text{in}}(g) \quad \forall g \in G$, it cannot be solved in a systematic fashion by decomposing it into irreducible subspaces, as done for kernels in Section 5.3.2 and for biases in Section 4.3.2. One is instead forced to design steerable local nonlinearities on a case-by-case basis. The following list summarizes some common design choices for specific families of field types that were proposed in the literature:

Trivial reps: For trivial representations ρ_{in} and ρ_{out} , the constraint $\rho_{\text{out}}(g) \circ = \circ \rho_{\text{in}}(g)$ becomes itself trivial. As a consequence, any nonlinearity is admissible for scalar fields.

Unitary reps: A general class of representations are unitary representations (Def. B.5.13), which preserve the norm of their representation space, that is, they satisfy $|\rho_{\text{unitary}}(g)F(x)| = |F(x)| \quad \forall g \in G$. As proven in [322], nonlinearities which solely act on the *norm of feature vectors* but preserve their orientation are equivariant w.r.t. unitary representation actions. They can in general be decomposed in $\sigma_{\text{norm}} : \mathbb{R}^c \rightarrow \mathbb{R}^c$, $F(x) \mapsto \eta(|F(x)|) \frac{F(x)}{|F(x)|}$ for some nonlinear function $\eta : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ acting on the norm of feature vectors. *Norm-ReLUs*, defined by $\eta(|F(x)|) = \text{ReLU}(|F(x)| - b)$ where $b \in \mathbb{R}_{\geq 0}$ is a learned bias, were used in [335, 323]. In [253], the authors consider *squashing nonlinearities* with $\eta(|F(x)|) = \frac{|F(x)|^2}{|F(x)|^2 + 1}$. *Gated nonlinearities* were proposed in [323] as conditional version of norm nonlinearities. They act by scaling the norm of a feature field by learned sigmoid gates $\frac{1}{1+e^{-s(x)}}$, parameterized by a scalar feature field s . Note that any representation of compact groups, including any $G \leq O(d)$, can by Theorem B.5.14 be considered as being unitary. The nonlinearities described here are in practice often used for irrep feature fields of such groups.

Permutation reps: Permutation representations act by permuting feature vector channels. This action commutes generally with any *element-wise nonlinearity*, for instance ReLU activations that are applied to each vector entry individually. Practically relevant examples are *regular representations* (Def. B.5.18), corresponding to group convolutions (Theorem 4.5.1), *quotient representations* (Def. B.5.20), and the *trivial representation*.

Tensor product reps: Feature vectors $F_1(x) \in \mathbb{R}^{c_1}$ and $F_2(x) \in \mathbb{R}^{c_2}$ of arbitrary types ρ_1 and ρ_2 can be combined to a tensor product feature $[F_1 \otimes F_2](x) \in \mathbb{R}^{c_1 c_2}$, whose type is the tensor product representation $\rho_1 \otimes \rho_2$; Def. B.5.4. The tensor product is a nonlinear (but still bilinear) operation, such that some authors choose not to apply further nonlinearities [161, 323, 163, 3].

All of these examples satisfy the equivariance constraint $\rho_{\text{out}}(g) s = s \rho_{\text{in}}(g) \forall g \in G$. Which particular nonlinearity works well in practice is, however, an empirical question. Section 6.5 presents a benchmarking of different field types and correspondingly steerable nonlinearities. The results are summarized in Table 6.6 – regular (permutation) representations with element-wise nonlinearities perform best overall, while gated nonlinearities perform best for unitary representations that do not act via permutations, like e.g. $O(d)$ -irreps. Franzen and Wand [94] provide a harmonic distortion analysis of nonlinearities that are applied to the channels of regular feature fields, or subrepresentations of it.

4.3.4 Steerable local pooling operations

As for translation equivariant CNNs, we discuss local and global pooling operations. The requirement for affine group equivariance imposes additional G -symmetry constraints on these operations, or may allow their application to specific field types ρ only.

Channel-wise local max pooling: Since *local max pooling* operates channel-wise, it commutes only with *permutation representations*. It requires furthermore an affine invariant choice of pooling regions, for instance balls of a certain radius for $G = SO(d)$.

Theorem 4.3.4 (Affine equivariance of local max pooling for permutation reps).

Let the field type $\rho : G \rightarrow GL(c)$ be a permutation representation, i.e. act by permuting field channels. The local max pooling operation in Eq. (3.30) is then affine group equivariant, that is,

$$\text{local_max_pool}[tg \triangleright_\rho F](x) = (tg \triangleright_\rho \text{local_max_pool}[F])(x) \quad (4.37)$$

for any $x \in \mathbb{R}^d$ and any $tg \in \text{Aff}(G)$, if and only if the pooling windows are affine group invariant (spatially shared and locally G -invariant), i.e. satisfy

$$(tg)^{-1}\mathcal{R}_x = \mathcal{R}_{(tg)^{-1}x} \quad \forall x \in \mathbb{R}^d, \quad tg \in \text{Aff}(G), \quad (4.38)$$

Proof: Using that permutation representations commute with the channel-wise application of local max pooling, and substituting the pooling region \mathcal{R}_x for $(tg)^{-1}\mathcal{R}_x$, the left hand side becomes

$$\begin{aligned} \text{local_max_pool}[tg \triangleright_\rho F](x) &= \max_{y \in \mathcal{R}_x} \rho(g)F((tg)^{-1}y) \\ &= \rho(g) \max_{y \in (tg)^{-1}\mathcal{R}_x} F(y). \end{aligned} \quad (4.39)$$

The statement follows by setting this equal with the right hand side, which is given by

$$\begin{aligned} (tg \triangleright_\rho \text{local_max_pool}[F])(x) &= \rho(g) \text{local_max_pool}[F]((tg)^{-1}x) \\ &= \rho(g) \max_{y \in (tg)^{-1}\mathcal{R}_x} F(y). \end{aligned} \quad (4.40) \quad \square$$

A typical example of permutation representations are regular representations.

Local norm max pooling: An alternative approach to max pool feature fields is *local norm max pooling*, which selects feature vectors based on their maximum norm within some pooling region. It applies to any field whose type ρ is a *unitary representation* since their actions preserve the feature vector norm. It is defined as a map

$$\text{local_norm_max_pool} : L^2(\mathbb{R}^d, \mathbb{R}^c) \rightarrow L^2(\mathbb{R}^d, \mathbb{R}^c) \quad (4.41)$$

that is for a pooling region $\mathcal{R}_x \subset \mathbb{R}^d$ around $x \in \mathbb{R}^d$ given by

$$\text{local_norm_max_pool}[F](x) = F\left(\arg \max_{y \in \mathcal{R}_x} |F(y)|\right). \quad (4.42)$$

If a stacked feature field $F = \bigoplus_i F_i$ is given by a direct sum of feature fields F_i of unitary type, `local_norm_max_pool` is usually applied to each summand individually.

Theorem 4.3.5 (Affine equivariance of local norm max pooling for unitary reps).

Let F be a feature field of unitary type $\rho : G \rightarrow \text{GL}(c)$, preserving the norm $|\rho(g)F(x)| = |F(x)|$ of feature vectors under G -actions. If the pooling regions are affine invariant, i.e.

$$(tg)^{-1}\mathcal{R}_x = \mathcal{R}_{(tg)^{-1}x} \quad \forall x \in \mathbb{R}^d, \quad tg \in \text{Aff}(G), \quad (4.43)$$

the local norm max pooling operation in Eq. (4.42) is then affine group equivariant, i.e.

$$\text{local_norm_max_pool}[tg \triangleright_\rho F](x) = (tg \triangleright_\rho \text{local_norm_max_pool}[F])(x) \quad (4.44)$$

for any $x \in \mathbb{R}^d$ and any $tg \in \text{Aff}(G)$.

Proof: Using the unitarity $|\rho(g)F(y)| = |F(y)|$ of ρ in the third step and a substitution of the pooling region, which requires a corresponding correction of the arg max result, in the fourth step, the left hand side of the equivariance constraint in Eq. (4.44) becomes:

$$\begin{aligned} & \text{local_norm_max_pool}[tg \triangleright_\rho F](x) \\ &= [tg \triangleright_\rho F]\left(\arg \max_{y \in \mathcal{R}_x} |[tg \triangleright_\rho F](y)|\right) \\ &= \rho(g)F\left((tg)^{-1} \arg \max_{y \in \mathcal{R}_x} |\rho(g)F((tg)^{-1}y)|\right) \\ &= \rho(g)F\left((tg)^{-1} \arg \max_{y \in \mathcal{R}_x} |F((tg)^{-1}y)|\right) \\ &= \rho(g)F\left((tg)^{-1}(tg) \arg \max_{y \in (tg)^{-1}\mathcal{R}_x} |F|(y)\right) \\ &= \rho(g)F\left(\arg \max_{y \in (tg)^{-1}\mathcal{R}_x} |F|(y)\right) \end{aligned} \quad (4.45)$$

The right hand side is given by

$$[tg \triangleright_\rho \text{local_norm_max_pool } F](x) = \rho(g)F\left(\arg \max_{y \in \mathcal{R}_{(tg)^{-1}x}} |F|(y)\right). \quad (4.46)$$

These expressions agree for any $tg \in \text{Aff}(G)$ and any unitary feature field $F \in L^2(\mathbb{R}^d, \mathbb{R}^c)$ if and only if the pooling regions are affine invariant, that is, satisfy $(tg)^{-1}\mathcal{R}_x = \mathcal{R}_{(tg)^{-1}x}$. \square

Local average pooling: We defined *local average pooling* in Eq. (3.35) as a channel-wise convolution with a scalar weighting kernel $\kappa : \mathbb{R}^d \rightarrow \mathbb{R}$. If this operation is to be affine equivariant, the weighting kernel is additionally required to be *G-steerable*. Local average pooling may be applied to fields of any type ρ since the convolution with a scalar kernel commutes with the G -action.

Theorem 4.3.6 (Affine equivariance of local average pooling). *Let $\kappa : \mathbb{R}^d \rightarrow \mathbb{R}$ be a scalar weighting kernel that is G -steerable in the sense that*

$$\kappa(gx) = \frac{1}{|\det g|} \kappa(x) \quad \forall x \in \mathbb{R}^d, g \in G. \quad (4.47)$$

The local average pooling operation in Eq. (4.42) is then affine group equivariant, i.e.

$$\text{local_avg_pool}[tg \triangleright_\rho F](x) = (tg \triangleright_\rho \text{local_avg_pool}[F])(x) \quad (4.48)$$

for any $x \in \mathbb{R}^d$, any $tg \in \text{Aff}(G)$ and any $F \in L^2(\mathbb{R}^d, \mathbb{R}^c)$.

Proof: The equality of both sides of the equivariance constraint is shown by

$$\begin{aligned} & \text{local_avg_pool}[tg \triangleright_\rho F](x) \\ &= \int_{\mathbb{R}^d} dy \kappa(x - y) [tg \triangleright_\rho F](y) \\ &= \int_{\mathbb{R}^d} dy \kappa(x - y) \rho(g) F((tg)^{-1}y) \\ &= \rho(g) \int_{\mathbb{R}^d} d\tilde{y} |\det g| \kappa(x - (tg)\tilde{y}) F(\tilde{y}) \\ &= \rho(g) \int_{\mathbb{R}^d} d\tilde{y} \kappa((tg)^{-1}x - \tilde{y}) F(\tilde{y}) \\ &= \rho(g) \text{local_avg_pool}[F]((tg)^{-1}x) \\ &= (tg \triangleright_\rho \text{local_avg_pool}[F])(x), \end{aligned} \quad (4.49)$$

where we substituted $\tilde{y} := (tg)^{-1}y$ and made use of the weighting kernel's G -steerability in the fourth step:

$$\kappa(x - (tg)\tilde{y}) = \kappa(x - g\tilde{y} + t) = \kappa(g(g^{-1}x - t) - \tilde{y}) \quad (4.50)$$

$$= \kappa(g((tg)^{-1}x - \tilde{y})) = \frac{1}{|\det g|} \kappa((tg)^{-1}x - \tilde{y}) \quad \square$$

Note that the convolution with a G -steerable weighting kernel is analogous to the use of affine invariant pooling regions.

As for regular translation equivariant CNNs, local pooling operations are in discretized affine equivariant steerable CNN implementations often followed by a subsampling step. For feature fields of regular representation type, this may be done using the method of Xu et al. [336], which subsamples the field on an equivariant choice of coset (Def. B.2.2).

4.3.5 Steerable global pooling operations

Global affine equivariant pooling operations result in a single feature vector that is *position-independent* but still *G-steerable*.

Channel-wise global max pooling: Channel-wise *global max pooling* was defined in Eq. (3.39). Applied to permutation representations, it is affine equivariant:

Theorem 4.3.7 (Affine equivariance of global max pooling for permutation reps).

If the field type ρ is a permutation representation, (channel wise) global max pooling is affine equivariant in the sense that

$$\text{global_max_pool}[tg \triangleright_\rho F] = \rho(g) \text{global_max_pool}[F] \quad (4.51)$$

holds for any $tg \in \text{Aff}(G)$ and any $F \in L^2(\mathbb{R}^d, \mathbb{R}^c)$.

Proof: Since the maximum of field values is taken for each channel separately, this operation commutes with the channel permutation action of the field type. Taking the maximum field value is furthermore position-independent. We therefore have

$$\begin{aligned} & \text{global_max_pool}[tg \triangleright_\rho F] \\ &= \max_{x \in \mathbb{R}^d} [tg \triangleright_\rho F](x) \\ &= \max_{x \in \mathbb{R}^d} \rho(g) F((tg)^{-1}x) \\ &= \rho(g) \max_{x \in \mathbb{R}^d} F((tg)^{-1}x) \\ &= \rho(g) \max_{x \in \mathbb{R}^d} F(x) \\ &= \rho(g) \text{global_max_pool}[F]. \end{aligned} \quad (4.52)$$

□

Global norm max pooling: We define *global norm max pooling* for unitary representation field types ρ as the map

$$\text{global_norm_max_pool} : L^2(\mathbb{R}^d, \mathbb{R}^c) \rightarrow \mathbb{R} \quad (4.53)$$

defined by

$$\text{global_norm_max_pool}[F](x) = F\left(\arg \max_{x \in \mathbb{R}^d} |F(x)|\right). \quad (4.54)$$

Theorem 4.3.8 (Affine equivariance of global norm max pooling for unitary reps).

Let the field type $\rho : G \rightarrow \text{GL}(c)$ be a unitary representation. The global norm max pooling operation in Eq. (4.54) is then affine group equivariant in the sense that

$$\text{global_norm_max_pool}[tg \triangleright_\rho F] = \rho(g) \text{global_norm_max_pool}[F] \quad (4.55)$$

holds for any $tg \in \text{Aff}(G)$ and any $F \in L^2(\mathbb{R}^d, \mathbb{R}^c)$.

Proof: Let $tg \in \text{Aff}(G)$ and $F \in L^2(\mathbb{R}^d, \mathbb{R}^c)$, then

$$\begin{aligned} & \text{global_norm_max_pool}[tg \triangleright_\rho F] \\ &= [tg \triangleright_\rho F]\left(\arg \max_{x \in \mathbb{R}^d} |[tg \triangleright_\rho F](x)|\right) \\ &= \rho(g) F\left((tg)^{-1} \arg \max_{x \in \mathbb{R}^d} |\rho(g) F((tg)^{-1}x)|\right) \\ &= \rho(g) F\left((tg)^{-1}(tg) \arg \max_{x \in \mathbb{R}^d} |F(x)|\right) \\ &= \rho(g) \text{global_norm_max_pool}[F], \end{aligned} \quad (4.56)$$

where the third step made use of the unitarity of ρ and expressed the argument of the spatially transformed field in terms of a spatially transformed argument of the original field. □

Global average pooling: If we consider the (channel wise) *global average pooling* operation from Eq. (3.41), we find that it is affine equivariant and results in a feature vector of type $|\det g|\rho(g)$. The determinant factor accounts hereby for non-volume preserving field transformations. Note that $g \mapsto |\det g|\rho(g)$ is a well defined G -representation (linear group homomorphism), i.e. satisfies $|\det(gh)|\rho(gh) = (|\det g|\rho(g))(|\det h|\rho(h))$ for any $g, h \in G$.

Theorem 4.3.9 (Affine equivariance of global average pooling). *Let $F \in L^2(\mathbb{R}^d, \mathbb{R}^c)$ be a feature field of arbitrary type ρ and $tg \in \text{Aff}(G)$, then*

$$\text{global_avg_pool}[tg \triangleright_\rho F] = |\det g|\rho(g)\text{global_avg_pool}[F]. \quad (4.57)$$

Proof: To prove the claim, we simply need to observe that $\rho(g)$ commutes with the integration and substitute $\tilde{x} := (tg)^{-1}x$:

$$\begin{aligned} & \text{global_avg_pool}[tg \triangleright_\rho F] \\ &= \int_{\mathbb{R}^d} dx [tg \triangleright_\rho F](x) \\ &= \int_{\mathbb{R}^d} dx \rho(g) F((tg)^{-1}x) \\ &= |\det g|\rho(g) \int_{\mathbb{R}^d} d\tilde{x} F(\tilde{x}) \\ &= |\det g|\rho(g) \text{global_avg_pool}[F]. \end{aligned} \quad (4.58)$$

□

4.4 Local symmetries and equivariance group restriction

The key idea of equivariant networks is to exploit symmetries in the distribution of characteristic patterns in data. Specifically for feature fields, the level of symmetry might vary over different length scales. For instance, natural images typically show small features like edges, intensity gradients or the blossom leafs in Fig. 4.8 in arbitrary orientations and reflections. On a larger length scale, however, the rotational symmetry is broken, as manifested in visual patterns that are exclusively appearing in upright rotation but still in different positions and reflections. Each individual layer of a convolutional network should therefore be adapted to the symmetries present at the length scale of its neurons' receptive fields.

A loss of symmetry can be implemented by *restricting* the equivariance constraints at a certain layer to a subgroup $\text{Aff}(H) < \text{Aff}(G)$, where $H < G$; e.g. from rotations and reflections $G = \text{O}(2)$ to mere reflections $H = \mathcal{R}$ in the natural image example above. Formally, this is achieved by the (forgetful) restriction functor (Def. B.5.2)

$$\begin{aligned} \text{Res}_{\text{Aff}(H)}^{\text{Aff}(G)} : \left(L^2(\mathbb{R}^d, \mathbb{R}^c), \text{Ind}_G^{\text{Aff}(G)}\rho\right) &\rightarrow \left(L^2(\mathbb{R}^d, \mathbb{R}^c), \text{Res}_{\text{Aff}(H)}^{\text{Aff}(G)} \text{Ind}_G^{\text{Aff}(G)}\rho\right) \\ &\cong \left(L^2(\mathbb{R}^d, \mathbb{R}^c), \text{Ind}_H^{\text{Aff}(H)} \text{Res}_H^G \rho\right), \end{aligned} \quad (4.59)$$

which maps feature fields with induced $\text{Aff}(G)$ -action to fields with induced $\text{Aff}(H)$ -action by simply forgetting the full group action and acting with subgroup elements only. The isomorphism on the right-hand side emphasizes that the *restriction of a feature field as a whole* (first row) can equivalently be viewed as a *restriction of its field type* to an H -representation

Figure 4.8: Natural images have preferred “up” and “down” directions on their *global scale*, however, *local patterns*, like the leafs of the sunflower blossoms, appear commonly in arbitrary rotations. This can be exploited by composing a full network from subnetworks with different levels of equivariance. Formally, the transition from an $\text{Aff}(G)$ -equivariant subnetwork to an $\text{Aff}(H)$ -equivariant subnetwork with $H < G$ is achieved by a *group restriction* of the original field type $\rho : G \rightarrow \text{GL}(c)$ to type $\text{Res}_H^G \rho : H \rightarrow \text{GL}(c)$.

(Image credit: FreeImages.com/bodee)

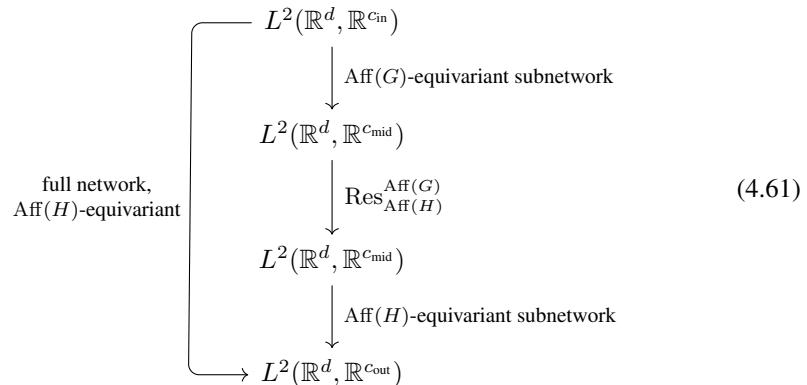


$\text{Res}_H^G \rho$. This equivalence is easily seen as follows.

$$\begin{aligned} \left[\operatorname{Res}_{\operatorname{Aff}(H)}^{\operatorname{Aff}(G)} \operatorname{Ind}_G^{\operatorname{Aff}(G)} \rho (th) F \right] (x) &= \left[\operatorname{Ind}_G^{\operatorname{Aff}(G)} \rho (th) F \right] (x) \\ &= \rho(h) F ((th)^{-1} x) \\ &= \left[\operatorname{Res}_H^G \rho \right] (h) F ((th)^{-1} x) \\ &= \left[\operatorname{Ind}_H^{\operatorname{Aff}(H)} \operatorname{Res}_H^G \rho (th) F \right] (x), \end{aligned} \quad (4.60)$$

where $F \in L^2(\mathbb{R}^d, \mathbb{R}^c)$, $x \in \mathbb{R}^d$ and $th \in \text{Aff}(H)$ are chosen arbitrarily. The removal and addition of the restriction functors in the first and third step was hereby possible since they are just a restriction of the respective representation's domain.

As a forgetful functor, the equivariance group restriction layer $\text{Res}_{\text{Aff}(H)}^{\text{Aff}(G)}$ does not change the actual data of the feature field on which it acts, and the resulting field will implicitly still transform according to the forgotten $\text{Aff}(G)$ -action. However, the restriction functor allows for the subsequent layers to be $\text{Aff}(H)$ -equivariant only, i.e. to break full $\text{Aff}(G)$ -equivariance. The resulting network *as a whole* will only be $\text{Aff}(H)$ -equivariant, while the $\text{Aff}(G)$ -equivariant subnetwork will nonetheless generalize its inference over a larger group of symmetries:



A simple example to illustrate these ideas are feature fields that transform under the regular representation (Def. B.5.18 and Remark B.5.19) of a dihedral group $G = D_N$. These fields have $c = 2N$ channels associated to N rotations in two reflections each. Upon restriction to

mere reflections $H = \mathcal{R} < D_N$, the $2N$ -dimensional feature vectors decouple into N two-dimensional fields, each transforming according to the regular representation of the reflection group. The D_N -action is implicitly still present, however, the reinterpretation as multiple reflection steerable fields puts less equivariance constraints on the following layer, allowing the network to break rotational equivariance (rotational relativity) by attending to different rotations in an absolute sense.

Conversely to the symmetry breaking on large length scales, it is also imaginable that local patterns are aligned, while patterns that emerge on a larger scale are more symmetrically distributed – think for instance about a solid that is locally magnetized (rotationally aligned) but globally amorphous (rotationally symmetric). This could analogously be exploited by merging $|G : H|$ fields of H -type ρ into a single field of induced G -type $\text{Ind}_H^G \rho$, where $H < G$.⁴ Other use cases would be autoencoders, which would require an induction step in their decoder subnetwork to mirror a group restriction in their encoder subnetwork, or generative adversarial networks (GANs).

An implementation and empirical evaluation of this network design was presented by Weiler and Cesa [322][38]. Romero and Lohit [249] proposed a related approach which relies on “equivariance” w.r.t. a learned (and not necessarily closed) subset of group elements. An example are networks that are “equivariant” w.r.t. rotations in an interval $[-\theta, \theta] \subseteq \text{SO}(2)$, where the angle $0 \leq \theta \leq \pi$ is learned.

4.5 Affine group convolutions as regular steerable convolutions

One of the most common and successful design for equivariant CNNs are *group convolutional neural networks* (GCNNs) [52, 162, 324, 10]. Here we argue that *group convolutions*⁵ correspond to *steerable convolutions with regular feature fields*, i.e. feature fields whose field type is the regular G -representation ρ_{reg}^G .

Group convolutions: Group convolutions process feature maps that are functions on symmetry groups and transform according to the (left) regular representations (Def. B.5.18) of these symmetry groups. In our specific application on affine equivariant CNNs, these would be the regular $\text{Aff}(G)$ -representations $\rho_{\text{reg}}^{\text{Aff}(G)}$, whose actions on functions

$$f : \text{Aff}(G) \rightarrow \mathbb{R}, \quad (4.62)$$

are for arbitrary tg and $\tilde{t}\tilde{g}$ in $\text{Aff}(G)$ given by

$$[\rho_{\text{reg}}^{\text{Aff}(G)}(tg)f](\tilde{t}\tilde{g}) := f((tg)^{-1}\tilde{t}\tilde{g}), \quad (4.63)$$

i.e. by a “shift” of f on $\text{Aff}(G)$ by tg .

An affine group convolution with a kernel $\kappa : \text{Aff}(G) \rightarrow \mathbb{R}$ on the group is then defined as

$$[f *_{\text{Aff}(G)} \kappa](tg) := \int_{\text{Aff}(G)} d(\tilde{t}\tilde{g}) f((\tilde{t}\tilde{g})) \kappa((\tilde{t}\tilde{g})^{-1}tg); \quad (4.64)$$

see Appendix F.1 for more details.

⁴ $|G : H|$ is the *index* (number of left cosets) of H in G , which equals $|G|/|H|$ for finite groups.

⁵Group convolutions are to be distinguished from “grouped convolutions” [290], which are just conventional convolutions with a sparse connectivity between “groups” of feature maps.

Equivalence: In order to show that $\text{Aff}(G)$ group convolutions are $\text{Aff}(G)$ steerable convolutions between G -regular feature fields, we leverage two properties of induced representations. Firstly, regular representations can be understood as being induced from the trivial representation $\rho_{\text{triv}}^{\{e\}}$ of the trivial group $\{e\}$. For instance, the regular G -representation is given by

$$\rho_{\text{reg}}^G \cong \text{Ind}_{\{e\}}^G \rho_{\text{triv}}^{\{e\}}. \quad (4.65)$$

Secondly, the induced representation satisfies for any nested subgroups $K \leq G \leq H$ and K -representation ρ^K the algebraic property

$$\text{Ind}_G^H \text{Ind}_K^G \rho^K \cong \text{Ind}_K^H \rho^K, \quad (4.66)$$

known as *induction in stages* [35]. Taken together, with $K := \{e\} \leq G < \text{Aff}(G) =: H$, this shows the equivalence of GCNN feature maps on $\text{Aff}(G)$ and G -regular feature fields on \mathbb{R}^d :

$$\begin{aligned} \rho_{\text{reg}}^{\text{Aff}(G)} &\cong \text{Ind}_{\{e\}}^{\text{Aff}(G)} \rho_{\text{triv}}^{\{e\}} && (\text{regular repr. induction from trivial, Eq. (4.65)}) \\ &\cong \text{Ind}_G^{\text{Aff}(G)} \text{Ind}_{\{e\}}^G \rho_{\text{triv}}^{\{e\}} && (\text{induction in stages, Eq. (4.66)}) \\ &\cong \text{Ind}_G^{\text{Aff}(G)} \rho_{\text{reg}}^G && (\text{regular repr. induction from trivial, Eq. (4.65)}) \end{aligned} \quad (4.67)$$

Since steerable convolutions are derived as the most general $\text{Aff}(G)$ -equivariant integral transforms between such feature spaces, they cover $\text{Aff}(G)$ group convolutions as a special case:

Theorem 4.5.1 ($\text{Aff}(G)$ group convolutions as G -regular steerable convolutions).

The feature spaces $L^2(\text{Aff}(G), \mathbb{R})$ and $L^2(\mathbb{R}^d, \mathbb{R}^{|G|})$ of $\text{Aff}(G)$ group convolutions and steerable convolutions between Euclidean feature fields of regular G -representation type, transforming according to $\rho_{\text{reg}}^{\text{Aff}(G)}$ and $\text{Ind}_G^{\text{Aff}(G)} \rho_{\text{reg}}^G$, respectively, are isomorphic.

$$(L^2(\text{Aff}(G), \mathbb{R}), \rho_{\text{reg}}^{\text{Aff}(G)}) \cong (L^2(\mathbb{R}^d, \mathbb{R}^{|G|}), \text{Ind}_G^{\text{Aff}(G)} \rho_{\text{reg}}^G) \quad (4.68)$$

The group convolutions and steerable convolutions between such feature spaces are equivalent.

Proof: The equivalence of the feature spaces (representation spaces) follows from 1) the induction in stages and 2) the induction of regular representations from trivial representations, as discussed above. That group convolutions and regular steerable convolutions are equivalent follows from the observation that both are the most general integral transforms between such feature spaces. \square

A similar equivalence holds between *convolutions on homogeneous spaces* on the one hand, described in Appendix F.2 and [162, 10], and *steerable convolutions with quotient representation fields* on the other hand.⁶

$$(4.69)$$

⁶In this case we use that G/K quotient representations (Def. B.5.20) are induced as $\rho_{\text{quot}}^{G/K} \cong \text{Ind}_K^G \rho_{\text{triv}}^K$, such that $\rho_{\text{quot}}^{\text{Aff}(G)/K} \cong \text{Ind}_K^{\text{Aff}(G)} \rho_{\text{triv}}^K \cong \text{Ind}_G^{\text{Aff}(G)} \text{Ind}_K^G \rho_{\text{triv}}^K \cong \text{Ind}_G^{\text{Aff}(G)} \rho_{\text{quot}}^{G/K} \forall K \leq G$.

Generality: Bekkers [10][11] argues that “*group convolutions are all you need*” to construct equivariant convolutional networks. Since group convolutions correspond specifically to steerable convolutions with regular feature fields, this raises the question in how far more general steerable convolutions with arbitrary field types are covered. In a nutshell, general feature fields for compact G may always be *embedded* into regular feature fields – implying that group convolutions are in this sense indeed sufficient to represent arbitrary steerable convolutions. However, an embedding of low-dimensional feature fields into high (or infinite) dimensional regular fields would consume an excessive amount of computational resources. Steerable convolutions, on the other hand, operate explicitly on feature fields in the relevant (sub)representation spaces. We would therefore argue that *steerable convolutions are all you need*, however, in this case *without excess subrepresentations*.

To make these arguments more precise, we make use of the Peter-Weyl theorem B.5.22, which guarantees that the regular representation of a compact group G decomposes into a direct sum of irreducible subspaces. Specifically, denoting the isomorphism classes of irreps of G by \widehat{G} , one has $\rho_{\text{reg}}^G \cong \widehat{\bigoplus}_{j \in \widehat{G}} \bigoplus_{i=1}^{m_j} \rho_j^G$, where $m_j \geq 1$ is the multiplicity of order j irreps ρ_j^G in the regular representation. The direct sum commutes with the induction functor [55], such that regular feature fields decompose into a direct sum of irrep feature fields:

$$\text{Ind}_G^{\text{Aff}(G)} \rho_{\text{reg}}^G \cong \text{Ind}_G^{\text{Aff}(G)} \widehat{\bigoplus}_{j \in \widehat{G}} \bigoplus_{i=1}^{m_j} \rho_j^G \cong \widehat{\bigoplus}_{j \in \widehat{G}} \bigoplus_{i=1}^{m_j} \text{Ind}_G^{\text{Aff}(G)} \rho_j^G \quad (4.70)$$

Note further that Theorem B.5.16 ensures the complete reducibility of finite dimensional representations of unitary groups G into irreps, which implies that arbitrary (finite dimensional) feature fields decompose in irrep fields as well. Since each irrep field appears in regular feature fields with positive multiplicity $m_j \geq 1$, a sufficient number of such regular fields allows to encode steerable feature fields of arbitrary (finite) type.

An example are scalar fields (trivial representation fields), which are in the regular representation embedded as constant functions on the group. However, instead of a single channel, this embedded encoding would require $|G|$ channels for finite and “infinite channels” for non-finite groups. As a second example, consider vector fields of cyclic groups $G = \mathbb{C}_N$, corresponding to the two-dimensional \mathbb{C}_N -irrep of order one. These vector fields could in principle be encoded into the N -dimensional regular representations of \mathbb{C}_N , however, with a linearly growing memory cost $O(N)$ and quadratically growing compute cost $O(N^2)$ instead of the fixed cost for steerable CNNs. We see therefore that while all steerable feature fields can in principle be embedded in regular feature fields, this approach might consume an excessive amount of memory and compute resources. In addition, the user would have to implement the embedding of general feature fields into regular fields manually to process e.g. vector field valued data like optical flow with regular GCNNs.

CHAPTER **5**

***G*-steerable convolution kernels**

Theorem 4.3.1 proved that $\text{Aff}(G)$ -equivariant convolutions generally require kernels $K : \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ satisfying the linear G -steerability constraint

$$K(gx) = \frac{1}{|\det g|} \rho_{\text{out}}(g) K(x) \rho_{\text{in}}(g)^{-1} \quad \forall x \in \mathbb{R}^d, g \in G, \quad (5.1)$$

which corresponds to the commutativity of the following equivariance diagram for any g in G :

$$\begin{array}{ccc} \mathbb{R}^d & \xrightarrow{K} & \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \\ g \cdot \downarrow & & \downarrow \frac{1}{|\det g|} \rho_{\text{out}}(g) [\cdot] \rho_{\text{in}}(g)^{-1} \\ \mathbb{R}^d & \xrightarrow{K} & \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \end{array} \quad (5.2)$$

As visualized in Fig. 4.6, a G -steerable kernel at $x \in \mathbb{R}^d$ is intuitively thought of as summarizing the field of input feature vectors around x such into an output feature vector at x that a local G -transformation of the input feature field by $g \in G$ results in a transformation of the resulting feature vector by $\rho_{\text{out}}(g)$.

This chapter investigates the nature of G -steerable kernels in greater detail, giving explicit examples and discussing general solution strategies for the kernel constraint. Section 5.1 starts off with the general observation that G -steerable kernels form a *vector space*. Any G -steerable kernel may consequently be parameterized and expanded in terms of a *basis* of G -steerable kernels. To build a first intuition for the workings of steerable kernels, Section 5.2 discusses the particularly simple and instructive example of *reflection steerable kernels*. The resulting kernels, visualized in Table. 5.1, exhibit some type of reflection symmetry, depending on the choice of input and output field types, shown in Fig. 5.1. Section 5.3 presents a generalized *Wigner-Eckart theorem for G -steerable kernels*, which gives a recipe to construct *complete steerable kernel bases* from 1) harmonic basis functions (Figs. 5.2 or 5.3), 2) irrep endomorphisms (reduced matrix elements), and 3) Clebsch-Gordan coefficients. It is applied to derive $\text{SO}(2)$ -steerable kernel bases, which are listed in Table 5.2. Section 5.4 gives a brief overview of alternative approaches that were proposed to parameterize equivariant convolution kernels.

Before coming to our investigation of G -steerable kernels, we want to mention that the G -steerability constraint may alternatively be derived in a more general *differential geometric* setting, as done in Sections 9.2.3 and 12.2.1 below. Kernels are here interpreted as

“local observers”, which measure a feature field relative to some choice of local reference frame. G -steerability is in this setting required to ensure that convolutional weight sharing is independent from particular choices of frames (gauges). This derivation emphasizes a compelling analogy between steerable – or frame relativistic – kernels on the one hand and relativistic laws of nature on the other hand: they are necessary for a group equivariant network inference and a Poincaré equivariant evolution of physical systems, respectively.

Implementations of G -steerable kernels for arbitrary representations of any subgroups $G \leq O(3)$ and $G \leq O(2)$ are available as part of the `escnn` library by Cesa et al. [39].

5.1 The vector space of G -steerable kernels and steerable basis expansion

Steerable kernel spaces: A first observation that we can make to parameterize steerable convolutions is that G -steerable kernels form a *vector space* and can therefore be *expanded in a steerable kernel basis*. To see this, consider the set

$$\mathcal{K} := \left\{ K : \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \right\} \quad (5.3)$$

of general, i.e. not necessarily G -equivariant, kernels. Equipped with the standard summation and scalar multiplication of functions, this set forms the usual vector space of convolution kernels. G -steerable kernels form the subset

$$\begin{aligned} \mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^G := \left\{ K : \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \mid K(gx) = \frac{1}{|\det g|} \rho_{\text{out}}(g) K(x) \rho_{\text{in}}(g)^{-1} \right. \\ \left. \forall x \in \mathbb{R}^d, g \in G \right\} \end{aligned} \quad (5.4)$$

of all convolution kernels that satisfy the G -steerability constraint in Eq. (5.1). As this constraint is linear, $\mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^G$ turns out to be a *linear vector subspace* of \mathcal{K} . It is therefore possible to solve for a *basis* of G -steerable kernels, in terms of which $\text{Aff}(G)$ -steerable convolutions can be parameterized. Note that the reduced dimensionality of the (sub)space of G -steerable kernels implies an *improved parameter efficiency* in comparison to conventional convolutions.

The remainder of this section will briefly elaborate on implementation aspects of steerable convolutions. We will thereby assume a basis of $\mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^G$ to be given, deferring a discussion of their construction to the following sections.

Steerable basis expansion: Working in the continuous setting, \mathcal{K} and $\mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^G$ are usually infinite dimensional function spaces, and would therefore require an infinite number of trainable parameters. To obtain a *finite steerable kernel basis* $\{K_1, \dots, K_N\}$, a given infinite basis of $\mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^G$ needs to be discretized subject to some smoothness conditions. For instance, $\text{SO}(2)$ -steerable kernels can generally be expanded in terms of the circular harmonics in Fig. 5.2 with an unconstrained radial part. A finite basis may then be obtained by restricting to 1) bandlimited circular harmonics up to some cutoff frequency and 2) smooth radial parts, e.g. the smooth rings in different rows of Fig. 5.2. Another example are reflection steerable kernels, Table. 5.1, which could be sampled on a finite number of points of a pixel grid.

Given a finite steerable kernel basis $\{K_1, \dots, K_N\}$, a general G -steerable kernel is expanded as

$$K = \sum_{i=1}^N w_i K_i, \quad (5.5)$$

where the expansion coefficients $\{w_1, \dots, w_N \mid w_i \in \mathbb{R}\}$ are the trainable parameters of the kernel. An $\text{Aff}(G)$ -steerable convolution operation is then just a conventional convolution operation after an additional kernel expansion step:

Algorithm: $\text{Aff}(G)$ -steerable convolution – forward pass

Input: G -steerable kernel basis $\{K_1, \dots, K_N\}$

trainable weights $\{w_1, \dots, w_N\}$

input feature field F_{in}

Output: output feature field F_{out}

expand learned kernel: $K \leftarrow \sum_{i=1}^N w_i K_i$

convolve to obtain output feature field: $F_{\text{out}} \leftarrow K * F_{\text{in}}$

return F_{out}

The cost of the additional kernel expansion operation is usually negligible compared to the cost of the convolution operation itself: when performing convolutions in continuous space (e.g. sampled at a point cloud), the continuous kernel needs to be expanded in some basis anyways. If the kernel is instead sampled on a square pixel grid, it does not need to be expanded, but is given directly as a parameter tensor of shape $(s_1, \dots, s_d, c_{\text{out}}, c_{\text{in}})$, where the $s_i \in \mathbb{N}$ are the extensions of the kernel pixel grid in different spatial dimensions. Consider furthermore a batch of $B \in \mathbb{N}$ feature fields, given as a tensor of shape $(B, X_1, \dots, X_d, c_{\text{in}})$, where $X_i \in \mathbb{N}$ are the spatial extensions of the fields' pixel grid. The computational cost of a convolution scales then as $\mathcal{O}(B \prod_i s_i X_i c_{\text{out}} c_{\text{in}})$.¹ In comparison, a sampled steerable kernel basis $\{K_1, \dots, K_N\}$ is represented by a tensor of shape $(N, s_1, \dots, s_d, c_{\text{out}}, c_{\text{in}})$, and the kernel expansion scales as $\mathcal{O}(N \prod_i s_i c_{\text{out}} c_{\text{in}})$. Since $N \ll B \prod_i X_i$ in usual applications, the kernel expansion's runtime is negligible in comparison to the convolution operation itself. Note furthermore that the kernel expansion is only necessary during training – as soon as the parameters are fixed (i.e. during test time), the kernel needs to be expanded only once and one is left solely with the cost of the convolution operation.

Similar arguments hold for the backward pass of $\text{Aff}(G)$ -steerable convolutions, where one needs to backpropagate additionally through the kernel expansion to obtain loss gradients $\frac{\partial \mathcal{L}}{\partial w_i}$ for the expansion coefficients w_i :

¹Here we are assuming a “spatial” implementation of the convolution operation, which is for the typically small kernels of convolutional networks cheaper than an implementation in terms of fast Fourier transforms.

Algorithm: Aff(G)-steerable convolution – backward pass

Input: G -steerable kernel basis $\{K_1, \dots, K_N\}$

trainable weights $\{w_1, \dots, w_N\}$

input feature field F_{in}

loss gradients w.r.t output feature field $\frac{\partial \mathcal{L}}{\partial F_{\text{out}}}$

Output: loss gradients w.r.t input feature field $\frac{\partial \mathcal{L}}{\partial F_{\text{in}}}$

loss gradients w.r.t trainable weights $\frac{\partial \mathcal{L}}{\partial w_i}, i = 1, \dots, N$

expand learned kernel: $K \leftarrow \sum_{i=1}^N w_i K_i$

backprop through convolution: $\frac{\partial \mathcal{L}}{\partial F_{\text{in}}}, \frac{\partial \mathcal{L}}{\partial K} \leftarrow \text{conv_backprop}(F_{\text{in}}, K, \frac{\partial \mathcal{L}}{\partial F_{\text{out}}})$

backprop through kernel expansion: $\frac{\partial \mathcal{L}}{\partial w_i} \leftarrow \frac{\partial K}{\partial w_i} \frac{\partial \mathcal{L}}{\partial K} = K_i \frac{\partial \mathcal{L}}{\partial K}$

return loss gradients $\frac{\partial \mathcal{L}}{\partial F_{\text{in}}}$ and $\frac{\partial \mathcal{L}}{\partial w_i}, i = 1, \dots, N$

The backward pass `conv_backprop` through the convolution is hereby performed as usual, i.e. it is given by convolutions of the output gradient field $\frac{\partial \mathcal{L}}{\partial F_{\text{out}}}$ with the (spatially reflected) kernel and input field. This operation is again dominating the overall computational cost in comparison to the cost of the kernel expansion and its backward pass. Note that the backpropagation through Aff(G)-steerable convolutions does not need to be implemented explicitly, since it is composed of differentiable primitives (kernel expansion and convolution), such that the backward pass is taken care of by differentiable programming languages.

We will in the following sections turn to analytical derivations of the steerable kernel bases themselves, starting with the example of reflection steerable kernels.

5.2 Simple example – reflection steerable kernels

Reflection steerable kernels are the arguably simplest example of G -steerable kernels. The reason for their simplicity is that the *reflection group* $\mathcal{R} = \{e, s\}$ consists of two elements, the identity e and the reflection (Spiegelung) s , only. They are composed according to the following simple multiplication table:

		e	s
e	e	s	
s	s	e	

(5.6)

The only nontrivial statement in this table is that two reflections annihilate, that is, $s^2 = e$, or, equivalently, $s^{-1} = s$. Here we assume the reflection group to be instantiated as a subgroup $\mathcal{R} =: G \leq \text{GL}(d)$ of the general linear group, such that it acts canonically on \mathbb{R}^d by reflecting points along some choice of reflection axis.

As a preparation for solving for reflection steerable kernels, Section 5.2.1 introduces some group representations of the reflection group, corresponding to the field types shown in Fig. 5.1. The steerable kernels that map between any pair of these field types are derived in Section 5.2.2. The reflection symmetries of the resulting kernels, shown in Table 5.1, guarantee the correct transformation behavior of feature fields and reduce the parameter cost by approximately a half.

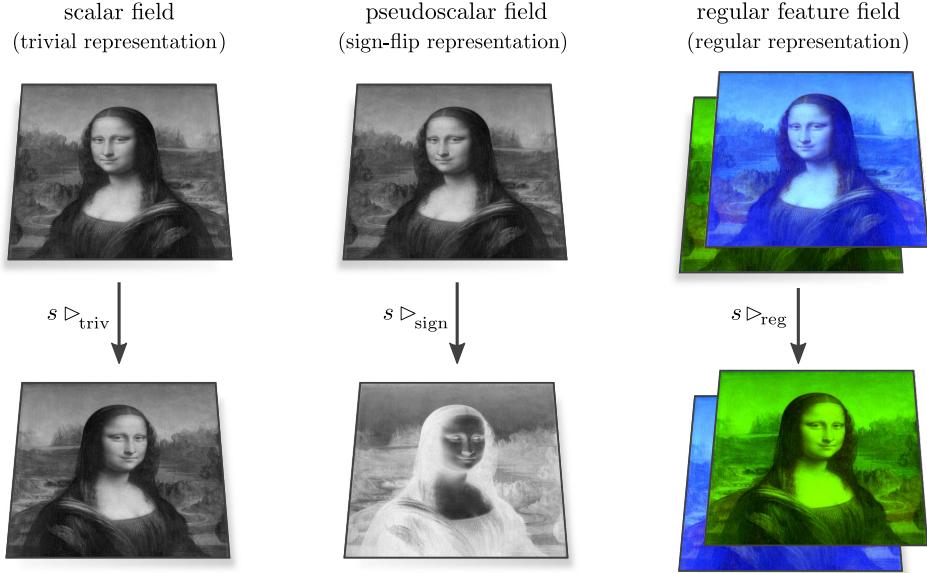


Figure 5.1: Examples of reflection steerable feature field types and their respective induced representation actions \triangleright_ρ . *Left:* Scalar fields correspond to the trivial representation ρ_{triv} , Eq. (5.7). The induced representation acts by reflecting the field spatially, but leaves the scalar values themselves invariant. *Middle:* The one-dimensional sign-flip representation ρ_{sign} , Eq. (5.9), is characterized by $\rho_{\text{sign}}(s) = -1$. It models pseudoscalar fields, whose values are negated under reflections. *Right:* The regular representation ρ_{reg} of the reflection group, Eq. (5.11), acts on two-dimensional feature vectors by swapping their two values. Regular feature fields consist therefore of two channels and the induced regular representation acts on such fields by 1) reflecting them spatially and 2) permuting their channels. Table 5.1 shows reflection steerable kernels that map between any pair of field types – the reader should validate intuitively that the given kernels will indeed guarantee the correct transformation behavior of the output field when being applied to transformed input fields.

Empirical results for convolutions with reflection steerable kernels are reported in Tables 6.2, 6.4, 6.5 and 6.6 where the reflection group $\mathcal{R} = D_1$ is sometimes identified with the dihedral group of order 1. Chapter 10 furthermore discusses orientation independent convolutions on the (non-orientable) Möbius strip, which necessarily need to apply reflection steerable kernels.

5.2.1 Reflection steerable feature fields

According to Def. 4.2.1, each feature field $F : \mathbb{R}^d \rightarrow \mathbb{R}^c$ is characterized by its geometric type – here a reflection group representation $\rho : \mathcal{R} \rightarrow \text{GL}(c)$ which specifies how the field's channels transform under reflections. We introduce three field types, scalar fields, pseudoscalar fields, and regular feature fields, which correspond to trivial, sign-flip and regular representations, respectively. The reader should check that the presented representations are in accordance with Def. B.5.1 indeed group homomorphisms (Def. B.1.3), i.e. satisfy $\rho(gh) = \rho(g)\rho(h) \quad \forall g, h \in \mathcal{R}$.

Scalar fields: The most basic example is the *trivial representation*

$$\rho_{\text{triv}} : \mathcal{R} \rightarrow \text{GL}(1) , \quad \begin{aligned} e &\mapsto [1] \\ s &\mapsto [1] \end{aligned} , \quad (5.7)$$

which assigns the 1×1 identity matrix to both group elements. It models *scalar fields* $F_{\text{triv}} : \mathbb{R}^d \rightarrow \mathbb{R}^1$, which reflect spatially, but whose scalar values stay invariant:

$$[s \triangleright_{\rho_{\text{triv}}} F_{\text{triv}}](x) = F_{\text{triv}}(sx) \quad (5.8)$$

Pseudoscalar fields: A second one-dimensional representation is the *sign-flip representation*

$$\rho_{\text{sign}} : \mathcal{R} \rightarrow \text{GL}(1) , \quad \begin{aligned} e &\mapsto [1] \\ s &\mapsto [-1] \end{aligned} , \quad (5.9)$$

which differs from the trivial representation by assigning the negative 1×1 identity matrix to reflections. The corresponding *pseudoscalar fields* $F_{\text{sign}} : \mathbb{R}^d \rightarrow \mathbb{R}^1$ change their sign when being reflected:

$$[s \triangleright_{\rho_{\text{sign}}} F_{\text{sign}}](x) = -F_{\text{sign}}(sx) \quad (5.10)$$

Since the trivial representation and the sign-flip representation are one-dimensional, they are both irreducible representations (irreps) of the reflection group. In fact, they are the only two irreps of the reflection group.

Regular feature fields: As a finite group, \mathcal{R} has a finite-dimensional (two-dimensional) *regular representation*

$$\rho_{\text{reg}} : \mathcal{R} \rightarrow \text{GL}(2) , \quad \begin{aligned} e &\mapsto \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ s &\mapsto \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \end{aligned} , \quad (5.11)$$

which represents the group elements by permutation matrices. By definition, the regular representation models the permutation of the group elements in \mathcal{R} when acting on themselves. Compare this to the columns of the multiplication table in Eq. (5.6): the middle column can be thought of as originating from the action of $\rho_{\text{reg}}(e)$ on the leftmost column, while the swapped group elements in the right column correspond to the permutation described by the action of $\rho_{\text{reg}}(s)$ on the left column.

The regular representation models *regular feature fields* $F_{\text{reg}} : \mathbb{R}^d \rightarrow \mathbb{R}^2$ of \mathcal{R} , which comprise two channels that are swapped under reflections:

$$[s \triangleright_{\rho_{\text{reg}}} F_{\text{reg}}](x) = \rho_{\text{reg}}(s)F_{\text{reg}}(sx) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} F_{\text{reg},1} \\ F_{\text{reg},2} \end{bmatrix}(sx) = \begin{bmatrix} F_{\text{reg},2} \\ F_{\text{reg},1} \end{bmatrix}(sx) \quad (5.12)$$

5.2.2 Reflection steerable convolutions

To solve for reflection steerable kernels for a given pair of input and output field type, consider the general *G*-steerability constraint in Eq. (5.1). For the reflection group, several things simplify:

ρ_{out}	ρ_{in}	trivial	sign-flip	regular
trivial		$K_{11}(sx) = K_{11}(x)$ 	$K_{11}(sx) = -K_{11}(x)$ 	$K_{11}(sx) = K_{12}(x)$
sign-flip		$K_{11}(sx) = -K_{11}(x)$ 	$K_{11}(sx) = K_{11}(x)$ 	$K_{11}(sx) = -K_{12}(x)$
regular		$K_{11}(sx) = K_{21}(x)$ 	$K_{11}(sx) = -K_{21}(x)$ 	$K_{11}(sx) = K_{22}(x)$ $K_{12}(sx) = K_{21}(x)$

Table 5.1: Visualization of reflection-steerable kernels for all considered pairs of input and output field types ρ_{in} and ρ_{out} and $d = 2$ spatial dimensions. In general, these kernels need to satisfy the \mathcal{R} -steerability kernel constraint $K(sx) = \rho_{\text{out}}(s)K(x)\rho_{\text{in}}(s)$ where $K : \mathbb{R}^2 \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$. Each entry of the table states the specific constraint for the corresponding input and output representations and visualizes one exemplary steerable kernel. Note that the constraint binds the reflected kernel $K(sx)$ to a linear transformation of the non-reflected kernel $K(x)$ by the input and output representation. It results therefore in reflectional symmetries of the kernels.

1. reflections are isometries, that is, the volume scaling factor $|\det g| = 1$ drops out.
2. the constraint needs to hold for any $g \in \mathcal{R}$, however, there are only two elements e and s , and the constraint is trivial for the former. We therefore only have to consider the reflection $g = s$.
3. since reflections are their own inverse, i.e. $s = s^{-1}$, we can replace $\rho_{\text{in}}(s)^{-1}$ with $\rho_{\text{in}}(s)$.

Overall, we obtain the simplified reflection steerability constraint

$$K(sx) = \rho_{\text{out}}(s) \cdot K(x) \cdot \rho_{\text{in}}(s) \quad \forall x \in \mathbb{R}^d, \quad (5.13)$$

stating that the spatially reflected kernel on the l.h.s. equals the non-reflected kernel on the r.h.s. after being left and right multiplied by the input and output representations, respectively.

We will in the following solve this constraint for all nine pairs of field types. The resulting kernels, all of which are in one or another sense symmetric under reflections, are visualized in Table 5.1.

- **scalar \leftarrow scalar:** Kernels $K = [K_{11}] : \mathbb{R}^d \rightarrow \mathbb{R}^{1 \times 1}$ which map between scalar fields are required to satisfy the constraint

$$[K_{11}](sx) = [1] \cdot [K_{11}](x) \cdot [1] = [K_{11}](x) \quad \forall x \in \mathbb{R}^d. \quad (5.14)$$

They are necessarily *symmetric* (invariant) under reflections; see the upper left entry in Table 5.1.

- **sign-flip \leftarrow scalar:** The kernels $K = [K_{11}] : \mathbb{R}^d \rightarrow \mathbb{R}^{1 \times 1}$ which map a scalar field to a sign-flip field need to satisfy

$$[K_{11}](sx) = [-1] \cdot [K_{11}](x) \cdot [1] = -[K_{11}](x) \quad \forall x \in \mathbb{R}^d. \quad (5.15)$$

This implies *antisymmetric* kernels as visualized in the middle row in the first column of Table 5.1.

- **regular \leftarrow scalar:** In order to map from a scalar field to a regular feature field one needs to apply kernels of the form $K = [K_{11}, K_{21}]^\top : \mathbb{R}^d \rightarrow \mathbb{R}^{2 \times 1}$, which map from one input channel to two output channels. The demanded permutation of the output channels is guaranteed if the kernel satisfies

$$\begin{bmatrix} K_{11} \\ K_{21} \end{bmatrix}(sx) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} K_{11} \\ K_{21} \end{bmatrix}(x) \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} K_{21} \\ K_{11} \end{bmatrix}(x) \quad \forall x \in \mathbb{R}^d. \quad (5.16)$$

This constraint requires that the two channels contain kernels which are *reflected copies* of each other, that is, $K_{11}(sx) = K_{21}(x)$ for all $x \in \mathbb{R}^d$ (this already covers the second line of the constraint in Eq. (5.16)). This case is visualized in the bottom left entry of Table 5.1.

- **scalar \leftarrow sign-flip:** Kernels $K = [K_{11}] : \mathbb{R}^d \rightarrow \mathbb{R}^{1 \times 1}$ that map from sign-flip to scalar fields are again *antisymmetric* since they need to satisfy the same constraint

$$[K_{11}](sx) = [1] \cdot [K_{11}](x) \cdot [-1] = -[K_{11}](x) \quad \forall x \in \mathbb{R}^d \quad (5.17)$$

like kernels which map in the opposite direction.

- **sign-flip \leftarrow sign-flip:** The kernels $K = [K_{11}] : \mathbb{R}^d \rightarrow \mathbb{R}^{1 \times 1}$ which preserve the transformation behavior of sign-flip fields are *symmetric* since the two sign inversions in the constraint

$$[K_{11}](sx) = [-1] \cdot [K_{11}](x) \cdot [-1] = [K_{11}](x) \quad \forall x \in \mathbb{R}^d \quad (5.18)$$

cancel out.

- **regular \leftarrow sign-flip:** In the case of kernels $K = [K_{11}, K_{21}]^\top : \mathbb{R}^d \rightarrow \mathbb{R}^{2 \times 1}$ which map from sign-flip to regular feature fields, we get the constraint

$$\begin{bmatrix} K_{11} \\ K_{21} \end{bmatrix}(sx) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} K_{11} \\ K_{21} \end{bmatrix}(x) \cdot \begin{bmatrix} -1 \\ 1 \end{bmatrix} = -\begin{bmatrix} K_{21} \\ K_{11} \end{bmatrix}(x) \quad \forall x \in \mathbb{R}^d. \quad (5.19)$$

The two lines imply each other, such that they can be summarized by the single kernel constraint $K_{11}(sx) = -K_{21}(x) \quad \forall x \in \mathbb{R}^d$. This constraint requires that the two channels of the kernel contain *reflected, negated copies* of each other; see the visualization in the middle of the bottom row of Table 5.1.

- **scalar \leftarrow regular:** The kernels which map regular feature fields to scalar fields have two input channels and one output channel and are therefore of the form $K = [K_{11}, K_{12}] : \mathbb{R}^d \rightarrow \mathbb{R}^{1 \times 2}$. The constraint

$$[K_{11}, K_{12}](sx) = [1] \cdot [K_{11}, K_{12}](x) \cdot \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = [K_{12}, K_{11}](x), \quad (5.20)$$

which can be reduced to the requirement $K_{11}(sx) = K_{12}(x) \quad \forall x \in \mathbb{R}^d$, again demands that the two entries of the kernel contain *reflected copies* of each other.

- **sign-flip \leftarrow regular:** Mappings from regular feature fields to sign-flip fields utilize kernels $K = [K_{11}, K_{12}] : \mathbb{R}^d \rightarrow \mathbb{R}^{1 \times 2}$ that satisfy

$$[K_{11}, K_{12}](sx) = [-1] \cdot [K_{11}, K_{12}](x) \cdot \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = -[K_{12}, K_{11}](x), \quad (5.21)$$

or, equivalently, $K_{11}(sx) = -K_{12}(x) \quad \forall x \in \mathbb{R}^d$. As probably already expected, they are made up from kernels whose two channels contain *reflected, negated copies* of another.

- **regular \leftarrow regular:** Lastly, we consider kernels $K = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} : \mathbb{R}^d \rightarrow \mathbb{R}^{2 \times 2}$ which map regular fields to regular fields and therefore have 2×2 matrices as codomain. Their constraint, coming from a left and right multiplication with the regular representation, becomes

$$\begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}(sx) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}(x) \cdot \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} K_{22} & K_{21} \\ K_{12} & K_{11} \end{bmatrix}(x) \quad (5.22)$$

for any $x \in \mathbb{R}^d$. This is equivalent to the two independent constraints

$$K_{11}(sx) = K_{22}(x) \quad \forall x \in \mathbb{R}^d \quad (5.23)$$

and

$$K_{12}(sx) = K_{21}(x) \quad \forall x \in \mathbb{R}^d, \quad (5.24)$$

which couple the four kernel entries such that there are *two pairs of mutually reflected kernels*. This case is visualized in the bottom right entry of Table 5.1.

While the derived results tell us how to map between individual feature fields, convolutional networks typically operate on feature spaces that consist of multiple, potentially differing feature fields. The kernels that map between these stacks of feature fields can be thought of as being built from blocks which map between the individual fields. To give an example, consider the case where both the input and output feature spaces contain one of the discussed representations each, that is, $\rho_{\text{in}} = \rho_{\text{out}} = \rho_{\text{triv}} \oplus \rho_{\text{sign}} \oplus \rho_{\text{reg}}$. The number of input and output channels is then $c_{\text{in}} = c_{\text{out}} = 1 + 1 + 2 = 4$, such that the full kernel is of the form $K : \mathbb{R}^d \rightarrow \mathbb{R}^{4 \times 4}$. Since the input and output representations are defined as direct sums, they are block diagonal. The full constraint decouples thus into nine independent constraints between all pairs of individual input and output fields, which correspond in this case exactly to the nine solutions presented above. The 4×4 entries of the full kernel will therefore be required to have the same symmetries as the 4×4 kernels which are visualized in Table 5.1 as a whole.

5.3 A generalized Wigner-Eckart theorem for G -steerable kernels

Let G be a *compact group*. The *Wigner-Eckart theorem* for steerable kernels by Lang and Weiler [173] and Cesa et al. [40], describes then the construction of a *complete basis* of G -steerable kernels on G -orbits on \mathbb{R}^d and which map between field types that are irreducible representations. General G -steerable kernels – on the whole of \mathbb{R}^d and which map between arbitrary finite dimensional field types – are easily assembled from these elementary solutions.

The ingredients from which the basis of G -steerable irrep kernels is constructed are

1. *harmonic basis functions* on G -orbits, e.g. the circular or spherical harmonics in Figs. 5.2 and 5.3,
2. *irrep endomorphisms*, Def. B.5.9, which allow to steer the harmonics in a learned manner, and
3. *Clebsch-Gordan coefficients*, Def. B.5.17, which determine the specific harmonics that are consistent with an equivariant mapping between the irreducible input and output field types.

While the harmonic basis and the Clebsch-Gordan decomposition are algebraically fixed, the endomorphisms form a vector space and constitute the learnable parameters of the kernel. They correspond to the *reduced matrix elements* in the original Wigner-Eckart theorem from quantum mechanics.

The following two Sections 5.3.1 and 5.3.2 lay the foundation of the Wigner-Eckart theorem by arguing that 1) the steerability constraint may always be restricted to individual G -orbits, and 2) one may w.l.o.g. consider *irreducible representations* as field types. The theorem itself is formulated in Section 5.3.3. Since a formal proof would require a deep dive into representation theory, we omit it here and point the reader to the original publications [173, 40]. Instead, Appendix H discusses a succession of increasingly complex constraints, which introduces the three ingredients and their role in steerable kernels step by step. Section 5.3.4 applies the theorem to derive $\text{SO}(2)$ -steerable kernels. Considerations regarding the sampling of continuous kernels on pixel grids are briefly discussed in Section 5.3.5.

5.3.1 Restriction to G -orbits

Inspecting the steerability constraint, we see that it relates kernel values at points $x \in \mathbb{R}^d$ to those at any other point gx on the group orbit $Gx = \{gx \mid g \in G\} \subset \mathbb{R}^d$; see Def. B.3.3 and Fig. B.3. Values on different orbits are not related, such that we can focus on solving the constraint on the individual orbits. By construction, the G -action is transitive (Def. B.3.8) on the orbits, making them *homogeneous G -spaces* (Def B.3.11) – we will therefore from now on consider G -steerable kernels $K : X \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ on homogeneous spaces $X \subset \mathbb{R}^d$.

A common example for homogeneous spaces for $\text{SO}(2)$ -steerable kernels are circles of different radii, visualized in Fig. 5.2. Similarly, for $G = \text{SO}(3)$, we get (the origin and) spheres of different radii as homogeneous spaces, allowing to solve for steerable kernels on S^2 instead of \mathbb{R}^3 , as shown in Fig. 5.3.

While one can solve the constraint independently on the orbits, it may be desirable to obtain a solution that is *smooth* when embedding the orbits in \mathbb{R}^d . Cesa et al. [40] explain how this smoothness can generally be achieved.

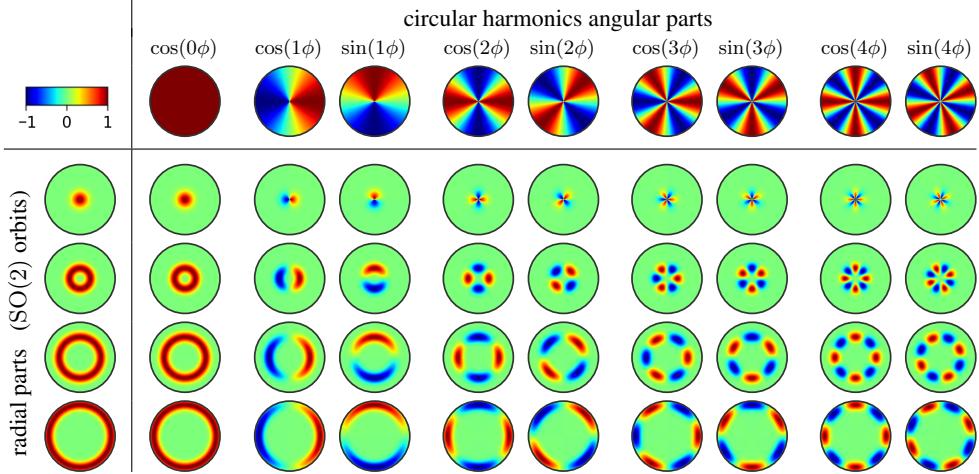


Figure 5.2: Visualization of *circular harmonics* angular parts ($\cos(j\phi)$, $\sin(j\phi)$) for different frequencies $j \geq 0$ (top row), modulated by different radial parts (left column) to form kernels on $\mathbb{R}^2 \cong S^1 \times \mathbb{R}_{>0} \cup \{0\}$. The circular harmonics span the (real) irreducible subrepresentations of the regular representation of $\text{SO}(2)$, i.e. square integrable functions on the circle $\text{SO}(2) \cong_{\text{top}} S^1$. The constant function for frequency zero corresponds to the one-dimensional trivial representation. All other (real) irreps are two-dimensional and transform according to frequency j rotation matrices (Eq. (5.33)). Any $\text{SO}(2)$ or $\text{O}(2)$ -steerable kernel can be expanded in terms of a *subspace of circular harmonics* as prescribed by the Wigner-Eckart theorem for G -steerable kernels [173, 40]; see Tables 5.2 and 5.3.

5.3.2 Restriction to irrep fields

If the structure group G is *compact*, we can w.l.o.g. consider *unitary representations*; see Appendix B.5.2. Theorem B.5.16 asserts the *complete reducibility* of finite unitary representations (of any group) into a *direct sum of irreps*, while the Peter-Weyl theorem B.5.22 proves the same property for regular and quotient representations of compact groups. We can for compact groups G furthermore drop the determinant factor $|\det g| = 1$.

Assuming the complete reducibility of the field types from now on, we show that it is sufficient to solve the kernel constraint for irreducible representations. Let $Q_{\text{in}} \in \text{GL}(c_{\text{in}})$ and $Q_{\text{out}} \in \text{GL}(c_{\text{out}})$ be the change of basis matrices that decouple the field types ρ_{in} and ρ_{out} into irreps, that is,

$$Q_{\text{in}} \rho_{\text{in}}(g) Q_{\text{in}}^{-1} = \bigoplus_{l \in I_{\text{in}}} \rho_l(g) \quad \text{and} \quad Q_{\text{out}} \rho_{\text{out}}(g) Q_{\text{out}}^{-1} = \bigoplus_{J \in I_{\text{out}}} \rho_J(g), \quad (5.25)$$

where I_{in} and I_{out} are index sets of the irreps ρ_l and ρ_J contained in ρ_{in} and ρ_{out} , respectively. Left and right multiplying the kernel constraint with Q_{out} and Q_{in}^{-1} , and inserting identities of the form $\text{id}_{\mathbb{R}^{c_{\text{out}}}} = Q_{\text{out}}^{-1} Q_{\text{out}}$ and $\text{id}_{\mathbb{R}^{c_{\text{in}}}} = Q_{\text{in}}^{-1} Q_{\text{in}}$ yields the equivalent constraint

$$\begin{aligned} Q_{\text{out}} K(gx) Q_{\text{in}}^{-1} &= Q_{\text{out}} \rho_{\text{out}}(g) [Q_{\text{out}}^{-1} Q_{\text{out}}] K(x) [Q_{\text{in}}^{-1} Q_{\text{in}}] \rho_{\text{in}}(g)^{-1} Q_{\text{in}}^{-1} \\ \iff K_{\text{irrep}}(gx) &= \left(\bigoplus_{J \in I_{\text{out}}} \rho_J(g) \right) K_{\text{irrep}}(x) \left(\bigoplus_{l \in I_{\text{in}}} \rho_l(g) \right)^{-1}, \end{aligned} \quad (5.26)$$

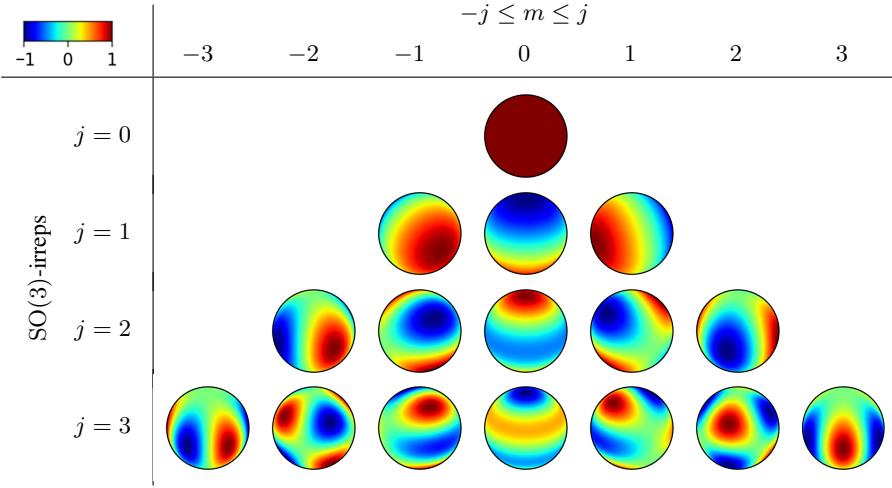


Figure 5.3: Visualization of (real) *spherical harmonics* $\vec{Y}_j : S^2 \rightarrow \mathbb{R}^{2j+1}$ (rows), where $j \in \mathbb{N}_0$, with components Y_j^m for $-j \leq m \leq j$ (columns). The spherical harmonics span the $(2j+1)$ -dimensional irreducible representation ρ_j of $\text{SO}(3)$ (Wigner D-matrices). They are the harmonic basis functions occurring in the Peter-Weyl decomposition (Theorem B.5.22) of the quotient representation $\rho_{\text{quot}}^{\text{SO}(3)/\text{SO}(2)}$ of $\text{SO}(3)$ on $\text{SO}(3)/\text{SO}(2) \cong_{\text{top}} S^2$. Any $\text{SO}(3)$ -steerable kernel on $\mathbb{R}^3 \cong S^2 \times \mathbb{R}_{>0} \cup \{0\}$ can be constructed in terms of spherical harmonics angular parts, modulated by an unconstrained radial part (the 3d analog of Fig. 5.2) [323].

for any $x \in X$ and $g \in G$ on the kernel $K_{\text{irrep}} := Q_{\text{out}} K Q_{\text{in}}^{-1}$ in the irrep basis. Since the direct sum representations are block diagonal, the irrep kernel decomposes into $|I_{\text{in}}| \cdot |I_{\text{out}}|$ blocks $K_{\text{irrep}}^{Jl} : X \rightarrow \mathbb{R}^{\dim \rho_J \times \dim \rho_l}$ that are independently required to satisfy the *kernel constraint for irreps*

$$K_{\text{irrep}}^{Jl}(gx) = \rho_J(g) K_{\text{irrep}}^{Jl}(x) \rho_l(g)^{-1} \quad \forall x \in X, g \in G. \quad (5.27)$$

This decomposition into irreducible kernel blocks is visualized in the following equation:

$$\underbrace{\begin{pmatrix} K_{\text{irrep}}^{J_1 l_1} & K_{\text{irrep}}^{J_1 l_2} & \dots \\ K_{\text{irrep}}^{J_2 l_1} & K_{\text{irrep}}^{J_2 l_2} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}}_{K_{\text{irrep}}(gx)} = \underbrace{\begin{pmatrix} \rho_{J_1}(g) & & & \\ & \rho_{J_2}(g) & & \\ & & \ddots & \\ & & & \ddots \end{pmatrix}}_{\bigoplus_{J \in I_{\text{out}}} \rho_J(g)} \underbrace{\begin{pmatrix} K_{\text{irrep}}^{J_1 l_1} & K_{\text{irrep}}^{J_1 l_2} & \dots \\ K_{\text{irrep}}^{J_2 l_1} & K_{\text{irrep}}^{J_2 l_2} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}}_{K_{\text{irrep}}(x)} \underbrace{\begin{pmatrix} \rho_{l_1}(g)^{-1} & & & \\ & \rho_{l_2}(g)^{-1} & & \\ & & \ddots & \\ & & & \ddots \end{pmatrix}}_{\bigoplus_{l \in I_{\text{in}}} \rho_l(g)^{-1}}$$

Given a basis for irrep steerable kernels K_{irrep}^{Jl} , the general steerable kernel basis is recovered by inserting these solutions into the right blocks and undoing the change of basis. For the remainder of this section, we will exclusively consider irrep fields $\rho_{\text{in}} = \rho_l$ and $\rho_{\text{out}} = \rho_J$, but drop the subscript ‘‘irrep’’ to reduce clutter. Note that, while the steerability constraint may w.l.o.g. be solved in the irrep basis, the specific choice of basis matters as soon as we apply nonlinear network operations.

5.3.3 Statement of the Wigner-Eckart theorem

This section states the Wigner-Eckart theorem for G -steerable kernels. It assumes some familiarity with representation theoretic concepts that are introduced in Appendix B.5. The

individual ingredients going into the construction of the theorem are motivated in a less formal setting in Appendix H.

To formulate the theorem, we fix some notation. Let (ρ_j, V_j) , $j \in \widehat{G}$, be the j -th order real irreducible representation of G and assume a basis to be chosen, such that we can identify the representation space V_j with \mathbb{R}^{\dim_j} , where $\dim_j := \dim V_j$. Denote the considered G -orbit by $X \subset \mathbb{R}^d$ and let $\vec{Y}_{ji} : X \rightarrow \mathbb{R}^{\dim_j}$ be the harmonic basis functions that span the irrep subspaces $V_{ji} \cong V_j$, $i = 1, \dots, m_j$ in the Peter-Weyl decomposition (Theorem B.5.22) of $L^2_{\mathbb{R}}(X) \cong \bigoplus_{j \in \widehat{G}} \bigoplus_{i=1}^{m_j} V_{ji}$. Consider the endomorphism space $\text{End}(V_j)$ of j -th order irreps (Def. B.5.9) and let $c_{jr} \in \text{End}(V_j)$, $r = 1, \dots, \dim \text{End}(V_j)$, be any basis of the endomorphism space. Let furthermore $\text{CG}_{lJ} : V_l \otimes V_J \rightarrow \bigoplus_{j \in \widehat{G}} \bigoplus_{s=1}^{m_{j,l}} V_j$ be the Clebsch-Gordan decomposition from Def. B.5.17, which decomposes the irrep tensor product $V_l \otimes V_J$ back into irreducible subspaces, where irrep order j occurs with multiplicity $m_{j,lJ}$. Denote the projection of $V_l \otimes V_J$ on the s -th irreducible subspace of order j by $\text{CG}_{lJ,js} := \text{proj}_{js} \circ \text{CG}_{lJ}$ and its pseudoinvserse, which embeds that subspace into the tensor product, by $\text{CG}_{lJ,js}^+$.

Given the assumed identification of V_j with \mathbb{R}^{\dim_j} , we can identify the endomorphisms c_{jr} with matrices in $\mathbb{R}^{\dim_j \times \dim_j}$ and the Clebsch-Gordan projectors $CG_{lJ,js}$ with matrices in $\mathbb{R}^{\dim_j \times (\dim_J \cdot \dim_l)}$. The pseudoinverses $CG_{lJ,js}^+$ are then $\mathbb{R}^{(\dim_J \cdot \dim_l) \times \dim_j}$ -matrices, which turn out to be given by the transpose of the matrices representing $CG_{lJ,js}$. Lastly, let $\text{vec} : \mathbb{R}^{\dim_J \times \dim_l} \rightarrow \mathbb{R}^{\dim_J \cdot \dim_l}$ be the vectorization operator, which stacks the columns of a $\dim_J \times \dim_l$ -matrix into a $\dim_J \cdot \dim_l$ -dimensional vector, and let unvec be its inverse.

Equipped with these prerequisites, we can state the Wigner-Eckart theorem for steerable vectors in the formulation of Cesa et al. [40]:²

Theorem 5.3.1 (Wigner-Eckart theorem for G -steerable kernels). Let $G \leq \mathrm{GL}(d)$ be compact and consider a G -orbit X and irreducible field types $\rho_{\text{in}} = \rho_l$ and $\rho_{\text{out}} = \rho_J$. The space of G -steerable irrep kernels

$$\mathcal{K}_{\rho_l, \rho_J}^{G, X} := \left\{ K^{Jl} : X \rightarrow \mathbb{R}^{\dim_J \times \dim_l} \mid K^{Jl}(gx) = \rho_J(g) \cdot K^{Jl}(x) \cdot \rho_l(g)^{-1} \quad (5.28) \right. \\ \left. \forall x \in X, g \in G \right\}$$

on X is then spanned by basis kernels

$$K_{srji}^{Jl} = \text{unvec } \text{CG}_{lJ,js}^+ c_{jr} \vec{Y}_{ji}, \quad (5.29)$$

that is,

$$\mathcal{K}_{\rho_i, \rho_j}^{G, X} = \text{span} \left\{ K_{srji}^{Jl} \mid j \in \widehat{G}, \ s \leq m_{j,lJ}, \ i \leq m_j, \ r \leq \dim \text{End}(V_j) \right\}. \quad (5.30)$$

The following diagram visualizes the definition of the steerable basis:

$$X \xrightarrow{\vec{Y}_{ji}} \mathbb{R}^{\dim_j} \xrightarrow{c_{jr}} \mathbb{R}^{\dim_j} \xrightarrow{\text{CG}_{lJ,js}^+} \mathbb{R}^{\dim_J \cdot \dim_l} \xrightarrow{\text{unvec}} \mathbb{R}^{\dim_J \times \dim_l}$$

K_{srji}^{Jl}

(5.31)

²The formulation by Lang and Weiler [173] is equivalent, but differs in that it decomposes a tensor product $V_l \otimes V_j$ into irreps V_J , instead of the decomposition of $V_l \otimes V_j$ into irreps V_j here.

Proof: For a formal proof, we point the reader to the original publications Lang and Weiler [173] and Cesa et al. [40]. A constructivistic motivation of the theorem is given in Appendix H. \square

A steerable kernel on \mathbb{R}^d as a whole can be assembled from these solutions for the individual G -orbits. Section 5.3.4 below solves for $\text{SO}(2)$ -steerable kernels on \mathbb{R}^2 as a practically relevant example application of the theorem.

Selection rules: Note that the Clebsch-Gordan coefficients $\text{CG}_{l,J,j}$ are generally sparse in j – implying that only certain harmonics frequencies j are allowed to map between field types of irrep orders l and J . This is a direct analog to the *selection rules* in quantum mechanics, which restrict the range of operators that may map between certain quantum states.

For instance, for $G = \text{SO}(3)$, the Clebsch-Gordan decomposition of $V_l \otimes V_J$ contains all of the $2 \min(l, J) + 1$ irreps V_j with indices $|l - J| \leq j \leq l + J$. $\text{SO}(3)$ -steerable kernels that map between fields of types V_l and V_J contain therefore only harmonics \vec{Y}_j with frequencies in that range. Fig. 5.4 visualizes the admissible transitions between different $\text{SO}(3)$ field types.

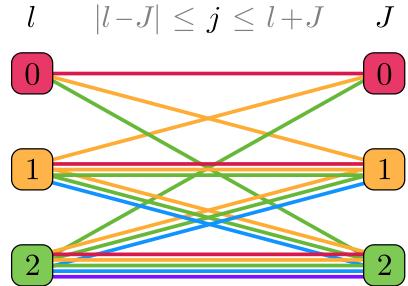


Figure 5.4: Visualization of the selection rules specifying which harmonics \vec{Y}_j may map between $\text{SO}(3)$ -irrep fields of orders l and J . The irrep orders are color coded in red (0), yellow (1), green (2), blue (3) and violet (4).

5.3.4 $\text{SO}(2)$ and $\text{O}(2)$ -steerable kernels

To illustrate the practical use of the Wigner-Eckart theorem, we derive complete bases for $\text{SO}(2)$ -steerable irrep kernels on \mathbb{R}^2 . The $\text{SO}(2)$ -action partitions the plane into orbits, which are the origin and rings at different radii. As the solutions at the origin are trivial (they turn out to be intertwiners), we consider the case of orbits $X \cong S^1$ that are circles, that is, we are interested in kernels with *angular parts*

$$K^{Jl} : S^1 \rightarrow \mathbb{R}^{\dim_J \times \dim_l} \quad \text{such that} \quad K^{Jl}(g\phi) = \rho_J(g)K^{Jl}(\phi)\rho_l(g)^{-1} \quad (5.32)$$

for any $\phi \in S^1$ and $g \in \text{SO}(2)$.

Full kernels on \mathbb{R}^2 follow by expanding this angular steerable kernel basis with an *unconstrained radial part*.

The following paragraph lists the ingredients required by the Wigner-Eckart theorem. Subsequently, we put these ingredients together to construct steerable kernel bases. The complete solution spaces of the steerability constraint for different $\text{SO}(2)$ -irreps of orders l and J are summarized in Table 5.2. Table 5.3 gives their analog for $G = \text{O}(2)$.

List of ingredients: To construct the steerable kernel basis according to the Wigner-Eckart theorem, we require the harmonics, endomorphism spaces and Clebsch-Gordan coefficients for all $\text{SO}(2)$ -irreps. The irreps of $\text{SO}(2)$ are the trivial irrep (ρ_0, V_0) with $V_0 = \mathbb{R}$ and the two-dimensional irreps (ρ_j, V_j) , $j \geq 1$, with $V_j = \mathbb{R}^2$, whose actions are given by the

identity matrix and frequency j rotation matrices, respectively:

$$\rho_0(\phi) = (1) \quad \text{and, for } j \geq 1, \quad \rho_j(\phi) = \begin{pmatrix} \cos(j\phi) & -\sin(j\phi) \\ \sin(j\phi) & \cos(j\phi) \end{pmatrix} \quad (5.33)$$

Their *endomorphism spaces* are

$$\text{End}(V_0) = \text{span}(c_{01}) \quad \text{and, for } j \geq 1, \quad \text{End}(V_j) = \text{span}(c_{j1}, c_{j2}) \quad (5.34)$$

$$\text{with } c_{01} = (1) \quad \text{with } c_{j1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } c_{j2} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

The corresponding harmonics are the *circular harmonics*

$$\vec{Y}_0(\phi) = (1) \quad \text{and, for } j \geq 1, \quad \vec{Y}_j(\phi) = \begin{pmatrix} \cos(j\phi) \\ \sin(j\phi) \end{pmatrix} \quad (5.35)$$

that are visualized in Fig. 5.2. Note that we dropped the index i , which is possible since all orders j occur with multiplicity $m_j = 1$ in the Peter-Weyl decomposition.

The *non-zero* Clebsch-Gordan coefficients $\text{CG}_{l,J}$ are in matrix form given by:

$l = J = 0$: if both irreps are trivial, we have $V_0 \otimes V_0 \cong V_0$ and $\text{CG}_{00,01} = (1)$

$l = 0, J \geq 1$: if V_l is trivial and V_J is not, then $V_0 \otimes V_J \cong V_J$ and $\text{CG}_{0J,J1} = \text{id}_{\mathbb{R}^2}$

$l \geq 1, J = 0$: for non-trivial V_l and trivial V_J , we have $V_l \otimes V_0 \cong V_l$ and $\text{CG}_{l0,l1} = \text{id}_{\mathbb{R}^2}$

$l, J \geq 1$: for non-trivial irreps V_l and V_J , one has to distinguish three different cases. The tensor product splits in all cases into multiple irreps, such that we need to consider different values for the indices j and s in $\text{CG}_{lJ,js}$.

$J > l \geq 1$: in this case, $V_l \otimes V_J \cong V_{J-l} \oplus V_{l+J}$ with

$$\text{CG}_{l,J} = \begin{pmatrix} \text{CG}_{lJ,J-l,1} \\ \text{CG}_{lJ,J+l,1} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \end{pmatrix}. \quad (5.36)$$

$l > J \geq 1$: here, $V_l \otimes V_J \cong V_{l-J} \oplus V_{l+J}$ with

$$\text{CG}_{l,J} = \begin{pmatrix} \text{CG}_{lJ,l-J,1} \\ \text{CG}_{lJ,l+J,1} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \end{pmatrix}. \quad (5.37)$$

$l = J \geq 1$: if the irrep orders are equal, $V_l \otimes V_J \cong V_0 \oplus V_0 \oplus V_{l+J}$ and

$$\text{CG}_{l,J} = \begin{pmatrix} \text{CG}_{lJ,0,1} \\ \hline \text{CG}_{lJ,0,2} \\ \hline \text{CG}_{lJ,l+J,1} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \quad (5.38)$$

Note that these Clebsch-Gordan coefficients are ultimately the same, just with the roles of l and J swapped in the first two cases and, in the third case $l = J$, with the projection $\text{CG}_{lJ,|l-J|,1}$ of $V_l \otimes V_J$ on $V_{|l-J|} = \mathbb{R}^2$ split further into two projections $\text{CG}_{lJ,0,1}$ and $\text{CG}_{lJ,0,2}$ on the two separate trivial subrepresentations $V_0 \oplus V_0 = \mathbb{R} \oplus \mathbb{R}$.

All other Clebsch-Gordan coefficients are zero. As stated above, the pseudoinverses $\text{CG}_{lJ,js}^+$ are given by transposition of $\text{CG}_{lJ,js}$.

Constructing solutions: With these ingredients we are ready to instantiate the $\text{SO}(2)$ -steerable bases with elements

$$K_{srj}^{Jl} := \text{unvec } \text{CG}_{lJ,js}^+ c_{jr} \vec{Y}_j \quad j \in \widehat{G}, \quad s \leq m_{j,lJ}, \quad r \leq \dim \text{End}(V_j), \quad (5.39)$$

where we dropped the index i as mentioned above. The following list constructs these bases for all qualitatively different combinations of input and output irreps. The admissible circular harmonic frequencies j are thereby determined by the *non-zero* Clebsch-Gordan coefficients listed above. All results are summarized in Table 5.2.

- $\mathbf{V}_0 \leftarrow \mathbf{V}_0$: Kernels $K^{00} : S^1 \rightarrow \mathbb{R}^{1 \times 1}$ that map between scalar fields, i.e. $l = J = 0$, have a one-dimensional (angular) basis

$$K_{110}^{00}(\phi) = \text{unvec } \text{CG}_{00,01}^+ c_{01} \vec{Y}_0 = \text{unvec } (1)(1)(1) = (1), \quad (5.40)$$

that is, they are necessarily rotation invariant. Fig. 16.5 visualizes such kernels when adding a learnable radial part.

- $\mathbf{V}_{J \geq 1} \leftarrow \mathbf{V}_0$: Convolutions that map scalar fields, $l = 0$, to non-trivial irrep fields, $J \geq 1$, rely on circular harmonics kernels $K^{J0} : S^1 \rightarrow \mathbb{R}^{2 \times 1}$ of frequency J . Their two-dimensional (angular) basis is spanned by

$$\begin{aligned} K_{11J}^{J0}(\phi) &= \text{unvec } \text{CG}_{0J,J1}^+ c_{J1} \vec{Y}_J \\ &= \text{unvec } \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(J\phi) \\ \sin(J\phi) \end{pmatrix} = \begin{pmatrix} \cos(J\phi) \\ \sin(J\phi) \end{pmatrix} \end{aligned} \quad (5.41)$$

and

$$\begin{aligned} K_{12J}^{J0}(\phi) &= \text{unvec } \text{CG}_{0J,J2}^+ c_{J2} \vec{Y}_J \\ &= \text{unvec } \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \cos(J\phi) \\ \sin(J\phi) \end{pmatrix} = \begin{pmatrix} -\sin(J\phi) \\ \cos(J\phi) \end{pmatrix}. \end{aligned} \quad (5.42)$$

Learned linear combinations of these two basis kernels correspond to a phase-shifted and amplitude scaled circular harmonics pair of frequency J . Fig. 4.1 visualizes the corresponding convolution for $l = 0$ and $J = 3$ (ignoring the learnable radial part and with arbitrary phase-shift).

- $\mathbf{V}_0 \leftarrow \mathbf{V}_{l \geq 1}$: If we are, conversely, mapping an order $l \geq 1$ irrep field to a scalar field, $J = 0$, this requires kernels $K^{0l} : S^1 \rightarrow \mathbb{R}^{1 \times 2}$ that are spanned by the transpose of the previous case, namely

$$\begin{aligned} K_{11l}^{0l}(\phi) &= \text{unvec } \text{CG}_{l0,l1}^+ c_{l1} \vec{Y}_l \\ &= \text{unvec } \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(l\phi) \\ \sin(l\phi) \end{pmatrix} = (\cos(l\phi) \quad \sin(l\phi)) \end{aligned} \quad (5.43)$$

and

$$\begin{aligned} K_{12l}^{0l}(\phi) &= \text{unvec } \text{CG}_{l0,l2}^+ c_{l2} \vec{Y}_l \\ &= \text{unvec } \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \cos(l\phi) \\ \sin(l\phi) \end{pmatrix} = (-\sin(l\phi) \quad \cos(l\phi)) \end{aligned} \quad (5.44)$$

Note that the unvec operator maps the vectorized kernel here to a 1×2 instead of a 2×1 kernel, as in the previous case.

- $\mathbf{V}_{J \geq 1} \leftarrow \mathbf{V}_{l \geq 1}$: If both the input and output field correspond to non-trivial irreps, $J, l \geq 1$, we have kernels $K^{Jl} : S^1 \rightarrow \mathbb{R}^{2 \times 2}$. As there are different Clebsch-Gordan coefficients for $J > l \geq 1$, $l > J \geq 1$ and $l = J \geq 1$, we need to make the same distinction in our construction of the steerable basis. As it turns out, the resulting solutions can ultimately be brought in the same form, which allows to ignore this case distinction in the solution Table 5.2.
- $J > l \geq 1$: In this case we have the two non-zero Clebsch-Gordan coefficients $\text{CG}_{lJ, J-l, 1}$ and $\text{CG}_{lJ, J+l, 1}$ from Eq. (5.36) and the two endomorphism basis elements c_{j1} and c_{j2} (for $j = J - l$ or $j = J + l$) from Eq. (5.34). This results in four (angular) basis kernels, two for the sum and two for the difference frequencies:

$$\begin{aligned} K_{11, J-l}^{Jl}(\phi) &= \text{unvec } \text{CG}_{lJ, J-l, 1}^+ c_{J-l, 1} \vec{Y}_{J-l} & (5.45) \\ &\propto \text{unvec} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos((J-l)\phi) \\ \sin((J-l)\phi) \end{pmatrix} \\ &= \begin{pmatrix} \cos((J-l)\phi) & -\sin((J-l)\phi) \\ \sin((J-l)\phi) & \cos((J-l)\phi) \end{pmatrix} \end{aligned}$$

$$\begin{aligned} K_{12, J-l}^{Jl}(\phi) &= \text{unvec } \text{CG}_{lJ, J-l, 1}^+ c_{J-l, 2} \vec{Y}_{J-l} & (5.46) \\ &\propto \text{unvec} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \cos((J-l)\phi) \\ \sin((J-l)\phi) \end{pmatrix} \\ &= \begin{pmatrix} -\sin((J-l)\phi) & -\cos((J-l)\phi) \\ \cos((J-l)\phi) & -\sin((J-l)\phi) \end{pmatrix} \end{aligned}$$

$$\begin{aligned} K_{11, l+J}^{Jl}(\phi) &= \text{unvec } \text{CG}_{lJ, l+J, 1}^+ c_{l+J, 1} \vec{Y}_{l+J} & (5.47) \\ &\propto \text{unvec} \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos((l+J)\phi) \\ \sin((l+J)\phi) \end{pmatrix} \\ &= \begin{pmatrix} \cos((l+J)\phi) & \sin((l+J)\phi) \\ \sin((l+J)\phi) & -\cos((l+J)\phi) \end{pmatrix} \end{aligned}$$

$$\begin{aligned} K_{12, l+J}^{Jl}(\phi) &= \text{unvec } \text{CG}_{lJ, l+J, 1}^+ c_{l+J, 2} \vec{Y}_{l+J} & (5.48) \\ &\propto \text{unvec} \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \cos((l+J)\phi) \\ \sin((l+J)\phi) \end{pmatrix} \\ &= \begin{pmatrix} -\sin((l+J)\phi) & \cos((l+J)\phi) \\ \cos((l+J)\phi) & \sin((l+J)\phi) \end{pmatrix} \end{aligned}$$

- $1 > \mathbf{J} \geq \mathbf{1}$: Here we have the same two-dimensional endomorphism basis from Eq. (5.34) but other Clebsch-Gordan coefficients, now from Eq. (5.37). As $\text{CG}_{l,J,l+J,1}$ occurred already before, we will obtain the same two basis kernels $K_{11,l+J}^{Jl}$ and $K_{12,l+J}^{Jl}$ from them. From $\text{CG}_{l,J,l-J,1}$ we get the other two basis elements:

$$\begin{aligned} K_{11,l-J}^{Jl}(\phi) &= \text{unvec } \text{CG}_{l,J,l-J,1}^+ c_{l-J,1} \vec{Y}_{l-J} & (5.49) \\ &\propto \text{unvec} \begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos((l-J)\phi) \\ \sin((l-J)\phi) \end{pmatrix} \\ &= \begin{pmatrix} \cos((l-J)\phi) & \sin((l-J)\phi) \\ -\sin((l-J)\phi) & \cos((l-J)\phi) \end{pmatrix} \\ &= \begin{pmatrix} \cos((J-l)\phi) & -\sin((J-l)\phi) \\ \sin((J-l)\phi) & \cos((J-l)\phi) \end{pmatrix} \end{aligned}$$

$$\begin{aligned} K_{12,l-J}^{Jl}(\phi) &= \text{unvec } \text{CG}_{l,J,l-J,1}^+ c_{l-J,2} \vec{Y}_{l-J} & (5.50) \\ &\propto \text{unvec} \begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \cos((l-J)\phi) \\ \sin((l-J)\phi) \end{pmatrix} \\ &= \begin{pmatrix} -\sin((l-J)\phi) & \cos((l-J)\phi) \\ -\cos((l-J)\phi) & -\sin((l-J)\phi) \end{pmatrix} \\ &= - \begin{pmatrix} -\sin((J-l)\phi) & -\cos((J-l)\phi) \\ \cos((J-l)\phi) & -\sin((J-l)\phi) \end{pmatrix} \end{aligned}$$

The last equalities bring the basis elements in a different form, emphasizing that $K_{11,l-J}^{Jl} = K_{11,J-l}^{Jl}$ agrees with the previous result, and that $K_{12,l-J}^{Jl} = -K_{12,J-l}^{Jl}$ is the negative of what we had before. Overall, this implies that these kernels span the same basis as in the previous case $J > l \geq 1$, such that they don't need to be distinguished further.

- $1 = \mathbf{J} \geq \mathbf{1}$: If the irrep orders are equal, there are three invariant subspaces in the decomposition of their tensor product, and accordingly three components in the Clebsch-Gordan coefficients in Eq. (5.38). The one corresponding to the sum frequency, $\text{CG}_{l,J,l+J,1}$, is again the same as before, and yields therefore the same two basis kernels $K_{11,l+J}^{Jl}$ and $K_{12,l+J}^{Jl}$. For the other two Clebsch-Gordan coefficients $\text{CG}_{l,J,0,1}$ and $\text{CG}_{l,J,0,2}$ we get two more basis elements:

$$\begin{aligned} K_{110}^{Jl}(\phi) &= \text{unvec } \text{CG}_{l,J01}^+ c_{01} \vec{Y}_0 \propto \text{unvec} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} (1)(1) & (5.51) \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \cos(0\phi) & \sin(0\phi) \\ \sin(0\phi) & -\cos(0\phi) \end{pmatrix} \end{aligned}$$

$$\begin{aligned}
K_{120}^{Jl}(\phi) &= \text{unvec } \text{CG}_{lJ02}^+ c_{01} \vec{Y}_0 \propto \text{unvec} \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix} (1)(1) \\
&= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -\sin(0\phi) & -\cos(0\phi) \\ \cos(0\phi) & -\sin(0\phi) \end{pmatrix}
\end{aligned} \tag{5.52}$$

The last equalities are again written to emphasize that these solutions turn out to be special cases of $K_{11,J-l}^{Jl}$ and $K_{12,J-l}^{Jl}$ for $J = l$.

Solution tables: The derived solutions of the $\text{SO}(2)$ -steerability constraint for irrep fields are summarized in Table 5.2.

Special Orthogonal Group $\text{SO}(2)$

in out	V_0	$V_l, l \geq 1$
	$[1]$	$[\cos(l\phi) \sin(l\phi)], [-\sin(l\phi) \cos(l\phi)]$
$V_J,$ $J \geq 1$	$\begin{bmatrix} \cos(j\phi) \\ \sin(j\phi) \end{bmatrix}, \begin{bmatrix} \cos((j-l)\phi) & -\sin((j-l)\phi) \\ \sin((j-l)\phi) & \cos((j-l)\phi) \end{bmatrix}, \begin{bmatrix} -\sin((j-l)\phi) & -\cos((j-l)\phi) \\ \cos((j-l)\phi) & -\sin((j-l)\phi) \end{bmatrix},$ $\begin{bmatrix} -\sin(j\phi) \\ \cos(j\phi) \end{bmatrix}, \begin{bmatrix} \cos((j+l)\phi) & \sin((j+l)\phi) \\ \sin((j+l)\phi) & -\cos((j+l)\phi) \end{bmatrix}, \begin{bmatrix} -\sin((j+l)\phi) & \cos((j+l)\phi) \\ \cos((j+l)\phi) & \sin((j+l)\phi) \end{bmatrix}$	

Table 5.2: Bases for the angular parts of $\text{SO}(2)$ -steerable kernels satisfying the irrep steerability constraint in Eq. (5.32). The $\text{SO}(2)$ -irreps (ρ_j, V_j) are given by Eq. (5.33). General $\text{SO}(2)$ -steerable kernels on \mathbb{R}^2 follow by adding an unconstrained radial part.

Table 5.3 gives analogous solutions for $\text{O}(2)$ -steerable irrep kernels, derived originally by Weiler and Cesa [322]. In contrast to the $\text{SO}(2)$ case, there is an additional sign-flip irrep of $\text{O}(2)$ and the higher order irreps involve an additional reflection action. Specifically, parameterizing $\text{O}(2)$ as

$$\text{O}(2) = \left\{ \begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & s \end{pmatrix} \mid \phi \in [0, 2\pi), s \in \{\pm 1\} \right\}, \tag{5.53}$$

where the additional parameter s models reflections, the irreps are given by:

$$\rho_0(\phi, s) = 1, \quad \rho_{\text{sign}}(\phi, s) = s \tag{5.54}$$

$$\text{and, for } j \geq 1, \quad \rho_j(\phi, s) \begin{pmatrix} \cos(j\phi) & -\sin(j\phi) \\ \sin(j\phi) & \cos(j\phi) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & s \end{pmatrix}$$

A major difference to $\text{SO}(2)$ is that the endomorphism spaces of $\text{O}(2)$ -irreps are all one-dimensional and consist of scaled identity matrices (homothetys). As a result, all bases involving higher order harmonics are of half the dimensionality as in the case of $\text{SO}(2)$.

5.3.5 Kernel sampling and anti-aliasing:

Feature fields and kernels are in practice commonly sampled on a pixel grid or evaluated at a finite set of sampling points. To prevent aliasing effects, it is necessary to bandlimit

Orthogonal Group $O(2)$			
out \ in	V_0	V_{sign}	$V_l, l \geq 1$
V_0	[1]	\emptyset	$[-\sin(l\phi) \cos(l\phi)]$
V_{sign}	\emptyset	[1]	$[\cos(l\phi) \sin(l\phi)]$
$V_J, J \geq 1$	$\begin{bmatrix} -\sin(J\phi) \\ \cos(J\phi) \end{bmatrix}$	$\begin{bmatrix} \cos(J\phi) \\ \sin(J\phi) \end{bmatrix}$	$\begin{bmatrix} \cos((J-l)\phi) & -\sin((J-l)\phi) \\ \sin((J-l)\phi) & \cos((J-l)\phi) \end{bmatrix}, \begin{bmatrix} \cos((J+l)\phi) & \sin((J+l)\phi) \\ \sin((J+l)\phi) & -\cos((J+l)\phi) \end{bmatrix}$

Table 5.3: Bases for the angular parts of $O(2)$ -steerable kernels satisfying the irrep steerability constraint for $G = O(2)$. The $O(2)$ -irreps are defined in Eq. (5.54). As the endomorphism spaces $\text{End}(V_j)$ for $j \geq 1$ are for $O(2)$ only one-dimensional, the dimensionalities of the kernel spaces such irrep orders are reduced by half in comparison to the $SO(2)$ solutions. General $O(2)$ -steerable kernels on \mathbb{R}^2 follow by adding an unconstrained radial part.

the continuous kernel before sampling it. This is luckily easily possible since our steerable kernel bases in Eq. (5.30) are already defined in terms of a harmonic (Fourier) basis, i.e. functions of a certain frequency.

Specifically for $G = SO(d)$ or $O(d)$, Weiler et al. [324] proposed to use a bandlimiting heuristic with a radially dependent cutoff frequency, allowing for higher order harmonics on orbits of larger radius only. This strategy was empirically shown to reduce the numerical errors in the analytically proven G -equivariance of steerable convolutions. Cesa et al. [40] furthermore discuss an adaptation of the Wigner-Eckart theorem for steerable kernels that allows for solutions that are smooth across orbits.

5.4 Alternative approaches to construct steerable kernels

For completeness, this section mentions alternative approaches to parameterize equivariant convolution kernels or solve the steerability constraint.

Heuristic approaches: While the vast majority of group equivariant CNNs relies implicitly on steerable kernels, most network architectures were proposed heuristically instead of being derived from first principles. For instance, Worrall et al. [335] proposed *harmonic networks*, which apply circular harmonic kernels to map between $SO(2)$ -irrep fields, Thomas et al. [301] proposed *tensor field networks*, which map via spherical harmonics between $SO(3)$ -irrep fields, and Schütt et al. [263] construct their $O(3)$ -invariant *Schnet* by applying isotropic kernels. Similarly, many authors, e.g. [195, 196, 71, 358], observed that *applying G -transformed copies of kernels* to a signal results in an equivariant response – which corresponds to group convolutions or convolutions between regular feature fields.

The difference to steerable CNNs is that these approaches do not derive the kernels from a symmetry constraint, but merely observe that the proposed construction yields equivariant responses. They are in particular not able to prove a notion of completeness of their

kernel space – in fact, it happens commonly that the authors parameterize only a subspace of steerable kernels that would be consisting with the feature spaces’ transformation laws [8, 232, 339]. Steerable CNNs allow furthermore to classify the zoo of equivariant models in terms of feature field types.

Group convolutions: Many equivariant convolutional networks apply *group convolutions*, which were in Section 4.5 shown to be equivalent to steerable convolutions between regular feature fields. Convolution kernels (and feature fields) are in this setting viewed as *unconstrained functions on the group*, e.g. $h : \text{Aff}(G) \rightarrow \mathbb{R}$. While this seems easier than implementing *regular representation constrained steerable kernels*, the complexity is here hidden in the definition of the group convolution operation itself, i.e. in Eq. (4.64).

We note that harmonic basis functions were prior to their use in steerable CNNs already utilized to parameterize group convolution kernels [324].

Expansion in harmonic basis functions: The G -steerability constraint in its current form was first derived by Weiler et al. [323], who considered $G = \text{SO}(3)$. The authors observed that the constraint may be restricted to spherical shells with an unconstrained radial part, which corresponds to the *restriction to G -orbits* from Section 5.3.1. Considering irrep fields, they found a spherical harmonics basis after applying a Clebsch-Gordan decomposition as in the Wigner-Eckart theorem for steerable kernels. Weiler and Cesa [322][38] extended this approach to arbitrary field types by introducing the *irrep decomposition of kernels* that was described in Section 5.27. Focusing on subgroups $G \leq \text{O}(2)$, the kernels were expanded in terms of circular harmonics, Fig. 5.2, and the authors found the solution Tables 5.2 and 5.3, as well as their analogs for cyclic and dihedral groups. Lang and Weiler [173] formalized these insights in terms of the *Wigner-Eckart theorem for steerable kernels*. This theorem was implemented by Cesa et al. [40][39], who adapted the theory such that the individual solutions on G -orbits vary smoothly across orbits.

Lie algebra representations: de Haan et al. [67] observed that it is already sufficient to satisfy the steerability constraint for the generators of the group. They reformulate the constraint accordingly in terms of *Lie algebra representations* and solve it symbolically with the help of a computer algebra system, again finding selection rules on harmonic subspaces.

Steering from orbit representatives: Since the G -steerability constraint relates kernel values on each G -orbits, the full kernel can be reconstructed by steering values from some choice of *orbit representatives* (Def. B.3.5) [56]. Specifically, let $G \backslash \mathbb{R}^d$ be the quotient space (Def. B.3.4) of \mathbb{R}^d by the left G -action, consisting of all possible G -orbits, let $r : G \backslash \mathbb{R}^d \rightarrow \mathbb{R}^d$ be a choice of orbit representatives and let $g_{r,x} \in G$ be any group element satisfying $g_{r,x}r(G.x) = x$. Then, any G -steerable kernel $K : \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ satisfies for any $x \in \mathbb{R}^d$

$$K(x) = K(g_{r,x}r(G.x)) = \frac{1}{|\det g_{r,x}|} \rho_{\text{in}}(g_{r,x}) K(r(G.x)) \rho_{\text{out}}(g_{r,x})^{-1}, \quad (5.55)$$

implying that it is possible to reconstruct the full kernel from its restriction

$$K|_{r(G \backslash \mathbb{R}^d)} : r(G \backslash \mathbb{R}^d) \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \quad (5.56)$$

to orbit representatives by steering it via $g_{r,x}$. The restricted kernel is still required to satisfy the reduced steerability constraint

$$K|_{r(G \backslash \mathbb{R}^d)}(\tau) = \frac{1}{|\det h|} \rho_{\text{in}}(h) K|_{r(G \backslash \mathbb{R}^d)}(\tau) \rho_{\text{out}}(h)^{-1} \quad (5.57)$$

for any $\tau \in r(G \backslash \mathbb{R}^d)$ and any $h \in \text{Stab}_\tau$, where $\text{Stab}_\tau \leq G$ is the stabilizer subgroup (Def. B.3.6) for τ .³

We want to mention that this approach is problematic if the kernel is to be sampled, since knowledge about the specific harmonics (frequency components) is required to prevent aliasing effects [324, 323, 322].

Numerical solutions: Assume that the structure group G is finite, and that the kernel is sampled at a finite number of points, the set of which is invariant under the G -action. The kernel constraint becomes then a (finite) system of linear equations, which can be solved numerically. Cohen and Welling [53] used this approach to solve for D_4 (dihedral group) or C_4 (cyclic group) steerable kernels on a square grid of $s \times s$ pixels.

MLP parametrizations: As an alternative to solving the steerability constraint, steerable kernels may be parameterized implicitly by means of G -equivariant MLPs (fully connected networks), mapping from \mathbb{R}^d to $\mathbb{R}^{c_{\text{out}} \cdot c_{\text{in}}} \cong \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ and satisfying the kernel constraint by construction. Finzi et al. [91] used a similar approach to parameterize regular group convolutional kernels.⁴ A thorough investigation of (non-steerable) MLP-parameterized kernels is found in [250].

³This constraint makes the construction in particular independent from the specific choice of $g_{r,x}$.

⁴Since MLPs map between vector spaces, Finzi et al. [91] would parameterize a group convolutional kernel $\kappa : \text{Aff}(G) \rightarrow \mathbb{R}$ in terms of a kernel $\kappa_{\text{Lie}} : \mathfrak{aff}(G) \rightarrow \mathbb{R}$ on the group's Lie algebra, associating the two via the Lie exponential and logarithmic map. Our formulation does not require this extra step since steerable kernels $K : \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ already map between vector spaces.

Empirical evaluation of steerable CNNs

This chapter investigates the properties and design choices of steerable CNNs empirically, thereby identifying general trends and caveats when working with equivariant models. In a nutshell, equivariant models are characterized by an enhanced data efficiency and convergence rate, leading to a significantly improved performance in comparison to their non-equivariant counterparts. Some care has to be taken when designing the model architecture, since for instance too much invariance or an unsuitable choice of field types or steerable nonlinearities may crucially impair the results. The following list gives a high level overview of our findings.

Generalization: Section 6.1 presents experiments that measure the *generalization of steerable CNNs over group orbits* by training on data in a fixed representative pose and testing over all other poses on the orbit. In theory, equivariant CNNs are guaranteed to generalize perfectly over orbits, i.e. achieve the same result for any transformed pose of a feature field. This guarantee holds indeed if the numerical discretization of the continuous model is invariant under the symmetry group action, but may be slightly broken if this is not the case (e.g. for continuous rotations on a pixel grid). Numerically broken equivariance is easily restored via data augmentation over the orbit. While conventional CNNs can also learn to be equivariant when being trained with augmentation, this requires additional learning capacity from them, such that they perform significantly worse.

Data efficiency and convergence: That equivariant models generalize their inference over group orbits leads to an improved data efficiency and convergence in comparison to non-equivariant models. Section 6.2 demonstrates this claim empirically by training models with different levels of equivariance on varying dataset sizes and observing that equivariant models on small datasets may achieve better results than non-equivariant models on significantly larger datasets. This is of particular importance in settings where data is scarce or expensive to collect, as is for instance the case with some medical imaging tasks.

Choice of symmetry group and group restriction: The performance of an equivariant model depends heavily on its equivariance group and the level of symmetries present in the data: too little equivariance does not make use of all available prior knowledge, while too much equivariance overconstrains the model, forcing it to generalize over transformations that are not respected by the ground truth itself. Section 6.3 presents experiments which vary the levels of symmetries in the data and models, and investigates the interplay of the two. We experiment additionally with group restricted

models, whose overall equivariance is adapted to the *global* symmetries of the data, but whose initial layers have a larger equivariance group, which allows them to exploit *local* symmetries. These models are consistently achieving the best results.

Drop-in replacement for conventional convolutions: As many contemporary signal processing pipelines are based on convolutional networks, practitioners may be interested in how far these pipelines are compatible with steerable convolutions. Our experiments in Section 6.4 take established image classifiers, replace their conventional with steerable convolutions, and train them subsequently with the original training procedure and choice of hyperparameters. The equivariant models significantly outperform their non-equivariant baselines, which suggests that steerable convolutions can be readily used as drop-in replacements of conventional convolutions.

Natural image datasets: The above mentioned experiments from Section 6.4 are performed on natural image datasets, which are characterized by a global rotational alignment due to a preferred gravity direction. Their results show therefore additionally that steerable convolutions yield substantial gains even if only local rotational symmetries are present in the data.

Field type and nonlinearity benchmarking: Section 6.5 presents a benchmarking study of the design choices of steerable CNNs, covering different symmetry groups $G \leq O(2)$, G -representations as field types, G -equivariant nonlinearities and G -invariant maps for a pose-independent classification. Table 6.6 summarizes the results on three transformed MNIST datasets with different levels of symmetries. Overall, we find that group or quotient space convolutions, corresponding to regular or quotient representations of finite subgroups of $O(2)$, achieve the best performance. An alternative are models whose field types are $O(2)$ or $SO(2)$ -irreps. These models are continuously rotation equivariant and have lower-dimensional feature vectors, however, their test errors are consistently higher than those of the regular or quotient representation based models. The best results among these models are achieved by learned gated nonlinearities [323].

Most of the results presented in this chapter were originally published by Weiler and Cesa [322], while the generalization experiments in Section 6.1 are inspired by Weiler et al. [324]. Chapter 10 presents additional empirical results for the analog of Euclidean reflection steerable CNNs on a Möbius strip, investigating in particular the performance of the field types from Fig. 5.1. Before coming to the actual experiments, the next two paragraphs describe the datasets and models that are used.

Datasets: All of our experiments are run on image datasets, i.e. signals on \mathbb{R}^2 , which is a practically relevant setting and is in contrast to higher dimensionalities computationally manageable. We focus on the supervised classification setting, which is computationally cheaper and easier to optimize than other tasks. The insights should qualitatively generalize to signals on higher dimensional spaces, to point cloud instead of pixel grid discretizations, and other learning tasks beyond classification.¹ A list of papers running experiments on Euclidean spaces of other dimensionalities and using other signal discretizations is found in Table 14.1. Literature using steerable CNNs for other learning tasks, like generative modelling, reinforcement learning or tracking was listed in the paragraph “Applications & literature review” in the introductory Chapter 1.

More specifically, we are running many experiments on G -transformed variants of the MNIST dataset, where G is either of the trivial group $\{e\}$, the reflection group \mathcal{R} , the rotation

¹For instance, fully convolutional image segmentation is equivalent to a sliding window pixel classification [189].

layer	output features
network input	1 scalar field
conv block (7×7 , pad 1)	16 regular fields
conv block (5×5 , pad 2)	24 regular fields
max pooling (2×2)	24 regular fields
conv block (5×5 , pad 2)	32 regular fields
conv block (5×5 , pad 2)	32 regular fields
max pooling (2×2)	32 regular fields
conv block (5×5 , pad 2)	48 regular fields
conv block (5×5 , pad 0)	64 regular fields
G -invariant projection	64 scalar fields
\mathbb{R}^2 average pooling	64 scalars
fully connected	64 scalars
fully connected + softmax	10 scalars

Table 6.1: Basic model architecture from which all models for the G -MNIST experiments in Sections 6.1, 6.2, 6.3 and 6.5. are derived. Each convolution block consists of a convolution layer, batch-normalization and a nonlinearity. The feature field types are in the first three sections regular representations of different discrete subgroups of $O(2)$, while the hyperparameter benchmark experiment in Section 6.5 uses the representations listed in Table 6.6. The first fully connected layer is followed by batch-normalization and ELU.

group $SO(2)$ or the orthogonal group $O(2)$. MNIST has the advantage that the level of symmetries in the dataset is well controllable since most digits appear in an upright rotation and have a preferred chirality (one would not use rotations or reflections for data augmentation in the original MNIST dataset). This is in contrast to e.g. natural images, whose statistics are usually reflection invariant. Following the construction of the official rotated MNIST dataset, we split the G -MNIST datasets into a training set consisting of 12000 images and test on the remaining 50000 images.

The experiments in Section 6.4 are conducted on CIFAR-10, CIFAR-100 and STL-10, all of which are datasets of natural images. They show a larger intra-class variability than MNIST, have, in the case of CIFAR-100 more classes, and, in the case of STL-10 a higher spatial resolution of 96×96 pixels.

Models: All of the experiments on the G -MNIST datasets rely on models that are variations of the architecture in Table 6.1. This baseline model is *group convolutional*, with feature vectors in \mathbb{R}^{16} that transform according to the *regular* C_{16} -representation; see Def. B.5.18 and Remark B.5.19. Regular representations act by permuting the $|G|$ feature vector entries according to the group's binary operation, as exemplified for C_4 in Table. B.1. Variations of this model that use other field types have their multiplicities of feature fields scaled such that the total number of model parameters is approximately held constant – depending on the parameter efficiency of the model, this leads to a different total number of channels. All experiments in Sections 6.1 - 6.4 stick with regular representations of different cyclic and dihedral groups C_N and D_N ,² while the variants in Section 6.5 adapt the field types as reported in Table 6.6. All permutation representation based networks, i.e. all regular and quotient representation based models, apply element-wise ELU nonlinearities [50], while the irrep based models apply different types of norm-nonlinearities or gated nonlinearities; see Section 4.3.3. Aff(G)-invariant models apply some G -invariant projection operation to scalar fields after the last convolution, which is followed by spatial average pooling, to produce invariant features for classification.

The models for CIFAR-10, CIFAR-100 and STL-10 in Section 6.4 are based on WideRes-Nets [344], whose conventional convolutions are replaced with regular representation steerable convolutions. The numbers of feature fields are thereby either adapted to match the

²The dihedral groups D_N comprise N rotations in two reflections each, implying that $D_1 = \mathcal{R}$ is the reflection group.

	{e}-MNIST train set		\mathcal{R} -MNIST train set	
	vanilla CNN	\mathcal{R} -steerable	vanilla CNN	\mathcal{R} -steerable
original test set	0.87 ± 0.05	0.98 ± 0.09	1.61 ± 0.08	0.96 ± 0.05
reflected test set	61.49 ± 0.91	0.98 ± 0.09	1.65 ± 0.12	0.96 ± 0.05

Table 6.2: Test errors of conventional and reflection steerable CNNs on differently reflected training and testing datasets. $\{e\}$ -MNIST refers to a training dataset with digits in their original orientation, while \mathcal{R} -MNIST applies (random) reflectional data augmentation during training. The “original test set” is again in the original orientation, while all digits in the reflected test set are (non-randomly) reflected. As expected, vanilla CNNs do not generalize over reflections, while \mathcal{R} -steerable CNNs do. When being trained on reflection augmented digits, vanilla CNNs achieve a worse performance than equivariant models.

number of parameters of the original model (resulting in more channels), or to match the original models’ number of channels (resulting in fewer parameters).

All models are implemented using the PyTorch library `e2cnn` [38] or its successor `escnn` [39].

6.1 Generalization over group orbits

The central idea of equivariant CNNs is that they generalize whatever they learn over their group orbits; see Fig. 4.2. Here we investigate the generalization capabilities of steerable CNNs by training a classifier network on a dataset of images in a fixed pose, but testing it on datasets that contain images in another, transformed pose on the group orbit, measuring how the classification error depends on the particular dataset transformation. As expected, steerable CNNs are found to generalize over group orbits, while non-equivariant CNNs do not. Non-equivariant CNNs that are trained with data augmentation are (more or less) generalizing over the augmentation orbit as well, however, as they explicitly need to learn the mapping for each pose, their performance is significantly worse.

We experiment in the following with reflection and rotation symmetries on square pixel grids. The qualitative difference between these two cases is that reflections are exact symmetries of the pixel grid, while the rotational symmetries are broken by this discretization. We disregard the translational generalization of steerable CNNs, since it is equivalent to that of conventional CNNs and was already investigated by Azulay and Weiss [6] – pooling layers are found to break translation equivariance, which can be alleviated via the approaches suggested in [348, 336]. An animation of the generalization of a rotation equivariant model is found at <https://github.com/QUVA-Lab/e2cnn#demo>.

Exact reflection symmetries: To measure the generalization of vanilla CNNs and steerable CNNs over reflections, we train and test them on differently transformed MNIST variants as summarized in Table 6.2.

Training on $\{e\}$ -MNIST (left two columns) means that all digits in the training set are presented in their original orientation. When testing such trained models on non-reflected digits (“original test set”), vanilla CNNs have a slight advantage over reflection invariant CNNs,

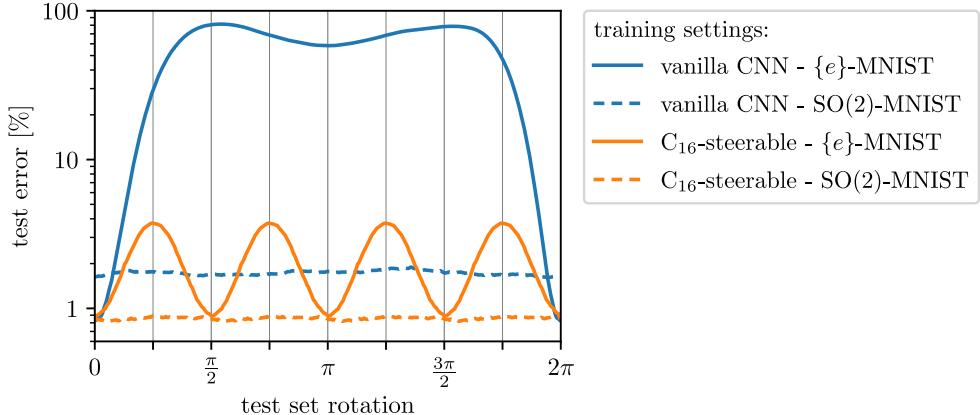


Figure 6.1: Rotational generalization of a conventional CNN (blue) and a rotation steerable CNN (orange) on a square pixel grid. The models are either trained on digits that are aligned upright ($\{e\}$ -MNIST, solid lines) or augmented with random rotations (SO(2)-MNIST, dashed lines). As expected, the vanilla CNN does not generalize over rotations. The theoretically predicted C_{16} -equivariance of the steerable CNN is in practice broken to C_4 -equivariance since the pixel grid is only invariant under rotations by multiples of $\frac{\pi}{2}$. While the test error for $\frac{\pi}{4}$ is far from random chance (mind the logarithmic y -axis), the rotational equivariance of the model can be improved further by using rotational data augmentation. The error rates of both models are approximately rotation independent when being trained on rotation augmented digits, but the error rate of the equivariant model is approximately half of that of the vanilla CNN.

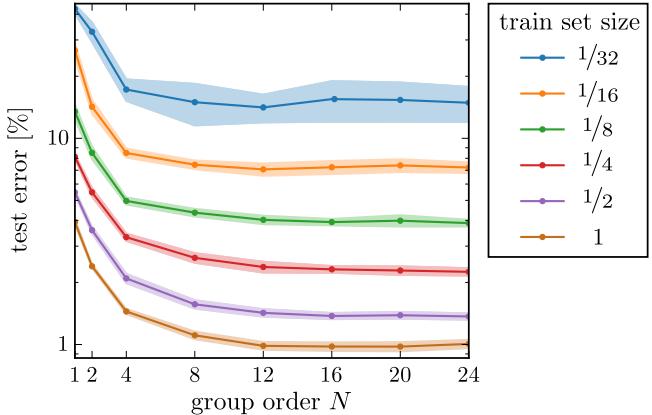
since the latter are unable to leverage any information about the digit orientation.³ However, when being tested on reflected digits (“reflected test set”), the vanilla CNN’s performance degrades significantly, since it never saw such samples during training. In contrast, the reflection steerable CNN generalizes perfectly to the unseen digit orientations, i.e. it achieves exactly the same test error and standard deviation on both test sets.

If the learning task is a-priori known to be reflection equivariant, the usual approach is to train the conventional CNN with *reflectional data augmentation*, denoted here as \mathcal{R} -MNIST. Augmentation leads to an approximately similar performance of the conventional CNN on both test set orientations. However, due to the greater intra-class variability of this classification task, the performance is quite a bit worse as compared to training on non-reflected digits. Reflection steerable CNNs are by design equivariant (here invariant), and therefore not affected at all by the reflectional augmentation. The test error of the reflection steerable CNNs is lower than that of augmentation trained vanilla CNNs, since they do not explicitly need to learn to be equivariant.

Numerically broken rotation symmetries: The exact analytical equivariance and generalization of steerable CNNs may be broken by the numerical implementation. An example of great practical relevance are rotation equivariant CNNs on square pixel grids. We investigate the networks’ generalization as before by training either on digits that are aligned upright ($\{e\}$ -MNIST) or randomly rotated (SO(2)-MNIST) and testing on a set of datasets that contain digits which are all rotated by the *same angle*. A plot of the resulting test errors for different test set rotation angles is shown in Fig. 6.1.

³This issue can be alleviated via group restriction operations, which help (locally) reflection steerable CNNs to achieve even better results than vanilla CNNs. Section 6.3 proves this claim empirically.

Figure 6.2: Dependence of the SO(2)-MNIST test error on the rotation group order N of regular C_N -steerable CNNs and on the (relative) training set size. Especially for low group orders, increasing N may result in larger gains than increasing the training set size by a factor of four, which demonstrates the improved data efficiency of equivariant CNNs.



When training on $\{e\}$ -MNIST, the vanilla CNN achieves a good performance on the non-rotated test set, but the performance degrades quickly to random chance for larger test set rotations. The small dip in the error rate for rotations around 180 degrees comes from digits like 0 and 8, which are approximately invariant under such transformations. The rotation steerable CNN is in theory expected to generalize its inference over test set rotations. As seen from the plot, this expectation is in our implementation violated, and the test error is $\frac{\pi}{2}$ -periodic, corresponding to the subgroup of rotational symmetries of the square pixel grid. While the generalization is with $\approx 3.7\%$ test error in the worst case already much better than random chance, there is still room for improvement.

The dashed lines correspond to models that were trained using augmentation with randomly rotated digits. Both curves are approximately independent from the test set angle, but the rotation steerable model has a lower test error. Rotational equivariance is therefore clearly beneficial, even though it should be combined with data augmentation if it is broken by the numerical discretization.

6.2 Data efficiency and convergence

The equivariance of steerable CNNs implies that they don't need to learn to process any transformed version of a feature field individually, but automatically generalize over group orbits. As a consequence, they *require less data to achieve better results* and *converge faster* than non-equivariant models.

Fig. 6.2 records the test errors of regular C_N -steerable CNNs on SO(2)-MNIST for varying rotation orders N and training set sizes. The test error decreases initially with growing N before it saturates at around $N = 12$.⁴ The claim that equivariant CNNs exhibit an improved data efficiency is apparent by the fact that an increase of the equivariance group order allows to reduce the test error further than a doubling or even quadrupling of the training set size.

The improvements in data efficiency are even more significant in higher dimensions [329, 8], since higher dimensional rotation groups consist of “more elements” and thus larger orbits to

⁴The exact order of saturation correlates with the chosen kernel size, which was here fixed to 5×5 pixels.

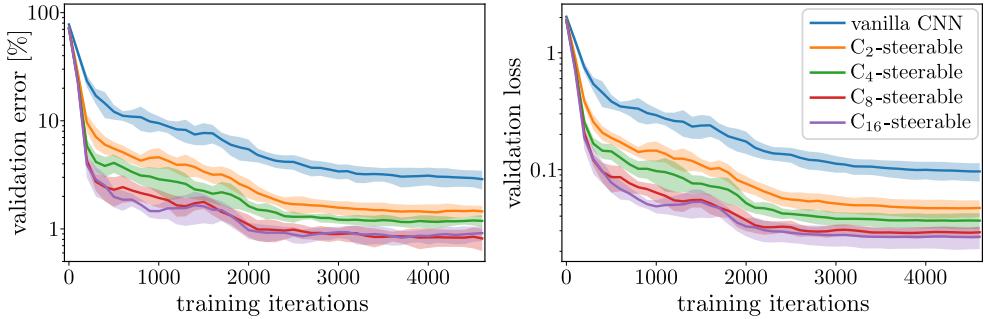


Figure 6.3: Validation errors and losses during the training of a conventional CNN and C_N -equivariant models on SO(2)-MNIST. Networks with higher levels of equivariance converge faster and achieve better final results.

generalize over. For instance, Winkels and Cohen [329] (Table 2) showed that incorporating equivariance w.r.t. cubic symmetries (24 rotations in two reflections each, i.e. 48 elements) may lead to larger gains in a detection task than a *tenfold* increase of the dataset size.

Fig. 6.5 shows a plot that is similar to Fig. 6.2, with the difference that the dataset size is here plotted on the x -axis and the experiment is performed on the STL-10 dataset of natural images. That the curves are straight lines in the log-log plot reflects the well-known fact that test errors depend on the data set size via a *power law* [120].⁵ More data efficient models result usually in shifted lines with the same slope, i.e. same power law exponent. Batzner et al. [8] noticed that these exponents, and therefore data efficiencies, are for locally rotation invariant models (scalar fields) lower than for locally faithfully rotation equivariant models (non-scalar fields). Observing that our equivariant and non-equivariant models all show the same slope suggests that this is an issue specific to locally invariant models. This issue with purely scalar field based models is supported by the fact that such models perform overall worst in our model benchmark in Table 6.6 (line 45, labeled as “irreps = 0”)

The evolution of validation errors and losses of C_N -steerable CNNs during training on SO(2)-MNIST is visualized in Fig. 6.3. One can see that higher rotation orders N do not only lead to lower final losses and errors, but also to a faster model convergence. The reason for this effect is again the models’ generalization over group orbits: conventional CNNs need to learn all transformed versions of an input image explicitly, which requires either more iterations or a larger batch size in comparison to equivariant models. Equivariant CNNs may therefore not only be more accurate than non-equivariant models, but also cheaper to train.

6.3 Choice of equivariance group and group restriction

The optimal choice of symmetry group w.r.t. which a model is equivariant depends on the level of symmetry present in the learning task. Assuming that the data distribution is invariant under $\text{Aff}(G_{\text{data}})$, and the model (globally) is $\text{Aff}(G_{\text{model}})$ -equivariant, we can distinguish three qualitatively different cases:

⁵The same effect manifests in Fig. 6.2 in the fact that the curves are equidistant on the (log scale) y -axis.

$G_{\text{model}} < G_{\text{data}}$: In this case the model does not fully exploit the symmetries present in the data and is forced to learn to generalize over G_{data} -orbits explicitly. As seen in the previous sections, this results in a suboptimal performance in comparison to equivariant models with $G_{\text{model}} = G_{\text{data}}$. A common example would be to use a conventional CNN, i.e. $G_{\text{model}} = \{e\}$, on satellite imaging data, which is often isometry invariant, i.e. $G_{\text{data}} = O(2)$. Other examples are the conventional CNNs on \mathcal{R} -MNIST in Table 6.2 and on $\text{SO}(2)$ -MNIST in Fig. 6.1.

$G_{\text{model}} > G_{\text{data}}$: If the model assumes more symmetries than actually present in the data, it generalizes over too large orbits and may draw wrong conclusions. For instance, the digits of the original MNIST dataset appear all in an upright rotation, i.e. $G_{\text{data}} = \{e\}$. When using a rotation invariant model, $G_{\text{model}} = \text{SO}(2)$, it confuses digits like 6 and 9, which are related by rotations. Similarly, the digits $\text{\texttt{4}}$ and $\text{\texttt{7}}$ (4 and 7) are related by a reflection and a rotation by $\pi/2$, and might therefore be confused by a $G_{\text{model}} = O(2)$ -invariant model. Another example is the \mathcal{R} -steerable CNN on $\{e\}$ -MNIST in Table 6.2, which performs worse than the vanilla CNN.

$G_{\text{model}} = G_{\text{data}}$: Having the symmetries of the data and model matching is obviously the best choice. This way we are neither giving away prior knowledge, nor overconstraining the model with too much equivariance. The examples in Table 6.2 and Fig. 6.1 are the vanilla CNN on $\{e\}$ -MNIST, the \mathcal{R} -steerable model on \mathcal{R} -MNIST and the rotation steerable model on $\text{SO}(2)$ -MNIST.

The above analysis considered the *global symmetries of the data* and the *overall equivariance of the model*. However, as described in Section 4.4, the statistics of local patterns in the data are often invariant under a larger symmetry group $G_{\text{data,local}} > G_{\text{data,global}}$ than that present at the global scale. Each convolution layer is usually accumulating features from a local receptive field, and its equivariance should be adapted to the symmetries present at that scale. This can be achieved with the *group restriction operation* $\text{Res}_{\text{Aff}(G_{\text{data,global}})}^{\text{Aff}(G_{\text{data,local}})}$ from Section 4.4, which we investigate here empirically. Our finding is that a higher level of equivariance up to the final network layer generally seem to be helpful – this way the model generalizes most of its inference over large $\text{Aff}(G_{\text{data,local}})$ -orbits, but can ultimately discriminate between features on different $\text{Aff}(G_{\text{data,global}})$ -orbits. We abbreviate the restriction of a model’s structure group G to a subgroup H after layer l in the following by $G|_l H$.

G-MNIST: To investigate the interplay of equivariance groups and dataset symmetries, we run fully D_N , C_N and $\{e\}$ -steerable models and group restricted variants of them on $O(2)$ -MNIST, $\text{SO}(2)$ -MNIST and $\{e\}$ -MNIST. All models are G_{model} -*invariant* classifiers, that is, they start confusing classes whenever $G_{\text{model}} > G_{\text{data,global}}$. Fig. 6.4 and rows 2-10 and 19-27 in Table 6.6 summarize the results of these experiments.

For $O(2)$ -MNIST, all models are in the regime $G_{\text{model}} \leq G_{\text{data,global}}$ and models with larger equivariance groups perform better than those with less. In particular, the D_N models perform consistently better than the C_N models of the same order N since they generalize additionally over the reflections that are present in the dataset.

On $\text{SO}(2)$ -MNIST, the C_N models and the conventional CNN are still in the regime $G_{\text{model}} \leq G_{\text{data,global}}$ and the test error decreases with growing rotation order N . The performance of these models improved relative to the corresponding results on $O(2)$ -MNIST since the intra-class variability in $\text{SO}(2)$ -MNIST is reduced. In contrast, the D_N models are harmed by their global reflection invariance – in fact, they are not able to distinguish between $O(2)$ -MNIST and $\text{SO}(2)$ -MNIST at all, thus achieving exactly the same result on both datasets. For $N = 1$ the dihedral model is purely reflection- but not rotation invariant, and therefore performs

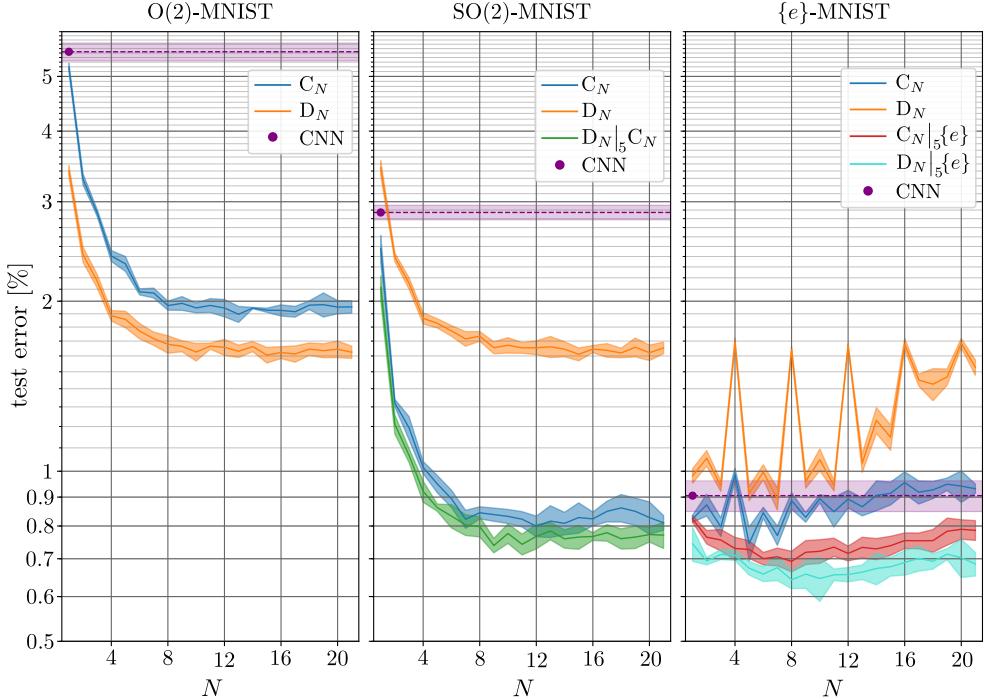


Figure 6.4: Test errors of D_N and C_N steerable and conventional CNNs on different G -MNIST variants. *Left:* All equivariant models improve upon the non-equivariant CNN baseline on $O(2)$ -MNIST. The error decreases before saturating at around 8 to 12 orientations. Since the dataset contains reflected digits, the D_N -equivariant models perform better than their C_N counterparts. *Middle:* Since the intra-class variability of $SO(2)$ -MNIST is reduced, the performances of the C_N model and the baseline CNN improve on this dataset. In contrast, the D_N models are invariant to global reflections such that they can't distinguish between the $O(2)$ -MNIST and $SO(2)$ -MNIST datasets. For $N = 1$ this leads to a worse performance than that of the CNN baseline. Restricted dihedral models, denoted by $D_N|_5 C_N$, make use of the local reflectional symmetries but are globally only rotation invariant. This makes them perform even better than the C_N models. *Right:* On $\{e\}$ -MNIST the globally invariant models C_N and D_N don't yield better results than the baseline, however, the restricted (i.e. non-invariant) models $C_N|_5 \{e\}$ and $D_N|_5 \{e\}$ do. For more details see the main text.

even worse than the CNN baseline. This issue is resolved by *restricting* the dihedral models after the penultimate convolution (at layer 5) to $C_N \leq D_N$, such that the group pooling after the final convolution results in *globally* C_N -invariant features. This model, denoted in the figure by $D_N|_5 C_N$, achieves a slightly better accuracy than the pure C_N -equivariant model since it can leverage *local* reflectional symmetries.

For $\{e\}$ -MNIST, the non-restricted D_N models perform again worse than the C_N models since they are insensitive to the chirality of the digits. In order to explain the non-monotonic trend of the curves of the C_N and D_N models, notice that some of the digits are approximately related by symmetry transformations⁶. If these transformations happen to be part of the equivariance group w.r.t. which the model is invariant the predictions are more likely to be confused. This is mostly the case for N being a multiple of 2 or 4 or for large orders N .

⁶E.g. 6 and 9 (6 and 9) or 2 and 3 (2 and 5) are related by a rotations by π and might therefore be confused by all models C_{2k} and D_{2k} for $k \in \mathbb{N}$. Similarly, 4 and 7 (4 and 7) are related by a reflection and a rotation by $\pi/2$ and might be confused by all models D_{4k} .

restriction depth	SO(2)-MNIST			{e}-MNIST		
	group	test error (%)	group	test error (%)	group	test error (%)
(0)	C ₁₆	0.82 ± 0.02	{e}	0.82 ± 0.01	{e}	0.82 ± 0.01
1	D ₁₆ ₁ C ₁₆	0.86 ± 0.05	D ₁₆ ₁ {e}	0.79 ± 0.03	C ₁₆ ₁ {e}	0.80 ± 0.03
2	D ₁₆ ₂ C ₁₆	0.82 ± 0.03	D ₁₆ ₂ {e}	0.74 ± 0.03	C ₁₆ ₂ {e}	0.77 ± 0.03
3	D ₁₆ ₃ C ₁₆	0.77 ± 0.03	D ₁₆ ₃ {e}	0.73 ± 0.03	C ₁₆ ₃ {e}	0.76 ± 0.03
4	D ₁₆ ₄ C ₁₆	0.79 ± 0.03	D ₁₆ ₄ {e}	0.72 ± 0.02	C ₁₆ ₄ {e}	0.77 ± 0.03
5	D ₁₆ ₅ C ₁₆	0.78 ± 0.04	D ₁₆ ₅ {e}	0.68 ± 0.04	C ₁₆ ₅ {e}	0.75 ± 0.02
no restriction	D ₁₆	1.65 ± 0.02	D ₁₆	1.68 ± 0.04	C ₁₆	0.95 ± 0.04

Table 6.3: Effect of the group restriction operation at different depths of the network on SO(2)-MNIST and {e}-MNIST. Before restriction, the models are equivariant to a larger symmetry group than the group of global symmetries of the corresponding dataset. A restriction at later layers leads to an improved accuracy. All restricted models perform better than non-restricted, and hence globally invariant, models.

Once again, the restricted models, here $D_N|_5\{e\}$ and $C_N|_5\{e\}$, show the best results since they exploit local symmetries but preserve information on the global pose of the digits. Since the restricted dihedral model generalizes additionally over local reflections, its performance is consistently better than that of the restricted cyclic model.

Restriction depth: The results in Fig. 6.4 demonstrate the benefit of using group restriction operations right before the final convolutional layer of the network. Table 6.3 extends on these experiments by varying the depth at which restriction is performed. The overall trend is that a restriction at later stages of the model improves the performance and that restricted models perform significantly better than the invariant models.

6.4 Drop-in replacement for conventional convolutions and natural image datasets

Lots of effort has been made to design conventional CNN architectures and to tune their hyperparameters, which raises the question in how far these findings transfer to steerable CNNs. Here we explore whether steerable convolutions can be used as a *drop-in replacement* for conventional convolutions *without any further adaptations or hyperparameter tuning*. To this end, we take widely established WideResNet architectures [344], upgrade them to be G -steerable, and optimize them according to the original training protocols and hyperparameter settings reported in the literature. The resulting steerable WideResNets outperform their baselines by a large margin, showing that steerable convolutions can be readily deployed for various signal processing tasks. All experiments in this section are conducted on *natural image datasets*⁷, thus demonstrating in addition that steerable CNNs are beneficial for this particular image modality.

CIFAR: A first set of experiments, conducted on CIFAR-10 and CIFAR-100 and reported in Table 6.4, replicates the WideResNet (wrn) architecture, training procedure and hyperparameters from [344] with steerable convolutions. The labels $G_1G_2G_3$ signify hereby

⁷The statistics of natural images are typically invariant under global translations and reflections, i.e. $\text{Aff}(D_1)$, but not under global rotations.

levels of equivariance in the three main blocks of the networks, which are separated by pooling layers. Regular representations are used throughout the whole model except for the last convolution which maps to a scalar field to produce invariant predictions. For a fair comparison we scale the width of all layers such that the number of parameters of the original wrn28/10 model is approximately preserved. Note that, due to their enhanced parameter efficiency, our models become wider than conventional CNNs. Since this implies a higher computational cost, we add an equivariant model, marked by an additional *, which has about the same number of channels as the non-equivariant wrn28/10. More details on the training procedure and hyperparameters are found in [322].

The results of the $D_1 D_1 D_1$ model in Table 6.4 confirm that incorporating the global symmetries of the data already yields a significant boost in accuracy. Interestingly, the $C_8 C_4 C_1$ model, which is purely rotation but not reflection-equivariant, achieves better results, which shows that it is worthwhile to leverage local rotational symmetries. Both symmetries are respected simultaneously by the wrn28/10 $D_8 D_4 D_1$ model. While this model performs better than the two previous ones on CIFAR-10, it surprisingly yields slightly worse result on CIFAR-100. This might be due to the higher dimensionality of its feature fields which, despite the model having more channels in total, leads to less independent fields. The best results (without using auto augment) are obtained by the $D_8 D_4 D_4$ model which suggests that rotational symmetries are useful even on a larger scale. The small wrn28/10* $D_8 D_4 D_1$ model shows a remarkable gain compared to the non-equivariant wrn28/10 baseline *despite not being computationally more expensive*. To investigate whether equivariance is useful even when a powerful data augmentation policy is available, we further rerun both $D_8 D_4 D_1$ models with *AutoAugment* (AA) [64]. As without AA, both the computationally cheap wrn28/10* model and the wider wrn28/10 version outperform the wrn28/10 baseline by a large margin.

STL-10: In order to test whether the previous results generalize to natural images of higher resolution we run additional experiments on STL-10 [51]. While this dataset was originally intended for semi-supervised learning tasks, its 5000 training images are also being used for supervised classification in the low data regime [69]. We adapt the experiments in [69] by replacing the non-equivariant convolutions of their wrn16/8 model, which was the previous supervised SOTA, with D_N -steerable convolutions. As in the CIFAR experiments, all intermediate features transform according to regular representations. A final, invariant prediction is generated via a convolution to scalar fields. We are again using steerable convolutions as a mere drop-in replacement, that is, we use the same training setting and hyperparameters as in the

model		CIFAR-10	CIFAR-100
wrn28/10	[344]	3.87	18.80
wrn28/10	$D_1 D_1 D_1$	3.36 ± 0.08	17.97 ± 0.11
wrn28/10*	$D_8 D_4 D_1$	3.28 ± 0.10	17.42 ± 0.33
wrn28/10	$C_8 C_4 C_1$	3.20 ± 0.04	16.47 ± 0.22
wrn28/10	$D_8 D_4 D_1$	3.13 ± 0.17	16.76 ± 0.40
wrn28/10	$D_8 D_4 D_4$	2.91 ± 0.13	16.22 ± 0.31
wrn28/10	[64] AA	2.6 ± 0.1	17.1 ± 0.3
wrn28/10*	$D_8 D_4 D_1$ AA	2.39 ± 0.11	15.55 ± 0.13
wrn28/10	$D_8 D_4 D_1$ AA	2.05 ± 0.03	14.30 ± 0.09

Table 6.4: Test errors of different equivariant models on the STL-10 dataset. Models with * are not scaled to the same number of parameters as the original model but preserve the number of channels of the baseline.

model	group	#params	test error (%)
wrn16/8 [69]	-	11M	12.74 ± 0.23
wrn16/8*	$D_1 D_1 D_1$	5M	11.05 ± 0.45
wrn16/8	$D_1 D_1 D_1$	10M	11.17 ± 0.60
wrn16/8*	$D_8 D_4 D_1$	4.2M	10.57 ± 0.70
wrn16/8	$D_8 D_4 D_1$	12M	9.80 ± 0.40

Table 6.5: Test errors of different equivariant models on the STL-10 dataset. Models with * are not scaled to the same number of parameters as the original model but preserve the number of channels of the baseline.

original paper. The four adapted models, reported in Table 6.5, are equivariant under either the action of D_1 in all blocks or the actions of D_8 , D_4 and D_1 in the respective blocks. For both choices we build a large model, whose width is scaled up to approximately match the number of parameters of the baseline, and a small model, which preserves the number of channels and thus compute and memory requirements, but is more parameter efficient.

As expected, all models improve significantly over the baseline with larger models outperforming smaller ones. However, due to their extended equivariance, the small $D_8 D_4 D_1$ model performs better than the large $D_1 D_1 D_1$ model. In comparison to the CIFAR experiments, rotational equivariance seems to give a more significant boost in accuracy. This is expected since the higher resolution of 96×96 pixels of the STL-10 images allows for more detailed local patterns which occur in arbitrary orientations.

Fig. 6.5 reports the results of a data ablation study which investigates the performance of the $D_8 D_4 D_1$ models for smaller training set sizes. The results validate that the gains from incorporating equivariance are consistent over all training sets. More information on the exact training procedures is given in [322].

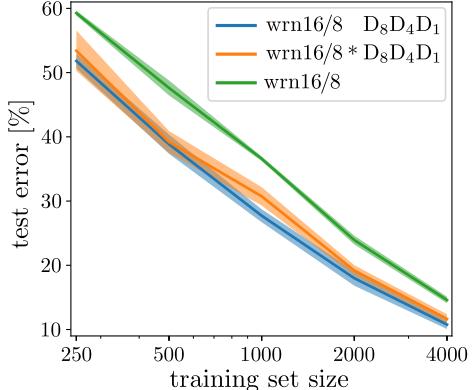


Figure 6.5: Data ablation study on STL-10.
The figure shows the test error percentage on the y-axis (ranging from 10 to 60) versus the training set size on the x-axis (ranging from 250 to 4000). Three data series are plotted: 'wrn16/8 D8D4D1' (blue line), 'wrn16/8 * D8D4D1' (orange line), and 'wrn16/8' (green line). All three series show a downward trend as the training set size increases. The 'wrn16/8' series consistently has the highest test error, followed by 'wrn16/8 D8D4D1', and then 'wrn16/8 * D8D4D1'. The error for all series drops sharply between 250 and 500 training samples and then levels off as the training set size increases further.

6.5 Field type, nonlinearity and symmetry group benchmarking

The framework of $\text{Aff}(G)$ -equivariant steerable CNNs comes with many design choices, like symmetry groups, field types or nonlinearities. While all of the above experiments used regular representations of C_N or D_N as field types and applied element-wise nonlinearities, the current section presents a benchmarking of alternative choices. The results are summarized in Table 6.6 and analyzed in detail below. The first four columns state the equivariance groups $G \leq O(2)$, G -representations, nonlinearities and G -invariant maps which distinguish the models. Column five cites related work that used the corresponding model design. As the performance of the (G -invariant) models depends heavily on the level of symmetry present in the data, we evaluate each model on $O(2)$ -MNIST, $SO(2)$ -MNIST and $\{e\}$ -MNIST. The statistics of each entry are averaged over (at least) 6 samples. All models in these experiments are derived from the base architecture described in Table 6.1. The actual width of each model is adapted such that the number of parameters is approximately preserved. Note that this results in different numbers of channels, depending on the parameter efficiency of the corresponding models. All models apply some form of G -invariant mapping to scalar fields followed by spatial pooling after the last convolutional layer such that the predictions are (up to discretization errors) guaranteed to be $\text{Aff}(G)$ -invariant. The number of invariant features passed to the fully connected classifier is approximately kept constant by adapting the width of the last convolutional layer to the invariant mapping used.

Before guiding through the results in detail, we give a high level overview: All models relying on regular or quotient representations discretize the continuous rotations in $SO(2)$

or $O(2)$ by N discrete rotations in the subgroups C_N or D_N , respectively.⁸ They apply element-wise acting ELU nonlinearities [50], which is a valid choice for any permutation representation; see Section 4.3.3. In addition, there are variants of the regular representation based models that are applying either a max-pooling operation over the feature vector entries to produce a G -invariant (scalar) response field or, for C_N , a vector pooling operation that maps the regular feature field to a vector field. As final G -invariant maps after the last convolution layer, all of these models compute scalars via a G -pooling or G/H -pooling operation over the regular or quotient feature vectors, respectively. For the continuous symmetry groups $O(2)$ and $SO(2)$, we benchmark different choices of irrep fields. These are unitary representations, and are acted on by different types of either norm-nonlinearities or gated nonlinearities, both of which were discussed in Section 4.3.3. An exception are scalar fields, for which we use conventional ELU nonlinearities. Specifically for $O(2)$, we also investigate field types that are defined by an induction of $SO(2)$ -irreps to $O(2)$ -representations. The invariant maps are all variants of either a linear convolution mapping to scalar fields, or a nonlinear mapping that takes the norm of non-trivial irrep features.

In a nutshell, regular representation based models, corresponding to group convolutions, perform best. Among the irrep based models, those using gated nonlinearities work better than those using norm-nonlinearities, and the $O(2)$ -induced $SO(2)$ -irreps outperform $O(2)$ -irreps. In the remainder of this subsection we will guide through the results presented in Table 6.6 in detail.

Regular steerable CNNs: Fig. 6.4 summarizes the results for all regular steerable CNNs on all variants of MNIST. The models in (rows 2-10 and 19-27 in Table 6.6) are steerable CNNs whose feature vectors are *regular* C_N or D_N -representations for varying rotation orders N ; see Def. B.5.18 and Remark B.5.19. As argued in Section 4.5, these models correspond to *group convolutional neural networks* [52, 324, 162, 10]. For the dihedral models we choose a vertical reflection axis. We apply element-wise ELU nonlinearities

$$G\text{-ELU} : \mathbb{R}^{|G|} \rightarrow \mathbb{R}^{|G|}, \quad f := \sum_{g \in G} f_g e_g \mapsto \sum_{g \in G} \text{ELU}(f_g) e_g \quad (6.1)$$

to the regular feature vectors $f \in \mathbb{R}^{|G|}$ and perform group pooling, defined by

$$G\text{-pool} : \mathbb{R}^{|G|} \rightarrow \mathbb{R}, \quad f := \sum_{g \in G} f_g e_g \mapsto \max_{g \in G} f_g, \quad (6.2)$$

as invariant map after the final convolution.

The relative performance of these regular steerable CNNs among each other was already discussed in Section 6.3 (paragraph *G-MNIST*) and Fig. 6.4 above. In comparison to other choices of group representations, regular steerable CNNs perform very well. The reason for this is that G -regular feature vectors in $\mathbb{R}^{|G|}$ can encode an independent response for each individual element of G and that the nonlinearities act localized (element-wise).⁹ The model for $C_1 = \{e\}$ corresponds to a conventional CNN, however, its performance is slightly better than that of the CNN baseline in row 1, which seems to be a result of the smoothed kernel parametrization in terms of (unconstrained) circular harmonics; see Chapter 5 and

⁸ Alternatively, one can use a Monte-Carlo sampling of the infinite-dimensional regular representations of the continuous Lie groups [91, 40].

⁹This is analogous to conventional CNNs, whose features are regular $(\mathbb{R}^d, +)$ -representations: their feature maps store an independent response (pixel) for each translation and nonlinearities act point-wise.

group	representation	nonlinearity	invariant map	citation	O(2)-MNIST	SO(2)-MNIST	$\{e\}$ -MNIST
1	$\{e\}$	trivial (conventional CNN)	ELU	-	5.53 ± 0.20	2.87 ± 0.09	0.91 ± 0.06
2	C_1			[324, 12]	5.19 ± 0.08	2.48 ± 0.13	0.82 ± 0.01
3	C_2			[324, 12]	3.29 ± 0.07	1.32 ± 0.02	0.87 ± 0.04
4	C_3			-	2.87 ± 0.04	1.19 ± 0.06	0.80 ± 0.03
5	C_4			[52, 53, 324, 12, 71]	2.40 ± 0.05	1.02 ± 0.03	0.99 ± 0.03
6	C_6	regular	ρ_{reg}	$G\text{-ELU}$	$G\text{-pooling}$	[125]	2.08 ± 0.03
7	C_8				[324, 12]	1.96 ± 0.04	0.84 ± 0.02
8	C_{12}				[324]	1.95 ± 0.07	0.80 ± 0.03
9	C_{16}				[324, 12]	1.93 ± 0.04	0.82 ± 0.02
10	C_{20}				[324]	1.95 ± 0.05	0.83 ± 0.05
11	C_4		$5\rho_{\text{reg}} \oplus 2\rho_{\text{quot}}^{\frac{C_4}{C_2}} \oplus 2\rho_0^{C_4}$		[53]	2.43 ± 0.05	1.03 ± 0.05
12	C_8		$5\rho_{\text{reg}} \oplus 2\rho_{\text{quot}}^{\frac{C_8}{C_2}} \oplus 2\rho_{\text{quot}}^{\frac{C_8}{C_4}} \oplus 2\rho_0^{C_8}$		-	2.03 ± 0.05	0.84 ± 0.05
13	C_{12}	quotient	$5\rho_{\text{reg}} \oplus 2\rho_{\text{quot}}^{\frac{C_{12}}{C_2}} \oplus 2\rho_{\text{quot}}^{\frac{C_{12}}{C_4}} \oplus 3\rho_0^{C_{12}}$	$G/H\text{-ELU}$	$G/H\text{-pooling}$	-	2.04 ± 0.04
14	C_{16}		$5\rho_{\text{reg}} \oplus 2\rho_{\text{quot}}^{\frac{C_{16}}{C_2}} \oplus 2\rho_{\text{quot}}^{\frac{C_{16}}{C_4}} \oplus 4\rho_0^{C_{16}}$		-	2.00 ± 0.01	0.86 ± 0.04
15	C_{20}		$5\rho_{\text{reg}} \oplus 2\rho_{\text{quot}}^{\frac{C_{20}}{C_2}} \oplus 2\rho_{\text{quot}}^{\frac{C_{20}}{C_4}} \oplus 5\rho_0^{C_{20}}$		-	2.01 ± 0.05	0.83 ± 0.03
16	regular/scalar		$\rho_0^{C_{16}} \xrightarrow{\text{conv}} \rho_{\text{reg}} \xrightarrow{G\text{-pool}} \rho_0^{C_{16}}$	ELU, $G\text{-pooling}$	[52, 195]	2.02 ± 0.02	0.90 ± 0.03
17	C_{16}	regular/vector	$\rho_1^{C_{16}} \xrightarrow{\text{conv}} \rho_{\text{reg}} \xrightarrow{\text{vector pool}} \rho_1^{C_{16}}$	vector field	$G\text{-pooling}$	[196, 198]	2.12 ± 0.02
18	mixed vector		$\rho_{\text{reg}} \oplus \rho_1 \xrightarrow{\text{conv}} 2\rho_{\text{reg}} \xrightarrow{\text{vector pool}} \rho_{\text{reg}} \oplus \rho_1$	ELU, vector field	-	1.87 ± 0.03	0.83 ± 0.02
19	$D_1 = \mathcal{R}$				-	3.40 ± 0.07	3.44 ± 0.10
20	D_2				-	2.42 ± 0.07	2.39 ± 0.04
21	D_3				-	2.17 ± 0.06	2.15 ± 0.05
22	D_4			[52, 71, 53, 307, 216]	1.88 ± 0.04	1.87 ± 0.04	1.69 ± 0.03
23	D_6	regular	ρ_{reg}	ELU	$G\text{-pooling}$	[125, 270]	1.77 ± 0.06
24	D_8				-	1.68 ± 0.06	1.73 ± 0.03
25	D_{12}				-	1.66 ± 0.05	1.65 ± 0.05
26	D_{16}				-	1.62 ± 0.04	1.65 ± 0.02
27	D_{20}				-	1.64 ± 0.06	1.62 ± 0.05
28	D_{16}	regular/scalar	$\rho_0^{D_{16}} \xrightarrow{\text{conv}} \rho_{\text{reg}} \xrightarrow{G\text{-pool}} \rho_0^{D_{16}}$	ELU, $G\text{-pooling}$	-	1.92 ± 0.03	1.88 ± 0.07
							1.74 ± 0.04

29	irreps ≤ 1	$\bigoplus_{j=0}^1 \rho_j^{\text{SO}(2)}$		-	2.98 ± 0.04	1.38 ± 0.09	1.29 ± 0.05		
30	irreps ≤ 3	$\bigoplus_{j=0}^3 \rho_j^{\text{SO}(2)}$		-	3.02 ± 0.18	1.38 ± 0.09	1.27 ± 0.03		
31	irreps ≤ 5	$\bigoplus_{j=0}^5 \rho_j^{\text{SO}(2)}$		-	3.24 ± 0.05	1.44 ± 0.10	1.36 ± 0.04		
32	irreps ≤ 7	$\bigoplus_{j=0}^7 \rho_j^{\text{SO}(2)}$		-	3.30 ± 0.11	1.51 ± 0.10	1.40 ± 0.07		
33	\mathbb{C} -irreps ≤ 1	$\bigoplus_{j=0}^1 \rho_j^{\text{SO}(2),\mathbb{C}}$	ELU, norm-ReLU	conv2triv	[335]	3.39 ± 0.10	1.47 ± 0.06	1.42 ± 0.04	
34	\mathbb{C} -irreps ≤ 3	$\bigoplus_{j=0}^3 \rho_j^{\text{SO}(2),\mathbb{C}}$			[335]	3.48 ± 0.16	1.51 ± 0.05	1.53 ± 0.07	
35	\mathbb{C} -irreps ≤ 5	$\bigoplus_{j=0}^5 \rho_j^{\text{SO}(2),\mathbb{C}}$			-	3.59 ± 0.08	1.59 ± 0.05	1.55 ± 0.06	
36	\mathbb{C} -irreps ≤ 7	$\bigoplus_{j=0}^7 \rho_j^{\text{SO}(2),\mathbb{C}}$			-	3.64 ± 0.12	1.61 ± 0.06	1.62 ± 0.03	
37	$\text{SO}(2)$	$\bigoplus_{j=0}^3 \rho_j^{\text{SO}(2)}$	ELU, squash		-	3.10 ± 0.09	1.41 ± 0.04	1.46 ± 0.05	
38			ELU, norm-ReLU		-	3.23 ± 0.08	1.38 ± 0.08	1.33 ± 0.03	
39	irreps ≤ 3	$\bigoplus_{j=0}^3 \rho_j^{\text{SO}(2)}$	ELU, shared norm-ReLU	norm	-	2.88 ± 0.11	1.15 ± 0.06	1.18 ± 0.03	
40			shared norm-ReLU		-	3.61 ± 0.09	1.57 ± 0.05	1.88 ± 0.05	
41	irreps ≤ 5	$\bigoplus_{j=0}^5 \rho_j^{\text{SO}(2)}$	ELU, gate	conv2triv	-	2.37 ± 0.06	1.09 ± 0.03	1.10 ± 0.02	
42			ELU, shared gate		-	2.33 ± 0.06	1.11 ± 0.03	1.12 ± 0.04	
43	irreps ≤ 7	$\bigoplus_{j=0}^7 \rho_j^{\text{SO}(2)}$	ELU, gate	norm	-	2.23 ± 0.09	1.04 ± 0.04	1.05 ± 0.06	
44			ELU, shared gate		-	2.20 ± 0.06	1.01 ± 0.03	1.03 ± 0.03	
45	irreps = 0	$\rho_0^{\text{O}(2)}$	ELU	-	[144]	5.46 ± 0.46	5.21 ± 0.29	3.98 ± 0.04	
46	irreps ≤ 1	$\rho_0^{\text{O}(2)} \oplus \rho_{\text{sign}}^{\text{O}(2)} \oplus 2\rho_1^{\text{O}(2)}$			-	3.31 ± 0.17	3.37 ± 0.18	3.05 ± 0.09	
47	irreps ≤ 3	$\rho_0^{\text{O}(2)} \oplus \rho_{\text{sign}}^{\text{O}(2)} \bigoplus_{j=1}^3 2\rho_j^{\text{O}(2)}$			-	3.42 ± 0.03	3.41 ± 0.10	3.86 ± 0.09	
48	irreps ≤ 5	$\rho_0^{\text{O}(2)} \oplus \rho_{\text{sign}}^{\text{O}(2)} \bigoplus_{j=1}^5 2\rho_j^{\text{O}(2)}$	ELU, norm-ReLU	O(2)-conv2triv	-	3.59 ± 0.13	3.78 ± 0.31	4.17 ± 0.15	
49	irreps ≤ 7	$\rho_0^{\text{O}(2)} \oplus \rho_{\text{sign}}^{\text{O}(2)} \bigoplus_{j=1}^7 2\rho_j^{\text{O}(2)}$			-	3.84 ± 0.25	3.90 ± 0.18	4.57 ± 0.27	
50	Ind-irreps ≤ 1	Ind $\rho_0^{\text{SO}(2)} \oplus \text{Ind } \rho_1^{\text{SO}(2)}$			-	2.72 ± 0.05	2.70 ± 0.11	2.39 ± 0.07	
51	O(2)	Ind-irreps ≤ 3	Ind $\rho_0^{\text{SO}(2)} \bigoplus_{j=1}^3 \text{Ind } \rho_j^{\text{SO}(2)}$	ELU, Ind-norm-ReLU	Ind-conv2triv	-	2.66 ± 0.07	2.65 ± 0.12	2.25 ± 0.06
52		Ind-irreps ≤ 5	Ind $\rho_0^{\text{SO}(2)} \bigoplus_{j=1}^5 \text{Ind } \rho_j^{\text{SO}(2)}$			-	2.71 ± 0.11	2.84 ± 0.10	2.39 ± 0.09
53	Ind-irreps ≤ 7	Ind $\rho_0^{\text{SO}(2)} \bigoplus_{j=1}^7 \text{Ind } \rho_j^{\text{SO}(2)}$				-	2.80 ± 0.12	2.85 ± 0.06	2.25 ± 0.08
54	irreps ≤ 3	$\rho_0^{\text{O}(2)} \oplus \rho_{\text{sign}}^{\text{O}(2)} \bigoplus_{j=1}^3 2\rho_j^{\text{O}(2)}$	ELU, gate	O(2)-conv2triv	-	2.39 ± 0.05	2.38 ± 0.07	2.28 ± 0.07	
55				norm	-	2.21 ± 0.09	2.24 ± 0.06	2.15 ± 0.03	
56	Ind-irreps ≤ 3	Ind $\rho_0^{\text{SO}(2)} \bigoplus_{j=1}^3 \text{Ind } \rho_j^{\text{SO}(2)}$	ELU, Ind-gate	Ind-conv2triv	-	2.13 ± 0.04	2.09 ± 0.05	2.05 ± 0.05	
57				Ind-norm	-	1.96 ± 0.06	1.95 ± 0.05	1.85 ± 0.07	

Table 6.6: Benchmarking of Euclidean steerable CNNs for different groups G , representations, nonlinearities and final G -invariant maps. Multiplicities of representations are reported in relative terms; the actual multiplicities are depth dependent integer multiples. The results are analyzed in Section 6.5.

Fig. 5.2. The main disadvantage of regular representations is that they are $\mathbb{R}^{|G|}$ -dimensional – while this dimensionality is with $|C_N| = N$ and $|D_N| = 2N$ for usual values of N well manageable, discrete subgroups of $SO(3)$ or $O(3)$ become prohibitively high-dimensional for current hardware.

Quotient representations: As an alternative to regular representations we experiment with some mixtures of G/H -quotient representations of $G = C_N$ (rows 11–15), which are explained in Def. B.5.20 and Remark B.5.21. These models differ from the regular models by enforcing more symmetries in the feature fields and thus kernels [322]. The individual feature fields are lower dimensional; however, by fixing the number of parameters, the models use more different fields which in this specific case leads to approximately the same number of channels and therefore compute and memory requirements. As nonlinearities, we are again applying ELUs element-wise to G/H -quotient features $f \in \mathbb{R}^{|G|/|H|}$, i.e.

$$G/H\text{-ELU} : \mathbb{R}^{|G|/|H|} \rightarrow \mathbb{R}^{|G|/|H|}, \quad f := \sum_{gH \in G/H} f_{gH} e_{gH} \mapsto \sum_{gH \in G/H} \text{ELU}(f_{gH}). \quad (6.3)$$

The invariant map after the final convolution is similarly a max-pooling over the quotient space:

$$G/H\text{-pool} : \mathbb{R}^{|G|/|H|} \rightarrow \mathbb{R}, \quad f := \sum_{gH \in G/H} f_{gH} e_{gH} \mapsto \max_{gH \in G/H} f_{gH} \quad (6.4)$$

We do not observe any significant difference in performance between regular and quotient representations.

G -pooling and vector field nonlinearities: For C_{16} we implement a group pooling network (row 16) and a vector field network (row 17). The former relies on steerable convolutions from scalar to regular feature fields, which are then *after each convolution* mapped back to scalar fields by applying the G -pooling operation from Eq. (6.2). Vector field networks [196] map the regular feature fields instead to vector fields, where the vectors’ norms are again determined by the maximal response, but an additional directional information is computed via an $\arg \max_{g \in G}$ operation. Both pooling operations compress the features in the regular fields, which can lead to lower memory and compute requirements. However, since we fix the number of parameters, the resulting models are ultimately much wider than the corresponding regular steerable CNNs. Since the pooling operations lead to a loss of information, both models perform worse than their purely regular counterpart on $O(2)$ -MNIST and $SO(2)$ -MNIST. Surprisingly, the group pooling network, whose features are orientation unaware, performs better than the vector field network. On $\{e\}$ -MNIST the group pooling network closes up with the regular steerable CNNs while the vector field network achieves an even better result. We further experiment with a model which applies vector field nonlinearities to only half of the regular fields and preserves the other half (row 18). This model is on par with the regular model on both transformed MNIST versions but achieves the overall best result on $\{e\}$ -MNIST. Similar to the case of C_{16} , the group pooling network for D_{16} (row 28) performs worse than the corresponding regular model, this time also on $\{e\}$ -MNIST.

$SO(2)$ irrep models: The feature fields of all $SO(2)$ -equivariant models that we consider are defined to transform according to irreducible representations $\rho_j^{SO(2)}$, which were defined in Eq. (5.33) in Section 5.3.4. Note that this covers scalar fields and vector fields for the trivial representation $\rho_0^{SO(2)}$ and defining representation $\rho_1^{SO(2)}$, respectively. Overall, these

models are not competitive compared to the regular steerable CNNs. This result is particularly important for $\text{SE}(3) = \text{Aff}(\text{SO}(3))$ -equivariant CNNs whose feature fields are often transforming according to $\text{SO}(3)$ -irreps [301, 323, 161, 163, 3].

The models in rows 29-32 are inspired by Harmonic Networks [335] and consist of irrep fields with the same multiplicity up to a certain threshold order. All models apply ELUs on scalar fields and norm-ReLUs (see Section 4.3.3) on higher order fields. The projection to invariant features is done via a convolution to scalar features (conv2triv) in the last convolutional layer. We find that irrep fields up to order 1 and 3 perform equally well while higher threshold orders yield worse results. The original implementation of Harmonic Networks considered complex irreps of $\text{SO}(2)$, which results in a lower dimensional steerable kernel basis as discussed in [322]. We reimplemented these models and found that their reduced kernel space leads to consistently worse results than ours (rows 33-36).

For the model containing irreps up to order 3 we implemented some alternative variants: The model in row 37 replaces the norm-ReLU activations by *squashing nonlinearities*, introduced in Section 4.3.3, which leads to a slightly worse performance. Row 38 shows a variation which sticks with norm-ReLUs, but computes the final G -invariant responses by taking the *norms* of all non-scalar fields. This network design does again not improve upon the baseline variant. Instead of applying the norm-ReLU nonlinearity individually to each higher order irrep field, the model in row 39 takes the norm of direct sums of higher order irrep features (each with multiplicity one), which is valid since this direct sum representation is unitary. This *shared norm-ReLU* nonlinearity performs significantly better than its baseline, but still far worse than regular steerable CNNs. Another variation in row 40 again takes the norm of direct sums of irrep features, now including trivial ones, which results in the overall worst results among all $\text{SO}(2)$ -models. The models in rows 41-44 apply learned *gated nonlinearities*, introduced in [323] and Section 4.3.3). These nonlinearities are either applied to each higher order irrep feature individually (gate), or to their direct sum (shared gate). While not beating regular steerable CNNs, these models close up to their performance. This insight might be interesting for $\text{SE}(3)$ or $\text{E}(3)$ -equivariant models, for which regular representations are quite high dimensional.

O(2) models: As for $\text{SO}(2)$, we are investigating steerable CNNs whose features transform according to $\text{O}(2)$ -irreps (Section 5.3.4, Eq. (5.54)) up to a certain order and apply norm-ReLUs (rows 46-49). In this case we choose twice the multiplicity for two-dimensional field types $\rho_{j \geq 1}^{\text{O}(2)}$ in comparison to that of one-dimensional types $\rho_0^{\text{O}(2)}$ and $\rho_{\text{sign}}^{\text{O}(2)}$, which reflects the multiplicities of irreps contained in the regular representation of $\text{O}(2)$. Invariant predictions are computed by convolving in equal proportion to fields which transform under trivial irreps $\rho_0^{\text{O}(2)}$ (scalar fields) and sign-flip irreps $\rho_{\text{sign}}^{\text{O}(2)}$ (pseudoscalar fields), followed by taking the absolute value of the latter ($\text{O}(2)$ -conv2triv).¹⁰ We again find that higher irrep thresholds yield worse results, this time already starting from threshold order 1. In particular, these models perform worse than their $\text{SO}(2)$ -equivariant counterparts, even on $\text{O}(2)$ -MNIST. This suggests that the kernel constraint for this particular choice of representations is too restrictive.

If only scalar fields, corresponding to the trivial irrep $\rho_0^{\text{O}(2)}$, are chosen, the kernel constraint becomes $K(gx) = K(x) \quad \forall g \in \text{O}(2)$, and therefore allows for isotropic kernels only. This limits the expressivity of the model so severely that it performs even worse than a conventional CNN on $\text{SO}(2)$ -MNIST and $\{e\}$ -MNIST, while being on par for $\text{O}(2)$ -MNIST;

¹⁰The motivation for $\text{O}(2)$ -conv2triv is that the steerable kernel space for immediate conv2triv convolutions from pseudoscalar fields ($\rho_{\text{sign}}^{\text{O}(2)}$) to scalar fields ($\rho_0^{\text{O}(2)}$) is empty; see Table 5.3.

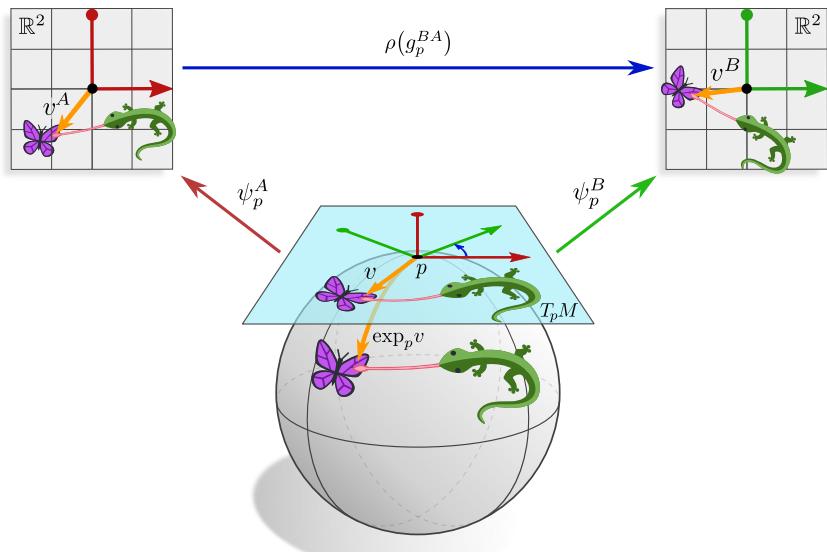
see row 45. Note that isotropic kernels correspond to vanilla graph convolutional networks, which are well known to *oversmooth* the signals they process [44, 31, 17].

In order to improve the performance of $O(2)$ -steerable CNNs, we propose to use $O(2)$ -representations $\text{Ind}_{SO(2)}^{O(2)} \rho_j^{SO(2)}$, induced from the irreps of $SO(2)$, as field types. By the definition of induction, this leads to pairs of fields which transform according to $\rho_j^{SO(2)}$ under rotations in $SO(2)$ but permute under reflections in $O(2)/SO(2) \cong_{\text{grp}} \mathcal{R}$ [322]. The multiplicity of the irreps of $O(2)$ contained in this induced representation coincides with the multiplicities chosen in the pure $O(2)$ -irrep models. However, the change of basis, relating both representations, does not commute with the nonlinearities, such that the networks behave differently. We apply Ind-norm-ReLU nonlinearities to the induced $O(2)$ models which compute the norm of each of the permuting subfields individually but share the norm-ReLU parameters (the bias) to guarantee equivariance. In order to project to final $O(2)$ -invariant feature vectors, we first apply a convolution producing $\text{Ind}_{SO(2)}^{O(2)} \rho_0^{SO(2)}$ fields (Ind-conv2triv). Since these transform like the regular representation of $\mathcal{R} \cong_{\text{grp}} O(2)/SO(2)$, we can simply apply G -pooling over the two reflections. The results, given in rows 50-53, show that these models perform significantly better than the $O(2)$ -irreps models and outperform the $SO(2)$ -irrep models on $O(2)$ -MNIST.

We again build models that apply gated nonlinearities. As for $SO(2)$, this leads to a greatly improved performance of the pure irrep models; see rows 54-55. In addition we adapt the gated nonlinearity to the *induced* irrep models (rows 56-57). Here we apply an independent gate to each of the two permuting sub-fields (Ind-gate). In order to be equivariant, the gates need to permute under reflections as well, which is easily achieved by deriving them from $\text{Ind}_{SO(2)}^{O(2)} \rho_0^{SO(2)}$ fields instead of scalar fields. The gated induced irrep model achieves the best results among all $O(2)$ -steerable networks, however, it is still not competitive compared to the D_N models with large N .

PART II

AN INTRODUCTION TO COORDINATE INDEPENDENT CNNS



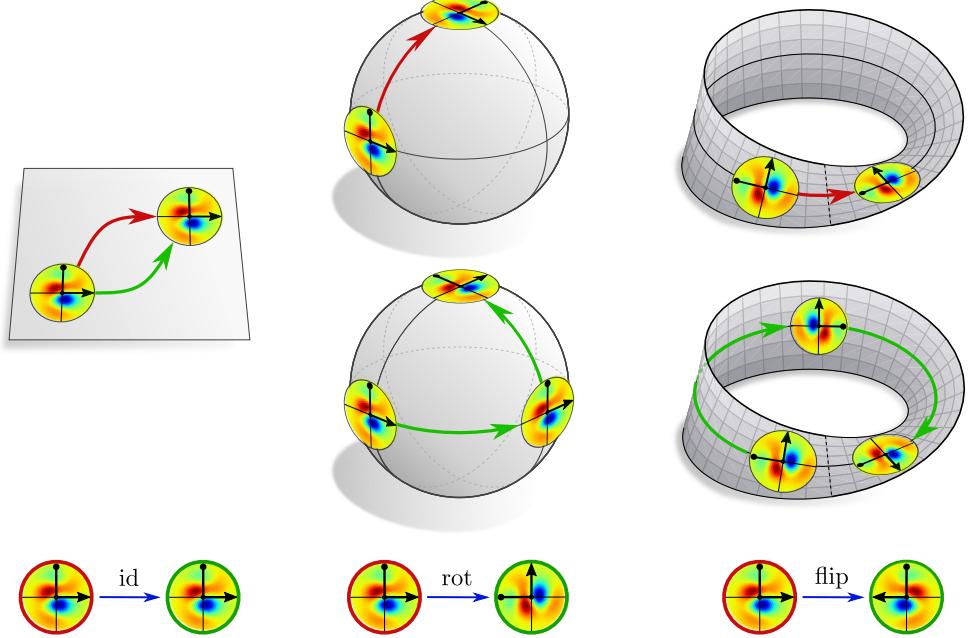


Figure II.1: An intuition on the inherent ambiguity of weight sharing on manifolds. *Left:* A common interpretation of weight sharing on the plane is to shift a kernel over the whole space. Since parallel transport is on Euclidean spaces path independent, this is unambiguous. *Middle:* On curved spaces, like the sphere, parallel transport is path dependent. Different paths result in kernels that are rotated relative to each other. *Right:* The Möbius strip is a non-orientable manifold. Different paths can therefore result in kernels that are reflected relative to each other. *Bottom:* We formalize different kernel alignments by different choices of local reference frames of the corresponding tangent spaces. It is well known that no choice of reference frames (gauge) is preferred on general manifolds. Different coordinatizations are related by gauge transformations, which take values in the structure group G of the manifold (the trivial group $G = \{e\}$ for the plane, rotation group $G = \text{SO}(2)$ for the sphere and reflection group $G = \mathcal{R}$ for the Möbius strip). Coordinate independent CNNs address the ambiguity of reference frames by applying G -steerable (gauge equivariant) convolution kernels.

Introduction & overview

There are many applications where the signal to be processed by a neural network is not supported on a flat Euclidean space, but on more general Riemannian manifolds; see for instance Fig. 1.4. Parts II and III of this work address such applications by generalizing the Euclidean steerable CNNs from Part I to the differential geometric setting. Since convolutions are essentially characterized by their spatial weight sharing property, a central question in this endeavor is *how convolution kernels should be shared over Riemannian manifolds.*¹¹

In the Euclidean setting, weight sharing could be derived by demanding models to be affine group equivariant. This approach generalizes to arbitrary homogeneous spaces (Def. B.3.11) like the sphere or torus [56]. However, it is bound to fail on general manifolds since they are potentially asymmetric, i.e. may have no transitive symmetries w.r.t. which one could demand networks to be equivariant. Weights would only be shared over isometry group orbits as visualized in Fig. 1.11 and no weight sharing at all would result for trivial isometry groups. An alternative intuition from the Euclidean setting is that kernels are shared by “shifting” them over the space. Since parallel transporters on Euclidean spaces are path independent, this results in an unambiguous alignment of kernels; see Fig. II.1 (left). However, transporters become on curved or non-orientable spaces path dependent, and thus unsuitable for sharing weights. Fig. II.1 (middle and right) exemplifies this issue for the sphere and the Möbius strip, where different paths lead to a different kernel alignment.

As it turns out, *the alignment of convolution kernels on manifolds is inherently ambiguous.* A natural solution to address a G -ambiguity in the kernel alignment is to use G -steerable kernels: different kernel alignments result then in equivalent responses, differing only by a predictable group action. Viewing kernels as local observers who are measuring features relative to their *local frames of reference*, this independence of the extracted information from the chosen kernel alignments (frames) can be interpreted as the networks’ *coordinate independence* (G -covariance).

The theory of coordinate independent CNNs is formalized in reverse order – its foundational principle is the models’ coordinate independence, from which the requirement for the kernels’ steerability is shown to follow. Chapter 7 introduces the underlying gauge theoretic formalism, which describes *coordinatizations of the tangent spaces* and gauge transformations between different choices of coordinates. Based on this, Chapter 8 defines *coordinate independent feature spaces*, in particular the gauge transformations, parallel transport and isometry pushforward of feature vectors. *Coordinate independent neural network layers* that map between such feature vector fields are developed in Chapter 9. An exemplary *instantiation* of such coordinate independent feature fields and network layers on the *Möbius strip* is presented in Chapter 10.

The aim of the current Part II is to introduce the theory of coordinate independent CNNs in an easily accessible language. It describes all quantities and layers in *local coordinates* (gauges), ensuring thereby that the particular choice of gauge remains irrelevant. This is in contrast to Part III, which develops the corresponding *coordinate free* formulation in terms of associated fiber bundles. The formulation presented here will be shown to follow from the global theory when expressing it in local bundle trivializations.

¹¹This question applies more generally to any shared template operation, including for instance biases or pointwise nonlinearities.

Gauges, gauge transformations and G -structures

Geometric quantities like tangent or feature vectors exist independently of coordinates, however, a (non-symbolic) computer implementation requires them to be expressed in terms of numerical coefficients in *some* gauge, i.e. relative to some choice of reference frames. The specific choice of coordinates is irrelevant – it represents just one of multiple *equivalent descriptions*. The appropriate mathematical framework to regulate such redundant degrees of freedom are *gauge theories*. A gauge theory accounts for the equivalence of different gauges by consistently relating them to each other via *gauge transformations*.

This chapter discusses the coordinatization of tangent spaces, from which coordinate expressions of associated geometric quantities like feature vectors will follow. In particular, Section 7.1 introduces gauges and gauge transformations of the tangent spaces as a formal way of describing choices of local reference frames and transformations between them. Section 7.2 explains how functions on tangent spaces are represented relative to different coordinatizations – this introduces the idea of coordinate independent mappings, which we use later to define coordinate independent network layers. Section 7.3 defines G -structures and G -atlases.

7.1 Tangent spaces and reference frames

A d -dimensional (smooth) manifold M has a tangent space $T_p M \cong \mathbb{R}^d$ attached to each point $p \in M$. The tangent spaces are d -dimensional vector spaces, however, in contrast to \mathbb{R}^d they do in general not come with any preferred choice of reference frame. A tangent vector $v \in T_p M$ is a *coordinate free* object and is thus not immediately represented numerically by a coordinate tuple $(v_1, \dots, v_d) \in \mathbb{R}^d$. More abstractly stated, each tangent space $T_p M$ is isomorphic to \mathbb{R}^d but in general no canonical isomorphism between them is given. Both spaces are therefore structurally equivalent but are not identified with each other in any preferred way.

A *gauge* (local trivialization of the tangent bundle) on $U^A \subseteq M$ is defined as a smoothly position-dependent collection of invertible linear maps

$$\psi_p^A : T_p M \rightarrow \mathbb{R}^d, \quad p \in U^A, \tag{7.1}$$

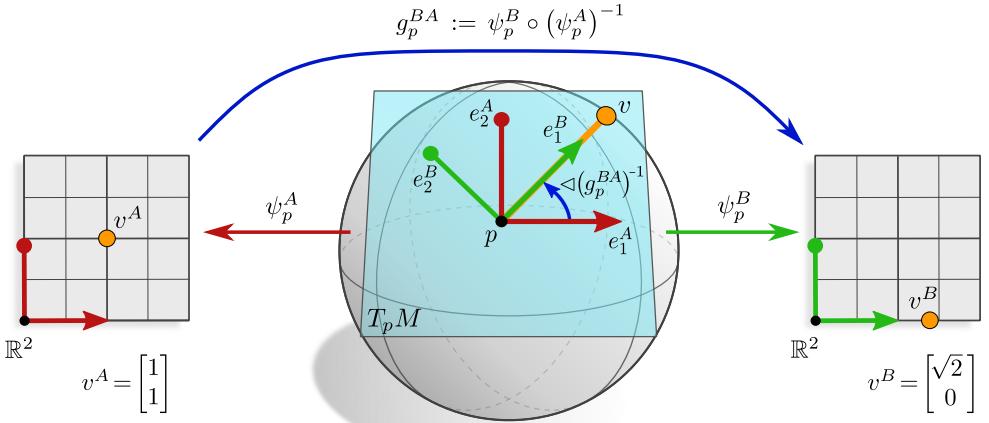


Figure 7.1: Identification of $T_p M \cong \mathbb{R}^2$ with \mathbb{R}^2 via different gauges. A (coordinate free) tangent vector $v \in T_p M$ (orange) can be represented numerically by a coordinate tuple $v^A = \psi_p^A(v) = (1, 1)^\top$ relative to gauge ψ_p^A (red) or, equivalently, by $v^B = \psi_p^B(v) = (\sqrt{2}, 0)^\top$ relative to gauge ψ_p^B (green). A choice of gauge corresponds to a choice $[e_1^A, e_2^A]$ or $[e_1^B, e_2^B]$ of reference frame. On a general manifold no choice of gauge or coordinatization is preferred a priori. Different gauges, and thus reference frames, are related by gauge transformations $g_p^{BA} := \psi_p^B \circ (\psi_p^A)^{-1}$ (blue) which take values in the thus defined structure group G . This figure is a graphical interpretation of the commutative diagrams in Eq. (7.8) and Fig. 11.5a. Note that gauges are immediately assigning coordinates to tangent spaces. Fig. 15.2 in Section 15.1 shows a similar diagram for (affine) charts, which assign coordinates to the manifold, thereby *inducing* gauges (“coordinate bases”).

specifying the missing vector space isomorphisms between $T_p M$ and \mathbb{R}^d . As visualized in Fig. 7.1, they coordinatize the tangent spaces by assigning a *coefficient vector*

$$v^A := \psi_p^A(v) \in \mathbb{R}^d \quad (7.2)$$

to each coordinate free tangent vector $v \in T_p M$. An inversion of this relation yields

$$v = (\psi_p^A)^{-1}(v^A) = (\psi_p^A)^{-1}\left(\sum_i v_i^A \epsilon_i\right) = \sum_i v_i^A (\psi_p^A)^{-1}(\epsilon_i) =: \sum_i v_i^A e_i^A, \quad (7.3)$$

where we denoted by $\{\epsilon_1, \dots, \epsilon_d\}$ the standard basis of \mathbb{R}^d and made use of the linearity of the gauge to pull out the summation. This shows that the gauge can be thought of as endowing each tangent space $T_p M$ with a *reference frame*

$$[e_1^A, \dots, e_d^A] := \left[(\psi_p^A)^{-1}(\epsilon_1), \dots, (\psi_p^A)^{-1}(\epsilon_d) \right], \quad (7.4)$$

defined as that d -tuple of linearly independent tangent vectors which results when mapping the standard frame of \mathbb{R}^d back through the inverse gauge map. For brevity, we will in the following use the shorthand notation $[e_i^A]_{i=1}^d$ for frames $[e_1^A, \dots, e_d^A]$. The coefficients v^A are the coordinates of v relative to this frame. The collection of frames induced by the ψ_p^A on U^A is called (smooth) *frame field*; see Fig 7.2 for a visualization.

Gauges ψ^X coordinatize tangent spaces only on local neighborhoods $U^X \subseteq M$, and can due to topological obstructions in general not be extended to the whole manifold without

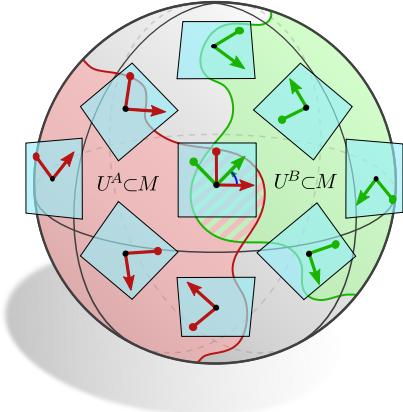


Figure 7.2: Each point p of a Riemannian manifold M has a tangent space $T_p M$ attached. A smooth gauge ψ^A on a suitably chosen subset $U^A \subseteq M$ (red) coordinatizes all tangent spaces $T_p M$ for p in U^A as shown in Fig. 7.1. It is equivalent to a choice of *smooth frame field* on U^A . Since it is in general not possible to extend a gauge globally over the whole manifold, it is necessary to consider a *G-atlas*, consisting of gauges which cover M . Different coordinatizations ψ^A on U^A (red) and ψ^B on U^B (green) are patched together via gauge transformations (or transition maps) $g^{BA} : U^A \cap U^B \rightarrow G$ which are defined on the overlap $U^A \cap U^B$ (striped) and take values in the structure group $G \leq \text{GL}(d)$.

violating the smoothness assumption. One therefore considers an *atlas*

$$\mathcal{A} = \{(U^X, \psi^X)\}_{X \in \mathfrak{X}}, \quad (7.5)$$

consisting of smooth gauges on a set of neighborhoods U^X covering the manifold, that is, satisfying $\bigcup_{X \in \mathfrak{X}} U^X = M$, where \mathfrak{X} is an index set.¹ On the overlaps $U^A \cap U^B \neq \emptyset$ of neighborhoods, different gauges ψ_p^A and ψ_p^B are stitched together by smooth *transition functions*

$$g^{BA} : U^A \cap U^B \rightarrow \text{GL}(d), \quad p \mapsto g_p^{BA} := \psi_p^B \circ (\psi_p^A)^{-1}. \quad (7.6)$$

Here we assume the codomain (for now) to be given by the *general linear group* $\text{GL}(d)$, consisting of all invertible matrices in $\mathbb{R}^{d \times d}$, which explain the relation between any pair of vector space isomorphisms (gauges) or reference frames. The action of such a transition function on a given gauge defines a *gauge transformation*

$$\psi_p^B = g_p^{BA} \cdot \psi_p^A. \quad (7.7)$$

In terms of a commutative diagram, the relation between different gauges is visualized as:

$$\begin{array}{ccccc} & & g_p^{BA} \cdot & & \\ & \swarrow & & \searrow & \\ \mathbb{R}^d & \xleftarrow{\psi_p^A} & T_p M & \xrightarrow{\psi_p^B} & \mathbb{R}^d \\ & \uparrow & & & \downarrow \\ & g_p^{AB} \cdot = (g_p^{BA})^{-1}. & & & \end{array} \quad (7.8)$$

Compare this diagram to its graphical interpretation in Fig 7.1.

A gauge transformation alters the coordinatization of the tangent spaces such that the same coordinate free tangent vector v is represented by a different component vector

$$v^B = g_p^{BA} v^A. \quad (7.9)$$

¹An atlas of gauges is very similar to usual atlases of charts of a manifold (Appendix C). The difference is that the here considered atlases directly assign coordinates to the tangent bundle TM instead of to the manifold M .

Since a gauge corresponds to a choice of frame field, a gauge transformation corresponds to a transformation between frame fields. Specifically, a frame $[e_i^A]_{i=1}^d = [e_1^A, \dots, e_d^A]$ at $p \in M$ transforms to another frame

$$\begin{aligned}
[e_i^B]_{i=1}^d &:= \left[(\psi_p^B)^{-1} (\epsilon_i) \right]_{i=1}^d && \text{(gauge induced frame, Eq. (7.4))} \\
&= \left[(g_p^{BA} \cdot \psi_p^A)^{-1} (\epsilon_i) \right]_{i=1}^d && \text{(gauge transformation, Eq. (7.7))} \\
&= \left[(\psi_p^A)^{-1} ((g_p^{BA})^{-1} \epsilon_i) \right]_{i=1}^d && \text{(expanded inverse)} \\
&= \left[(\psi_p^A)^{-1} \left(\sum_j \epsilon_j \epsilon_j^\top (g_p^{BA})^{-1} \epsilon_i \right) \right]_{i=1}^d && \text{(inserted identity } \mathbb{1} = \sum_j \epsilon_j \epsilon_j^\top \text{)} \\
&= \left[(\psi_p^A)^{-1} \left(\sum_j \epsilon_j \left((g_p^{BA})^{-1} \right)_{ji} \right) \right]_{i=1}^d && \text{(matrix elements of } (g_p^{BA})^{-1} \text{)} \\
&= \left[\sum_j (\psi_p^A)^{-1} (\epsilon_j) \left((g_p^{BA})^{-1} \right)_{ji} \right]_{i=1}^d && \text{(linearity of } \psi_p^A \text{)} \\
&= \left[\sum_j e_j^A \left((g_p^{BA})^{-1} \right)_{ji} \right]_{i=1}^d && \text{(gauge induced frame, Eq. (7.4))} \\
&=: [e_i^A]_{i=1}^d \triangleleft (g_p^{BA})^{-1} && \text{(7.10)}
\end{aligned}$$

via the thus defined *right action*

$$\triangleleft : ([e_i]_{i=1}^d, g) \mapsto [e_i]_{i=1}^d \triangleleft g := \left[\sum_j e_j g_{ji} \right]_{i=1}^d \quad (7.11)$$

of group elements on frames. Note that the inverse in this action in Eq. (7.10) is due to the definition of Eq. (7.7) without inverse.² One usually refers to the transformation behavior of reference frames as *covariant* transformation while the transformation of gauges and vector coefficients is denoted as *contravariant* transformation; see Appendix C.

Since the transformation behavior of the coefficients in Eq. (7.9) and the basis in Eq. (7.10) are inverse to each other they compensate, that is, they leave the tangent vector $v = \sum_i v_i^A e_i^A = \sum_i v_i^B e_i^B$ invariant:

$$\begin{aligned}
v &= \sum_i v_i^B e_i^B = \sum_i v_i^B \sum_j e_j^A \left((g_p^{BA})^{-1} \right)_{ji} \\
&= \sum_j \left(\sum_i \left((g_p^{BA})^{-1} \right)_{ji} v_i^B \right) e_j^A \\
&= \sum_j v_j^A e_j^A. \quad (7.12)
\end{aligned}$$

This construction ensures that any calculation is ultimately independent of the chosen gauge, which is usually denoted as *coordinate independence*. In general, any coordinate representation of a coordinate free object or function is for consistency reasons required to be coordinate independent.

²Other conventions might flip the choice of inverses in $\psi^B = g^{BA} \psi^A$ and $[e_i^B]_{i=1}^d = [e_i^A]_{i=1}^d \triangleleft (g^{BA})^{-1}$. An inverse in either of the two equations is necessary to make the left action \cdot on gauges and right action \triangleleft on frames compatible.

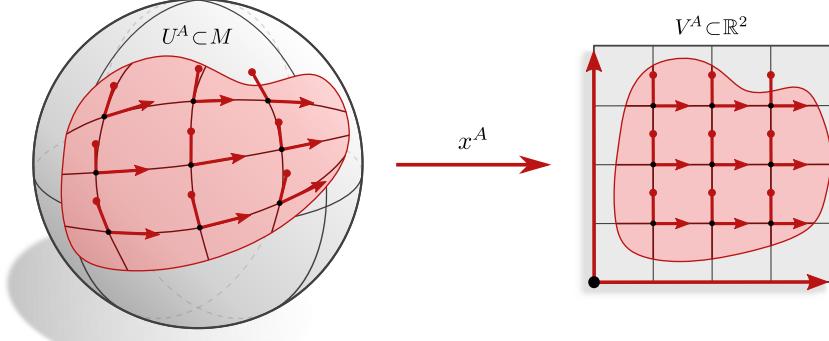


Figure 7.3: A *chart* $x^A : U^A \rightarrow V^A$ assigns coordinates $V^A \subseteq \mathbb{R}^d$ to regions $U^A \subseteq M$ of the manifold. It induces *coordinate bases* $\left[\frac{\partial}{\partial x_1^A} \Big|_p, \dots, \frac{\partial}{\partial x_d^A} \Big|_p \right]$ and corresponding gauges $\psi_p^A = \hat{dx}_p^A$ of the tangent spaces $T_p M$ over U^A . We will mostly not work with charts but rather refer to points $p \in M$ in a *coordinate free* manner. Gauges (frames) are then directly assigned to the tangent spaces instead of being induced.

For completeness we want to mention that the here presented formalism defines general bases of the tangent spaces, sometimes referred to as *non-coordinate bases* (non-holonomic bases), in terms of local gauges. A very popular but less general alternative are *coordinate bases* (holonomic bases)

$$\left[\frac{\partial}{\partial x_1^A} \Big|_p, \dots, \frac{\partial}{\partial x_d^A} \Big|_p \right], \quad (7.13)$$

which are induced by *coordinate charts*

$$x^A : U^A \rightarrow V^A \subseteq \mathbb{R}^d \quad (7.14)$$

of the manifold [221]. The corresponding gauges are given by the the *chart differentials*, that is,

$$\psi_p^A = \hat{dx}_p^A = (\hat{dx}_{p,1}^A, \dots, \hat{dx}_{p,d}^A)^\top : T_p M \rightarrow \mathbb{R}^d. \quad (7.15)$$

Gauge transformations coincide in this setting with the *Jacobians*

$$g_p^{BA} = \left. \frac{\partial x^B}{\partial x^A} \right|_{x^A(p)} \in \mathrm{GL}(d) \quad (7.16)$$

of chart transition maps. An exemplary chart and its induced coordinate bases are visualized in Fig. 7.3. Appendix C discusses the relationship between both formalisms in detail; an overview is given in Table C.1.

In the remainder of this paper we will mainly work in the gauge formalism, which assigns reference frames immediately to the tangent spaces instead of inducing them from charts. Exceptions are the Möbius convolutions in Chapter 10, Euclidean CNNs in Chapter 15, log-polar coordinates in Section 16.2 and icosahedral CNNs in Section 17.4. In all of these cases the manifolds are *locally flat* and the charts are *isometric*, such that they induce *orthonormal frames*. GM-convolutions on U^A can then be computed in an efficient manner by running Euclidean convolutions with G -steerable kernels on the charts' codomains V^A .

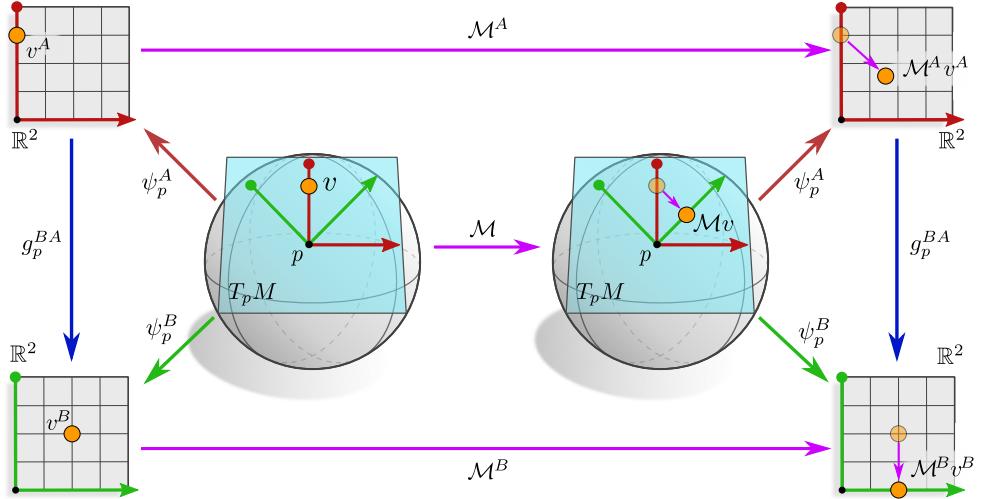


Figure 7.4: Graphical interpretation of the commutative diagram in Eq. (7.23). A coordinate free map $\mathcal{M} : T_p M \rightarrow T_p M$ can be equivalently represented by functions $\mathcal{M}^A : \mathbb{R}^d \rightarrow \mathbb{R}^d$ or $\mathcal{M}^B : \mathbb{R}^d \rightarrow \mathbb{R}^d$ relative to different gauges ψ_p^A or ψ_p^B , respectively. These coordinatizations of \mathcal{M} are defined by a pre- and postcomposition with gauges in the domain and codomain, for instance, following the arrows, $\mathcal{M}^A := \psi_p^A \circ \mathcal{M} \circ (\psi_p^A)^{-1}$. As a consequence, gauge transformations $\mathcal{M}^B = g_p^{BA} \mathcal{M}^A (g_p^{BA})^{-1}$ between coordinatizations are given by a pre- and postcomposition with transition maps g_p^{BA} in the domain and codomain. All quantities and mappings in this work will either be coordinate free (like \mathcal{M}) or will be expressed in a coordinate independent way in different gauges (like \mathcal{M}^A and \mathcal{M}^B). We will therefore need to define (or derive) transformation laws for any quantity and function.

7.2 Coordinate independent functions on tangent spaces

Just as the vectors $v \in T_p M$, *functions* on the tangent spaces are coordinate free, that is, they are defined without referring to any reference frame. A chosen gauge allows to represent such coordinate free mappings by functions which operate on coefficient vectors in \mathbb{R}^d . Similar to the coefficient vectors, coordinatizations of functions need to transform in a specific way under gauge transformations in order to be consistently defined, i.e. to respect coordinate independence. We will later apply the here presented concept of expressing coordinate free mappings in terms of local coordinates to define GM -coordinate independent convolutions.

As a simple example for a coordinate free operation, let us consider the case of a *linear map*

$$\mathcal{M} : T_p M \rightarrow T_p M. \quad (7.17)$$

Let $v_{\text{in}} \in T_p M$ be a tangent vector which is by \mathcal{M} being mapped to $v_{\text{out}} = \mathcal{M}v_{\text{in}} \in T_p M$. Linear maps are in numerical implementations usually modeled by *coefficient matrices* which map between *coefficient vectors* relative to some choice of reference frame. To make this precise, assume some gauge ψ_p^A to be given such that the coordinate free vectors v_{in} and v_{out} in $T_p M$ are represented by coefficient vectors $v_{\text{in}}^A = \psi_p^A(v_{\text{in}})$ and $v_{\text{out}}^A = \psi_p^A(v_{\text{out}})$ in \mathbb{R}^d . The linear map \mathcal{M} is in this gauge represented by the matrix

$$\mathcal{M}^A := \psi_p^A \circ \mathcal{M} \circ (\psi_p^A)^{-1} \in \mathbb{R}^{d \times d} \quad (7.18)$$

whose definition is visualized by the commutative diagram below:

$$\begin{array}{ccccc} \mathbb{R}^d & \xleftarrow{\psi_p^A} & T_p M & \xrightarrow{\mathcal{M}} & T_p M & \xrightarrow{\psi_p^A} & \mathbb{R}^d \\ & & \underbrace{\qquad\qquad\qquad}_{\mathcal{M}^A} & & & & \uparrow \\ & & & & & & \end{array} \quad (7.19)$$

The matrix is consistent with the coordinate free mapping since both imply each other:

$$\begin{aligned} \mathcal{M}^A v_{\text{in}}^A &= [\psi_p^A \circ \mathcal{M} \circ (\psi_p^A)^{-1}] \circ [\psi_p^A(v_{\text{in}})] \\ &= \psi_p^A(\mathcal{M} v_{\text{in}}) \\ &= \psi_p^A(v_{\text{out}}) \\ &= v_{\text{out}}^A \end{aligned} \quad (7.20)$$

Of course one could have represented \mathcal{M} relative to any other choice of gauge ψ_p^B . We know from Eq. (7.9) that coefficient vectors in different gauges are related by $v^B = g_p^{BA} v^A$. Similarly, \mathcal{M}^B relates to \mathcal{M}^A by the gauge transformation

$$\begin{aligned} \mathcal{M}^B &= \psi_p^B \circ \mathcal{M} \circ (\psi_p^B)^{-1} \\ &= \psi_p^B \circ (\psi_p^A)^{-1} \circ \mathcal{M}^A \circ \psi_p^A \circ (\psi_p^B)^{-1} \\ &= g_p^{BA} \mathcal{M}^A (g_p^{BA})^{-1}, \end{aligned} \quad (7.21)$$

which acts here both on the domain and codomain.³ This transformation law is again seen to be consistent by the mutual transformations canceling out:

$$\begin{aligned} \mathcal{M}^B v_{\text{in}}^B &= [g_p^{BA} \mathcal{M}^A (g_p^{BA})^{-1}] [g_p^{BA} v_{\text{in}}^A] \\ &= g_p^{BA} \mathcal{M}^A v_{\text{in}}^A \\ &= g_p^{BA} v_{\text{out}}^A \\ &= v_{\text{out}}^B \end{aligned} \quad (7.22)$$

The derived gauge transformations therefore assert that all coordinatized computations are ultimately coordinate independent. The relations between the coordinate free mapping and its coordinatizations is summarized by the following commutative diagram

$$\begin{array}{ccccccc} \mathbb{R}^d & & \mathcal{M}^A & & \mathbb{R}^d \\ & \swarrow & & \searrow & & \\ & \psi_p^A & & & \psi_p^A & & \\ & \downarrow g_p^{BA} & & & \downarrow g_p^{BA} & & \\ \mathbb{R}^d & & T_p M & \xrightarrow{\mathcal{M}} & T_p M & \xrightarrow{\psi_p^B} & \mathbb{R}^d \\ & \swarrow & & \searrow & & \swarrow & \\ & \psi_p^B & & & \psi_p^B & & \end{array} \quad (7.23)$$

which is graphically interpreted in Fig. 7.4.

³The transformation of the matrix coefficients via the left and right multiplication with g^{BA} and $(g^{BA})^{-1}$, respectively, identify the linear map as a tensor of type $(1, 1)$.

In practice one can not instantiate the coordinate free linear map \mathcal{M} numerically without referring to a choice of coordinatization. However, its existence is implied if (and only if) its coordinatizations relate to each other as specified by Eq. (7.21), which ensures that the correct transformation behavior of the input and output vector coefficients in Eq. (7.9) is preserved.

7.3 Structure groups, G -structures and G -atlases

We will later on require neural networks to operate in a coordinate independent manner, that is, we demand their inference to be independent from arbitrary choices of reference frames. This raises the question to which extent the choice of reference frames on a manifold is arbitrary. In the previous Sections 7.1 and 7.2 we allowed for any possible choice of gauge or reference frame, which were thus related by general $GL(d)$ -valued gauge transformations. In many applications the manifold does, however, come with additional structure which allows to distinguish a *preferred subset of reference frames* or gauges, whose transition functions take values in a *reduced structure group* $G \leq GL(d)$. Such geometric structures – or rather the subsets of preferred reference frames themselves, which encode equivalent information – are denoted as *G -structures*.

G -structures are best understood by considering some specific examples. The following list gives such examples, classified by their structure group $G \leq GL(d)$:

- $O(d)$: Consider the *metric* structure of a Riemannian manifold, which allows to measure distances and angles, and therefore to distinguish orthonormal frames, that is, those frames that satisfy $\eta(e_i, e_j) = \delta_{ij}$ for any $i, j = 1, \dots, d$. Correspondingly, a Riemannian metric allows to talk about isometric gauges ψ_p^A , which identify the metric of \mathbb{R}^d with that of $T_p M$, i.e. which satisfy $\eta(v, w) = \langle \psi_p^A(v), \psi_p^A(w) \rangle_{\mathbb{R}^d}$ for any $v, w \in T_p M$. Since orthonormal frames and isometric gauges are defined up to rotations and reflections, any gauge transformation between them will take values in the orthogonal group $O(d)$, which is that subgroup of $GL(d)$ that preserves angles and distances.
- $GL^+(d)$: Similarly, an *orientation* of the manifold distinguishes left-handed from right-handed frames and orientation preserving gauges from orientation reversing gauges. Gauge transformations between frames of a given handedness take values in $GL^+(d)$, that is, that subgroup of $GL(d)$ which preserves orientations.
- $SO(d)$: Together, a given *metric and orientation* specify orthonormal frames of a certain handedness. Gauge transformations between such frames are guaranteed to lie in the subgroup $SO(d)$ of $GL(d)$.
- $\{e\}$: A *globally smooth frame field* defines an $\{e\}$ -structure on M . In this case there is only one single distinguished frame at each position, such that gauge transformations lie in the trivial group $\{e\} \leq GL(d)$.
- $GL(d)$: If no additional structure is imposed, *any* reference frame of the tangent spaces is equally valid. Gauge transformations are in this case general invertible liner maps in $GL(d)$ and the corresponding G -structure is just the frame bundle FM .

structure group G	G -structure GM	equivalent structure on M
$\text{GL}(d)$	all reference frames, i.e. $GM = FM$	smooth structure only
$\text{GL}^+(d)$	positively oriented frames	orientation of M
$\text{SL}(d)$	unit volume frames	volume form
$\text{CO}(d)$	conformal frames	—
$\text{Sp}(d)$	symplectic frames	—
$O(d)$	orthonormal frames	Riemannian metric
$O(d - n, n)$	pseudo-orthonormal frames	pseudo-Riemannian metric
$\text{SO}(d)$	positively oriented orthonormal frames	Riemannian metric + orientation
$\{e\}$	parallelization (global frame field)	—

Table 7.1: Examples of G -structures GM on M and their corresponding reduced structure groups $G \leq \text{GL}(d)$ [159]. A G -structure is defined as a smoothly varying subset of reference frames (a principal G -subbundle of the frame bundle FM), where the frames of any tangent space are mutually related by G -valued gauge transformations. While this is a quite abstract definition, it allows to view many geometric structures on M in a unified way. For instance, a Riemannian metric on M allows to distinguish orthonormal frames. Conversely, a specification of orthonormality uniquely implies a metric. A Riemannian metric and an orthonormal structure are thus equivalent to each other. Similarly, there is a one-to-one correspondence between volume forms and unit volume frames. Note that a choice of structure group G does not uniquely specify a G -structure. For example, different Riemannian metrics could be chosen as $O(d)$ -structure, different volume forms as $\text{SL}(d)$ -structure or different global frame fields as $\{e\}$ -structure. Coordinate independent CNNs are designed to respect a given G -structure – which particular structure this depends on the learning task.

The common theme in those motivating examples is that they are all defined by

1. a (spatially smoothly varying) subset of distinguished reference frames,
2. a corresponding subset of preferred gauges and
3. a subgroup $G \leq \text{GL}(d)$ of gauge transformations which preserve the distinguished notion of frames and gauges.

Such smoothly varying subsets of reference frames are denoted as *G -structures* GM on M and the group G is denoted as (reduced) *structure group* – see Section 11.3 for a more rigorous definition.⁴ The process of specifying a G -structure is known as a *reduction of the structure group* from $\text{GL}(d)$ to G . An atlas $\mathcal{A}^G = \{(U^X, \psi^X)\}_{X \in \mathfrak{X}}$ is denoted as *G -atlas* if all of its transition functions

$$g^{BA} : U^A \cap U^B \rightarrow G, \quad p \mapsto g_p^{BA} := \psi_p^B \circ (\psi_p^A)^{-1} \quad (7.24)$$

lie in a reduced structure group $G \leq \text{GL}(d)$ (cf. Eq. (7.6)). The relation between reference frames and gauges in Eq. (7.4) implies that any G -atlas encodes a corresponding G -structure.

Multiple choices of G -structures may exist for a given structure group G . To connect to the examples above: different Riemannian metrics specify different subsets of reference frames as being orthonormal, that is, they correspond to different $O(d)$ -structures OM . A choice of metric is therefore equivalent to a choice of $O(d)$ -structure. Similarly, different choices of orientations of an orientable manifold specify a different set of frames as being

⁴Formally, GM is defined as a principal G -subbundle of the frame bundle FM , which is a principal $\text{GL}(d)$ -bundle.

right-handed. The two possible choices of orientations therefore correspond to two possible choices of $\mathrm{GL}^+(d)$ -structures GL^+M . $\mathrm{SO}(d)$ -structures SOM may differ in both the choice of orientation and metric. A further example are $\{e\}$ -structure $\{e\}M$. They do not allow for (non-trivial) gauge transformations and therefore correspond to choices of smooth, global frame fields on M . Table 7.1 gives more examples of structure groups G and the corresponding G -structures.

A reduction of the structure group to G , i.e. the existence of a G -structure, might be obstructed by the topology of the manifold. This implies that there is an “*irreducible*” *structure group* beyond which the ambiguity of reference frames can not be resolved without violating the smoothness (or even continuity) assumption of the G -structure. For example, the Möbius strip is non-orientable, which means that it does not admit a globally consistent, smooth definition of frame handedness and thus $\{e\}$ -structure (globally smooth frame field). As visualized in Fig. 10.1, a G -atlas of gauges covering the Möbius strip will unavoidably require a reflection in one of the transition maps, implying an irreducible structure group $G = \mathcal{R}$. Coordinate independent CNNs on the Möbius strip is therefore required to be at least reflection equivariant. Similarly, the structure group of the sphere can not be reduced further than $G = \mathrm{SO}(2)$. Smooth spherical CNNs are thus necessarily based on locally rotation equivariant kernels.

Note that *any* (differentiable) manifold comes with *some* G -structure. For instance, a raw differentiable manifold has a $\mathrm{GL}(d)$ -structure (containing any possible frame), a Riemannian manifold an $\mathrm{O}(d)$ -structure and \mathbb{R}^d is canonically equipped with an $\{e\}$ -structure, visualized in Fig. 13.3a. We will therefore without loss of generality refine the term “coordinate independence” to *GM-coordinate independence*, i.e. the independence w.r.t. choices of reference frames in the G -structure given on M . Throughout, we will assume that gauges are part of some G -atlas

$$\mathcal{A}^G = \{(U^X, \psi^X)\}_{X \in \mathfrak{X}} \quad \text{such that} \quad g_p^{BA} \in G \quad \forall \psi_p^A, \psi_p^B \in \mathcal{A}^G, \quad p \in U^A \cap U^B, \quad (7.25)$$

corresponding to the given G -structure. Any quantity or function can be expressed relative to any gauge from this atlas⁵, and the coordinatizations in different gauges relate uniquely by some G -valued gauge transformation. Guaranteeing the coordinate independence of all constructions, they will always correspond to some coordinate-free counterparts, in terms of which we will formulate the global theory in Part III.

⁵This is a non-trivial statement since not any quantity can be expressed relative to arbitrary $\mathrm{GL}(d)$ -related reference frames. For instance, the feature fields, introduced in Section 8.1, will only admit G -valued gauge transformations and are therefore only defined relative to the preferred frames in *GM*. As an intuitive example, consider the feature vectors of a conventional (non-equivariant) CNN on \mathbb{R}^d , which are extracted relative to the canonical $\{e\}$ -structure of \mathbb{R}^d and do *not* carry information about the kernel responses relative to other reference frames.

Coordinate independent feature vector fields

The feature spaces of coordinate independent CNNs are spaces of feature vector fields. Similar to the case of tangent vector coefficients, the numerical coefficients of feature vectors are required to transform consistently under gauge transformations. The specific transformation law (group representation) of a feature field does hereby specify its *field type*, – common examples are scalar fields, tangent vector fields, tensor fields, regular feature fields or irrep fields. A field's type determines furthermore how feature vectors are parallel transported or acted on by isometries.

The goal of this chapter is to define coordinate independent feature fields and their geometric properties. Section 8.1 introduces feature vectors and their *gauge transformation laws*. *Parallel transporters* of feature vectors and their representation relative to different coordinatizations are introduced in Section 8.2. Section 8.3 discusses *isometries* and their action on geometric quantities like tangent and feature vectors.

The coordinate independent feature fields described here are the differential geometric generalization of the Euclidean feature fields from Section 4.2. The coordinate free definition of feature vector bundles is given in Section 11.3.3 below.

8.1 Gauge transformations of feature vectors

Convolutional feature fields assign a feature vector, encoding information inferred from a local neighborhood of the input signal, to each point of the manifold. The spatial accumulation of information is performed by a convolutional kernel which is *measuring feature fields* in its surrounding *relative to its local reference frame*. We are thus assuming a gauge ψ^A which specifies the kernel alignments on a neighborhood U^A . Relative to this gauge the kernel will yield a smooth local field of responses (observations)

$$f^A : U^A \rightarrow \mathbb{R}^c, \quad (8.1)$$

given by a c -dimensional numerical feature vector $f^A(p)$ at each position $p \in U^A$. Assume a second response field $f^B : U^B \rightarrow \mathbb{R}^c$, inferred relative to gauge ψ^B on U^B , to be given. Since the response of a kernel depends in general on its alignment, it is to be expected that f^A and f^B do not agree on the overlap $U^A \cap U^B$. Without further restrictions the responses of a convolution kernel will be arbitrarily gauge dependent.

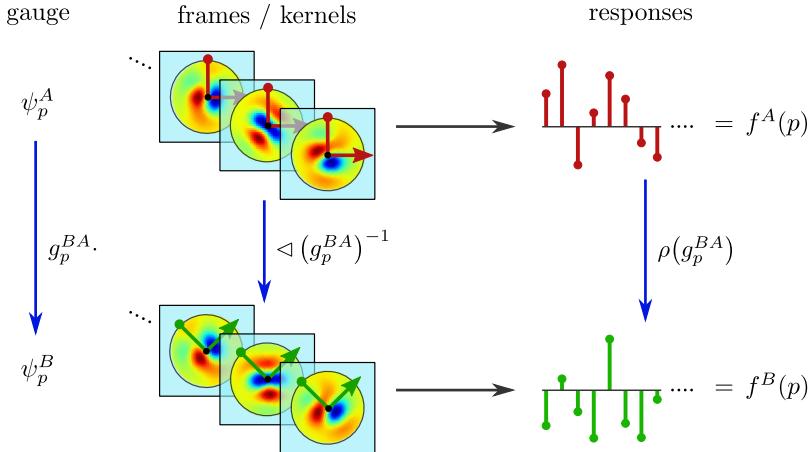


Figure 8.1: The numerical responses $f^A(p) \in \mathbb{R}^c$ and $f^B(p) \in \mathbb{R}^c$ of kernels that are oriented according to different frames do in general not coincide. In order to represent numerical coefficients of the *same coordinate independent feature vector* relative to the chosen gauge, they are required to be related by gauge transformations $\rho(g_p^{BA})$ if the gauges are related by g_p^{BA} . As derived in Chapter 9, this requirement imposes a gauge equivariance constraint on the convolution kernels.

The *principle of covariance*, proposed by Albert Einstein [80, 79], states that:

“Universal laws of nature are to be expressed by equations which hold good for all systems of coordinates, that is, are covariant with respect to any substitutions whatever.”

We believe that a similar principle should hold in geometric deep learning as well, that is, the inference should be independent from any arbitrariness in the choice of reference frames. Given that this arbitrariness in coordinatizations is precisely captured by the given G -structure GM , this requires in particular that *features should be GM -coordinate independent* geometric objects.¹ We thus design convolution kernels such that their responses f^A and f^B encode fields of *feature vector coefficients* which *represent a coordinate free feature vector field f* locally in different gauges. A collection of such numerical coefficient fields f^X , expressed relative to a G -atlas of gauges ψ^X on neighborhoods U^X covering M , is equivalent to the global, coordinate free feature field f on M .

In order for this coordinate free feature field to be well defined, i.e. GM -coordinate independent, the local coefficient fields (or kernel responses) are required to be consistently stitched together via G -valued transition maps. They must therefore transform in a principled manner under gauge transformations. Since we are dealing with feature vector spaces, these transformations are typically taken to be linear, that is, they are modeled by G -representations

$$\rho : G \rightarrow \mathrm{GL}(c). \quad (8.2)$$

¹In this point we deviate from Einstein’s *general covariance*, which always considers $\mathrm{GL}(d)$ -valued gauge transformations (corresponding to diffeomorphism covariance). His setting is in our formulation included for $G = \mathrm{GL}(d)$, however, we keep the assumed structure group flexible since most applications will assume a reduced structure group.

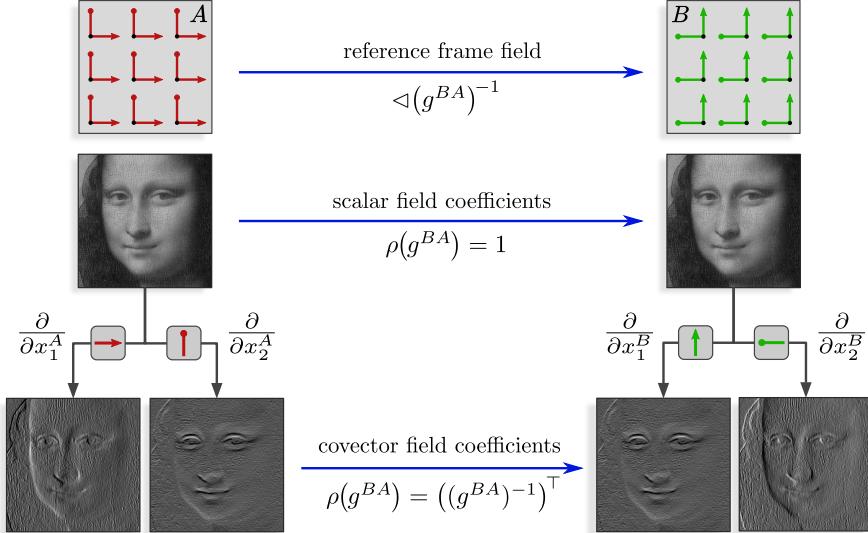


Figure 8.2: Examples of feature coefficient fields on $M = \mathbb{R}^2$ from classical image processing. *Top:* For simplicity we assume a “parallel” frame field and consider the same gauge transformation, a rotation by $\pi/2$, at each point $p \in M$. *Middle:* The intensity values of a grayscale image are independent from the choice of reference frames. They are therefore modeled by scalar fields, characterized by the trivial representation $\rho(g) = 1 \forall g \in G$. *Bottom:* The two coefficient channels of a gradient image are calculated from a scalar image by taking the derivatives along the frame axes – they are therefore gauge dependent. Gradient images w.r.t. different gauges are related by the group representation $\rho(g) = (g^{-1})^\top$ and are therefore identified as covector fields (tensor fields of type $(0, 1)$ or 1-forms). For the visualized rotation by $\pi/2$ this leads to a new first channel $(\partial/\partial x_1^B)$ equivalent to the old second channel $(\partial/\partial x_2^A)$ and a new second channel $(\partial/\partial x_2^B)$ equivalent to the negative old first channel $(\partial/\partial x_1^A)$. Relative to their respective reference frames, both coefficient fields encode the same (coordinate free) gradient field. The description is therefore automatically coordinate independent.

Similar to the transformation of tangent vector coefficients in Eq. (7.9), the feature vector coefficients are then defined to transform under a G -valued gauge transformation g_p^{BA} according to

$$f^B(p) := \rho(g_p^{BA}) f^A(p), \quad (8.3)$$

where $p \in U^A \cap U^B$; see Fig. 8.1 for a visualization. Being constructed to transform synchronously, the spaces of reference frames, tangent vector coefficients and feature vector coefficients are said to be G -associated to each other. Note that the construction via a G -representation ρ does in general *not* describe $GL(d)$ -valued gauge transformations, i.e. fully coordinate independent features. The extracted feature vectors will therefore only have a well defined expression relative to the frames in the considered G -structure GM , which is captured by the term “ GM -coordinate independence” (or G -covariance).

Bundle description: For completeness we briefly mention that coordinate free feature vector fields will in Part III be more formally defined as smooth sections $f \in \Gamma(\mathcal{A})$ of a *feature vector bundle* $\mathcal{A} \xrightarrow{\pi_A} M$ which is associated to the G -structure GM and has the feature vector coefficient spaces \mathbb{R}^c as typical fibers. The coefficient vectors $f^A(p)$ and $f^B(p)$ in \mathbb{R}^c are local trivializations of a coordinate free feature vector $f(p) \in \mathcal{A}_p \cong \mathbb{R}^c$, and are similarly

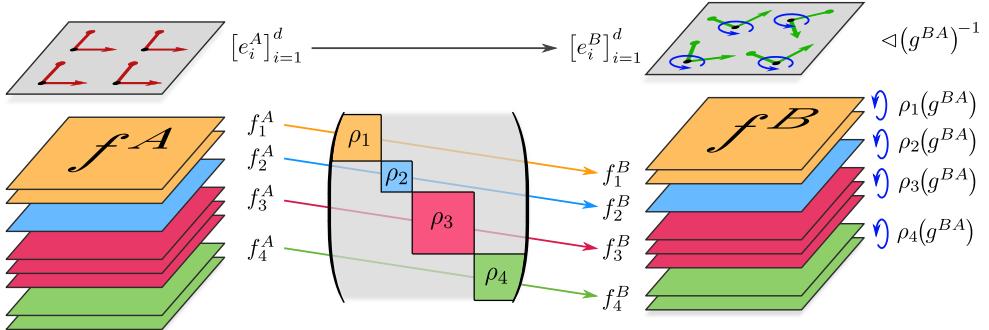


Figure 8.3: As for steerable CNNs (Chapter 4), the full feature spaces of coordinate independent CNNs consist of multiple independent feature fields f_i of potentially different types ρ_i and dimensionalities c_i . The individual fields are in gauge ψ^A locally represented by coefficient fields $f_i^A : U^A \rightarrow \mathbb{R}^{c_i}$ and relate to coefficient fields in another gauge ψ^B via local gauge transformations $f_i^B = \rho_i(g^{BA})f_i^A$. The representation modeling the whole feature space is given by a direct sum, here $\bigoplus_i \rho_i = \rho_1 \oplus \rho_2 \oplus \rho_3 \oplus \rho_4$, which reflects that the coefficients of each individual field transform independently (i.e. do not mix). This independent transformation behavior of the individual coefficient fields applies to active transformations (isometry pushforwards) as well; cf. Fig. 4.4.

defined as the coefficients $v^A = \psi_p^A(v)$ and $v^B = \psi_p^B(v)$ of a tangent vector $v \in T_p M$. Note that, while being isomorphic, the feature spaces $\mathcal{A}_p \cong \mathcal{A}_q$ at different points $p \neq q$ of M are distinct from each other, such that their elements can not be summed together. The parallel transporters, discussed in Sections 8.2 and 11.5, provide isomorphisms between different feature vector spaces, which allows the summation of features (after transporting them into the same vector space). Since these definitions are quite technical, we skip their details for now and refer the interested reader to Section 11.3.

Stacked feature fields and coordinate independent feature spaces: As for steerable CNNs, we define the feature spaces of coordinate independent CNNs as comprising multiple feature fields f_i of potentially different types ρ_i and dimensionalities c_i . A full field of activations of a feature space is therefore defined as the direct sum $f = \bigoplus_i f_i$. Its local numerical representations $f^X = \bigoplus_i f_i^X$ transform according to the direct sum $\rho = \bigoplus_i \rho_i$ of the individual field types. The individual fields transform by construction independently from each other:

$$\rho(g^{BA})f^A = [\bigoplus_i \rho_i(g^{BA})][\bigoplus_i f_i^A] = \bigoplus_i [\rho_i(g^{BA})f_i^A] \quad (8.4)$$

Fig. 8.3 visualizes the independent passive transformation of a direct sum of coefficient fields under local gauge transformations. Compare this with the situation in Fig. 4.4, where we instead visualized the active transformation of a feature coefficient field in a fixed gauge. This active transformation viewpoint is in the differential geometric setting described by pushforward actions of the isometry group, which are covered in Section 8.3 below.

For specific examples of stacked coordinate independent feature fields, we refer the reader back to the end of Section 4.2.

8.2 Parallel transport of feature vectors

The kernels of convolutional networks accumulate features from all points q in a neighborhood around each point p of the manifold. Since features at different points live in different feature vector spaces they need to be *parallel transported* along some path γ from q to p before they can be processed further. We first discuss the transport of tangent vectors, which is formalized by a parallel transport map

$$\mathcal{P}_\gamma : T_q M \rightarrow T_p M. \quad (8.5)$$

This transporter is often computed from the canonical Levi-Civita connection of the manifold, however, it might in some applications correspond to an alternative (G -compatible) connection, as further discussed below and in the discussion of applications in Part IV. A transporter of (G -associated) feature vectors follows from that of the tangent vectors if the transport is G -compatible.

8.2.1 Tangent vector transporters

It is didactically reasonable to start with the specific case of Levi-Civita transporters on Euclidean spaces, depicted in Fig. 8.4a, before proceeding to more general transporters and manifolds. In this case the parallel transport is independent from the chosen path γ and keeps the transported vector parallel in the usual sense on Euclidean spaces. Note that the transporter \mathcal{P}_γ is a *coordinate free* map between the tangent spaces $T_q M$ and $T_p M$. It can, however, be expressed relative to coordinates, then operating on numerical coefficient vectors instead of tangent vectors. An intuition is given in Fig. 8.4a, where the frames at q and p are not parallel² such that the coefficients $(1, 1)^\top$ at q and $(\sqrt{2}, 0)^\top$ at p differ even though the corresponding (coordinate free) tangent vectors are parallel to each other. To make this more precise, consider gauges $\psi_q^{\tilde{A}}$ and ψ_p^A to be given on neighborhoods $U^{\tilde{A}}$ of q (red) and U^A of p (green). Let a vector $v = (\psi_q^{\tilde{A}})^{-1}(v^{\tilde{A}}) \in T_q M$ be given by its coefficients $v^{\tilde{A}} \in \mathbb{R}^d$. The coefficients of the transported vector $\mathcal{P}_\gamma v$ at p are then given by $\psi_p^A \circ \mathcal{P}_\gamma(v) = \psi_p^A \circ \mathcal{P}_\gamma \circ (\psi_q^{\tilde{A}})^{-1}(v^{\tilde{A}})$. It follows that the coordinate expression of a transporter is relative to gauges \tilde{A} and A expressed as:³

$$g_\gamma^{A\tilde{A}} := \psi_p^A \circ \mathcal{P}_\gamma \circ (\psi_q^{\tilde{A}})^{-1} \in \mathrm{GL}(d) \quad (8.6)$$

The group element $g_\gamma^{A\tilde{A}}$ accounts for non-parallel choices of reference frames at q and p . On \mathbb{R}^d , one typically assumes all frames to be parallel such that all coordinatizations of Levi-Civita transporters become trivial – conventional and steerable Euclidean CNNs implicitly make this assumption of parallel frames (Fig. 1.6a) and trivial transporters, which explains why they don't appear in their mathematical formulation.

As the transporter in Eq. (8.6) is coordinate dependent, we are interested in its gauge transformations. Denote by $\psi_q^{\tilde{B}}$ and ψ_p^B two alternative gauges on neighborhoods of q and p .

²In contrast to general manifolds, \mathbb{R}^d comes with a canonical notion of parallelism of frames.

³ $g_\gamma^{A\tilde{A}}$ takes values in $\mathrm{GL}(d)$ if we assume arbitrary ($\mathfrak{gl}(d)$ -valued) connections and general structure groups $G \leq \mathrm{GL}(d)$. For the $\mathfrak{so}(d)$ -valued Levi-Civita connection and orthonormal frames, i.e. $G = \mathrm{O}(d)$, one has $g_\gamma^{A\tilde{A}} \in \mathrm{O}(d)$.

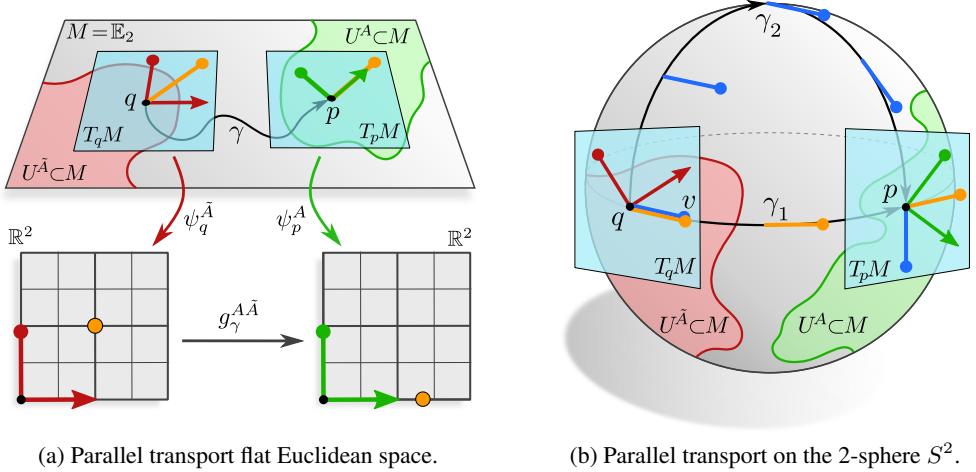


Figure 8.4: Parallel transport of tangent vectors $v \in T_q M$ at q to $\mathcal{P}_\gamma v \in T_p M$ at p . Fig. 8.4a visualizes the special case of Levi-Civita transporters on flat Euclidean spaces $M = \mathbb{E}_2$. Independently from the chosen path γ , the Levi-Civita transport keeps the vector (orange) parallel in the usual sense in Euclidean spaces. Gauges $\psi_q^{\tilde{A}}$ (red) and ψ_p^A (green) allow to express the coordinate free transporter by a group element $g_\gamma^{AA} = \psi_p^A \circ \mathcal{P}_\gamma \circ (\psi_q^{\tilde{A}})^{-1} \in \mathrm{GL}(d)$ which accounts for the change of vector coefficients if the target frame does not agree with the transported source frame. Fig. 8.4b shows the Levi-Civita transport on the 2-sphere S^2 , Eq. (17.18). The transporters \mathcal{P}_{γ_1} and \mathcal{P}_{γ_2} along different paths γ_1 and γ_2 disagree in general. As before, the coordinate free transporters can be expressed by group elements that operate on coefficients relative to the coordinate frames at q and p .

From the commutative diagram

$$\begin{array}{ccccc}
 & & g_\gamma^{A\tilde{A}}. & & \\
 & \swarrow & & \searrow & \\
 \mathbb{R}^d & & T_q M & & \mathbb{R}^d \\
 & \downarrow g_q^{\tilde{B}\tilde{A}}. & \xrightarrow{\mathcal{P}_\gamma} & \downarrow g_p^{BA}. & \\
 & \uparrow \psi_q^{\tilde{A}} & & \uparrow \psi_p^A & \\
 & \uparrow \psi_q^{\tilde{B}} & & \uparrow \psi_p^B & \\
 & & g_\gamma^{B\tilde{B}}. & &
 \end{array} \tag{8.7}$$

one can then read off that the transporters in the different gauges are related by

$$g_\gamma^{B\tilde{B}} = g_p^{BA} g_\gamma^{A\tilde{A}} (g_q^{\tilde{B}\tilde{A}})^{-1} \tag{8.8}$$

Note the similarity of this transformation law and commutative diagram to those in Eqs. (7.21) and (7.23). The difference between both is that the transporter has a different domain $T_q M$ and codomain $T_p M$, which are trivialized by different, independent gauges and transform therefore independently.

In general, the parallel transport of tangent vectors is determined by some choice of connection, for instance (but not necessarily) by the canonical Levi-Civita connection of a Riemannian manifold. A connection can be seen as a collection of infinitesimal transporters between

adjacent tangent spaces, such that the full transporter \mathcal{P}_γ is given by integrating the connection along the path γ . The transporters along different paths γ_1 and γ_2 from q to p need not agree, which is in Fig. 8.4b exemplified by Levi-Civita transporters on the 2-sphere S^2 , cf. Eq. (17.18). As for flat spaces, the coordinate free transporters can by Eq. (8.6) be expressed relative to gauges. The gauge transformations of such coordinatized transporters are again given by Eq. (8.8). The transporters on a given manifold can in principle be calculated analytically from the connection [100, 221] and can sometimes be expressed in closed form, for instance for the sphere S^2 , Eq. (17.18). Several numerical algorithms exist to compute parallel transporters on meshes; see Section 18.1.2. We will not go into more details on how to compute tangent vector transporters \mathcal{P}_γ but simply assume them to be given.

8.2.2 Feature vector transporters

Eq. (8.3) defines the transformation law of feature vector coefficients by their *field type* ρ . Their parallel transporter, expressed relative to gauges $\psi_g^{\tilde{A}}$ and ψ_p^A , is analogously given by wrapping the tangent vector coefficient transporter into this field representation, that is, by

$$\rho(g_\gamma^{A\tilde{A}}). \quad (8.9)$$

Note that – since $\rho : G \rightarrow \text{GL}(c)$ is a G -representation – this construction is only then well defined when all transporters $g_\gamma^{A\tilde{A}}$ (for arbitrary paths γ and frames A, \tilde{A}) are actually taking values in the chosen structure group G . Whether this is the case depends both on the particular choice of G -structure (or G -atlas) and the transporters (or connection) considered – they need to be *compatible* [326].

All convolutional networks accumulate (thus transport) feature vectors in one way or the other, and assume therefore some choice of connection and G -structure. If the chosen G -structure is incompatible with the Levi-Civita connection, this implies that these models are – often implicitly – assuming an alternative, G -compatible connection to accumulate features. The reader should for now not worry about the specific choices of connections, which will become more clear when reviewing specific applications in Part IV. In the remainder of this section, we will elaborate more on the G -compatibility of connections and G -structures. Assuming that *feature transporters will in the following always be well defined*, this part can be safely ignored at a first reading.

A more rigorous, coordinate free discussion of transporters on the associated feature vector bundles can be found in Section 11.5.

8.2.3 Compatibility of connections and G -structures

A connection is said to be *G -compatible* with a G -structure GM if the coordinate expressions $g_\gamma^{A\tilde{A}}$ of its transporters \mathcal{P}_γ relative to any frames A, \tilde{A} of GM take values in the structure group G [326].⁴ A G -compatible connection gives rise to transporters of G -associated feature vectors.

⁴Equivalently, the *connection 1-form* of the connection, expressed relative to frames of GM , is required to be \mathfrak{g} -valued, where \mathfrak{g} denotes the Lie algebra of G . More abstractly, we are interested in *principal Ehresmann connections* on the principal G -bundle GM .

To illuminate this somewhat abstract compatibility condition, we discuss a few specific examples. A simple example is that of the Levi-Civita connection on \mathbb{R}^2 , Fig. 8.4a, Consider the two $\{e\}$ -structures on \mathbb{R}^2 that are shown in Figs. 13.3a and 13.3b. Here $G = \{e\}$, which means that the field type $\rho : \{e\} \rightarrow \text{GL}(c)$ is a $\{e\}$ -representation, such that the parallel transport of feature vectors can only be defined if the coordinate expressions $g_\gamma^{A\tilde{A}}$ take values in $\{e\}$, i.e. are trivial. As the $\{e\}$ -structure in Fig. 13.3a consists of “parallel” frames, this is indeed the case – the Levi-Civita connection is thus compatible with this $\{e\}$ -structure. In contrast, the frames of the $\{e\}$ -structure in Fig. 13.3b are “rotated” relative to each other, resulting in non-trivial coordinate expressions $g_\gamma^{A\tilde{A}}$ that take values in $\text{SO}(2)$ (visualized in Fig. 8.4a). Since the field type $\rho : \{e\} \rightarrow \text{GL}(c)$ does not handle rotations, it is not possible to define the Levi-Civita transport of features associated to this $\{e\}$ -structure – they are incompatible. As a second example, consider the Levi-Civita connection on S^2 , shown in Fig. 8.4b. The transport will in this case always be path dependent and lead to differently rotated vectors, implying that $g_\gamma^{A\tilde{A}}$ will take values in $\text{SO}(2)$. Feature vectors to be transported according to Levi-Civita connection need therefore to be of some type $\rho : \text{SO}(2) \rightarrow \text{GL}(c)$ that is an $\text{SO}(2)$ -representations. This requires at least the $\text{SO}(2)$ -structure on S^2 that is shown in Fig. 17.2a. The $\{e\}$ -structure on S^2 from Fig. 17.2b is incompatible with the Levi-Civita connection.

Since the Levi-Civita connection is a metric connection, it preserves lengths of and angles between tangent vectors, and thus transports orthonormal frames to orthonormal frames. It follows that the Levi-Civita connection is *always* compatible with the $\text{O}(d)$ -structure of orthonormal frames, relative to which $g_\gamma^{A\tilde{A}}$ takes values in $\text{O}(d)$. If the manifold is orientable, the frame-handedness is preserved by Levi-Civita transporters, which means that they are guaranteed to be compatible with $\text{SO}(d)$ -structures of orthonormal, right-handed frames on M . All of the convolutional networks in our literature review in Part IV that are based on $\text{SO}(d)$ -structures accumulate features via Levi-Civita transporters.

If a given G -structure is incompatible with the Levi-Civita connection, one needs to define an alternative, G -compatible connection to transport the feature vectors. The most prominent example in our literature review is that of *trivial connections* on $\{e\}$ -structures. A trivial connection is characterized by the property that its transport is *path independent* [62]. Any $\{e\}$ -structure implies a unique trivial connection, which transports tangent vectors such that they keep the same angle to the reference frames of the $\{e\}$ -structure. This implies $g_\gamma^{A\tilde{A}} = e$, i.e. they transport coefficient vectors in \mathbb{R}^c (relative to frames of the $\{e\}$ -structure) without transforming their numerical values. Such transporters are used in convolutional networks that do not explicitly model non-trivial transporters – which applies to *all* networks with $G = \{e\}$ in Table 14.1, specifically those in Sections 17.3 and 18.3. Note that the trivial connection is the only connection that is compatible with an $\{e\}$ -structure.

As stated above, *any* convolutional network assumes some choice of compatible G -structure and connection, most often Levi-Civita connections or trivial connections.

Section 11.5 elaborates on the compatibility of transporters and G -structures from a coordinate free viewpoint.

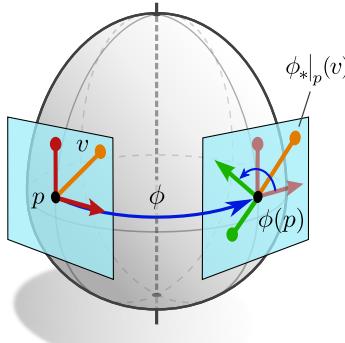


Figure 8.5: Visualization of the coordinate free pushforward of tangent vectors and its coordinate expression relative to given reference frames at the source and target location. The coordinate free pushforward $\phi_*|_p : T_p M \rightarrow T_{\phi(p)} M$ moves tangent vectors $v \in T_p M$ to $\phi_*|_p(v) \in T_{\phi(p)} M$ (orange). Let $\psi_p^{\tilde{A}}$ be the gauge at p that corresponds to the red reference frame and $\psi_{\phi(p)}^A$ the gauge at $\phi(p)$ that corresponds to the green reference frame. They explain the vectors before and after the pushforward by numerical coefficients $\psi_p^{\tilde{A}}(v) = (1, 1)^\top$ and $\psi_{\phi(p)}^A(\phi_*|_p(v)) = (0, -\sqrt{2})^\top$. This transformation of vector coefficients is described by the isometry induced gauge transformation $g_\phi^{A\tilde{A}}(p) \in \text{GL}(d)$, that is, $\psi_{\phi(p)}^A(\phi_*|_p(v)) = g_\phi^{A\tilde{A}}(p) \cdot \psi_p^{\tilde{A}}(v)$. The coefficients of feature vectors transform analogously according to $\rho(g_\phi^{A\tilde{A}}(p))$ if $g_\phi^{A\tilde{A}}(p) \in G$.

8.3 Isometry actions and induced gauge transformations

Until now our discussion focused exclusively on the local gauge symmetries in the coordinatization of tangent spaces. A manifold might, however, come with non-trivial symmetries itself, which are in the case of a Riemannian manifold M forming its *isometry group* $\text{Isom}(M)$. This section discusses isometries and their action on manifolds, tangent vectors, reference frames and feature fields in a nutshell, summarizing results which are more rigorously derived in Section 13.1. We will thereby highlight the equivalence of *active* isometry actions and their *passive* interpretation in terms of isometry induced gauge transformations. This equivalence will later on allow us to describe the isometry equivariance of *GM*-convolutions.

Isometries are defined as the symmetries of Riemannian manifolds, that is, those maps (diffeomorphisms)

$$\phi : M \rightarrow M, \quad (8.10)$$

that preserve the metric and thus distances on M . The set of all isometries of a Riemannian manifold M forms its isometry group, which we denote as $\text{Isom}(M)$. For instance, the Euclidean group $E(d)$ is the isometry group of Euclidean spaces \mathbb{E}_d . It consists of translations, rotations and reflections, all of which preserve the standard metric of \mathbb{E}_d . The isometry group of the 2-sphere S^2 is given by the orthogonal group $O(3)$, consisting of rotations and reflections. Fig. 8.5 shows an egg-shaped manifold, whose isometries are rotations and reflections in $O(2)$ around the vertical axis.

8.3.1 Pushforward of tangent vectors

Any isometry $\phi \in \text{Isom}(M)$ acts via its *pushforward* (or differential)

$$\phi_*|_p : T_p M \rightarrow T_{\phi(p)} M, \quad (8.11)$$

naturally on tangent vectors. The pushforward can intuitively be thought of as carrying tangent vectors along with the action of the isometry on the underlying manifold M . A formal definition of the pushforward on TM is given in Appendix C.2, however, the given

intuition is sufficient for our purpose. Since the pushforward is a coordinate free, linear map between tangent spaces, its action is in coordinates represented by some $d \times d$ matrix. Assuming gauges $\psi_p^{\tilde{A}}$ and $\psi_{\phi(p)}^A$ at the source and target location, respectively, this matrix is given by

$$g_{\phi}^{A\tilde{A}}(p) := \psi_{\phi(p)}^A \circ \phi_*|_p \circ (\psi_p^{\tilde{A}})^{-1} \in \text{GL}(d). \quad (8.12)$$

It explains the transformation from the numerical coefficients of an original vector $v \in T_p M$ in the source gauge and its pushforward $\phi_*|_p(v) \in T_{\phi(p)} M$ in the target gauge, that is, $\psi_{\phi(p)}^A(\phi_*|_p(v)) = g_{\phi}^{A\tilde{A}}(p) \cdot \psi_p^{\tilde{A}}(v)$. The commutative diagram

$$\begin{array}{ccccc} & & g_{\phi}^{A\tilde{A}}(p) \cdot & & \\ & \swarrow & & \searrow & \\ \mathbb{R}^d & & T_p M & & \mathbb{R}^d \\ \downarrow g_p^{\tilde{B}\tilde{A}} \cdot & \nearrow \psi_p^{\tilde{A}} & \xrightarrow{\phi_*|_p} & \nearrow \psi_{\phi(p)}^A & \downarrow g_{\phi(p)}^{BA} \cdot \\ & \swarrow \psi_p^{\tilde{B}} & & \searrow \psi_{\phi(p)}^B & \\ & \mathbb{R}^d & & & \mathbb{R}^d \\ & & g_{\phi}^{B\tilde{B}}(p) \cdot & & \end{array}, \quad (8.13)$$

which is conceptually similar to that in Eq. (8.7), visualizes the definition of the tangent vector pushforward's coordinate expression. It furthermore implies that the gauge transformations between different coordinatizations are given by

$$g_{\phi}^{B\tilde{B}} = g_{\phi(p)}^{BA} g_{\phi}^{A\tilde{A}}(g_p^{\tilde{B}\tilde{A}})^{-1}, \quad (8.14)$$

which is the conceptual analog to Eq. (8.8).

8.3.2 Pushforward of reference frames and symmetries of the G -structure

Since reference frames are just d -tuples of linearly independent frame vectors, the pushforward of tangent vectors induces a pushforward of reference frames by *pushing the individual frame axes forward*. Specifically, the pushforward of a frame $[e_i]_{i=1}^d$ at p is defined as the frame $[\phi_*|_p(e_i)]_{i=1}^d$ at $\phi(p)$.

This pushforward of frames is always well defined, however, it might not be compatible with the G -structure, that is, there is in general no guarantee that frames in GM remain in GM when being pushed forward. Take for instance the $\{e\}$ -structure in Fig. 1.12 (top left), which is preserved by horizontal translations but not by vertical translations or any other isometry of \mathbb{R}^2 . Similarly, the \mathcal{R} -structure in Fig. 1.12 (bottom left) is preserved by translations and horizontal reflections, but not by rotations. We consider therefore the subgroup

$$\text{Isom}_{GM} := \left\{ \phi \in \text{Isom}(M) \mid [\phi_*(e_i)]_{i=1}^d \in GM \quad \forall [e_i]_{i=1}^d \in GM \right\} \leq \text{Isom}(M) \quad (8.15)$$

of *isometries which are symmetries of the G -structure*, i.e. which are guaranteed to map any frame in GM to another frame that is also contained in GM .⁵ Note that Isom_{GM} depends in

⁵More formally stated, such isometries are (or induce) *principal bundle automorphisms* of the G -structure.

general on the specific choice of G -structure GM , not only on the structure group G . For the special case that $G \geq O(d)$, it is guaranteed that $\text{Isom}_{GM} = \text{Isom}(M)$ coincide since isometries are guaranteed to map orthonormal frames to orthonormal frames. We are interested in the subgroup Isom_{GM} since only those isometries will induce a well defined pushforward of GM -coordinate independent feature vectors, as discussed further in the following section.

Before proceeding to the isometry action on feature vectors, we discuss what we call *isometry induced gauge transformations*. For this purpose, let $[e_i^{\tilde{A}}]_{i=1}^d$ be that frame at p that corresponds to some source gauge $\psi_p^{\tilde{A}}$ and let $[e_i^A]_{i=1}^d$ be that frame at $\phi(p)$ that corresponds to some target gauge $\psi_{\phi(p)}^A$, as shown in Fig. 8.5 in red (left) and green (right), respectively.

The pushforward $[\phi_*|_p(e_i^{\tilde{A}})]_{i=1}^d$ of the source frame from p to $\phi(p)$ (translucent red, right) does in general not coincide with the target frame. However, as proven in Section 13.1.3, the two frames are related by the isometry induced gauge transformation

$$[\phi_*|_p(e_i^{\tilde{A}})]_{i=1}^d = [e_i^A]_{i=1}^d \triangleleft g_{\phi}^{A\tilde{A}}(p), \quad (8.16)$$

where $g_{\phi}^{A\tilde{A}}(p)$ is the group element from Eq. (8.12) and \triangleleft is the right action from Eq. (7.11). The term “isometry induced gauge transformation” makes in so far sense that the geometries around p and $\phi(p)$ are indistinguishable since ϕ is an isometry, i.e. a symmetry of M . Identifying the two points with each other, one can therefore reinterpret the *active* action of ϕ on a geometric quantity as a *passive* gauge transformation, i.e. an induced change from the source to the target frame.

Theorem 13.1.3 in Section 13.1 asserts that G -structure preserving isometries in Isom_{GM} and G -valued induced gauge transformations imply each other, that is,

$$\phi \in \text{Isom}_{GM} \iff g_{\phi}^{A\tilde{A}}(p) \in G \quad \forall p \in M \quad (8.17)$$

holds for arbitrary gauges $\psi_p^{\tilde{A}}$ and $\psi_{\phi(p)}^A$ of the G -atlas. The reader should verify these claims at our examples in Fig. 1.12.

8.3.3 Pushforward of feature vectors

If (and only if) an isometry is a symmetry of the G -structure, it gives rise to a *pushforward of feature vectors*. Intuitively, this pushforward moves feature vectors from points p to $\phi(p)$. When being expressed relative to the two reference frames at p and $\phi(p)$, it is given by the induced gauge transformation

$$\rho(g_{\phi}^{A\tilde{A}}(p)). \quad (8.18)$$

Note that this transformation is well defined for any $\phi \in \text{Isom}_{GM}$, since the induced gauge transformations $g_{\phi}^{A\tilde{A}}(p)$ will in this case take values in G and ρ is a G -representation. In contrast, if ϕ is not a symmetry of the G -structure, it is *impossible* to define a corresponding feature vector pushforward. This statement relates to the fact that the features of conventional CNNs have no specified transformation behavior under rotations or reflections in the Euclidean group $E(d)$.

The pushforward of individual feature vectors implies an action on the whole feature field f , which we denote by $\phi \triangleright f$. Relative to coordinates, this action is expressed as

$$[\phi \triangleright f]^A(\phi(p)) = \rho(g_{\phi}^{A\tilde{A}}(p)) f^{\tilde{A}}(p). \quad (8.19)$$

Note the similarity of this action to the *induced representation action* on Euclidean feature fields in Eq. (4.3).

We will later prove that coordinate independent CNNs are equivariant w.r.t. the action of isometries in Isom_{GM} on feature fields; see Fig. 9.3. This property relies on the fact that the active isometry action on feature fields can by Eq. (8.19) be understood as a mere passive gauge transformation of feature vector coefficients.

Coordinate independent networks and *GM*-convolutions

Neural networks process data by applying a series of parameterized mappings (layers) to an input signal – in our case to a set of feature fields on a Riemannian manifold. The *principle of covariance* requires thereby that the individual network layers should be *GM*-coordinate independent operations. The coordinate representations of such layers will therefore have to transform such that they respect the transformation laws of their input- and output feature field. Except from this consistency requirement, general coordinate independent layers remain *unconstrained*.

A common design principle of neural networks which operate on spatial signals (feature fields) is that they are in some generalized sense convolutional. The main characteristic which most generalizations of the convolution operation share is that their inference is *position-independent*. This is achieved by *sharing template functions* (neural connectivity), for instance convolution kernels or biases, between different locations. Whenever the structure group G is non-trivial, the weight sharing process is ambiguous since template functions could be shared relative to different reference frames. As we will prove in this chapter, this ambiguity requires the shared template functions to be *equivariant under G -valued gauge transformations* (G -steerable). Gauge steerable template functions will be indifferent to the specific reference frame in which they are applied and therefore allow for a coordinate independent weight sharing. Intuitively, G -steerability can be thought of as extending spatial weight sharing to a weight sharing over the full G -structure, i.e. additionally over G -transformations of frames.

In essence, neither coordinate independence nor weight sharing *alone* require the steerability of the neural connectivity, but *together* they do:

$$\left. \begin{array}{l} \text{GM-coordinate independence} \\ \text{spatial weight sharing} \end{array} \right\} \rightarrow G\text{-steerability (gauge equivariance)}$$

To emphasize this distinction, all of the network layers in this chapter are introduced in a two step process: first, we only demand their coordinate independence, i.e. investigate their coordinate representations relative to different frames. Subsequently, we require spatial weight sharing, from which steerability constraints follow.

In this chapter we will consider network layers which take fields f_{in} of type ρ_{in} as input and produce field f_{out} of type ρ_{out} as output. Section 9.1 discusses the specific case of layers which operate *pointwise*, that is, whose output $f_{\text{out}}(p)$ at any $p \in M$ depends only on the single input feature vector $f_{\text{in}}(p)$ at the same location. The practically relevant examples

considered here are gauge equivariant 1×1 -convolutions in Section 9.1.1, bias summation in Section 9.1.2 and nonlinearities in Section 9.1.3.

The more complicated case of *convolutions with spatially extended kernels* is treated in Section 9.2. As a preparation, Section 9.2.1 discusses feature fields as seen from the viewpoints of local observers (reference frames), relative to which the (convolution) kernels will be applied. Such observations are formalized as a pullback of the feature field to an observer's tangent space; see Fig. 9.1. Section 9.2.2 introduces so-called *kernel field transforms*, which are similar to convolutions but do not assume spatial weight sharing and are therefore parameterized by a (smoothly varying) kernel field on M . The actual *GM*-convolutions are in Section 9.2.3 defined as those kernel field transforms that are parameterized by a single, shared template kernel. In order to ensure the coordinate independence of the weight sharing process, the convolution kernels are required to be *G-steerable*, i.e. to satisfy a gauge equivariance constraint.

Section 9.3 shows that *GM*-convolutions are automatically equivariant under those *isometries* that are symmetries of the *G*-structure (Isom_{GM} -equivariant). This means that *GM*-convolutions commute with the action of isometries on feature fields as visualized in Fig. 9.3.

9.1 Pointwise gauge equivariant operations

To begin with, we consider some neural network operations for which the constraints coming from the required coordinate independence and weight sharing are particularly easy to derive. All of these operations have in common that they act pointwise on feature vectors, that is, they compute output feature vectors $f_{\text{out}}(p)$ at $p \in M$ solely based on the input feature vectors $f_{\text{in}}(p)$ at the same location. In order to satisfy the principle of *G*-covariance, the coordinatizations of these operations are all required to transform according to a precomposition with ρ_{in} and a postcomposition with ρ_{out} . When demanding that the operations are determined in terms of shared weights, these transformation laws imply a requirement for the gauge equivariance (or invariance) of the operations.

The derivations for the different pointwise operations in the following Sections 9.1.1, 9.1.2 and 9.1.3 are in the first steps mostly analogous and lead to essentially the same covariance and equivariance constraints on the template functions. They could therefore be treated together, keeping the particular operation (or template function) abstract. However, since the implications of the resulting constraints differ for the particular instantiations, and since we want to keep the discussion close to the application, we will omit such an abstract formulation and directly consider particular instantiations.

9.1.1 Gauge equivariant 1×1 -convolutions

As a first example of pointwise operations, we consider the action of a family of *linear maps* \mathcal{C}_p , which send the input feature vector $f_{\text{in}}(p)$ at each $p \in M$ to an output feature vector

$$f_{\text{out}}(p) := \mathcal{C}_p f_{\text{in}}(p). \quad (9.1)$$

If we add the assumption of spatial weight sharing, the linear maps \mathcal{C}_p and \mathcal{C}_q at different locations p and q will be coupled, and the operation can be seen as a convolution with a linear operator-valued Dirac delta kernel. This operation is quite common in computer vision, where it is usually denoted as *1×1 -convolution*, since the spatial discretization of a

linear Dirac kernel which operates on two-dimensional images is given by a (matrix-valued) kernel with a spatial extent of 1×1 pixels. We will in the following derive that the demand for spatial weight sharing will result in a constraint, which forces the matrix-valued template kernels to be *intertwiners*, that is, gauge equivariant matrices.

Prior to the assumption of weight sharing, the coordinate expressions of the linear maps \mathcal{C}_p and the gauge transformations between them behave very similar to those of the linear maps on $T_p M$, which were discussed in Section 7.2. Since the input and output feature vectors are in coordinates represented by coefficient vectors $f_{\text{in}}^A(p) \in \mathbb{R}^{c_{\text{in}}}$ and $f_{\text{out}}^A(p) \in \mathbb{R}^{c_{\text{out}}}$, the linear map is naturally represented by that matrix $\mathcal{C}_p^A \in \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ that satisfies

$$f_{\text{out}}^A(p) = \mathcal{C}_p^A \cdot f_{\text{in}}^A(p). \quad (9.2)$$

This relation does of course hold for arbitrary coordinatizations, such that we have $f_{\text{out}}^B(p) = \mathcal{C}_p^B \cdot f_{\text{in}}^B(p)$ for any other gauge, labeled by B . The transformation law which relates \mathcal{C}_p^B to \mathcal{C}_p^A follows by the principle of covariance from the transformation laws of the input and output features. Since these are given by $f_{\text{in}}^B(p) = \rho_{\text{in}}(g_p^{BA}) f_{\text{in}}^A(p)$ and $f_{\text{out}}^B(p) = \rho_{\text{out}}(g_p^{BA}) f_{\text{out}}^A(p)$, one has

$$\begin{aligned} f_{\text{out}}^B(p) &= \mathcal{C}_p^B \cdot f_{\text{in}}^B(p) \\ \Leftrightarrow \quad \rho_{\text{out}}(g_p^{BA}) f_{\text{out}}^A(p) &= \mathcal{C}_p^B \rho_{\text{in}}(g_p^{BA}) f_{\text{in}}^A(p) \\ \Leftrightarrow \quad f_{\text{out}}^A(p) &= \rho_{\text{out}}(g_p^{BA})^{-1} \mathcal{C}_p^B \rho_{\text{in}}(g_p^{BA}) f_{\text{in}}^A(p). \end{aligned} \quad (9.3)$$

A comparison with Eq. (9.2) implies that the two coordinate expressions of \mathcal{C}_p are necessarily related by

$$\mathcal{C}_p^B = \rho_{\text{out}}(g_p^{BA}) \mathcal{C}_p^A \rho_{\text{in}}(g_p^{BA})^{-1} \quad (9.4)$$

if they should respect the transformation laws of the feature vectors. As usual, these considerations are concisely captured by a commutative diagram:

$$\begin{array}{ccc} \mathbb{R}^{c_{\text{in}}} & \xrightarrow{\mathcal{C}_p^A} & \mathbb{R}^{c_{\text{out}}} \\ \rho_{\text{in}}(g_p^{BA}) \cdot \downarrow & & \downarrow \rho_{\text{out}}(g_p^{BA}) \cdot \\ \mathbb{R}^{c_{\text{in}}} & \xrightarrow{\mathcal{C}_p^B} & \mathbb{R}^{c_{\text{out}}} \end{array} \quad (9.5)$$

The important practical implication of this result so far is that the linear map \mathcal{C}_p is not restricted in any way. Differently formulated: as long as the coordinate expressions in different gauges are related by Eq. (9.4), one is free to parameterize \mathcal{C}_p in an arbitrary, fixed gauge A by an *unconstrained* matrix \mathcal{C}_p^A . As we will see, the situation changes when requiring the linear maps to share weights.

Consider now the case where the linear maps \mathcal{C}_p and \mathcal{C}_q share weights. This means that we assume them to be parameterized by a shared set of parameters, given by a 1×1 -convolution template kernel $K_{1 \times 1} \in \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$. The open question is how exactly the coordinate free maps should be parameterized in terms of this template kernel. Our requirement for *GM*-coordinate independence demands that we do not prefer any particular reference frame in the weight sharing process, that is, that we treat all coordinatizations in the same manner. It is therefore necessary to *share the template kernel with all coordinatizations at the same time*, that is, to set

$$\mathcal{C}_p^X = K_{1 \times 1} \quad \text{for any gauge } (U^X, \psi^X) \in \mathcal{A}^G \text{ with } p \in U^X, \quad (9.6)$$

where \mathcal{A}^G is the (maximal) G -atlas corresponding to the considered G -structure; see Eq. (7.25). As the covariance constraint in Eq. (9.4) needs to hold for arbitrary G -related gauges, and the coordinatizations $\mathcal{C}_p^A = \mathcal{C}_p^B = K_{1 \times 1}$ of the linear maps do all coincide, the joint demand for weight sharing and GM -coordinate independence is seen to imply a constraint

$$K_{1 \times 1} = \rho_{\text{out}}(g) K_{1 \times 1} \rho_{\text{in}}(g)^{-1} \quad \forall g \in G \quad (9.7)$$

on the template kernel. The corresponding adaptation of the commutative diagram in Eq. (9.5) with weight sharing is for any g in G given by:

$$\begin{array}{ccc} \mathbb{R}^{c_{\text{in}}} & \xrightarrow{K_{1 \times 1} \cdot} & \mathbb{R}^{c_{\text{out}}} \\ \rho_{\text{in}}(g) \cdot \downarrow & & \downarrow \rho_{\text{out}}(g) \cdot \\ \mathbb{R}^{c_{\text{in}}} & \xrightarrow{K_{1 \times 1} \cdot} & \mathbb{R}^{c_{\text{out}}} \end{array} \quad (9.8)$$

The conclusion of this analysis is that the template kernels which can be *unambiguously shared* are exactly those which are *invariant (equivariant) under the gauge action*. The vector space of such gauge invariant 1×1 -convolution kernels is simply the space of *intertwining maps* (Def. B.5.7) between the representations ρ_{in} and ρ_{out} , that is,

$$\text{Hom}_G(\rho_{\text{in}}, \rho_{\text{out}}) := \left\{ K_{1 \times 1} \in \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \mid K_{1 \times 1} = \rho_{\text{out}}(g) K_{1 \times 1} \rho_{\text{in}}(g)^{-1} \quad \forall g \in G \right\}. \quad (9.9)$$

Note that, according to *Schur's Lemma* B.5.10, the requirement on $K_{1 \times 1}$ to be an intertwiner prevents a mapping between fields that transform under non-isomorphic irreducible representations via 1×1 -convolutions. This severe restriction is unavoidable with 1×1 -convolution kernels but will be resolved later when allowing for spatially extended kernels.

At this point we want to mention that we use the terms “gauge equivariant template function” and “gauge invariant template function” interchangeably. This is justified by the observation that the invariance constraint in Eq. (9.7) can be written as an equivariance constraint $K_{1 \times 1} \rho_{\text{in}}(g) = \rho_{\text{out}}(g) K_{1 \times 1} \quad \forall g \in G$. It is in general possible to view functions which are equivariant w.r.t. some group action in their domain and codomain as the invariants of the corresponding action on the function itself; see Eq. (B.28) in Appendix B.4. In our application, the equivariance viewpoint highlights that a transformation of the input field will lead to a corresponding transformation of the output field, which ensures that all involved quantities transform covariantly with each other. On the other hand, the invariance viewpoint emphasizes that the template function can be shared in an arbitrary gauge.

9.1.2 Gauge equivariant bias summation

After applying a convolution operation, it is common to sum a (shared) bias vector to the individual feature vectors. Together with the requirement of coordinate independence, weight sharing will again lead to a linear constraint. This constraint will only allow for biases to be summed to the invariant subspaces of the gauge action on the input feature field.

As before, we first consider the bias summation without requiring weight sharing. We thus have biases b_p , depending on the position p on the manifold, which are summed to an input feature vector to produce an output feature vector

$$f_{\text{out}}(p) = f_{\text{in}}(p) + b_p. \quad (9.10)$$

Relative to gauges ψ_p^A and ψ_p^B , the bias is represented by those coefficient vectors β_p^A and β_p^B in \mathbb{R}^c that satisfy $f_{\text{out}}^A(p) = f_{\text{in}}^A(p) + \beta_p^A$ and $f_{\text{out}}^B(p) = f_{\text{in}}^B(p) + \beta_p^B$. Since the summation of vectors does not allow to change their transformation laws, the group representations associated with the input and output feature necessarily agree, that is,

$$\rho_{\text{in}} = \rho_{\text{out}} =: \rho. \quad (9.11)$$

Together with the requirement for coordinate independence, this implies that the diagram

$$\begin{array}{ccc} \mathbb{R}^c & \xrightarrow{+\beta_p^A} & \mathbb{R}^c \\ \rho(g_p^{BA}) \cdot \downarrow & & \downarrow \rho(g_p^{BA}) \cdot \\ \mathbb{R}^c & \xrightarrow{+\beta_p^B} & \mathbb{R}^c \end{array}, \quad (9.12)$$

which is the analog of that in Eq. (9.5), needs to commute. Written out as an equation, this demands the relation $\rho(g_p^{BA})f_p^A + \beta_p^B = \rho(g_p^{BA})(f_p^A + \beta_p^A)$ to hold. Since the linearity of $\rho(g)$ allows to rewrite the right-hand side as $\rho(g_p^{BA})f_p^A + \rho(g_p^{BA})\beta_p^A$, a subtraction of the input feature vector leads to

$$\beta_p^B = \rho(g_p^{BA})\beta_p^A. \quad (9.13)$$

The coefficient vectors which represent a coordinate independent bias relative to different gauges therefore need to transform exactly like the feature vectors to which they are summed. As in the case of 1×1 -convolutions, the coordinate independence does *not* restrict the bias β_p in any way, but only requires different coordinatizations of the same bias to be consistent with each other. An implementation could therefore pick an arbitrary gauge and freely parameterize the bias in that gauge by parameters in $\mathbb{R}^{c_{\text{in}}}$.

The situation changes again when asking for spatial weight sharing. Let $b \in \mathbb{R}^{c_{\text{in}}}$ be a template bias vector to be shared over the manifold. Since the only way to do this without arbitrarily preferring any coordinatization is to share the bias vector in all gauges simultaneously, we have to require

$$\beta_p^X = b \quad \text{for any gauge } (U^X, \psi^X) \in \mathcal{A}^G \text{ with } p \in U^X. \quad (9.14)$$

in analogy to Eq. (9.6). The combination of the covariance constraint in Eq. (9.13) with this gauge independent weight sharing then leads to the invariance constraint

$$b = \rho(g)b \quad \forall g \in G \quad (9.15)$$

on the bias vector template. This is exactly what we found in Theorem 4.3.2 for $\text{Aff}(G)$ -steerable bias summation layers on Euclidean spaces, but here derived from a passive instead of active transformation viewpoint. To complete the analogy to the case of 1×1 -convolutions, we show the adapted version of the commutative diagram in Eq. (9.12) with shared weights:

$$\begin{array}{ccc} \mathbb{R}^c & \xrightarrow{+b} & \mathbb{R}^c \\ \rho(g) \cdot \downarrow & & \downarrow \rho(g) \cdot \\ \mathbb{R}^c & \xrightarrow{+b} & \mathbb{R}^c \end{array}, \quad (9.16)$$

Section 4.3.2 argued that the vector space

$$\mathcal{B}_\rho^G := \{b \in \mathbb{R}^c \mid b = \rho(g)b \quad \forall g \in G\} \quad (9.17)$$

of such G -invariant bias templates coincides with the *trivial subrepresentations* in the irrep decomposition of ρ and gave some specific examples. Examples for the reflection group are explicitly derived in Section 10.3.1.

9.1.3 Gauge equivariant nonlinearities

Except from linear (convolution) operations and bias summations, the most basic operations used in any neural network are nonlinearities. We will here consider the usual case of nonlinearities σ_p which act in a spatially localized way, that is, which compute output feature vectors as $f_{\text{out}}(p) = \sigma_p(f_{\text{in}}(p))$. A shared nonlinearity will again be required to be gauge equivariant. As the reasoning which leads to this conclusion is essentially equivalent to that in the previous cases, we will only summarize it shortly.

Similar to before, any coordinate free nonlinearity σ_p is relative to gauges A and B given by coordinate expressions $\sigma_p^A : \mathbb{R}^{C_{\text{in}}} \rightarrow \mathbb{R}^{C_{\text{out}}}$ and $\sigma_p^B : \mathbb{R}^{C_{\text{in}}} \rightarrow \mathbb{R}^{C_{\text{out}}}$, which are by the demand for coordinate independence required to be related by $\sigma_p^B = \rho_{\text{out}}(g_p^{BA}) \circ \sigma_p^A \circ \rho_{\text{in}}(g_p^{BA})^{-1}$. A nonlinear template function $s : \mathbb{R}^{C_{\text{in}}} \rightarrow \mathbb{R}^{C_{\text{out}}}$ can only be shared in a coordinate independent way when sharing it with all gauges simultaneously. This turns the covariance constraint in an invariance constraint $s = \rho_{\text{out}}(g) \circ s \circ \rho_{\text{in}}(g)^{-1} \quad \forall g \in G$ on the template function, or, equivalently, in the corresponding equivariance constraint

$$\rho_{\text{out}}(g) \circ s = s \circ \rho_{\text{in}}(g)^{-1} \quad \forall g \in G. \quad (9.18)$$

This recovers the constraint on $\text{Aff}(G)$ -equivariant local nonlinearities found previously in Theorem 4.3.3. Due to the generality of nonlinear maps it is impossible to derive linear solution spaces, as done for 1×1 -convolutions and biases in Eqs. (9.9) and (9.17), respectively. For specific examples and a benchmarking of steerable nonlinearities we refer back to Sections 4.3.3 and 6.5.

9.2 Kernel field transforms and *GM*-convolutions

The central operation of convolutional networks is the convolution operation, which linearly accumulates characteristic patterns of features from a local neighborhood around each point $p \in M$ into a new feature vector $f_{\text{out}}(p)$. A spatially extended convolution kernel determines thereby the specifics of this accumulation. The principle of covariance requires coordinate independence, and therefore a specific transformation law of kernels under gauge transformations. As in the previous examples, an additional demand for weight sharing results in a requirement on the template kernel to be gauge equivariant (G -steerable).

In accordance with the previous section, we clearly distinguish between the requirements for coordinate independence and weight sharing. Section 9.2.2 starts therefore by discussing fields of kernels and their transformation laws without demanding the kernels at individual positions to be tied together. Such unrestricted kernel fields give rise to *kernel field transforms*, which are integral transforms that can be seen as precursors of convolutions. The actual *GM*-convolutions, which are parameterized by a shared, necessarily gauge equivariant template kernel, are defined in Section 9.2.3. As a preparation, we will in the following

Section 9.2.1 describe local representations of feature fields on the tangent spaces, where they will be matched with the convolution kernels.

9.2.1 A local observer's view on feature fields

In contrast to Euclidean spaces or more general homogeneous spaces like the sphere, the local geometry of a general Riemannian manifold varies from point to point. It is therefore not immediately clear how a convolution kernel should be defined on M and how it could be shared between different locations. A common solution is to define the kernel as usual on a flat, Euclidean vector space \mathbb{R}^d , and to share it over the tangent spaces instead of the manifold itself; see Sections 9.2.2 and 9.2.3 or prior work [204, 232, 293, 58, 57, 327, 67, 339]. Subsequently, the kernel can via the Riemannian exponential map be mapped down to the manifold. It can be thought of as being applied by a local observer, who is measuring features in its surrounding relative to its local reference frame. We will in this section shortly elaborate on how feature fields are perceived from the perspective of different local observers. Mathematically, this is formalized as the pullback and parallel transport of the feature field to the tangent spaces; see Fig. 9.1 for a visualization.

In order to map between the tangent spaces and the manifold, we consider the *Riemannian exponential map* (corresponding to the Levi-Civita connection).¹ Assuming the manifold for simplicity to be geodesically complete², the exponential map at a specific point $p \in M$ is a map

$$\exp_p : T_p M \rightarrow M. \quad (9.19)$$

It identifies vectors $v \in T_p M$ with those points $\exp_p(v) \in M$ that are reached when following the geodesic through p with an initial velocity of v for one unit of time. While preserving radial distances, the exponential map does in general distort angles and fails to be injective. For instance, if the manifold is a sphere, the exponential maps wrap their corresponding tangent space infinitely often around it. It is, however, guaranteed that the exponential map is a local diffeomorphism if its domain is restricted to distances shorter than the distance to the cut locus (where injectivity fails).

Given the exponential maps, one can pull feature fields on the manifold back to the tangent spaces. Specifically, let f be some feature field on M , then the *pullback* $\exp_p^* f := f \circ \exp_p$ is defined as that map that assigns the feature vector $f(\exp_p(v))$ from $\exp_p(v)$ to $v \in T_p M$. Note that, due to the missing injectivity of the exponential map, each tangent vector might be assigned to multiple tangent vectors v_1 and v_2 if $\exp_p(v_1) = \exp_p(v_2)$ – this is somewhat similar to gravitational lensing effects in physics. For the case that the exponential map is injective, or when restricting it to its injectivity radius, the pullback corresponds to an expression of the feature fields in *geodesic normal coordinates* [204].

Recall that the purpose of pulling the feature vectors back to the tangent spaces is to enable that they can be accumulated by a convolution kernel. Unfortunately, this is not immediately possible since the feature vectors at different locations live in different vector spaces and are expressed relative to different gauges.³ It is therefore necessary to express all feature

¹Even models which assume an *alternative (G-compatible) connection to transport features* utilize usually the canonical *Levi-Civita connection to compute geodesics* and exponential maps.

²The assumption that M is *geodesically complete* means that the exponential maps \exp_p are for any $p \in M$ defined on the whole tangent space $T_p M$. In cases where this assumption is violated one can resort to *zero padding*, which is commonly used in convolutional networks for finitely supported images.

³A very similar circumstance motivates the definition of *covariant derivatives*, which also needs to combine geometric objects that live in different spaces.

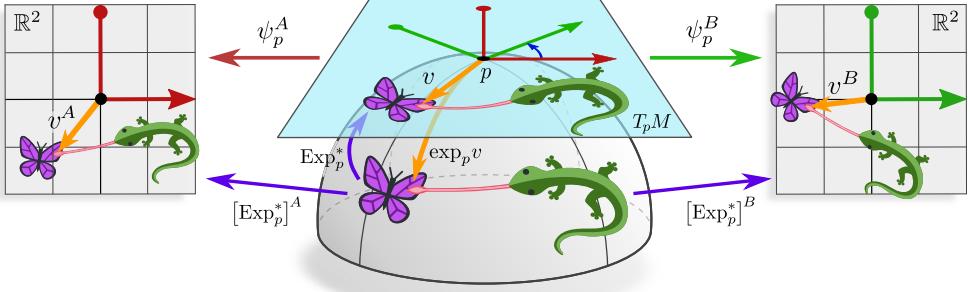


Figure 9.1: A feature field f on M and its local representation $\text{Exp}_p^* f$ on $T_p M$ via the *transporter pullback* Exp_p^* . Just like the usual pullback $\exp_p^* f$ of f along the exponential map $\exp_p : T_p M \rightarrow M$, the transporter pullback assigns feature vectors $f(\exp_p(v))$ to tangent vectors $v \in T_p M$. However, as we aim to accumulate the pulled back features by means of a convolution kernel, they need to be given in the same space and be expressed relative to the same gauge at p . The transporter pullback therefore additionally applies the (G -compatible) parallel transporter along the geodesic from $\exp_p(v)$ to p . Via a gauge ψ_p^X , the transporter pullback of f on $T_p M$ can be expressed on \mathbb{R}^d as $[\text{Exp}_p^* f]^X : \mathbb{R}^d \rightarrow \mathbb{R}^c$ – different choices of reference frames (observers) correspond hereby to different linear deformations of the feature field. Kernel field transforms and *GM*-convolutions compute an output feature $f_{\text{out}}(p)$ at p by matching a kernel \mathcal{K}_p on $T_p M$ with $\text{Exp}_p^* f$ (i.e. integrate their product over the tangent space; see Eq. (9.30)).

(Lizards and butterflies adapted under the Creative Commons Attribution 4.0 International license by courtesy of Twitter.)

vectors $[\text{Exp}_p^* f](v)$ in the same vector space and relative to the same gauge. A natural idea, proposed by Poulenard and Ovsjanikov [232], is to do this by *parallel transporting* the feature vectors along the geodesics that define the exponential map from $\exp_p(v)$ to v .⁴ We denote this pullback of f with additional transport as $\text{Exp}_p^* f$ to emphasize its close relation to the usual pullback $\exp_p^* f$ to $T_p M$. Fig. 9.1 gives a visual idea of this *transporter pullback* of feature fields to the tangent space and its representations $[\text{Exp}_p^* f]^A$ and $[\text{Exp}_p^* f]^B$ on \mathbb{R}^d relative to different coordinatizations.

We formalize $\text{Exp}_p^* f$ by defining it in terms of its coordinate expression relative to some choice of gauge. To this end, let ψ_p^A be a gauge at p , relative to which the transported features will ultimately be expressed and let $\psi_{\exp_p(v)}^{\tilde{A}}$ be an arbitrary gauge at $\exp_p(v)$, which represents the feature vector at that location by a coefficient vector $f^{\tilde{A}}(\exp_p(v)) \in \mathbb{R}^c$. Denote by

$$\rho(g_{p \leftarrow \exp_p(v)}^{A \tilde{A}}) \quad (9.20)$$

the G -compatible parallel transporter of feature vector coefficients along the geodesic from $\exp_p(v)$ to p . Then we define the *transporter pullback* in coordinates as

$$[\text{Exp}_p^* f]^A : \mathbb{R}^d \rightarrow \mathbb{R}^c, \quad v^A \mapsto [\text{Exp}_p^* f]^A(v^A) \quad (9.21)$$

$$:= \rho(g_{p \leftarrow \exp_p(\psi_p^A)^{-1}(v^A)}^{A \tilde{A}}) \cdot f^{\tilde{A}}(\exp_p(\psi_p^A)^{-1}(v^A)),$$

⁴The parallel transport along any other path would be equally valid.

where $v = (\psi_p^A)^{-1}(v^A) \in T_p M$ is the coordinate free tangent vector referred to by the coefficients v^A via ψ_p^A . As claimed before, the choice of gauge $\psi_{\exp_p(v)}^{\tilde{A}}$ at $\exp_p(v)$ is by the coordinate independence of all equations irrelevant and cancels out. Specifically, one could have used any other gauge $\psi_{\exp_p(v)}^{\tilde{B}}$ at $\exp_p(v)$, implying gauge transformations $\rho(g_p^{A\tilde{B}}) = \rho(g_p^{A\tilde{A}}) \rho(g_{\exp_p(v)}^{\tilde{B}\tilde{A}})^{-1}$ of the transporter by Eq. (8.8) and $f^{\tilde{B}}(\exp_p(v)) = \rho(g_{\exp_p(v)}^{\tilde{B}\tilde{A}}) f^{\tilde{A}}(\exp_p(v))$ of the feature vector coefficients by Eq. (8.3), which annihilate when composing both expressions.

The transporter pullback $[\text{Exp}_p^* f]^A$ depends, however, still on the gauge at p , and therefore transforms under gauge transformations g_p^{BA} at p . As for any coordinatized function, its transformation law is determined by the gauge transformations on its domain \mathbb{R}^d and codomain \mathbb{R}^c . It is therefore given by

$$[\text{Exp}_p^* f]^B = \rho(g_p^{BA}) \circ [\text{Exp}_p^* f]^A \circ (g_p^{BA})^{-1}, \quad (9.22)$$

which is summarized by the following commutative diagram:

$$\begin{array}{ccc} \mathbb{R}^d & \xrightarrow{[\text{Exp}_p^* f]^A} & \mathbb{R}^c \\ g_p^{BA} \downarrow & & \downarrow \rho(g_p^{BA}) \\ \mathbb{R}^d & \xrightarrow{[\text{Exp}_p^* f]^B} & \mathbb{R}^c \end{array} \quad (9.23)$$

Note that this is essentially the induced representation action on ρ -fields from Section 4.2, however, here restricted to G and viewed passively on a tangent space. As visualized in Fig. 9.1, $[\text{Exp}_p^* f]^A$ and $[\text{Exp}_p^* f]^B$ should be thought of as the perspective of different local observers (reference frames) on the feature field.

In principle, one could consider alternative constructions for the pullback of feature fields from M to $T_p M$. Our definition of kernel field transforms and *GM*-convolutions in Sections 9.2.2 and 9.2.3 below is independent from this particular choice.

9.2.2 Coordinate independent kernels and kernel field transforms

GM-convolutions are coordinate independent operations which apply the same, shared kernel at each point of the manifold. To clearly separate the assumptions being made, we first discuss more general *kernel field transforms*, which are coordinate independent operations but drop the requirement of weight sharing. They are therefore similar to *GM*-convolutions but apply a potentially different kernel \mathcal{K}_p to each point p of the manifold. In order to respect the principle of covariance, the coordinate expressions of those kernels are required to transform in a principled manner, however, the kernels themselves are left unconstrained.

Coordinate independent kernels: Since convolutions in deep learning map between fields of feature vectors of dimensionalities $\mathbb{R}^{c_{\text{in}}}$ and $\mathbb{R}^{c_{\text{out}}}$, the convolution kernels are $c_{\text{out}} \times c_{\text{in}}$ matrix-valued. Discretized implementations of d -dimensional convolutions on Euclidean spaces typically represent such kernels as arrays of shape $(s_1, \dots, s_d, c_{\text{out}}, c_{\text{in}})$. The

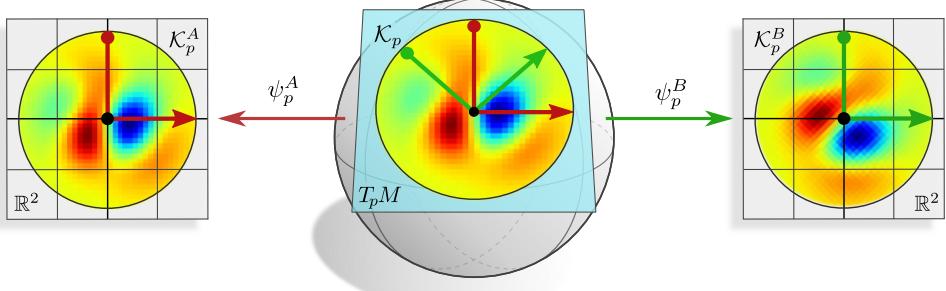


Figure 9.2: A coordinate free kernel \mathcal{K}_p on $T_p M$ and its coordinate expressions $\mathcal{K}_p^X : \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ relative to gauges ψ_p^X (only one of the $c_{\text{out}} \times c_{\text{in}}$ kernel channels is shown). The gauge transformations that relate different coordinatizations of a kernel follow from the transformation laws of their domain \mathbb{R}^d and codomain $\mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$. They are therefore for any $v \in \mathbb{R}^d$ given by $\mathcal{K}_p^B(g_p^{BA}v) = \rho_{\text{out}}(g_p^{BA}) \mathcal{K}_p^A(v) \rho_{\text{in}}(g_p^{BA})^{-1}$. A kernel field \mathcal{K} on M is a smooth assignment of kernels over the tangent spaces (Def. 12.2.1). Note that we are here assuming the kernel on $T_p M$ to be given and express it subsequently relative to different gauges on \mathbb{R}^d . This is conceptually different from the situation depicted in Figs. 1.5, 1.7, 8.1 and F.2, where we assume a template kernel K to be given on \mathbb{R}^d and subsequently define \mathcal{K}_p on $T_p M$ via convolutional weight sharing relative to some reference frame. In order to preserve coordinate independence during the weight sharing process, the shared kernel needs to be *invariant* (or equivariant) under gauge transformations; see Section 9.2.3 and Appendix G.

first d axes represent hereby a spatial grid of $s_1 \times \dots \times s_d$ pixels, each of which is assigned a $c_{\text{out}} \times c_{\text{in}}$ matrix, encoded in the last two axes.⁵ In the continuous, Euclidean setting, such kernels can be described as maps

$$K : \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}, \quad (9.24)$$

which assign a $c_{\text{out}} \times c_{\text{in}}$ matrix to each point of \mathbb{R}^d . As mentioned in the previous Section 9.2.1, we define *GM*-convolutions as matching the transporter pullback $\text{Exp}_p^* f_{\text{in}}$ on the tangent space $T_p M$ with a kernel \mathcal{K}_p on $T_p M$. Since the tangent spaces are flat, it is natural to define this matching as in the usual, fully Euclidean setting. We do therefore define the kernels \mathcal{K}_p via their coordinate expressions, which take the form in Eq. (9.24), that is,

$$\mathcal{K}_p^A : \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}. \quad (9.25)$$

Fig. 9.2 shows a *given* coordinate free kernel on $T_p M$ and its representations on \mathbb{R}^d relative to different reference frames.⁶

The transformation law between the coordinate representations \mathcal{K}_p^A and \mathcal{K}_p^B of a kernel \mathcal{K}_p on $T_p M$ follows as usual from the transformation laws of their domain and codomain. On the domain \mathbb{R}^d the transformation law is given by g_p^{BA} , while the transformation law of $\mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ is, as in Eq. (9.4), given by a simultaneous left multiplication with $\rho_{\text{out}}(g_p^{BA})$ and

⁵The actual memory layout depends on the particular deep learning framework in consideration.

⁶We emphasize that we are here assuming a coordinate free kernel \mathcal{K}_p which is *given* on $T_p M$ and consider its coordinate expressions \mathcal{K}_p^X on \mathbb{R}^d relative to reference frames X . Convolutional weight sharing will later pose us with the question of how to *define* a coordinate free kernel \mathcal{K}_p on $T_p M$ given a template kernel K on \mathbb{R}^d . Appendix G elaborates on these two concepts and their relation to the kernel's *G*-steerability.

right multiplication with $\rho_{\text{in}}(g_p^{BA})^{-1}$. The two coordinatizations of the kernel \mathcal{K}_p relate thus for any $v \in \mathbb{R}^d$ by

$$\mathcal{K}_p^B(g_p^{BA}v) = \rho_{\text{out}}(g_p^{BA}) \cdot \mathcal{K}_p^A(v) \cdot \rho_{\text{in}}(g_p^{BA})^{-1}, \quad (9.26)$$

which is visualized by the following commutative diagram:

$$\begin{array}{ccc} \mathbb{R}^d & \xrightarrow{\mathcal{K}_p^A} & \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \\ g_p^{BA} \downarrow & & \downarrow \rho_{\text{out}}(g_p^{BA}) [\cdot] \rho_{\text{in}}(g_p^{BA})^{-1} \\ \mathbb{R}^d & \xrightarrow{\mathcal{K}_p^B} & \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \end{array} \quad (9.27)$$

As in the examples from Section 9.1, the principle of covariance only requires a consistent transformation behavior between different kernel coordinatizations but does not lead to a constraint on the kernel itself. One might therefore parameterize the kernels \mathcal{K}_p for any $p \in M$ and an arbitrary gauge at p by some unrestricted, matrix-valued kernel. We denote smooth fields of such kernels as kernel fields, which play a major role in our analysis of the isometry equivariance of *GM*-convolutions in Chapter 13.

Coordinate independent kernel field transforms: Given a smooth kernel field \mathcal{K} , we can define *kernel field transforms*, which are similar to convolutions but differ in that they might apply a different kernel at each spatial position. They compute a field of output feature vectors $f_{\text{out}}(p)$ by integrating the product of the corresponding kernel \mathcal{K}_p and transporter pullback $\text{Exp}_p^* f_{\text{in}}$ of f_{in} over $T_p M$, that is,

$$f_{\text{out}}(p) = \int_{T_p M} \mathcal{K}_p(v) \text{Exp}_p^* f_{\text{in}}(v) dv. \quad (9.28)$$

To express this coordinate free definition in terms of coordinates, one has to replace all quantities by their coordinate expressions and to pull the integration via the chosen gauge from $T_p M$ to \mathbb{R}^d . As described in Appendix D, the appropriate (gauge invariant) Riemannian volume element is for a gauge ψ_p^A given by

$$\sqrt{|\eta_p^A|} dv^A, \quad (9.29)$$

where the factor $\sqrt{|\eta_p^A|}$, defined in Eq. (D.6), is the (positive) volume spanned by the reference frame $[e_i^A]_{i=1}^d$ at p . The coordinate expression of the kernel field transform thus reads

$$f_{\text{out}}^A(p) = \int_{\mathbb{R}^d} \mathcal{K}_p^A(v^A) [\text{Exp}_p^* f_{\text{in}}]^A(v^A) \sqrt{|\eta_p^A|} dv^A. \quad (9.30)$$

The coordinate independence of the kernel field transform is asserted by expressing it relative to an alternative gauge ψ_p^B and showing that the resulting output field transforms as expected,

which is indeed the case:

$$\begin{aligned}
f_{\text{out}}^B(p) &\stackrel{(1)}{=} \int_{\mathbb{R}^d} \mathcal{K}_p^B(v^B) [\text{Exp}_p^* f_{\text{in}}]^B(v^B) \sqrt{|\eta_p^B|} dv^B \\
&\stackrel{(2)}{=} \int_{\mathbb{R}^d} \left[\rho_{\text{out}}(g_p^{BA}) \mathcal{K}_p^A((g_p^{BA})^{-1} v^B) \right] \rho_{\text{in}}(g_p^{BA})^{-1} [\text{Exp}_p^* f_{\text{in}}]^B(g_p^{BA} v^A) \sqrt{|\eta_p^A|} dv^A \\
&\stackrel{(3)}{=} \rho_{\text{out}}(g_p^{BA}) \int_{\mathbb{R}^d} \mathcal{K}_p^A(v^A) \left[\rho_{\text{in}}(g_p^{BA})^{-1} [\text{Exp}_p^* f_{\text{in}}]^B(g_p^{BA} v^A) \right] \sqrt{|\eta_p^A|} dv^A \\
&\stackrel{(4)}{=} \rho_{\text{out}}(g_p^{BA}) \int_{\mathbb{R}^d} \mathcal{K}_p^A(v^A) [\text{Exp}_p^* f_{\text{in}}]^A(v^A) \sqrt{|\eta_p^A|} dv^A \\
&\stackrel{(5)}{=} \rho_{\text{out}}(g_p^{BA}) f_{\text{out}}^A(p)
\end{aligned} \tag{9.31}$$

Here we used the definition of kernel field transforms and the transformation law of kernels (Eq. (9.26)) in the first two steps. The third step follows by pulling ρ_{out} out of the integral and substituting v^B with $v^A = (g_p^{BA})^{-1} v^B$, using that the volume element $\sqrt{|\eta_p^B|} dv^B = \sqrt{|\eta_p^A|} dv^A$ is by design gauge invariant. The last two steps follow then by identifying the transformation law of the transporter pullback of the feature field in Eq. (9.22) and the definition of the kernel field transform in gauge ψ_p^A . Note that the coordinate independence of the kernel field transform affirms the correctness of the kernel transformation law in Eq. (9.26).

A kernel field transform is only well defined if the integrals over the tangent spaces converge, which is more rigorously discussed in Section 12.2 and Appendix I. Theorem 12.2.6 proves that a compact support of the kernels \mathcal{K}_p is sufficient to guarantee this well-definedness. It further proves that kernel field transforms that are based on smooth kernel fields will map smooth input feature fields to smooth output feature fields.

9.2.3 *GM*-convolutions and *G*-steerable kernels

The freedom of kernel field transforms to apply a different kernel at each location does not allow them to generalize learned inference over different locations and thus makes them data inefficient. One does therefore typically consider convolutions, which can be seen as those specific kernel field transforms that are based on convolutional kernel fields, i.e. kernel fields that are parameterized by a single, shared template kernel. As before, a coordinate independent weight sharing requires the template kernels to be gauge equivariant (*G*-steerable). This gauge equivariance of the template kernels implies that patterns which appear in different, *G*-related geometric poses are guaranteed to evoke the same response up to a corresponding transformation of the feature vector via ρ_{out} .

Convolutional weight sharing: Let $K : \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ be a template kernel to be shared over all tangent spaces. In order to not prefer any particular gauge – which would contradict our requirement for coordinate independence – we are forced to share the kernel with coordinatizations in all gauges simultaneously. Naively, this seems to suggest to share the template kernel by setting $\mathcal{K}_p^X = K$ for any point $p \in M$ and any gauge ψ_p^X at p . While such a definition of kernel sharing seems reasonable, it does not follow our principle of sharing local template functions in a strict sense: instead of directly sharing the kernel, it is important to share the *whole* local operation – which is here the whole integral transform in Eq. (9.30). Since this operation is parameterized in terms of the kernel field \mathcal{K} , this leads indirectly to a

sharing of the template kernel, however, with a slightly different result as the naive sharing considered above.

To find the correct definition of *GM*-convolutional kernel fields according to our principle of sharing local template functions, we first need to identify these local operations. We do this by abstracting kernel field transforms (in coordinates) as a collection of local integral operators of the form

$$\mathcal{G}_{\mathcal{K},p}^A : C^\infty(\mathbb{R}^d, \mathbb{R}^c) \rightarrow \mathbb{R}^c, \quad F \mapsto \int_{\mathbb{R}^d} \mathcal{K}_p^A(v) F(v) \sqrt{|\eta_p^A|} dv, \quad (9.32)$$

where $C^\infty(\mathbb{R}^d, \mathbb{R}^c)$ denotes the space of smooth maps from \mathbb{R}^d to \mathbb{R}^c . In our application, these smooth maps are just the local feature field representations $[\text{Exp}_p^* f]^A : \mathbb{R}^d \rightarrow \mathbb{R}^c$ as seen from the tangent spaces at p , which are by the kernel field transform mapped to an output feature vector $f_{\text{out}}^A(p) = \mathcal{G}_{\mathcal{K},p}^A([\text{Exp}_p^* f]^A)$ at p . Given our template kernel $K : \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$, we define a corresponding integral operator template

$$\mathcal{I}_K : C^\infty(\mathbb{R}^d, \mathbb{R}^c) \rightarrow \mathbb{R}^c, \quad F \mapsto \int_{\mathbb{R}^d} K(v) F(v) dv, \quad (9.33)$$

which multiplies a local field representation F with the template kernel K and then integrates their product. Note that \mathcal{I}_K is as a template function necessarily agnostic to specific choices of gauges and does therefore not involve a frame volume factor. A *GM*-coordinate independent convolutional weight sharing scheme is imposed by demanding that this template functional agrees with all the individual integral operators at any point and in any gauge, that is,

$$\mathcal{G}_{\mathcal{K},p}^X = \mathcal{I}_K \quad \text{for any gauge } (U^X, \psi^X) \in \mathcal{A}^G \text{ with } p \in U^X, \quad (9.34)$$

where \mathcal{A}^G is the (maximal) G -atlas corresponding to the considered G -structure; see Eq. (7.25). This is equivalent to directly sharing the local template kernel according to

$$\mathcal{K}_p^X = \frac{K}{\sqrt{|\eta_p^X|}} \quad \text{for any gauge } (U^X, \psi^X) \in \mathcal{A}^G \text{ with } p \in U^X, \quad (9.35)$$

where the normalization factor reduces the ‘‘kernel density’’ by the reference frame volume $\sqrt{|\eta_p^X|}$. This volume factor will result in the factor $|\det g|$ in the G -steerability constraint, which is necessary for 1) recovering Euclidean steerable CNNs from Part I and 2) guaranteeing the active gauge equivariance of coordinate independent CNNs for non-volume-preserving symmetry groups, as discussed below.

We denote kernel fields which are parameterized by a shared kernel K according to Eq. (9.35) as *GM-convolutional kernel fields*. The simultaneous requirement for weight sharing and coordinate independence leads to an equivariance constraint on the template kernels. To derive this constraint, insert the kernel sharing in Eq. (9.35) into the kernel transformation law in Eq. (9.26), which results in

$$\frac{1}{\sqrt{|\eta_p^B|}} K(g_p^{BA} v) = \frac{1}{\sqrt{|\eta_p^A|}} \rho_{\text{out}}(g_p^{BA}) \cdot K(v) \cdot \rho_{\text{in}}(g_p^{BA})^{-1}. \quad (9.36)$$

Since the volumes of different reference frames are related by $\sqrt{|\eta_p^A|} = |\det(g_p^{BA})| \sqrt{|\eta_p^B|}$ and since the transformation law needs to hold for arbitrary G -related gauges, this implies the *G-steerability* constraint

$$K(gv) = \frac{1}{|\det g|} \rho_{\text{out}}(g) \cdot K(v) \cdot \rho_{\text{in}}(g)^{-1} \quad \forall v \in \mathbb{R}^d, g \in G. \quad (9.37)$$

object	\mathcal{K}_p^X	K	$[\text{Exp}_p^* f]^X$ and F	$\sqrt{ \eta^X }$	dv^X and dv
density s	0	-1	0	-1	1

Table 9.1: An overview of the density exponents s of different objects involved in general kernel field transforms and *GM*-convolutions. The coordinate expression of an s -density transforms with a factor of $|\det g|^s$ when the coordinates are transformed via $g \in G$. A general matrix-valued kernel \mathcal{K}_p^X is according to Eq. (9.26) a 0-density. The same holds for feature fields and their pullbacks, whose transformation laws are given in Eqs. (8.3) and (9.22). The whole integrand $\mathcal{K}_p^X(v^X)[\text{Exp}_p^* f]^X(v^X)\sqrt{|\eta^X|}dv^X$ of a general kernel field transforms in Eq. (9.30) is seen to be a 0-density as well – note that this is necessary for its coordinate independence as demonstrated in Eq. (9.31). As the integral operator template \mathfrak{I}_K in Eq. (9.33) is agnostic of any choice of gauge, it does not involve the frame volume factor $\sqrt{|\eta^X|}$. Since it should nonetheless behave like the integral operators $\mathcal{I}_{K,p}^X$ underlying kernel field transforms, the whole integrand $K(v)F(v)dv$ of $\mathfrak{I}_K(F)$ is required to be a 0-density. This necessitates the shared template kernels K themselves to transform like -1-densities, which is reflected in the G -steerability constraint in Eq. (9.37). Note that this transformation law of template kernels is strictly necessary for the local G -equivariance of *GM*-convolutions if the output features should transform like densities of weight 0; see Eq. (9.42). For an alternative perspective, we point the interested reader to Corollary 1 in [10], where the determinant factor is derived from Haar measures on Lie groups.

on template kernels, which was already found in Theorem 4.3.1 on Euclidean steerable convolutions. Diagrammatically, a G -steerable kernels K is required to satisfy the commutativity of

$$\begin{array}{ccc} \mathbb{R}^d & \xrightarrow{K} & \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \\ g \cdot \downarrow & & \downarrow \frac{1}{|\det g|} \rho_{\text{out}}(g) [\cdot] \rho_{\text{in}}(g)^{-1} \\ \mathbb{R}^d & \xrightarrow{K} & \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \end{array} \quad (9.38)$$

for any $g \in G$. Note that the inverse determinant factor $|\det g|$ in the kernel’s transformation law makes it transform like a *matrix-valued -1-density*; see Table 9.1 for more details. Intuitively, G -steerable kernels are exactly those kernels that can be shared relative to arbitrary G -related reference frames without that the particular choice of gauge would influence the result.⁷ The ambiguity of kernel alignments – which motivated our investigation of coordinate independent CNNs in the first place – is thus resolved by additional weight sharing over all the equivalent reference frames (all gauges) in the considered G -structure *GM*. For more details on steerable kernels and their solution spaces we refer back to Chapter 5.

***GM*-coordinate independent convolutions:** Given a G -steerable template kernel $K \in \mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^G$, the *GM*-convolution $K \star$ with this kernel is defined as the kernel field transform with the corresponding *GM*-convolutional kernel field, satisfying $\mathcal{K}_p^X = K / \sqrt{|\eta^X|}$ for any point $p \in M$ and any gauge ψ_p^X . By inserting the *GM*-convolutional kernel field into the

⁷The G -steerability constraint can be rewritten as $K(v) = |\det g|^{-1} \rho_{\text{out}}(g) \cdot K(g^{-1}v) \cdot \rho_{\text{in}}(g)^{-1} \forall v \in \mathbb{R}^d, g \in G$, which emphasizes that G -steerable kernels are the *invariants* under the gauge action on the right-hand-side. Being invariant under gauge transformations, a G -steerable kernel leads to the same coordinate free kernel \mathcal{K}_p at p when being shared relative to any reference frame in $G_p M$.

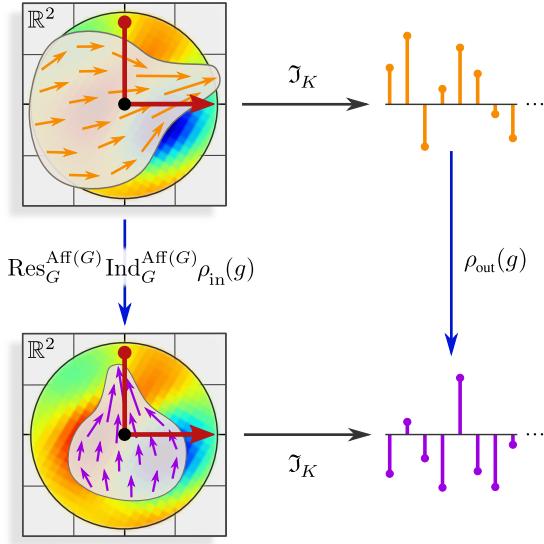


Figure 9.2: Local G -equivariance of the shared integral operator template \mathfrak{J}_K underlying a GM -convolution $K\star$. An active G -transformation $\text{Res}_G^{\text{Aff}(G)} \text{Ind}_G^{\text{Aff}(G)} \rho$ of a local field representation on \mathbb{R}^d moves feature vectors from $g^{-1}v$ to v and transforms them additionally via $\rho_{\text{in}}(g)$. While the former moves features spatially, the latter transforms their numerical coefficients (visualized as rotation and scaling of the individual (tangent) vectors in the figure). The application of \mathfrak{J}_K to both inputs results in different output feature vectors, however, by the G -equivariance of \mathfrak{J}_K , the responses are guaranteed to be related by $\rho_{\text{out}}(g)$; see Eq. (9.42). An active G -transformation of an input field therefore results in a corresponding active G -transformation of the output feature vector. Note that the G -equivariance of \mathfrak{J}_K is a direct consequence of the G -steerability of K .

kernel field transform, Eq. (9.30), the coordinate expression of the GM -convolution boils down to

$$f_{\text{out}}^A(p) = [K \star f_{\text{in}}]^A(p) := \int_{\mathbb{R}^d} K(v) [\text{Exp}_p^* f_{\text{in}}]^A(v) dv = \mathfrak{J}_K([\text{Exp}_p^* f_{\text{in}}]^A). \quad (9.39)$$

It is thus simply given by matching the transporter pullback $[\text{Exp}_p^* f_{\text{in}}]^A$ of the feature field in an *arbitrarily chosen gauge* ψ_p^A with the *gauge independent* convolution kernel K . GM -coordinate independent convolutions are therefore easily implemented by 1) choosing arbitrary reference frames, 2) pulling (and transporting) the feature fields back to the tangent space coordinatizations and 3) contracting them there with a (trainable) G -steerable kernel.

GM -convolutions exhibit multiple related symmetry properties, an overview of which is given in Fig. 1.9:

GM-coordinate independence: As specific instances of kernel field transforms, GM -convolutions are (passively) coordinate independent (or G -covariant), i.e. Eq. (9.31) applies to them.

global isometry equivariance: They are equivariant under the *active, global* action of G -structure preserving isometries in Isom_{GM} on feature fields. Sections 9.3 and specifically Chapter 13 discuss this property in detail.

local G -equivariance: The integral operator template \mathfrak{J}_K is by the G -steerability of K itself G -equivariant. Any G -transformation of a local feature field representation on \mathbb{R}^d will therefore result in a corresponding transformation of the resulting feature vector; see Fig. 9.2. Independent G -transformations of patterns that are centered at different points $p_i \in M$ will therefore lead to independent output feature transformations at these points (this holds *only* at these points and requires compactly supported kernels whose entire *field of view* transforms according to the G -transformation).

To make the last point precise, we need to define active G -transformations of local feature field representations as seen from a kernel's viewpoint, i.e. G -actions on $C^\infty(\mathbb{R}^d, \mathbb{R}^c)$.

These are naturally given by the restriction of the induced representation (Eq. (4.6)) back to G ,⁸ that is, by

$$\text{Res}_G^{\text{Aff}(G)} \text{Ind}_G^{\text{Aff}(G)} \rho : G \rightarrow \text{GL}(C^\infty(\mathbb{R}^d, \mathbb{R}^c)), \quad (9.40)$$

whose actions are defined by

$$[\text{Res}_G^{\text{Aff}(G)} \text{Ind}_G^{\text{Aff}(G)} \rho](g) F := \rho(g) \circ F \circ g^{-1}. \quad (9.41)$$

Note that this is the active counterpart of the passive local gauge transformation of transporter pullbacks from Eq. (9.22) above. Fig. 1.10 visualizes both, showing in particular that they are indistinguishable from a kernel's viewpoint.

The claimed G -equivariance of \mathfrak{I}_K is easily seen by applying it to a transformed input, followed by a substitution and making use of the G -steerability of K :

$$\begin{aligned} & \mathfrak{I}_K \left(\text{Res}_G^{\text{Aff}(G)} \text{Ind}_G^{\text{Aff}(G)} \rho_{\text{in}}(g) F \right) && (9.42) \\ &= \mathfrak{I}_K (\rho_{\text{in}}(g) \circ F \circ g^{-1}) && (\text{def. of } \text{Res}_G^{\text{Aff}(G)} \text{Ind}_G^{\text{Aff}(G)} \rho, \text{Eq. (9.41)}) \\ &= \int_{\mathbb{R}^d} K(v) \rho_{\text{in}}(g) F(g^{-1}v) dv && (\text{def. of } \mathfrak{I}_K, \text{Eq. (9.33)}) \\ &= \int_{\mathbb{R}^d} K(g\tilde{v}) \rho_{\text{in}}(g) F(\tilde{v}) |\det g| d\tilde{v} && (\text{substitution of } \tilde{v} = g^{-1}v) \\ &= \int_{\mathbb{R}^d} \rho_{\text{out}}(g) K(\tilde{v}) F(\tilde{v}) d\tilde{v} && (G\text{-steerability of } K, \text{Eq. (9.37)}) \\ &= \rho_{\text{out}}(g) \mathfrak{I}_K(F) && (\text{def. of } \mathfrak{I}_K, \text{Eq. (9.33)}) \end{aligned}$$

An active transformation of a local feature field representation F on some tangent space coordinatization by $\text{Res}_G^{\text{Aff}(G)} \text{Ind}_G^{\text{Aff}(G)} \rho_{\text{in}}(g)$ is therefore guaranteed to lead to a transformation of the resulting output feature vector by $\rho_{\text{out}}(g)$. In other words, features which appear in different G -related geometric poses will evoke the same response up to a transformation via ρ_{out} . In terms of a commutative diagram, this is concisely summarized as:

$$\begin{array}{ccc} C^\infty(\mathbb{R}^d, \mathbb{R}^{c_{\text{in}}}) & \xrightarrow{\mathfrak{I}_K} & \mathbb{R}^{c_{\text{out}}} \\ \text{Res}_G^{\text{Aff}(G)} \text{Ind}_G^{\text{Aff}(G)} \rho_{\text{in}}(g) \downarrow & & \downarrow \rho_{\text{out}}(g) \\ C^\infty(\mathbb{R}^d, \mathbb{R}^{c_{\text{in}}}) & \xrightarrow{\mathfrak{I}_K} & \mathbb{R}^{c_{\text{out}}} \end{array} \quad (9.43)$$

Fig. 9.2 gives a visual interpretation of this equivariance property of \mathfrak{I}_K .

Note that the equivariance under local G -transformations in Eq. (9.42) requires the G -steerability constraint exactly as it is in Eq. (9.37), that is, in particular, *with* the determinant factor $|\det g|^{-1}$ which makes the kernel transform like a -1 -density. This factor is traced back to our definition of convolutional weight sharing in Eq. (9.35) *with* the normalization by the reference frame volumes $\sqrt{|\eta_p^X|}$. The naive weight sharing mentioned in the beginning of this section would therefore not have lead to the desired transformation behavior. In other words: both the naive and the normalized version of the kernel sharing are coordinate independent and behave therefore both consistently under passive gauge transformations

⁸The restriction back to G suppresses the translations in $\text{Aff}(G)$.

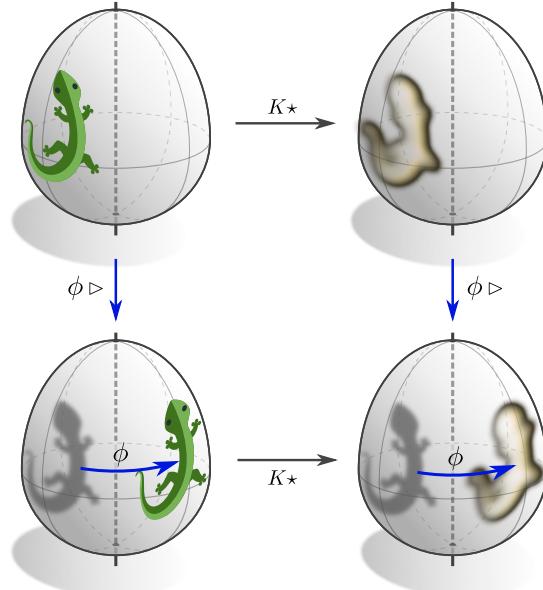


Figure 9.3: A network layer is said to be isometry equivariant when it commutes with their action on feature fields, i.e. satisfies $K \star (\phi \triangleright f_{\text{in}}) = \phi \triangleright (K \star f_{\text{in}})$ for any feature field f_{in} and isometry ϕ . GM -convolutions are by design equivariant w.r.t. the subgroup Isom_{GM} of isometries that are symmetries of the G -structure.

That GM -convolutions are Isom_{GM} -equivariant relies on the facts that 1) kernels are shared over the whole manifold, 2) isometries preserve the transporter pullback of feature fields and 3) that Isom_{GM} induces G -valued gauge transformations, which are accounted for by the kernel's G -steerability. (Lizards adapted under the Creative Commons Attribution 4.0 International license by courtesy of Twitter.)

– in particular such which change the frame volume. However, in the case of the naive kernel sharing, this is taken care of by the invariance of the Riemannian volume element $\sqrt{|\eta_p^A|} dv^A = \sqrt{|\eta_p^B|} dv^B$. By canceling this factor in the normalized weight sharing, the consistency of the transformation behavior is not guaranteed by the integration measure itself anymore – which requires the G -steerable kernels themselves to explain volume changes via the determinant factor. Only the latter generalizes to active transformations, where only the feature field is transformed, while the integration measure stays invariant.

As our definition of GM -convolutions allows for arbitrary Riemannian manifolds, G -structures and field types, it is quite general and covers a wide range of applications. We substantiate this claim in Part IV, where we explain many CNNs on Euclidean affine spaces \mathbb{E}_d , the sphere S^2 and general manifolds or meshes as specific instantiations of Eq. (9.39). For an overview and a classification of these models, we refer to Table 14.1.

9.3 Isometry equivariance

Given that a manifold exhibits symmetries it is usually desirable that neural networks respect these symmetries, i.e. are equivariant under their action on feature fields. GM -convolutions are by design guaranteed to be Isom_{GM} -equivariant, which means that they commute with the action of isometries in Isom_{GM} (Eq. (8.15)) on feature fields, as visualized in Fig. 9.3.⁹

⁹Recall that an action on GM -coordinate independent feature fields can only be defined for the G -structure preserving isometries in Isom_{GM} . It is therefore not even possible to define a notion of isometry equivariance for isometries that are not symmetries the G -structure. Note that this is without loss of generality since one can always choose a structure group $G = O(d)$, for which $\text{Isom}_{GM} = \text{Isom}(M)$ coincides with the full isometry group.

Expressed in equations, the *GM*-convolution $K \star$ is equivariant when it satisfies the relation

$$K \star (\phi \triangleright f_{\text{in}}) = \phi \triangleright (K \star f_{\text{in}}) \quad \forall \phi \in \text{Isom}_{GM} \quad (9.44)$$

for any possible input field f_{in} , that is, when the following diagram commutes:

$$\begin{array}{ccc} f_{\text{in}} & \xrightarrow{K \star} & f_{\text{out}} \\ \phi \triangleright \downarrow & & \downarrow \phi \triangleright \\ \phi \triangleright f_{\text{in}} & \xrightarrow{K \star} & \phi \triangleright f_{\text{out}} \end{array} \quad (9.45)$$

As a first step towards proving the isometry equivariance of *GM*-convolutions, recall that they are pointwise defined as the contraction of a kernel K with the transporter pullback $[\text{Exp}_p^* f_{\text{in}}]^A$ of the input field f_{in} . Since isometries preserve the Riemannian geometry of M by definition, they preserve in particular the Riemannian exponential map and Levi-Civita transporters; see Section 13.1.4 and Fig. 13.5.¹⁰ This implies that the transporter pullback of the pushforward field $\phi \triangleright f_{\text{in}}$ at $\phi(p)$ will only differ from the transporter pullback of the original field f_{in} at p by the isometry induced gauge transformation, that is,

$$[\text{Exp}_{\phi(p)}^*(\phi \triangleright f_{\text{in}})]^A = \rho_{\text{in}}(g_\phi^{A\tilde{A}}(p)) \circ [\text{Exp}_p^* f_{\text{in}}]^{\tilde{A}} \circ g_\phi^{A\tilde{A}}(p)^{-1}; \quad (9.46)$$

cf. Eq. (9.22) and, for the coordinate free formulation and a proof, Theorem 13.1.4.

Given this identity, the isometry equivariance of *GM*-convolutions is proven by the following simple calculation, which crucially leverages the G -steerability of the template kernel K to explain away the isometry induced gauge action:

$$\begin{aligned} & [K \star (\phi \triangleright f_{\text{in}})]^A(\phi(p)) \\ & \stackrel{(1)}{=} \int_{\mathbb{R}^d} K(v) [\text{Exp}_{\phi(p)}^*(\phi \triangleright f_{\text{in}})]^A(v) dv \\ & \stackrel{(2)}{=} \int_{\mathbb{R}^d} K(v) \left[\rho_{\text{in}}(g_\phi^{A\tilde{A}}(p)) [\text{Exp}_p^* f_{\text{in}}]^{\tilde{A}}(g_\phi^{A\tilde{A}}(p)^{-1}v) \right] dv \\ & \stackrel{(3)}{=} \int_{\mathbb{R}^d} \left[K(g_\phi^{A\tilde{A}}(p)\tilde{v}) \rho_{\text{in}}(g_\phi^{A\tilde{A}}(p)) \right] [\text{Exp}_p^* f_{\text{in}}]^{\tilde{A}}(\tilde{v}) |\det g_\phi^{A\tilde{A}}(p)| d\tilde{v} \\ & \stackrel{(4)}{=} \int_{\mathbb{R}^d} \left[\rho_{\text{out}}(g_\phi^{A\tilde{A}}(p)) K(\tilde{v}) \right] [\text{Exp}_p^* f_{\text{in}}]^{\tilde{A}}(\tilde{v}) d\tilde{v} \\ & \stackrel{(5)}{=} \rho_{\text{out}}(g_\phi^{A\tilde{A}}(p)) \cdot f_{\text{out}}^{\tilde{A}}(p) \\ & \stackrel{(6)}{=} [\phi \triangleright f_{\text{out}}]^A(\phi(p)) \\ & \stackrel{(7)}{=} [\phi \triangleright (K \star f_{\text{in}})]^A(\phi(p)) \end{aligned} \quad (9.47)$$

The first step follows hereby from the definition of *GM*-convolutions in Eq. (9.39) while the second step inserted the induced gauge transformation according to Eq. (9.46). A substitution from v to $\tilde{v} = g_\phi^{A\tilde{A}}(p)^{-1}v$ justifies step three. In the fourth step the G -steerability

¹⁰More generally, whenever an alternative G -compatible connection is chosen to transport feature vectors, we assume this connection to be invariant under the action of Isom_{GM} ; see Section 13.1.4. This assumption is satisfied for all models that are covered in the literature review in Part IV.

of the template kernel, i.e. Eq. (9.37), is applied (recall Eq. (8.17), which states that the Isom_{GM} -induced gauge transformations are G -valued). What follows is that the resulting output feature vector is transformed by the induced gauge transformation. After identifying this as the coordinate expression of the pushforward of the output field in Eq. (8.19), the statement follows. As all steps are valid for arbitrary isometries in Isom_{GM} , we see that *GM-convolutions are automatically equivariant w.r.t. any G-structure preserving isometry*. They are *not* necessarily equivariant w.r.t. general isometries in $\text{Isom}(M)$, which might disrespect the G -structure, however, full isometry equivariance is guaranteed for orthonormal structure groups $G = \text{O}(d)$ (or supergroups of it).

Invariant kernel fields: A more in depth analysis of the isometry equivariance of general kernel field transforms can be found in Sections 13.2 and 13.3. The central result of this investigation is Theorem 13.2.4, which states that *the isometry equivariance of a kernel field transform implies the isometry invariance of its kernel field* and vice versa. Fig 13.6 visualizes such an invariant kernel field, which is required to share weights over the orbits of the isometry action. The required invariance of the kernel field is intuitively plausible since isometry equivariance certainly requires the inference of the network to be the constant on each orbit. This abstract results implies the isometry equivariance of GM-convolutions by observing that GM-convolutional kernel fields – which are determined by a single, shared template kernel – are invariant under isometries in Isom_{GM} ; see Theorem 13.2.5 and Fig. 1.12. The template kernel's G -steerability accounts thereby for the invariance of kernels under the action of stabilizer subgroups of the isometry group.

Homogeneous spaces: While the demand for isometry equivariance requires kernels to be shared over orbits of the isometry group, it does in general not require convolutional weight sharing over the whole manifold. An important exception is the case of manifolds that are *homogeneous spaces* of their isometry group, like for instance \mathbb{R}^d or the sphere S^2 . By definition, the isometry action is on such spaces transitive, that is, there exists only one single orbit. Consequently, there will only be one independent kernel, which is via the action of the isometry group being shared over the whole space. Theorem 13.3.3 in Section 13.3 proves that *isometry equivariant kernel field transforms on homogeneous spaces are necessarily coordinate independent convolutions*. This observation establishes a formal link between our theory and prior work on convolutional networks on homogeneous spaces by Kondor and Trivedi [162], Cohen et al. [56] and Bekkers [10], who are defining convolutions via their equivariance w.r.t. global symmetries of the underlying space.

Diffeomorphism equivariance: The reader might wonder whether it is possible to make our coordinate independent CNNs fully diffeomorphism equivariant. As one can easily see, the pointwise operations from Section 9.1, i.e. 1×1 -convolutions, biases and nonlinearities, are already diffeomorphism equivariant. Specifically, let

$$\text{Diff}_{GM} := \left\{ \phi \in \text{Diff}(M) \mid [\phi_*(e_i)]_{i=1}^d \in GM \quad \forall [e_i]_{i=1}^d \in GM \right\} \leq \text{Diff}(M) \quad (9.48)$$

be the subgroup of G -structure preserving diffeomorphisms, i.e. the analog to Eq. (8.15) without the requirement on ϕ to be an isometry. Similarly to Eq. (8.17) and Theorem 13.1.3, the coordinate expressions (induced gauge transformations) of G -structure preserving diffeomorphisms are guaranteed to take values in G , that is,

$$\phi \in \text{Diff}_{GM} \iff g_\phi^{A\tilde{A}}(p) \in G \quad \forall p \in M. \quad (9.49)$$

The G -equivariance of the shared pointwise template functions will guarantee that they commute with these Diff_{GM} -induced gauge transformations – and therefore with the active diffeomorphism action itself.

GM -convolutions with spatially extended kernels, on the other hand, are in general *not* equivariant w.r.t. diffeomorphisms. The reason for this is that the transporter pullback $\text{Exp}_p^* f$ relies on exponential maps, which are inherently Riemannian constructions that do not commute with diffeomorphisms. However, as the kernels are G -steerable, Diff_{GM} -equivariance should nonetheless hold in the limit of the kernel support going to zero, where they correspond to *steerable partial differential operators* [137]. Given that convolution kernels are in typical deep learning applications quite small, diffeomorphism equivariance should in practice hold approximately.

Affine equivariance: Euclidean spaces constitute a special case since they allow for GM -convolutions that are equivariant under the action of *affine groups* $\text{Aff}(G)$. That this is the case relies on the fact that the exponential map commutes on Euclidean spaces not only with the action of isometries but more generally with affine transformations. The affine group equivariance of Euclidean GM -convolutions, which correspond to *Euclidean steerable CNNs*, is proven in Section 15.2.

Reflection steerable Möbius CNNs

To make the theoretical considerations in the previous chapters more tangible, we turn now to an exemplary application. While not being of immediate practical importance, GM -convolutions on the Möbius strip are a suitable toy model since its geometry and the involved representation theory are particularly simple. Due to its non-orientability, reference frames can only be (smoothly) preferred up to reflections. As expected, coordinate independent CNNs, applying reflection equivariant template functions, outperform a naive, coordinate dependent implementation. They are furthermore shown to be equivariant under the action of the Möbius strip's isometry group.

The following Section 10.1 discusses the geometry of the flat Möbius strip. Due to its twist, its structure group can not be reduced further than to the reflection group $G = \mathcal{R}$, such that one needs to consider an \mathcal{R} -atlas of gauges as visualized in Fig. 10.1. The isometry group is given by rotations along the strip and induces \mathcal{R} -valued gauge transformations. RM -coordinate independent feature fields, some of which are discussed in Section 10.2, necessarily have to transform according to some representation of the reflection group. Section 10.3 discusses orientation independent convolutional network operations like RM -convolutions, bias summation and nonlinearities. The RM -convolutions rely hereby on the \mathcal{R} -steerable kernels from Section 5.2 and Table 5.1. A numerical implementation of the proposed model family is discussed in Section 10.4 and evaluated in Section 10.5. The code is publicly available at <https://github.com/mauriceweiler/MobiusCNNs>.

10.1 Geometry of the Möbius strip

The manifold M under consideration is the flat Möbius strip with boundary, as shown in Fig. 10.1. It can be thought of as being constructed by taking a rectangular subset $[0, X] \times [0, Y]$ of \mathbb{R}^2 and gluing two opposing ends together in a twisted way. Such defined, the Möbius strip inherits the canonical metric of \mathbb{R}^2 , which endows it with a Riemannian structure. The metric specifies in particular a Levi-Civita connection and therefore exponential maps and parallel transporters, which are further discussed below.

A first question to answer when constructing a coordinate independent CNN is to which extent the choice of reference frames is ambiguous. Given the Riemannian metric on the strip, we can restrict our attention to orthonormal frames. One can furthermore single out one of the two directions *along* the strip to (smoothly) disambiguate the rotation of the reference

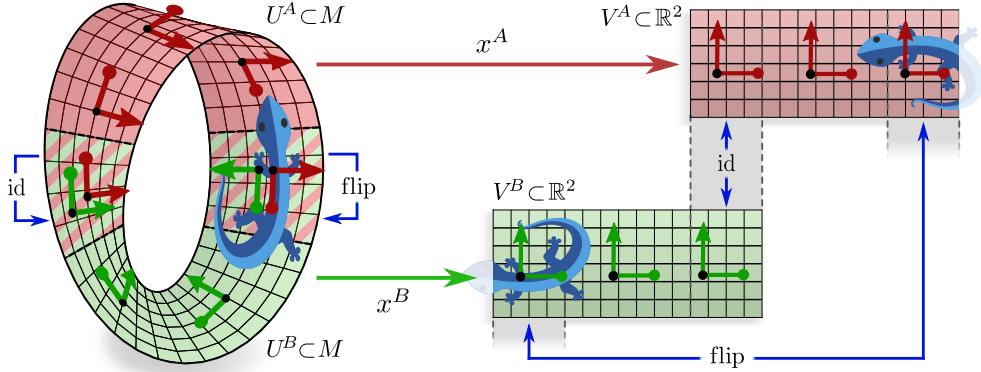


Figure 10.1: The flat geometry of the Möbius strip allows for local subsets which can be isometrically identified with corresponding subsets of \mathbb{R}^2 . We fix an isometric atlas, consisting of two charts x^A and x^B on U^A (red) and U^B (green), which cover the whole strip. Gauges $\psi_p^X = \hat{d}x_p^X : T_p M \rightarrow \mathbb{R}^d$ for $p \in U^A$ are induced as chart differentials. Due to the twist of the Möbius strip, the transition functions g_p^{BA} will at one of the overlapping regions be trivial, while the other region will necessarily transition between gauges via flips s . The chosen atlas of charts therefore induces an \mathcal{R} -atlas of gauges and implies a corresponding \mathcal{R} -structure \mathcal{RM} , consisting of two reflected frames at each point of M . Each of the charts x^X induces a smooth local frame field, given by the coordinate bases $\left[\frac{\partial}{\partial x_i^X} \right]_{|p}^d$. The flip in the transition functions at one overlap shows in a reflection of frames.

(Lizards adapted under the Creative Commons Attribution 4.0 International license by courtesy of Twitter.)

frames by aligning their first axes with this direction. This leaves us with an ambiguity of frame handedness, with the two orientations corresponding to the two possible directions of the second frame axis perpendicular to the strip. Being a non-orientable manifold, the Möbius strip does not admit a globally smooth (or even continuous) choice of frame orientations. To get an intuition about this statement, consider the attempt of constructing a smooth frame field by picking an arbitrary frame at a random position and to smoothly extend this choice over the whole strip. After one revolution around the strip the constructed frames will unavoidably be reflected w.r.t. the initial frames, and therefore contradict the desired smoothness. It is thus topologically *impossible* to define an $\{e\}$ -structure, i.e. a globally smooth field of frames, on the Möbius strip. We are thus left with an irreducible structure group

$$G = \mathcal{R} \cong \mathbb{Z}/2\mathbb{Z}, \quad (10.1)$$

which models the reflection of frames. As already discussed in Section 5.2 on \mathcal{R} -steerable kernels, the reflection group contains only two elements, the identity e and the reflection (Spiegelung) s . They are composed according to the following simple multiplication table:

	e	s
e	e	s
s	s	e

(10.2)

The only nontrivial statement in this table is that two reflections annihilate, that is, $s^2 = e$, or, equivalently, $s^{-1} = s$. Given the irreducibility of the structure group \mathcal{R} , we will in the following need to consider the corresponding \mathcal{R} -structure \mathcal{RM} which consists of two frames of opposing handedness at each point on the Möbius strip.

To encode smooth \mathcal{RM} -coordinate independent feature fields on M , one needs to specify an \mathcal{R} -atlas, consisting of \mathcal{R} -related gauges that cover the whole strip. We choose to do this by fixing an atlas of *charts*

$$x^X : U^X \rightarrow V^X \subset \mathbb{R}^2 \quad (10.3)$$

which cover the strip, and subsequently induce the gauges from it. Fig. 10.1 visualizes such an atlas, consisting of two charts x^A and x^B on U^A (red) and U^B (green) which map two overlapping halves of the strip isometrically to corresponding rectangular regions of \mathbb{R}^2 . As described in Appendix C.3, the charts induce gauges, which are given by the chart differentials, that is,

$$\psi_p^X := \hat{dx}_p^X : T_p M \rightarrow \mathbb{R}^2 \quad \text{for any } p \in U^X \text{ and } X = A, B. \quad (10.4)$$

The transition functions coincide then with the Jacobians $g^{BA} = \frac{\partial x^B}{\partial x^A}$. Due to the twist, the transition maps are at one of the two overlapping regions all trivial, that is, $g_p^{BA} = e$, and on the other end necessarily reflected, i.e. $g_p^{BA} = s$. The induced atlas of gauges is therefore indeed identified as an \mathcal{R} -atlas. Being derived from coordinate charts, the smooth local frame fields corresponding to the gauges are just the usual coordinate bases, that is, the frames $[e_i^X]_{i=1}^d$ at $p \in U^X$ are given by $[\frac{\partial}{\partial x_i^X}]_{i=1}^d$. Since the charts are isometric, the induced frame field is automatically orthonormal. However, the two rectangular regions V^A and V^B in \mathbb{R}^2 must not be rotated relative to each other in order to induce an \mathcal{R} -atlas and a corresponding \mathcal{R} -structure \mathcal{RM} .

We need to emphasize that the approach of inducing gauges via coordinate charts is not strictly necessary – it is just a convenient option since the *flat* Möbius strip is locally identified with regions of \mathbb{R}^2 in an *isometric* way. This will later allow us to transfer regular sampling grids from \mathbb{R}^2 , like for instance the pixel grid \mathbb{Z}^2 , to regular sampling grids on the strip. As this is not possible for manifolds that are not locally flat, for instance meshes in computer graphics, most implementations on general manifolds (or meshes) assign coordinates immediately to the tangent spaces; see Chapter 18.

The canonical Levi-Civita connection on the Möbius strip defines a notion of parallel transport of tangent vectors. Since the strip is locally isometric to the plane \mathbb{R}^2 , this transport can on local patches be understood as flattening these patches out into a plane and moving the vectors as usual on \mathbb{R}^2 . If no single patch can cover a path γ , there will be an open covering such that the full transport is explained by a sequence of transporters over the local patches. It is easy to see that the transport will relative to frames of the chosen \mathcal{R} -structure take values $g_\gamma^{A\tilde{A}}$ in the reflection group \mathcal{R} . This means that the Levi-Civita connection is \mathcal{R} -compatible with \mathcal{RM} . It does therefore imply well defined transporters $\rho(g_\gamma^{A\tilde{A}})$ of \mathcal{R} -associated feature vectors.

The group $\text{Isom}_{\mathcal{RM}}$ of isometries that preserve the \mathcal{R} -structure contains all rotations which shift the strip along itself. Note that a rotation once around the strip, which we denote by an angle of 2π , does *not* correspond to the identity but rather maps the strip in a reflected way on itself. Only a rotation by 4π , i.e. two full revolutions, map the strip back to itself.¹ The action of the isometry group on the manifold and on reference frames is visualized in Fig. 10.2. Relative to coordinates, the isometry action will induce \mathcal{R} -valued gauge transformations.

¹The Möbius strip is therefore seen to have the cylinder as double cover.

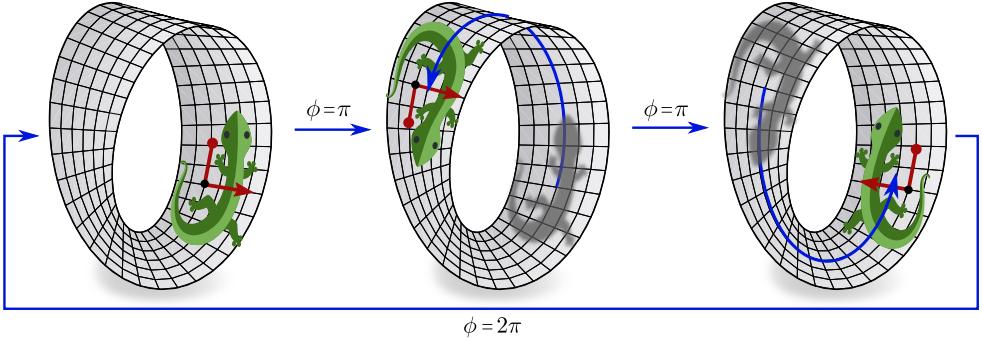


Figure 10.2: Visualization of the group of \mathcal{R} -structure preserving isometries Isom_{RM} of the Möbius strip, which is isomorphic to $\text{SO}(2)$. It consists of all rotations along the strip. Due to the twist, a rotation by 2π , i.e. once around the strip, does not yet map it back to itself but results in a reflection. After a second revolution, that is, a total rotation of 4π , the strip is mapped back to itself. Induced gauge transformations take values in \mathcal{R} .

(Lizards adapted under the Creative Commons Attribution 4.0 International license by courtesy of Twitter.)

10.2 Orientation independent feature fields

The principle of covariance requires the feature fields on the Möbius strip to be \mathcal{R} -coordinate independent, that is, they need to be equivalently expressible relative to frames of either handedness. They are therefore characterized by a choice of group representation $\rho : \mathcal{R} \rightarrow \text{GL}(c)$ of the reflection group, which specifies the transformation of numerical feature vectors when switching between the two orientations. As in Section 5.2.1 on reflection steerable feature fields on Euclidean spaces, we will consider scalar, pseudoscalar and regular feature fields, corresponding to trivial, sign-flip and regular representation, respectively. In contrast to these Euclidean feature fields, which were defined in terms of their global transformation law according to induced representations $\text{Ind}_{\mathcal{R}}^{\text{Aff}(\mathcal{R})}\rho$, the orientation independent feature fields on the Möbius strip are defined by their local gauge transformations when reflecting frames. We will therefore reintroduce \mathcal{R} -steerable feature fields from this more general viewpoint. In addition, we discuss irrep decompositions of regular feature fields. Our numerical implementation of such feature fields is described in Section 10.4.

Local gauge transformations of orientation independent features: *Scalar fields* are fields of one-dimensional feature vectors whose coefficients $f_{\text{triv}}^A(p) \in \mathbb{R}^1$ transform according to the *trivial representation* ρ_{triv} , Eq. (5.7), i.e. stay *invariant under frame reflections*:

$$\rho_{\text{triv}}(s)f_{\text{triv}}^A(p) = f_{\text{triv}}^A(p) \quad (10.5)$$

The *sign-flip representation* ρ_{sign} , Eq. (5.9), assigns the negative 1×1 identity matrix to reflections. It models *pseudoscalar fields*, which are one-dimensional and are characterized by the property that their numerical coefficients $f_{\text{sign}}^A(p) \in \mathbb{R}^1$ *change their sign under reflections*, that is,

$$\rho_{\text{sign}}(s)f_{\text{sign}}^A(p) = -f_{\text{sign}}^A(p). \quad (10.6)$$

Scalar and pseudoscalar fields are the two irreducible field types of the reflection group – any other feature field can be decomposed into a direct sum of these types, as explained below.

Regular feature fields correspond to the two dimensional *regular representation* of the reflection group, defined in Eq. (5.11). Their feature vector coefficients $f_{\text{reg}}^A(p) \in \mathbb{R}^2$ have two numerical values, which are *swapped under frame reflections*:

$$\rho_{\text{reg}}(s)f_{\text{reg}}^A(p) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} f_{\text{reg},1}^A \\ f_{\text{reg},2}^A \end{bmatrix}(p) = \begin{bmatrix} f_{\text{reg},2}^A \\ f_{\text{reg},1}^A \end{bmatrix}(p) \quad (10.7)$$

The description in terms of local gauge transformations allows to model such feature fields on any manifold admitting an \mathcal{R} -structure. For completeness, we want to point out that the Euclidean \mathcal{R} -steerable feature fields in Section 5.2.1 correspond to the \mathcal{R} -structure in Fig. 1.6d. The induced representations $\text{Ind}_{\mathcal{R}}^{\text{Aff}(\mathcal{R})}\rho$ acting on such fields is in the differential geometric setting recovered as the pushforward action $\phi \triangleright$ of isometries ϕ in $\text{Isom}_{\mathcal{RM}} = \text{Aff}(\mathcal{R}) = (\mathbb{R}^d, +) \rtimes \mathcal{R}$ on feature fields; see Section 8.3.3.

Irrep decomposition of feature fields: Any finite-dimensional representations of compact (including finite) groups is *completely reducible* into a direct sum of irreps; see Theorem B.5.16 or [102, 72, 267]. This suggests that any covariant feature vector, transforming under a compact structure group, can up to a change of basis be constructed from irrep features. As argued in Section 5.3.2 and [322], it is in this case possible to reduce any affine network operation (i.e. linear or summation layers) to equivalent operations between irrep fields, which simplifies the construction of the space of G -steerable kernels and G -invariant biases.

An example are regular feature fields, which contain two proper invariant subspaces corresponding to a scalar and a pseudoscalar field. Specifically, the regular representation can be thought of as being constructed of the direct sum $\rho_{\text{triv}} \oplus \rho_{\text{sign}}$ of the trivial and sign-flip irrep and a change of basis Q :

$$\rho_{\text{reg}}(g) = Q (\rho_{\text{triv}} \oplus \rho_{\text{sign}})(g) Q^\top \quad \text{where} \quad Q = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \quad (10.8)$$

The validity of this statement is easily asserted by inserting the right-hand side for both group elements:

$$Q (\rho_{\text{triv}} \oplus \rho_{\text{sign}})(e) Q^\top = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \rho_{\text{reg}}(e) \quad (10.9)$$

$$Q (\rho_{\text{triv}} \oplus \rho_{\text{sign}})(s) Q^\top = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \rho_{\text{reg}}(s) \quad (10.10)$$

We will in the following Section 10.3.1 show how the orientation independent bias summation operation can either be derived directly in the regular representation basis or in the irrep basis.

While the equivariance constraints may be equivalently solved in any basis, *nonlinear* network operations are sensitive to the chosen basis. This statement is in Section 10.5 shown empirically by comparing convolutions that use regular feature fields and those that use the corresponding irrep fields.

10.3 Orientation independent convolutional networks

In order to construct orientation independent CNNs on the Möbius strip we need to instantiate the gauge equivariant layers from Chapter 9 for the reflection group \mathcal{R} . More specifically,

each of the shared equivariant template functions defining the orientation independent layers needs to be instantiated for any choice of the considered field types ρ_{triv} , ρ_{sign} and ρ_{reg} . Section 10.3.1 starts by solving for the spaces $\mathcal{B}_\rho^{\mathcal{R}}$ of gauge invariant bias templates from Eq. (9.17). Some admissible choices of gauge equivariant nonlinearities for the different field types are proposed in Section 10.3.2. Section 10.3.3 comments on \mathcal{RM} -convolutions, which rely on the \mathcal{R} -steerable kernels that were already derived in Section 5.2. More general isometry equivariant kernel field transforms are discussed in Section 10.3.4. While this section will mainly consist of theoretical derivations, the following Section 10.4 will cover more practical implementation details.

10.3.1 Orientation independent bias summation

The space of bias templates that can be summed to a field of type ρ without interfering with the coordinate independence assumption was in Section 9.1.2 shown to be given by

$$\mathcal{B}_\rho^{\mathcal{R}} := \{b \in \mathbb{R}^c \mid b = \rho(g)b \quad \forall g \in \mathcal{R}\}. \quad (10.11)$$

For the case of the reflection group, there are only two group elements and thus two constraints. The constraint for the identity element $g = e$ is trivially satisfied since $\rho(e) = \text{id}_{\mathbb{R}^c}$ is by definition always the identity on \mathbb{R}^c . In the following it is therefore sufficient to restrict attention to the constraint $b = \rho(s)b$ coming from the reflection $g = s$.

We start with the case of scalar fields, i.e. the trivial representation. The reflection constraint then reads $b = \rho_{\text{triv}}(s)b = b$, which is always satisfied. It follows that the space of bias templates

$$\mathcal{B}_{\rho_{\text{triv}}}^{\mathcal{R}} = \mathbb{R} \quad (10.12)$$

remains unconstrained such that arbitrary real-valued biases can be summed to scalar fields. For the sign-flip representation the reflection constraint becomes $b = \rho_{\text{triv}}(s)b = -b$ and is therefore only satisfied for biases which are zero:

$$\mathcal{B}_{\rho_{\text{sign}}}^{\mathcal{R}} = \{0\} \quad (10.13)$$

It is thus impossible to sum biases to sign-flip fields while maintaining coordinate independence. Our third exemplary field type is the two-dimensional regular representation. The corresponding reflectional constraint on $b \in \mathbb{R}^2$ reads

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = b = \rho_{\text{reg}}(s)b = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} b_2 \\ b_1 \end{bmatrix} \quad (10.14)$$

and leads to the one-dimensional solution space

$$\mathcal{B}_{\rho_{\text{reg}}}^{\mathcal{R}} = \{b \in \mathbb{R}^2 \mid b_1 = b_2\} = \left\{ \begin{bmatrix} \beta \\ \beta \end{bmatrix} \mid \beta \in \mathbb{R} \right\}. \quad (10.15)$$

The coordinate independence of this constraint is intuitively clear: since the regular representation swaps the two channels which make up the field, the bias summation is only then coordinate independent when the values summed to both channels are equal, such that their order does not matter.

As argued in Section 4.3.2, the solution space $\mathcal{B}_\rho^{\mathcal{R}}$ for a representation ρ coincides exactly with its trivial subrepresentations. This is certainly true for the trivial representation, to which one can sum any bias, and the sign-flip representation, which has itself no trivial

subrepresentation and therefore does not admit biases at all. A more interesting example is the regular representation, which was in Eq. (10.8) shown to decompose into a direct sum of the trivial and the sign-flip representation. The one-dimensional solution space in Eq. (10.15) corresponds exactly to the single trivial subrepresentation contained in ρ_{reg} . To check the validity of this statement, note that the admissible biases for the direct sum representation $\rho_{\text{triv}} \oplus \rho_{\text{sign}}$ are of the form $(\beta, 0)^\top$, where $\beta \in \mathbb{R}$. This results can via the change of basis Q be translated back to the regular representation, which indeed recovers our solution in Eq. (10.15):

$$Q \cdot \begin{bmatrix} \beta \\ 0 \end{bmatrix} \propto \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \beta \\ 0 \end{bmatrix} = \begin{bmatrix} \beta \\ \beta \end{bmatrix} \quad (10.16)$$

10.3.2 Orientation independent nonlinearities

To construct a deep network, we need to come up with equivariant nonlinearities for each of the field types. As already discussed in Section 4.3.3, scalar fields can due to their invariance under gauge transformations be acted on by any nonlinearity $\sigma_{\text{triv}} : \mathbb{R} \rightarrow \mathbb{R}$. Usual choices are ReLU or ELU nonlinearities.

For the sign-flip fields one might take the absolute value $\|f_{\text{sign}}^A(p)\|$ of feature vectors, which maps the sign-flip field to a scalar field. In our implementation below we instead use nonlinearities of the form

$$\sigma_{\text{sign}} : f \mapsto \text{ReLU}(\|f\| - b) \cdot \frac{f}{\|f\|}, \quad (10.17)$$

where $b \in \mathbb{R}_{\geq 0}$ is a learnable bias parameter. This choice is easily seen to map sign-flip fields to sign-flip fields since the first multiplicand is acting on the gauge invariant norm of feature vectors while the second multiplicand is preserving the feature vector's sign.

As a permutation representation, the regular representation allows for any nonlinearity that acts on each field channel individually, without changing the field type. An example are ReLU nonlinearities, that are applied element-wise to the entries of a feature vector:

$$\begin{aligned} \rho_{\text{reg}}(s) \circ \sigma_{\text{reg}} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \text{ReLU}(f_1) \\ \text{ReLU}(f_2) \end{bmatrix} = \begin{bmatrix} \text{ReLU}(f_2) \\ \text{ReLU}(f_1) \end{bmatrix} \\ &= \sigma_{\text{reg}} \begin{bmatrix} f_2 \\ f_1 \end{bmatrix} = \sigma_{\text{reg}} \circ \rho_{\text{reg}}(s) \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \end{aligned} \quad (10.18)$$

While the regular representation is linearly equivalent to $\rho_{\text{triv}} \oplus \rho_{\text{sign}}$, we can not apply independent element-wise nonlinearities to the two channels in the irrep basis. This substantiates the claim that nonlinearities make the networks sensitive to the particular choice of basis of the representation.

10.3.3 Orientation independent convolutions

The last operations that we instantiate here are reflection equivariant convolutions. This requires us on the one hand to explain the exponential map and parallel transport on the strip, and on the other hand to solve for the \mathcal{R} -steerable kernel spaces.

Due to the locally flat geometry of the strip, the computation of exponential maps and transporters is almost trivial since we can choose isometric charts and identify the operations with their counterparts on \mathbb{R}^2 . More specifically, consider an isometric chart $x^A : U^A \rightarrow V^A \subset \mathbb{R}^2$ and assume first the trivial case where the whole geodesic of the exponential map $\exp_p(v)$ is contained in the chart domain, in particular $p \in U^A$ and $\exp_p(v) \in U^A$. The exponential map is then in the chart codomain given by the sum $x^A(\exp_p(v)) = x^A(p) + \hat{d}x_p^A(v)$ on \mathbb{R}^2 and the transporter is trivial. If the geodesic leaves the chart's domain U^A , one can split it in a first exponential map up to the boundary of U^A , and then transition to another chart x^B on U^B and continue there. Transporters along such geodesics are composed of the gauge transformations $\rho(g^{BA})$ at all chart transitions.

The \mathcal{R} -steerable kernel spaces were already derived in Section 5.2. They consist of kernels with certain reflectional symmetries, where the specific type of symmetry depends on the field types; see Table 5.1.

It follows that \mathcal{RM} -convolutions are within a chart (co)domain simply Euclidean convolutions with reflection steerable kernels. We handle the boundaries of the chart (co)domain, where the kernel transitions between charts, via the “transport padding” operation shown in Fig. 10.3. More details on our numerical implementation of \mathcal{RM} -convolutions on the Möbius strip are given in Section 10.4 below.

10.3.4 General isometry equivariant kernel field transforms on the Möbius strip

For completeness, we briefly elaborate on general kernel field transforms and isometry equivariant kernel field transforms on the Möbius strip. In the general case the smooth kernel field remains entirely unrestricted, that is, no weights need to be shared and the individual kernels are not required to have any reflectional symmetries whatsoever. In order for the kernel field transform to be equivariant w.r.t. isometries, the applied kernel field is required to be invariant under isometry actions. This requires weights to be shared over the isometry orbits, which come in two different types.

The first type corresponds to the single orbit lying exactly in the middle of the strip. Points on this orbit return back to themselves after being shifted one revolution around the strip, while the strip itself ends up reflected over this central orbit. An $\text{Isom}_{\mathcal{RM}}$ -invariant kernel field will therefore have some kernel shared over this central orbit. Since the kernel is after one revolution mapped back on itself in a reflected manner, it kernel is additionally required to have some reflectional symmetry like the \mathcal{R} -steerable kernels in Table 5.1.²

Any other orbit is of the second orbit type. Consider some point at a given distance from the central orbit. The isometry action will move this point at this distance from the center along the strip. Due to the strip's twist, it will not return to the initial point after one revolution but to that point which lies at the same distance on the opposite of the central orbit. Only after a second revolution around the strip the orbit will close. The demanded isometry invariance of the kernel field will thus require kernels to be shared over all points with the same distance in either direction of the strip's center (but allows for different kernels at different distances). In contrast to the central orbit's kernel, these kernels are not required to be reflection equivariant themselves, but the shared kernels will be reflected over the central orbit.

²Section 13.3 describes such situations in a more general setting as *stabilizer subgroup* constraints of the isometry group. In the current case, the subgroup of rotations once around the strip stabilizes the points on the central orbit. It is isomorphic to the reflection group and therefore leads to reflectional symmetries in the kernels.

This analysis shows that any isometry equivariant kernel field transform requires \mathcal{R} -steerable kernels, although strictly only on the central orbit. Conversely, the convolutional kernel field, corresponding to the application of the same \mathcal{R} -steerable kernels on the whole manifold, is certainly invariant under isometries. The orientation independent convolution on the Möbius strip is therefore $\text{Isom}_{\mathcal{RM}}$ -equivariant, which is empirically confirmed below.

10.4 Numerical implementation of Möbius convolutions

Being prepared with the analytical derivations in the previous sections we are ready to discuss a numerical implementation of orientation independent CNNs on the Möbius strip. The implementation is publicly available at <https://github.com/mauriceweiler/MobiusCNNs>.

Feature spaces: The first question to answer when implementing convolutions on the Möbius strip is how the feature fields should be represented numerically. Since the Möbius strip is a flat manifold, we can conveniently inherit (subsets of) the regular sampling grid \mathbb{Z}^2 from \mathbb{R}^2 over to the strip. This intuition is formalized by the pullbacks $f^X \circ (x^X)^{-1} : V^X \rightarrow \mathbb{R}^c$ of the local feature field coordinatizations $f^X : U^X \rightarrow \mathbb{R}^c$ via the (inverse) charts $(x^X)^{-1} : V^X \rightarrow U^X$ to a new domain $V^X \subset \mathbb{R}^2$, where $X = A, B$. The numerical discretization is then defined as a restriction $f^X \circ (x^X)^{-1}|_{V^X \cap \mathbb{Z}^2}$ of the pullback to the sampling grid, which is in Fig. 10.1 shown as an overlay.

Note that this representation is due to the overlap $U^A \cap U^B \neq \emptyset$ of the charts redundant. To remove this redundancy one needs to identify those regions that are represented twice and store only one, shared copy of the corresponding feature vectors. One possible scheme to do so, which we use in our numerical implementation, is to store the feature fields in the multidimensional array corresponding to the magenta rectangle in Fig. 10.3. It can be thought of as being defined by “gluing” those regions in V^A and V^B which are identified by *trivial* gauge transformations g^{BA} together (“id” in Fig. 10.1 and central four pixels in Fig. 10.3). What remains is a redundancy of feature vectors at the second overlapping region with reflecting gauge transformations (“flip” in Fig. 10.1). It is resolved by assigning those feature vectors in equal parts to either end of the “glued” local field representation (orange pixels in Fig. 10.3). Together, the pixels in the magenta box represent the feature space in a non-redundant way by assigning a c -dimensional feature vector to each of them. The ring of two pixels around the magenta rectangle is *not* part of the feature space but visualizes a padding region which will only be used during the forward pass of the convolution operation as discussed below.

The actual dimensions (shape) of the array that encodes a feature space depend on the chosen field multiplicities. Let m_{triv} , m_{sign} and m_{reg} be those integer multiplicities of feature fields which make up a feature space. Since the scalar and sign-flip fields are one-dimensional and the regular feature fields are two-dimensional, the overall number of channels (or dimensionality of stacked feature vectors) is given by $c = m_{\text{triv}} + m_{\text{sign}} + 2m_{\text{reg}}$. Assume further that the spatial resolution of the magenta rectangle is $X \times Y$ pixels and assume a batch size of N samples. The array that encodes a feature space is then of shape (N, c, X, Y) , as usual in image processing. Note that this numerical representation of the feature space is both agnostic to the twisted geometry of the strip and the actual type of the contained feature fields (except for their dimensionality). The actual geometric information is therefore solely carried by the network layers which process the fields.

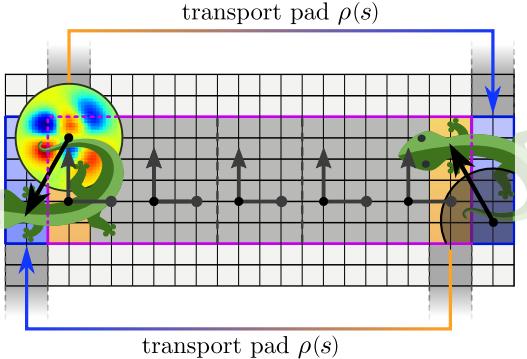


Figure 10.3: Numerical representation of feature fields on the Möbius strip and Levi-Civita transport of feature vectors during the convolution. The flat geometry of the strip allows to cut it open and flatten it out isometrically to the inner magenta rectangle. When assigning the canonical reference frames of \mathbb{R}^2 this corresponds to a gluing of the two charts V^A and V^B from Fig. 10.1 at their overlap with trivial gauge transformations (“id”). In order to avoid redundancies, we assign half of the width of

the second chart overlap with reflective gauge transformations (“flip”) to either end of the flattened magenta strip (orange pixels). Feature fields are stored as an array with spatial dimensions corresponding to the magenta box and c channels. During the convolution operation, the kernel collects features from all pixels that it covers. Choosing a kernel size of 5×5 pixels, we need to specify all values in a radius of 2 pixels around its center, which overall requires to pad a border region of 2 pixels around the magenta rectangle. The border at the top and bottom correspond to the boundary of the strip. Since no valid feature values can be assigned there, we zero-pad the array as commonly done in computer vision. The left and right border of the flattened strip are glued together with a twist. We implement the parallel transport of those features by cutting an area of two pixels width from either end of the strip (orange) and padding them in a reflected way to the opposite ends (blue). As the twist implies a gauge transformation, the feature fields need to be acted on by $\rho(s)$ when being reflected. After padding, the convolution is run with “valid” boundary conditions, such that its output again has the size of the magenta box. Operations which act pointwise do not require the padding but can be applied to the magenta array right away.

(Lizards adapted under the Creative Commons Attribution 4.0 International [license](#) by courtesy of Twitter.)

Bias summation: To implement the orientation independent bias summation, recall the results from Section 10.3.1 that the vector spaces of reflection equivariant template biases are for scalar fields and regular fields one-dimensional and for sign-flip fields zero-dimensional. At initialization of the bias module we therefore allocate an m_{triv} -dimensional parameter vector β_{triv} and an m_{reg} -dimensional parameter vector β_{reg} . During the forward pass we expand these parameters into a c -dimensional bias vector b_{full} , which is to be summed to the full stack of feature fields. This is done by allocating a c -dimensional array of zeros and filling the first m_{triv} elements with the scalar field bias parameters and the last $2m_{\text{reg}}$ elements with the m_{reg} regular field bias parameters, each repeated twice to satisfy the structure of the solution space in Eq. (10.15). After this expansion the full bias vector

$$b_{\text{full}} = \left[\underbrace{\beta_{\text{triv},1}, \dots, \beta_{\text{triv},m_{\text{triv}}}}_{m_{\text{triv}}}, \underbrace{0, \dots, 0}_{m_{\text{sign}}}, \underbrace{\beta_{\text{reg},1}, \beta_{\text{reg},1}, \dots, \beta_{\text{reg},m_{\text{reg}}}, \beta_{\text{reg},m_{\text{reg}}}}_{2m_{\text{reg}}} \right]^T \quad (10.19)$$

in \mathbb{R}^c is summed to the feature field array as usual. Its orientation independence (gauge invariance) justifies the summation to the array in Fig. 10.3, despite it being glued from feature vectors in two different gauges.

Nonlinearities: The nonlinearities can be implemented straightforwardly as defined in Section 10.3.2. We do this by splitting the full stack of feature fields into three stacks of fields of the same type, applying the respective reflection equivariant nonlinearities to them, and finally concatenating the results. Due to the definition of the nonlinearity for sign-flip

fields in Eq. (10.17) with a learnable bias, the nonlinearity module has m_{sign} trainable parameters.

GM-convolutions: Since convolution operations do not operate pointwise but accumulate all features covered by the kernel, their implementation is less trivial. The forward pass is split in three steps, namely 1) the expansion of reflection symmetric kernels from parameter arrays, 2) a Levi-Civita transport of feature vectors and 3) the actual convolution.

Recall that the space $\mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^{\mathcal{R}}$ of \mathcal{R} -steerable kernels is a linear subspace of the space of unconstrained kernels \mathcal{K} in Eq. (5.3). To parameterize \mathcal{R} -steerable kernels it is necessary to choose a basis of $\mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^{\mathcal{R}}$, in terms of which the kernels are expanded. The trainable parameters of the convolution operation are the expansion coefficients in this basis. Our implementation parameterizes all kernels that correspond to the same pair of input and output field types jointly since they share the same symmetries and thus basis. Considering the nine pairs of field types shown in Table 5.1, this means that the convolution module holds nine corresponding parameter arrays. The actual kernels are then expanded from these parameters during each forward pass. To give an example, consider the subset of kernel channels that map from m_{triv} scalar fields to m_{sign} sign-flip fields and assume a kernel size of $s \times s$ pixels. The corresponding parameter array is then of shape $(m_{\text{sign}}, m_{\text{triv}}, \frac{s}{2}, s)$ and represents the $m_{\text{triv}} \times m_{\text{sign}}$ individual kernel channels with a basis of $\frac{s}{2} \times s$ *antisymmetric* kernels each. The expansion is implemented as filling the upper $\frac{s}{2} \times s$ pixels with the unaltered parameters while the bottom $\frac{s}{2} \times s$ pixels are filled with the negated and spatially reflected parameters. As a second example, consider the kernel channels that map from m_{reg} regular fields to m_{triv} scalar fields. The parameter array for this case is of shape $(m_{\text{triv}}, m_{\text{reg}}, s, s)$ and stores one of the two kernel channels per input and output field. The second, symmetric channels are during the forward pass expanded by spatially reflecting the first kernel channels as shown in Table 5.1. After expanding the full kernel in this fashion from all of its sub-blocks corresponding to the different combinations of field types, it has the usual shape of kernels in deep learning but is guaranteed to respect the symmetries derived in Section 5.2. Note that the kernel symmetries make \mathcal{RM} -convolutions more parameter efficient than a corresponding non-equivariant CNN with the same number of channels c . Specifically for \mathcal{R} -steerable kernel the number of parameters reduced by a factor of two³.

After expanding the kernels, they are convolved with the feature fields. This requires an implementation of the exponential map and the \mathcal{R} -compatible Levi-Civita transporters on the Möbius strip – or rather on its numerical representation by the magenta array from Fig. 10.3. The flat geometry of the Möbius strip makes the implementation almost trivial, however, its boundaries and circular connectivity require some special care. We therefore need to distinguish between three qualitatively different cases, which correspond to 1) exponential maps that lie completely within the magenta array, 2) exponential maps that would cross a boundary and are therefore not well defined and 3) exponential maps whose geodesics run out at one end of the array and enter it (twisted) at the other end. The first case is trivial and corresponds to the exponential map on \mathbb{R}^2 itself. Since the strip is flat and the reference frames within the array are all parallel, the transport along these geodesics is trivial. Within the interior region of the array, where the (finitely supported) kernels do not protrude out of it, one can therefore implement the convolution as usual on \mathbb{R}^2 . The second case concerns the top and bottom rows of the array where the exponential maps might cross the boundary

³The improved parameter efficiency of \mathcal{R} -steerable kernels by a factor of 2 is exact for continuous kernels or for even kernel sizes s . If s is odd, the number of parameters scales for symmetric kernels like $s(s+1)/2$ and for antisymmetric kernels like $s(s-1)/2$ since the former are freely parameterizing the central row of pixels while the latter need to set them to zero.

of the strip (or array). This is analogous to the boundary problems for usual flat, rectangular images, where the issue is most commonly solved via zero-padding. Adopting this solution, we pad the array with rows of zeros, shown as the two light gray strips above and below the magenta rectangle in Fig. 10.3. Given a kernel size of $s \times s$ pixels with s being odd, one needs to pad $(s - 1)/2$ rows of zeros at both sides. The third case occurs at the left and right end of the array, where the strip was cut open to flatten it out. Fig. 10.3 visualizes an exemplary geodesic which crosses the cut line and therefore enters the array in a reflected direction at the opposite side. Due to the reflection, the parallel transporter across the cut is given by $\rho(s)$. In order to be able to run a conventional convolution routine, we implement the transport across the cut by copying a region of $(s - 1)/2$ pixels from both ends of the array (orange), reflecting them upside down to model the twist, acting on them with $\rho(s)$ to account for the reflected gauges and finally appending them to the opposite side of the array (blue). Having padded the array in this way, all relevant geodesics and transporters are reduced to their trivial counterparts on \mathbb{R}^2 .

Overall, our implementation of the convolution operation applies the three steps mentioned above. It first expands the \mathcal{R} -steerable kernels and pads the magenta feature field array with zeros and the field values which are transported over the cut. The expanded kernel is then convolved with the padded feature fields, calling a conventional convolution routine for flat images. We use “valid” boundary settings for the convolution, which means that the operation does not implicitly pad further zeros and only computes feature vectors for those points where the kernel does not protrude beyond the boundaries of our manually padded array. The resulting feature field will therefore again have the same spatial dimensions as the original magenta rectangle.

Pooling: Conventional CNNs usually apply spatial pooling operations which summarize feature vectors from a given pooling window, for instance a region of 2×2 pixels, into a new feature vector. Such operations reduce the spatial resolution, which lowers the computational cost and increases the effective field of view of the convolution kernels. A common way of pooling is the so-called “max-pooling”, which takes the maximum value of each individual feature channel in the pooling region. This operation can be applied to scalar fields right away since they are gauge invariant. It is further admissible for regular feature fields since taking the maximum commutes with the permutation of channels. However, as sign-flip fields change their sign under gauge transformations, max-pooling is not equivariant w.r.t. their transformation law. An equivariant alternative is average pooling, which takes the average of features in the pooling region and therefore commutes with a change of sign. Another option, that we use in our experiments below since it performs slightly better, is to pool sign-flip fields based on their absolute value, which is again invariant under sign inversions. We then multiply the sign of the pooled field values with maximum norm back in to preserve the original transformation law.

While such defined pooling operations are equivariant w.r.t. gauge transformations, their design principle interferes fundamentally with the desired isometry equivariance. This is the case since they reduce the spatial resolution of the numerical discretization, such that the output is only exactly equivariant w.r.t. the subgroup of symmetries of the lower resolution grid. This effect is well known for conventional CNNs [6]. Even though some attempts to rectify the situation were made [348], the partial loss of translation (or isometry) equivariance to a subgroup is usually accepted as it is.

Unit tests: All of the proposed coordinate independent operations are unit tested in order to guarantee their gauge equivariance and isometry equivariance. The gauge equivariance tests pass for all of the proposed operations as well as for the whole networks described in

the following section. For the convolution, bias summation and nonlinearities, our unit tests confirm isometry equivariance to hold exactly. As expected, the spatial pooling operations are not exactly equivariant w.r.t. the symmetries of the high resolution grid.⁴ However, we confirm their isometry equivariance for that subgroup of isometries which are simultaneously a symmetry of the lower resolution grid. Our empirical results, which we discuss next, suggest that the inexact isometry equivariance affects the isometry invariance of a full network’s classification predictions in most cases only marginally.

10.5 Empirical evaluation of Möbius convolutions

We evaluate the coordinate independent operations and their claimed equivariance properties on a simple classification task of MNIST images which are projected on the Möbius strip. Different combinations of field types are compared by instantiating similar model architectures for them. As a baseline we train a non coordinate independent CNN on the Möbius strip, which is significantly outperformed by the equivariant models.

The Möbius MNIST dataset is constructed by taking the standard MNIST digits of 28×28 pixels and projecting them on the strip by identifying the left and right border with an additional twist. In compliance with the rotated MNIST dataset, which is a standard benchmark for rotation equivariant Euclidean CNNs, we reduce the training set size to 12000 samples [324, 322]. Since MNIST contains single channel grayscale digits, which are invariant under gauge transformations, its samples are identified as scalar fields. Each sample is therefore represented by an array of shape $(1, 28, 28)$, corresponding to the magenta rectangle in Fig. 10.3. Note that the identification of the left and right border does not lead to any discontinuities for the specific case of MNIST digits since their background color is constant black (i.e. zero). In order to demonstrate the induced isometry equivariance of the coordinate independent CNNs, we construct two versions of this dataset. The first one contains digits which are all *centered*, that is, which occur at the same position on the strip. The second dataset puts the digits at random positions around the strip, i.e. *shifts* them by randomly sampled isometries as visualized in Fig. 10.2. Any isometry equivariant model is expected to generalize their inference from the dataset of centered digits to the isometry shifted dataset, which is confirmed by our experiments. While Möbius MNIST clearly is a toy dataset, it exhibits all the theoretical properties which we are interested in and serves as a convenient test case to demonstrate the difference between conventional CNNs and coordinate independent CNNs.

All network architectures are as usual constructed as a series of convolutional layers, followed by a global pooling operation and an invariant, fully connected classifier; see Table 10.1 for a comparison. The convolutional parts are built from six convolutional blocks with spatial pooling operations after the second and fourth convolution block. The convolution blocks are pretty basic and consist of only one convolutional layer followed by a bias summation and a nonlinearity layer. All intermediate pooling operations utilize pooling windows of 2×2 pixels and therefore halve the spatial resolution. In the case of reflection equivariant models, the last convolutional layer maps to 64 scalar fields. Their invariance under gauge transformations guarantees that the subsequent global max-pooling operation produces *both position and gauge invariant* features. An MLP with a final softmax activation takes those features to produce invariant predictions. It consists for all models of the same two MLP blocks, which apply a batch-normalization, ELU nonlinearity, dropout with

⁴Note that this issue is inherent for pooling operations and applies to conventional CNNs as well [6, 348].

layer	output field multiplicities ($m_{\text{triv}}, m_{\text{sign}}, m_{\text{reg}}$) / channels / neurons					
	scalar	sign-flip	regular	irreps	mixed	CNN
network input	(1, 0, 0)	(1, 0, 0)	(1, 0, 0)	(1, 0, 0)	(1, 0, 0)	1
conv block	(16, 0, 0)	(0, 16, 0)	(0, 0, 8)	(8, 8, 0)	(4, 4, 2)	$\lfloor 16/\sqrt{\alpha} \rfloor$
conv block	(32, 0, 0)	(0, 32, 0)	(0, 0, 16)	(16, 16, 0)	(8, 8, 4)	$\lfloor 32/\sqrt{\alpha} \rfloor$
pooling	— —	— —	— —	— —	— —	— —
conv block	(64, 0, 0)	(0, 64, 0)	(0, 0, 32)	(32, 32, 0)	(16, 16, 8)	$\lfloor 64/\sqrt{\alpha} \rfloor$
conv block	(128, 0, 0)	(0, 128, 0)	(0, 0, 64)	(64, 64, 0)	(32, 32, 16)	$\lfloor 128/\sqrt{\alpha} \rfloor$
pooling	— —	— —	— —	— —	— —	— —
conv block	(256, 0, 0)	(0, 256, 0)	(0, 0, 128)	(128, 128, 0)	(64, 64, 32)	$\lfloor 256/\sqrt{\alpha} \rfloor$
conv block	(64, 0, 0)	(64, 0, 0)	(64, 0, 0)	(64, 0, 0)	(64, 0, 0)	64
global max-pool	64	64	64	64	64	64
MLP	32	32	32	32	32	32
MLP + softmax	10	10	10	10	10	10

Table 10.1: Overview of the compared model architectures. All models consist of a convolutional part on the Möbius strip, followed by a global max-pooling operation and an MLP classifier. The five orientation independent CNNs differ in their multiplicities ($m_{\text{triv}}, m_{\text{sign}}, m_{\text{reg}}$) of field types but agree exactly in their number of channels and approximately in their number of parameters. Their inputs, i.e. the MNIST digits, are assumed to be scalar fields. All orientation independent models map in their last convolution to 64 gauge invariant scalar fields. A subsequent global pooling operation therefore produces position and coordinate independent features. The baseline CNN model comes in two flavors, which differ by their factor of $\sqrt{\alpha}$ in the number of channels. A first version assumes $\alpha = 1$, and therefore utilizes the same number of channels like the coordinate independent models. Due to the inferior parameter efficiency of non-equivariant CNNs, this model uses approximately twice as many parameters. For a fair comparison we add a second version with $\alpha = 2$ and therefore approximately the same number of parameters like the equivariant models.

30% dropping probability and a linear (or affine) layer, whose number of output neurons is listed in Table 10.1. The differences between the different models are therefore restricted to the convolutional part.

The five coordinate independent models that we instantiate differ in the utilized field types: there are three pure models, denoted by “scalar”, “sign-flip” and “regular”, which assume only the suggested field type. Due to their higher dimensionality, the field multiplicities of the regular feature fields are halved in comparison to those of scalar and sign-flip fields. A fourth model, denoted by “irrep”, uses a mixture of scalar and sign-flip fields in equal proportions. Note that the feature fields of this model are linearly equivalent to those of the “regular” models since the change of basis from Eq. (10.8) translates between both. A fifth, “mixed” model applies all three field types. The nonlinearities in use for the different field types are those described in Section 10.3.2. As stated before, all models assume scalar inputs and outputs.

All coordinate independent layers are unit tests and found to be exactly gauge equivariant, implying that the models are overall exactly gauge invariant. Since they apply two pooling steps, which reduce the spatial resolution by a factor of 2 each before the global pooling, the isometry equivariance (invariance) holds only for the subgroup of shifts by multiples of 4 pixels. The theoretically claimed properties therefore hold as expected.

As a baseline, we compare the reflection equivariant models to conventional coordinate dependent CNNs on the Möbius strip. In order to respect the topology of the strip, we apply a naive version of the transport padding operation. Since CNNs are agnostic to field types, this is done by taking the orange strips of two pixels from Fig. 10.3 and padding them to the

model	field types ρ_i			params	test error (%)	
	trivial	sign-flip	regular		shifted train digits	centered train digits
CNN (channels)	—	—	—	1501 k	1.97 ± 0.11	42.99 ± 2.65
CNN (params)	—	—	—	832 k	2.08 ± 0.10	43.68 ± 2.85
gauge CNN (scalar)	✓	✗	✗	902 k	1.60 ± 0.10	1.60 ± 0.09
gauge CNN (sign-flip)	✗	✓	✗	820 k	4.27 ± 0.24	4.89 ± 0.36
gauge CNN (regular)	✗	✗	✓	752 k	1.24 ± 0.08	1.23 ± 0.07
gauge CNN (irreps)	✓	✓	✗	752 k	1.65 ± 0.09	1.64 ± 0.12
gauge CNN (mixed)	✓	✓	✓	752 k	1.43 ± 0.09	1.42 ± 0.10

Table 10.2: Test errors of the different network architectures, each averaged over 32 runs. The column “shifted train digits” reports the performance for a setting where both the training and test samples are placed at random locations on the strip. While not being \mathcal{RM} -coordinate independent, the conventional CNNs are able to learn to detect the digits as seen from their discontinuous frame field. Almost all coordinate independent CNNs achieve significantly better results. The inferior performance of the sign-flip model shows that coordinate independent CNNs might not work very well when bad choices of field types or nonlinearities are made. The training digits in the column “centered train digits” are all placed at the same position on the strip while the test digits remain randomly shifted. The coordinate independent CNNs are able to generalize their inference between both situations which affirms their isometry equivariance. In contrast, the performance of the conventional CNNs deteriorates, which reflects their missing equivariance under isometries.

opposite side of the array after applying a reflection but without acting with the unspecified group representation – formally, this corresponds to transporting the features according to a trivial connection. Since the non-equivariant operations are less parameter efficient, we consider two different versions: the first version uses the same number of *channels* like the coordinate independent CNNs, and therefore requires approximately twice as many parameters. The number of channels of the second version is scaled down by a factor $\sqrt{2}$ such that the number of *parameters* is approximately equivalent to that of the orientation independent models.

All models are trained for 20 epochs with a batch size of 128 samples, a weight decay of 10^{-6} and using the Adam optimizer [151]. The initial learning rate of $5 \cdot 10^{-3}$ is chosen as high as possible without leading to a divergence of the training process. A fixed learning rate decay schedule reduces the step size every 4 epochs by a factor of 2.

Table 10.2 shows the resulting test errors of all models, each averaged over 32 runs. The first setting, reported in the column “shifted train digits”, uses randomly located digits both in the training and test dataset. Both versions of the non-equivariant CNN achieve approximately the same test error. In contrast, most coordinate independent CNNs achieve a significantly better result. Only the model which is purely based on sign-flip fields performs worse – this suggests that the utilized combination of sign-flip fields and nonlinearities is not a good choice, despite being coordinate independent. Bad choices of feature fields and nonlinearities are therefore seen to harm the model performance. The model achieving the best results is based on regular feature fields. This observation is in alignment with previous findings, for instance the systematic comparison of field types in [322]. Our interpretation of this result is that the kernel constraints involving regular feature fields allow for essentially unconstrained kernel channels, with the additional requirement of applying two reflected copies of them – view this in contrast to the \mathcal{R} -steerable kernels between irrep fields, which are required to be symmetric *within* one kernel channel. The model based on scalar fields achieves an intermediate performance between the conventional CNNs and the regular field model. Both models

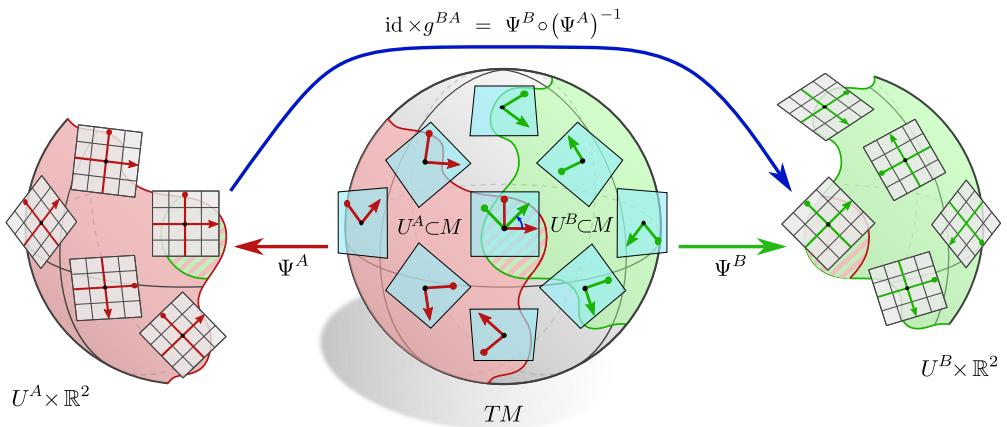
which use mixed field types have performances lying between those of the field types of the mix. We want to emphasize again that the regular model and the irrep model contain exactly the same irrep field types but are expressed in a different basis. Since this change of basis could be interpreted as part of the applied nonlinearities, this result implies that the used nonlinearities have a major impact on the model performance. Despite being investigated in [322], the landscape of equivariant nonlinearities is still largely unexplored territory.

The second training setting, reported in the column “centered train digits”, investigates the capability of the models to generalize over all poses that are related by isometries. All models are trained on digits which occur at the same location on the strip but test on randomly shifted digits. As expected, the conventional CNNs’ performances degrade significantly in this setting – this implies that they are indeed not equivariant under the isometries of the Möbius strip. In contrast, the performance of most coordinate independent CNNs stays within the standard deviation unchanged. Despite only being exactly equivariant (invariant) to the subgroup of isometries which shifts by multiples of 4 pixels, the full isometry invariance of the models therefore seems to hold very well. While the sign-flip model becomes significantly worse in comparison to the first training setting, it is still approximately isometry equivariant and therefore performs much better than the conventional CNNs.

In conclusion, the conducted experiments confirm the claimed properties of coordinate independent CNNs and show their superiority over coordinate dependent models.

PART III

FIBER BUNDLE THEORY OF COORDINATE INDEPENDENT CNNS



Introduction & overview

The previous Part II introduced coordinate independent feature fields and network layers in terms of their coordinate expressions relative to some choice of gauge on *local* neighborhoods $U \subseteq M$. As the existence of *global* gauges is in general topologically obstructed, global coordinate representations of feature fields do in general not exist. Part II addressed this issue by assembling the global content of feature fields from their local coordinate expressions relative to an atlas of gauges that cover M . A more elegant alternative is to define global feature fields and network layers in an abstract, *coordinate free* formalism in terms of fiber bundles. Bundle trivializations allow to recover the local coordinate expressions of feature fields and network layers.

This part develops a global, coordinate free description of the neural networks and feature spaces from Part II. No new models or layers are introduced, but the theory of coordinate independent CNNs is presented in a more formal language. A series of theorems justifies so far unproven claims and presents new results, including the existence and smoothness of kernel field transforms and *GM*-convolutions. The global bundle formulation allows specifically for a much more thorough investigation of the networks' equivariance under *global* isometry actions. An interesting result in this regard is that isometry equivariant kernel field transforms on homogeneous spaces are necessarily *GM*-convolutions. An overview of all theorems and definitions is given in Appendix A.

Chapter 11 introduces the theory of fiber bundles in general and discusses the tangent bundle, G -structures and G -associated feature vector bundles in particular. Neural network operations like kernel field transforms and *GM*-convolutions are defined in Chapter 12. Chapter 13 investigates the isometry equivariance of these operations.

Associated bundles and coordinate free feature fields

Fields of geometric quantities on manifolds are formalized as “sections” of fiber bundles (Eq. (11.19)). Any smooth manifold is naturally endowed with its tangent bundle and frame bundle. A choice of G -structure, which is a G -bundle of reference frames, allows to define G -associated feature vector bundles. The feature spaces of our coordinate independent neural networks are spaces of feature fields, i.e. sections of these feature vector bundles.

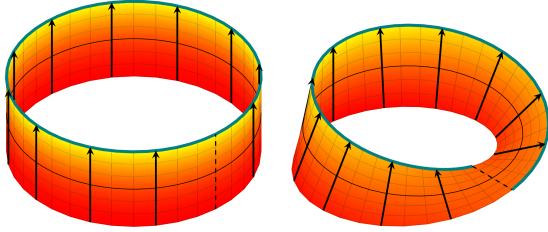
Fiber bundles in general are reviewed in Section 11.1. Section 11.2 discusses the tangent bundle TM and the frame bundle FM . G -structures GM , which are subsets of reference frames that are distinguished by the given geometric structure on the manifold, are introduced in Section 11.3. Associated G -bundles, including the feature vector bundles \mathcal{A} , are constructed from the G -structure. Section 11.4 gives details on the local trivializations (gauges) of TM , FM , GM and \mathcal{A} , which reintroduces coordinates and recovers the formulation in Part II. The mutual transformation of the trivialized feature fields with trivialized tangent vector coefficients and reference frames follows thereby from the coordinate free formulation via associated G -bundles. Section 11.5 discusses parallel transporters on the associated bundles, in particular how they induce each other.

All concepts presented here are well established in differential geometry and can easily be found in the literature [262, 221, 132, 283, 273, 202, 326, 284, 230, 59]. Our contribution is to give a comprehensive exposition which bridges between the mathematical theory and its application in geometric deep learning.

11.1 A brief introduction to fiber bundles

Intuitively, a *fiber bundle* can be thought of as a space which is constructed by taking a so-called *base space*, in our case the manifold M , and attaching another space F , denoted as *typical fiber*, to each of its points. A trivial example would be the direct product $M \times F$. However, the fibers can in general be connected in a twisted way such that the resulting bundle is topologically different from a product. For instance, let the base space be the circle $M = S^1$ and let the fiber be the line segment $F = [-1, 1]$. Their direct product $S^1 \times [-1, 1]$ then forms a cylinder; see Fig. 11.1 (left). In contrast, if the fibers are attached such that they are twisted “upside down” after one revolution around the circle, one obtains the Möbius strip, a non-trivial fiber bundle which is topologically different from the cylinder;

Figure 11.1: A cylinder and a Möbius strip. Both bundles share the circle S^1 as base space and line segments $[-1, 1]$ as fibers, however, their topological structure differs by a twist in the fibers. (Figure based on Jake's code from tex.stackexchange.com.)



see Fig. 11.1 (right).^{1,2} Note that the Möbius strip *locally* looks like a direct product $U \times F$ of a line element $U \subsetneq S^1$ with the fiber F . As discussed below, fiber bundles can by definition always be locally trivialized to direct products.

We are interested in fiber bundles since they allow for a global description of fields on manifolds. For instance, a wind field on the globe $M = S^2$ is a tangent vector field which assigns a tangent vector in $T_p M$ to each point p of M . The corresponding fiber bundle is the tangent bundle TM which connects all the tangent spaces together and is therefore identified as a fiber bundle with base space $M = S^2$ and fiber $\mathbb{R}^d \cong T_p M$. Similar to the fibers of a Möbius strip, the tangent spaces of a curved manifold are in general not connected in a canonical way but are inherently twisted relative to each other. The tangent bundle is therefore in general topologically distinct from a product, that is, $TM \not\cong M \times \mathbb{R}^d$. In order to define c -dimensional feature vector fields, we will later consider bundles with base space M and feature vector spaces \mathbb{R}^c as fibers.

11.1.1 Fiber bundles in general

Formally, a fiber bundle is a structure (E, M, π, F) consisting of topological spaces E (total space), M (base space) and F (typical fiber) and a continuous surjective projection map $\pi : E \rightarrow M$. A fiber bundle is *locally trivializable*, which means that for each point $p \in M$ there exists a local neighborhood $U \subseteq M$ of p , restricted to which the bundle looks like a direct product $U \times F$. The local triviality is formalized by homeomorphisms³ $\Psi : \pi^{-1}(U) \rightarrow U \times F$ satisfying the commutative diagram below

$$\begin{array}{ccc} E & \supseteq & \pi^{-1}(U) \xrightarrow{\Psi} U \times F \\ & & \downarrow \pi \\ M & \supseteq & U \xleftarrow{\text{proj}_1} \end{array}, \quad (11.1)$$

that is,

$$\pi = \text{proj}_1 \circ \Psi, \quad (11.2)$$

¹To prevent confusion, we emphasize that this example considers the Möbius strip as a fiber bundle with base space (manifold) $M = S^1$. In contrast, all previous figures that contained the Möbius strip considered it as the base space (manifold) M to convolve over, and to which fibers were attached.

²We furthermore need to mention that the arrows shown in the figure are just meant to emphasize the twist in the Möbius strip. They do *not* imply a gluing direction as in *gluing diagrams*.

³A *homeomorphism* is a topological isomorphism, i.e. a continuous, invertible map between topological spaces with continuous inverse.

where $\text{proj}_1 : U \times F \rightarrow U$ denotes the natural projection on the first factor. A bundle which is globally homeomorphic to the product $M \times F$ is called *trivial*. Bundles are often shortly written $E \xrightarrow{\pi} M$ or just E with the typical fiber and base space left implicit. Since we are considering smooth frame fields, we assume E, M and F to be smooth manifolds and π and Ψ to be smooth maps (diffeomorphisms).

The local triviality of $E \xrightarrow{\pi} M$ implies that the preimage $E_p := \pi^{-1}(p)$ of any point $p \in M$, called the *fiber over p*, is diffeomorphic to the typical fiber F . As in Chapter 7, we denote the diffeomorphisms which identify the fibers over different points with the typical fiber by $\psi_p : E_p \rightarrow F$. The local trivializations are then in terms of these diffeomorphisms given by

$$\Psi : \pi^{-1}(U) \rightarrow U \times F, \quad e \mapsto (\pi(e), \psi_{\pi(e)}(e)). \quad (11.3)$$

If the typical fiber F and the fibers E_p over p carry additional structure, the diffeomorphisms $\psi_p : E_p \rightarrow F$ are required to respect this structure, i.e. to be isomorphisms.⁴ For instance, if F and E_p carry a vector space structure, then ψ_p is required to be linear.

In general the specific choice of local trivializations (or diffeomorphisms) over U is not canonically specified by the bundle. One therefore has to consider different choices (gauges) and *transition functions* (gauge transformations) between them. To make this precise, consider two overlapping trivializing neighborhoods U^A and U^B with local trivializations Ψ^A and Ψ^B . From Eq. (11.3) it follows that the transition between both local trivializations is on $U^{AB} := U^A \cap U^B \neq \emptyset$ given by

$$\begin{aligned} \Psi^B \circ (\Psi^A)^{-1} : U^{AB} \times F &\rightarrow U^{AB} \times F, \\ (p, f) &\mapsto \left(p, [\psi_p^B \circ (\psi_p^A)^{-1}](f) \right) =: (p, g_p^{BA} \blacktriangleright f) \end{aligned} \quad (11.4)$$

where we implicitly defined the smooth *transition functions*⁵

$$g^{BA} : U^{AB} \rightarrow \text{Aut}(F), \quad p \mapsto g_p^{BA} := \psi_p^B \circ (\psi_p^A)^{-1} \quad (11.5)$$

and their left action

$$\blacktriangleright : \text{Aut}(F) \times F \rightarrow F, \quad (g_p^{BA}, f) \mapsto g_p^{BA} \blacktriangleright f := [\psi_p^B \circ (\psi_p^A)^{-1}](f). \quad (11.6)$$

on the typical fiber F ; cf. Eqs. (7.7) and (7.6). To see that the first factor in Eq. (11.4) is indeed given by the identity, note that, for any $p \in U^{AB}$ and any $f \in F$, the repeated application of Eq. (11.2) implies $[\text{proj}_1 \circ \Psi^B \circ (\Psi^A)^{-1}](p, f) = [\pi \circ (\Psi^A)^{-1}](p, f) = \text{proj}_1(p, f) = p$. The transition between different trivializations is visualized by the fol-

⁴Alternatively, assume that F carries structure which is respected by the transition functions $\psi_p^B \circ (\psi_p^A)^{-1} = g^{BA}(p) \in \text{Aut}(F)$ (see the next paragraph). Then the trivializations $\psi_p^X : E_p \rightarrow F$ consistently *induce* the structure of F on E_p and are automatically isomorphisms.

⁵The automorphism group $\text{Aut}(F)$ of a space F consists of all invertible, structure preserving maps (isomorphisms) from F to itself. For instance, if $F = \mathbb{R}^n$ is a vector space, the automorphism group is the general linear group $\text{GL}(n)$, which consists of all invertible $n \times n$ matrices.

lowing (commuting) extension of the commutative diagram in Eq. (11.1):

$$\begin{array}{ccccc}
 & & \text{id} \times g^{BA} \blacktriangleright & & \\
 & \swarrow & & \searrow & \\
 U^{AB} \times F & \xleftarrow{\Psi^A} & \pi^{-1}(U^{AB}) & \xrightarrow{\Psi^B} & U^{AB} \times F \\
 & \searrow \text{proj}_1 & \downarrow \pi & \swarrow \text{proj}_1 & \\
 & & U^{AB} & &
 \end{array} \tag{11.7}$$

Compare this diagram to that in Eq. (7.8), which applies to a single point $p \in U^{AB}$ only (and considers specifically the tangent bundle). A graphical interpretation of the full local trivialization diagram in Eq. (11.7) is given in Fig. 11.3.

By definition, the transition functions in Eq. (11.5) satisfy the following three conditions:⁶

$$i) \quad g_p^{AA} = e \quad \forall p \in U^A \tag{11.8}$$

$$ii) \quad g_p^{BA} = (g_p^{AB})^{-1} \quad \forall p \in U^A \cap U^B \tag{11.9}$$

$$iii) \quad g_p^{CB} g_p^{BA} = g_p^{CA} \quad \forall p \in U^A \cap U^B \cap U^C \quad (\text{cocycle condition}) \tag{11.10}$$

By the *fiber bundle construction theorem*, any fiber bundle can be fully specified globally in terms of an *atlas* $\mathcal{A} = \{(U^X, \Psi^X) \mid X \in \mathfrak{X}\}$ of local trivializations (U^X, Ψ^X) which cover M and whose transition functions satisfy Eqs. (11.8), (11.9) and (11.10) (here \mathfrak{X} denotes some index set). The individual trivializations can be thought of as being “glued together” by the transition maps, which is visualized in Fig. 11.2. Note that this is similar to the global description of a manifold in terms of an atlas of local charts.

11.1.2 Vector bundles

Several more specific notions of fiber bundles, carrying additional mathematical structure, exist. An important example are *vector bundles*, which, as the name suggests, are bundles consisting of vector spaces attached to a manifold. Formally, a (real) vector bundle of rank k is a bundle $(E, M, \pi, \mathbb{R}^k)$ with typical fiber \mathbb{R}^k and fibers $E_p \cong \mathbb{R}^k$ over p such that the local trivializations are fiber wise vector space isomorphisms (linear maps). The transition functions $\psi_p^B \circ (\psi_p^A)^{-1} \in \text{Aut}(\mathbb{R}^k) = \text{GL}(k)$ then take values in the general linear group.

Alternatively, given the fiber \mathbb{R}^k and an atlas of local trivializations whose transition functions take values in $\text{Aut}(\mathbb{R}^k) = \text{GL}(k)$, a vector space structure of E_p is induced by setting

$$\alpha v + \beta w := (\psi_p^A)^{-1}(\alpha\psi_p^A(v) + \beta\psi_p^A(w)) \quad \forall v, w \in E_p, \quad \alpha, \beta \in \mathbb{R} \tag{11.11}$$

for an arbitrary gauge $\psi_p^A : E_p \rightarrow \mathbb{R}^k$ from the $\text{GL}(k)$ -atlas. That the vector space structure is consistently defined is clear as

$$\begin{aligned}
 & (\psi_p^B)^{-1}(\alpha\psi_p^B(v) + \beta\psi_p^B(w)) \\
 &= (\psi_p^A)^{-1}((g_p^{BA})^{-1}(\alpha g_p^{BA}\psi_p^A(v) + \beta g_p^{BA}\psi_p^A(w))) \\
 &= (\psi_p^A)^{-1}(\alpha\psi_p^A(v) + \beta\psi_p^A(w))
 \end{aligned} \tag{11.12}$$

⁶Conditions *i*) and *ii*) follow from the cocycle condition *iii*) but are often stated explicitly.

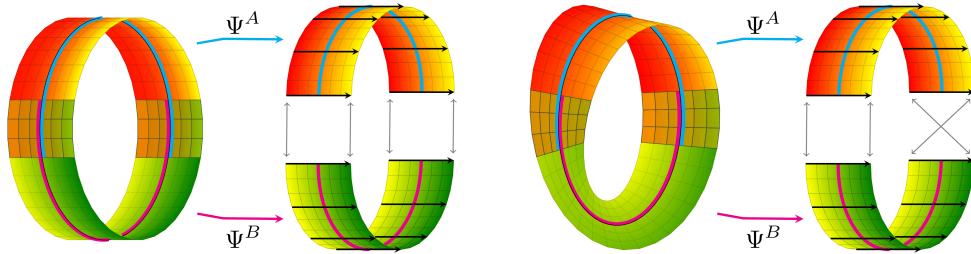


Figure 11.2: Description of the cylinder and the Möbius strip in terms of G -atlases consisting of two local trivializations each. *Left:* Since the cylinder is a trivial bundle, all transition functions can be chosen to be identity maps such that the structure group is reduced to the trivial group $G = \{e\}$. Differing from the visualized situation, it is possible to choose a single, global trivialization. *Right:* The topology of the Möbius strip forces the transition functions at one of the overlaps to glue the fibers together in an inverted way. The structure group can therefore not be reduced further than the group $G = \mathcal{R}$ which models the reflection of fibers. Global trivializations of the Möbius strip do therefore not exist. Note that the arrows on the Möbius strip should not be confused with the arrows in gluing diagrams, that is, the twist glues the vectors at one of the cuts in opposite direction.

yields the same result. Note that the last step required the linearity of $g_p^{BA} \in \mathrm{GL}(d)$. The gauges ψ_p^A or ψ_p^B are then automatically vector space isomorphisms.

The most relevant examples for us are the tangent bundle and feature vector bundles, which are introduced in the following sections.

11.1.3 G -bundles

Depending on the topology of the bundle, it might be possible to define an atlas of local trivializations $\mathcal{A}^G = \{(U^X, \Psi^X) \mid X \in \mathfrak{X}\}$ whose *transition functions are restricted to a subgroup $G \leq \mathrm{Aut}(F)$* , that is, they satisfy

$$g_p^{BA} \in G \quad \text{for all } A, B \in \mathfrak{X} \text{ and all } p \in U^A \cap U^B. \quad (11.13)$$

Any such atlas is called *G -atlas* and G is denoted as *structure group* of the bundle. Two different G -atlases are equivalent (or compatible), if their union is again a G -atlas. A bundle equipped with an equivalence class of G -atlases is known as a *G -bundle*.⁷

The topology of a bundle determines how far its structure group can be reduced. For instance, the cylinder in Fig. 11.2 (or any other trivial bundle) can be described by an $\{e\}$ -atlas, consisting of local trivializations with identity transition functions only. This corresponds to a reduction to a trivial structure group $G = \{e\}$. In contrast, the twisted topology of the Möbius strip requires any G -atlas to contain transition functions which glue the fibers together in an inverted orientation; see Fig 11.2 (right). The structure group of the Möbius strip can therefore not be restricted further than the group $G = \mathcal{R}$ which models the reflection of fibers. On Riemannian manifolds the structure group of the tangent bundle TM , and thus the associated feature vector bundles, can in general not be reduced further than to

⁷The equivalence class ensures thereby that no single of the equivalent G -atlases is preferred. Equivalently, one could take the *maximal G -atlas*, defined as the unique G -atlas in which any other compatible G -atlas is contained. Note that an equivalence class of G -atlases is uniquely implied by a single given G -atlas.

an orthogonal structure group $O(d)$ which motivated this work on coordinate independent CNNs in the first place.

11.1.4 Associated G -bundles

Two G -bundles are said to be *associated* to each other if they share the same base space, structure group and, most importantly, *same transition functions*. Associated bundles (E, M, π, F) and $(\tilde{E}, M, \tilde{\pi}, \tilde{F})$ with structure group G might differ in their typical fibers F and \tilde{F} and therefore also in their left actions $\blacktriangleright : G \times F \rightarrow F$ and $\tilde{\blacktriangleright} : G \times \tilde{F} \rightarrow \tilde{F}$ of G on the respective fiber. Given two G -atlases $\{(U^X, \Psi^X) \mid X \in \mathfrak{X}\}$ and $\{(U^X, \tilde{\Psi}^X) \mid X \in \mathfrak{X}\}$ of the bundles over the same open cover of M , the requirement for the equivalence of the transition functions (up to the different left actions) means:

$$\Psi^B \circ (\Psi^A)^{-1} = (\text{id} \times g^{BA} \blacktriangleright) \iff \tilde{\Psi}^B \circ (\tilde{\Psi}^A)^{-1} = (\text{id} \times g^{BA} \tilde{\blacktriangleright}) \quad (11.14)$$

Intuitively, the typical fibers F and \tilde{F} of E and \tilde{E} are “glued together” in the same way over M .

An important example of bundles which are $GL(d)$ -associated to each other are the tangent bundle TM , the cotangent bundle T^*M , any other tensor bundle $T_s^r M$ and the tangent frame bundle FM , (the first and the latter are introduced in Section 11.2). The associatedness of these bundles is reflected in that their components relative to chosen bases transform according to the same gauge transformation (e.g. Jacobian $g_{\mu\nu}^{BA} = \frac{\partial x_\mu^B}{\partial x_\nu^A}$, see Appendix C). The different actions of a gauge transformation on the respective fibers is in this example denoted as being a contravariant transformation (TM), covariant transformation (T^*M), r -times contra- and s -times covariant transformation ($T_s^r M$) and, again, covariant transformation (FM), respectively. We will later on introduce the G -structure GM , the tangent bundle TM and the feature vector bundles \mathcal{A} as associated G -bundles. The associatedness does in this case come from the fact that changes of reference frames in GM lead to a simultaneous transformations of the tangent vector coefficients and feature vector coefficients.

We want to mention that any associated bundles are additionally associated to a uniquely specified principal G -bundle (defined in the next paragraph). In turn, any associated bundle can be constructed from the respective associated principal bundle – we will make heavy use of this construction to define feature vector bundles in Section 11.3.

11.1.5 Principal G -bundles

A fiber bundle (P, M, π, G) is called a (smooth) *principal G -bundle* $(P, M, \pi, G, \triangleleft)$ if 1) its typical fiber coincides with its structure group G and 2) it is endowed with a smooth *right G -action*

$$\triangleleft : P \times G \rightarrow G, \quad (p, g) \mapsto p \triangleleft g \quad (11.15)$$

which preserves the fibers, that is,

$$\pi(p \triangleleft g) = \pi(p) \quad \forall p \in P, g \in G \quad (11.16)$$

and acts *transitively* and *freely* on them.⁸ The last two conditions (transitivity and freedom) together require that the fibers of a principal G -bundle are G -torsors (or principal homogeneous G -spaces), which intuitively means that they “look like G ” but come without any specified origin or identity element.⁹ The local trivializations $\Psi : \pi^{-1}(U) \rightarrow U \times G$ are required to respect the right G -action, that is, to be right G -equivariant

$$\Psi(p \triangleleft g) = \Psi(p)(\text{id} \times \cdot g) \quad \forall p \in P, g \in G, \quad (11.17)$$

or, expressed locally,

$$\psi_{\pi(p)}(p \triangleleft g) = \psi_{\pi(p)}(p) \cdot g \quad \forall p \in P, g \in G,$$

where $\cdot g$ denotes the canonical right multiplication with group elements on the typical fiber G . This extends the diagram in Eq. (11.1) to the diagram

$$\begin{array}{ccc} \pi^{-1}(U) & \xrightarrow{\Psi} & U \times G \\ \downarrow \triangleleft g & & \uparrow (\text{id} \times \cdot g) \\ \pi^{-1}(U) & \xrightarrow{\Psi} & U \times G \\ \downarrow \pi & \nearrow \text{proj}_1 & \\ U & & \end{array}, \quad (11.18)$$

which is required to commute for any $g \in G$.

Principal G -bundles are of great relevance for the study of general G -bundles. In particular, any G -bundle (E, M, π_E, F) is associated to some (unique) principal G -bundle $(P, M, \pi_P, G, \triangleleft)$ over M and any associated G -bundle can be constructed from P . In the following sections we will present the frame bundle FM and G -structures GM as specific instances of principal bundles, which will make the claims made here less abstract and uncover some consequences of them.

11.1.6 Sections and fields

Smooth F -valued fields over M are formalized as smooth *sections* σ of a bundle $E \xrightarrow{\pi} M$ with fiber F . A smooth section is thereby defined as a smooth map $\sigma : M \rightarrow E$ that assigns to each point p of the base space an element in the fiber E_p over p , that is, it satisfies $\pi \circ \sigma = \text{id}_M$, which the following commutative diagram visualizes:

$$\begin{array}{ccccc} M & \xrightarrow{\sigma} & E & \xrightarrow{\pi} & M \\ & \underbrace{\hspace{10em}}_{\text{id}_M} & & & \uparrow \end{array} \quad (11.19)$$

⁸A (right) group action $\phi : X \times G \rightarrow X$, $(x, g) \mapsto x \cdot g$ is called *transitive* if any point of X can be mapped to any other point i.e. if for each $x, y \in X$ there exists a $g \in G$ such that $y = x \cdot g$. It is called (fixed point) *free* if for any $x \in X$ the equation $x = x \cdot g$ implies that $g = e$, that is, if only the action of the identity element leaves p invariant. Equivalent statements can be made for left actions; see Defs. B.3.8 and B.3.10.

⁹Formally, a (right) G -torsor P satisfies $P \times G \cong P \times_M P$ where the isomorphism is given by $(p, g) \mapsto (p, p \cdot g)$. This condition implies that there is a *unique* group element connecting *any* two points in the torsor. See also Def. B.3.13.

An important example are tangent vector fields, which are modeled as sections $v : M \rightarrow TM$ that assign a tangent vector $v(p) \in T_p M$ to each point p in M . Note that the projection map is, by its nature, non-invertible, such that $\sigma \circ \pi \neq \text{id}_E$. The following diagram does therefore in general *not* commute:

$$\begin{array}{ccccc} E & \xrightarrow{\pi} & M & \xrightarrow{\sigma} & E \\ & & \Downarrow & & \\ & & \text{id}_E & & \end{array} \quad (11.20)$$

In cases below where a diagram does not commute, which is mostly the case for sections, we emphasize this visually by adding the symbol \Downarrow . Smooth sections do not necessarily exist globally but can always be defined on trivializing neighborhoods $U \subseteq M$. Via a local trivialization, a local section can be identified with a function $s : U \rightarrow F$ by setting $s(p) = \psi_p(\sigma(p))$ for $p \in U$. We denote the space of global sections by $\Gamma(E)$ while the space of local sections is written $\Gamma(U, E)$.

11.1.7 Bundle morphisms

The morphisms (maps) in the category of fiber bundles are called *bundle morphisms* or bundle maps. They differ from mere diffeomorphisms between the total spaces in that they are additionally required to respect the bundle structure, i.e. to *map fibers to fibers*. In general, a smooth bundle map between two smooth fiber bundles (E, M, π, F) and $(\tilde{E}, \tilde{M}, \tilde{\pi}, \tilde{F})$ is a smooth map $\phi : E \rightarrow \tilde{E}$ between the total spaces such that there exists a second smooth map $\hat{\phi} : M \rightarrow \tilde{M}$ between the base spaces which satisfies $\tilde{\pi} \circ \phi = \hat{\phi} \circ \pi$, that is, the following diagram is required to commute:

$$\begin{array}{ccc} E & \xrightarrow{\phi} & \tilde{E} \\ \pi \downarrow & & \downarrow \tilde{\pi} \\ M & \xrightarrow{\hat{\phi}} & \tilde{M} \end{array} \quad (11.21)$$

The map on the base space ensures that the bundle morphism maps fibers at $p \in M$ to fibers at $\hat{\phi}(p) \in \tilde{M}$ instead of “shearing them apart”. Obvious generalizations to bundle *isomorphisms* and bundle *automorphism* exist. For instance, bundle isomorphisms require ϕ and $\hat{\phi}$ to be invertible, i.e. diffeomorphisms (and to respect further structure if defined).

The specific kind of bundle map under consideration can be narrowed down further by demanding additional requirements. A *bundle M -morphism* between two bundles (E, M, π, F) and $(\tilde{E}, \tilde{M}, \tilde{\pi}, \tilde{F})$ over the same base space M is required to map fibers E_p over any $p \in M$ to fibers \tilde{E}_p over the same point p , that is, $\hat{\phi} = \text{id}_M$. In terms of a commutative diagram this reads:

$$\begin{array}{ccc} E & \xrightarrow{\phi} & \tilde{E} \\ \pi \searrow & & \swarrow \tilde{\pi} \\ & M & \end{array} \quad (11.22)$$

From this perspective, we identify the bundle trivialization in Diagram (11.1) as a bundle U -morphism Ψ between the trivial bundles $\pi^{-1}(U)$ and $U \times F$ over U .

If the fibers carry additional structure, this structure is typically required to be preserved by the bundle map. For instance, *vector bundle morphisms* ϕ between $(E, M, \pi, \mathbb{R}^k)$ and $(\tilde{E}, \tilde{M}, \tilde{\pi}, \mathbb{R}^{\tilde{k}})$ are demanded to respect the vector space structure on the fibers, and therefore to restrict to *fiber wise linear maps* $\phi|_p : E_p \rightarrow \tilde{E}_{\phi(p)}$. Similarly, *principal bundle morphisms* are required to respect the property of the fibers to be right G -torsors, i.e. to be right G -equivariant. Given two principal bundles $(P, M, \pi, G, \triangleleft)$ and $(\tilde{P}, \tilde{M}, \tilde{\pi}, \tilde{G}, \tilde{\triangleleft})$ and some group homomorphism $\theta : G \rightarrow \tilde{G}$, a principal bundle morphism is required to make the following diagram commute for any $g \in G$:

$$\begin{array}{ccc}
 P & \xrightarrow{\phi} & \tilde{P} \\
 \triangleleft g \uparrow & & \uparrow \tilde{\triangleleft} \theta(g) \\
 P & \xrightarrow{\phi} & \tilde{P} \\
 \pi \downarrow & & \downarrow \tilde{\pi} \\
 M & \xrightarrow{\hat{\phi}} & \tilde{M}
 \end{array} \tag{11.23}$$

The local trivialization of principal bundles in Diagram (11.18) is therefore identified as a principal bundle U -morphism Ψ between $\pi^{-1}(U)$ and $U \times G$, where the group homomorphism $\theta : G \rightarrow G$, $g \mapsto g$ is given by the identity on G .

Bundle morphisms are of particular importance in Chapter 13, where they describe the transformation of bundles and feature fields under the action of isometries. Coordinate independent CNNs are proven to be equivariant w.r.t. these actions on bundles and their sections.

For more background on fiber bundles in general we refer to [262, 221, 132, 283, 273, 202, 326].

11.2 The tangent bundle TM and frame bundle FM

Any differentiable (and thus any Riemannian) manifold M is canonically equipped with its tangent bundle TM and the (general) frame bundle FM , consisting of all local reference frames of the tangent spaces. The two bundles are naturally associated to each other, with their structure group a-priori given by $\text{Aut}(\mathbb{R}^d) = \text{GL}(d)$. This fact will be emphasized by “reconstructing” TM from FM via an associated bundle construction which will later allow us to define associated feature vector bundles. To clearly separate the concepts introduced and assumptions made, we will describe TM and FM here as $\text{GL}(d)$ -bundles. The following Section 11.3 will additionally assume a G -structure imposed on TM and FM , which will establish them as G -bundles. While bundles are locally trivializable by definition, we will take the specific trivializations for now as granted and postpone their exact definition to Section 11.4.

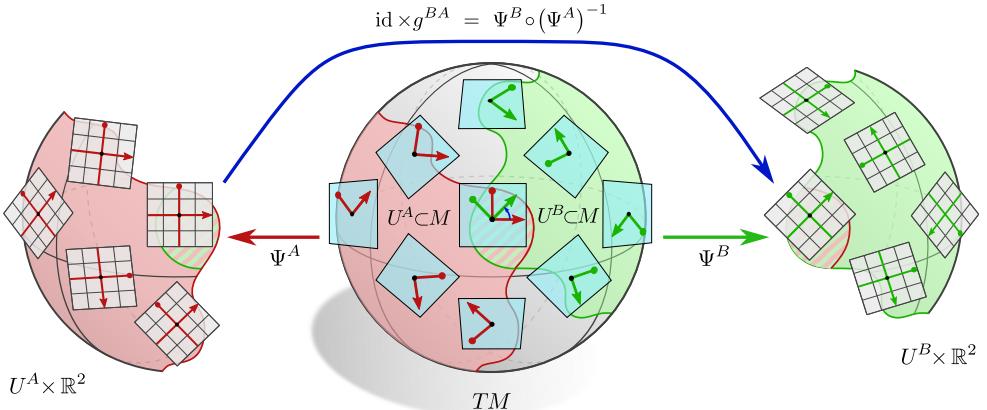


Figure 11.3: The tangent bundle TM is the bundle formed by all tangent spaces of a smooth manifold M . The tangent spaces are vector spaces isomorphic to \mathbb{R}^d , which implies local trivializations of the form $\Psi_{TM} : \pi_{TM}^{-1}(U) \rightarrow U \times \mathbb{R}^d$. Transition maps $\Psi^B_{TM} \circ (\Psi^A_{TM})^{-1} =: \text{id} \times g^{BA}$ can be viewed as a field of local gauge transformations $g_p^{BA} \in \text{GL}(d)$, translating between different identifications of the tangent spaces $T_p M$ with \mathbb{R}^d . Trivializations of TM are canonically induced by coordinate charts; see Appendix C for more details.

11.2.1 Tangent bundle TM

Any smooth manifold M comes with a set of tangent spaces $T_p M \cong \mathbb{R}^d$. Their disjoint union¹⁰

$$TM := \coprod_{p \in M} T_p M, \quad (11.24)$$

together with a canonically given smooth structure and projection map, defines a smooth fiber bundle known as the *tangent bundle*. The projection map $\pi_{TM} : TM \rightarrow M$ is thereby given by the obvious choice $\pi_{TM}(v) = p$ for $v \in T_p M$.

Local trivializations $\Psi_{TM} : \pi_{TM}^{-1}(U) \rightarrow U \times \mathbb{R}^d$ identify the tangent bundle over a trivializing neighborhood U with a product $U \times \mathbb{R}^d$; visualized in Fig. 11.3. As derived in Appendix C, any coordinate chart $x : U \rightarrow V \subseteq \mathbb{R}^d$ of the manifold induces a corresponding local trivialization, denoted as *coordinate basis*. We can therefore take the trivializability of TM for now as granted and postpone a detailed discussion to Section 11.4. A smooth structure on TM is induced from the smooth structure of M via the above mentioned trivializations from charts. We skip the technicalities on this construction and refer the interested reader to [262, 221].

The thus defined tangent bundle is a *vector bundle* since its typical fiber \mathbb{R}^d is a vector space. Tangent vector fields, describing for instance a flow on M , are formalized as sections $\sigma : M \rightarrow TM$ of the tangent bundle. Smooth global sections of vector bundles always exist; a standard example is the zero section which assigns the zero vector of $T_p M$ to each $p \in M$. We want to emphasize that the tangent spaces – and therefore the tangent bundle –

¹⁰The disjoint union $\coprod_{p \in M} T_p M = \bigcup_{p \in M} \{(p, v) \mid v \in T_p M\}$ of tangent spaces can be thought of as “remembering” from which particular tangent space $T_p M$ a certain vector $v \in TM$ originates, which is necessary for the definition of the projection map π_{TM} .

are defined without reference to coordinate frames, such that sections describe vector fields in a coordinate free way.

After introducing the tangent frame bundle FM below, we will come back to the tangent bundle and its explicit construction as associated $GL(d)$ -bundle which emphasizes its coordinate free nature. In Section 11.3 we will analogously construct TM as a associated G -bundle to a G -structure GM .

11.2.2 Frame bundle FM

The space of local reference frames of all tangent spaces $T_p M$ forms the (tangent) *frame bundle*. Consider the spaces of reference frames (ordered bases) of the individual tangent spaces $T_p M$:

$$F_p M := \{[e_1, \dots, e_d] \mid \{e_1, \dots, e_d\} \text{ is a basis of } T_p M\} \quad (11.25)$$

The frame bundle is defined as their disjoint union $FM := \coprod_{p \in M} F_p M$ together with the projection map $\pi_{FM} : FM \rightarrow M$ which sends frames in $F_p M$ to p and a smooth structure induced from TM . The typical fiber of the frame bundle is the general linear group $GL(d) \cong F_p M$, i.e. the group of invertible $d \times d$ matrices whose linearly independent columns can be thought of as defining a frame of \mathbb{R}^d . As the frame bundle is constructed from the tangent bundle, its local trivializations $\Psi_{FM} : \pi_{FM}^{-1}(U) \rightarrow U \times GL(d)$ are immediately induced from those of TM ; see Section 11.4. Fig. 11.4 shows a graphical interpretation of the frame bundle.

Smooth local sections $\sigma : U \rightarrow \pi_{FM}^{-1}(U) \subseteq FM$ of the frame bundle map points $p \in U$ to frames in $F_p M$. They define smooth local frame fields, that is, smoothly varying choices of reference frames for $T_p M$, $p \in U$; visualized in Fig. 7.2. As argued in Eq. (7.4), a choice of frame field on U is *equivalent* to a choice of gauge or local trivialization on U . This implies that global frame fields exist only if FM – and thus TM – are trivial. We will discuss this equivalence in more depth in Section 11.4.

A transitive and free right action on the individual fibers $F_p M \cong GL(d)$ of the frame bundle is naturally given by the change of frames defined in Eq. (7.10) [262]. The corresponding action

$$\begin{aligned} \lhd : FM \times GL(d) &\rightarrow FM, \\ ([e_i]_{i=1}^d, g) &\mapsto [e_i]_{i=1}^d \lhd g := \left[\sum_j e_j g_{ji} \right]_{i=1}^d \end{aligned} \quad (11.26)$$

on FM as a whole makes the frame bundle a *principal $GL(d)$ -bundle* as defined in Section 11.1.5. The lack of origin or preferred identity element of the fibers $F_p M$ as $GL(d)$ -torsors reflects the inherent ambiguity of reference frames.

11.2.3 TM as $GL(d)$ -associated vector bundle $(FM \times \mathbb{R}^d)/GL(d)$

In Section 7.1 we expressed tangent vectors in $T_p M$ in terms of their coefficients in \mathbb{R}^d relative to some reference frame. The particular choice of frames was thereby irrelevant since the transformation of the coefficients in Eq. (7.9) cancels with the transformation of reference frames in Eq. (7.10) such that $v = \sum_i v_i^A e_i^A = \sum_i v_i^B e_i^B$ are equivalent coordinate

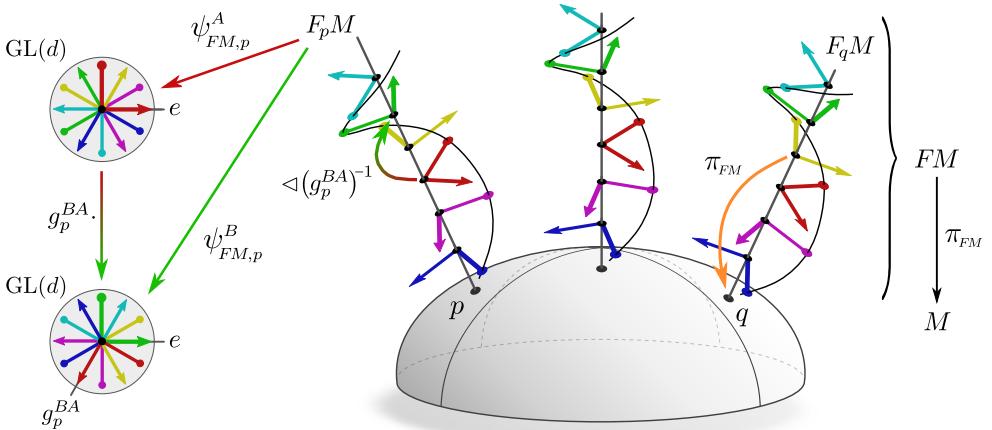


Figure 11.4: A graphical interpretation of the frame bundle FM over M and its trivializations. The fiber $F_p M$ over p is defined as the space of all possible reference frames of $T_p M$. All frames in $F_p M$ are by the projection map π_{FM} being mapped to that point p in M to which the fiber is attached. The fibers $F_p M$ are isomorphic to $GL(d)$, but come without an origin which would distinguish a preferred choice of reference frame. Gauges $\psi_{FM,p}^A : F_p M \rightarrow GL(d)$ or $\psi_{FM,p}^B : F_p M \rightarrow GL(d)$, introduced in Section 11.4 below, identify the fibers with $GL(d)$, thereby specifying a preferred frame. Different gauges are related by gauge transformations $g_p^{BA} \in GL(d)$. We need to warn the reader about two potential misconceptions: Firstly, the frames in different fibers are a-priori not identified with each other in a canonical way, which the redundant colors might suggest. Secondly, to minimize clutter, the visualization shows only right-handed, orthonormal frames instead of all possible reference frames. As we will discuss in the following Section 11.3, the shown orthonormal, right-handed frames would correspond to a G -structure GM (a principal G -subbundle of FM) for the structure group $G = SO(2)$.

representations of the same coordinate free vector $v \in T_p M$. Following this idea, one can construct the tangent bundle from the frame bundle by pairing reference frames with coefficient vectors and taking a quotient to collapse the resulting redundant descriptions of tangent vectors relative to different frames to one unique element.

In order to construct the tangent bundle in this way, consider the product $FM \times \mathbb{R}^d$ which can be seen as a fiber bundle with base space M and a typical fiber $GL(d) \times \mathbb{R}^d$. This bundle consists of pairs of (mutually unrelated) reference frames and coefficients. Motivated by the equivalent expression of tangent vectors in different reference frames we define the *equivalence relation*¹¹

$$([e_i]_{i=1}^d, v) \sim_{GL(d)} ([e_i]_{i=1}^d \triangleleft g^{-1}, g \cdot v) \quad \forall g \in GL(d) \quad (11.27)$$

on $FM \times \mathbb{R}^d$. As an equivalence relation, it partitions $FM \times \mathbb{R}^d$ into *equivalence classes* $[[e_i]_{i=1}^d, v]$. The space of these equivalence classes is the quotient space $(FM \times \mathbb{R}^d) / GL(d)$. The projection map

$$\pi_{\sim_{GL(d)}} : (FM \times \mathbb{R}^d) / GL(d) \rightarrow M, \quad [[e_i]_{i=1}^d, v] \mapsto \pi_{FM}([e_i]_{i=1}^d), \quad (11.28)$$

¹¹An *equivalence relation* on a set X is a binary relation \sim which is *reflexive* ($x \sim x$), *symmetric* ($x \sim y \Leftrightarrow y \sim x$) and *transitive* ($x \sim y \wedge y \sim z \Rightarrow x \sim z$). It defines a partitioning of X into *equivalence classes* $[x] := \{y \in X | x \sim y\}$ of elements $x \in X$. The space of equivalence classes $X/\sim := \{[x] | x \in X\}$ is called the *quotient set* of X by \sim .

which is induced from that of the frame bundle, turns $(FM \times \mathbb{R}^d)/\text{GL}(d)$ into a fiber bundle with base space M and typical fiber \mathbb{R}^d . Note that the projection map in Eq. (11.28) is well defined since it is independent of the representative of the equivalence class, i.e. $\pi_{\sim_{\text{GL}(d)}}([e_i]_{i=1}^d \triangleleft g^{-1}, g \cdot v]) := \pi_{FM}([e_i]_{i=1}^d \triangleleft g^{-1}) = \pi_{FM}([e_i]_{i=1}^d)$, where we used that the right action \triangleleft preserves the fibers of FM . The vector space structure of \mathbb{R}^d makes $(FM \times \mathbb{R}^d)/\text{GL}(d)$ to a vector bundle with linear combinations within the same fiber being defined by

$$\alpha[[e_i]_{i=1}^d, v] + \beta[[e_i]_{i=1}^d, w] := [[e_i]_{i=1}^d, \alpha v + \beta w], \quad (11.29)$$

for arbitrary $\alpha, \beta \in \mathbb{R}$ and $v, w \in \mathbb{R}^d$. This definition is easily checked to be independent of the choice of representative in both summands.

The thus defined bundle is isomorphic to the tangent bundle,

$$TM \cong (FM \times \mathbb{R}^d)/\text{GL}(d), \quad (11.30)$$

with the vector bundle M -isomorphism given by the fiber wise linear map

$$\chi : (FM \times \mathbb{R}^d)/\text{GL}(d) \rightarrow TM, \quad [[e_i]_{i=1}^d, v] \mapsto \sum_{i=1}^d v_i e_i \quad (11.31)$$

which takes some representative tuple of frame and coefficient vector from the equivalence class and maps them to the corresponding tangent vector. By the definition of the equivalence relation $\sim_{\text{GL}(d)}$, this function is independent of the choice of representative, that is, $\forall g \in \text{GL}(d) : \chi(([e_i]_{i=1}^d \triangleleft g^{-1}, g \cdot v)) = \sum_i (g \cdot v)_i ([e_j]_{j=1}^d \triangleleft g^{-1})_i = \sum_i v_i e_i$; cf. Eq. (7.12). As discussed in [262], the inverse is given by taking a tangent vector, projecting it on an arbitrary frame and taking the equivalence class.

The bundle $(FM \times \mathbb{R}^d)/\text{GL}(d)$ is by construction associated to FM as $\text{GL}(d)$ -bundle, that is, it has the same transition functions in $\text{GL}(d)$ as FM , as we will derive in Section 11.4. The construction of TM as quotient $(FM \times \mathbb{R}^d)/\text{GL}(d)$ emphasizes the *coordinate free* nature of the tangent bundle in a very intuitive way: it considers all possible choices of coordinatizations of the tangent spaces and treats them as being equivalent by taking a quotient.

11.3 G -structures GM and associated feature vector bundles \mathcal{A}

We will now introduce G -structures GM as distinguished subsets of frames in FM , which encode additional geometric structure on M that is to be respected by coordinate independent CNNs. The tangent bundle is via a similar associated bundle construction to that in the last section reintroduced as an associated G -bundle. This approach can be generalized to construct any other associated G -bundle, which we use to define the feature vector bundles \mathcal{A} . All such constructed bundles are associated to each other, that is, they differ only in their fiber F but share the same base space M , structure group G and transition functions g^{BA} between trivializing neighborhoods. The local trivializations of the bundles and their mutual gauge transformations are discussed in detail in the next Section 11.4.

11.3.1 G -structures GM

As discussed in Section 7.3 and Table 7.1, it is often possible to work with a *distinguished subset of reference frames* which are related by the action of a *reduced structure group*

$G \leq \mathrm{GL}(d)$. This is best understood by discussing a few examples before coming to a technical definition below. For instance, a restriction to orthonormal frames

$$\begin{aligned} \mathrm{O}_p M &:= \left\{ [e_1, \dots, e_d] \mid \{e_1, \dots, e_d\} \text{ is an orthonormal basis of } T_p M \text{ w.r.t. } \eta \right\} \\ &\cong \mathrm{O}(d) \end{aligned} \quad (11.32)$$

gives rise to a principal subbundle OM of FM with structure group $\mathrm{O}(d)$. Note that the orthonormality of reference frames is judged by the metric η on M – different choices of metrics on a manifold therefore correspond to different subsets of preferred reference frames for the same structure group $\mathrm{O}(d)$. As a second example, consider a choice of orientation on an orientable manifold, which allows to specify a preferred notion of frames¹²

$$\begin{aligned} \mathrm{GL}_p^+ M &:= \left\{ [e_1, \dots, e_d] \mid \{e_1, \dots, e_d\} \text{ is a positively oriented basis of } T_p M \right\} \\ &\cong \mathrm{GL}^+(d) \end{aligned} \quad (11.33)$$

and a corresponding principal subbundle $\mathrm{GL}^+(d)M$ of FM with structure group $\mathrm{GL}^+(d)$. Again, the two different choices of orientations correspond to two different choices of subbundles of accordingly oriented frames. Combining both requirements for the orthonormality and right-handedness of frames results in an $\mathrm{SO}(d)$ -structure with fibers

$$\begin{aligned} \mathrm{SO}_p M &:= \left\{ [e_1, \dots, e_d] \mid \{e_1, \dots, e_d\} \text{ is a positively oriented,} \right. \\ &\quad \left. \text{orthonormal basis of } T_p M \right\} \cong \mathrm{SO}(d), \end{aligned} \quad (11.34)$$

Fig. 11.4 can be thought of as showing an $\mathrm{SO}(2)$ -structure since only right-handed, orthonormal frames are shown (the typical fiber $\mathrm{GL}(d)$ should then be labeled $\mathrm{SO}(2)$). Different choices of $\mathrm{SO}(d)$ -structures correspond either to an opposite handedness of frames, sticking to the same notion of orthonormality, or to a different choice of metric (or both). The exact same pattern repeats for volume forms ω (on orientable manifolds M): they allow to specify a preferred notion of frames

$$\begin{aligned} \mathrm{SL}_p M &:= \left\{ [e_1, \dots, e_d] \mid \{e_1, \dots, e_d\} \text{ is a basis of } T_p M \text{ with unit volume w.r.t. } \omega \right\} \\ &\cong \mathrm{SL}(d) \end{aligned} \quad (11.35)$$

and thus principal subbundles $\mathrm{SL}M$ of FM with structure group $\mathrm{SL}(d)$. The specific set of frames which are preferred depends here on the specific choice of volume form. As a last example, consider $\{e\}$ -structures, corresponding to a trivial structure group $G = \{e\}$ and therefore consisting of one single frame at each point p . By definition, $\{e\}$ -structures are equivalent to global (smooth) frame fields $\sigma \in \Gamma(FM)$:

$$\{e\}_p M := \left\{ [e_1, \dots, e_d] = \sigma(p) \right\} \cong \{e\} \quad (11.36)$$

They do therefore only exist on trivial manifolds. Figs. 13.3a and 13.3b visualize two different choices of $\{e\}$ -structures $\{e\}M$ on $M = \mathbb{R}^2$.

All of these examples represent specific choices of G -structures GM on M . In general, a G -structure on M is a principal G -subbundle of FM , that is, a “smoothly varying” choice of subsets $G_p M \subseteq F_p M$ which are right G -torsors w.r.t. \lhd for any $p \in M$ [284, 230, 59].¹³ The smoothness can hereby be formalized by requiring that around each frame $[e_i]_{i=1}^d \in$

¹²Conversely, non-orientable manifolds do not allow for a reduction of structure group to $\mathrm{GL}^+(d)$.

¹³As $F_p M$ is a right $\mathrm{GL}(d)$ -torsor, any G -orbit $G_p M$ in $F_p M$ is automatically guaranteed to be a right G -torsor.

$G_p M$ there exists a neighborhood U of p on which a smooth section $\sigma : U \mapsto \pi_{GM}^{-1}(U) \subseteq GM$ with $\sigma(p) = [e_i]_{i=1}^d$ exists. The projection

$$\pi_{GM} := \pi_{FM}|_{GM} : GM \rightarrow M \quad (11.37)$$

of GM is hereby simply given by the restriction of the projection map of FM to GM . Together with the restriction

$$\triangleleft : GM \times G \rightarrow GM, \quad ([e_i]_{i=1}^d, g) \mapsto [e_i]_{i=1}^d \triangleleft g := \left[\sum_j e_j g_{ji} \right]_{i=1}^d \quad (11.38)$$

of the right action of $GL(d)$ on FM in Eq. (11.26) to an action of $G \leq GL(d)$ on $GM \subseteq FM$, this makes the G -structure to a *principal G -bundle* $GM \xrightarrow{\pi_{GM}} M$. However, it is important to note that there are *multiple choices* of such subbundles, corresponding to different G -structures for the same structure group G ; compare this claim with the examples above. As discussed earlier, the topology of a bundle might obstruct the reduction to a structure group G , and thus the existence of a corresponding G -structure GM .

While the above definition of G -structures would be sufficient, it is instructive to briefly review some alternative, equivalent definitions. The claim that GM is a principal G -subbundle of FM is made precise by defining it as a tuple (P, \mathcal{E}) consisting of a choice of an (also non-unique) principal G -bundle P over M together with a smooth, right G -equivariant embedding $\mathcal{E} : P \rightarrow FM$ (over M).¹⁴ This is visualized by the following diagram, which is required to commute for any $g \in G$:

$$\begin{array}{ccc} P & \xleftarrow{\mathcal{E}} & FM \\ \triangleleft_P g \uparrow & & \uparrow \triangleleft g \\ P & \xleftarrow{\mathcal{E}} & FM \\ & \searrow \pi_P & \swarrow \pi_{FM} \\ & M & \end{array} \quad (11.39)$$

Different subsets of preferred frames correspond in this viewpoint to different choices of embeddings $GM = \mathcal{E}(P)$ of P in FM . G -structures are furthermore equivalent to sections of the form $s : M \mapsto FM/G$ with $GM = s(M)$, which emphasizes that $G_p M = s(p) \in F_p M/G$ is indeed a choice of G -orbit in $F_p M$ as stated in footnote 13. Yet another definition of G -structures is in terms of (equivalence classes of) G -atlases [326]. As this is the viewpoint which might be taken in an implementation of GM -convolutions, we discuss it in more detail in the following Section 11.4. For the interested reader we want to mention that G -structures are a specific case of the more general concept of a *reduction (or lift) of structure groups* [284, 230, 59].

G -structures are of pivotal importance for the theory of GM -convolutions. The particular choice of G -structure determines the specific set of reference frames over which the G -steerable template kernel is shared. By the gauge equivariance of the kernels, GM -convolutions are guaranteed to respect the G -structure, i.e. to be *GM -coordinate independent*. As derived in Chapter 13, the isometries with respect to which a GM -convolution is equivariant are exactly those which preserve the G -structure (i.e. those which induce automorphisms of GM).

¹⁴The embedding is a principal G -bundle M -morphism as introduced in Section 11.1.7, with the group homomorphism $\theta : G \rightarrow GL(d)$ being the canonical inclusion of the subgroup $G \leq GL(d)$ into $GL(d)$.

11.3.2 TM as G -associated vector bundle $(GM \times \mathbb{R}^d)/G$

Given a G -structure GM , one can adapt the associated $\mathrm{GL}(d)$ -bundle construction of TM from FM in Section 11.2 to a similar associated G -bundle construction of TM based on GM . Instead of expressing tangent vectors relative to general frames in FM , they will thereby be expressed relative to the distinguished frames in GM and the quotient is taken w.r.t. the reduced structure group G instead of $\mathrm{GL}(d)$. The resulting bundle is by design associated to GM (or to FM with a G -atlas, which is equivalent as explained in the next section) and therefore has transition functions which take values in G . The restriction of χ in Eq. (11.31) to $(GM \times \mathbb{R}^d)/G$ yields a vector bundle isomorphism

$$TM \cong (GM \times \mathbb{R}^d)/G. \quad (11.40)$$

While all three bundles TM , $(FM \times \mathbb{R}^d)/\mathrm{GL}(d)$ and $(GM \times \mathbb{R}^d)/G$ are thus isomorphic as vector bundles, they are only isomorphic as associated G -bundles if TM and $(FM \times \mathbb{R}^d)/\mathrm{GL}(d)$ are endowed with a G -structure (or G -atlas), which is a-priori not the case. In contrast, the bundle $(GM \times \mathbb{R}^d)/G$ comes with a G -structure by design. For a precise definition of associated G -bundle isomorphisms we refer to [262].

11.3.3 Associated feature vector bundles \mathcal{A}

The associated G -bundle construction $(GM \times \mathbb{R}^d)/G$ can be generalized to attach other fibers with other group actions to the G -structure GM . Indeed, *any* bundle associated to GM can be constructed in this way. Important examples in differential geometry are the cotangent bundle T^*M with its typical fiber being the dual \mathbb{R}^{d*} of \mathbb{R}^d , acted on by the dual action, or the (r, s) tensor bundles $T_s^r M$ with fibers $(\mathbb{R}^d)^{\otimes r} \otimes (\mathbb{R}^{d*})^{\otimes s}$ being acted on by the corresponding tensor product representation of G .

In the following we consider *associated feature vector bundles* with feature vector coefficients \mathbb{R}^c as typical fibers. Under gauge transformations, these fibers are acted on from the left by a multiplication with a group representation $\rho : G \rightarrow \mathrm{GL}(c)$, that is, Eq. (11.6) is instantiated with $\blacktriangleright_\rho : G \times \mathbb{R}^c \rightarrow \mathbb{R}^c$, $(g, f) \mapsto \rho(g)f$. Similar to before, feature vector bundles are then constructed as a quotient

$$\mathcal{A} := (GM \times \mathbb{R}^c)/\sim_\rho \quad (11.41)$$

with the equivalence relation \sim_ρ here given by

$$([e_i]_{i=1}^d, f) \sim_\rho ([e_i]_{i=1}^d \triangleleft g^{-1}, \rho(g)f) \quad \forall g \in G. \quad (11.42)$$

The elements of \mathcal{A} are the equivalence classes $[[e_i]_{i=1}^d, f]$ of feature vector coefficients relative to reference frames and are therefore *coordinate free*. A (well defined) projection map is again induced from the projection of the G -structure:

$$\pi_{\mathcal{A}} : \mathcal{A} \rightarrow M, [[e_i]_{i=1}^d, f] \mapsto \pi_{GM}([e_i]_{i=1}^d) \quad (11.43)$$

Linear combinations on the fibers are defined in analogy to Eq. (11.29). Since such defined feature vector bundles are associated to GM , their structure group is $G \leq \mathrm{GL}(d)$, as we will explicitly derive in the next Section 11.4.¹⁵ Note that this definition includes tangent vector

¹⁵The transition functions are actually taking values in $\rho(G) \leq \mathrm{GL}(c)$ instead of $G \leq \mathrm{GL}(d)$, however, the transitions are still “ G -valued” in that they are defined via a G -action, as required in Section 11.1.4.

fields and scalar fields, which can of course be processed as feature fields, for $\rho(g) = g$ and $\rho(g) = 1$, respectively.

The construction of \mathcal{A} as an associated G -bundle models GM -coordinate independent feature vectors on M – features $f_p \in \mathcal{A}$ are equivalently expressed relative to arbitrary frames in GM , with feature coefficients in different coordinatizations being related via Eq. (11.42). Such features do, however, not have a well defined coordinate expression relative to other frames that are not contained in GM . From an engineering viewpoint, the G -bundle construction is reflected in the G -steerability of convolution kernels, which ensures that measurements of features are performed *relative* to arbitrary frames in GM , but allows to discriminate between patterns whose poses are not related by a G -valued gauge transformation in an *absolute* sense.

11.3.4 Associated feature vector field and feature spaces

Smooth, coordinate free feature fields are defined as smooth global sections $f \in \Gamma(\mathcal{A})$ of the feature vector bundles, that is, as smooth maps $f : M \rightarrow \mathcal{A}$ satisfying $\pi_{\mathcal{A}} \circ f = \text{id}_M$. As discussed before, such feature fields are guaranteed to exist since vector bundles always admit smooth global sections. In the following Section 11.4 we show how a local bundle trivialization over U^A allows to represent f by a field $f^A : U^A \rightarrow \mathbb{R}^c$ of feature vector coefficients. A different trivialization over U^B will lead to a different coefficient field $f^B : U^B \rightarrow \mathbb{R}^c$ representing f locally. From the transition maps between bundle trivializations it will follow that both coefficient fields are on the overlap $U^{AB} = U^A \cap U^B$ of their domains related by $f^B(p) = \rho(g_p^{BA})f^A(p)$. The commutative diagram in Fig. 11.7 visualizes the relations between feature vector fields and their local trivializations.

The feature spaces of coordinate independent CNNs usually consist of multiple independent feature fields over the same base space. The bundle describing a feature space as a whole is the *Whitney sum* $\bigoplus_i \mathcal{A}_i$ of the feature vector bundles $\mathcal{A}_i \xrightarrow{\pi_{\mathcal{A}_i}} M$ underlying its individual fields. As such it has the same base space M , a typical fiber $\bigoplus_i \mathbb{R}^{c_i} \cong \mathbb{R}^{\sum_i c_i}$ defined as direct sum of the individual fields' fibers and is equipped with the obvious projection map. It is associated to TM , FM , GM and the \mathcal{A}_i as G -bundles and can therefore equivalently be defined as

$$\bigoplus_i \mathcal{A}_i \cong \left(GM \times \mathbb{R}^{\sum_i c_i} \right) / \sim_{\bigoplus_i \rho_i} \quad (11.44)$$

Note that the direct sum $\bigoplus_i \rho_i$ of representations ρ_i defining \mathcal{A}_i guarantees that the transition maps of $\bigoplus_i \mathcal{A}_i$ transform each individual field independently. The feature spaces are then defined as the spaces $\Gamma(\bigoplus_i \mathcal{A}_i)$ of global sections of the Whitney sum bundle.

11.4 Local bundle trivializations of TM , FM , GM and \mathcal{A}

While the global theory of coordinate independent CNNs is elegantly formalized in terms of coordinate free fiber bundles, a numerical implementation requires coordinate free feature vectors $f(p) \in \mathcal{A}_p$ to be expressed by coefficient vectors $f^A(p) := \psi_{\mathcal{A},p}^A(f(p)) \in \mathbb{R}^c$ relative to some choice of reference frame $[e_i^A]_{i=1}^d \in G_p M$ as described in Section 8.1. In the language of fiber bundles, this corresponds to a choice of local trivializations or gauges Ψ_{GM}^A , Ψ_{TM}^A , Ψ_{FM}^A and $\Psi_{\mathcal{A}}^A$, all of which transform simultaneously if GM , TM , FM and \mathcal{A} are

taken to be G -associated to each other. Recall that a local description and thus implementation via a G -atlas, consisting of local trivializations which cover M and satisfy the three conditions (11.8), (11.9) and (11.10), is fully equivalent to the global, coordinate free theory.

In this section we work out the associated trivializations of TM , FM , GM and \mathcal{A} and their synchronous gauge transformations. We start out by assuming trivializations of TM to be given and discuss how they induce trivializations of FM and corresponding local frame fields. If a G -atlas is chosen for TM and thus FM , it gives rise to a G -structure GM whose G -atlas agrees with that of FM . The local trivializations of any associated G -bundle, in particular those of the feature vector bundles \mathcal{A} , follow from those of GM . These trivializations recover the transformation law of feature fields from Section 8.1.

11.4.1 Trivializations of TM

As the tangent bundle has \mathbb{R}^d as typical fiber, its local trivializations are given by maps of the form

$$\Psi_{TM} : \pi_{TM}^{-1}(U) \rightarrow U \times \mathbb{R}^d, \quad (11.45)$$

which are visualized in Fig. 11.3. These trivializations correspond to the (spatially smoothly varying) pointwise gauges

$$\psi_{TM,p} : T_p M \rightarrow \mathbb{R}^d \quad (11.46)$$

from Eq. (7.1) by identifying $\Psi_{TM}(v) = (\pi_{TM}(v), \psi_{TM,p}(v))$ for $p = \pi_{TM}(v)$. In order to respect the vector space structures of the fiber \mathbb{R}^d and the tangent spaces $T_p M$, the trivializations Ψ_{TM} are defined as *vector bundle isomorphisms* between $\pi_{TM}^{-1}(U)$ and $U \times \mathbb{R}^d$, that is, the maps $\psi_{TM,p}$ are required to be linear and invertible (i.e. vector space isomorphisms). The transition maps between different trivializations of TM will in general take values in the general linear group $GL(d)$, the (linear) automorphism group of \mathbb{R}^d .

If further structure is specified on the tangent bundle, the trivializations are required to respect this structure. For instance, if a metric is defined on M and thus TM , the maps $\psi_{TM,p}$ are required to be isometric, i.e. to map vectors in $T_p M$ in such a way to vectors in \mathbb{R}^d that norms and angles are preserved. As the trivializations are then only allowed to differ in their direction and orientation, different trivializations are guaranteed to be related by a reduced structure group $O(d)$, corresponding to the metric as $O(d)$ -structure. More generally, a G -structure on TM requires – or is implied by – a choice of G -atlas $\{(U^X, \Psi_{TM}^X)\}_{X \in \mathfrak{X}}$. Two different trivializations Ψ_{TM}^A and Ψ_{TM}^B of such a G -atlas are on $U^A \cap U^B$ related by $\Psi_{TM}^B \circ (\Psi_{TM}^A)^{-1}$ as defined in Eq. (11.4) with G -valued transition functions

$$g^{BA} : U^A \cap U^B \rightarrow G, \quad p \mapsto \psi_{TM,p}^B \circ (\psi_{TM,p}^A)^{-1}, \quad (11.47)$$

which define the left action $\blacktriangleright : G \times \mathbb{R}^d \rightarrow \mathbb{R}^d$, $(g, v) \mapsto g \cdot v$ on the typical fiber. For a graphical intuition on the pointwise action of the transition functions on individual fibers we refer back to Fig. 7.1. A diagrammatic visualization of local trivializations of TM and their transitions is given in Fig. 11.5a.

11.4.2 Induced trivializations of FM and frame fields

Any atlas $\{(U^X, \Psi_{TM}^X)\}_{X \in \mathfrak{X}}$ of the tangent bundle is in one-to-one correspondence to an atlas $\{(U^X, \Psi_{FM}^X)\}_{X \in \mathfrak{X}}$ of the frame bundle. Specifically, given a local trivialization Ψ_{TM}^A

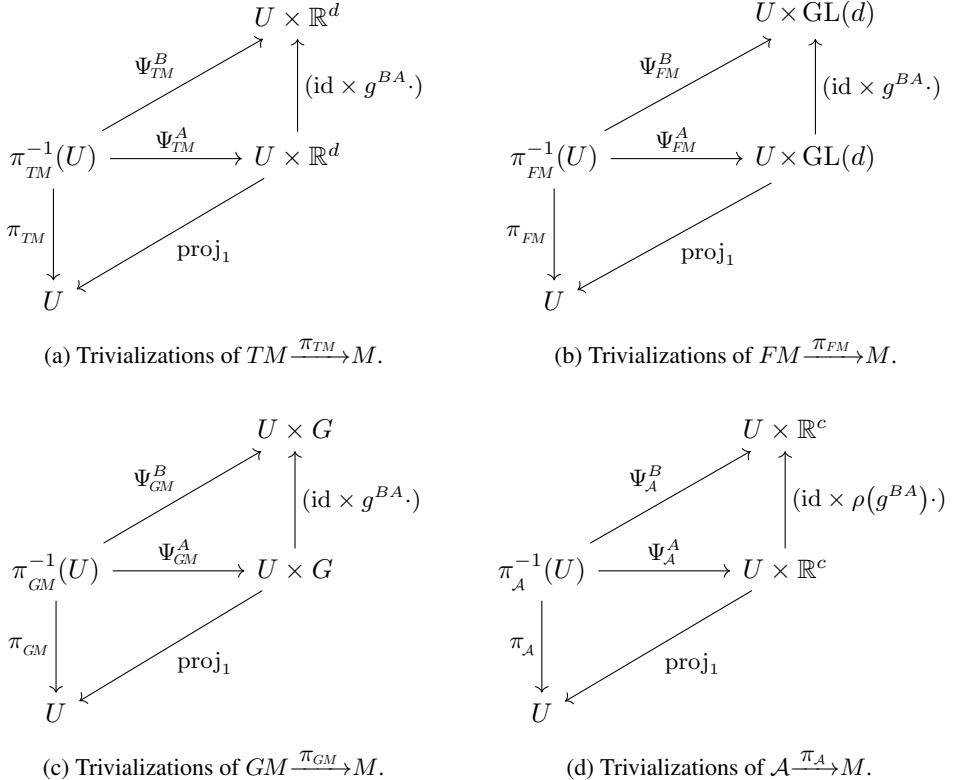


Figure 11.5: Visualization of the local trivializations of the associated G -bundles TM , FM , GM and \mathcal{A} in terms of commutative diagrams where we abbreviate $U := U^A \cap U^B$. A G -atlas $\{U^X, \Psi_{TM}^X\}$ of the tangent bundle with transition maps $g^{BA} : U \rightarrow G$ implies a G -structure GM and induces G -atlases for FM , GM and \mathcal{A} with compatible transition functions. More detailed commutative diagrams which show sections $\sigma : U \rightarrow \pi_{FM}^{-1}(U)$ and the right action \triangleleft on the frame bundle are given in Figs. 11.6a and 11.6b. Feature fields, modeled as sections $f : M \rightarrow \mathcal{A}$ of the associated feature vector bundle \mathcal{A} , and their local trivializations $f^A : U^A \rightarrow \mathbb{R}^c$ are shown in Fig. 11.7. Graphical interpretations of the commutative diagram for TM are given in Figs. 7.1 and 11.3.

of TM , a corresponding local trivialization

$$\Psi_{FM}^A : \pi_{FM}^{-1}(U^A) \rightarrow U^A \times \text{GL}(d), \quad [e_i]_{i=1}^d \mapsto (p, \psi_{FM,p}^A([e_i]_{i=1}^d)), \quad (11.48)$$

of FM , where we abbreviated $p = \pi_{FM}([e_i]_{i=1}^d)$, is induced by defining

$$\psi_{FM,p}^A : F_p M \rightarrow \text{GL}(d), \quad [e_i]_{i=1}^d \mapsto \psi_{FM,p}^A([e_i]_{i=1}^d) := (\psi_{TM,p}^A(e_i))_{i=1}^d \quad (11.49)$$

as a map from tangent frames to invertible $d \times d$ matrices whose i -th column is given by the trivialization $\psi_{TM,p}^A(e_i) \in \mathbb{R}^d$ of the i -th frame axis $e_i \in T_p M$. As required for associated

bundles, the trivializations of TM and FM share the *same transition functions*,

$$\begin{aligned}\psi_{FM,p}^B([e_i]_{i=1}^d) &= (\psi_{TM,p}^B(e_i))_{i=1}^d \\ &= (g_p^{BA} \psi_{TM,p}^A(e_i))_{i=1}^d \\ &= g_p^{BA} (\psi_{TM,p}^A(e_i))_{i=1}^d \\ &= g_p^{BA} \psi_{FM,p}^A([e_i]_{i=1}^d),\end{aligned}\tag{11.50}$$

since the action of g^{BA} on the individual trivialized frame axes in the second line agrees with its action on the trivialized frame matrix in the third line. Furthermore, as claimed for principal bundles in Eq. (11.17), the trivializations of the frame bundle are *right $GL(d)$ -equivariant*, that is, for any $h \in GL(d)$ one has:

$$\begin{aligned}\psi_{FM,p}^A([e_i]_{i=1}^d \triangleleft h) &= \psi_{FM,p}^A \left(\left(\sum_j e_j h_{ji} \right)_{i=1}^d \right) \\ &= \left(\psi_{TM,p}^A \left(\sum_j e_j h_{ji} \right) \right)_{i=1}^d \\ &= \left(\sum_j \psi_{TM,p}^A(e_j) h_{ji} \right)_{i=1}^d \\ &= (\psi_{TM,p}^A(e_i))_{i=1}^d \cdot h \\ &= \psi_{FM,p}^A([e_i]_{i=1}^d) \cdot h\end{aligned}\tag{11.51}$$

Here we used the linearity of $\psi_{TM,p}^A$ in the third step and identified the index expression as a right matrix multiplication in the fourth step. Fig. 11.6a summarizes the left action on the trivialization via transition functions $\Psi_{FM}^B \circ (\Psi_{FM}^A)^{-1} = (\text{id} \times g^{BA} \cdot)$ as derived in Eq. (11.50) and the right equivariance $\Psi_{FM}^A \circ (\triangleleft h) = (\text{id} \times \cdot h) \circ \Psi_{FM}^A$ of the trivializations as derived in Eq. (11.51).

As indicated in Eq. (7.4) and visualized in Figs. 7.1 and 7.2, a smooth local trivialization Ψ_{TM}^A on U^A of the tangent bundle induces a *frame field* on U^A . It is formalized as a smooth *local section*

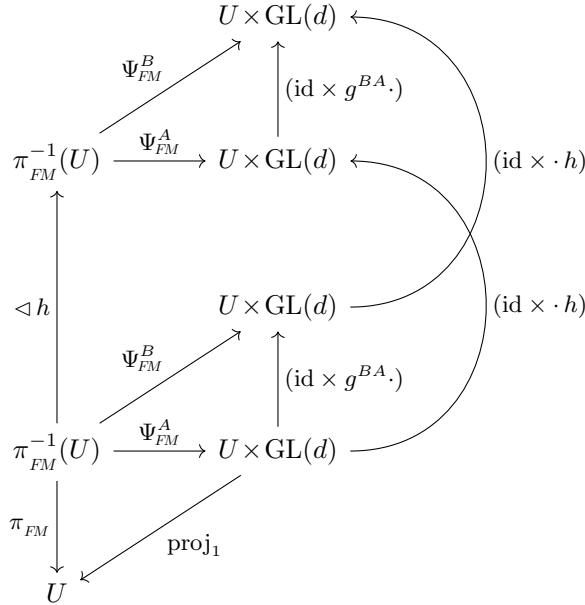
$$\sigma^A : U^A \rightarrow \pi_{FM}^{-1}(U^A), \quad p \mapsto \left[(\psi_{TM,p}^A)^{-1}(\epsilon_i) \right]_{i=1}^d\tag{11.52}$$

of the frame bundle, defined by mapping the standard frame vectors ϵ_i of \mathbb{R}^d back to the tangent spaces in $\pi_{TM}^{-1}(U^A) \subseteq TM$. Following Eq. (7.10), a gauge transformation from Ψ_{TM}^A to $\Psi_{TM}^B = (\text{id} \times g^{BA} \cdot) \Psi_{TM}^A$ corresponds to a transformation

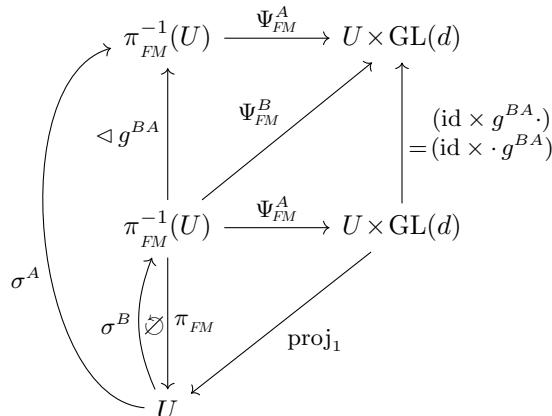
$$\sigma^B(p) = \sigma^A(p) \triangleleft (g_p^{BA})^{-1}\tag{11.53}$$

of sections on U^{AB} . Being defined in terms of Ψ_{TM}^A , the trivializations Ψ_{FM}^A of FM have the nice property that they map the corresponding sections σ^A to the identity frame $e \in GL(d) \subset \mathbb{R}^{d \times d}$ of \mathbb{R}^d , which can be seen by inserting both definitions:

$$\begin{aligned}\psi_{FM,p}^A \circ \sigma^A(p) &= \psi_{FM,p}^A \left(\left[(\psi_{TM,p}^A)^{-1}(\epsilon_i) \right]_{i=1}^d \right) \\ &= \left(\psi_{TM,p}^A \circ (\psi_{TM,p}^A)^{-1}(\epsilon_i) \right)_{i=1}^d = (\epsilon_i)_{i=1}^d = e\end{aligned}\tag{11.54}$$



(a) The trivializations of the frame bundle are right equivariant, i.e. they satisfy $\Psi_{FM} \circ \triangleleft h = (\text{id} \times \cdot h) \circ \Psi_{FM}$ for any $h \in \text{GL}(d)$.



(b) If identity sections σ^A and σ^B are added to the diagram, the left and right actions agree with each other since $\psi_{FM,p}^A \circ \sigma^A(p) = e$ and $g \cdot e = e \cdot g \quad \forall g \in \text{GL}(d)$.

Figure 11.6: Extended diagrams of the frame bundle trivializations which capture the interplay of the transition functions g^{BA} , the right actions $\triangleleft h$ and $\cdot h$ and the identity sections σ^A and σ^B . As before, we abbreviate $U = U^{AB} = U^A \cap U^B$. Except for $\sigma^A \circ \pi_{FM} \neq \text{id}_{FM}$ and $\sigma^B \circ \pi_{FM} \neq \text{id}_{FM}$, the diagrams commute. If the trivializations are part of some G -atlas, similar diagrams, with FM and $\text{GL}(d)$ being replaced by GM and G , apply to the corresponding G -structure.

This property is often used to define sections of FM given trivializations Ψ_{FM}^A as

$$\sigma^A : U^A \rightarrow \pi_{FM}^{-1}(U^A), \quad p \mapsto (\Psi_{FM}^A)^{-1}(p, e) = (\psi_{FM,p}^A)^{-1}(e), \quad (11.55)$$

which ultimately coincides with our definition in Eq. (11.52). Since σ^A and Ψ_{FM}^A constructed this way imply each other they are sometimes called *identity sections* and *canonical local trivializations*. Extending the diagram in Fig. 11.6a with identity sections σ^A and σ^B , related by Eq. (11.53), fixes $h = g^{BA}$ and thus leads to the commutative diagram in Fig. 11.6b. The left and right multiplications with g^{BA} on the typical fiber $GL(d)$ coincide hereby only since $\psi_{FM,p}^A \circ \sigma^A = \psi_{FM,p}^B \circ \sigma^B = e$ for which $g^{BA} \cdot e = g^{BA} = e \cdot g^{BA}$. Compare Fig. 11.6b to Fig. 11.4, which shows the left gauge action g_p^{BA} on $GL(d)$ and the right action $\triangleleft(g_p^{BA})^{-1}$ of the inverse group element which transforms between the corresponding identity section frames.

11.4.3 G -atlas induced G -structure GM

The agreement of the transition functions of the tangent bundle and the frame bundle in Eq. (11.50) implies that a G -atlas of TM induces a G -atlas for FM . As we will derive in the following, such G -atlases fix a corresponding G -structure GM , i.e. a principal G -subbundle of FM , consisting of preferred frames.

To motivate the definition of GM in terms of a given G -atlas $\{(U^X, \Psi_{FM}^X)\}_{X \in \mathfrak{X}}$ of FM , consider two of its local trivializations Ψ_{FM}^A and Ψ_{FM}^B with overlapping domains and let $p \in U^A \cap U^B$. The trivializations define reference frames $\sigma^A(p)$ and $\sigma^B(p)$ in $F_p M$, which are according to Eq. (11.53) related by the right action of some element g_p^{BA} of the reduced structure group $G \leq GL(d)$. Any such defined frame is therefore seen to be an element of a G -orbit $G_p M \cong G$ in $F_p M \cong GL(d)$. Specifically, expressing the identity sections via Eq. (11.55) as $\sigma^A(p) = (\psi_{FM,p}^A)^{-1}(e)$ and $\sigma^B(p) = (\psi_{FM,p}^B)^{-1}(e) = (g_p^{BA} \psi_{FM,p}^A)^{-1}(e) = (\psi_{FM,p}^A)^{-1}((g_p^{BA})^{-1})$ suggests the pointwise definition of the G -structure in terms of inverse images of G by (arbitrary) gauge maps:

$$G_p M := \{(\psi_{FM,p}^A)^{-1}(g) \mid g \in G\} = (\psi_{FM,p}^A)^{-1}(G) \quad (11.56)$$

The independence from the chosen gauge of the G -atlas is clear as any other choice $(\psi_{FM,p}^B)^{-1}(G) = (\psi_{FM,p}^A)^{-1}((g_p^{BA})^{-1}G) = (\psi_{FM,p}^A)^{-1}(G)$ would yield the same result. As one can easily check, $G_p M$ is indeed a right G -torsor since G is a right G -torsor and $\psi_{FM,p}^A$ is by Eq. (11.51) a right $GL(d)$ -equivariant – and thus in particular right G -equivariant – isomorphism. The required smoothness of $GM = \coprod_{p \in M} G_p M$ follows from the smoothness of the trivializations Ψ_{FM}^A .

A G -atlas of local trivializations of GM is given by restricting the trivializations in the G -atlas of FM to frames in GM , that is,

$$\Psi_{GM}^A := \Psi_{FM}^A|_{\pi_{GM}^{-1}(U^A)} : \pi_{GM}^{-1}(U^A) \rightarrow U^A \times G, \quad (11.57)$$

or, locally,

$$\psi_{GM,p}^A := \psi_{FM,p}^A|_{G_p M} : G_p M \rightarrow G. \quad (11.58)$$

It follows immediately that the G -valued transition functions agree with those of TM and FM , that is,

$$\psi_{GM,p}^B([e_i]_{i=1}^d) = g_p^{BA} \psi_{GM,p}^A([e_i]_{i=1}^d), \quad (11.59)$$

and that the trivializations are right G -equivariant:

$$\psi_{GM,p}^A([e_i]_{i=1}^d \triangleleft h) = \psi_{GM,p}^A([e_i]_{i=1}^d) \cdot h \quad \forall h \in G \quad (11.60)$$

The frame fields are also given by an equivalent expression

$$\sigma^A(p) = (\psi_{GM,p}^A)^{-1}(e) \quad (11.61)$$

to that in Eq. (11.55). The commutative diagrams in Figs. 11.6a and 11.6b hold as well when replacing FM with GM and $GL(d)$ with G .

11.4.4 Induced trivializations of associated bundles \mathcal{A}

A G -atlas $\{(U^X, \Psi_{\mathcal{A}}^X)\}_{X \in \mathfrak{X}}$ consisting of local trivializations $\Psi_{\mathcal{A}}^X : \pi_{\mathcal{A}}^{-1}(U^X) \rightarrow U^X \times \mathbb{R}^c$ of the associated feature vector bundles $\mathcal{A} = (GM \times \mathbb{R}^c)/\sim_\rho$ is induced from the corresponding trivializations Ψ_{GM}^X of the G -structure. In order to construct these trivializations, recall that \mathcal{A} is defined in terms of equivalence classes $[[e_i]_{i=1}^d, f]$ consisting of pairs of reference frames and feature coefficient vectors which are related by the equivalence relation \sim_ρ defined in Eq. (11.42). A natural idea is thus to trivialize $[[e_i]_{i=1}^d, f] \in \mathcal{A}_p$ by picking one representative of its equivalent coefficient vectors in \mathbb{R}^c . A preferred choice of representative is hereby given by that coefficient vector belonging to the identity section frame $\sigma^A(p)$ corresponding to Ψ_{GM}^A .

Let $[e_i]_{i=1}^d := \sigma^A(p) \triangleleft h \in G_p M$ be some frame that is defined by an offset $h \in G$ relative to section σ^A . This offset can be recovered by the trivialization of the G -structure:

$$\psi_{GM,p}^A([e_i]_{i=1}^d) = \psi_{GM,p}^A(\sigma^A(p) \triangleleft h) = \psi_{GM,p}^A(\sigma^A(p)) \cdot h = h \quad (11.62)$$

Here we used the right G -equivariance of $\psi_{GM,p}^A$ and that σ^A is defined as identity section; see Eqs. (11.60) and (11.54), the latter adapted to $\psi_{GM,p}^A$. We can therefore rewrite any frame via its offset as:

$$[e_i]_{i=1}^d = \sigma^A(p) \triangleleft \psi_{GM,p}^A([e_i]_{i=1}^d) \quad (11.63)$$

Similarly, we can rewrite any feature vector $[[e_i]_{i=1}^d, f] \in \mathcal{A}_p$ by different representatives of the equivalence class:

$$[[e_i]_{i=1}^d, f] = [\sigma^A(p) \triangleleft \psi_{GM,p}^A([e_i]_{i=1}^d), f] = [\sigma^A(p), \rho(\psi_{GM,p}^A([e_i]_{i=1}^d)) f] \quad (11.64)$$

Based on these insights we define induced trivializations of \mathcal{A} by setting

$$\begin{aligned} \Psi_{\mathcal{A}}^A : \pi_{\mathcal{A}}^{-1}(U^A) &\rightarrow U^A \times \mathbb{R}^c, \\ [[e_i]_{i=1}^d, f] &\mapsto \left(\pi_{GM}([e_i]_{i=1}^d), \psi_{\mathcal{A},p}^A([[e_i]_{i=1}^d, f]) \right), \end{aligned} \quad (11.65)$$

$$\begin{array}{ccccc}
& & U \times \mathbb{R}^c & \xrightarrow{\text{proj}_2} & \mathbb{R}^c \\
& \nearrow \Psi_A^B & \uparrow (\text{id} \times \rho(g^{BA}) \cdot) & & \uparrow \rho(g^{BA}). \\
\pi_A^{-1}(U) & \xrightarrow{\Psi_A^A} & U \times \mathbb{R}^c & \xrightarrow{\text{proj}_2} & \mathbb{R}^c \\
\downarrow f \circ \pi_A & \nearrow \text{proj}_1 & \nearrow f^B & & \\
U & \xrightarrow{f^A} & & &
\end{array}$$

Figure 11.7: Coordinate free feature fields are defined as global sections $f \in \Gamma(\mathcal{A})$. On local neighborhoods U^A and U^B they trivialize to fields of feature coefficient vectors $f^A : U^A \mapsto \mathbb{R}^c$ and $f^B : U^B \mapsto \mathbb{R}^c$ which are on $U = U^A \cap U^B$ related by $f^B(p) = \rho(g_p^{BA})f^A(p)$. Except for $f \circ \pi_A \neq \text{id}_{\mathcal{A}}$, the diagram commutes.

with

$$\psi_{\mathcal{A},p}^A : \mathcal{A}_p \rightarrow \mathbb{R}^c, \quad (11.66)$$

$$[[e_i]_{i=1}^d, f] = [\sigma^A(p), \rho(\psi_{GM,p}^A([e_i]_{i=1}^d))f] \mapsto \rho(\psi_{GM,p}^A([e_i]_{i=1}^d))f,$$

which picks that specific representative coefficient vector $f^A = \rho(\psi_{GM,p}^A([e_i]_{i=1}^d))f \in \mathbb{R}^c$ that is distinguished by the frame $\sigma^A(p)$ corresponding to the chosen gauge. For later convenience we note that this implies in particular that the inverse of Eq. (11.66) is given by

$$(\psi_{\mathcal{A},p}^A)^{-1} : \mathbb{R}^c \rightarrow \mathcal{A}_p : f \mapsto [\sigma^A(p), f]. \quad (11.67)$$

The such defined trivialization is independent of the chosen representative since for any $k \in G$ we have:

$$\begin{aligned}
\psi_{\mathcal{A},p}^A \left([[e_i]_{i=1}^d \triangleleft k^{-1}, \rho(k)f] \right) &= \rho(\psi_{GM,p}^A([e_i]_{i=1}^d \triangleleft k^{-1}))\rho(k)f \\
&= \rho(\psi_{GM,p}^A([e_i]_{i=1}^d) \cdot k^{-1})\rho(k)f \\
&= \rho(\psi_{GM,p}^A([e_i]_{i=1}^d))f \\
&= \psi_{\mathcal{A},p}^A \left([[e_i]_{i=1}^d, f] \right)
\end{aligned} \quad (11.68)$$

By construction, the transition functions are given by $\rho(g_p^{BA})$:

$$\begin{aligned}
\psi_{\mathcal{A},p}^B \left([[e_i]_{i=1}^d, f] \right) &= \rho(\psi_{GM,p}^B([e_i]_{i=1}^d))f \\
&= \rho(g_p^{BA}\psi_{GM,p}^A([e_i]_{i=1}^d))f \\
&= \rho(g_p^{BA})\rho(\psi_{GM,p}^A([e_i]_{i=1}^d))f \\
&= \rho(g_p^{BA})\psi_{\mathcal{A},p}^A \left([[e_i]_{i=1}^d, f] \right)
\end{aligned} \quad (11.69)$$

If the tangent bundle is taken as a G -associated vector bundle $TM \cong (GM \times \mathbb{R}^d)/G$, its trivializations are recovered from Eq. (11.65) for the specific choice $\rho(g) = g$.

Assume a coordinate free feature field $f \in \Gamma(\mathcal{A})$ to be given. Relative to gauge $\Psi_{\mathcal{A}}^A$, it can be locally represented as a coefficient vector field $f^A : U^A \rightarrow \mathbb{R}^c$ by defining

$$f^A := \text{proj}_2 \circ \Psi_{\mathcal{A}}^A \circ f \quad (11.70)$$

which is equivalent to the pointwise definition

$$f^A(p) = \psi_{\mathcal{A},p}^A \circ f(p). \quad (11.71)$$

As apparent from the commutative diagram in Fig. 11.7, the transition functions in Eq. (11.69) carry over to the local coefficient fields such that we get

$$f^B(p) = \rho(g_p^{BA}) f^A(p) \quad (11.72)$$

for $p \in U^A \cap U^B$. This agrees with and justifies our definition of the gauge transformations of feature coefficient vectors in Part II, Eq. (8.3).

11.4.5 Summarizing remarks

The here defined local trivializations and transition functions formalize and justify the definitions of gauges and gauge transformations from Section 8.1. Local trivializations of TM and FM were shown to induce each other. If a G -atlas is chosen for either of both, it defines a G -structure GM , whose G -atlas essentially coincides with that of FM . It furthermore induces a G -atlas for any other associated bundle, including \mathcal{A} . As visualized in Fig. 11.5, the transition functions of all G -atlases for TM , FM , GM and \mathcal{A} agree, making the bundles G -associated to each other. Specifically, when switching from gauge A to gauge B , the trivializations of TM , FM and GM transform according to a left multiplication with g^{BA} while the feature vector bundle trivializations transform according to a left multiplication with $\rho(g^{BA})$; see Eqs. (11.47), (11.50), (11.59) and (Eq. (11.69)). At the same time, frame fields transform according to the right action $\triangleleft(g^{BA})^{-1}$ (Eq. (11.53)).

11.5 Parallel transporters on associated bundles

Section 8.2 gave an intuitive introduction to the parallel transport of tangent vectors and feature vectors along a path γ from $q \in M$ to $p \in M$. Here we briefly discuss how coordinate free parallel transporters on the fiber bundles induce each other and derive coordinate expressions relative to given trivializations for them. We start by assuming coordinate free transporters

$$\mathcal{P}_{TM,\gamma} : T_q M \rightarrow T_p M \quad (11.73)$$

on the tangent bundle TM to be given and explain how they *induce* transporters

$$\mathcal{P}_{FM,\gamma} : F_q M \rightarrow F_p M \quad (11.74)$$

on the frame bundle FM . If these transporters are *G-compatible* with the chosen G -structure, as discussed below, they further induce transporters

$$\mathcal{P}_{GM,\gamma} : G_q M \rightarrow G_p M \quad (11.75)$$

$$\mathcal{P}_{\mathcal{A},\gamma} : \mathcal{A}_q \rightarrow \mathcal{A}_p \quad (11.76)$$

on the associated G -bundles GM and \mathcal{A} . In practice, most convolutional networks assume either transporters that are based on the Levi-Civita connection or some trivial connection. Chapter 14 gives an overview of the transporters occurring in the applications in Part IV.

A more formal definition of bundle transporters might take a different route, starting by introducing a so-called principal Ehresmann connection on the principal G -bundle GM (which would by definition be G -compatible). Such an Ehresmann connection can either be defined by a choice of horizontal subbundle HGM of the tangent bundle TGM of GM or, equivalently, by a Lie algebra-valued connection 1-form $\omega : TGM \rightarrow \mathfrak{g}$ on GM . The transport on GM would subsequently be defined via the horizontal lift $\gamma^\uparrow : [0, 1] \rightarrow GM$ of curves $\gamma : [0, 1] \rightarrow M$ on the base space such that the tangent vectors of the lift in GM are horizontal, i.e. $\dot{\gamma}^\uparrow \in HGM$. All transporters on TM , FM and \mathcal{A} as associated G -bundles would then be induced from the transporters on the G -structure. Instead of following this formal approach, which would be rather technical and can be found in the literature [262, 326, 132, 221, 202, 273], we focus on how the different transporters interrelate by inducing each other.

11.5.1 Transport on TM

To this end, we take a shortcut by assuming the coordinate free transporters $\mathcal{P}_{TM,\gamma}$ on TM to be given. Recall that, given gauges $\Psi_{TM}^{\tilde{A}}$ on a neighborhood $U^{\tilde{A}}$ of q and Ψ_{TM}^A on a neighborhood U^A of p , the tangent vector transporter is coordinatized according to Eq. (8.6), that is,

$$g_\gamma^{A\tilde{A}} := \psi_{TM,p}^A \circ \mathcal{P}_{TM,\gamma} \circ (\psi_{TM,q}^{\tilde{A}})^{-1} \in \text{GL}(d), \quad (11.77)$$

and that its coordinatizations transform under gauge transformations at q and p according to Eq. (8.8):

$$g_\gamma^{B\tilde{B}} = g_p^{BA} g_\gamma^{A\tilde{A}} (g_q^{\tilde{B}\tilde{A}})^{-1} \quad (11.78)$$

We refer back to Eq. (8.7) for a visualization of these definitions in terms of a commutative diagram.

11.5.2 Transport on FM

Given the transporter on the tangent bundle, the transporter on the frame bundle follows immediately from the transport of individual frame axes. In equations, let $[e_i]_{i=1}^d \in F_q M$ be a frame at q , then the individual axes e_i for $i = 1, \dots, d$ are tangent vectors in $T_q M$ which can be transported via $\mathcal{P}_{TM,\gamma}$. We thus define the transporter on the frame bundle as:¹⁶

$$\mathcal{P}_{FM,\gamma} : F_q M \rightarrow F_p M, \quad [e_i]_{i=1}^d \mapsto \mathcal{P}_{FM,\gamma}([e_i]_{i=1}^d) := [\mathcal{P}_{TM,\gamma}(e_i)]_{i=1}^d \quad (11.79)$$

¹⁶The transport of a frame along γ describes a curve γ^\uparrow (horizontal lift) in FM . The space spanned by all tangent vectors $\dot{\gamma}^\uparrow$ in TFM along such curves is the horizontal subbundle HFM of TFM , mentioned above.

In order to derive the explicit form of its coordinatization $\psi_{FM,p}^A \circ \mathcal{P}_{FM,\gamma} \circ (\psi_{FM,q}^{\tilde{A}})^{-1} \in \text{GL}(d)$, consider its action on a group element $h \in \text{GL}(d)$, representing a trivialized frame of \mathbb{R}^d which is spanned by the matrix columns $h_{:,i} \in \mathbb{R}^d$, $i = 1, \dots, d$:

$$\begin{aligned}
& \left[\psi_{FM,p}^A \circ \mathcal{P}_{FM,\gamma} \circ (\psi_{FM,q}^{\tilde{A}})^{-1} \right] (h) & (11.80) \\
&= \left[\psi_{FM,p}^A \circ \mathcal{P}_{FM,\gamma} \right] \left(\left[(\psi_{TM,q}^{\tilde{A}})^{-1} (h_{:,i}) \right]_{i=1}^d \right) & (\text{def. of } \psi_{FM,p}^{\tilde{A}}, \text{Eq. (11.49)}) \\
&= \psi_{FM,p}^A \left(\left[\mathcal{P}_{TM,\gamma} \circ (\psi_{TM,q}^{\tilde{A}})^{-1} (h_{:,i}) \right]_{i=1}^d \right) & (\text{def. of } \mathcal{P}_{FM,\gamma}, \text{Eq. (11.79)}) \\
&= \left(\psi_{TM,p}^A \circ \mathcal{P}_{TM,\gamma} \circ (\psi_{TM,q}^{\tilde{A}})^{-1} (h_{:,i}) \right)_{i=1}^d & (\text{def. of } \psi_{TM,p}^A, \text{Eq. (11.49)}) \\
&= \left(g_{\gamma}^{A\tilde{A}} (h_{:,i}) \right)_{i=1}^d & (\text{triv. of } \mathcal{P}_{TM,\gamma}, \text{Eq. (11.77)}) \\
&= g_{\gamma}^{A\tilde{A}} h
\end{aligned}$$

The coordinatizations of the frame transporters are therefore equivalent to those of the tangent vector transporters in Eq. (11.77) but act on trivialized frames in $\text{GL}(d)$ instead of acting on coefficient vectors in \mathbb{R}^d . Their gauge transformations are from the commutative diagram

$$\begin{array}{ccccc}
\text{GL}(d) & \xrightarrow{g_{\gamma}^{A\tilde{A}}.} & & & \text{GL}(d) \\
\downarrow g_q^{\tilde{B}\tilde{A}}. & \swarrow \psi_{FM,q}^{\tilde{A}} & F_q M & \xrightarrow{\mathcal{P}_{FM,\gamma}} & F_p M \\
& \downarrow \psi_{FM,q}^{\tilde{B}} & & & \downarrow \psi_{FM,p}^A \\
\text{GL}(d) & & \xrightarrow{g_{\gamma}^{B\tilde{B}}.} & & \text{GL}(d)
\end{array} \quad (11.81)$$

seen to coincide with those of the coordinatized transporters on TM in Eq. (11.78).

11.5.3 Compatibility of connections and G -structures

Not any choice of connection or definition of transporters on the $\text{GL}(d)$ -bundles TM and FM is compatible with any G -structure. Specifically, a G -structure might not be closed under the transport of frames, that is, while a frame in $G_q M \subseteq F_q M$ will by $\mathcal{P}_{FM,\gamma}$ be transported to some frame in $F_p M$, this frame is *not* necessarily contained in $G_p M$.¹⁷ Relative to trivializations of GM , such an incompatibility would reflect in coordinatized transporters $g_{\gamma}^{A\tilde{A}} \notin G$, whose left multiplication is well defined on the fibers \mathbb{R}^d and $\text{GL}(d)$ of the $\text{GL}(d)$ -bundles TM and FM , but not on the fiber G of GM . If the subbundle GM is not

¹⁷In terms of a principal Ehresmann connection on FM , this is the case if the horizontal subbundle $HFM \subseteq TFM$ is not contained in $TGM \subseteq TFM$. An immediate definition of parallel transport in terms of a choice of horizontal subbundle HGM on the G -structure will always (by definition) lead to a well defined transport on GM .

closed under the parallel transport on FM , this means that no well defined corresponding transport on GM – and thus on any associated G -bundles \mathcal{A} – exists.

As an example, consider the Levi-Civita connection on Euclidean spaces, whose transporters keep tangent vectors and frames parallel in the usual sense on \mathbb{E}_d . The $\{e\}$ -structure (frame field) in Fig. 13.3a is closed under this transport, and therefore compatible. The $\{e\}$ -structure in Fig. 13.3b, on the other hand, is not closed under the transport, and thus incompatible with the Levi-Civita connection. Similarly, the $SO(2)$ -structure on S^2 in Fig. 17.2a is compatible with the Levi-Civita connection on the sphere, while the $\{e\}$ -structure in Fig. 17.2b is not.

The reader might wonder which general statements about the compatibility of connections (or transporters) and G -structures can be made. In general, the Levi-Civita connection, or any other metric connection, are compatible with the $O(d)$ -structure OM that corresponds to the metric.¹⁸ If the manifold is orientable, the Levi-Civita connection is furthermore compatible with any $SO(d)$ -structure that corresponds to the metric. An example is the $SO(2)$ -structure on S^2 in Fig. 17.2a. A necessary (but not sufficient) condition for a G -structure to be compatible with a given connection is that the holonomy group of the connection is a subgroup of the structure group G .

An important special case is that of $\{e\}$ -structures, since they imply a *unique trivial connection*.^{19,20} The corresponding transporters move frames in such a way that the stay parallel with the frames of the $\{e\}$ -structure. Trivial connections might not seem to be of particular importance for the theory of GM -convolutions, however, they are actually utilized by many convolutional networks. Specifically, any network that relies on an $\{e\}$ -structure is implicitly assuming a trivial connection. This includes all of the models in Table 14.1 with $G = \{e\}$, specifically those which are reviewed in Sections 17.3 and 18.3.²¹ Note that these models assume the trivial connection only for their feature vector transport but compute geodesics for the transporter pullback, Eq. (9.21), based on the original Levi-Civita connection.

11.5.4 Transport on GM

Assuming that GM is compatible with (i.e. closed under) the transport on FM , a well defined transporter is given by restricting the frame bundle transporter to the G -structure:

$$\mathcal{P}_{GM,\gamma} := \mathcal{P}_{FM,\gamma}|_{GM} : G_q M \rightarrow G_p M \quad (11.82)$$

The transition functions between different coordinatizations of $\mathcal{P}_{GM,\gamma}$ do then agree with those of $\mathcal{P}_{FM,\gamma}$ and thus also $\mathcal{P}_{TM,\gamma}$. We obtain the following commutative diagram, which visualizes the restriction of the diagram in Eq. (11.81) from $F_q M$, $F_p M$ and $GL(d)$ to $G_q M$,

¹⁸This statement holds by definition since metric connections preserve angles and lengths between vectors and thus the orthonormality of frames. One can furthermore define metric connections as principal Ehresmann connections on OM .

¹⁹A connection is *trivial* if its holonomy group, i.e. its parallel transport around any closed loop, is trivial [62].

²⁰Only one principal Ehresmann connection $H\{e\}M = T\{e\}M$ can be chosen on $\{e\}M$ since the vertical subbundle $V\{e\}M$ is the zero-section of $T\{e\}M$.

²¹These models are *implicitly* assuming a trivial connection by not modeling non-trivial transporters of feature vectors: they accumulate feature vector coefficients without transforming them.

$G_p M$ and G :

$$\begin{array}{ccc}
 G & \xrightarrow{g_\gamma^{A\tilde{A}}} & G \\
 \downarrow g_q^{\tilde{B}\tilde{A}} & \nearrow \psi_{GM,q}^{\tilde{A}} & \downarrow g_p^{BA} \\
 G_q M & \xrightarrow{\mathcal{P}_{GM,\gamma}} & G_p M \\
 \downarrow \psi_{GM,q}^{\tilde{B}} & & \downarrow \psi_{GM,p}^A \\
 G & \xrightarrow{g_\gamma^{B\tilde{B}}} & G
 \end{array} \quad (11.83)$$

We will in the remainder of this work assume that the transport on GM is well defined.

11.5.5 Transport on \mathcal{A}

If the transporters of a connection are well defined on GM , they induce transporters on any associated G -bundle, including the feature vector bundles $\mathcal{A} = (GM \times \mathbb{R}^c)/\sim_\rho$. Let $f_q := [[e_i]_{i=1}^d, f]$ be a coordinate free feature vector in \mathcal{A}_q . Its parallel transport is given by that equivalence class defined by keeping some representative coefficients $f \in \mathbb{R}^c$ fixed and transporting the corresponding frame $[e_i]_{i=1}^d$:

$$\mathcal{P}_{\mathcal{A},\gamma} : \mathcal{A}_q \rightarrow \mathcal{A}_p, \quad f_q \mapsto \mathcal{P}_{\mathcal{A},\gamma}(f_q) := [\mathcal{P}_{GM,\gamma}([e_i]_{i=1}^d), f] \quad (11.84)$$

In Section 8.2 we claimed that the transporter of numerical feature vector coefficients is given by $\rho(g_\gamma^{A\tilde{A}})$ provided that $g_\gamma^{A\tilde{A}} \in G$, which is the case if the transport on GM is well defined. This coordinate expression of $\mathcal{P}_{\mathcal{A},\gamma}$ can be derived by evaluating the action of $\psi_{\mathcal{A},p}^A \circ \mathcal{P}_{\mathcal{A},\gamma} \circ (\psi_{\mathcal{A},q}^{\tilde{A}})^{-1} \in \rho(G) \leq \text{GL}(c)$ on a feature coefficient vector $f \in \mathbb{R}^c$ step by step:

$$\begin{aligned}
 & [\psi_{\mathcal{A},p}^A \circ \mathcal{P}_{\mathcal{A},\gamma} \circ (\psi_{\mathcal{A},q}^{\tilde{A}})^{-1}](f) \\
 &= [\psi_{\mathcal{A},p}^A \circ \mathcal{P}_{\mathcal{A},\gamma}]([\sigma^{\tilde{A}}(q), f]) \quad (\text{def. of } (\psi_{\mathcal{A},p}^{\tilde{A}})^{-1}, \text{Eq. (11.67)}) \\
 &= \psi_{\mathcal{A},p}^A([\mathcal{P}_{GM,\gamma}(\sigma^{\tilde{A}}(q)), f]) \quad (\text{def. of } \mathcal{P}_{\mathcal{A},\gamma}, \text{Eq. (11.84)}) \\
 &= \rho(\psi_{GM,p}^A \circ \mathcal{P}_{GM,\gamma} \circ \sigma^{\tilde{A}}(q)) \cdot f \quad (\text{def. of } \psi_{\mathcal{A},p}^A, \text{Eq. (11.66)}) \\
 &= \rho(\psi_{GM,p}^A \circ \mathcal{P}_{GM,\gamma} \circ (\psi_{GM,q}^{\tilde{A}})^{-1}(e)) \cdot f \quad (\text{def. of identity section } \sigma^{\tilde{A}}, \text{Eq. (11.55)}) \\
 &= \rho(g_\gamma^{A\tilde{A}}) \cdot f \quad (\mathcal{P}_{GM,\gamma} \text{ in coordinates Eq. (11.83)})
 \end{aligned} \quad (11.85)$$

The commutative diagram

$$\begin{array}{ccccc}
 & & \rho(g_{\gamma}^{A\tilde{A}}) & & \\
 & \swarrow & & \searrow & \\
 \mathbb{R}^c & & \mathcal{A}_q & & \mathbb{R}^c \\
 \downarrow \rho(g_q^{\tilde{B}\tilde{A}}) & \nearrow \psi_{A,q}^{\tilde{A}} & \xrightarrow{\mathcal{P}_{A,\gamma}} & \nearrow \psi_{A,p}^A & \downarrow \rho(g_p^{BA}) \\
 & \searrow \psi_{A,q}^{\tilde{B}} & & \swarrow \psi_{A,p}^B & \\
 & & \rho(g_{\gamma}^{B\tilde{B}}) & & \mathbb{R}^c
 \end{array} \quad (11.86)$$

implies that the gauge transformations of the coordinatized feature vector transporters read:

$$\rho(g_{\gamma}^{B\tilde{B}}) = \rho(g_p^{BA})\rho(g_{\gamma}^{A\tilde{A}})\rho(g_q^{\tilde{B}\tilde{A}})^{-1} \quad (11.87)$$

Note that this transformation law is in agreement with that in Eq. (11.78).

Coordinate free formulation of kernel field transforms and *GM*-convolutions

The associated G -bundles introduced in Chapter 11 allow to describe feature fields – and therefore convolutional networks – on a global level. Given a sequence

$$\mathcal{A}_0 \xrightarrow{\pi_{\mathcal{A}_0}} M, \dots, \mathcal{A}_N \xrightarrow{\pi_{\mathcal{A}_N}} M \quad (12.1)$$

of G -associated feature vector bundles over M , we describe coordinate free convolutional networks as sequences

$$\Gamma(\mathcal{A}_0) \xrightarrow{L_1} \Gamma(\mathcal{A}_1) \xrightarrow{L_2} \dots \xrightarrow{L_N} \Gamma(\mathcal{A}_N) \quad (12.2)$$

of parameterized layers L_1, \dots, L_N which map between the corresponding feature spaces $\Gamma(\mathcal{A}_0), \dots, \Gamma(\mathcal{A}_N)$, i.e. between feature fields. While the field types (or transformation laws) $\rho_i : G \rightarrow \mathrm{GL}(c_i)$ of the intermediate bundles $\mathcal{A}_i := (GM \times \mathbb{R}^{c_i}) / \sim_{\rho_i}$ for $i = 1, \dots, N - 1$ have to be specified by the user as a hyperparameter, the field types $\rho_0 : G \rightarrow \mathrm{GL}(c_0)$ and $\rho_N : G \rightarrow \mathrm{GL}(c_N)$ of the network input and output are typically determined by the learning task. The modular construction of neural networks allows to restrict attention to individual layers, mapping between feature spaces $\Gamma(\mathcal{A}_{\text{in}})$ and $\Gamma(\mathcal{A}_{\text{out}})$ of dimensionality c_{in} and c_{out} and type ρ_{in} and ρ_{out} .

The main goal of this chapter is to introduce the coordinate free formulation of *GM*-convolutions, which are the central building blocks of *GM*-coordinate independent networks on Riemannian manifolds. To get started, and to introduce concepts that are required later on, we will in Section 12.1 first focus on the simpler case of 1×1 *GM-convolutions*, which apply point-like kernels. Section 12.2 shifts the focus to *GM*-convolutions and kernel field transforms with spatially extended kernels. They are parameterized in terms of smooth, *global kernel fields*, which are introduced in Section 12.2.1. *GM-convolutional kernel fields* are required to share weights between different spatial positions. In order for this weight sharing to be *GM*-coordinate independent, the template kernels that parameterize *GM*-convolutional kernel fields are required to be *G-steerable* (Eq. (12.28)). The actual kernel field transforms and *GM*-convolutions are introduced in Section 12.2.2. Their global definition is guided by replacing the local coordinate expressions from Section 9.2 with their global, coordinate free counterparts. As shown in Section 12.2.3, these coordinate free definitions reduce in local trivializations to the coordinate expressions from Chapter 9 in Part II.

12.1 $1 \times 1 GM$ -convolutions

$1 \times 1 GM$ -convolutions map input feature fields $f_{\text{in}} \in \Gamma(\mathcal{A}_{\text{in}})$ to output feature fields $f_{\text{out}} \in \Gamma(\mathcal{A}_{\text{out}})$ by linearly mapping each individual input feature vector $f_{\text{in}}(p) \in \mathcal{A}_{\text{in},p} \cong \mathbb{R}^{c_{\text{in}}}$ to an output feature vector $f_{\text{out}}(p) \in \mathcal{A}_{\text{out},p} \cong \mathbb{R}^{c_{\text{out}}}$ at the same location $p \in M$. The convolutional character is implemented by *sharing* the linear map from $\mathcal{A}_{\text{in},p}$ to $\mathcal{A}_{\text{out},p}$ between different spatial locations. However, while the feature spaces $\mathcal{A}_{\text{in},p}$ and $\mathcal{A}_{\text{in},q}$ as well as $\mathcal{A}_{\text{out},p}$ and $\mathcal{A}_{\text{out},q}$ are for different $p, q \in M$ isomorphic to each other, there is no canonical isomorphism between them given if the considered structure group G is non-trivial. It is therefore not obvious how the linear map could be shared between different locations. As already suggested in the introduction of this chapter, this issue is resolved by considering G -equivariant kernels which are indifferent to the specific choice of isomorphism or gauge. The arbitrariness of the trivialization which is chosen from the G -atlas reflects the GM -coordinate independence of $1 \times 1 GM$ -convolutions.

Mathematically, $1 \times 1 GM$ -convolutions can be formulated either as specific vector bundle M -morphisms or via the corresponding sections of (associated) homomorphism bundles $\text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}})$. Since we require both concepts later on, we will introduce both viewpoints in the following Sections 12.1.1 and 12.1.2.

12.1.1 $1 \times 1 GM$ -convolutions as vector bundle M -morphisms

$1 \times 1 GM$ -convolutions can be formalized in terms of specific smooth *vector bundle M -morphisms* which share weights over spatial positions. Ignoring the requirement for shared weights for now, such a vector bundle M -morphism \mathcal{C} is a smooth bundle map satisfying the following commutative diagram:

$$\begin{array}{ccc} \mathcal{A}_{\text{in}} & \xrightarrow{\mathcal{C}} & \mathcal{A}_{\text{out}} \\ \pi_{\mathcal{A}_{\text{in}}} \searrow & & \swarrow \pi_{\mathcal{A}_{\text{out}}} \\ & M & \end{array} \quad (12.3)$$

The commutativity $\pi_{\mathcal{A}_{\text{in}}} = \pi_{\mathcal{A}_{\text{out}}} \circ \mathcal{C}$ ensures that each fiber $\mathcal{A}_{\text{in},p}$ is mapped to the fiber $\mathcal{A}_{\text{out},p}$ over the same point $p \in M$ (which gives rise to the “ M ” in the term M -morphism). As a vector bundle morphism, the restriction $\mathcal{C}|_p : \mathcal{A}_{\text{in},p} \rightarrow \mathcal{A}_{\text{out},p}$ to a single fiber is further defined to be linear. Relative to a local trivialization $\Psi_{\mathcal{A}_{\text{in}}}^A$ of \mathcal{A}_{in} and $\Psi_{\mathcal{A}_{\text{out}}}^A$ of \mathcal{A}_{out} , the bundle map is therefore at each point $p \in U^A$ represented by a matrix

$$\mathcal{C}^A|_p := \psi_{\mathcal{A}_{\text{out}},p}^A \circ \mathcal{C}|_p \circ (\psi_{\mathcal{A}_{\text{in}},p}^A)^{-1} \in \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}. \quad (12.4)$$

Its relationship to a second coordinatization \mathcal{C}^B is at $p \in U^A \cap U^B$ given by

$$\mathcal{C}^B|_p = \rho_{\text{out}}(g_p^{BA}) \mathcal{C}^A|_p \rho_{\text{in}}(g_p^{BA})^{-1}, \quad (12.5)$$

which is evident from the commutative diagram below:

$$\begin{array}{ccccc}
& & \mathcal{C}^A|_p & & \\
\mathbb{R}^{c_{\text{in}}} & \xrightarrow{\quad} & & \xrightarrow{\quad} & \mathbb{R}^{c_{\text{out}}} \\
\downarrow \rho_{\text{in}}(g_p^{BA}) & \swarrow \psi_{\mathcal{A}_{\text{in}}, p}^A & \mathcal{A}_{\text{in}, p} & \xrightarrow{\mathcal{C}|_p} & \mathcal{A}_{\text{out}, p} & \searrow \psi_{\mathcal{A}_{\text{out}}, p}^A \\
& \swarrow \psi_{\mathcal{A}_{\text{in}}, p}^B & & & \searrow \psi_{\mathcal{A}_{\text{out}}, p}^B & \\
& \mathbb{R}^{c_{\text{in}}} & \xrightarrow{\quad} \mathcal{C}^B|_p & \xrightarrow{\quad} & \mathbb{R}^{c_{\text{out}}} & \downarrow \rho_{\text{out}}(g_p^{BA})
\end{array} \tag{12.6}$$

The bundle map \mathcal{C} acts on input feature fields $f_{\text{in}} \in \Gamma(\mathcal{A}_{\text{in}})$ to produce output feature fields

$$f_{\text{out}} = \mathcal{C} \circ f_{\text{in}} \in \Gamma(\mathcal{A}_{\text{out}}). \tag{12.7}$$

In terms of a commutative diagram, this mapping is visualized as:

$$\begin{array}{ccc}
\mathcal{A}_{\text{in}} & \xrightarrow{\mathcal{C}} & \mathcal{A}_{\text{out}} \\
f_{\text{in}} \swarrow & & \nearrow f_{\text{out}} \\
M
\end{array}, \tag{12.8}$$

In order for a vector bundle M -morphism $\mathcal{C}_{K_{1 \times 1}}$ to represent a 1×1 GM-convolution, it needs to be parameterized in terms of a 1×1 GM-convolution kernel template $K_{1 \times 1} \in \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ which is shared with coordinatizations at all spatial positions. As argued before in Section 9.1.1, no particular gauge must thereby be preferred in order to ensure *GM-coordinate independence*. It is therefore necessary to *share the weights with all trivializations $X \in \mathfrak{X}$ of the G-atlas \mathcal{A}^G simultaneously*, that is, to require:

$$\mathcal{C}_{K_{1 \times 1}}|_p = K_{1 \times 1} \quad \text{for any gauge } X \in \mathfrak{X} \text{ with } p \in U^X. \tag{12.9}$$

From the transformation behavior between different coordinatizations in Eq. (12.5) it follows that the kernel template has to satisfy the linear constraint

$$\rho_{\text{out}}(g) K_{1 \times 1} \rho_{\text{in}}(g)^{-1} = K_{1 \times 1} \quad \forall g \in G, \tag{12.10}$$

that is, it has to be an intertwiner (an equivariant linear map, Def. B.5.7). The vector space

$$\text{Hom}_G(\rho_{\text{in}}, \rho_{\text{out}}) := \left\{ K_{1 \times 1} \in \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \mid K_{1 \times 1} \rho_{\text{in}}(g) = \rho_{\text{out}}(g) K_{1 \times 1} \quad \forall g \in G \right\} \tag{12.11}$$

of intertwining maps characterizes the space of GM-coordinate independent 1×1 -convolution kernels fully. As already mentioned in Section 9.1.1, *Schur's Lemma* B.5.10 implies that the requirement on $K_{1 \times 1}$ to be an intertwiner prevents a mapping between fields which transform under non-isomorphic irreducible representations via 1×1 GM-convolutions. The more general GM-convolutions with spatially extended kernels, defined in Section 12.2, will resolve this issue.

With these preparations we are ready to give a concise definition of 1×1 GM-convolutions:

Definition 12.1.1 (1×1 GM-convolution). A 1×1 GM-convolution is a map

$$K_{1 \times 1} \circledast : \Gamma(\mathcal{A}_{\text{in}}) \rightarrow \Gamma(\mathcal{A}_{\text{out}}), \quad f_{\text{in}} \mapsto K_{1 \times 1} \circledast f_{\text{in}} := \mathcal{C}_{K_{1 \times 1}} \circ f_{\text{in}} \quad (12.12)$$

which is parameterized by an intertwining 1×1 GM-convolution kernel $K_{1 \times 1} \in \text{Hom}_G(\rho_{\text{in}}, \rho_{\text{out}})$. Here $\mathcal{C}_{K_{1 \times 1}}$ is the unique smooth vector bundle M -morphism between \mathcal{A}_{in} and \mathcal{A}_{out} which is in arbitrary gauges $\psi_{\mathcal{A}_{\text{in}}, p}$ and $\psi_{\mathcal{A}_{\text{out}}, p}$ from the considered G -atlas pointwise defined by

$$\mathcal{C}_{K_{1 \times 1}}|_p := \psi_{\mathcal{A}_{\text{out}}, p}^{-1} \circ K_{1 \times 1} \circ \psi_{\mathcal{A}_{\text{in}}, p}. \quad (12.13)$$

The independence of the chosen gauges (GM-coordinate independence) is guaranteed by $K_{1 \times 1}$ being an intertwiner.

To show the independence of the chosen gauge explicitly, consider any G -related trivializations $\rho_{\text{in}}(g) \psi_{\mathcal{A}_{\text{in}}, p}$ and $\rho_{\text{out}}(g) \psi_{\mathcal{A}_{\text{out}}, p}$ for an arbitrary structure group element $g \in G$, which leave the construction of

$$\begin{aligned} \mathcal{C}_{K_{1 \times 1}}|_p &= (\rho_{\text{out}}(g) \psi_{\mathcal{A}_{\text{out}}, p})^{-1} \circ K_{1 \times 1} \circ (\rho_{\text{in}}(g) \psi_{\mathcal{A}_{\text{in}}, p}) \\ &= \psi_{\mathcal{A}_{\text{out}}, p}^{-1} \circ (\rho_{\text{out}}(g)^{-1} K_{1 \times 1} \rho_{\text{in}}(g)) \circ \psi_{\mathcal{A}_{\text{in}}, p} \\ &= \psi_{\mathcal{A}_{\text{out}}, p}^{-1} \circ K_{1 \times 1} \circ \psi_{\mathcal{A}_{\text{in}}, p} \end{aligned} \quad (12.14)$$

invariant. That such defined 1×1 GM-convolutions are indeed mapping to sections in $\Gamma(\mathcal{A}_{\text{out}})$ follows from $\mathcal{C}_{K_{1 \times 1}}$ being a bundle map. An overview of local coordinatizations of 1×1 GM-convolutions is given in Fig. 12.1.

12.1.2 1×1 GM-convolutions as homomorphism bundle sections

While the vector bundle M -morphism with gauge independent coordinatizations from Def. 12.1.1 and Fig. 12.1 fully specifies a 1×1 GM-convolution, we will now adopt an alternative viewpoint which describes 1×1 GM-convolutions in terms of the *homomorphism bundle* $\text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}}) \xrightarrow{\pi_{\text{Hom}}} M$. To this end, recall that the vector bundle morphism \mathcal{C} in Eq. (12.3) restricts to linear maps $\mathcal{C}|_p : \mathcal{A}_{\text{in}, p} \rightarrow \mathcal{A}_{\text{out}, p}$ over each $p \in M$. The set of such linear maps (or vector space homomorphisms) between $\mathcal{A}_{\text{in}, p}$ and $\mathcal{A}_{\text{out}, p}$ is denoted as $\text{Hom}(\mathcal{A}_{\text{in}, p}, \mathcal{A}_{\text{out}, p})$. Since it is closed under linear combinations, it forms itself a vector space. It can be shown that the disjoint union

$$\text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}}) := \coprod_{p \in M} \text{Hom}(\mathcal{A}_{\text{in}, p}, \mathcal{A}_{\text{out}, p}) \quad (12.15)$$

of these homomorphism spaces forms a vector bundle, the homomorphism bundle between \mathcal{A}_{in} and \mathcal{A}_{out} , when being equipped with the projection map $\pi_{\text{Hom}} : \text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}}) \rightarrow M$ which sends elements in $\text{Hom}(\mathcal{A}_{\text{in}, p}, \mathcal{A}_{\text{out}, p})$ to p and a smooth structure induced from that of \mathcal{A}_{in} and \mathcal{A}_{out} [74]. The fibers over p satisfy $\text{Hom}(\mathcal{A}_{\text{in}, p}, \mathcal{A}_{\text{out}, p}) \cong \text{Hom}(\mathbb{R}^{c_{\text{in}}}, \mathbb{R}^{c_{\text{out}}}) \cong \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ such that we can take the typical fiber to be the vector space of real-valued $c_{\text{out}} \times c_{\text{in}}$ matrices.

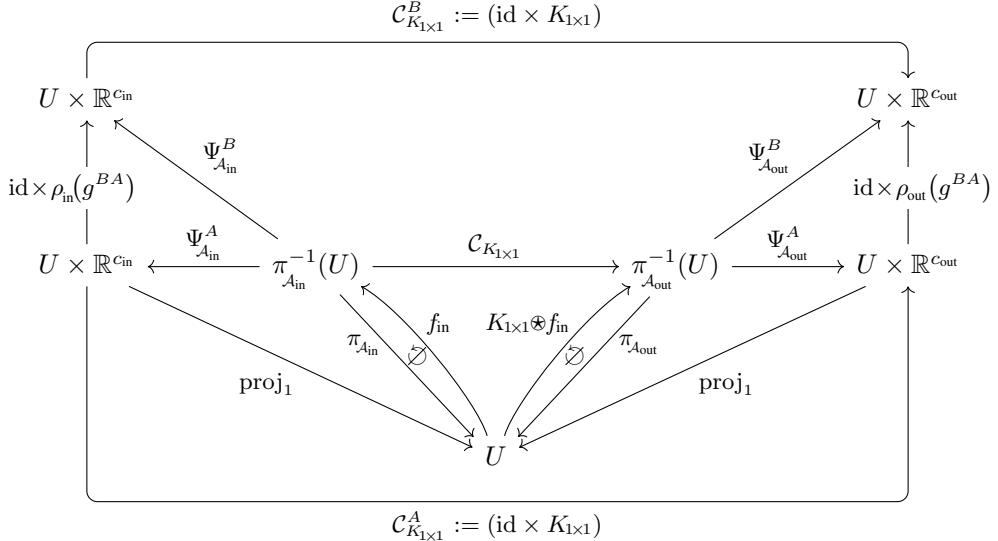


Figure 12.1: Coordinatization of an 1×1 GM-convolution $K_{1x1} \otimes : \Gamma(\mathcal{A}_{in}) \rightarrow \Gamma(\mathcal{A}_{out})$ and its corresponding vector bundle M -morphism $\mathcal{C}_{K_{1x1}}$. The convolutional character is encoded into the morphism by sharing a kernel matrix $K_{1x1} \in \mathbb{R}^{c_{out} \times c_{in}}$ over different spatial positions $p \in M$. Since no gauge is to be preferred, the kernel is furthermore shared over different trivializations $\mathcal{C}_{K_{1x1}}^A$ and $\mathcal{C}_{K_{1x1}}^B$. The commutativity of the diagram for any choices $\Psi_{\mathcal{A}_{in}}^A$, $\Psi_{\mathcal{A}_{out}}^A$ and $\Psi_{\mathcal{A}_{in}}^B$, $\Psi_{\mathcal{A}_{out}}^B$ therefore enforces the constraint $\rho_{out}(g)K_{1x1}\rho_{in}(g)^{-1} = K_{1x1} \quad \forall g \in G$ which restricts the kernel matrix to be an intertwiner (an equivariant linear map), that is, $K_{1x1} \in \text{Hom}_G(\rho_{in}, \rho_{out}) \subseteq \mathbb{R}^{c_{out} \times c_{in}}$. Except for $f_{in} \circ \pi_{\mathcal{A}_{in}} \neq \text{id}_{\mathcal{A}_{in}}$ and $[K_{1x1} \otimes f_{in}] \circ \pi_{\mathcal{A}_{out}} \neq \text{id}_{\mathcal{A}_{out}}$, the diagram commutes.

The trivializations

$$\Psi_{\text{Hom}} : \pi_{\text{Hom}}^{-1}(U) \rightarrow U \times \mathbb{R}^{c_{out} \times c_{in}}, \quad H \mapsto (p, \psi_{\text{Hom},p}(H)), \quad (12.16)$$

where we abbreviated $p = \pi_{\text{Hom}}(H)$, are *induced* from the trivializations of \mathcal{A}_{in} and \mathcal{A}_{out} by defining

$$\psi_{\text{Hom},p} : \text{Hom}(\mathcal{A}_{in,p}, \mathcal{A}_{out,p}) \rightarrow \mathbb{R}^{c_{out} \times c_{in}}, \quad H \mapsto \psi_{\mathcal{A}_{out,p}} \circ H \circ (\psi_{\mathcal{A}_{in,p}})^{-1} \quad (12.17)$$

in analogy to Eqs. (12.4) and (7.18). This implies transition maps

$$\begin{aligned} H^B &= \psi_{\mathcal{A}_{out,p}}^B \circ H \circ (\psi_{\mathcal{A}_{in,p}}^B)^{-1} \\ &= \psi_{\mathcal{A}_{out,p}}^B \circ (\psi_{\mathcal{A}_{out,p}}^A)^{-1} H^A \psi_{\mathcal{A}_{in,p}}^A \circ (\psi_{\mathcal{A}_{in,p}}^A)^{-1} \\ &= \rho_{out}(g^{BA}) H^A \rho_{in}(g^{BA})^{-1} \\ &=: \rho_{\text{Hom}}(g^{BA}) H^A \end{aligned} \quad (12.18)$$

between gauges Ψ_{Hom}^A and Ψ_{Hom}^B on $U^A \cap U^B$, where we introduced the homomorphism group representation $\rho_{\text{Hom}} : G \rightarrow \text{GL}(\mathbb{R}^{c_{out} \times c_{in}})$ as left and right multiplication with ρ_{out} and ρ_{in} for notational convenience.¹ The homomorphism bundle $\text{Hom}(\mathcal{A}_{in}, \mathcal{A}_{out})$ is by construction

¹In general, a homomorphism bundle between two *non-associated* vector bundles with structure groups G_1 and G_2 would have a structure group $G_1 \times G_2$. Since \mathcal{A}_{in} and \mathcal{A}_{out} are associated, they

associated to TM , GM , \mathcal{A}_{in} and \mathcal{A}_{out} , that is, its trivializations transform synchronously with those of the other bundles. As a G -associated vector bundle, it can be identified with $(GM \times \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}) / \sim_{\rho_{\text{Hom}}}$. Fig. 12.2a gives an overview of the local trivializations of $\text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}})$. Note the similarity to the trivializations of the other associated G -bundles in Fig. 11.5.

From the viewpoint of homomorphism bundles, unconstrained bundle maps as in Eq. (12.3) correspond to the action of unconstrained smooth homomorphism bundle sections

$$\sigma_{\text{Hom}} : M \mapsto \text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}}) \quad \text{such that} \quad \pi_{\text{Hom}} \circ \sigma_{\text{Hom}} = \text{id}_M , \quad (12.19)$$

which can be interpreted as 1×1 *kernel fields* that do not share weights. Their global existence is guaranteed by $\text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}})$ being a vector bundle. Sections corresponding to 1×1 *GM-convolutions* require in addition that the linear transformations $\sigma_{\text{Hom}}(p) \in \text{Hom}(\mathcal{A}_{\text{in},p}, \mathcal{A}_{\text{out},p})$ are determined by a template kernel $K_{1 \times 1} \in \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ which is shared over different positions $p \in M$ and any choice of gauge. They can therefore for any $p \in M$ be defined as

$$\sigma_{K_{1 \times 1}}(p) := \psi_{\text{Hom},p}^{-1}(K_{1 \times 1}), \quad K_{1 \times 1} \in \text{Hom}_G(\rho_{\text{in}}, \rho_{\text{out}}) , \quad (12.20)$$

where the chosen trivialization Ψ_{Hom} is arbitrary if (and only if) $K_{1 \times 1}$ satisfies the intertwiner constraint

$$\rho_{\text{Hom}}(g)K_{1 \times 1} = K_{1 \times 1} \quad \forall g \in G , \quad (12.21)$$

which is equivalent to Eq. (12.10).² The gauge irrelevance of such sections is visualized in the commutative diagram in Fig. 12.2b (compare this to the equivalent bundle map trivialization in Fig. 12.1).

Summarizing remarks: A smooth 1×1 *GM-convolution layer* $K_{1 \times 1} \otimes : \Gamma(\mathcal{A}_{\text{in}}) \rightarrow \Gamma(\mathcal{A}_{\text{out}})$, $f_{\text{in}} \mapsto f_{\text{out}}$ can equivalently be defined via a smooth bundle map as $f_{\text{out}}(p) := C_{K_{1 \times 1}} \circ f_{\text{in}}(p)$ or via a smooth homomorphism bundle section as $f_{\text{out}}(p) := \sigma_{K_{1 \times 1}}(p) \circ f_{\text{in}}(p)$. By definition, both trivialize in an arbitrarily chosen gauge Ψ_{Hom}^A to $f_{\text{out}}^A(p) = K_{1 \times 1}f_{\text{in}}^A(p)$. The *GM*-coordinate independence of this definition is guaranteed by the intertwining property of the kernel in Eq. (12.10) or, equivalently, Eq. (12.21). This can be seen by considering a different trivialization via Ψ_{Hom}^B :

$$\begin{aligned} K_{1 \times 1}f_{\text{in}}^B(p) &= K_{1 \times 1}(\rho_{\text{in}}(g_p^{BA})f_{\text{in}}^A(p)) \\ &= \rho_{\text{out}}(g_p^{BA})K_{1 \times 1}f_{\text{in}}^A(p) \\ &= \rho_{\text{out}}(g_p^{BA})f_{\text{out}}^A(p) \\ &= f_{\text{out}}^B(p) \end{aligned} \quad (12.22)$$

transform synchronously under the same structure group $G_1 = G_2 = G$ such that their transition maps take values in the diagonal subgroup G of $G \times G$.

²The required smoothness of the section follows from the smoothness of the local trivializations.

$$\begin{array}{ccc}
& U \times \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} & \\
\Psi_{\text{Hom}}^B \nearrow & & \uparrow (\text{id} \times \rho_{\text{Hom}}(g^{BA}) \cdot) \\
\pi_{\text{Hom}}^{-1}(U) & \xrightarrow{\Psi_{\text{Hom}}^A} & U \times \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \\
\pi_{\text{Hom}} \downarrow & & \swarrow \text{proj}_1 \\
U & &
\end{array}$$

(a) Trivialization of $\text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}})$. Being associated to TM , GM , \mathcal{A}_{in} and \mathcal{A}_{out} , the transition maps of the homomorphism bundle are determined by the same group element g^{BA} of the shared structure group G (compare this to Fig. 11.5). The homomorphism representation ρ_{Hom} is defined in Eq. (12.18). Unconstrained vector bundle M -morphisms as shown in Eq. (12.3) correspond to unconstrained smooth sections of $\text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}})$.

$$\begin{array}{ccc}
& (\text{id} \times \rho_{\text{Hom}}(g^{BA}) \cdot) & \\
& \curvearrowright & \\
\pi_{\text{Hom}}^{-1}(U) & \xrightarrow[\Psi_{\text{Hom}}^A]{\Psi_{\text{Hom}}^B} & U \times \underbrace{\text{Hom}_G(\rho_{\text{in}}, \rho_{\text{out}})}_{\subseteq \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}} \\
\sigma_{K_{1 \times 1}} \nearrow \otimes \quad \pi_{\text{Hom}} \downarrow & & \swarrow \text{proj}_1 \\
U & &
\end{array}$$

(b) The sections $\sigma_{K_{1 \times 1}} : M \rightarrow \text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}})$ of the homomorphism bundle which correspond to 1×1 GM-convolutions are exactly those which trivialize to the same (intertwining) matrix $K_{1 \times 1} \in \text{Hom}_G(\rho_{\text{in}}, \rho_{\text{out}}) \subseteq \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ in all gauges. Such sections correspond to bundle maps which trivialize as specified in Fig. 12.1.

Figure 12.2: Local trivializations of the homomorphism bundle $\text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}})$, which is the vector bundle of linear maps between the spaces $\mathcal{A}_{\text{in},p}$ and $\mathcal{A}_{\text{out},p}$ for any $p \in M$. As usual we abbreviate $U = U^A \cap U^B$. Except for $\sigma_{K_{1 \times 1}} \circ \pi_{\text{Hom}} \neq \text{id}_{\text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}})}$ the diagrams commute.

12.2 Kernel field transforms and *GM*-convolutions

We now turn to kernel field transforms and *GM*-convolutions with spatially extended kernels. Section 12.2.1 introduces general, unconstrained kernel fields and more specific *GM*-convolutional kernel fields, the latter defined in terms of a shared, *G*-steerable template kernel. General kernel field transforms and *GM*-convolutions are introduced in Section 12.2.2. As both are defined *globally*, their formulation is necessarily *coordinate free*. Section 12.2.3 expresses both operations relative to local trivializations, recovering our local definitions from Section 9.2.

12.2.1 Coordinate free kernels fields and *G*-steerable kernels

To detect spatial patterns in feature fields, convolutional networks apply spatially extended kernels which linearly accumulate features from a local neighborhood around each point. In Eq. (9.24) we defined (unconstrained) template kernels for a d -dimensional manifold and c_{in} - and c_{out} -dimensional input and output feature fields as maps $K : \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ which assign a $c_{\text{out}} \times c_{\text{in}}$ matrix to each point of their domain. The definition of convolution kernels as maps with domain $\mathbb{R}^d \cong T_p M$ and codomain $\mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \cong \text{Hom}(\mathcal{A}_{\text{in},p}, \mathcal{A}_{\text{out},p})$ suggests a coordinate free definition of kernels as maps between the tangent spaces and the corresponding homomorphism spaces:

Definition 12.2.1 (Kernel field). We define (unconstrained) kernel fields of type $\rho_{\text{in}}, \rho_{\text{out}}$ on a manifold M as smooth bundle M -morphisms between the tangent bundle TM and the feature vector homomorphism bundle $\text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}})$. By its definition as an M -morphism, a kernel field \mathcal{K} lets the following diagram commute:

$$\begin{array}{ccc} TM & \xrightarrow{\mathcal{K}} & \text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}}) \\ \pi_{TM} \searrow & & \swarrow \pi_{\text{Hom}} \\ & M & \end{array} \quad (12.23)$$

Despite smoothly mapping between two vector bundles, \mathcal{K} is not assumed to be a vector bundle morphism, that is, the restrictions $\mathcal{K}_p : T_p M \rightarrow \text{Hom}(\mathcal{A}_{\text{in},p}, \mathcal{A}_{\text{out},p})$ are not assumed to be linear.³

The name *kernel field* is motivated by the fact that such defined bundle maps \mathcal{K} assign a (potentially different) coordinate free kernel $\mathcal{K}_p : T_p M \rightarrow \text{Hom}(\mathcal{A}_{\text{in},p}, \mathcal{A}_{\text{out},p})$ to each point p of the manifold.⁴ In practice, kernels \mathcal{K}_p are often designed to detect local patterns around p and are therefore assumed to be compactly supported around the origin of $T_p M$.

A coordinate free kernel \mathcal{K}_p at p is relative to gauges $\psi_{TM,p}^A$ and $\psi_{\text{Hom},p}^A$ of the *G*-atlases given by the map

$$\mathcal{K}_p^A : \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}, \quad \mathcal{K}_p^A := \psi_{\text{Hom},p}^A \circ \mathcal{K}_p \circ (\psi_{TM,p}^A)^{-1}. \quad (12.24)$$

³This reflects that convolution kernels are in general not linear as maps $K : \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$. Note that this does not interfere with the linearity of $K(v) \in \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ (as map $\mathbb{R}^{c_{\text{in}}} \rightarrow \mathbb{R}^{c_{\text{out}}}$) for any $v \in \mathbb{R}^d$ or, here, the linearity of $\mathcal{K}_p(v) \in \text{Hom}(\mathcal{A}_{\text{in},p}, \mathcal{A}_{\text{out},p})$ (as map $\mathcal{A}_{\text{in},p} \rightarrow \mathcal{A}_{\text{out},p}$) for any $v \in T_p M$.

⁴We expect that it is possible to work out a well defined notion of *kernel bundles* whose sections are in one-to-one correspondence to our definition of kernel fields as bundle maps (this reformulation would mirror the transition from Eq. (12.3) to Eq. (12.19)).

Fig. 9.2 visualizes a coordinate free kernel on $T_p M$ and its coordinatizations on \mathbb{R}^d relative to different gauges. From the commutative diagram

$$\begin{array}{ccccc}
 & & \mathcal{K}_p^A & & \\
 & \swarrow \psi_{TM,p}^A & & \nearrow \psi_{Hom,p}^A & \\
 \mathbb{R}^d & \xrightarrow{\quad T_p M \quad} & Hom(\mathcal{A}_{in}|_p, \mathcal{A}_{out}|_p) & & \mathbb{R}^{c_{out} \times c_{in}} \\
 g_p^{BA} \downarrow & & & & \rho_{Hom}(g_p^{BA}) \\
 & \searrow \psi_{TM,p}^B & & \swarrow \psi_{Hom,p}^B & \\
 & & \mathcal{K}_p^B & & \mathbb{R}^{c_{out} \times c_{in}}
 \end{array} \quad (12.25)$$

it follows that different kernel coordinatizations are related by

$$\mathcal{K}_p^B = \rho_{Hom}(g_p^{BA}) \circ \mathcal{K}_p^A \circ (g_p^{BA})^{-1}. \quad (12.26)$$

Note that this relation only implies GM -coordinate independence but does not constrain the coordinate free kernel in any way. As before, the situation changes when sharing weights over spatial positions.

In order for a kernel field \mathcal{K}_K to correspond to a convolution, it needs to be fully specified by a single template kernel $K : \mathbb{R}^d \rightarrow \mathbb{R}^{c_{out} \times c_{in}}$ which is shared over all spatial positions. We are again forced to share weights with all gauges $X \in \mathfrak{X}$ simultaneously in order to preserve their equivalence and thus GM -coordinate independence. As argued in Section 9.2.3, the appropriate way of sharing K with kernel coordinatizations $\mathcal{K}_{K,p}^X$ involves a normalization by the reference frame volume $\sqrt{|\eta_p^X|}$ and is defined by

$$\mathcal{K}_{K,p}^X = \frac{K}{\sqrt{|\eta_p^X|}} \quad \text{for any gauge } X \in \mathfrak{X} \text{ with } p \in U^X. \quad (12.27)$$

The reason for the frame normalization factor is that convolutions will later be defined in terms of integrals over the tangent spaces. We are therefore actually required to share the integral operator itself in different coordinatizations, which is equivalent to identifying the matrix-valued integration measures $\mathcal{K}_{K,p}^X(v) \sqrt{|\eta_p^X|} dv$ for any gauge $X \in \mathfrak{X}$ at $p \in M$ with a template measure $K(v) dv$. The form of the kernel sharing in Eq. (12.27) follows by equating both expressions.

Together with the relation $\sqrt{|\eta_p^A|} = |\det(g_p^{BA})| \sqrt{|\eta_p^B|}$ between different frame volumes, the kernel transformation law in Eq. (12.26) and the weight sharing in Eq. (12.27) imply the *G-steerability kernel constraint*

$$\frac{1}{|\det g|} \rho_{Hom}(g) \circ K \circ g^{-1} = K \quad \forall g \in G. \quad (12.28)$$

Valid template kernels are thus given by the invariants under the simultaneous gauge action of $|\det g|^{-1}$, $\rho_{Hom}(g)$ and g^{-1} . Writing out the representation ρ_{Hom} , acting on $\mathbb{R}^{c_{out} \times c_{in}}$ via multiplication with ρ_{out} and ρ_{in}^{-1} from the left and right, respectively, the constraint in Eq. (12.28) is seen to be equivalent to that in Eq. (9.37), i.e. $K(gv) = |\det g|^{-1} \rho_{out}(g) K(v) \rho_{in}(g)^{-1} \forall g \in G, v \in \mathbb{R}^d$.

We cast these insights into definitions:

Definition 12.2.2 (G -steerable kernel). *G -steerable kernels are characterized by their invariance under the gauge action. The vector space of smooth G -steerable kernels that map between field types ρ_{in} and ρ_{out} is defined by*

$$\mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^G := \left\{ K: \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \text{ smooth } \mid \frac{1}{|\det g|} \rho_{\text{Hom}}(g) \circ K \circ g^{-1} = K \quad \forall g \in G \right\} \quad (12.29)$$

$$= \left\{ K: \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \text{ smooth } \mid \frac{1}{|\det g|} \rho_{\text{out}}(g) K(g^{-1} \mathbf{v}) \rho_{\text{in}}(g)^{-1} = K(\mathbf{v}) \quad \forall g \in G, \mathbf{v} \in \mathbb{R}^d \right\}, \quad (12.30)$$

where $\rho_{\text{Hom}}(g)H := \rho_{\text{out}}(g)H\rho_{\text{in}}(g)^{-1}$ for any $H \in \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ and $G \leq \text{GL}(d)$. The gauge invariance of G -steerable kernels allows for GM-coordinate independent weight sharing.

G -steerable kernels were in [53] introduced to equivariant deep learning, where finite groups were assumed. The current formulation in Def. 12.2.2 was proposed in [323]. A complete solution for the G -steerable kernel spaces for arbitrary representations ρ_{in} and ρ_{out} of structure groups $G \leq \text{O}(2)$ has been derived in [322], an implementation is publicly available at [38]. Mathematically, steerable kernel are equivalent to *representation operators* like for instance the spherical tensor operators from quantum mechanics. A generalization of the *Wigner-Eckart theorem* describes G -steerable kernels as being composed from harmonic basis functions, Clebsch-Gordan coefficients and endomorphisms of irreducible representations [173]. An implementation of this algorithm was described in [40] and is available at [39]. More details on steerable kernels are found in Chapter 5.

Definition 12.2.3 (GM-convolutional kernel field). *A GM-convolutional kernel field \mathcal{K}_K of type $\rho_{\text{in}}, \rho_{\text{out}}$ is a kernel field which is determined by a shared, G -steerable template kernel $K \in \mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^G$. It is in arbitrary gauges $\psi_{TM,p}^X$ and $\psi_{Hom,p}^X$ from the considered G -atlas pointwise defined by:*

$$\mathcal{K}_{K,p} := (\psi_{Hom,p}^X)^{-1} \circ \frac{K}{\sqrt{|\eta_p^X|}} \circ \psi_{TM,p}^X \quad (12.31)$$

The smoothness of \mathcal{K}_K follows from the smoothness of the gauges, the metric and the template kernel.

As in the case of 1×1 GM-convolutions, the arbitrariness of the particular choice of gauge in Eq. (12.31) – and therefore the GM-coordinate independence of the definition – is guaranteed by the G -steerability of $K \in \mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^G$. To show this explicitly, one may define the kernel field relative to some gauge B and then apply a transformation to any other gauge A ,

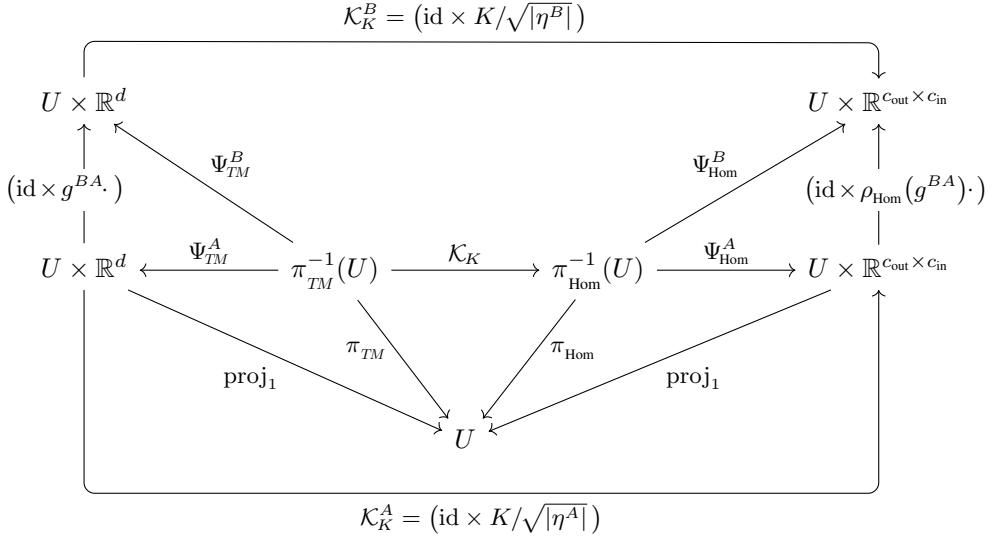


Figure 12.3: Commutative diagram showing local coordinatizations of a *GM-convolutional kernel field* \mathcal{K}_K as defined in Def. 12.2.3. Convolutional weight sharing requires the coordinate expression of the kernel field \mathcal{K}_K at any point $p \in M$ and any gauge X at p to be determined by the shared template kernel $K : \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ as $\mathcal{K}_{K,p}^X = K / \sqrt{|\eta_p^X|}$. The commutativity of the diagram then implies the G -steerability constraint $|\det g|^{-1} \rho_{\text{Hom}}(g) \circ K \circ g^{-1} = K \quad \forall g \in G$ on the space $\mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^G$ of template kernels. We want to emphasize that, despite looking similar to the diagram in Fig. 12.1, the diagram in the current figure should be seen as analog to that in Fig. 12.2b. The difference between the current diagram and that in Fig. 12.2b is that the linear maps in the homomorphism bundle are via $\mathcal{K}_K : TM \rightarrow \text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}})$ determined by an element of the tangent bundle TM instead of the section $\sigma_{K_{1 \times 1}} : M \rightarrow \text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}})$.

which cancels out and therefore leads to an equivalent expression:

$$\begin{aligned}
 \mathcal{K}_{K,p} &= (\psi_{\text{Hom},p}^B)^{-1} \circ \frac{K}{\sqrt{|\eta^B|}} \circ \psi_{TM,p}^B \\
 &= (\rho_{\text{Hom}}(g_p^{BA}) \psi_{\text{Hom},p}^A)^{-1} \circ \frac{K}{\sqrt{|\eta^A|} / |\det(g_p^{BA})|} \circ (g_p^{BA} \cdot \psi_{TM,p}^A) \\
 &= (\psi_{\text{Hom},p}^A)^{-1} \circ \frac{|\det(g_p^{BA})| \rho_{\text{Hom}}(g_p^{BA})^{-1} \circ K \circ g_p^{BA}}{\sqrt{|\eta^A|}} \circ \psi_{TM,p}^A \\
 &= (\psi_{\text{Hom},p}^A)^{-1} \circ \frac{K}{\sqrt{|\eta^A|}} \circ \psi_{TM,p}^A
 \end{aligned} \tag{12.32}$$

Fig. 12.3 gives an overview of the local trivializations of *GM*-convolutional kernel fields in terms of a commutative diagram.

Note that the G -steerability constraint in Eq. (12.30) or (12.29) reduces to the constraint on 1×1 *GM*-convolution kernels in Eq. (12.10) or (12.21) when being evaluated at the origin $v = 0$ of \mathbb{R}^d , which is invariant under the action of any $g \in G$. The results on 1×1 *GM*-convolutions, derived in the previous section, are therefore seen to be a special

case for the choice of point-like kernels.⁵ We further want to mention that the constraint on spatially extended kernels does in general not require their codomain to be restricted to $\text{Hom}_G(\rho_{\text{in}}, \rho_{\text{out}})$, i.e. the space of intertwiners. In contrast to 1×1 GM -convolutions, this allows GM -convolutions with spatially extended kernels to map between fields that transform according to non-isomorphic irreducible representations.

12.2.2 Kernel field transforms and GM -convolutions

Having defined both feature fields and kernel fields, we are ready to introduce kernel field transforms and GM -convolutions. They are pointwise defined in terms of integral operators which compute output feature vectors $f_{\text{out}}(p)$ at points $p \in M$ by matching the kernel \mathcal{K}_p at p with the feature field f_{in} “as seen from p ”.

The local representation of an input field “as seen from p ” is formally given by its *transporter pullback*, which is visualized in Fig. 9.1. It is defined as the usual pullback from M to TM via the Riemannian exponential map⁶ with the additional application of a parallel transporter (Eq. (11.84)), which is necessary in order to express the pulled back features in $\mathcal{A}_{\text{in}, \exp(v)}$ as features in $\mathcal{A}_{\text{in}, p}$. Denoting this parallel transporter along the geodesic path $\gamma_v(t) := \exp((1-t)v)$ between $\gamma(0) = \exp(v)$ and $\gamma(1) = \pi(v) =: p$ by

$$\mathcal{P}_{\mathcal{A}, p \leftarrow \exp(v)} : \mathcal{A}_{\exp(v)} \rightarrow \mathcal{A}_p , \quad (12.33)$$

we thus define the pulled back feature field representations on the tangent spaces as follows:

Definition 12.2.4 (Transporter pullback of feature field to TM). Given a feature field $f \in \Gamma(\mathcal{A})$, we define its (redundant) representation on the tangent bundle as

$$\text{Exp}^* f : TM \rightarrow \mathcal{A}, \quad v \mapsto \mathcal{P}_{\mathcal{A}, \pi_{TM}(v) \leftarrow \exp(v)} \circ f \circ \exp(v) . \quad (12.34)$$

The Riemannian exponential map \exp corresponds hereby to the Levi-Civita connection, while the transporter $\mathcal{P}_{\mathcal{A}, \pi_{TM}(v) \leftarrow \exp(v)}$ relies on some G -compatible connection; see Sections 8.2 and 11.5.

From the construction it is clear that $\text{Exp}^* f(v) \in \mathcal{A}_p$ for any $v \in T_p M$, that is, $\text{Exp}^* f$ is a bundle M -morphism, satisfying the following commutative diagram:

$$\begin{array}{ccc} TM & \xrightarrow{\text{Exp}^* f} & \mathcal{A} \\ \pi_{TM} \searrow & & \swarrow \pi_{\mathcal{A}} \\ & M & \end{array} \quad (12.35)$$

Despite smoothly mapping between two vector bundles, $\text{Exp}^* f$ is not assumed to be a vector bundle morphism, that is, the restrictions $\text{Exp}_p^* f : T_p M \rightarrow \mathcal{A}_p$ are usually not linear.

⁵To make this statement precise, one would have to generalize Def. 12.2.2 to operator-valued distributions and define 1×1 GM -convolution kernels as operator-valued Dirac deltas.

⁶We define the exponential map on the full tangent bundle as $\exp : TM \rightarrow M$, $v \mapsto \exp_{\pi_{TM}(v)}(v)$. Recall that we assumed the manifold to be geodesically complete, such that the exponential map is well defined on the whole tangent bundle (and resort to zero-padding if this assumption fails to hold).

The restriction $\text{Exp}_p^* f := \text{Exp}^* f|_{T_p M}$ of the transporter pullback's domain to $T_p M$ captures the feature field from the perspective of an observer at p as shown in Fig. 9.1. Note that this definition resembles a local representation of the feature field in terms of *geodesic normal coordinates*, with the difference that it is not restricted to the injectivity radius of the exponential map.⁷ We furthermore want to mention that the transporter may be replaced with any other isomorphism between $\mathcal{A}_{\text{exp}(v)}$ and \mathcal{A}_p , as done for instance in [280].

As stated before, kernel field transforms and *GM*-convolutions are defined as matching the local feature field representations on the tangent spaces with kernels. Working towards these definitions, note that the bundle M -morphisms of kernels $\mathcal{K} : TM \rightarrow \text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}})$ and local field representations $\text{Exp}^* f_{\text{in}} : TM \rightarrow \mathcal{A}_{\text{in}}$, can be combined to yet another (nonlinear) M -morphism from TM to \mathcal{A}_{out} ,

$$\begin{array}{ccccc} TM & \xrightarrow{\mathcal{K} \times \text{Exp}^* f_{\text{in}}} & \text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}}) \times \mathcal{A}_{\text{in}} & \xrightarrow{\text{ev}} & \mathcal{A}_{\text{out}} \\ & \searrow \pi_{TM} & & \swarrow \pi_{\mathcal{A}_{\text{out}}} & \\ & & M & & \end{array}, \quad (12.36)$$

where $\text{ev} : (\mathcal{K}(v), \text{Exp}^* f_{\text{in}}(v)) \mapsto \mathcal{K}(v) \text{Exp}^* f_{\text{in}}(v)$ is the evaluation map on $\text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}}) \times \mathcal{A}_{\text{in}}$. Kernel field transforms compute output feature vectors at p by integrating this product of kernels and input fields over the respective tangent space $T_p M$:

Definition 12.2.5 (Kernel field transform). Let \mathcal{K} be any smooth kernel field (Def. 12.2.1). The corresponding kernel field transform is a smooth integral transform

$$\mathcal{T}_{\mathcal{K}} : \Gamma(\mathcal{A}_{\text{in}}) \rightarrow \Gamma(\mathcal{A}_{\text{out}}) \quad (12.37)$$

which is pointwise defined by⁸

$$\begin{aligned} [\mathcal{T}_{\mathcal{K}}(f_{\text{in}})](p) &:= \int_{T_p M} \mathcal{K}(v) \text{Exp}^* f_{\text{in}}(v) dv \\ &= \int_{T_p M} \mathcal{K}(v) \mathcal{P}_{\mathcal{A}_{\text{in}}, p \leftarrow \text{exp}_p v} f_{\text{in}}(\text{exp}_p v) dv. \end{aligned} \quad (12.38)$$

In order to be well defined, the integral needs to exist and the resulting output field $\mathcal{T}_{\mathcal{K}}(f)$ needs to be smooth. This requires \mathcal{K} to be chosen suitably, e.g. by assuming it to decay rapidly or to be compactly supported.

Note that general kernel field transforms do not necessarily model convolutions as they do not assume synapse weights (kernels) to be shared between spatial positions. Such general kernel field transforms will become handy in the following Chapter 13, where we derive a requirement for spatial weight sharing from the requirement for isometry equivariance.

⁷Any feature vector $f(q)$ might therefore be represented multiple times on the same tangent space $T_p M$, once for each $v \in T_p M$ with $\text{exp}(v) = q$. If this is not desired, one may restrict the kernel support to the injectivity radius of the exponential map, such that only the geodesically nearest occurrence will be measured.

⁸The integration over $T_p M$ via the Riemannian volume density dv is discussed in Appendix D.

Appendix I discusses the existence and smoothness of kernel field transforms. A sufficient condition for kernel field transforms to be well defined is the restriction of kernel supports to balls of a fixed radius $R > 0$:

Theorem 12.2.6 (Kernel field transform existence for compactly supported kernels).

Let \mathcal{K} be a kernel field whose individual kernels \mathcal{K}_p at any $p \in M$ are (at most) supported on a closed ball of radius $R > 0$ around the origin of $T_p M$, that is,

$$\text{supp}(\mathcal{K}_p) \subseteq \{v \in T_p M \mid \|v\| \leq R\} \quad \forall p \in M. \quad (12.39)$$

The corresponding kernel field transform $\mathcal{T}_{\mathcal{K}}$ is then guaranteed to be well defined, i.e. the integral in Eq. (12.38) exists and the output field $\mathcal{T}_{\mathcal{K}}(f) \in \Gamma(\mathcal{A}_{\text{out}})$ is smooth for any smooth input field $f \in \Gamma(\mathcal{A}_{\text{in}})$.

Proof: See Appendices I and I.

□

The requirement to restrict the kernel support to a closed ball of certain radius is common practice in deep learning. Note, however, that a compactly supported kernel is at odds with scale equivariant convolutions, which, by the corresponding G -steerability kernel constraints, require infinitely far extending kernels. Current implementations of scale equivariant convolutions usually approximate scale equivariant kernel spaces by restricting their support [197, 334, 107, 359, 10, 281, 220] and are therefore covered by Theorem 12.2.6.

Based on general kernel field transforms, we define *coordinate free GM-convolutions* by adding the assumption of spatial weight sharing, i.e. by assuming *GM-convolutional kernel fields*:

Definition 12.2.7 (GM-convolution). Let \mathcal{A}_{in} and \mathcal{A}_{out} be G -associated feature vector bundles with types ρ_{in} and ρ_{out} , respectively. We define the GM-convolution with a G -steerable kernel $K \in \mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^G$ as the kernel field transform with the corresponding GM-convolutional kernel field \mathcal{K}_K (Def. 12.2.3):

$$K \star : \Gamma(\mathcal{A}_{\text{in}}) \rightarrow \Gamma(\mathcal{A}_{\text{out}}), \\ f_{\text{in}} \mapsto K \star f_{\text{in}} := \mathcal{T}_{\mathcal{K}_K}(f_{\text{in}}) = \int_{T_p M} \mathcal{K}_K(v) \text{Exp}^* f_{\text{in}}(v) dv \quad (12.40)$$

As GM-convolutions do not prefer any reference frame in the G -structure, they are guaranteed to generalize their inference over all poses of patterns which are related by the action of the structure group G ; see Eq. (9.42) and Fig. 9.2.

12.2.3 Kernel field transforms and GM-convolutions in local coordinates

What is left to show is that the coordinate free definitions of transporter pullbacks, kernel field transforms and GM-convolutions introduced in this section reduce to the coordinate expressions from Part II when being expressed relative to some local trivialization.

The local coordinate expression of the transporter pullback $\text{Exp}^* f$ of a feature field f is, as usual, defined by pre- and post-composing it with local trivializations of the corresponding bundles, that is:

$$\begin{aligned} [\text{Exp}^* f]^A : U \times \mathbb{R}^d &\rightarrow U \times \mathbb{R}^c, \\ (p, v) &\mapsto \Psi_{\mathcal{A}}^A \circ \text{Exp}^* f \circ (\Psi_{TM}^A)^{-1}(p, v) \\ &= (p, \psi_{\mathcal{A}, p}^A \circ \text{Exp}_p^* f \circ (\psi_{TM, p}^A)^{-1}(v)) \end{aligned} \quad (12.41)$$

Local gauge transformations at $p \in M$ are from this definition seen to be given by

$$[\text{Exp}_p^* f]^B = \rho(g_p^{BA}) \circ [\text{Exp}_p^* f]^A \circ (g_p^{BA})^{-1}. \quad (12.42)$$

We visualize these coordinate expressions in terms of a commutative diagram, which is very similar to that for the local trivializations of kernel fields in Fig. 12.3:

$$\begin{array}{ccccc} & & [\text{Exp}^* f]^B & & \\ & \swarrow & & \searrow & \\ U \times \mathbb{R}^d & & & & U \times \mathbb{R}^c \\ \uparrow & \nearrow & & & \uparrow \\ (\text{id} \times g^{BA}) & & \Psi_{TM}^B & & (\text{id} \times \rho(g^{BA})) \\ \downarrow & & \downarrow & & \downarrow \\ U \times \mathbb{R}^d & \xleftarrow{\Psi_{TM}^A} & \pi_{TM}^{-1}(U) & \xrightarrow{\text{Exp}^* f} & \pi_{\mathcal{A}}^{-1}(U) \xrightarrow{\Psi_{\mathcal{A}}^A} U \times \mathbb{R}^c \\ \uparrow & & \uparrow & & \uparrow \\ \text{proj}_1 & & \pi_{TM} & & \pi_{\mathcal{A}} \\ \downarrow & & \downarrow & & \downarrow \\ U & & & & U \end{array} \quad (12.43)$$

For an implementation it is useful to further resolve the coordinate expression of the transporter pullback into those of its individual components, i.e. of the transporter $\mathcal{P}_{\mathcal{A}, p \leftarrow \exp(v)}$, the feature field f and the exponential map \exp . This is achieved by expanding it with an identity of the form $\text{id}_{\mathcal{A}_{\exp(v)}} = (\psi_{\mathcal{A}, \exp(v)}^{\tilde{A}})^{-1} \circ \psi_{\mathcal{A}, \exp(v)}^{\tilde{A}}$, where the choice of gauge \tilde{A} at $\exp(v)$ is irrelevant as it ultimately drops out:

$$\begin{aligned} [\text{Exp}_p^* f]^A(v) &= [\psi_{\mathcal{A}, p}^A \circ \text{Exp}_p^* f \circ (\psi_{TM, p}^A)^{-1}](v) \\ &= \psi_{\mathcal{A}, p}^A \circ \mathcal{P}_{\mathcal{A}, p \leftarrow \exp(v)} \circ f(\exp \circ (\psi_{TM, p}^A)^{-1}(v)) \\ &= \psi_{\mathcal{A}, p}^A \circ \mathcal{P}_{\mathcal{A}, p \leftarrow \exp(v)} \circ (\psi_{\mathcal{A}, \exp(v)}^{\tilde{A}_v})^{-1} \circ \psi_{\mathcal{A}, \exp(v)}^{\tilde{A}_v} \circ f(\exp \circ (\psi_{TM, p}^A)^{-1}(v)) \\ &= \rho(g_{p \leftarrow \exp \circ (\psi_{TM, p}^A)^{-1}(v)}^{A \tilde{A}}) \cdot f^{\tilde{A}}(\exp \circ (\psi_{TM, p}^A)^{-1}(v)) \end{aligned} \quad (12.44)$$

As expected, we recover our definition from Eq. (9.21) in Section 9.2.1, which approves that Def. 12.2.4 is indeed its coordinate free counterpart.

The coordinate expression of a kernel field transform, which coincides with Eq. (9.30) in Section 9.2.1, is given by the following theorem:

Theorem 12.2.8 (Kernel field transform in coordinates). *Relative to some gauge A at $p \in U^A$, the kernel field transform is given by the coordinate expression*

$$\begin{aligned} [\mathcal{J}_K(f_{\text{in}})]^A(p) &= \int_{\mathbb{R}^d} \mathcal{K}_p^A(v^A) [\text{Exp}_p^* f_{\text{in}}]^A(v^A) \sqrt{|\eta_p^A|} dv^A \\ &= \int_{\mathbb{R}^d} \mathcal{K}_p^A(v^A) \rho(g_{p \leftarrow \exp \circ (\psi_{TM,p}^A)^{-1}(v^A)}^{A\tilde{A}}) \cdot f_{\text{in}}^{\tilde{A}}(\exp \circ (\psi_{TM,p}^A)^{-1}(v^A)) \sqrt{|\eta_p^A|} dv^A, \end{aligned} \quad (12.45)$$

where the gauges \tilde{A} at $\exp(v)$ are chosen arbitrarily as they cancel out.⁹

Proof: The first expression is derived by a simple calculation which translates all involved quantities into their corresponding coordinate expressions:

$$\begin{aligned} &[\mathcal{J}_K(f_{\text{in}})]^A(p) \\ &\stackrel{(1)}{=} \psi_{A_{\text{out}},p}^A [\mathcal{J}_K(f_{\text{in}})](p) \\ &\stackrel{(2)}{=} \psi_{A_{\text{out}},p}^A \int_{T_p M} \mathcal{K}_p(v) [\text{Exp}_p^* f_{\text{in}}](v) dv \\ &\stackrel{(3)}{=} \psi_{A_{\text{out}},p}^A \int_{\mathbb{R}^d} \mathcal{K}_p((\psi_{TM,p}^A)^{-1}(v^A)) [\text{Exp}_p^* f_{\text{in}}]((\psi_{TM,p}^A)^{-1}(v^A)) \sqrt{|\eta_p^A|} dv^A \\ &\stackrel{(4)}{=} \int_{\mathbb{R}^d} [\psi_{A_{\text{out}},p}^A \circ \mathcal{K}_p((\psi_{TM,p}^A)^{-1}(v^A)) \circ (\psi_{A_{\text{in}},p}^A)^{-1}] \\ &\quad [\psi_{A_{\text{in}},p}^A \circ [\text{Exp}_p^* f_{\text{in}}] \circ (\psi_{TM,p}^A)^{-1}](v^A) \sqrt{|\eta_p^A|} dv^A \\ &\stackrel{(5)}{=} \int_{\mathbb{R}^d} [\psi_{\text{Hom},p}^A \circ \mathcal{K}_p \circ (\psi_{TM,p}^A)^{-1}](v^A) [\psi_{A_{\text{in}},p}^A \circ [\text{Exp}_p^* f_{\text{in}}] \circ (\psi_{TM,p}^A)^{-1}](v^A) \sqrt{|\eta_p^A|} dv^A \\ &\stackrel{(6)}{=} \int_{\mathbb{R}^d} \mathcal{K}_p^A(v^A) [\text{Exp}_p^* f_{\text{in}}]^A(v^A) \sqrt{|\eta_p^A|} dv^A \end{aligned} \quad (12.46)$$

Step (1) expresses the output feature vector at p explicitly in terms of gauge $\psi_{A_{\text{out}},p}^A$, acting on the coordinate free kernel field transform. This coordinate free expression is in step (2) expanded as defined in Def. 12.2.5. Step (3) pulls the integral over $T_p M$ via the chosen gauge back to \mathbb{R}^d , which is in more detail described in Appendix D. Step (4) inserts an identity map of the form $\text{id} = (\psi_{A_{\text{in}},p}^A)^{-1} \circ \psi_{A_{\text{in}},p}^A$ and pulls $\psi_{A_{\text{out}},p}^A$ into the integral while step (5) identifies the definition of $\psi_{\text{Hom},p}^A$ from Eq. (12.17). Lastly, we identify the coordinate expressions of \mathcal{K}_p and $\text{Exp}_p^* f_{\text{in}}$ from Eqs. (12.24) and (12.41).

The second expression follows from the first one by expanding the coordinate expression of the transporter pullback according to Eq. (12.44). \square

⁹Note that the gauges at $\exp(v)$ might differ for different $v \in T_p M$ and should more correctly be labeled by \tilde{A}_v . We suppress this dependency for brevity.

The coordinate expression for the coordinate free GM -convolutions follows immediately:

Theorem 12.2.9 (GM -convolutions in coordinates). *A coordinate free GM -convolution $K \star : \Gamma(\mathcal{A}_{\text{in}}) \rightarrow \Gamma(\mathcal{A}_{\text{out}})$ with a G -steerable kernel $K \in \mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^G$ is relative to some gauge A at $p \in U^A$ given by*

$$[K \star f]^A(p) = [\mathcal{J}_{\mathcal{K}_K}(f)]^A(p) = \int_{\mathbb{R}^d} K(v^A) [\text{Exp}_p^* f]^A(v^A) dv^A, \quad (12.47)$$

that is, by the coordinate expression that was introduced in Eq. (9.39). This expression may be written out further as done for general kernel field transforms in Eq. (12.45).

Proof: The result follows from Theorem 12.2.8 by observing that the coordinate free GM -convolution $K \star$ is just a kernel field transform with the corresponding GM -convolutional kernel field \mathcal{K}_K ; see Def. 12.2.7. Specifically, the coordinate expression of a GM -convolutional kernel field \mathcal{K}_K is according to Def. 12.2.3 given by the frame volume normalized G -steerable kernel K , that is, $\mathcal{K}_{K,p}^A = K / \sqrt{|\eta_p^A|}$. Inserting this identity in Eq. (12.45) leads to the claimed coordinate expression for GM -convolutions. \square

This result assures that a global, coordinate free GM -convolution can be implemented in terms of its local coordinate expressions relative to some G -atlas of local trivializations that cover M .

Isometry equivariance

A main characteristic of the convolution operation and its various generalizations is their equivariance w.r.t. symmetries of the underlying manifold. For instance, the conventional convolution on Euclidean spaces is translation equivariant while spherical convolutions are rotation equivariant. More generally, any locally compact group and their homogeneous spaces admit group convolutions [114, 165, 48, 102], which were recently picked up by the deep learning community to generalize convolutional networks to such spaces [52, 162, 56, 10]. However, as these approaches rely fundamentally on the global, *transitive* symmetries of the homogeneous space, they do not immediately apply to general Riemannian manifolds.

GM-convolutions on the other hand shift the focus from *global symmetries of the space itself* to *local symmetries in the coordinatization of the space*. As it turns out, the local gauge equivariance of *GM*-convolutions, together with convolutional weight sharing, induces their equivariance under the action of global symmetries. Stated more precisely, *GM*-convolutions are equivariant under the action of *G-structure preserving isometries* (Def. 13.1.1), which form a subgroup $\text{Isom}_{GM} \leq \text{Isom}(M)$ of the full isometry group. The requirement on the symmetry to be an isometry (i.e. to preserve the metric) comes hereby from the use of exponential maps, which rely on the Levi-Civita connection and thus Riemannian metric. The additional requirement on these isometries to preserve the *G*-structure is a consequence of the definition of feature vector bundles as associated *G*-bundles, whose elements have a well defined meaning only relative to those reference frames that are contained in *GM*. Note that the latter is not really a restriction, as one may always choose structure groups $G \geq O(d)$, for which *any* isometry respects the corresponding *G*-structure. On the contrary, this design allows for a precise control of the level of isometry equivariance. For instance, the conventional convolution on Euclidean vector spaces relies on the canonical $\{e\}$ -structure of \mathbb{R}^d , visualized in Fig. 13.3a, and is therefore solely translation equivariant. An $SO(d)$ -structure on \mathbb{R}^d , visualized in Fig. 13.4a, is additionally preserved by rotations, and thus corresponds to $SE(d)$ -equivariant convolutions. Equivariance under the full isometry group $E(d)$ of \mathbb{R}^d is implied when choosing an $O(d)$ -structure on \mathbb{R}^d .

The goal of this chapter is to derive theorems which formally characterize the isometry equivariance of *GM*-convolutions and kernel field transforms. Section 13.1 lays the foundations of this investigation by introducing isometry groups of Riemannian manifolds and discussing a range of well-known relations and constructions which they induce. Specifically, Section 13.1.1 introduces isometries and isometry groups while Section 13.1.2 defines their induced action (“pushforwards”) on the associated bundles in a coordinate free setting. In Section 13.1.3 we express these actions on bundles relative to local trivializations and

discuss their passive interpretation as isometry induced gauge transformations, visualized in Fig. 8.5. Section 13.1.4 briefly states how the quantities involved in kernel field transforms behave under the action of isometries.

Based on these properties, we study the isometry equivariance of kernel field transforms and *GM*-convolutions in Section 13.2. After defining the term “isometry equivariance” formally, Section 13.2.1 proves a central result, which asserts that the demand for *isometry equivariance requires the invariance of the kernel field under isometries*; see Fig. 13.6. Section 13.2.2 considers the more specific *GM*-convolutions and proves that they are by design equivariant under any isometry that preserves the *G*-structure. This result implies in particular, that *OM*-convolutions are equivariant w.r.t. any isometry.

The invariance constraint on kernel fields enforces that they share weights over the orbits of the isometry group. This suggests that invariant kernel fields can equivalently be described by representative kernels on orbit representatives, which we formalize in Section 13.3. Section 13.3.1 discussed isometry induced quotient spaces and their representatives. In Section 13.3.2 we use these mathematical definitions to prove that the space of isometry invariant kernel fields is indeed isomorphic to kernel fields on quotient representatives. This implies in particular that isometry equivariant kernel field transforms on homogeneous spaces are necessarily convolutions, which closes the loop to prior work.

13.1 Isometries and their action on manifolds, bundles and fields

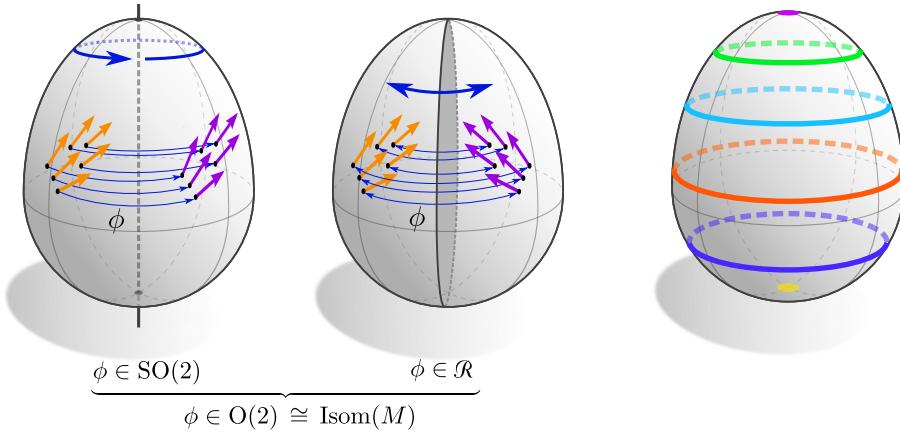
This section introduces most of the mathematical concepts required for our study of the isometry equivariance of kernel field transforms and *GM*-convolutions. After defining isometries in Section 13.1.1, we discuss in Section 13.1.2 how they induce natural actions on tangent vectors and reference frames. For structure groups $G < O(d)$, not any isometry is compatible with any *G*-structure. We define the subgroup $\text{Isom}_{GM} \leq \text{Isom}(M)$ of those isometries which do act on (induce automorphisms of) a *G*-structure *GM* and their *G*-associated feature bundles. While these constructions are kept coordinate free, Section 13.1.3 expresses the action of isometries on fiber bundles relative to local bundle trivializations. In preparation for our investigation of isometry equivariant kernel field transforms later on, Section 13.1.4 discusses how isometries commute with exponential maps and parallel transporters, which allows to derive how they act on transporter pullbacks $\text{Exp}_p^* f$ of feature fields f . While mostly staying mathematical, we draw connections to the application wherever possible.

13.1.1 Isometry groups

A (global) *isometry* $\phi : M \rightarrow \hat{M}$ is a diffeomorphism between Riemannian manifolds (M, η) and $(\hat{M}, \hat{\eta})$, which *preserves the metric*. In terms of the pushforward (differential) $\phi_{*,TM} : TM \rightarrow T\hat{M}$ of tangent vectors, which we introduce in Appendix C.2 and in Section 13.1.2 below, this statement is made precise by requiring that isometries satisfy

$$\eta_p(v, w) = \hat{\eta}_{\phi(p)}(\phi_{*,TM}v, \phi_{*,TM}w) \quad \forall p \in M, v, w \in T_p M, \quad (13.1)$$

i.e. that they preserve distances and angles between tangent vectors. Intuitively, an isometry is thought of as a distance preserving map between manifolds. Note that the inverse of an isometry is necessarily an isometry as well. Since isometries (and their inverses) respect the metric, they constitute the *isomorphisms in the category of Riemannian manifolds*.



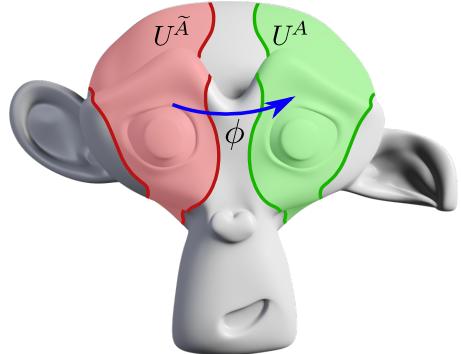
(a) Action of different subgroups of the isometry group on fields. (b) Orbits of the isometry group.

Figure 13.1: Visualizations of the isometry group $\text{Isom}(M) \cong \text{O}(2)$ of an egg M , which we will use throughout this chapter to exemplify different concepts and constructions relating to isometries. Fig. 13.1a shows the action of the isometry group on (tangent or feature) vector fields. It can be thought of as consisting of the subgroups of rotations in $\text{SO}(2)$ and reflections in \mathcal{R} . The action of the isometry group partitions the egg into orbits $\text{Isom}(M).p = \{\phi(p) \mid \phi \in \text{Isom}(M)\}$ of points $p \in M$, shown in Fig. 13.1b in different colors. Note that not all orbits are homeomorphic to each other – the orbits at the poles are single points while any other orbit traces out a circle around the egg. The isometry group of the egg acts non-transitively on it, that is, not every point can be reached from any other point. A kernel field transform is isometry equivariant if it commutes with the isometry action on feature fields. We show that isometry equivariance is guaranteed if and only if the kernel field is invariant under the action of isometries. This implies in particular that isometry equivariance require weight sharing along the isometry orbits; see Fig. 13.6.

The set of all isometries $\phi : M \rightarrow M$ from a Riemannian manifold to itself, equipped with the usual function composition $\circ : (\phi_1, \phi_2) \mapsto \phi_1 \circ \phi_2$, defines a group, known as *isometry group* $\text{Isom}(M)$ of M . This group is the automorphism group of a Riemannian manifold, which contains all of its (metric) “symmetries”. It is a subgroup of the diffeomorphism group $\text{Diff}(M)$ of M . The full isometry group might have non-trivial subgroups, which we will in the following denote by $\mathcal{I} \leq \text{Isom}(M)$. An example is given in Fig. 13.1a, which visualizes the isometry group $\text{Isom}(M) \cong \text{O}(2)$ of an egg. The full isometry group splits (for instance) into the subgroups of rotations in $\mathcal{I}_1 \cong \text{SO}(2)$ and reflections in $\mathcal{I}_2 \cong \mathcal{R}$.

In general, the isometry group of a manifold is non-transitive (Def. B.3.8), that is, not every point of M can be reached from any other point by its action. The manifold is then partitioned into disjoint *orbits* (Def. B.3.3), visualized for the example of M being an (Easter) egg in Fig. 13.1b. The isometry group of a manifold M might be trivial, given that M is sufficiently asymmetric. In this case there might still exist non-trivial isometries between open subsets U^A and U^B of M , restricted to which Eq. (13.1) holds. Fig. 13.2 shows an example of a manifold which is globally asymmetric but has non-trivial isometries between local subsets of itself. We will in the following only consider global isometries of M , however, all concepts of the current Section 13.1 generalize in an obvious way to isometries between local subsets. Without proof, we claim that the same holds for the isometry equivariance of any neural network operation which acts pointwise, for instance 1×1 -convolutions, nonlinearities or bias summation. The equivariance of kernel field transforms with spatially extended kernels holds up to boundary effects.

Figure 13.2: An asymmetric manifold, whose global isometry group is trivial. Since the asymmetry is limited to the ears and the mouth of “Suzanne”, the monkey, there are non-trivial localized symmetries left. For instance, the smooth map $\phi : U^{\tilde{A}} \rightarrow U^A$ between the red and green highlighted subsets preserves the metric locally. All concepts developed in Section 13.1 as well as the isometry equivariance of point-wise operations like 1×1 -convolutions generalize immediately to such isometries between local subsets. The isometry equivariance of kernel field transforms with spatially extended kernels generalizes up to boundary effects.



13.1.2 Isometry action on fiber bundles

Isometries act naturally on tangent vectors in TM and reference frames in FM by ‘‘carrying them along’’ with the group action as visualized in Fig. 13.1a. If an isometry is in addition compatible with the G -structure, that is, if it gives rise to an automorphism of GM , it furthermore acts on any associated G -bundle, in particular the feature vector bundles \mathcal{A} . We discuss these actions of isometries on the associated bundles and on feature fields in the following.

Isometry action on the tangent bundle TM : Any isometry $\phi \in \text{Isom}(M)$ gives rise to a *pushforward*

$$\phi_{*,TM} : TM \rightarrow TM, \quad \phi \in \text{Isom}(M) \quad (13.2)$$

on the tangent bundle, which is just the differential of ϕ as introduced in Appendix C.2. It can at each point $p \in M$ be thought of as a *linear* approximation of ϕ , which maps vectors $v \in T_p M$ to $\phi_{*,TM}(v) \in T_{\phi(p)} M$, that is, it satisfies

$$\pi_{TM} \circ \phi_{*,TM} = \phi \circ \pi_{TM}. \quad (13.3)$$

As argued in Appendix C.2, the pushforward is invertible with $(\phi_{*,TM})^{-1} = (\phi^{-1})_{*,TM}$, for which we will unambiguously write $\phi_{*,TM}^{-1}$.¹ The pushforward of an element ϕ of the isometry group is therefore seen to be an (isometric) vector bundle automorphism of TM over ϕ , satisfying the following commutative diagram:

$$\begin{array}{ccc} TM & \xrightleftharpoons[\phi_{*,TM}^{-1}]{\phi_{*,TM}} & TM \\ \pi_{TM} \downarrow & & \downarrow \pi_{TM} \\ M & \xrightleftharpoons[\phi^{-1}]{\phi} & M \end{array} \quad (13.4)$$

By the *definition of isometries*, their pushforward preserves distances and angles, that is,

$$\eta_{\phi(p)}(\phi_{*,TM} v, \phi_{*,TM} w) = \eta_p(v, w) \quad \forall p \in M, v, w \in T_p M, \phi \in \text{Isom}(M). \quad (13.5)$$

¹The invertibility does not hold for pushforwards in general but only for those of diffeomorphisms and thus isometries, which are themselves invertible.

More details about pushforwards between tangent bundles are easily found in the literature, for instance in [262].

Isometry action on the frame bundle FM : The pushforward on TM immediately induces a corresponding principal bundle automorphism $\phi_{*,FM}$ on FM by pushing forward the individual frame vectors:

$$\phi_{*,FM} : FM \rightarrow FM, \quad [e_i]_{i=1}^d \mapsto \phi_{*,FM}([e_i]_{i=1}^d) := [\phi_{*,TM}(e_i)]_{i=1}^d, \quad \phi \in \text{Isom}(M) \quad (13.6)$$

It maps frames in $F_p M$ for arbitrary $p \in M$ to frames at $F_{\phi(p)} M$, that is, it satisfies $\pi_{FM} \circ \phi_{*,FM} = \phi \circ \pi_{FM}$. To see this, let $[e_p]_{i=1}^d \in F_p M$, then $\phi \circ \pi_{FM}([e_i]_{i=1}^d) = \phi(p)$ and $\pi_{FM} \circ \phi_{*,FM}([e_i]_{i=1}^d) = \pi_{FM}([\phi_{*,TM}(e_i)]_{i=1}^d) = \pi_{TM} \circ \phi_{*,TM}(e_j) = \phi \circ \pi_{TM}(e_j) = \phi(p)$ for any $j = 1, \dots, d$. It can further be checked to be invertible with $(\phi_{*,FM})^{-1} = (\phi^{-1})_{*,FM}$, again abbreviated by $\phi_{*,FM}^{-1}$. The *left action* of the $\phi_{*,FM}$ on the frame bundle commutes with the *right action* \triangleleft on its fibers, that is, for arbitrary $g \in \text{GL}(d)$ and $\phi \in \text{Isom}(M)$ we have that:

$$\begin{aligned} \left(\phi_{*,FM}([e_i]_{i=1}^d) \right) \triangleleft g &= [\phi_{*,TM}(e_i)]_{i=1}^d \triangleleft g && \text{(def. of } \phi_{*,FM}, \text{ Eq. (13.6)}) \\ &= \left[\sum_j \phi_{*,TM}(e_j) g_{ji} \right]_{i=1}^d && \text{(def. of } \triangleleft, \text{ Eq. (11.26))} \\ &= \left[\phi_{*,TM} \left(\sum_j e_j g_{ji} \right) \right]_{i=1}^d && \text{(linearity of } \phi_{*,TM} \text{)} \\ &= \phi_{*,FM} \left(\left[\sum_j e_j g_{ji} \right] \right)_{i=1}^d && \text{(def. of } \phi_{*,FM}, \text{ Eq. (13.6))} \\ &= \phi_{*,FM}([e_i]_{i=1}^d \triangleleft g) && \text{(def. of } \triangleleft, \text{ Eq. (11.26))} \quad (13.7) \end{aligned}$$

A gauge transformation of a frame at $p \in M$ by $g \in \text{GL}(d)$, followed by a pushforward to $\phi(p)$, is therefore equal to a pushforward of the untransformed frame, followed by a gauge transformation by the same group element g but at $\phi(p)$. Different frames in the fiber $F_p M$ are hence mapped in such a way to frames at $F_{\phi(p)} M$ that their relative offset is preserved. The derived properties of $\phi_{*,FM}$ are summarized by the statement that the diagram

$$\begin{array}{ccc} FM & \xrightarrow{\phi_{*,FM}} & FM \\ \uparrow \triangleleft g & & \uparrow \triangleleft g \\ FM & \xrightarrow{\phi_{*,FM}} & FM \\ \downarrow \pi_{FM} & & \downarrow \pi_{FM} \\ M & \xrightarrow{\phi} & M \end{array} \quad (13.8)$$

commutes for any $\phi \in \text{Isom}(M)$ and any $g \in \text{GL}(d)$. Satisfying the commutativity of this diagram, the pushforward $\phi_{*,FM}$ on the frame bundle is identified as a *principal bundle automorphism*² over ϕ .

²I.e. a principal bundle isomorphism from the frame bundle to itself; cf. Eq. (11.23).

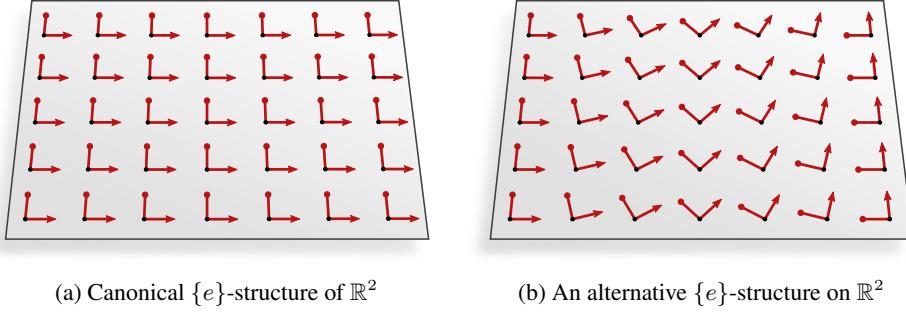


Figure 13.3: Two specific choices of $\{e\}$ -structures (global frame fields) $\{e\}M$ on $M = \mathbb{R}^2$, which we use to visualize the concept of *G-structure preserving isometries*. The full isometry group of M is the Euclidean group $\text{Isom}(M) = E(2)$, consisting of translations, rotations and reflections. Fig. 13.3a shows the canonical $\{e\}$ -structure of \mathbb{R}^2 , which is invariant under translations but not under rotations or reflections. More abstractly stated, translations make up the subgroup $\text{Isom}_{\{e\}}M = (\mathbb{R}^2, +)$ of isometries that induce automorphisms of $\{e\}M$. In contrast, rotations or reflections map frames in $\{e\}_p M$ to frames in $F_{\phi(p)}M$ but fail to send them to $\{e\}_{\phi(p)}M$. They do therefore not induce automorphisms of the $\{e\}$ -structure and are not part of $\text{Isom}_{\{e\}}M$. Group actions of such isometries on $\{e\}M$ or any of its $\{e\}$ -associated bundles *are not defined*. Fig. 13.3b shows an alternative choice of $\{e\}$ -structure on $M = \mathbb{R}^2$ (or $M = \mathbb{E}_2$), which is only invariant under translations in the “up-down” direction, i.e. $\text{Isom}_{\{e\}}M \cong (\mathbb{R}, +)$. The examples in Figs. 13.3a and 13.3b exemplify that the *G*-structure automorphisms do not only depend on the structure group G but on the particular choice of *G*-structure GM . The general case for G being non-trivial is harder to visualize since $G_p M$ will then not be a single frame but a set of frames; see e.g. Fig. 13.4.

Note that the inverses, which are shown explicitly in the diagram (13.4), are omitted to reduce clutter.

Isometry action on *G*-structures GM : As *G*-structures are principal subbundles of the frame bundle, one can consider the restriction of the domain of the pushforward on FM to GM , that is,

$$\phi_{*,FM}|_{GM} : GM \rightarrow FM, \quad \phi \in \text{Isom}(M). \quad (13.9)$$

It is hereby necessary to keep the full frame bundle FM as codomain since there is in general no guarantee that frames in $G_p M$ are mapped to frames of $G_{\phi(p)}M$ but only to $F_{\phi(p)}M$. Since *G*-structures are in general not closed under the action of isometries on FM , it might be *impossible* to define a group action of the full isometry group on GM or any other associated *G*-bundle. To remedy this shortcoming, we will in the following consider the subgroup of those isometries that respect the *G*-structure, i.e. which map preferred frames in GM to frames in GM .

Definition 13.1.1 (*G*-structure preserving isometries). *Given a *G*-structure GM , we define the corresponding subgroup of *G*-structure preserving isometries Isom_{GM} as*

$$\text{Isom}_{GM} := \{\phi \in \text{Isom}(M) \mid \phi_{*,FM}(G_p M) = G_{\phi(p)}M \quad \forall p \in M\} \leq \text{Isom}(M). \quad (13.10)$$

For such isometries, we define the induced action on GM as

$$\phi_{*,GM} := \phi_{*,FM}|_{GM} : GM \rightarrow GM, \quad \phi \in \text{Isom}_{GM}. \quad (13.11)$$

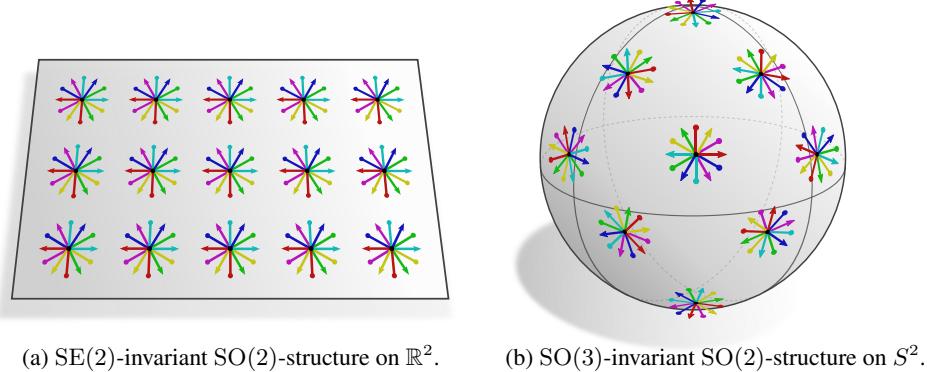


Figure 13.4: Two examples of $\text{SO}(2)$ -structures SOM over the plane $M = \mathbb{R}^2$ and the sphere $M = S^2$. For $M = \mathbb{R}^2$, shown in Fig. 13.4a, the $\text{SO}(2)$ -structure is invariant under translations and rotations. As it consists only of right-handed frames (mind the arrow tips on the first and the circle tips on the second frame axes) it is not invariant under reflections. The isometries which preserve SOM therefore form the group $\text{Isom}_{\text{SOM}} = \text{SE}(2)$, which is a subgroup of the full isometry group $\text{Isom}(M) = \text{E}(2)$. In the case of $M = S^2$, shown in Fig. 13.4b, the $\text{SO}(2)$ -structure is invariant under rotations but not under reflections. The $\text{SO}(2)$ -structure automorphisms are here $\text{Isom}_{\text{SOM}} = \text{SO}(3)$ while the full isometry group is $\text{Isom}(M) = \text{O}(3)$.

Such defined actions for $\phi \in \text{Isom}_{GM}$ are *G-structure automorphisms*, that is, they make the following diagram commute for any $g \in G$ (which follows by restricting Eq (13.8) from FM to GM and $\text{GL}(d)$ to G):

$$\begin{array}{ccc}
 GM & \xrightarrow{\phi_{*,GM}} & GM \\
 \uparrow \triangleleft g & & \uparrow \triangleleft g \\
 GM & \xrightarrow{\phi_{*,GM}} & GM \\
 \downarrow \pi_{GM} & & \downarrow \pi_{GM} \\
 M & \xrightarrow{\phi} & M
 \end{array} \tag{13.12}$$

Fig. 13.3 shows two examples of $\{e\}$ -structures on $M = \mathbb{R}^2$, i.e. global frame fields. From these examples it is apparent that the subgroups Isom_{GM} do really depend on the particular choice of G -structure GM , not only on the structure group G . In Fig. 13.4a we visualize an $\text{SO}(2)$ -structure on $M = \mathbb{R}^2$. Its isometry group $\text{Isom}_{\text{SOM}} = \text{SE}(2)$ is larger than those of the $\{e\}$ -structures in Fig. 13.3. An $\text{SO}(2)$ -structure on the sphere S^2 , which is preserved by all rotations $\text{Isom}_{\text{SOM}} = \text{SO}(3)$, is shown in Fig. 13.4a.

For specific choices of structure groups G it is possible to make more general statements about which isometries are contained in the subgroup Isom_{GM} . Most importantly, for orthonormal structure groups $G = \text{O}(d)$ (which are compatible with η) any isometry will induce an automorphism of OM , that is, one always has $\text{Isom}_{\text{OM}} = \text{Isom}(M)$. To prove this claim, let $[e_i]_{i=1}^d \in \text{O}_p M \subset F_p M$ be an orthonormal frame, which is by an arbitrary isometry $\phi \in \text{Isom}(M)$ being sent to $\phi_{*,FM}|_{\text{OM}} [e_i]_{i=1}^d = [\phi_{*,TM} e_i]_{i=1}^d$; see Eq. (13.6). Ap-

pling Eq. (13.5) to the individual axes of the pushforward frame yields

$$\eta(\phi_{*,TM} e_i, \phi_{*,TM} e_j) = \eta(e_i, e_j) = \delta_{ij} \quad \forall i, j \in 1, \dots, d, \quad (13.13)$$

which implies the orthonormality of the pushforward frame $\phi_{*,FM}|_{OM}[e_i]_{i=1}^d \in O_{\phi(p)}M$ and therefore allows to define $\phi_{*,OM} := \phi_{*,FM}|_{OM}$ for any $\phi \in \text{Isom}(M)$. More generally, this result implies:

$$\text{Isom}_{GM} = \text{Isom}(M) \quad \forall G \geq O(d) \quad (13.14)$$

It is similarly possible to show

$$\text{Isom}_{SOM} = \text{Isom}_+(M), \quad (13.15)$$

that is, that any orientation preserving isometry in $\text{Isom}_+(M)$ induces an automorphism of an $SO(d)$ -structure SOM . Note that these statements all depend only on the structure group G but are independent from the specific choice of G -structure. This is ultimately a result of only considering isometries, which are adapted to $O(d)$ -structures by definition, instead of considering more general diffeomorphisms. As mentioned before, the subgroup Isom_{GM} does in general depend on the specific choice of G -structure GM , not only the structure group G .

Isometry action on associated vector bundles \mathcal{A} : From the pushforward of isometries in Isom_{GM} on GM one can construct a pushforward $\phi_{*,\mathcal{A}}$ on any G -associated vector bundle $\mathcal{A} = (GM \times \mathbb{R}^c)/\sim_\rho$ by defining

$$\phi_{*,\mathcal{A}} : \mathcal{A} \rightarrow \mathcal{A}, \quad [[e_i]_{i=1}^d, f] \mapsto \phi_{*,\mathcal{A}}\left([[e_i]_{i=1}^d, f]\right) := [\phi_{*,GM}([e_i]_{i=1}^d), f], \quad \phi \in \text{Isom}_{GM}. \quad (13.16)$$

This action is well defined since the construction is by the right G -equivariance of $\phi_{*,GM}$ in Eq. (13.12) independent from the chosen representative of the equivalence class. Similar to before, one has $\pi_{\mathcal{A}} \circ \phi_{*,\mathcal{A}} = \phi \circ \pi_{\mathcal{A}}$, that is, $\phi_{*,\mathcal{A}}$ maps feature vectors at \mathcal{A}_p to feature vectors at $\mathcal{A}_{\phi(p)}$, which can be checked by acting on a feature vector and using the corresponding property of $\phi_{*,GM}$. Since $\phi_{*,\mathcal{A}}$ is defined by the action of $\phi_{*,GM}$ on the first factor in $(GM \times \mathbb{R}^c)/\sim_\rho$, it does not interfere with linear combinations, which act on the second factor as defined in Eq. (11.29). This implies that the pushforward on associated bundles maps linearly between their fibers. The invertibility of $\phi_{*,\mathcal{A}}$ follows from the invertibility of $\phi_{*,GM}$ such that one gets again $(\phi_{*,\mathcal{A}})^{-1} = (\phi^{-1})_{*,\mathcal{A}}$, which we write as $\phi_{*,\mathcal{A}}^{-1}$. These properties, together with the fact that $\phi_{*,GM} \in \text{Aut}(GM)$ is in particular a principal bundle automorphism, identify $\phi_{*,\mathcal{A}}$ as an *associated vector bundle automorphism*, satisfying the following commutative diagram:

$$\begin{array}{ccc} \mathcal{A} & \xrightarrow{\phi_{*,\mathcal{A}}} & \mathcal{A} \\ \pi_{\mathcal{A}} \downarrow & & \downarrow \pi_{\mathcal{A}} \\ M & \xrightarrow{\phi} & M \end{array} \quad (13.17)$$

The associated bundle resulting from the specific choices of $\rho(g) = g$ as group representation and \mathbb{R}^d as typical fiber is via the bundle morphism $\chi : (GM \times \mathbb{R}^d)/\sim \rightarrow TM$ from Eq. (11.31) isomorphic to the tangent bundle TM (as G -bundle). Our definition of pushforwards on associated G -bundles is consistent with this identification since $\chi \circ \phi_{*,\mathcal{A}} = \phi_{*,TM} \circ \chi$.

To see this, let $[[e_i]_{i=1}^d, v] \in (GM \times \mathbb{R}^d)/\sim$ be an element of the isomorphic associated bundle that is mapped to $\chi([[e_i]_{i=1}^d, v]) = \sum_i e_i v_i$. Then we have $\chi \circ \phi_{*,A}([[e_i]_{i=1}^d, v]) = \chi([\phi_{*,TM}(e_i)]_{i=1}^d, v) = \sum_i \phi_{*,TM}(e_i)v_i = \phi_{*,TM}(\sum_i e_i v_i) = \phi_{*,TM} \circ \chi([[e_i]_{i=1}^d, v])$, which shows the consistency of the definitions.

As an associated bundle, the pushforward $\phi_{*,\text{Hom}}$ on the homomorphism bundle $\text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}}) \cong (GM \times \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}})/\sim_{\rho_{\text{Hom}}}$ is specified by Eq. (13.16) as well. However, we will later on require an expression of $\phi_{*,\text{Hom}}$ in terms of the pushforwards $\phi_{*,\mathcal{A}_{\text{in}}}$ and $\phi_{*,\mathcal{A}_{\text{out}}}$ of \mathcal{A}_{in} and \mathcal{A}_{out} , respectively, which we will shortly derive here. For that purpose, let $H \in \text{Hom}(\mathcal{A}_{\text{in}}|_p, \mathcal{A}_{\text{out}}|_p)$ be a homomorphism at p and $f_p \in \mathcal{A}_{\text{in},p}$ be a feature vector at p . Then $H(f_p)$ is by definition a feature vector in $\mathcal{A}_{\text{out},p}$. In order to be consistently defined, the pushforward of the input feature vector f_p , being acted on by the pushforward of the homomorphism H , needs to agree with the pushforward of the output feature vector $H(f_p)$. This implies

$$\phi_{*,\mathcal{A}_{\text{out}}} [H(f_p)] = [\phi_{*,\mathcal{A}_{\text{out}}} H \phi_{*,\mathcal{A}_{\text{in}}}^{-1}] (\phi_{*,\mathcal{A}_{\text{in}}} f_p) =: [\phi_{*,\text{Hom}} H] (\phi_{*,\mathcal{A}_{\text{in}}} f_p), \quad (13.18)$$

where we defined the pushforward on the homomorphism bundle as:

$$\phi_{*,\text{Hom}} : \text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}}) \rightarrow \text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}}), \quad H \mapsto \phi_{*,\mathcal{A}_{\text{out}}} H \phi_{*,\mathcal{A}_{\text{in}}}^{-1}, \quad \phi \in \text{Isom}_{GM} \quad (13.19)$$

Note that the composition of an element $H \in \text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}})$ with $\phi_{*,\mathcal{A}_{\text{out}}}$ on the left and with $\phi_{*,\mathcal{A}_{\text{in}}}^{-1}$ on the right mirrors the style of Eq. (12.17).

Isometry action on feature fields: The actions of isometries in Isom_{GM} on the associated bundles give rise to actions on their sections, in particular on feature fields. This pushforward of sections is defined as follows:

Definition 13.1.2 (Isometry pushforward of feature field:). Let $f \in \Gamma(\mathcal{A})$ be a feature field and let $\phi \in \text{Isom}_{GM}$ be a G -structure preserving isometry. The isometry acts on the feature field via the pushforward³

$$\triangleright : \text{Isom}_{GM} \times \Gamma(\mathcal{A}) \rightarrow \Gamma(\mathcal{A}), \quad (\phi, f) \mapsto \phi \triangleright f := \phi_{*,\mathcal{A}} \circ f \circ \phi^{-1}. \quad (13.20)$$

In terms of a commutative diagram, this definition is visualized as:

$$\begin{array}{ccc} \mathcal{A} & \xrightarrow{\phi_{*,\mathcal{A}}} & \mathcal{A} \\ f \uparrow & & \uparrow \phi \triangleright f := \phi_{*,\mathcal{A}} \circ f \circ \phi^{-1} \\ M & \xrightarrow{\phi} & M \end{array} \quad (13.21)$$

Intuitively, this definition states that the pushforward section $\phi \triangleright f$, evaluated at $p \in M$, returns the feature vector of f from $\phi^{-1}(p)$, pushed forward to p via $\phi_{*,\mathcal{A}}$. Note that such pushforwards do indeed yield well defined sections which satisfy

$$\begin{aligned} \pi_{\mathcal{A}} \circ (\phi \triangleright f) &= \pi_{\mathcal{A}} \circ \phi_{*,\mathcal{A}} \circ f \circ \phi^{-1} \\ &= \phi \circ \pi_{\mathcal{A}} \circ f \circ \phi^{-1} \\ &= \phi \circ \text{id}_M \circ \phi^{-1} \\ &= \text{id}_M \end{aligned} \quad (13.22)$$

³Note the similarity of this definition to that of the *induced representation action* on Euclidean feature fields in Def. 4.2.1, which can be viewed as coordinate expression of the pushforward.

as required by Eq. (11.19). Fig. 13.1a visualizes the action of isometries on fields. The action of isometries on the transporter pullback $\text{Exp}_p^* f$ of a fields f is derived in Section 13.1.4 below.

13.1.3 Isometry action in local coordinates

Most of the derivations on the isometry equivariance of kernel field transforms in Sections 13.2 and 13.3 will be kept in a coordinate free setting. However, since GM -convolutions are defined relative to a choice of G -atlases of the associated bundles, the investigation of their isometry equivariance will require us to study coordinate expressions of the isometry pushforwards $\phi_{*,TM}$, $\phi_{*,FM}$, $\phi_{*,GM}$ and $\phi_{*,A}$ relative to local bundle trivializations. Coordinate expressions of the isometry action are furthermore useful in numerical implementations, which are necessarily encoding feature fields relative to fields of reference frames.

In the following, we assume gauges $\Psi_{TM}^{\tilde{A}}$ and Ψ_{TM}^A on neighborhoods $U^{\tilde{A}}$ of p and U^A of $\phi(p)$ to be given. For convenience, let $U^A = \phi(U^{\tilde{A}})$ coincide with the image of $U^{\tilde{A}}$ under the isometry, which is always possible without losing generality.

Pushforward on TM in coordinates: Recall that the pushforward on the tangent bundle is a linear map from vectors $v \in T_p M$ to vectors $\phi_{*,TM} v \in T_{\phi(p)} M$. Relative to the given gauges, the pushforward is therefore coordinatized by a field of matrices⁴

$$g_\phi^{A\tilde{A}} : U^{\tilde{A}} \rightarrow \text{GL}(d), \quad p \mapsto g_\phi^{A\tilde{A}}(p) := \psi_{TM,\phi(p)}^A \circ \phi_{*,TM} \circ (\psi_{TM,p}^{\tilde{A}})^{-1}, \quad \phi \in \text{Isom}(M), \quad (13.23)$$

which transforms between the corresponding numerical coefficients $\psi_{TM,p}^{\tilde{A}}(v)$ of v at p and $\psi_{TM,\phi(p)}^A(\phi_{*,TM} v) = g_\phi^{A\tilde{A}}(p) \psi_{TM,p}^{\tilde{A}}(v)$ of $\phi_{*,TM} v$ at $\phi(p)$. More precisely, $g_\phi^{A\tilde{A}}$ takes values in the subgroup $\langle G \cup \text{O}(d) \rangle$ of $\text{GL}(d)$, which is generated by the elements of $\text{O}(d)$ (due to $\phi_{*,TM}$ preserving the metric) and G (since the transition functions might form a supergroup of $\text{O}(d)$). The definition of the pushforward in local coordinates is visualized by the following commutative diagram:

$$\begin{array}{ccccccc} \mathbb{R}^d & \xleftarrow{\psi_{TM,p}^{\tilde{A}}} & T_p M & \xrightarrow{\phi_{*,TM}} & T_{\phi(p)} M & \xrightarrow{\psi_{TM,\phi(p)}^A} & \mathbb{R}^d \\ & \underbrace{\hspace{10em}}_{g_\phi^{A\tilde{A}}(p)} & & & & & \uparrow \end{array} \quad (13.24)$$

Fig. 8.5 gives a graphical interpretation of the pushforward in coordinates.

Pushforward on FM in coordinates: The coordinatization of the pushforward on the frame bundle is defined in analogy to Eq. (13.23). It turns out to be given by the left action

⁴Given charts $x^{\tilde{A}} : U^{\tilde{A}} \rightarrow x^{\tilde{A}}(U^{\tilde{A}}) \subseteq \mathbb{R}^d$ and $x^A : U^A \rightarrow x^A(U^A) \subseteq \mathbb{R}^d$ of M , an isometry ϕ can be locally represented by a map $x^A \circ \phi \circ (x^{\tilde{A}})^{-1} : x^{\tilde{A}}(U^{\tilde{A}}) \rightarrow x^A(U^A)$ between coordinates. For the special case that the gauges at p and $\phi(p)$ correspond to the coordinate bases of those charts, $g_\phi^{A\tilde{A}}$ is simply given by the Jacobian of $x^A \circ \phi \circ (x^{\tilde{A}})^{-1}$.

of the same group element $g_\phi^{A\tilde{A}}$ on trivialized frames as shown in the commutative diagram below:

$$\begin{array}{ccccc} \text{GL}(d) & \xleftarrow{\psi_{FM,p}^{\tilde{A}}} & F_p M & \xrightarrow{\phi_{*,FM}} & F_{\phi(p)} M & \xrightarrow{\psi_{FM,\phi(p)}^A} & \text{GL}(d) \\ & \downarrow & & & & & \uparrow \\ & & g_\phi^{A\tilde{A}}(p). & & & & \end{array} \quad (13.25)$$

To prove this claim, we compute the action on a trivialized frame, given by a matrix $h \in \text{GL}(d)$ whose i -th column $h_{:,i}$ represents the i -th frame vector:

$$\begin{aligned} & \left[\psi_{FM,\phi(p)}^A \circ \phi_{*,FM} \circ (\psi_{FM,p}^{\tilde{A}})^{-1} \right] (h) \\ &= \left[\psi_{FM,\phi(p)}^A \circ \phi_{*,FM} \right] \left(\left((\psi_{TM,p}^{\tilde{A}})^{-1} (h_{:,i}) \right)_{i=1}^d \right) \quad (\text{def. of } \psi_{FM,p}^{\tilde{A}}, \text{Eq. (11.49)}) \\ &= \psi_{FM,\phi(p)}^A \left(\left(\phi_{*,TM} \circ (\psi_{TM,p}^{\tilde{A}})^{-1} (h_{:,i}) \right)_{i=1}^d \right) \quad (\text{def. of } \phi_{*,FM}, \text{Eq. (13.6)}) \\ &= \left(\left(\psi_{TM,\phi(p)}^A \circ \phi_{*,TM} \circ (\psi_{TM,p}^{\tilde{A}})^{-1} (h_{:,i}) \right)_{i=1}^d \right) \quad (\text{def. of } \psi_{FM,\phi(p)}^A, \text{Eq. (11.49)}) \\ &= \left(g_\phi^{A\tilde{A}}(p) \cdot h_{:,i} \right)_{i=1}^d \quad (\text{def. of } g_\phi^{A\tilde{A}}, \text{Eq. (13.23)}) \\ &= g_\phi^{A\tilde{A}}(p) \cdot h \end{aligned} \quad (13.26)$$

The action of the pushforward on local trivializations can be thought of as *inducing a gauge transformation*. A graphical intuition for this statement was given in Fig. 8.5 where the initial gauges at p and $\phi(p)$ are visualized by choices of reference frames. A pushforward of the frame at p to $\phi(p)$ (red) does in general not agree with the original frame at $\phi(p)$ (green). The transition between these two frames is the induced gauge transformation by at $\phi(p)$. We will in the following construct this transformation; first in terms of local trivializations, then in terms of the corresponding frame fields.

From the commutative diagram in Eq. (13.25) it is clear that the gauge $\psi_{FM,p}^{\tilde{A}} : F_p M \rightarrow \text{GL}(d)$ at p can via $\phi_{*,FM}^{-1}$ be pulled back to a gauge at $\phi(p)$, which is given by

$$\psi_{FM,p}^{\tilde{A}} \circ \phi_{*,FM}^{-1} = (g_\phi^{A\tilde{A}}(p))^{-1} \psi_{FM,\phi(p)}^A : F_{\phi(p)} M \rightarrow \text{GL}(d). \quad (13.27)$$

The corresponding extension of the commutative diagram in Eq. (13.25) visualizes the equivalence of both expressions and makes an algebraic proof superfluous:

$$\begin{array}{ccccc} \psi_{FM,p}^{\tilde{A}} \phi_{*,FM}^{-1} = (g_\phi^{A\tilde{A}}(p))^{-1} \cdot \psi_{FM,\phi(p)}^A & & & & \\ \downarrow & \xleftarrow{\psi_{FM,p}^{\tilde{A}}} & F_p M & \xrightarrow{\phi_{*,FM}} & F_{\phi(p)} M & \xrightarrow{\psi_{FM,\phi(p)}^A} & \text{GL}(d) \\ \text{GL}(d) & & & & & & & \uparrow \\ & & g_\phi^{A\tilde{A}}(p). & & & & & \end{array} \quad (13.28)$$

The transition map (gauge transformation) between the isometry induced gauge $\psi_{FM,p}^{\tilde{A}} \phi_{*,FM}^{-1}$ and the original gauge $\psi_{FM,\phi(p)}^A$ at $\phi(p)$ is read off to be given by the inverse⁵ group element

$$(\psi_{FM,p}^{\tilde{A}} \phi_{*,FM}^{-1}) \circ (\psi_{FM,\phi(p)}^A)^{-1} = (g_\phi^{A\tilde{A}}(p))^{-1} \in \langle G \cup O(d) \rangle \leq GL(d). \quad (13.29)$$

Note that this group element does for $G \leq O(d)$ not necessarily lie in the structure group, that is, the isometry induced gauge might not be G -compatible (it can not be added to an existing G -atlas of FM). In the next paragraph on G -structures we will show that this happens exactly when $\phi \notin \text{Isom}_{GM}$, i.e. for isometries which do not respect the G -structure.

To derive the isometry action on frame fields, consider the identity sections $\sigma^{\tilde{A}} : U^{\tilde{A}} \rightarrow \pi_{FM}^{-1}(U^{\tilde{A}})$ over $U^{\tilde{A}}$ and $\sigma^A : U^A \rightarrow \pi_{FM}^{-1}(U^A)$ over U^A . These sections model the original frame fields from Fig. 8.5. The new frame field is then given by the pushforward section

$$\phi \triangleright \sigma^{\tilde{A}} := \phi_{*,FM} \circ \sigma^{\tilde{A}} \circ \phi^{-1} : U^A \rightarrow \pi_{FM}^{-1}(U^A), \quad (13.30)$$

which is equivalently defined to that in Eq. (13.20). An alternative expression for the push-forward frame field in terms of the right action of $g_\phi^{A\tilde{A}}$ is found by applying $\psi_{FM,\phi(p)}^A$:

$$\begin{aligned} & \psi_{FM,\phi(p)}^A ([\phi \triangleright \sigma^{\tilde{A}}](\phi(p))) \\ &= \psi_{FM,\phi(p)}^A \phi_{*,FM} \sigma^{\tilde{A}}(p) && (\text{def. of } \phi \triangleright \sigma^{\tilde{A}}, \text{Eq. (13.30)}) \\ &= g_\phi^{A\tilde{A}}(p) \psi_{FM,p}^{\tilde{A}} \sigma^{\tilde{A}}(p) && (\text{equivalent expressions in Eq. (13.27)}) \\ &= g_\phi^{A\tilde{A}}(p) && (\text{identity section } \sigma^{\tilde{A}}, \text{Eq. (11.54)}) \\ &= \psi_{FM,\phi(p)}^A (\sigma^A(\phi(p))) g_\phi^{A\tilde{A}}(p) && (\text{identity section } \sigma^A, \text{Eq. (11.54)}) \\ &= \psi_{FM,\phi(p)}^A (\sigma^A(\phi(p)) \triangleleft g_\phi^{A\tilde{A}}(p)) && (\text{right } GL(d) \text{ equivariance, Eq. (11.51)}) \end{aligned} \quad (13.31)$$

Since $\psi_{FM,\phi(p)}^A$ is an isomorphism, it follows that

$$(\phi \triangleright \sigma^{\tilde{A}})(\phi(p)) = \phi_{*,FM} \sigma^{\tilde{A}}(p) = \sigma^A(\phi(p)) \triangleleft g_\phi^{A\tilde{A}}(p), \quad (13.32)$$

that is, $g_\phi^{A\tilde{A}}(p)$ does as expected describe the transformation between identity sections. This isometry induced transformation between reference frames is in Fig. 8.5 visualized by the blue arrow between the (translucent) red and green frame.

The isometry transformed gauge $\psi_{FM,p}^{\tilde{A}} \phi_{*,FM}^{-1}$ and the pushforward section $\phi \triangleright \sigma^{\tilde{A}}$ correspond to each other in so far that the latter is the identity section of the former:

$$\psi_{FM,p}^{\tilde{A}} \phi_{*,FM}^{-1} [\phi \triangleright \sigma^{\tilde{A}}](\phi(p)) = \psi_{FM,p}^{\tilde{A}} \phi_{*,FM}^{-1} \phi_{*,FM} \sigma^{\tilde{A}}(p) = \psi_{FM,p}^{\tilde{A}} \sigma^{\tilde{A}}(p) = e \quad (13.33)$$

Pushforward on GM in coordinates: As argued in the previous Section 13.1.2, the push-forward on GM is only well defined for isometries ϕ in a subgroup Isom_{GM} . Not surprisingly, the corresponding isometry induced gauge transformations take values in the structure group G :

⁵The inverse is a matter of convention. It arises here since we defined $g_\phi^{A\tilde{A}}$ as coordinate expression of the covariant pushforward of frames while gauges transform contravariantly.

Theorem 13.1.3 (Isom_{GM} in local trivializations). *Let $\phi \in \text{Isom}(M)$ be any isometry of M . Then the following three statements are equivalent:*

1. *ϕ is G -structure preserving, that is, $\phi \in \text{Isom}_{GM}$.*
2. *The isometry pullback $\psi_{FM,p}^{\tilde{A}} \phi_{*,FM}^{-1}$ of any gauge $\psi_{FM,p}^{\tilde{A}}$ of the G -atlas of FM that defines GM is G -compatible with that G -atlas.*
3. *The coordinate expression of $\phi_{*,FM}$ relative to any gauges $\psi_{FM,p}^{\tilde{A}}$ and $\psi_{FM,\phi(p)}^A$ from the G -atlas of FM takes values in the structure group, that is, $g_{\phi}^{A\tilde{A}}(p) \in G$ for any p in M .*

Proof: The defining property of a G -structure preserving isometry $\phi \in \text{Isom}_{GM}$ is that it satisfies $\phi_{*,FM}(G_p M) = G_{\phi(p)} M$ for any $p \in M$; see Eq. (13.10). In terms of a given G -atlas of FM , Eq. (11.56) defined the G -structure at $p \in M$ as $G_p M := (\psi_{FM,p}^{\tilde{A}})^{-1}(G)$ where $\psi_{FM,p}^{\tilde{A}}$ is an arbitrary gauge of the G -atlas. With this expression we expand the left-hand side of the defining property of Isom_{GM} :

$$\begin{aligned}\phi_{*,FM}(G_p M) &= \phi_{*,FM}(\psi_{FM,p}^{\tilde{A}})^{-1}(G) \\ &= (\psi_{FM,p}^{\tilde{A}} \phi_{*,FM})^{-1}(G)\end{aligned}\quad (13.34)$$

Relative to any gauge $\psi_{FM,\phi(p)}^A$ of the G -atlas at $\phi(p)$, this can be further manipulated to

$$\begin{aligned}\phi_{*,FM}(G_p M) &= ((g_{\phi}^{A\tilde{A}}(p))^{-1} \psi_{FM,\phi(p)}^A)^{-1}(G) \\ &= (\psi_{FM,\phi(p)}^A)^{-1}(g_{\phi}^{A\tilde{A}}(p) G).\end{aligned}\quad (13.35)$$

The right-hand side of the defining property of Isom_{GM} is in terms of $\psi_{FM,\phi(p)}^A$ given by

$$G_{\phi(p)} M = (\psi_{FM,\phi(p)}^A)^{-1}(G).\quad (13.36)$$

Setting both sides equal and using that $\psi_{FM,\phi(p)}^A$ is an isomorphism implies $g_{\phi}^{A\tilde{A}}(p) G = G$ which leads to the claimed equivalence

$$\phi_{*,FM}(G_p M) = G_{\phi(p)} M \iff g_{\phi}^{A\tilde{A}}(p) \in G\quad (13.37)$$

of statements 1. and 3. To prove the equivalence to statement 2., recall that $g_{\phi}^{A\tilde{A}}(p)$ is by Eq. (13.29) equal to the gauge transformation from $\psi_{FM,p}^{\tilde{A}} \phi_{*,FM}^{-1}$ to $\psi_{FM,\phi(p)}^A$. As G -atlases have by definition transition functions in the structure group G , the implications $(2 \leftrightarrow 3.)$ follow, such that all three statements are seen to be equivalent. \square

These results are of central importance for our later study of the isometry equivariance of GM -convolutions. We will be able to show that such convolutions are equivariant under the action of $\phi \in \text{Isom}_{GM}$ on feature fields, which relies on the fact that the G -steerability of the convolution kernels accounts for the isometry induced gauge transformations $g_{\phi}^{A\tilde{A}}(p) \in G$.

For G -structure automorphism inducing isometries $\phi \in \text{Isom}_{GM}$, we can adapt the commutative diagram for FM in Eq. (13.28) to its cousin for GM :

$$\begin{array}{ccccccc}
 & & \psi_{GM,p}^{\tilde{A}} \phi_{*,GM}^{-1} = (g_{\phi}^{A\tilde{A}}(p))^{-1} \cdot \psi_{GM,\phi(p)}^A & & & & \\
 & \downarrow & & & & & \\
 G & \xleftarrow{\psi_{GM,p}^{\tilde{A}}} & G_p M & \xrightarrow{\phi_{*,GM}} & G_{\phi(p)} M & \xrightarrow{\psi_{GM,\phi(p)}^A} & G \\
 & \boxed{ & & & & & } & & & & \\
 & & & & & & \\
 & & g_{\phi}^{A\tilde{A}}(p) \cdot & & & & \\
 & & & & & &
 \end{array} \quad (13.38)$$

Pushforward on \mathcal{A} in coordinates: The pushforward of $\phi \in \text{Isom}_{GM}$ on associated G -bundles is similarly coordinatized as those of the other bundles. In terms of a commutative diagram we get, not surprisingly,

$$\begin{array}{ccccc}
 \mathbb{R}^c & \xleftarrow{\psi_{\mathcal{A},p}^{\tilde{A}}} & \mathcal{A}_p & \xrightarrow{\phi_{*,\mathcal{A}}} & \mathcal{A}_{\phi(p)} & \xrightarrow{\psi_{\mathcal{A},\phi(p)}^A} & \mathbb{R}^c \\
 & \boxed{ & & & & & } & & & & \\
 & & \rho(g_{\phi}^{A\tilde{A}}(p)) & & & & \\
 & & & & & &
 \end{array}, \quad (13.39)$$

which follows when acting on feature vector coefficients $f \in \mathbb{R}^c$:

$$\begin{aligned}
 & \left[\psi_{\mathcal{A},\phi(p)}^A \circ \phi_{*,\mathcal{A}} \circ (\psi_{\mathcal{A},p}^{\tilde{A}})^{-1} \right] (f) \\
 &= \left[\psi_{\mathcal{A},\phi(p)}^A \circ \phi_{*,\mathcal{A}} \right] \left([\sigma^{\tilde{A}}(p), f] \right) && \left(\text{def. of } (\psi_{\mathcal{A},p}^{\tilde{A}})^{-1}, \text{Eq. (11.67)} \right) \\
 &= \psi_{\mathcal{A},\phi(p)}^A \left([\phi_{*,FM}(\sigma^{\tilde{A}}(p)), f] \right) && \left(\text{def. of } \phi_{*,\mathcal{A}}, \text{Eq. (13.16)} \right) \\
 &= \psi_{\mathcal{A},\phi(p)}^A \left([\sigma^A(\phi(p)) \triangleleft g_{\phi}^{A\tilde{A}}(p), f] \right) && \left(\text{induced gauge transformation, Eq. (13.32)} \right) \\
 &= \psi_{\mathcal{A},\phi(p)}^A \left([\sigma^A(\phi(p)), \rho(g_{\phi}^{A\tilde{A}}(p))f] \right) && \left(\text{def. of } \sim_{\rho}, \text{Eq. (11.42)} \right) \\
 &= \rho \left(g_{\phi}^{A\tilde{A}}(p) \right) \cdot f && \left(\text{def. of } \psi_{\mathcal{A},p}, \text{Eq. (11.66)} \right)
 \end{aligned} \quad (13.40)$$

Note that the expression $\rho(g_{\phi}^{A\tilde{A}}(p))$ requires $g_{\phi}^{A\tilde{A}}(p)$ to be a structure group element since ρ is a G -representation. This shows once again from another perspective that pushforwards on \mathcal{A} can only be defined for isometries in Isom_{GM} .

For completeness, we give the following local trivialization of the commutative diagram from Eq. (13.21), which might be useful when implementing coordinate independent CNNs

and testing their Isom_{GM} -equivariance:

$$\begin{array}{ccccccc}
 & & & \phi \times \rho(g_\phi^{A\tilde{A}}) \cdot & & & \\
 & \swarrow & \downarrow \Psi_{\mathcal{A}}^{\tilde{A}} & & \downarrow \phi_{*,\mathcal{A}} & \searrow & \\
 U^{\tilde{A}} \times \mathbb{R}^c & \xleftarrow{\quad} & \pi_{\mathcal{A}}^{-1}(U^{\tilde{A}}) & \xrightarrow{\quad} & \pi_{\mathcal{A}}^{-1}(U^A) & \xrightarrow{\quad} & U^A \times \mathbb{R}^c \\
 & \searrow & \downarrow \pi_{\mathcal{A}} \circledast f & & \downarrow \phi \triangleright f \circledast \pi_{\mathcal{A}} & \swarrow & \\
 & & U^{\tilde{A}} & \xrightarrow{\quad} & U^A & \xrightarrow{\quad} & \\
 & & \downarrow \phi & & \uparrow \phi & & \\
 & & U^{\tilde{A}} & \xrightarrow{\quad} & U^A & \xrightarrow{\quad} & \\
 & \nearrow \text{proj}_1 & & & \nearrow \text{proj}_1 & & \\
 & & & & & &
 \end{array} \tag{13.41}$$

13.1.4 Commutativity of isometry actions with the exponential map and transporters

In the following section we will need an expression for the action of isometries on transporter pullbacks $\text{Exp}_p^* f$ of feature fields f , which we derive here. For this purpose, we discuss the behavior of the exponential map and parallel transporters under the action of isometries.

Isometries and the exponential map: As proven in [100], isometries map geodesics to geodesics and do therefore in particular commute with the exponential map.⁶ More specifically, the identity

$$\exp_{\phi(p)} \circ \phi_{*,TM}(v) = \phi \circ \exp_p(v) \quad \forall v \in T_p M, \quad \phi \in \text{Isom}(M), \tag{13.42}$$

holds for any isometry and any tangent vector at p (still assuming a geodesically complete manifold). It states that the result of the exponential map at p , evaluated with some vector v and then being mapped through the isometry, equals the exponential map at $\phi(p)$ when being evaluated with the pushforward of v as visualized in Fig. 13.5 (left). This statement is diagrammatically expressed by the commutativity of (the upper square of) the following diagram:

$$\begin{array}{ccc}
 M & \xrightarrow{\phi} & M \\
 \uparrow \exp & & \uparrow \exp \\
 TM & \xrightarrow{\phi_{*,TM}} & TM \\
 \downarrow \pi_{TM} & & \downarrow \pi_{TM} \\
 M & \xrightarrow{\phi} & M
 \end{array} \tag{13.43}$$

⁶ The proof relies on the fact that the Levi-Civita connection $\nabla : \Gamma(TM) \times \Gamma(TM) \rightarrow \Gamma(TM)$, $(X, Y) \mapsto \nabla_X Y$, on which the Riemannian exponential map is based, commutes with isometries: $\phi \triangleright (\nabla_X Y) = \nabla_{\phi \triangleright X} (\phi \triangleright Y)$; see [100]

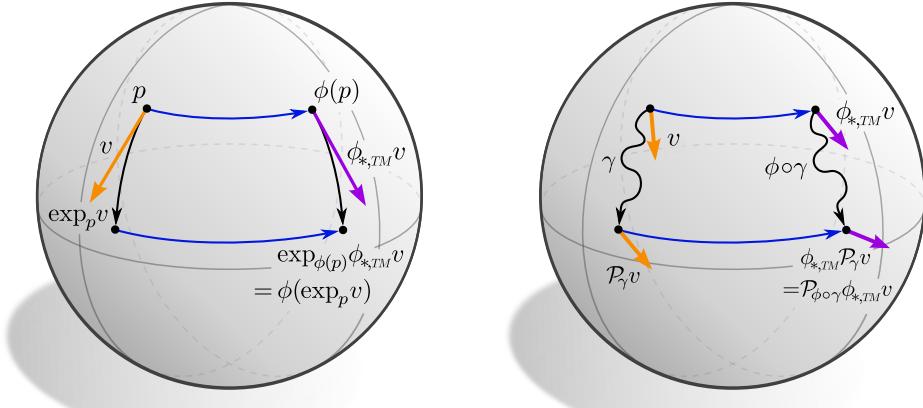


Figure 13.5: *Left:* Isometries commute with the exponential map, that is, $\exp_{\phi(p)} \circ \phi_{*,TM}(v) = \phi \circ \exp_p(v)$ for any vector $v \in T_p M$ and isometry $\phi \in \text{Isom}(M)$. *Right:* Isometries also commute with the Levi-Civita transport of tangent vectors and feature vectors, that is, $\phi_{*,A} \circ \mathcal{P}_{A,\gamma} = \mathcal{P}_{A,\phi \circ \gamma} \circ \phi_{*,A}$ for arbitrary paths $\gamma : [0, 1] \rightarrow M$ and isometries $\phi \in \text{Isom}(M)$. If an alternative, G -compatible connection is used, we demand that the same commutativity property holds for them. The isometry-invariance of exponential maps and transporters allows GM -convolutions to be equivariant under the action of isometries.

Isometries and parallel transporters: The pushforward on the tangent bundle was in [100] further argued to commute with the corresponding Levi-Civita transporters, as visualized in Fig. 13.5 (right). If an alternative, G -compatible connection is chosen to transport feature vectors, we *demand* that it commutes with the action of isometries as well. Since the transporters and the pushforwards on FM , GM and A are induced from those on TM , one can easily show that this property translates to them. Specifically for the associated feature vector bundles this means that for arbitrary isometries $\phi \in \text{Isom}_{GM}$ and paths γ we assume the relation

$$\phi_{*,A} \circ \mathcal{P}_{A,\gamma} = \mathcal{P}_{A,\phi \circ \gamma} \circ \phi_{*,A} \quad (13.44)$$

to hold, such that the following diagram commutes:

$$\begin{array}{ccc} \mathcal{A}_{\gamma(0)} & \xrightarrow{\phi_{*,A}} & \mathcal{A}_{\phi \circ \gamma(0)} \\ \mathcal{P}_{A,\gamma} \downarrow & & \downarrow \mathcal{P}_{A,\phi \circ \gamma} \\ \mathcal{A}_{\gamma(1)} & \xrightarrow{\phi_{*,A}} & \mathcal{A}_{\phi \circ \gamma(1)} \end{array} \quad (13.45)$$

Isometries and transporter pullbacks of feature fields: Knowing the transformation laws of exponential maps and transporters under the action of isometries, we have everything at hand that is required to derive the transformation law of transporter pullbacks $\text{Exp}_p^* f$ of feature fields f :

Theorem 13.1.4 (Isometry action on transporter pullbacks of feature fields). Let $f \in \Gamma(A)$ be any feature field and let $\phi \in \text{Isom}_{GM}$ be any G -structure preserving isometry.

Assume the feature vector transporters to commute with the action of Isom_{GM} , that is, that Eq. (13.44) holds (which is automatically guaranteed for the Levi-Civita connection). The transporter pullback (Def. 12.2.4) of the pushforward field $\phi \triangleright f$ (Def. 13.1.2) is then given by:

$$\text{Exp}_p^*(\phi \triangleright f) = \phi_{*, \mathcal{A}_{\text{out}}} \circ [\text{Exp}_{\phi^{-1}(p)}^* f] \circ \phi_{*, TM}^{-1} \quad (13.46)$$

Proof: We start by letting the right-hand side act on an arbitrary vector $v \in T_p M$ and work progressively to the left-hand side by using the properties derived in this section:

$$\begin{aligned} & \phi_{*, \mathcal{A}_{\text{out}}} [\text{Exp}_{\phi^{-1}(p)}^* f] \phi_{*, TM}^{-1} (v) && (13.47) \\ &= \phi_{*, \mathcal{A}_{\text{out}}} \mathcal{P}_{\mathcal{A}_{\text{in}}, \phi^{-1}(p) \leftarrow \exp_{\phi^{-1}(p)} \circ \phi_{*, TM}^{-1} (v)} \circ f \circ \exp_{\phi^{-1}(p)} \circ \phi_{*, TM}^{-1} (v) && (\text{Def. 12.2.4}) \\ &= \phi_{*, \mathcal{A}_{\text{out}}} \mathcal{P}_{\mathcal{A}_{\text{in}}, \phi^{-1}(p) \leftarrow \phi^{-1} \circ \exp_p (v)} \circ f \circ \phi^{-1} \circ \exp_p (v) && (\text{isom. action on exp, Eq. (13.42)}) \\ &= \mathcal{P}_{\mathcal{A}_{\text{in}}, p \leftarrow \exp_p (v)} \circ \phi_{*, \mathcal{A}_{\text{in}}} \circ f \circ \phi^{-1} \circ \exp_p (v) && (\text{isometry action on } \mathcal{P}_{\mathcal{A}_{\text{in}}}, \text{ Eq. (13.44)}) \\ &= \mathcal{P}_{\mathcal{A}_{\text{in}}, p \leftarrow \exp_p (v)} \circ (\phi \triangleright f) \circ \exp_p (v) && (\text{pushforward of fields, Eq. (13.20)}) \\ &= [\text{Exp}_p^*(\phi \triangleright f)](v) && (\text{transporter pullback, Def. 12.2.4}) \end{aligned}$$

□

Intuitively, this result just states that the transporter pullback of a pushforward field equals the pushforward of the original field's transporter pullback. Relative to local trivializations, this pushforward can be interpreted as an isometry induced gauge transformation, which was stated in Eq. (9.46). We will in the following assume that the G -compatible connection which is chosen to transport feature vectors will always be Isom_{GM} -invariant, and thus that Eq. (13.46) holds.

That the transporter pullback and the isometry pushforward commute is a consequence of the commutativity of the exponential map and parallel transporter, in terms of which the transporter pullback is defined. Note that general diffeomorphisms do not preserve the metric and thus the exponential map and the transporter pullback of feature fields. Being based on these constructions, kernel field transforms and GM -convolutions can only be isometry equivariant but not fully diffeomorphism equivariant.

13.2 Isometry equivariance of kernel field transforms and GM -convolutions

We now turn to investigate under which conditions kernel field transforms and GM -convolutions are equivariant w.r.t. the action of isometries on feature fields. As the action on the G -associated feature vector bundles is only defined for G -structure preserving isometries, we will formulate all statements for the subgroup $\text{Isom}_{GM} \leq \text{Isom}(M)$ or subgroups $\mathcal{I} \leq \text{Isom}_{GM}$ thereof. One can of course always consider structure groups $G \geq \text{O}(d)$, for which $\text{Isom}_{GM} = \text{Isom}(M)$.

The equivariance of a kernel field transform, and thus GM -convolutions, is defined as follows:

Definition 13.2.1 (Isometry equivariant kernel field transform).

Let $\mathcal{T}_{\mathcal{K}} : \Gamma(\mathcal{A}_{\text{in}}) \rightarrow \Gamma(\mathcal{A}_{\text{out}})$ be a kernel field transform. Then $\mathcal{T}_{\mathcal{K}}$ is said to be equivariant w.r.t the action of isometries in a subgroup $\mathcal{I} \leq \text{Isom}_{GM}$ if it commutes with this

action, that is, if the following property holds:

$$\mathcal{T}_{\mathcal{K}}(\phi \triangleright f) = \phi \triangleright (\mathcal{T}_{\mathcal{K}}(f)) \quad \forall f \in \Gamma(\mathcal{A}_{\text{in}}), \phi \in \mathcal{I} \quad (13.48)$$

In terms of a diagram, $\mathcal{T}_{\mathcal{K}}$ is equivariant w.r.t isometries in \mathcal{I} if

$$\begin{array}{ccc} \Gamma(\mathcal{A}_{\text{in}}) & \xrightarrow{\mathcal{T}_{\mathcal{K}}} & \Gamma(\mathcal{A}_{\text{out}}) \\ \phi \triangleright \downarrow & & \downarrow \phi \triangleright \\ \Gamma(\mathcal{A}_{\text{in}}) & \xrightarrow{\mathcal{T}_{\mathcal{K}}} & \Gamma(\mathcal{A}_{\text{out}}) \end{array} \quad (13.49)$$

commutes for all $\phi \in \mathcal{I}$.

A visualization of this definition is given in Fig. 9.3. In the following Section 13.2.1 we will derive a constraint on kernel fields in order for the corresponding kernel field transform to be isometry equivariant. The geometrically intuitive result which we obtain is that the kernel field itself is required to be invariant under the action of isometries, which implies a form of weight sharing over the isometry orbits; see Fig. 13.6. Section 13.2.2 applies these insights to the more specific case of *GM*-convolutions and *GM*-convolutional kernel fields. It turns out that *GM*-convolutions are by the *G*-steerability of their template kernel automatically equivariant with respect to *any* isometry in Isom_{GM} .

13.2.1 Isometry equivariance of general kernel field transforms

The main result of this section, Theorem 13.2.4, states that *a kernel field transform $\mathcal{T}_{\mathcal{K}}$ is isometry equivariant if and only if its underlying kernel field \mathcal{K} is invariant under isometries*. To make sense of this statement we start by defining the transformation behavior of kernel fields when being acted on by isometries.

Definition 13.2.2 (Isometry action on kernel fields). Let $\mathcal{K} : TM \rightarrow \text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}})$ be a kernel field as defined in Def. 12.2.1. An isometry $\phi \in \text{Isom}_{GM}$ acts on \mathcal{K} via the kernel field pushforward

$$\phi_{*,\mathcal{K}} \mathcal{K} := \phi_{*,\text{Hom}} \circ \mathcal{K} \circ \phi_{*,TM}^{-1}. \quad (13.50)$$

Intuitively, this pushforward of kernel fields can be thought of as moving the individual kernels \mathcal{K}_p at points $p \in M$ along the orbits of the isometry group to $\phi(p)$.

Since kernel fields are defined to be bundle M -morphisms, that is, to satisfy $\pi_{\text{Hom}} \mathcal{K} = \pi_{TM}$, their pushforward is only well defined if it preserves this property. This is guaranteed since the pushforward on the tangent bundle and homomorphism bundle are bundle maps, satisfying $\pi_{TM} \circ \phi_{*,TM} = \phi \circ \pi_{TM}$ (Eq. (13.4)) and $\pi_{\text{Hom}} \circ \phi_{*,\text{Hom}} = \phi \circ \pi_{\text{Hom}}$ (Eq. (13.17)), respectively:

$$\begin{aligned} \pi_{\text{Hom}} \phi_{*,\mathcal{K}} \mathcal{K} &= \pi_{\text{Hom}} \phi_{*,\text{Hom}} \mathcal{K} \phi_{*,TM}^{-1} \\ &= \phi \pi_{\text{Hom}} \mathcal{K} \phi_{*,TM}^{-1} \\ &= \phi \pi_{TM} \phi_{*,TM}^{-1} \\ &= \phi \phi^{-1} \pi_{TM} \\ &= \pi_{TM} \end{aligned} \quad (13.51)$$

We visualize the definition of the isometry action on kernel fields by a commutative diagram:

$$\begin{array}{ccccc}
 & & M & & \\
 & \nearrow \pi_{TM} & & \searrow \pi_{Hom} & \\
 TM & \xrightarrow{\phi_{*,\mathcal{K}}\mathcal{K}} & Hom(\mathcal{A}_{in}, \mathcal{A}_{out}) & & \\
 \uparrow \phi_{*,TM} & & \uparrow \phi_{*,Hom} & & \\
 TM & \xrightarrow{\mathcal{K}} & Hom(\mathcal{A}_{in}, \mathcal{A}_{out}) & & \\
 & \searrow \pi_{TM} & & \swarrow \pi_{Hom} & \\
 & M & & &
 \end{array} \tag{13.52}$$

The bottom part of this diagram shows the coordinate free kernel field \mathcal{K} from the diagram in Eq. (12.23) while the upper part shows its pushforward $\phi_{*,\mathcal{K}}\mathcal{K} = \phi_{*,Hom} \circ \mathcal{K} \circ \phi_{*,TM}^{-1}$ by $\phi \in \text{Isom}_{GM}$. The commutativity of the leftmost arrow, which asserts that $\phi_{*,\mathcal{K}}$ moves kernels from p to $\phi(p)$, follows from $\phi_{*,TM}$ and $\phi_{*,Hom}$ both being bundle maps over ϕ .

We proceed by defining isometry invariant kernel fields – a visualization is found in Fig. 13.6.

Definition 13.2.3 (Isometry invariant kernel fields). A kernel field \mathcal{K} is said to be invariant⁷ under isometries in $\mathcal{I} \leq \text{Isom}_{GM}$ if it satisfies the constraint $\phi_{*,\mathcal{K}}\mathcal{K} = \mathcal{K}$ for all $\phi \in \mathcal{I}$. We denote the space of isometry invariant kernel fields by

$$\mathcal{K}_{invar}^{\mathcal{I}} := \left\{ \mathcal{K} : TM \rightarrow \text{Hom}(\mathcal{A}_{in}, \mathcal{A}_{out}) \text{ smooth } \mid \begin{array}{l} \pi_{Hom} \circ \mathcal{K} = \pi_{TM}, \\ \phi_{*,\mathcal{K}}\mathcal{K} = \mathcal{K} \quad \forall \phi \in \mathcal{I} \end{array} \right\}. \tag{13.53}$$

By writing out $\phi_{*,\mathcal{K}}$, the invariance constraint reads

$$\phi_{*,Hom} \circ \mathcal{K} \circ \phi_{*,TM}^{-1} = \mathcal{K} \quad \forall \phi \in \mathcal{I}, \tag{13.54}$$

which, after further expanding $\phi_{*,Hom}$ as defined in Eq. (13.19), becomes:

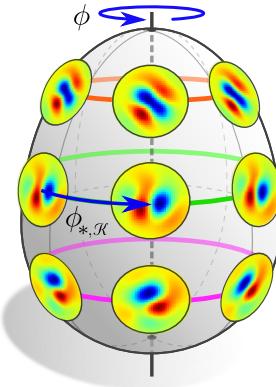
$$\phi_{*,\mathcal{A}_{out}} \mathcal{K}(\phi_{*,TM}^{-1} v) \phi_{*,\mathcal{A}_{in}}^{-1} = \mathcal{K}(v) \quad \forall v \in TM, \quad \forall \phi \in \mathcal{I} \tag{13.55}$$

Note the similarity of these kernel field constraints in Eqs. (13.54) and (13.55) with the G -steerability constraint on template kernels in Eqs. (12.29) and (12.30), respectively. Indeed, both constraints are closely related and imply each other to a certain extent as we will show in the following Section 13.2.2 on isometry equivariant GM -convolutions.

The following theorem proves that kernel fields which are invariant under isometries do indeed correspond to isometry equivariant kernel field transforms:

⁷Instead of saying that \mathcal{K} is *invariant*, one could call it *equivariant* since it satisfies $\phi_{*,Hom} \circ \mathcal{K} = \mathcal{K} \circ \phi_{*,TM}^{-1} \forall \phi \in \mathcal{I}$. This claim holds more generally, see Eq. (B.28) in Appendix B.4.

Figure 13.6: Visualization of an invariant kernel field \mathcal{K} for the case of an isometry (sub)group $\mathcal{I} = \text{SO}(2)$. The invariance constraint requires $\phi_{*,\mathcal{K}}\mathcal{K} := \phi_{*,\text{Hom}}\mathcal{K}\phi_{*,TM}^{-1} = \mathcal{K}$ for any ϕ in \mathcal{I} . It enforces kernels to be shared over the orbits $\mathcal{I}.p := \{\phi(p) \mid \phi \in \mathcal{I}\}$ of the action but allows for different kernels on different orbits. Theorem 13.2.4 proves that invariant kernel fields and equivariant kernel field transforms imply each other. This is intuitively clear since a specific pattern in the feature field at $p \in M$ will evoke the same response when being transported to $\phi(p)$ if and only if the kernels at both points coincide. For the choice of $\mathcal{I} = \text{O}(2)$ as isometry group, the kernels would additionally have to satisfy a reflectional constraint; see Fig. 13.7.



Theorem 13.2.4 (Equivariant kernel field transform \Leftrightarrow invariant kernel field).

A kernel field transform $\mathcal{T}_{\mathcal{K}} : \Gamma(\mathcal{A}_{\text{in}}) \rightarrow \Gamma(\mathcal{A}_{\text{out}})$ (Def. 12.2.5) is equivariant w.r.t. isometries in $\mathcal{I} \leq \text{Isom}_{GM}$ according to Def. 13.2.1 if and only if the underlying kernel field \mathcal{K} is invariant under isometries according to Def. 13.2.3, that is,

$$\mathcal{T}_{\mathcal{K}}(\phi \triangleright f) = \phi \triangleright \mathcal{T}_{\mathcal{K}}(f) \iff \phi_{*,\mathcal{K}}\mathcal{K} = \mathcal{K} \quad (13.56)$$

for any $\phi \in \mathcal{I}$ and any $f \in \Gamma(\mathcal{A}_{\text{in}})$.

Proof: To prove this theorem, we write out the kernel field transforms and feature field pushforwards on both sides of the isometry equivariance condition in Eq. (13.48). The statement follows from a comparison of both sides after a few algebraic manipulations.

We start with the right-hand side of Eq. (13.48):

$$\begin{aligned}
 [\phi \triangleright \mathcal{T}_{\mathcal{K}}(f)](p) &\stackrel{(1)}{=} \phi_{*,\mathcal{A}_{\text{out}}} [\mathcal{T}_{\mathcal{K}}(f)](\phi^{-1}(p)) \\
 &\stackrel{(2)}{=} \phi_{*,\mathcal{A}_{\text{out}}} \int_{T_{\phi^{-1}(p)}M} \mathcal{K}(v) [\text{Exp}_{\phi^{-1}(p)}^* f](v) dv \\
 &\stackrel{(3)}{=} \phi_{*,\mathcal{A}_{\text{out}}} \int_{T_{\phi^{-1}(p)}M} \mathcal{K}(v) \phi_{*,\mathcal{A}_{\text{in}}}^{-1} [\text{Exp}_p^*(\phi \triangleright f)](\phi_{*,TM}(v)) dv \\
 &\stackrel{(4)}{=} \int_{T_pM} \phi_{*,\mathcal{A}_{\text{out}}} \mathcal{K}(\phi_{*,TM}^{-1} \tilde{v}) \phi_{*,\mathcal{A}_{\text{in}}}^{-1} [\text{Exp}_p^*(\phi \triangleright f)](\tilde{v}) d\tilde{v} \\
 &\stackrel{(5)}{=} \int_{T_pM} [\phi_{*,\text{Hom}} \mathcal{K} \phi_{*,TM}^{-1}](\tilde{v}) [\text{Exp}_p^*(\phi \triangleright f)](\tilde{v}) d\tilde{v} \\
 &\stackrel{(6)}{=} \int_{T_pM} [\phi_{*,\mathcal{K}}\mathcal{K}](\tilde{v}) [\text{Exp}_p^*(\phi \triangleright f)](\tilde{v}) d\tilde{v} \quad (13.57)
 \end{aligned}$$

Steps (1) and (2) expand the isometry action \triangleright on feature fields (Def. 13.1.2) and the kernel field transform (Def. 12.2.5). The transformation law of the field's transporter pullback in Theorem 13.1.4, which relies on the Isom_{GM} -invariance of the G -compatible connection, justifies step (3). Step (4) substitutes v with $\tilde{v} = \phi_{*,TM}v$. Since

ϕ is an isometry, the change of volume equates to 1. Steps (5) and (6) identify the action of the kernel pushforward $\phi_{*,\mathcal{K}}$, Def. 13.2.2. The resulting statement is quite intuitive: A transformation of the kernel field transform's output corresponds to a simultaneous transformation of its input *and* kernel field.

Writing out the left-hand side yields

$$[\mathcal{T}_{\mathcal{K}}(\phi \triangleright f)](p) = \int_{T_p M} \mathcal{K}(v) [\text{Exp}_p^*(\phi \triangleright f)](v) dv, \quad (13.58)$$

which is equivalent to the right-hand side *up to* the transformation of the kernel field.

Isometry equivariance requires both expressions to agree for arbitrary fields $f \in \Gamma(\mathcal{A}_{\text{in}})$, points $p \in M$ and isometries $\phi \in \mathcal{I}$. This is the case if and only if $\phi_{*,\mathcal{K}}\mathcal{K} = \mathcal{K}$ holds for any $\phi \in \mathcal{I}$, i.e. if the kernel field is invariant under the action of isometries. \square

Note that this proof would have been very cumbersome to work out in (a G -atlas of) local trivializations. The global, coordinate free description of kernel field transforms allows for a simple proof without having to worry that the isometries move features between different local trivializations.

At this point we could proceed with a further investigation of isometry invariant kernel fields: since the invariance constraint implies kernels to be shared over orbits of the isometry group, a description of the entire kernel field on the full manifold is redundant. It is therefore possible to reduce the description of such kernel fields to kernel fields on quotient spaces. As this analysis is not required to prove the isometry equivariance of *GM*-convolutions and requires some technical definitions, we postpone it to Section 13.3

13.2.2 Isometry equivariance of *GM*-convolutions

Recall that *GM*-convolutions (Def. 12.2.7) were defined as specific kernel field transforms with *GM*-convolutional kernel fields (Def. 12.2.3). The results on the isometry equivariance of kernel field transforms therefore immediately apply to *GM*-convolutions as well. However, in addition to the isometry invariance constraint in Eq. (13.54), *GM*-convolutional kernel fields need to satisfy the G -steerability constraint on the template kernel from Eq. (12.29) and share weights over the G -structure according to Eq. (12.31). In order for the *GM*-convolution to be isometry equivariant, all of these constraints have to be satisfied simultaneously. Intuitively, this implies that the convolutional weight sharing needs to agree with the isometry induced weight sharing over orbits. Luckily it turns out that this is automatically the case for the isometries under consideration: *GM*-convolutions share weights over the G -structure and the isometries in Isom_{GM} preserve the G -structure such that *GM-convolutional kernel fields are guaranteed to be Isom_{GM} invariant*. In coordinates, this reflects in the Isom_{GM} -induced gauge transformations $g_{\phi}^{A\bar{A}}(p)$ taking values in the structure group G , such that they are explained away by the G -steerability of the template kernels.

To make these arguments more rigorous, consider a *GM*-convolution $K\star : \Gamma(\mathcal{A}_{\text{in}}) \rightarrow \Gamma(\mathcal{A}_{\text{out}})$ with some G -steerable kernel $K \in \mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^G$, which is by Def. 12.2.7 just the kernel field transform $\mathcal{T}_{\mathcal{K}_K}$ with the *GM*-convolutional kernel field \mathcal{K}_K . By Theorem 13.2.4, the *GM*-convolution is therefore exactly then Isom_{GM} -equivariant if \mathcal{K}_K is Isom_{GM} -invariant, i.e. when it satisfies $\phi_{*,\mathcal{K}}\mathcal{K}_K = \phi_{*,\text{Hom}} \circ \mathcal{K}_K \circ \phi_{*,\text{TM}}^{-1} = \mathcal{K}_K$ for any $\phi \in \text{Isom}_{GM}$. This

constraint on the full kernel field is equivalently expressed by a set of constraints on the individual convolution kernels that make up the field:

$$\phi_{*,\text{Hom}} \circ \mathcal{K}_{K,p} \circ \phi_{*,\text{TM}}^{-1} = \mathcal{K}_{K,\phi(p)} \quad \forall p \in M, \phi \in \text{Isom}_{GM} \quad (13.59)$$

Considering a specific point $p \in M$, we choose arbitrary gauges \tilde{A} at p and A at $\phi(p)$ from the G -atlas. The GM -convolutional kernel field is by Def. 12.2.3 at p and $\phi(p)$ given by

$$\mathcal{K}_{K,p} := (\psi_{\text{Hom},p}^{\tilde{A}})^{-1} \circ \frac{K}{\sqrt{|\eta_p^{\tilde{A}}|}} \circ \psi_{\text{TM},p}^{\tilde{A}} \quad (13.60)$$

$$\text{and} \quad \mathcal{K}_{K,\phi(p)} := (\psi_{\text{Hom},\phi(p)}^A)^{-1} \circ \frac{K}{\sqrt{|\eta_{\phi(p)}^A|}} \circ \psi_{\text{TM},\phi(p)}^A. \quad (13.61)$$

Plugging these expressions into the constraint in Eq. (13.59) for the fixed p and identifying $\phi_{*,\text{Hom}}(\psi_{\text{Hom},p}^{\tilde{A}})^{-1}$ with $(\psi_{\text{Hom},p}^{\tilde{A}} \phi_{*,\text{Hom}}^{-1})^{-1}$ yields, for any $\phi \in \text{Isom}_{GM}$:

$$(\psi_{\text{Hom},p}^{\tilde{A}} \phi_{*,\text{Hom}}^{-1})^{-1} \circ \frac{K}{\sqrt{|\eta_p^{\tilde{A}}|}} \circ \psi_{\text{TM},p}^{\tilde{A}} \phi_{*,\text{TM}}^{-1} = (\psi_{\text{Hom},\phi(p)}^A)^{-1} \circ \frac{K}{\sqrt{|\eta_{\phi(p)}^A|}} \circ \psi_{\text{TM},\phi(p)}^A \quad (13.62)$$

The isometry equivariance will therefore hold if the weight sharing via the isometry induced gauges $\psi_{(\cdot),p}^{\tilde{A}} \phi_{*,(\cdot)}$ agrees with the weight sharing via the original gauges $\psi_{(\cdot),\phi(p)}^A$ from $\phi(p)$. Recall that the isometry induced gauges are by Theorem 13.1.3 for isometries in Isom_{GM} guaranteed to be compatible with the G -atlases (of the corresponding bundle). As shown in Eq. (12.32), the particular choice of gauge, relative to which the G -steerable template kernel is oriented, is irrelevant, as long as the gauges are G -compatible. Since all derivations were independent from the chosen point p and the particular choice of gauges, this implies that GM -convolutions are by design guaranteed to be Isom_{GM} -equivariant.

To gain a better intuition for this result it is worth to make the induced, G -valued gauge transformations $g_\phi^{A\tilde{A}}(p)$ explicit. To this end, note that the commutativity of the diagrams in Eqs. (13.39) and (13.24) implies $\psi_{\text{Hom},p}^{\tilde{A}} \phi_{*,\text{Hom}}^{-1} = \rho_{\text{Hom}}(g_\phi^{A\tilde{A}}(p))^{-1} \psi_{\text{Hom},\phi(p)}^A$ and $\psi_{\text{TM},p}^{\tilde{A}} \phi_{*,\text{TM}}^{-1} = (g_\phi^{A\tilde{A}}(p))^{-1} \psi_{\text{TM},\phi(p)}^A$. Inserting these coordinate expressions into the constraint in Eq. (13.62) leads to the requirement that

$$\begin{aligned} & \left(\rho_{\text{Hom}}(g_\phi^{A\tilde{A}}(p))^{-1} \psi_{\text{Hom},\phi(p)}^A \right)^{-1} \circ \frac{K}{\sqrt{|\eta_p^{\tilde{A}}|}} \circ (g_\phi^{A\tilde{A}}(p))^{-1} \psi_{\text{TM},\phi(p)}^A \\ &= (\psi_{\text{Hom},\phi(p)}^A)^{-1} \circ \frac{K}{\sqrt{|\eta_{\phi(p)}^A|}} \circ \psi_{\text{TM},\phi(p)}^A \end{aligned} \quad (13.63)$$

needs to hold for any isometry ϕ in Isom_{GM} . By expanding the inverse on the left-hand side, using that $\sqrt{|\eta_p^{\tilde{A}}|} = \sqrt{|\eta_{\phi(p)}^A|} \cdot |\det g_\phi^{A\tilde{A}}(p)|$ and dropping the gauges, which is possible since they are isomorphisms, we end up with the constraint

$$\frac{1}{|\det g_\phi^{A\tilde{A}}(p)|} \rho_{\text{Hom}}(g_\phi^{A\tilde{A}}(p)) \circ K \circ (g_\phi^{A\tilde{A}}(p))^{-1} = K \quad \forall \phi \in \text{Isom}_{GM}, \quad (13.64)$$

which looks *exactly* like the G -steerability kernel constraint on K from Def. 12.2.2. Recall that the isometry induced gauge transformations $g_\phi^{A\tilde{A}}(p)$ are by Theorem 13.1.3 guaranteed to be G -valued if ϕ is an element of Isom_{GM} . The constraint in Eq. (13.64) is therefore always satisfied by the G -steerability of K .

The derived results on the Isom_{GM} -invariance of GM -convolutional kernel fields \mathcal{K}_K are concisely summarized by the statement that the following diagram is guaranteed to be commutative if K is G -steerable and if $\phi \in \text{Isom}_{GM}$ is G -structure preserving:

$$\begin{array}{ccccccc}
 & & U^A & & & & \\
 & & \pi_{TM} \nearrow & \nwarrow \pi_{\text{Hom}} & & & \\
 U^A \times \mathbb{R}^d & \xleftarrow{\Psi_{TM}^A} & \pi_{TM}^{-1}(U^A) & \xrightarrow{\phi_{*,\mathcal{R}} \mathcal{K}_K} & \pi_{\text{Hom}}^{-1}(U^A) & \xrightarrow{\Psi_{\text{Hom}}^A} & U^A \times \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \\
 \uparrow \phi \times g_\phi^{A\tilde{A}}. & & \uparrow \phi_{*,TM} & & \uparrow \phi_{*,\text{Hom}} & & \uparrow \phi \times \rho_{\text{Hom}}(g_\phi^{A\tilde{A}}) \\
 U^{\tilde{A}} \times \mathbb{R}^d & \xleftarrow{\Psi_{TM}^{\tilde{A}}} & \pi_{TM}^{-1}(U^{\tilde{A}}) & \xrightarrow{\mathcal{K}_K} & \pi_{\text{Hom}}^{-1}(U^{\tilde{A}}) & \xrightarrow{\Psi_{\text{Hom}}^{\tilde{A}}} & U^{\tilde{A}} \times \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \\
 & & \searrow \pi_{TM} & \swarrow \pi_{\text{Hom}} & & & \\
 & & U^{\tilde{A}} & & & &
 \end{array}$$

$\boxed{\quad}$

$\text{id} \times K / \sqrt{|\eta^{\tilde{A}}|}$

Here we defined the pullback $g_\phi'^{A\tilde{A}} := g_\phi^{A\tilde{A}} \circ \phi^{-1} : U^A \rightarrow G$ of the isometry pushforward coordinatization from $U^{\tilde{A}}$ to U^A for notational convenience.

Together with Theorem 13.2.4, the Isom_{GM} -invariance of GM -convolutional kernel fields implies the Isom_{GM} -equivariance of GM -convolutions:

Theorem 13.2.5 (Isometry equivariance of GM -convolutions). A GM -convolution $K \star : \Gamma(\mathcal{A}_{\text{in}}) \rightarrow \Gamma(\mathcal{A}_{\text{out}})$ with a G -steerable kernel $K \in \mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^G$ is equivariant with respect to all G -structure preserving isometries $\phi \in \text{Isom}_{GM}$, that is,

$$K \star (\phi \triangleright f) = \phi \triangleright (K \star f) \quad \forall f \in \Gamma(\mathcal{A}_{\text{in}}), \phi \in \text{Isom}_{GM}. \quad (13.65)$$

The following diagram commutes therefore for every $\phi \in \text{Isom}_{GM}$:

$$\begin{array}{ccc}
 \Gamma(\mathcal{A}_{\text{in}}) & \xrightarrow{K \star} & \Gamma(\mathcal{A}_{\text{out}}) \\
 \phi \triangleright \downarrow & & \downarrow \phi \triangleright \\
 \Gamma(\mathcal{A}_{\text{in}}) & \xrightarrow{K \star} & \Gamma(\mathcal{A}_{\text{out}})
 \end{array}$$

(13.66)

Proof: The proof was given in the discussion prior to the theorem. \square

Having this general result on GM -convolutions derived, we will now discuss some special cases for specific choices of structure groups G . Firstly, for orthogonal structure groups $G = O(d)$ (or supergroups of it), the convolution will commute with *any* isometry:

Theorem 13.2.6 (Full isometry equivariance of OM-convolutions). *OM-convolutions are equivariant w.r.t. the action of any isometry $\phi \in \text{Isom}(M)$ on feature fields. More generally, any GM-convolution for G-structures with structure groups $G \geq O(d)$ is fully isometry equivariant.*

Proof: The statement follows from Theorem (13.2.5) by observing that $\text{Isom}_{GM} = \text{Isom}(M)$ is guaranteed for structure groups $G \geq O(d)$. The latter was discussed in Eq. (13.14). \square

This result relies essentially on the fact that isometries are defined as that subgroup of diffeomorphisms on M which induce $O(d)$ -structure automorphisms. Less abstractly stated, $\text{Isom}(M)$ is by definition that subgroup of diffeomorphisms which respect the Riemannian metric η of M and the corresponding $O(d)$ -structure OM is equivalent information to the metric.

On orientable Riemannian manifolds one can furthermore pick an orientation (frame handedness), which together with the metric defines an $SO(d)$ -structure SOM . The corresponding isometries which lift to $SO(d)$ structure automorphisms are the orientation preserving isometries in $\text{Isom}_+(M)$.

Theorem 13.2.7 ($\text{Isom}_+(M)$ equivariance of SOM-convolutions). *SOM-convolutions are equivariant w.r.t. the action of orientation preserving isometries $\phi \in \text{Isom}_+(M)$ on feature fields.*

Proof: This result follows from Theorem (13.2.5) by observing that $\text{Isom}_{SOM} = \text{Isom}_+(M)$. \square

For instance, an SOM-convolution for $M = \mathbb{R}^2$, corresponding to Fig 13.4a, is equivariant w.r.t. the action of the special Euclidean group $\text{Isom}_+(\mathbb{R}^2) = SE(2)$. Similarly, an SOM-convolution for $M = S^2$, corresponding to Fig 13.4b, is rotation equivariant with $\text{Isom}_+(S^2) = SO(3)$.

Note that the results of Theorems 13.2.6 and 13.2.7 depend only on the structure group G but not on the particular choice of G -structure. For subgroups G of $O(d)$ (or $SO(d)$) things become more complicated. In these cases the subgroups Isom_{GM} of $\text{Isom}(M)$ depend on the specific embedding of the G -structure GM into FM . This was for $G = \{e\}$ visualized in Fig. 13.3. Specifically, Fig. 13.3a shows the canonical $\{e\}$ -structure of \mathbb{R}^2 , which is fully translation equivariant, that is, $\text{Isom}_{\{e\}M} = (\mathbb{R}^2, +)$. In contrast, Fig. 13.3b shows an $\{e\}$ -structure of \mathbb{R}^2 which is only translation equivariant along one axis such that $\text{Isom}_{\{e\}M} \cong (\mathbb{R}^1, +)$. From the viewpoint of convolutional networks this result is very intuitive: The $\{e\}$ -steerable kernels in these examples are unconstrained, i.e. conventional convolution kernels. They do therefore in general not carry any information about their responses when being applied relative to gauge transformed reference frames. Since the frames, and therefore kernels, in Fig. 13.3b are differently rotated along the “left-right” direction, the kernel responses change unpredictably when translating a signal in that direction. If the template kernels would, however be $SO(2)$ -steerable, they could account for the rotation of frames. This case corresponds to the situation in Fig 13.4a, i.e. an SOM-convolution.

13.3 Quotient kernel fields

Theorem 13.2.4 showed that the isometry equivariance of a kernel field transform requires the invariance of the corresponding kernel field. Since the invariance constraint implies kernels to be shared over orbits as visualized in Fig. 13.6, the mathematical description of such

invariant kernel fields is redundant: a single kernel at one orbit representative is sufficient to reconstruct the kernel field on the whole orbit. In Section 13.3.2 we derive equivalent, reduced descriptions of invariant kernel fields in terms of kernels on orbit representatives. These representative kernels are themselves constrained by the action of the stabilizer subgroup of the orbit representative. We propose a (unique) lifting from representative kernels to invariant kernel fields, which establishes an isomorphism between both descriptions. This lifting isomorphism suggest a way of parameterizing and constructing isometry equivariant kernel field transforms in an implementation. Before deriving these results in Section 13.3.2, the following Section 13.3.1 sets up the mathematical framework.

The derivations and results of this section are close in spirit to the theory of *steerable CNNs on homogeneous spaces* [55, 56], however, we generalize their results from homogeneous spaces to general manifolds. When sticking to homogeneous spaces M , we prove that isometry equivariant kernel field transforms are equivalent to GM -convolutions.

13.3.1 Isometry induced quotient spaces

The action of a symmetry group on a space partitions it into orbits, defined as the sets of all points which are connected by the group action. The space of such orbits is the *quotient space* w.r.t. this group action. In the following we will discuss the quotient spaces arising from the actions of some isometry group $\mathcal{I} \leq \text{Isom}_{GM}$ both on the manifold and on the fiber bundles. These definitions will later allow us to share weights over orbits by acting with isometries on kernels.

Manifold quotients: Any point $p \in M$ traces out an *orbit* (Def. B.3.3)

$$\mathcal{I}.p := \{\phi(p) \mid \phi \in \mathcal{I}\} \subseteq M, \quad (13.67)$$

which is defined as the set of all points reached by acting on p with any isometry in $\mathcal{I} \leq \text{Isom}(M)$. One can easily check that the relation “ p and q are elements of the same orbit” is an equivalence relation (see footnote 11) and thus *partitions* the manifold as visualized in Fig. 13.7a. The quotient space (Def. B.3.4)

$$\mathcal{I}\backslash M := \{\mathcal{I}.p \mid p \in M\} \quad (13.68)$$

with respect to this equivalence relation is the space of all orbits, that is, each element of $\mathcal{I}\backslash M$ corresponds to a full orbit in M .⁸ The corresponding *quotient map*

$$Q_M : M \rightarrow \mathcal{I}\backslash M, \quad p \mapsto \mathcal{I}.p \quad (13.69)$$

identifies a point $p \in M$ with its orbit $\mathcal{I}.p \in \mathcal{I}\backslash M$. For each orbit one can select an arbitrary *orbit representative* (Def. B.3.5), formally determined by a *section*

$$r_M : \mathcal{I}\backslash M \rightarrow M \quad \text{such that} \quad Q_M \circ r_M = \text{id}_{\mathcal{I}\backslash M}, \quad (13.70)$$

where the last condition ensures that the representative $r_M(\mathcal{I}.p)$ is indeed an element of the orbit $\mathcal{I}.p$. One is often interested in continuous (or smooth) sections, however, these do in general not exist. We will therefore in the following *not* demand the orbit representatives to be chosen continuously and make up for this shortcoming post-hoc if necessary. As usual for sections, they are in general only right inverses of the quotient map but not left inverses,

⁸We write $\mathcal{I}\backslash M$ as a left quotient since \mathcal{I} acts on M from the left.

that is, $r_M \circ Q_M \neq \text{id}_M$. This is visualized by a commutative diagram

$$\begin{array}{ccccc} \mathcal{I}\backslash M & \xrightarrow{r_M} & M & \xrightarrow{Q_M} & \mathcal{I}\backslash M \\ & \underbrace{\hspace{10em}}_{\text{id}_{\mathcal{I}\backslash M}} & & & \uparrow \end{array} \quad (13.71)$$

similar to that in Eq. (11.19) and a non-commutative diagram

$$\begin{array}{ccccc} M & \xrightarrow{Q_M} & \mathcal{I}\backslash M & \xrightarrow{r_M} & M \\ & \underbrace{\hspace{4em}}_{\emptyset} & & & \uparrow \\ & & \text{id}_M & & \end{array} \quad (13.72)$$

similar to that in Eq. (11.20). The individual fibers $\text{preim}_{Q_M}(\mathcal{I}.p) = \mathcal{I}.p \subseteq M$ of the quotient map Q_M are given by the orbits themselves. Note that $M \xrightarrow{Q_M} \mathcal{I}\backslash M$ is in general *not* a fiber bundle since the orbits are not necessarily homeomorphic to each other and can therefore not be locally trivialized with a shared typical fiber F , as required by the commutative diagram in Eq. (11.1). Each orbit therefore has an own *type* which is in close relation to the stabilizer subgroups of the points on that particular orbit. The *stabilizer subgroup* (Def. B.3.6)

$$\text{Stab}_p := \{\xi \in \mathcal{I} \mid \xi(p) = p\} \leq \mathcal{I} \quad (13.73)$$

of a point $p \in M$ is thereby defined as that subgroup of the isometry group which leaves p fixed. In terms of the stabilizer subgroup, it holds that the orbit of a point is identified with

$$\mathcal{I}.p \cong \mathcal{I}/\text{Stab}_p, \quad (13.74)$$

which is known as orbit-stabilizer theorem B.3.7. To see this claim, let $f_p : \mathcal{I} \rightarrow \mathcal{I}.p$, $\phi \mapsto \phi(p)$ for some $p \in M$ and observe that $f_p(\phi \circ \xi) = \phi \circ \xi(p) = \phi(p) = f_p(\phi)$ for any $\xi \in \text{Stab}_p$. It can easily be shown that indeed $\text{preim}_{f_p}(\phi(p)) = \phi$. Stab_p is a coset of the stabilizer subgroup of p and thus that f_p establishes the claimed isomorphism $\mathcal{I}.p \cong \mathcal{I}/\text{Stab}_p$.

To make these constructions more intuitive, consider the example in Fig. 13.7a with $\mathcal{I} \cong O(2)$. The orbits $\mathcal{I}.n = \{n\}$ and $\mathcal{I}.s = \{s\}$ of the north and south pole are just points, which are fixed by \mathcal{I} . This agrees with, for instance, $\mathcal{I}.n \cong \mathcal{I}/\text{Stab}_n = \mathcal{I}/\mathcal{I} \cong \{n\}$ since $\text{Stab}_n = \mathcal{I}$ coincides with the full isometry group. For any other point $p \in M$, the orbits $\mathcal{I}.p$ are circles. We have reflections $\text{Stab}_p \cong \mathcal{R}$ (flipping over p) as stabilizer subgroup and thus indeed get the circle $\mathcal{I}/\text{Stab}_p \cong O(2)/\mathcal{R} \cong S^1$ as orbit type. The quotient map $Q_M : M \rightarrow \mathcal{I}\backslash M$ sends points $q \in M$ to their orbits $Q_M(q) = \mathcal{I}.q$ in the quotient space $\mathcal{I}\backslash M$, shown on the right. Since the orbits can be traversed from the north to the south pole, the quotient space $\mathcal{I}\backslash M$ has the topology of a line segment. The section $r_M : \mathcal{I}\backslash M \rightarrow M$ picks one representative point $r_M(o) \in M$ for any orbit $o \in \mathcal{I}\backslash M$. In general, this orbit representative does not recover a projected point. For instance, we have that $r_M Q_M(p) \neq p$. One can interpret the section as embedding the quotient space $\mathcal{I}\backslash M$ into the manifold, shown as the black line $r_M(\mathcal{I}\backslash M)$ from the north to the south pole.

Bundle quotients: Since the isometry group acts not only on the manifold itself but via pushforwards also on the associated bundles, these bundles are in a similar manner partitioned into orbits. To keep the discussion general, we are in the following considering

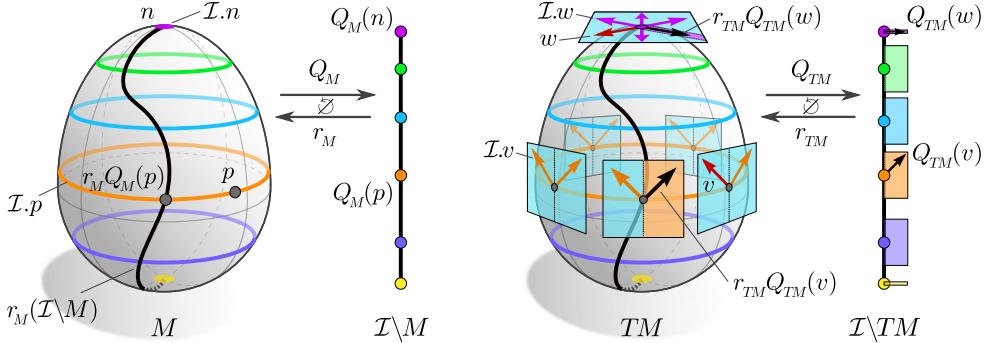
(a) Quotient map and orbit representatives for M . (b) Quotient map and orbit representatives for TM .

Figure 13.7: Quotient maps Q_M and Q_{TM} and orbit representatives (sections) r_M and r_{TM} for the actions of the isometry group $\mathcal{I} = O(2)$ on the manifold M in Fig. 13.7a and on the tangent bundle TM in Fig. 13.7b. A detailed description of both visualizations is given in the main text.

a generic associated bundle $E \xrightarrow{\pi_E} M$, which could stand for TM , FM , GM , \mathcal{A} or $\text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}})$. We denote elements of the total space as $e \in E$ and let $\phi_{*,E}$ be the push-forward of ϕ on E as introduced in Section 13.1.2. The orbit of an element of the bundle is then in analogy to Eq. (13.67) given by

$$\mathcal{I}.e = \{\phi_{*,E}(e) \mid \phi \in \mathcal{I}\} \quad (13.75)$$

while the quotient space, consisting of bundle orbits, is analogously to Eq. (13.68) defined as

$$\mathcal{I}\backslash E = \{\mathcal{I}.e \mid e \in E\}. \quad (13.76)$$

Similar to before, the (canonical) quotient map sends bundle elements to their orbit:

$$Q_E : E \mapsto \mathcal{I}\backslash E, \quad e \mapsto \mathcal{I}.e \quad (13.77)$$

We define a (uniquely determined) projection map

$$\pi_{\mathcal{I}\backslash E} : \mathcal{I}\backslash E \rightarrow \mathcal{I}\backslash M, \quad Q_E(e) \mapsto Q_M \circ \pi_E(e) \quad (13.78)$$

between the bundle quotients and manifold quotient as visualized in the following commutative diagram:

$$\begin{array}{ccc} \mathcal{I}\backslash E & \xleftarrow{Q_E} & E \\ \pi_{\mathcal{I}\backslash E} \downarrow & & \downarrow \pi_E \\ \mathcal{I}\backslash M & \xleftarrow{Q_M} & M \end{array} \quad (13.79)$$

Note that the definition in Eq. (13.78) does not depend on the particular choice of orbit representative since for any other $\phi_{*,E}(e) \in Q_E(e)$ we obtain the same result: $Q_M \circ \pi_E \circ \phi_{*,E}(e) = Q_M \circ \phi \circ \pi_E(e) = Q_M \circ \pi_E(e)$. Orbit representatives are formally determined by a choice of section

$$r_E : \mathcal{I}\backslash E \rightarrow E \quad \text{such that} \quad Q_E \circ r_E = \text{id}_{\mathcal{I}\backslash E}, \quad (13.80)$$

which we again do not demand to be continuous. However, for convenience we demand the representatives of bundle orbits to lie above the representatives $r_M(\mathcal{I}\backslash M)$ in the base space, that is, to satisfy

$$\pi_E \circ r_E = r_M \circ \pi_{\mathcal{I}\backslash E} \quad (13.81)$$

as shown in the commutative diagram below:

$$\begin{array}{ccc} \mathcal{I}\backslash E & \xrightarrow{r_E} & E \\ \pi_{\mathcal{I}\backslash E} \downarrow & & \downarrow \pi_E \\ \mathcal{I}\backslash M & \xrightarrow{r_M} & M \end{array} \quad (13.82)$$

The stabilizer subgroup of a bundle element $e \in E$ is defined as

$$\text{Stab}_e := \{\xi \in \mathcal{I} \mid \xi_{*,E} e = e\} \leq \text{Stab}_{\pi_E(e)} \leq \mathcal{I}. \quad (13.83)$$

It is necessarily a subgroup of the stabilizer subgroup $\text{Stab}_{\pi_E(e)}$ of the point $\pi_E(e)$ in the base space, which is easily seen by $\xi \in \text{Stab}_e \Leftrightarrow \xi_{*,E} e = e \Rightarrow \pi_E(\xi_{*,E} e) = \xi \pi_E(e) = \pi_E(e) \Leftrightarrow \xi \in \text{Stab}_{\pi_E(e)}$. As before, the relation $\mathcal{I}.e \cong \mathcal{I}/\text{Stab}_e$ holds.

We extend our example from Fig. 13.7a by considering the action of $\mathcal{I} \cong O(2)$ on the tangent bundle TM of the egg M in Fig. 13.7b. The orbit (violet) of a non-zero vector $0 \neq w \in T_n M$ (red) at the north pole n describes a circle in $T_n M$. This is consistent with $\mathcal{I}.w \cong \mathcal{I}/\text{Stab}_w \cong O(2)/\mathcal{R} \cong S^1$ since such a vector is stabilized by reflections $\text{Stab}_w \cong \mathcal{R}$ along its axis. The orbit of $0 \in T_n M$ is a single point in TM , which is stabilized by any isometry. Any other vector $v \in T_p M$ (red), living in a tangent space at a point $p \in M$ different from the poles, is by the action of the isometry group rotated and reflected to other tangent spaces $T_{\phi(p)} M$ on the orbit $\mathcal{I}.p$ of p . The orbit $\mathcal{I}.v$ (orange) of any such vector, if not pointing exactly to the north or south, is given by an eastward and a westward pointing copy of the vector in each of the tangent spaces over $\mathcal{I}.p$. We have $\text{Stab}_v = \{e\}$ for such vectors and indeed the orbit $\mathcal{I}.v \cong \mathcal{I}/\text{Stab}_v \cong O(2)/\{e\}$ is homeomorphic to $O(2)$ (or two circles). Vectors $v' \in T_p M$ which do point exactly north- or southwards are stabilized by reflections over the axis which they define, that is, $\text{Stab}_{v'} \cong \mathcal{R}$. Their orbit is homeomorphic to a circle $\mathcal{I}.v' \cong \mathcal{I}/\text{Stab}_{v'} \cong O(2)/\mathcal{R} \cong S^1$.

The quotient map $Q_{TM} : TM \rightarrow \mathcal{I}\backslash TM$ projects the tangent bundle to the bundle quotient $\mathcal{I}\backslash TM$, shown in the right half of Fig. 13.7b. To understand its structure, we consider all qualitatively different cases: Firstly, note that the orbits of vectors at the poles correspond to circles of a certain radius, such that the set of such orbits forms a line $\pi_{\mathcal{I}\backslash E}^{-1}(\mathcal{I}.n) \cong \mathbb{R}_{\geq 0}$ (pink ray under the black arrow). Similarly, the orbits of vectors at any other point $p \in M$ intersect all tangent spaces $T_{\phi(p)} M$ over $\mathcal{I}.p$ in two reflections and therefore form a half plane $\pi_{\mathcal{I}\backslash E}^{-1}(\mathcal{I}.p) \cong \mathbb{R} \times \mathbb{R}_{\geq 0}$ (orange). The section $r_{TM} : \mathcal{I}\backslash TM \rightarrow TM$ sends each bundle quotient element to some representative in TM . By the requirement in Eq. (13.81), these representatives are required to lie in the same fiber over the representatives $r_M(\mathcal{I}\backslash M)$ of the manifold quotient $\mathcal{I}\backslash M$, shown as the black line. For instance, $v \in T_p M$ (red) is by the quotient map sent to $Q_{TM}(v) \in \mathcal{I}\backslash TM$ (black). The section represents $Q_{TM}(v)$ by $r_{TM} Q_{TM}(v)$ (also black), which is an element of $T_{r_M Q_M(p)} M$ and does in general differ from v .

13.3.2 Quotient representative kernel fields and stabilizer constraints

To motivate the construction of quotient representative kernel fields and stabilizer constraints, consider the more explicit formulation

$$\phi_{*,\text{Hom}} \circ \mathcal{K}_p \circ \phi_{*,TM}^{-1} = \mathcal{K}_{\phi(p)} \quad \forall p \in M, \phi \in \mathcal{I}. \quad (13.84)$$

of the isometry invariance constraint from Def. 13.2.3, which follows by writing out Eq. (13.54) for any point $p \in M$ individually. This formulation emphasizes that the constraint leads to *shared weights along the manifold orbits* $\mathcal{I}.p \in \mathcal{I}\backslash M$ as visualized in Figs. 13.6 and 13.7. It implies that the kernel \mathcal{K}_r at an *arbitrary representative point* $r = r_M(o)$ of any orbit $o = \mathcal{I}.r$ fully specifies the kernel field on the rest of the orbit, i.e. at all points $\phi(r)$ where $\phi \in \mathcal{I}$. The kernel \mathcal{K}_r at the representative point r is itself constrained by the stabilizer subgroup of r :

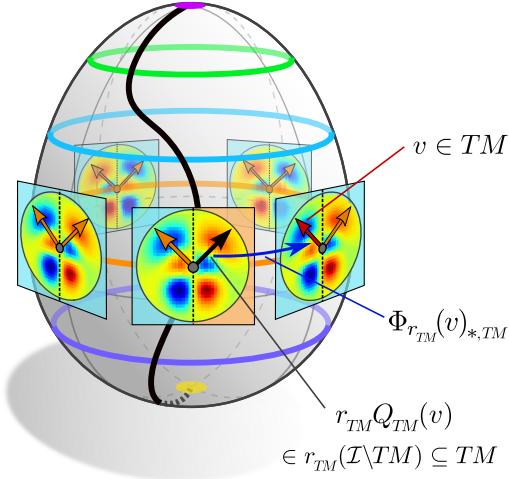
$$\xi_{*,\text{Hom}} \circ \mathcal{K}_r \circ \xi_{*,TM}^{-1} = \mathcal{K}_r \quad \forall \xi \in \text{Stab}_r. \quad (13.85)$$

This implies that any isometry invariant kernel field can be parameterized in terms of a field of kernels on manifold orbit representatives $r \in r_M(\mathcal{I}\backslash M)$ which satisfy Eq. (13.85).

In case that the stabilizer subgroup at r happens to be non-trivial, the stabilizer constraint in Eq. (13.85) implies further symmetries of the kernel \mathcal{K}_r at r itself. For instance, in the example in Fig. 13.7 one has the stabilizer subgroup $\text{Stab}_r \cong \mathcal{R}$ on the highlighted orbit, enforcing a reflectional symmetry of the kernels. Such stabilizer symmetries allows to compress the description of isometry invariant kernel fields further: it turns out to be sufficient to know the values $\mathcal{K}(w)$ of the kernel field on the tangent bundle quotient representatives $w \in r_{TM}(\mathcal{I}\backslash TM) \subseteq TM$ only. In Fig. 13.7 this corresponds to knowing the kernel values on the orange highlighted half space, from which the full field on the orbit can be reconstructed by the reflectional and rotational symmetries in $\mathcal{I} \cong O(2)$.

Theorem 13.3.1 below makes the latter claim precise by proving that the space $\mathcal{K}_{\text{invar}}^{\mathcal{I}}$ of isometry invariant kernel fields is isomorphic to a space $\mathcal{K}_{\text{quot}}^{\mathcal{I}}$ of kernel fields on tangent bundle orbit representatives $r_{TM}(\mathcal{I}\backslash TM)$. $\mathcal{K}_{\text{quot}}^{\mathcal{I}}$ is characterized by maximally reduced constraints and thus encodes the kernel fields in $\mathcal{K}_{\text{invar}}^{\mathcal{I}}$ in a non-redundant way. It can therefore be viewed as the distilled degrees of freedom contained in $\mathcal{K}_{\text{invar}}^{\mathcal{I}}$. In Theorem 13.3.2 we formulate a third isomorphic space $\hat{\mathcal{K}}_{\text{quot}}^{\mathcal{I}}$, which equivalently describes isometry invariant kernel fields in terms of the stabilizer subgroup constrained kernels \mathcal{K}_r from Eq. (13.85). While the formulation of isometry invariant kernel fields in terms of $\hat{\mathcal{K}}_{\text{quot}}^{\mathcal{I}}$ involves stronger constraints than that in terms of $\mathcal{K}_{\text{quot}}^{\mathcal{I}}$, it might be more convenient for implementations, since it describes kernels on full tangent spaces instead of kernels on quotients of tangent spaces.

Reconstruction isometries: In order to reconstruct full invariant kernel fields in $\mathcal{K}_{\text{invar}}^{\mathcal{I}}$ from single kernels on orbit representatives, the representative kernels need to be redistributed over the full manifold by applying the kernel pushforward in Eq. (13.84) with $p = r$ fixed to the chosen representative points. For the kernel reconstruction at some point $q \in M$, this requires some isometry ϕ which maps the orbit representative $r_M Q_M(q) \in r_M(\mathcal{I}\backslash M) \subseteq M$ back to $q \in M$, that is, which satisfies $\phi(r_M Q_M(q)) = q$. To make this more precise, recall that kernel fields $\mathcal{K} : TM \rightarrow \text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}})$ are defined as maps with domain TM , encoding the *kernel alignments* in addition to their position. We therefore need to consider more specific isometries which push tangent bundle orbit representatives



reconstruction isometry $\Phi_{r_{TM}}(v) \in \mathcal{I}$, defined in Eq. (13.86), back to vectors $v \in TM$. We want to mention that the visualized *antisymmetric* kernels would result when mapping between feature fields of even and odd parity, while kernels between feature fields of the same parity would be symmetric.

$r_{TM}Q_{TM}(v) \in r_{TM}(\mathcal{I}\setminus TM) \subseteq TM$ back to vectors $v \in TM$. These *reconstruction isometries* are defined by:⁹

$$\Phi_{r_{TM}} : TM \rightarrow \mathcal{I} \quad \text{such that} \quad \Phi_{r_{TM}}(v)_{*,TM} r_{TM}Q_{TM}(v) = v \quad \forall v \in TM \quad (13.86)$$

We recommend to consult Fig. 13.7 to get an intuition for the reconstruction isometries: graphically, $\Phi_{r_{TM}}(v)$ is defined as *any* isometry which pushes the black vector $r_{TM}Q_{TM}(v)$ on the orange orbit $\mathcal{I}.v$ back to the red vector v on the same orbit. Note that $\Phi_{r_{TM}}$ is only unique up to the stabilizer subgroups of the orbit representatives since for any $\xi \in \text{Stab}_{r_{TM}Q_{TM}(v)}$ it follows that $\Phi_{r_{TM}}(v)\xi$ satisfies the defining constraint in Eq. (13.86) as well:¹⁰ $[\Phi_{r_{TM}}(v)\xi]_{*,TM} r_{TM}Q_{TM}(v) = \Phi_{r_{TM}}(v)_{*,TM} r_{TM}Q_{TM}(v) = v$. All of the following constructions are shown to be independent from this ambiguity. The action of reconstruction isometries on the base space M follows by applying the tangent bundle projection to both sides of the defining constraint in Eq. (13.86):

$$\begin{aligned} \pi_{TM}(v) &= \pi_{TM}\Phi_{r_{TM}}(v)_{*,TM} r_{TM}Q_{TM}(v) && (\text{Def. of } \Phi_{r_{TM}}, \text{Eq. (13.86)}) \\ &= \Phi_{r_{TM}}(v)\pi_{TM}r_{TM}Q_{TM}(v) && (\text{Pushforward is a bundle map, Eq. (13.3)}) \\ &= \Phi_{r_{TM}}(v)r_M\pi_{\mathcal{I}\setminus TM}Q_{TM}(v) && (\text{Def. of bundle sections, Eq. (13.81)}) \\ &= \Phi_{r_{TM}}(v)r_MQ_M\pi_{TM}(v) && (\text{Def. of } \pi_{\mathcal{I}\setminus E}, \text{Eq. (13.79)}) \end{aligned} \quad (13.87)$$

⁹Since the sections r_{TM} are in general not continuous, $\Phi_{r_{TM}}$ can in general not be demanded to be continuous either.

¹⁰: Furthermore, the defining constraint on $\Phi_{r_{TM}}$ is fulfilled when *left* multiplying $\Phi_{r_{TM}}(v)$ with any $\zeta \in \text{Stab}_v$. This does, however, not add any new degrees of freedom since $\text{Stab}_v \cong \text{Stab}_{r_{TM}Q_{TM}(v)}$ and $\zeta\Phi_{r_{TM}}(v) = \Phi_{r_{TM}}(v)[\Phi_{r_{TM}}(v)^{-1}\zeta\Phi_{r_{TM}}(v)] =: \Phi_{r_{TM}}(v)\tilde{\zeta}$ with $\tilde{\zeta} \in \text{Stab}_{r_{TM}Q_{TM}(v)}$.

A visual summary of the properties of $\Phi_{r_{TM}}$, that is, its actions on TM and M , is given in the following commutative diagram

$$\begin{array}{ccccc}
 & & \text{id}_{TM} & & \\
 & \swarrow & & \searrow & \\
 TM & \xrightarrow{\Phi_{r_{TM}} \times r_{TM} \circ Q_{TM}} & \mathcal{I} \times r_{TM}(\mathcal{I} \setminus TM) & \xrightarrow{\text{ev}} & TM \\
 \pi_{TM} \downarrow & & \text{id}_{\mathcal{I}} \times \pi_{TM} \downarrow & & \pi_{TM} \downarrow \\
 M & & \mathcal{I} \times r_M(\mathcal{I} \setminus M) & \xrightarrow{\text{ev}} & M \\
 & \uparrow & & & \uparrow \text{id}_M
 \end{array} \quad (13.88)$$

where the *evaluation maps* ev are, overloading the notation, given by $\text{ev} : \mathcal{I} \times M \rightarrow M$, $(\phi, p) \mapsto \phi(p)$ and $\text{ev} : \mathcal{I} \times TM \rightarrow TM$, $(\phi, v) \mapsto \phi_{*,TM}(v)$, respectively.

Quotient representative kernel fields: As argued above, the symmetries which are present in an isometry invariant feature field $\mathcal{K} \in \mathcal{K}_{\text{invar}}^{\mathcal{I}}$ should allow for its full reconstruction from its restriction $\mathcal{K}|_{r_{TM}(\mathcal{I} \setminus TM)} : r_{TM}(\mathcal{I} \setminus TM) \rightarrow r_{\text{Hom}}(\mathcal{I} \setminus \text{Hom})$ to tangent bundle orbit representatives $r_{TM}(\mathcal{I} \setminus TM) \subseteq TM$.¹¹ To construct a (unique) *lift* Λ which recovers $\mathcal{K} = \Lambda(\mathcal{K}|_{r_{TM}(\mathcal{I} \setminus TM)})$ from $\mathcal{K}|_{r_{TM}(\mathcal{I} \setminus TM)}$, we expand tangent vectors v in the domain of \mathcal{K} via the reconstruction isometry $\Phi_{r_{TM}}$ from Eq. (13.86) and make use of the invariance (equivariance) of the kernel field in Eq. (13.54). This leads to:

$$\begin{aligned}
 \mathcal{K}(v) &= \mathcal{K} \Phi_{r_{TM}}(v)_{*,TM} r_{TM} Q_{TM}(v) \\
 &= \Phi_{r_{TM}}(v)_{*,\text{Hom}} \mathcal{K} r_{TM} Q_{TM}(v) \\
 &= \Phi_{r_{TM}}(v)_{*,\text{Hom}} \mathcal{K}|_{r_{TM}(\mathcal{I} \setminus TM)} r_{TM} Q_{TM}(v) \\
 &=: [\Lambda(\mathcal{K}|_{r_{TM}(\mathcal{I} \setminus TM)})](v)
 \end{aligned} \quad (13.89)$$

Note that this construction is well defined despite the ambiguity of $\Phi_{r_{TM}}$ w.r.t. the right multiplication with elements in $\text{Stab}_{r_{TM} Q_{TM}(v)}$. This is easily seen by observing that for any $w \in TM$, any $\xi \in \text{Stab}_w$ and any $\mathcal{K} \in \mathcal{K}_{\text{invar}}^{\mathcal{I}}$ one has $\xi_{*,\text{Hom}} \mathcal{K}(w) = \mathcal{K}(\xi_{*,TM} w) = \mathcal{K}(w)$, which implies that $\text{Stab}_{\mathcal{K}(w)} \geq \text{Stab}_w$, and thus that the final result does not depend on the particular choice of the ambiguous $\Phi_{r_{TM}}$.

Since the lift Λ recovers invariant kernel fields from their restriction to tangent bundle orbit representatives, it can be viewed as the *inverse* map of the restriction (of invariant kernel fields). This viewpoint implies that the lift establishes an *isomorphism* $\Lambda : \mathcal{K}_{\text{quot}}^{\mathcal{I}} \rightarrow \mathcal{K}_{\text{invar}}^{\mathcal{I}}$ between the image of the restriction $\mathcal{K}_{\text{quot}}^{\mathcal{I}}$, which we still need to characterize, and $\mathcal{K}_{\text{invar}}^{\mathcal{I}}$:

$$\begin{array}{ccc}
 & \overset{\Lambda}{\curvearrowright} & \\
 \mathcal{K}_{\text{quot}}^{\mathcal{I}} & \leftarrow \curvearrowright & \mathcal{K}_{\text{invar}}^{\mathcal{I}}
 \end{array} \quad (13.90)$$

$$\Lambda^{-1} = (\cdot)|_{r_{TM}(\mathcal{I} \setminus TM)}$$

¹¹In the following we might abbreviate $\text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}})$ and $\mathcal{I} \setminus \text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}})$ with Hom and $\mathcal{I} \setminus \text{Hom}$, respectively.

In order to characterize the space $\mathcal{K}_{\text{quot}}^{\mathcal{I}}$ which makes Λ to an isomorphism, it is sufficient to list the properties of restricted fields $\mathcal{Q} := \mathcal{K}|_{r_{TM}(\mathcal{I} \setminus TM)} \in \mathcal{K}_{\text{quot}}^{\mathcal{I}}$ for $\mathcal{K} \in \mathcal{K}_{\text{invar}}^{\mathcal{I}}$:

- First of all, since Λ^{-1} is given by the restriction of the domain to $r_{TM}(\mathcal{I} \setminus TM)$, it is clear that any $\mathcal{Q} \in \mathcal{K}_{\text{quot}}^{\mathcal{I}}$ is required to be of the form $\mathcal{Q} : r_{TM}(\mathcal{I} \setminus TM) \rightarrow r_{\text{Hom}}(\mathcal{I} \setminus \text{Hom})$.
- Secondly, the property of kernel fields to be bundle M -morphisms translates under the restriction Λ^{-1} to the requirement on \mathcal{Q} to satisfy $\pi_{\text{Hom}} \circ \mathcal{Q}(w) = \pi_{TM}(w)$ for any $w \in r_{TM}(\mathcal{I} \setminus TM)$.
- Thirdly, \mathcal{Q} is required to satisfy the (*vector*) *stabilizer constraint* $\xi_{*,\text{Hom}} \mathcal{Q}(w) = \mathcal{Q}(w)$ for any representative vector $w \in r_{TM}(\mathcal{I} \setminus TM)$ and any $\xi \in \text{Stab}_w$. This requirement is a residual from the invariance constraint in Eq. (13.54), surviving the restriction.

It can be deduced by considering the full constraint $\phi_{*,\text{Hom}} \mathcal{Q} \phi_{*,TM}^{-1}(w) = \mathcal{Q}(w)$ for any $w \in r_{TM}(\mathcal{I} \setminus TM)$ and any isometry $\phi \in \mathcal{I}$ which additionally satisfies that $\phi_{*,TM}(w) \in r_{TM}(\mathcal{I} \setminus TM)$, i.e. that the pushforward $\phi_{*,TM}(w)$ stays within the restricted domain of \mathcal{Q} . Note that $\phi_{*,TM}(w) \in \mathcal{I}.w$ and that $r_{TM}(\mathcal{I} \setminus TM)$ intersects each orbit exactly once. This implies that $\mathcal{I}.w \cap r_{TM}(\mathcal{I} \setminus TM) = \{w\}$ such that $\phi \in \mathcal{I}$ is required to satisfy $\phi_{*,TM}(w) = w$, that is, $\phi \in \text{Stab}_w$. The claimed (vector) stabilizer constraint follows from these considerations.

For an intuition we refer back to Fig. 13.7 where the black representative vector $w = r_{TM} Q_{TM}(v)$ is stabilized only by the trivial isometry $\xi = \{e\}$, implying that the corresponding value of \mathcal{Q} is unconstrained. Vectors $w' \in r_{TM}(\mathcal{I} \setminus TM)$ which point exactly north- or southwards, i.e. which lie on the dashed reflection axis, are stabilized by reflections in $\text{Stab}_{w'} \cong \mathcal{R}$, implying a constraint on the corresponding kernel values.¹²

- As a last requirement, \mathcal{Q} needs to lift to a *smooth* kernel field, that is, $\Lambda(\mathcal{Q})$ is required to be smooth. Unfortunately, the smoothness (or even continuity) of $\Lambda(\mathcal{Q})$ does not automatically follow from the smoothness (continuity) of \mathcal{Q} since Λ is defined in terms of r_{TM} and $\Phi_{r_{TM}}$, which can in general not be demanded to be smooth (continuous).

Before summarizing and proving these claims rigorously in Theorem 13.3.1 below, we give a visual overview of the relation between $\mathcal{Q} = \mathcal{K}|_{r_{TM}(\mathcal{I} \setminus TM)} \in \mathcal{K}_{\text{quot}}^{\mathcal{I}}$ and its lift $\mathcal{K} = \Lambda(\mathcal{Q}) \in \mathcal{K}_{\text{invar}}^{\mathcal{I}}$ in terms of commutative diagrams

$$\begin{array}{ccccc}
 & & \mathcal{K} = \Lambda(\mathcal{Q}) & & \\
 & \swarrow & & \searrow & \\
 TM & \xrightarrow[\Phi_{r_{TM}} \times \mathcal{Q} \circ r_{TM} \circ Q_{TM}]{} & \mathcal{I} \times r_{\text{Hom}}(\mathcal{I} \setminus \text{Hom}) & \xrightarrow{\text{ev}} & \text{Hom}
 \end{array} \tag{13.91}$$

¹²The exact constraint depends on the action $\xi_{*,\text{Hom}}$ on $\text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}})$, which depends on ρ_{Hom} and thus on ρ_{in} and ρ_{out} . The visualized kernel in Fig. (13.7) would correspond to ρ_{Hom} being the sign-flip (odd parity) representation of the reflection group, which enforces antisymmetric kernels. The antisymmetry requires \mathcal{Q} to be constrained to $\mathcal{Q}(w') = -\mathcal{Q}(w) = 0$ for w' on the reflection axis; cf. Table 5.1

and

$$\begin{array}{ccccc}
 & & \mathcal{K} = \Lambda(\mathcal{Q}) & & \\
 TM & \xrightarrow{\quad r_{TM} \circ Q_{TM} \quad} & & \xrightarrow{\quad r_{Hom} \circ Q_{Hom} \quad} & Hom \\
 & \searrow & \downarrow \mathcal{Q} & \swarrow & \\
 r_{TM}(\mathcal{I}\setminus TM) & & r_{Hom}(\mathcal{I}\setminus Hom) & & \\
 & \pi_{TM} \searrow & \downarrow \pi_{Hom} & \swarrow \pi_{Hom} & \\
 & & r_M(\mathcal{I}\setminus M) & & \\
 & & \uparrow r_M \circ Q_M & & \\
 & & M & &
 \end{array} \tag{13.92}$$

In the last diagram, the commutativity of the top square follows by inserting the definition of the lift, which yields $r_{Hom} Q_{Hom} \Lambda(\mathcal{Q}) = r_{Hom} Q_{Hom} \Phi_{r_{TM}}(v)_{*,Hom} \mathcal{Q} r_{TM} Q_{TM} = \mathcal{Q} r_{TM} Q_{TM}$. The commutative of the bottom left and right squares follows from Eqs. 13.81 and 13.78.

Theorem 13.3.1 (Tangent quotient representative kernel fields). *The space of isometry invariant kernel fields $\mathcal{K}_{invar}^{\mathcal{I}}$ from Def. 13.2.3 is isomorphic to the space $\mathcal{K}_{quot}^{\mathcal{I}}$ of (vector) stabilizer subgroup constrained kernel fields on tangent bundle quotient representatives, defined as:¹³*

$$\mathcal{K}_{quot}^{\mathcal{I}} := \left\{ \mathcal{Q} : r_{TM}(\mathcal{I}\setminus TM) \rightarrow r_{Hom}(\mathcal{I}\setminus Hom) \mid \begin{aligned} \pi_{Hom} \circ \mathcal{Q} &= \pi_{TM}, & \Lambda(\mathcal{Q}) &\text{ smooth,} \\ \xi_{*,Hom} \mathcal{Q}(w) &= \mathcal{Q}(w) & \forall w \in r_{TM}(\mathcal{I}\setminus TM), \xi \in \text{Stab}_w \end{aligned} \right\} \tag{13.93}$$

The (unique) lifting isomorphism $\Lambda : \mathcal{K}_{quot}^{\mathcal{I}} \rightarrow \mathcal{K}_{invar}^{\mathcal{I}}$ between both spaces is hereby given by

$$\begin{aligned}
 \Lambda(\mathcal{Q}) : TM &\rightarrow \text{Hom}(\mathcal{A}_{in}, \mathcal{A}_{out}), \\
 v &\mapsto [\Lambda(\mathcal{Q})](v) := \Phi_{r_{TM}}(v)_{*,Hom} \mathcal{Q} r_{TM} Q_{TM}(v).
 \end{aligned} \tag{13.94}$$

Its inverse $\Lambda^{-1} : \mathcal{K}_{invar}^{\mathcal{I}} \rightarrow \mathcal{K}_{quot}^{\mathcal{I}}$ is given by the restriction of invariant kernel fields to the bundle quotient representatives $r_{TM}(\mathcal{I}\setminus TM) \subseteq TM$:

$$\begin{aligned}
 \Lambda^{-1}(\mathcal{K}) : r_{TM}(\mathcal{I}\setminus TM) &\rightarrow r_{Hom}(\mathcal{I}\setminus Hom), \\
 w &\mapsto [\Lambda^{-1}(\mathcal{K})](w) := \mathcal{K}|_{r_{TM}(\mathcal{I}\setminus TM)}(w)
 \end{aligned} \tag{13.95}$$

¹³This definition of $\mathcal{K}_{quot}^{\mathcal{I}}$ is in cyclic dependency with that of Λ in Eq. (13.94). This could be avoided on the expense of 1) having to define spaces $\widetilde{\mathcal{K}_{quot}^{\mathcal{I}}}$ and $\widetilde{\mathcal{K}_{invar}^{\mathcal{I}}}$ without smoothness requirements, in terms of which 2) $\widetilde{\Lambda} : \widetilde{\mathcal{K}_{quot}^{\mathcal{I}}} \rightarrow \widetilde{\mathcal{K}_{invar}^{\mathcal{I}}}$ could be defined, which would 3) allow to demand the smoothness requirements in $\mathcal{K}_{quot}^{\mathcal{I}}$ in terms of $\widetilde{\Lambda}$.

Proof: In order to prove that $\Lambda : \mathcal{K}_{\text{quot}}^{\mathcal{I}} \rightarrow \mathcal{K}_{\text{invar}}^{\mathcal{I}}$ is an isomorphism, we need to show that

1) Λ^{-1} is indeed an inverse of Λ , that 2) the defining properties of $\mathcal{K}_{\text{invar}}^{\mathcal{I}}$ and $\mathcal{K}_{\text{quot}}^{\mathcal{I}}$ are satisfied after lifting and restricting and that 3) the constructions do not depend on arbitrary choices. In order to not overload this section, we outsource the full proof to Appendix K.1. The individual steps of the proof are listed below:

$$1a) \quad \Lambda \circ \Lambda^{-1} = \text{id}_{\mathcal{K}_{\text{invar}}^{\mathcal{I}}}, \text{ that is, } \Lambda^{-1} \text{ is a right inverse of } \Lambda$$

$$1b) \quad \Lambda^{-1} \circ \Lambda = \text{id}_{\mathcal{K}_{\text{quot}}^{\mathcal{I}}}, \text{ that is, } \Lambda^{-1} \text{ is a left inverse of } \Lambda$$

$$2a) \quad \pi_{\text{Hom}} \circ \Lambda(\mathcal{Q}) = \pi_{TM} \text{ for any } \mathcal{Q} \in \mathcal{K}_{\text{quot}}^{\mathcal{I}}, \text{ that is, the lift } \Lambda(\mathcal{Q}) \text{ is a bundle } M\text{-morphism}$$

$$2b) \quad \pi_{\text{Hom}} \circ \Lambda^{-1}(\mathcal{K}) = \pi_{TM} \text{ for any } \mathcal{K} \in \mathcal{K}_{\text{invar}}^{\mathcal{I}}$$

$$2c) \quad \phi_{*,\text{Hom}} \Lambda(\mathcal{Q}) \phi_{*,TM}^{-1} = \Lambda(\mathcal{Q}) \quad \forall \phi \in \mathcal{I}, \text{ that is, } \Lambda(\mathcal{Q}) \text{ satisfies the full isometry invariance (equivariance) constraint}$$

$$2d) \quad \xi_{*,\text{Hom}} [\Lambda^{-1}(\mathcal{K})](w) = [\Lambda^{-1}(\mathcal{K})](w) \quad \forall w \in r_{TM}(\mathcal{I} \setminus TM), \quad \xi \in \text{Stab}_w, \text{ that is, } \Lambda^{-1}(\mathcal{K}) \text{ satisfies the stabilizer constraint}$$

$$3) \quad \text{All constructions and proofs are independent from the particular choice of } \Phi_{r_{TM}}$$

The smoothness of lifted quotient representative kernel fields holds by definition. \square

The arbitrariness in the choice of section r_{TM} allows for different, isomorphic quotient kernel fields, expressed on different bundle quotient representatives.

Instead of maximally restricting the kernel field to bundle orbit representatives in $r_{TM}(\mathcal{I} \setminus M)$, one could choose to restrict the description to $\pi_{TM}^{-1}(r_M(\mathcal{I} \setminus M))$ only, i.e. to *complete tangent spaces* $T_r M$ for any $r \in r_M(\mathcal{I} \setminus M)$. In Fig. (13.7), this would correspond to modeling the (reflection symmetric) kernel on the full tangent space shown in the front instead of only one half. The requirements on such restricted kernels can be derived by following the same rationale as before and results in the constraint in Eq. (13.85). We obtain a similar theorem to Theorem (13.3.1):

Theorem 13.3.2 (Manifold quotient representative kernel fields). *The space of isometry invariant kernel fields $\mathcal{K}_{\text{invar}}^{\mathcal{I}}$ from Def. 13.2.3 is isomorphic to the space $\widehat{\mathcal{K}}_{\text{quot}}^{\mathcal{I}}$ of (manifold) stabilizer subgroup constrained kernel fields on the tangent spaces over manifold quotient representatives $r_M(\mathcal{I} \setminus M)$, defined as:*

$$\begin{aligned} \widehat{\mathcal{K}}_{\text{quot}}^{\mathcal{I}} := & \left\{ \widehat{\mathcal{Q}} : \pi_{TM}^{-1}(r_M(\mathcal{I} \setminus M)) \rightarrow \pi_{\text{Hom}}^{-1}(r_M(\mathcal{I} \setminus M)) \mid \pi_{\text{Hom}} \circ \widehat{\mathcal{Q}} = \pi_{TM}, \quad \widehat{\Lambda}(\widehat{\mathcal{Q}}) \text{ smooth,} \right. \\ & \left. \xi_{*,\text{Hom}} \widehat{\mathcal{Q}}|_r \xi_{*,TM}^{-1} = \widehat{\mathcal{Q}}|_r \quad \forall r \in r_M(\mathcal{I} \setminus M), \quad \xi \in \text{Stab}_r \right\} \end{aligned} \quad (13.96)$$

The lifting isomorphism $\widehat{\Lambda} : \widehat{\mathcal{K}}_{\text{quot}}^{\mathcal{I}} \rightarrow \mathcal{K}_{\text{invar}}^{\mathcal{I}}$ is in terms of Λ and a restriction defined as

$$\widehat{\Lambda} := \Lambda \circ (\cdot)|_{r_{TM}(\mathcal{I} \setminus TM)} \quad (13.97)$$

and therefore essentially agrees with Λ :

$$\begin{aligned}\widehat{\Lambda}(\widehat{\mathcal{Q}}) : TM &\rightarrow \text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}}), \\ v &\mapsto [\widehat{\Lambda}(\widehat{\mathcal{Q}})](v) := \Phi_{r_{TM}}(v)_{*,\text{Hom}} \widehat{\mathcal{Q}} r_{TM} Q_{TM}(v)\end{aligned}\quad (13.98)$$

Its inverse $\widehat{\Lambda}^{-1} : \mathcal{K}_{\text{invar}}^{\mathcal{I}} \rightarrow \widehat{\mathcal{K}}_{\text{quot}}^{\mathcal{I}}$ is given by the restriction of invariant kernel fields to the tangent spaces over manifold quotient representatives: $\pi_{TM}^{-1}(r_M(\mathcal{I}\setminus M)) \subseteq TM$:

$$\begin{aligned}\widehat{\Lambda}^{-1}(\mathcal{K}) : \pi_{TM}^{-1}(r_M(\mathcal{I}\setminus M)) &\rightarrow \pi_{\text{Hom}}^{-1}(r_M(\mathcal{I}\setminus M)), \\ \widehat{w} &\mapsto [\widehat{\Lambda}^{-1}(\mathcal{K})](\widehat{w}) := \mathcal{K}|_{\pi_{TM}^{-1}(r_M(\mathcal{I}\setminus M))}(\widehat{w})\end{aligned}\quad (13.99)$$

Proof: The proof is essentially analogous to that of Theorem 13.3.1 with the slight difference that the stronger constraint $\xi_{*,\text{Hom}} \widehat{\mathcal{Q}}|_r \xi_{*,TM}^{-1} = \widehat{\mathcal{Q}}|_r \quad \forall r \in r_M(\mathcal{I}\setminus M), \xi \in \text{Stab}_r$ is required. Since it would not add much in addition to what is presented in Appendix K.1, we omit the proof. \square

The following commutative diagram shows the isomorphisms between the three equivalent descriptions of invariant kernel fields:

$$\begin{array}{ccccc} & & \Lambda & & \\ & \swarrow \Omega & & \searrow \widehat{\Lambda} & \\ \mathcal{K}_{\text{quot}}^{\mathcal{I}} & & \widehat{\mathcal{K}}_{\text{quot}}^{\mathcal{I}} & & \mathcal{K}_{\text{invar}}^{\mathcal{I}} \\ \uparrow \Omega^{-1} = (\cdot)|_{r_{TM}(\mathcal{I}\setminus TM)} & & \downarrow \widehat{\Lambda}^{-1} = (\cdot)|_{\pi_{TM}^{-1}(r_M(\mathcal{I}\setminus M))} & & \\ & & \Lambda^{-1} = (\cdot)|_{r_{TM}(\mathcal{I}\setminus TM)} & & \end{array} \quad (13.100)$$

Relation to GM-convolutions: The difference between Isom_{GM} -equivariant GM-convolutions and general Isom_{GM} -equivariant kernel field transforms via Isom_{GM} -invariant kernel fields is that the former share G -steerable kernels over the whole manifold while the latter are only required to share Stab_p -steerable kernels over orbits $\text{Isom}_{GM} \cdot p \in \text{Isom}_{GM} \setminus M$. The requirement to share weights over the whole manifold is not strictly necessary but is – supported by Occam’s razor – likely to be a good inductive bias in practice. It can be viewed as an analog to the assumption that the same physical laws apply throughout the whole universe.

Assume now that M is a *homogeneous space* (Def. B.3.11) with respect to the action of some isometry group $\mathcal{I} \leq \text{Isom}_{GM}$, that is, for any two points $p, q \in M$ there is at least one isometry $\phi \in \mathcal{I}$ that connects both points, i.e. $q = \phi(p)$. In this case there is only one single orbit $\mathcal{I} \cdot p$, which is just M itself, and the stabilizers Stab_p of all points $p \in M$ coincide up to isomorphism. The quotient space $\mathcal{I} \setminus M$ is a singleton which is represented by a single representative point $r = r_M(\mathcal{I} \setminus M)$ in M . By Theorem 13.3.2, the space of \mathcal{I} -invariant kernel fields is equivalently expressed by a kernel field on orbit representatives. Since we have only a single representative point r for homogeneous spaces, the full isometry invariant kernel field is in this case equivalent to a single kernel on $T_r M$. This representative kernel is required to satisfy the stabilizer subgroup constraint in Eq. (13.96). Via the lifting isomorphism $\widehat{\Lambda}$ in Eq. (13.98), the representative kernel is shared over the whole manifold.

This sounds very similar to the definition of GM -convolutions, which share a single, G -steerability constrained kernel over the whole manifold. Theorem 13.3.3 below asserts that there is indeed an equivalence between convolutions and equivariant kernel field transforms on homogeneous spaces. The coordinate free Stab_r -steerability from the stabilizer constraint thereby translates (non-canonically) to the H -steerability of template kernels, where $H \cong \text{Stab}_r$ with $H \leq G$ is an isomorphic representation of Stab_r relative to some coordinatization. One can view the isometry (sub)group $\mathcal{I} \leq \text{Isom}_{GM}$ as a principal Stab_r -bundle over M , whose (non-canonical) embedding into GM gives rise to a H -(sub)structure HM of GM . The sharing of a Stab_r -steerable kernel via the lifting isomorphism, which operates per action of \mathcal{I} , then corresponds exactly to the sharing of an H -steerable kernel over HM . This implies that \mathcal{I} -equivariant kernel fields transforms on homogeneous spaces do indeed correspond to some HM -convolution.

Theorem 13.3.3 (Equivariance on homogeneous M implies convolution). *Let M be a manifold equipped with a G -structure GM . Assume that there is an isometry group $\mathcal{I} \leq \text{Isom}_{GM}$ which acts transitively on M , making it a homogeneous space. Let $r \in M$ be an arbitrary representative point of M and $\text{Stab}_r \leq \mathcal{I}$ its stabilizer. Then*

- 1) *There exists a H -(sub)structure $HM \subseteq GM$ on M with:*
 - $H \cong \text{Stab}_r \leq \mathcal{I}$ is a subgroup of $G \cap \text{O}(d)$
 - HM is an embedding of \mathcal{I} (as principal Stab_r -bundle $\mathcal{I} \rightarrow \mathcal{I}/\text{Stab}_r$) into GM , which is preserved by \mathcal{I} , that is, $\text{Isom}_{HM} = \mathcal{I}$
- 2) *Any \mathcal{I} -equivariant kernel field transform shares a single H -steerable kernel over the whole space M and is equivalent to a HM -convolution with that kernel.*

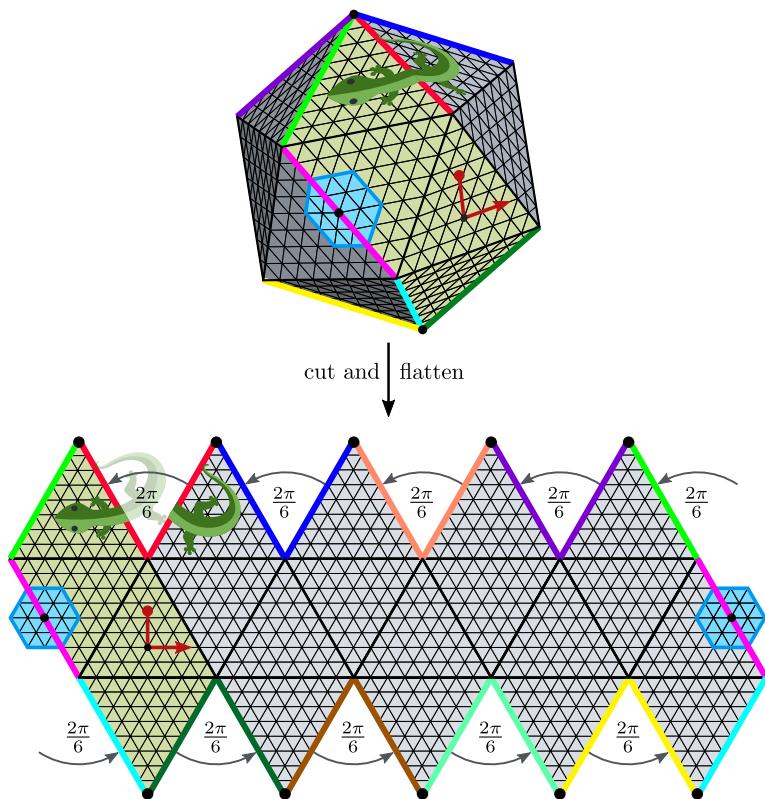
The specific choice of H -structure depends on the chosen isomorphism $H \cong \text{Stab}_r$ but is irrelevant since \mathcal{I} -equivariant kernel field transforms can be equivalently expressed in any such choice.

Proof: The proof is found in Appendix K.2. □

Our definition of isometry equivariant kernel field transforms is on homogeneous spaces essentially equivalent to the *steerable convolutions on homogeneous spaces* as proposed by Cohen et al. [55][56], with $\text{Aff}(G)$ -equivariant Euclidean steerable CNNs from Part I as a special case. The proven equivalence between isometry equivariant kernel field transforms and HM -convolutions on homogeneous spaces therefore asserts that HM -convolutions and steerable convolutions are essentially similar in this case. However, while steerable convolutions are only defined on homogeneous spaces, HM -convolutions generalize to general Riemannian manifolds. More details on convolutions on homogeneous spaces are discussed in the related work Appendix F.

PART IV

APPLICATIONS & LITERATURE REVIEW



Introduction & overview

The formulation of coordinate independent CNNs in terms of associated G -bundles over Riemannian manifolds is quite general and covers a wide range of possible model instantiations. To substantiate this claim, we review a large body of convolutional models from the literature and explain them from the unifying viewpoint of coordinate independent CNNs. Most of the papers in the literature do not explicitly formulate their models in terms of G -structures and associated G -bundles. The *implicitly assumed* G -structures and group representations are therefore *deduced* from the models' weight sharing patterns, kernel symmetries and equivariance properties; see for instance Fig. 16.4. Table 14.1 on page 273 summarizes the resulting taxonomy of coordinate independent CNNs. The following chapters discuss the covered models and their properties in detail.

Chapter 15 starts with a general discussion of the design choices and implementation aspects of coordinate independent CNNs and provides an overview of the models covered in the following chapters. Chapter 15 revisits the $\text{Aff}(G)$ -equivariant (steerable) convolutions on Euclidean spaces \mathbb{E}_d from Part I in the differential geometric formulation. These models rely on $\text{Aff}(G)$ -invariant G -structures as shown in Fig. 15.3. Chapter 16 covers models that operate on punctured Euclidean spaces $\mathbb{E}_d \setminus \{0\}$; see Figs. 16.1, 16.3 or 16.4. They are equivariant w.r.t. rotations around the chosen origin $\{0\}$ but are not translation equivariant. Spherical and icosahedral CNNs are discussed in Chapter 17. Most of these models assume the G -structures that are visualized in Figs. 17.2a and 17.2b and are therefore $\text{SO}(3)$ or $\text{SO}(2)$ -equivariant, respectively. Chapter 18 reviews GM -convolutions on general surfaces, which are mostly discretized as meshes.

Design choices and overview

A coordinate independent CNN is *in theory* fully specified by

- 1) a choice of *Riemannian manifold* (M, η)
- 2) its *G-structure GM*,
- 3) a *G-compatible connection* which specifies feature transporters $\mathcal{P}_{A,\gamma}$,
- 4) the *field types* or *G-representations* ρ of each feature space, and
- 5) a choice of *G-equivariant nonlinearities*.

The *geodesics*, *exponential* and *logarithmic maps* follow from the canonical Levi-Civita connection on M .¹ The *isometry group* Isom_{GM} w.r.t. which the network is equivariant follows from the metric η and the *G-structure GM*. All *kernel spaces* $\mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^G$ are determined by the group representations of the feature spaces between which they map. *Weight sharing* is performed by placing a *G-steerable template kernel* relative to an arbitrary *G-frame* in $G_p M$ for each point $p \in M$.

In practice, the user is faced with additional design questions, for instance concerning the discretization of the geometry, the encoding of feature fields, numerical algorithms for computing geodesics and transporters, etc. This chapter gives a high level overview of all relevant design choices. More specific details are found in the following Chapters 15, 16, 17 and 18.

Discretizations of manifolds and feature fields: The implementations differ in their representation of the manifolds and sampling of the feature fields.

Euclidean spaces \mathbb{E}_d admit regular pixels grids, for instance \mathbb{Z}^d or the hexagonal grid [125]. More generally, locally regular grids are suitable for locally flat manifolds like the Möbius strip and the icosahedron; see Figs. 10.3 and 17.7. Feature fields on Euclidean spaces may furthermore be sampled on a non-regular point cloud. This is for instance useful when processing atomic environments, where the atom positions serve as sampling locations [301].

An important difference between the two approaches is that regular pixel grids are not equivariant w.r.t. continuous translations in $(\mathbb{R}^d, +)$, but only w.r.t. the subgroup of discrete

¹It might seem strange to compute geodesics and feature transporters based on potentially different connections. When the transporter connection differs from Levi-Civita, this is usually due to the Levi-Civita connection not being *G-compatible* with the chosen *G-structure* when $G < \mathcal{O}(d)$. Some examples are given in the paragraph on *G-compatible* connections below.

translations which preserves the grid, for instance $(\mathbb{Z}^d, +)$. CNNs on regular grids are furthermore usually applying spatial pooling operations which reduce the models' equivariance even further. Specifically, given that the pooling operation has a stride of n pixels, it is equivariant w.r.t. translations in $(n\mathbb{Z}^d, +)$. After L pooling layers in a convolutional network, this implies that the model as a whole is only equivariant w.r.t. translations in $(n^L\mathbb{Z}^d, +)$ – this issue was empirically investigated in [6]. Zhang [348] propose to remedy this issue by replacing stride n pooling layers with stride 1 pooling layers (with the same pooling window size), a low-pass filtering, and an n -pixel subsampling. The additional low-pass filtering between the pooling and subsampling operations prevents aliasing effects, which is shown to make the networks sufficiently more stable under translations which are not elements of $(n\mathbb{Z}^d, +)$.

Curved spaces like the 2-sphere S^2 do in general not admit regular sampling grids. A seemingly obvious discretization is in terms of a regular sampling grid in spherical coordinates (Eq. (17.7) and Fig. 17.3), however, as these coordinates are non-isometric, they oversample the signal towards the poles [351, 295]. Approximately uniform sampling grids on S^2 are the “generalized spiral set” [58] or the icospherical grid [139, 148]. Alternatively, feature fields may be discretized in the spectral domain. For the sphere, this is done via an expansion in terms of spherical harmonics for scalar fields, spin-weighted spherical harmonics for irrep fields or Wigner D-matrices for general feature fields [83, 86, 54, 163].

General surfaces are most commonly represented by triangle meshes; see Section 18.1.2. Feature fields can then be sampled on the mesh vertices, edges or faces [65]. A higher resolution of the feature fields can be achieved by encoding them via texture maps [180, 131]. Alternatively, surfaces may be represented as point clouds [294, 141].

G-structures GM and structure groups G : The specific choice of G -structure to be respected by the network depends on the learning task and the topology of M (if continuity or smoothness of the convolution is demanded). In general, M comes equipped with an $O(d)$ -structure, i.e. a bundle of orthonormal reference frames with respect to the given Riemannian metric. A *lift* to structure groups G with $O(d) < G \leq GL(d)$ is uniquely determined by G -valued gauge transformations of orthonormal frames. *Reductions* of the structure group to $G < O(d)$ are, in contrast, not necessarily unique, and encode additional geometric information. For instance, a reduction to $G = SO(d)$ requires an orientation on the manifold.² The following chapters discuss further (mostly implicitly made) choices of G -structures found in the literature; see for instance Figs. 15.3, 16.1, 16.2, 16.3, 16.4, 17.2, or 17.6. They are either determined by a demand for the equivariance under the isometry group Isom_{GM} , canonically given on the manifold or, specifically for $\{e\}$ -structures, algorithmically fixed via some heuristic. Recall that $\{e\}$ -structures are on non-parallelizable manifolds (by definition) necessarily discontinuous.

The most commonly encountered structure groups in the literature are the following:

- *trivial* group $\{e\}$, corresponding to non-coordinate independent CNNs with unconstrained kernels
- *reflection* group $\mathcal{R} \cong \mathbb{Z}/2\mathbb{Z}$, flipping the first frame axis
- *special orthogonal* groups $SO(d)$ (continuous rotations)

²For a single, connected manifold, this choice is arbitrary as long as the kernel initialization is symmetric w.r.t. both orientations. In this case the network will simply learn reflected kernels for different orientations. When considering a dataset consisting of multiple manifolds, their relative orientation is relevant for a correct generalization.

- *orthogonal groups* $O(d)$ (continuous rotations and reflections)
- *scaling group* $\mathcal{S} \cong (\mathbb{R}_{\geq 0}, *)$,

Since the last three groups are continuous Lie groups, they are in numerical implementations sometimes approximated by finite subgroups. For instance, $SO(2)$ and $O(2)$ are often modeled by *cyclic* groups C_N or *dihedral* groups D_N , while three-dimensional rotations and reflections in $O(3)$ can be approximated by *polyhedral* groups (symmetry groups of Platonic solids, e.g. the icosahedron). To reduce the complexity of the classification of models in Table 14.1 we chose to not distinguish between the continuous symmetries and their approximations by finite subgroups. We will, however, state such approximations in our detailed discussion of the models in the following chapters.

G-compatible connections: All of the models consider either the canonical *Levi-Civita* connection on M or the unique *trivial connection* which is induced by an $\{e\}$ -structure. The choice of connection becomes irrelevant (thus unspecified) for networks which operate solely on *scalar fields*, whose transport is always trivial.

More specifically, all Euclidean CNNs from Chapter 15 use Levi-Civita transporters, which transport vectors such that they remain parallel in the usual sense on Euclidean spaces \mathbb{E}_d ; see Fig. 8.4a. This is possible since the Levi-Civita connection is G -compatible with the models' G -structures (defined in Eq. (15.1) and visualized in Fig. 15.3).³

The models on the punctured Euclidean spaces $\mathbb{E}_d \setminus \{0\}$ from Chapter 16 are either based on $\{e\}$ -structures and/or consider scalar fields. They utilize therefore trivial connections which differ from the canonical Levi-Civita connection on $\mathbb{E}_d \setminus \{0\}$.

All spherical CNNs that rely on the $SO(2)$ -structure in Fig. 17.2a (reviewed in Section 17.2) transport features according to the Levi-Civita connection on S^2 (Fig. 8.4b). Those which operate on the $\{e\}$ -structure in Fig. 17.2b (reviewed in Section 17.3) are again considering a trivial connection since the spherical Levi-Civita connection is incompatible with this $\{e\}$ -structure. The icosahedral CNN with C_6 -structure, Fig. 17.6c, transports features according to the C_6 -compatible icosahedral Levi-Civita connection.⁴

All CNNs on general surfaces that are listed in rows (41-43) of Table 14.1 assume oriented surfaces that are equipped with an $SO(2)$ -structure. They transport features with the $SO(2)$ -compatible Levi-Civita connection of the surfaces. The other surface CNNs are based on $\{e\}$ -structures and/or operate on scalar fields – their feature transport is therefore trivial.

Our Möbius strip convolutions transport features via the Levi-Civita connection, which is compatible with the assumed \mathcal{R} -structure.

Recall that the Levi-Civita connection is uniquely determined by the metric, and is therefore generally isometry invariant; cf. footnote 6 in Section 13.1.4. As trivial $\{e\}$ -compatible connections are uniquely specified by the $\{e\}$ -structure they share its symmetries, that is, they are invariant under the action of $\text{Isom}_{\{e\}M}$. This implies by Theorem 13.2.5 that the GM -convolutions, which are based on these connections, are Isom_{GM} -equivariant.

Transporter pullbacks and alternative projections to $T_p M$: The transporter pullback $\text{Exp}_p^* f$, defined in Def. 12.2.4 and Eq. (9.21), represents a feature field f in a geodesic parametrization on the tangent space $T_p M$. The transportation part of the operation

³In contrast, the Euclidean $\{e\}$ -structure in Fig. 13.3b would be incompatible with the Levi-Civita connection on \mathbb{E}_2 .

⁴Discrete Levi-Civita connections on meshes are discussed in Section 18.1.2 and [60, 62].

is determined by the G -compatible connection. Geodesics – and therefore exponential maps $\exp_p : T_p M \rightarrow M$ – have closed form expressions on Euclidean spaces \mathbb{E}_d and the sphere S^2 . Specifically, the exponential maps on \mathbb{E}_d reduce in Cartesian coordinates to the vector summation in Eq. (15.31), such that Euclidean GM -convolutions reduce to conventional convolutions on \mathbb{R}^d ; see Theorem 15.2.1. Geodesics on S^2 are well known to be given by the great circles of the sphere. If the sphere is viewed as being embedded in \mathbb{R}^3 , the exponential map is explicitly given by Eq. (17.10). The geodesics on general surface meshes are not described by closed form solutions but are computed numerically; see Section 18.1.2. In contrast to the smooth setting, one needs to distinguish between “shortest” and “straightest” geodesics on meshes [231].

The pullback of feature fields into geodesic normal coordinates is not the only way of representing feature fields on the tangent spaces. In the literature on spherical CNNs it is rather common to use gnomonic projections, which are visualized in Fig. 17.2. Theorem 17.3.1 shows that this projection can be viewed as a special case of our more general geodesic parameterization after applying a radial warp to the kernels. The corresponding models are therefore exactly identified as GM -convolutions. Surfaces which are embedded in an ambient space like \mathbb{R}^3 might furthermore rely on various projections in the embedding space; see for instance the last three models that are discussed in Section 18.3. Note that these approaches are truly different from ours, i.e. these three models are not exactly GM -convolutions.

G -representations and nonlinearities: Almost all models consider either of the trivial representation, irreducible representations or regular representations as field types. Exceptions are quotient representations, more general induced representations, tensor product representations and, specifically for $G = SO(3)$, the quaternion representation. Infinite-dimensional representations, in particular regular and quotient representations of Lie groups, are in implementations discretized. This can either happen via Monte Carlo sampling or by falling back to the corresponding representations of finite subgroups as discussed above.

The nonlinearities are required to be equivariant w.r.t. the action of the chosen G -representations. Since scalar fields are G -invariant, they are acted on by usual nonlinearities like ReLU. Feature fields that transform according to permutation representations, most importantly regular representations, are acted on channel-wise. All other field types require custom-tailored nonlinearities – we refer back to Section 6.5 and [322] for a more detailed discussion of specific choices.

G -steerable kernel spaces: GM -convolutions map input fields of type ρ_{in} to output fields of type ρ_{out} by convolving them with G -steerable kernels $K \in \mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^G$. Since the space $\mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^G$ of G -steerable kernels, Def. 12.2.2, is a vector space, it is usually parameterized in terms of a basis $\{K_1, \dots, K_N\}$ of $\mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^G$. Before computing the convolution, the learned kernel $K = \sum_{i=1}^N w_i K_i$ is expanded in this basis, where $\{w_1, \dots, w_N\}$ are real-valued weights to be optimized. Provably complete kernel spaces for the groups $G \leq O(2)$ were implemented in [322, 38] and for $G \leq O(3)$ in [40, 39]. A generalization of the Wigner-Eckart theorem characterizes the kernel space bases for general compact structure groups G [173]. We refer the reader for more details on steerable kernels back to Chapter 5.

In practice, the majority of authors does not use a representation theoretic formulation of feature fields and steerable kernels, but formulate them based on intuition. Specifically, most authors assume a given input field type and propose various convolution operations which are engineered such that the resulting output field transforms in an equivariant (or coordinate

independent) manner.⁵ While these approaches propose certain G -steerable kernels that map between ρ_{in} -fields and ρ_{out} -fields, these kernels do sometimes not span the complete space of possible kernels. This applies for instance to the *MDGCNNs* and *PFCNNs*, which are discussed in Section 18.2.

⁵This is opposed to our approach, which fixes the input and output fields and subsequently asks for the resulting constraint on convolution kernels.

	manifold M	structure group G	global symmetry Aff_{GM} or Isom_{GM}	representation ρ	citation
1	\mathbb{E}_d	$\{e\}$	$(\mathbb{R}^d, +)$	trivial	conventional CNNs [175, 348]
2		\mathcal{S}	$(\mathbb{R}^1, +) \rtimes \mathcal{S}$	regular	[248]
3	\mathbb{E}_1	\mathcal{R}	$(\mathbb{R}^1, +) \rtimes \mathcal{R}$	regular	[193]
4				irreps	[193]
5		\mathcal{R}	$(\mathbb{R}^2, +) \rtimes \mathcal{R}$	regular	[322]
6				irreps	[335, 322]
7		$\text{SO}(2)$	$\text{SE}(2)$	regular	[71, 52, 358, 53, 324, 12, 125, 258, 275, 27, 70, 76] [322, 110, 170, 279, 317, 247, 215, 276, 277, 37, 23] [270, 10, 91, 306, 113, 216, 311, 122, 223, 43, 116]
8				quotients	[53, 322]
9				regular $\xrightarrow{\text{pool}}$ trivial	[52, 195, 322]
10				regular $\xrightarrow{\text{pool}}$ vector	[196, 322]
11	\mathbb{E}_2			trivial	[144, 322]
12				irreps	[322]
13		$\text{O}(2)$	$\text{E}(2)$	regular	[71, 52, 125, 53, 322] [216, 110, 270, 23]
14				quotients	[53]
15				regular $\xrightarrow{\text{pool}}$ trivial	[322]
16				induced $\text{SO}(2)$ -irreps	[322]
17		\mathcal{S}	$(\mathbb{R}^2, +) \rtimes \mathcal{S}$	regular	[334, 281, 10, 359]
18				regular $\xrightarrow{\text{pool}}$ trivial	[107]
19		$\text{SO}(2) \times \mathcal{S}$	$(\mathbb{R}^2, +) \rtimes (\text{SO}(2) \times \mathcal{S})$	regular	[349]
20				irreps	[323, 301, 211, 161, 3, 184]
21		$\text{SO}(3)$	$\text{SE}(3)$	quaternion	[345]
22				regular	[91, 329, 333]
23	\mathbb{E}_3			regular $\xrightarrow{\text{pool}}$ trivial	[4]
24		$\text{O}(3)$	$\text{E}(3)$	irreps	[8]
25				regular	[329]

26		$O(3)$	$E(3)$	quotient $O(3)/O(2)$	[136]
27				irrep $\xrightarrow{\text{norm}}$ trivial	[233]
28	\mathbb{E}_3	C_4	$(\mathbb{R}^3, +) \rtimes C_4$	regular	[289]
29		D_4	$(\mathbb{R}^3, +) \rtimes D_4$	regular	[289]
30	$\mathbb{E}_{d-1,1}$	$SO(d-1, 1)$	$(\mathbb{R}^d, +) \rtimes SO(d-1, 1)$	irreps	[274]
31	$\mathbb{E}_2 \setminus \{0\}$	$\{e\}$	$SO(2)$	trivial	[47, 91]
32			$SO(2) \times \mathcal{S}$		[84, 91]
33	$\mathbb{E}_3 \setminus \{0\}$	$O(2)$	$O(3)$	trivial	[238]
34		$\{e\}$	$\{e\}$	trivial	[20]
35		$SO(2)$	$SO(3)$	irreps	[163, 86]
36	S^2			regular	[54, 148]
37		$O(2)$	$O(3)$	trivial	[83, 228, 338]
38	$S^2 \setminus \text{poles}$	$\{e\}$	$SO(2)$	trivial	[58, 295, 351, 203, 139, 287, 288, 77, 178]
39	icosahedron	C_6	$I (\approx SO(3))$	regular	[57]
40	ico \ poles	$\{e\}$	$C_5 (\approx SO(2))$	trivial	[346, 187]
41				irreps	[327]
42	surface ($d=2$)	$SO(2)$	$\text{Isom}_+(M)$	regular	[232, 293, 339, 67]
43	(e.g. meshes)			regular $\xrightarrow{\text{pool}}$ trivial	[204, 205, 217, 293]
44		D_4	$\text{Isom}_{D_4 M}$	trivial	[131]
45		$\{e\}$	$\text{Isom}_{\{e\} M}$	trivial	[217, 259, 140, 294, 180]
46	Möbius strip	\mathcal{R}	$SO(2)$	irreps	Chapter 10
47				regular	

Table 14.1: Classification of convolutional networks in the literature from the viewpoint of coordinate independent CNNs. Bold lines separate different geometries. The affine group equivariant convolutions on Euclidean spaces \mathbb{E}_d (rows 1-30) are reviewed in Chapter 15. Chapter 16 discusses *GM*-convolutions on punctured Euclidean spaces $\mathbb{E}_d \setminus \{0\} \cong S^{d-1} \times \mathbb{R}_{>0}$ (rows 31-34). Details on spherical CNNs (rows 35-40) are found in Chapter 17. The models in rows (41-45) operate on general surfaces, mostly represented by triangle meshes; see Chapter 18. The last two lines list our Möbius convolutions from Chapter 10. $(\mathbb{R}^d, +)$, \mathcal{R} and \mathcal{S} denote the translation, reflection and scaling group, respectively, while C_N and D_N are cyclic and dihedral groups. Infinite-dimensional representations are in implementations discretized or sampled. For instance, the regular representations of $SO(2)$ or $O(2)$ are typically approximated by the regular representations of cyclic or dihedral groups C_N or D_N .

Euclidean coordinate independent CNNs

This chapter considers equivariant convolutions on Euclidean spaces, which are undoubtedly of greatest practical relevance [166]. Convolutional networks on Euclidean spaces are applied for analyzing planar and volumetric images, audio signals, videos, physical events in (pseudo-Euclidean) Minkowski spacetime or planar environments in reinforcement learning. The prototypical convolutional model architecture – both on Euclidean spaces and in general – is the conventional translation equivariant CNN by LeCun et al. [175], which was covered in Chapter 3. These models were in Chapter 4 generalized to *Euclidean steerable CNNs*, which are equivariant under actions of affine symmetry groups. The current chapter demonstrates that both conventional and steerable CNNs are special cases of coordinate independent CNNs on Euclidean spaces, equipped with $\text{Aff}(G)$ -invariant G -structures as visualized in Fig. 15.3.

A major difference in the formulation of steerable and coordinate independent CNNs is that they focus on *global affine group actions* and *local gauge transformations*, respectively. That steerable CNNs can be described as coordinate independent CNNs implies that they are actually also locally gauge equivariant, despite not deliberately being designed for it. That this is the case was intuitively clear since steerable kernels are G -equivariant w.r.t. transformations of their field of view, as shown in Fig. 4.6, but could not be proven in the non-gauge theoretic framework of steerable CNNs.

The equivariance of coordinate independent CNNs under global *isometries*, visualized for Euclidean spaces in Fig. 15.1, was proven in Theorem 13.2.5. Theorem 15.2.2 in Section 15.2 below asserts that the stronger statement of *affine group equivariance* can be made on Euclidean spaces. The underlying reason for this result is that the geodesics and Levi-Civita transporters are on Euclidean spaces not only preserved by isometries, but also by general affine transformations.

Convolutions on Euclidean spaces are classically formulated *in coordinates* (Fig. 15.2, left and right), that is, considering *Euclidean vector spaces* \mathbb{R}^d instead of more general (coordinate free) *Euclidean affine spaces* \mathbb{E}_d (Fig. 15.2, middle). An advantage of formulating convolutions this way is that \mathbb{R}^d comes with all mathematical structure that is required for the definitions. However, \mathbb{R}^d is equipped with an *excess of structure*, for instance a choice of origin or its canonical $\{e\}$ -structure. By designing neural networks to be equivariant, their inference is *post-hoc made independent from this structure*. Specifically, translation equivariance equalizes the particular choice of origin, while G -steerability guarantees that the frames of the canonical $\{e\}$ -structure are just a specific choice of gauge fixing among structurally equivalent G -transformed frames. Steerable CNNs could therefore be viewed as

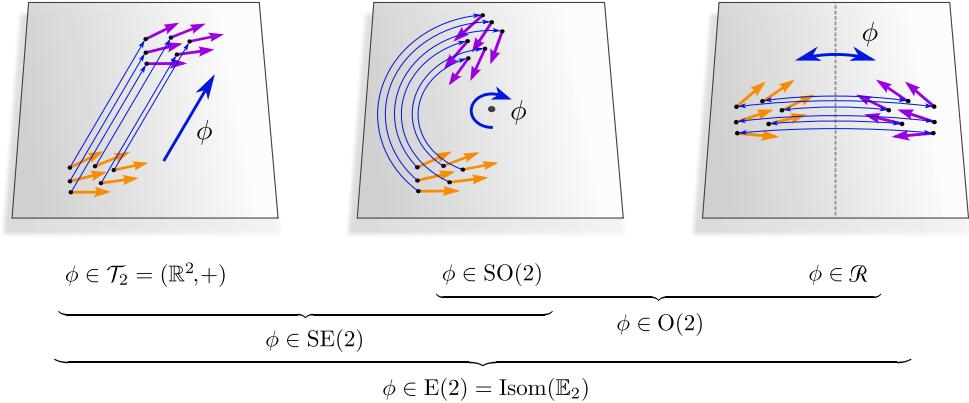


Figure 15.1: Visualization of the full isometry group $\text{Isom}(\mathbb{E}_d) = \text{E}(d) = (\mathbb{R}^d, +) \rtimes \text{O}(d)$ of Euclidean spaces \mathbb{E}_d for $d = 2$. It contains subgroups of translations $(\mathbb{R}^d, +)$, rotations $\text{SO}(d)$ and reflections \mathcal{R} . Rotations and reflections form the orthogonal group $\text{O}(d) = \text{SO}(d) \rtimes \mathcal{R}$ while translations and rotations form the special Euclidean group $\text{SE}(d) = (\mathbb{R}^d, +) \rtimes \text{SO}(d)$. The models in this chapter are not only isometry equivariant, but more generally under affine groups $\text{Aff}(G) = (\mathbb{R}^d, +) \rtimes G$.

operating on G -structures that are (canonical) G -lifts

$$GM = \{e\}M \triangleleft G := \left\{ [e_i]_{i=1}^d \triangleleft g \mid [e_i]_{i=1}^d \in \{e\}M, g \in G \right\} \quad (15.1)$$

of the canonical $\{e\}$ -structure $\{e\}M$ of $M = \mathbb{R}^d$. Intuitively, these lifted G -structures are defined by augmenting every canonical reference frame in $\{e\}M$ with any other G -related frame (its right G -orbit in FM); see Fig. 15.3.

An alternative and more clean approach is to not include the undesired excess structure of \mathbb{R}^d in the first place, but to model convolutions directly on Euclidean affine spaces \mathbb{E}_d . These come naturally with an *affine and metric structure* – if more geometric structure is desired, it can be added by specifying an *atlas of affine charts* $x^A : \mathbb{E}_d \rightarrow \mathbb{R}^d$, from which a G -structure is induced. This viewpoint implies that *steerable convolutions* on \mathbb{R}^d are just the *coordinate representations of* (affine group equivariant) GM -convolutions on \mathbb{E}_d , which is formally proven in Theorem 15.2.1. The particular choice of chart corresponds in an implementation on the (arbitrary) choice of pixel grid.

The main objective of this chapter is to recover the Euclidean steerable CNNs from Chapter 4 from the differential geometric formulation of coordinate independent CNNs. Section 15.1 discusses thereby the affine geometry of \mathbb{E}_d and explains how atlases of affine charts with transition maps in $\text{Aff}(G)$ (Fig. 15.2) induce $\text{Aff}(G)$ -invariant G -structures (Fig. 15.3). Section 15.2 considers GM -convolutions on these G -structures and proves their global $\text{Aff}(G)$ -equivariance. Specific instantiations of such models in the literature, listed in rows (1-30) of Table 14.1, are discussed in Section 15.3.

15.1 Affine geometry of Euclidean spaces \mathbb{E}_d

Before discussing coordinate free convolutions on Euclidean spaces, we need to understand the underlying geometry. Euclidean spaces \mathbb{E}_d are by definition *affine* spaces, that is, they

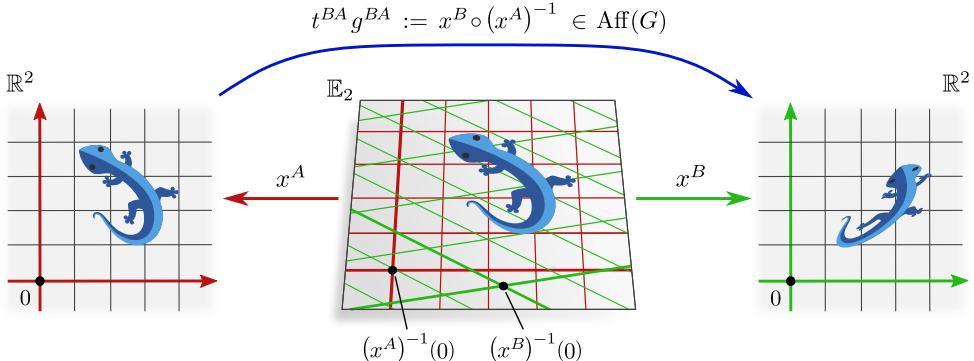


Figure 15.2: Visualization of affine charts $x^X : \mathbb{E}_d \rightarrow \mathbb{R}^d$, which assign global coordinates to Euclidean spaces. Both \mathbb{E}_d and \mathbb{R}^d are affine spaces, such that one can demand the charts to be affine maps, which preserve collinearity and ratios of distances. We define an $\text{Aff}(G)$ -atlas $\mathcal{A}_{\mathbb{E}_d}^{\text{Aff}(G)}$ as consisting of charts that are related by transition functions $t^{BA} g^{BA} := x^B \circ (x^A)^{-1}$ that are elements in $\text{Aff}(G)$. Charts in an $\text{Aff}(G)$ -atlas differ at most in their choice of origin $(x^X)^{-1}(0)$ and a G -transformation. A choice of an $\text{Aff}(G)$ -atlas, consisting of charts x^X , induces a G -atlas \mathcal{A}^G of gauges \hat{dx}^X . The corresponding G -structure GM , which is in Fig. 15.3 exemplified for different groups G , is invariant under the action of $\text{Aff}(G)$. Theorem 15.2.2 proves that GM -convolutions on such G -structures are $\text{Aff}(G)$ -equivariant.

(Lizards adapted under the Creative Commons Attribution 4.0 International license by courtesy of Twitter.)

come with an associated vector space of dimension d , which defines *translations* on \mathbb{E}_d . In addition to being affine spaces, Euclidean spaces are endowed with an *Euclidean metric* (distance function). This distance function corresponds to a Riemannian metric η , i.e. an $O(d)$ -structure OM on the (Riemannian) manifold $M = \mathbb{E}_d$. This metric has the property that its curvature vanishes everywhere, that is, \mathbb{E}_d is globally flat.

A standard example for Euclidean spaces are the *vector spaces* \mathbb{R}^d , however, general Euclidean spaces consider less structure. In particular, they do not come with a vector space structure and do thus not have a preferred origin. Furthermore, they are in general not equipped with Cartesian coordinates. We will therefore start with bare Euclidean spaces \mathbb{E}_d and discuss how the relevant geometric structure is added to them. One could in principle consider any G -structure, however, we are specifically interested in those G -structures that recover the classical steerable CNNs from the Chapter 4, which explain all models in rows (1-30) of Table 14.1. Such $\text{Aff}(G)$ -invariant G -structures are induced from $\text{Aff}(G)$ -atlases, consisting of charts of \mathbb{E}_d whose transition functions take values in $\text{Aff}(G)$; see Fig. 15.2. More infos about the relation between coordinate charts and gauges can be found in Appendix C.

15.1.1 Affine charts and $\text{Aff}(G)$ -atlases

A Euclidean space \mathbb{E}_d of dimension d is homeomorphic to \mathbb{R}^d , and admits therefore global charts

$$x^A : \mathbb{E}_d \rightarrow \mathbb{R}^d. \quad (15.2)$$

In the following we will always require these charts to be affine maps, i.e. isomorphisms of affine spaces, which preserve collinearity (i.e. they map straight lines to straight lines) and

ratios of distances. Since compositions of affine maps are affine maps, it follows that the chart transition functions

$$x^B \circ (x^A)^{-1} : \mathbb{R}^d \rightarrow \mathbb{R}^d \quad (15.3)$$

are affine transformations of \mathbb{R}^d , i.e. elements in $\text{Aff}(\text{GL}(d))$. The transition functions decompose therefore uniquely into a translation $t^{BA} \in (\mathbb{R}^d, +)$ and an element $g^{BA} \in \text{GL}(d)$:

$$t^{BA} g^{BA} := x^B \circ (x^A)^{-1} \quad (15.4)$$

The notation g^{BA} is hereby not accidental as these group elements agree with the gauge transformations that are induced by chart transitions, which is proven in Theorem 15.1.2 below.

Given a choice of affine group $\text{Aff}(G)$, we define $\text{Aff}(G)$ -atlases of \mathbb{E}_d as those atlases of global charts from \mathbb{E}_d to \mathbb{R}^d , whose chart transition functions take values in $\text{Aff}(G)$:

Definition 15.1.1 ($\text{Aff}(G)$ -atlas of Euclidean space). Let \mathfrak{X} be an index set labeling charts and, for any $X \in \mathfrak{X}$, let $x^X : \mathbb{E}_d \rightarrow \mathbb{R}^d$ be a global affine chart of \mathbb{E}_d . The atlas

$$\mathcal{A}_{\mathbb{E}_d}^{\text{Aff}(G)} = \{(\mathbb{E}_d, x^X) \mid X \in \mathfrak{X}\} \quad (15.5)$$

is said to be an $\text{Aff}(G)$ -atlas if all of its chart transition functions take values in $\text{Aff}(G)$, that is, if

$$x^B \circ (x^A)^{-1} \in \text{Aff}(G) \quad \forall A, B \in \mathfrak{X}. \quad (15.6)$$

Fig. 15.2 visualizes affine charts and the $\text{Aff}(G)$ -valued chart transition maps between them.

15.1.2 Induced G -atlases and G -structures

Any global coordinate chart $x^A : \mathbb{E}_d \rightarrow \mathbb{R}^d$ induces a global gauge, which is pointwise given by the chart gradients

$$\psi_{TM,p}^A := \hat{dx}_p^A : T_p M \rightarrow \mathbb{R}^d, \quad (15.7)$$

see Eq. (C.30) in Appendix C.3 and Table C.1. An atlas of charts corresponds therefore to an atlas of gauges. In particular, given that the charts form an $\text{Aff}(G)$ -atlas, it is guaranteed that the gauge transformations are G -valued, that is, that the induced gauges form a G -atlas:

Theorem 15.1.2 ($\text{Aff}(G)$ -atlases of charts induce G -atlases of gauges).

Let $\mathcal{A}_{\mathbb{E}_d}^{\text{Aff}(G)} = \{(\mathbb{E}_d, x^X) \mid X \in \mathfrak{X}\}$ be an $\text{Aff}(G)$ -atlas of charts. The induced atlas of gauges

$$\mathcal{A}^G = \{(\mathbb{E}_d, \hat{dx}^X) \mid X \in \mathfrak{X}\} \quad (15.8)$$

is then guaranteed to be a G -atlas. In particular, if the chart transition maps are given by $x^B \circ (x^A)^{-1} = t^{BA} g^{BA}$, the transition maps between gauges are at any point $p \in \mathbb{E}_d$ given by $g_p^{BA} = g^{BA} \in G$.

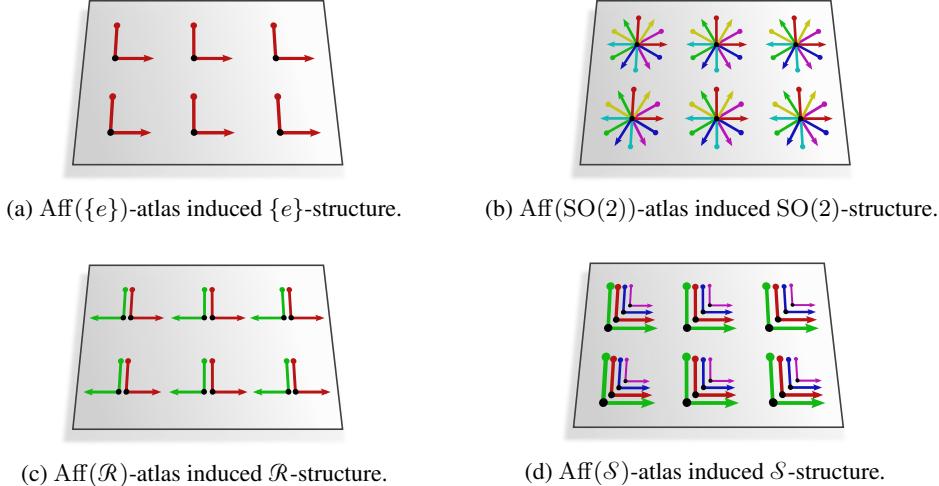


Figure 15.3: Visualization of different G -structures GM on Euclidean spaces $M = \mathbb{E}_2$ which are induced by an $\text{Aff}(G)$ -atlas of charts (Def. 15.1.1). Fig. 15.3a shows the translation invariant $\{e\}M$ that corresponds to conventional Euclidean convolutions. The other three G -structures correspond to non-trivial G -steerable CNNs as defined in Chapter 4. They generalize locally over all poses that are related by the specific set of reference frames in $G_p M$. As the G -structures are $\text{Aff}(G)$ -invariant (implied by Theorem 15.1.6), the G -steerable convolutions are globally equivariant w.r.t. $\text{Aff}(G)$ (Theorem 15.2.2). Instead of defining the G -structures via an $\text{Aff}(G)$ -atlas of charts, one could define them via a G -lift $GM := \{e\}M \triangleleft G$ of the canonical $\{e\}$ -structure of \mathbb{R}^d (Eq. (15.1)), which augments the frames in the $\{e\}$ -structure with all G -related frames.

Proof: The transition functions between chart induced gauges coincide by Eq. (C.45) with the Jacobian of the chart transition maps, that is,

$$g_p^{BA} = \hat{dx}_p^B \circ (\hat{dx}_p^A)^{-1} = \frac{\partial x^B}{\partial x^A} \Big|_{x^A(p)}. \quad (15.9)$$

The last expression is the usual abuse of notation for Jacobians of chart transition maps, which was in Eq. (C.39) in components defined as

$$\frac{\partial x_\mu^B}{\partial x_\nu^A} \Big|_{x^A(p)} := \partial_\nu (x_\mu^B \circ (x^A)^{-1}) \Big|_{x^A(p)}. \quad (15.10)$$

Using that the chart transition maps are given by $x^B \circ (x^A)^{-1} = t^{BA} g^{BA}$, this implies

$$(g_p^{BA})_{\mu\nu} = \partial_\nu (x_\mu^B \circ (x^A)^{-1})(x) \Big|_{x^A(p)} = \partial_\nu (g^{BA} x + t^{BA})_\mu \Big|_{x^A(p)} = g_{\mu\nu}^{BA}, \quad (15.11)$$

that is, that the induced gauge transformations g_p^{BA} are G -valued and agree with g^{BA} (which justifies the notation). As this argument holds for any $p \in \mathbb{E}_d$ and any charts $A, B \in \mathfrak{X}$, this implies that the induced atlas of gauges is a G -atlas. \square

As discussed in Section 11.4, any G -atlas of gauges implies a G -structure GM . According to Eq. (11.56), GM is pointwise determined by

$$G_p M := (\psi_{FM,p}^A)^{-1}(G), \quad (15.12)$$

where the particular choice of gauge $A \in \mathfrak{X}$ is arbitrary. The frames in $G_p M$ are the coordinate bases $\left[\frac{\partial}{\partial x_\mu^A} \Big|_p \right]_{\mu=1}^d = \left[(\hat{dx}_p^A)^{-1}(\epsilon_\mu) \right]_{\mu=1}^d$ and all G -transformations of them. As the maximal $\text{Aff}(G)$ -atlas is by definition $\text{Aff}(G)$ -invariant, the same holds for the induced G -structure (with the action defined via any chart, as clarified and proven below). Fig. 15.3 shows such G -structures for different affine groups. In the next section we prove that the corresponding GM -convolutions are equivariant under the action of $\text{Aff}(G)$.

As it turns out, $GM \xrightarrow{\pi_{GM}} \mathbb{E}_d$ is (non-canonically) isomorphic to $\text{Aff}(G) \xrightarrow{q} \text{Aff}(G)/G \cong \mathbb{R}^d$ as a principal bundle, where

$$q : \text{Aff}(G) \rightarrow \mathbb{R}^d, \quad tg \mapsto t \quad (15.13)$$

is the canonical quotient map of the affine group (after identifying cosets tG with translations t).¹ Non-surprisingly, this principal bundle isomorphism depends on the choice of chart.

Theorem 15.1.3 (Principal bundle isomorphism between $\text{Aff}(G)$ and GM). *Let GM be an $\text{Aff}(G)$ -atlas induced G -structure on \mathbb{E}_d . Then GM is isomorphic to $\text{Aff}(G) \xrightarrow{q} \mathbb{R}^d$ as a principal bundle, i.e. there are isomorphisms*

$$\alpha^A : \text{Aff}(G) \rightarrow GM, \quad tg \mapsto (\psi_{GM,(x^A)^{-1}(t)})^{-1}(g) \quad (15.14)$$

and

$$(x^A)^{-1} : \mathbb{R}^d \rightarrow \mathbb{E}_d \quad (15.15)$$

such that the following diagram commutes:

$$\begin{array}{ccc} \text{Aff}(G) \times G & \xrightarrow{\alpha^A \times \text{id}_G} & GM \times G \\ \downarrow \cdot & & \downarrow \triangleleft \\ \text{Aff}(G) & \xrightarrow{\alpha^A} & GM \\ \downarrow q & & \downarrow \pi_{GM} \\ \mathbb{R}^d & \xrightarrow{(x^A)^{-1}} & \mathbb{E}_d \end{array} \quad (15.16)$$

The inverse of α^A is hereby given by

$$(\alpha^A)^{-1} : GM \rightarrow \text{Aff}(G), \quad [e_i]_{i=1}^d \mapsto tg \quad \text{where} \quad \begin{cases} t = x^A \circ \pi_{GM}([e_i]_{i=1}^d) \\ g = \psi_{GM,\pi_{GM}([e_i]_{i=1}^d)}^A([e_i]_{i=1}^d) \end{cases} \quad (15.17)$$

Note that the isomorphisms are in one-to-one correspondence to the $\text{Aff}(G)$ -compatible charts of the considered atlas.

¹We implicitly employ a canonical isomorphism $\text{Aff}(G)/G \xrightarrow{\sim} \mathbb{R}^d$, $tG \mapsto t$, where t denotes a translation group element in $(\mathbb{R}^d, +)$ on the left-hand side and a vector in \mathbb{R}^d on the right-hand side.

Proof: To prove the statement, we need to show that α^A and $(\alpha^A)^{-1}$ are indeed inverse to each other, that α^A is a bundle map over $(x^A)^{-1}$ and that α^A is right G -equivariant. That $(\alpha^A)^{-1}$ is both a well defined left and right inverse of α^A is easily checked by the reader. That α^A is a bundle map over $(x^A)^{-1}$ means that the bottom square of the diagram commutes. This is seen by observing that $(x^A)^{-1} \circ q(tg) = (x^A)^{-1}(t)$ and $\pi_{GM} \circ \alpha^A(tg) = \pi_{GM} \circ (\psi_{GM, (x^A)^{-1}(t)})^{-1}(g) = (x^A)^{-1}(t)$ agree for any $tg \in \text{Aff}(G)$. The commutativity of the upper square in the diagram, i.e. the right G -equivariance of α^A , follows from the fact that $\alpha^A(tg \cdot \tilde{g}) = (\psi_{GM, (x^A)^{-1}(t)})^{-1}(g\tilde{g}) = (\psi_{GM, (x^A)^{-1}(t)})^{-1}(g) \triangleleft \tilde{g} = \alpha^A(tg) \triangleleft \tilde{g}$ holds for any $tg \in \text{Aff}(G)$ and any $\tilde{g} \in G$. The second step made use of the fact that $\psi_{GM, (x^A)^{-1}(t)}$ is right G -equivariant (Eq. (11.60)), which implies the equivariance of its inverse. Together, these properties show that α^A is a principal bundle isomorphism. \square

15.1.3 Coordinate free affine transformations

As we want to prove the equivariance of GM -convolutions under affine transformations in a coordinate free setting, we need to introduce groups of *affine transformations* of \mathbb{E}_d , instead of \mathbb{R}^d as above. The charts will relate these coordinate free affine groups to the affine groups $\text{Aff}(G)$ of \mathbb{R}^d .

We start with the full group

$$\text{Aff}(\mathbb{E}_d) := \{\phi : \mathbb{E}_d \rightarrow \mathbb{E}_d \mid \phi \text{ is an affine transformation of } \mathbb{E}_d\} \quad (15.18)$$

of affine transformations of a Euclidean space \mathbb{E}_d . It is easy to prove that $\text{Aff}(\mathbb{E}_d)$ is isomorphic to $\text{Aff}(\text{GL}(d))$, with isomorphisms given by $\phi \mapsto x^A \phi (x^A)^{-1}$ for an arbitrary choice of chart x^A . This statement is proven in a more general setting in Theorem 15.1.6 below.

As in the case of isometries, we define subgroups $\text{Aff}_{GM} \leq \text{Aff}(\mathbb{E}_d)$ of G -structure preserving affine transformations:

Definition 15.1.4 (G-structure preserving affine transformations). Let GM be any G -structure on \mathbb{E}_d . We define the corresponding subgroup of G -structure preserving affine transformations as

$$\text{Aff}_{GM} := \{\phi \in \text{Aff}(\mathbb{E}_d) \mid \phi_{*,FM} G_p M = G_{\phi(p)} M \quad \forall p \in \mathbb{E}_d\} \leq \text{Aff}(\mathbb{E}_d). \quad (15.19)$$

Compare this definition to that of Isom_{GM} in Def. 13.1.1. As in the case of Isom_{GM} , the gauge transformations that are induced by affine transformations in Aff_{GM} are guaranteed to be G -valued. This statement is formalized by the following theorem, which is essentially analogous to Theorem 13.1.3:

Theorem 15.1.5 (Aff_{GM} in local trivializations). Let $\phi \in \text{Aff}(\mathbb{E}_d)$ be any isometry of $M = \mathbb{E}_d$. Then the following three statements are equivalent:

1. ϕ is G -structure preserving, that is, $\phi \in \text{Aff}_{GM}$.
2. The affine transformation pullback $\psi_{FM,p}^{\widetilde{A}} \phi_{*,FM}^{-1}$ of any gauge $\psi_{FM,p}^{\widetilde{A}}$ of the G -atlas of FM that defines GM is G -compatible with that G -atlas.

3. The coordinate expression of $\phi_{*,FM}$ relative to any gauges $\psi_{FM,p}^{\tilde{A}}$ and $\psi_{FM,\phi(p)}^A$ from the G -atlas of FM takes values in the structure group, that is, $g_\phi^{A\tilde{A}}(p) := \psi_{FM,\phi(p)}^A \phi_{*,FM} (\psi_{FM,p}^{\tilde{A}})^{-1} = \hat{dx}_{\phi(p)}^A \phi_{*,TM} (\hat{dx}_p^{\tilde{A}})^{-1}$ is G -valued.

Proof: The proof is analogous to that of Theorem 13.1.3. More generally, the statement holds for arbitrary G -structure preserving diffeomorphisms. \square

Affine group action induced gauge transformations describe the transformation of the coordinate expressions of bundle elements, e.g. tangent or feature vector coefficients. The action of the affine transformation ϕ on the manifold \mathbb{E}_d itself can also be described in coordinates \mathbb{R}^d . This is achieved by a left and right multiplication of ϕ with any (affine) chart, which we can w.l.o.g. take to be equal at the source and target location since we are only considering global charts. The resulting coordinate expression $t_\phi^{AA} g_\phi^{AA}$, defined by the following commutative diagram,

$$\begin{array}{ccccc} \mathbb{R}^d & \xleftarrow{x^A} & \mathbb{E}_d & \xrightarrow{\phi} & \mathbb{E}_d & \xrightarrow{x^A} & \mathbb{R}^c \\ & & \underbrace{\qquad\qquad\qquad}_{t_\phi^{AA} g_\phi^{AA}} & & & & \uparrow \end{array} \quad (15.20)$$

is guaranteed to take values in $\text{Aff}(G)$ if ϕ preserves the G -structure.

Theorem 15.1.6 (Aff_{GM} in global affine charts). Let GM be the G -structure induced by some $\text{Aff}(G)$ -atlas and let $x^A : \mathbb{E}_d \rightarrow \mathbb{R}^d$ be a chart of this atlas. The coordinate expression of an element $\phi \in \text{Aff}_{GM}$ relative to x^A is then given by

$$x^A \phi (x^A)^{-1} =: t_\phi^{AA} g_\phi^{AA} \in \text{Aff}(G), \quad (15.21)$$

where

$$t_\phi^{AA} := x^A \phi (x^A)^{-1}(0) \in (\mathbb{R}^d, +) \quad (15.22)$$

$$\text{and} \quad g_\phi^{AA} := \hat{dx}_{\phi(p)}^A \phi_{*,TM} (\hat{dx}_p^A)^{-1} \in G \quad (15.23)$$

The element $g_\phi^{AA} \in G$ in the coordinate expression coincides hereby with the induced gauge transformation $g_\phi^{AA}(p) \in G$ from Theorem 15.1.5 at any point $p \in \mathbb{E}_d$.

Furthermore, the coordinatization map

$$\text{Aff}_{GM} \rightarrow \text{Aff}(G), \quad \phi \mapsto x^A \phi (x^A)^{-1}, \quad (15.24)$$

is a group isomorphism.

Proof: Since x^A and ϕ are affine maps, $x^A \phi (x^A)^{-1} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is an affine transformation of \mathbb{R}^d , i.e. an element of $\text{Aff}(\text{GL}(d))$ (or some subgroup of it). This implies that a first order Taylor expansion of the expression is exact. The application of the coordinate expression to an arbitrary element $x \in \mathbb{R}^d$ can therefore be written in terms of the following Taylor expansion around the origin 0 of \mathbb{R}^d :

$$\begin{aligned} [x^A \phi (x^A)^{-1}](x) &= [x^A \phi (x^A)^{-1}](0) + \frac{\partial}{\partial x'} [x^A \phi (x^A)^{-1}]_{x'=0} \cdot x \\ &= t_\phi^{AA} + g_\phi^{AA} \cdot x \\ &= (t_\phi^{AA} g_\phi^{AA}) x \end{aligned} \quad (15.25)$$

Here we implicitly defined the translation $t_\phi^{AA} \in \mathbb{R}^d$ and the Jacobian $g_\phi^{AA} \in \mathbb{R}^{d \times d}$ and identified them with group elements, which is possible since all involved morphisms are invertible.

That the Jacobian g_ϕ^{AA} agrees with the induced gauge transformation $g_\phi^{AA}(p)$ at an arbitrary point $p \in M$ is shown by rewriting it via Eq. (C.21) in terms of differentials:

$$\begin{aligned} g_\phi^{AA} &= \frac{\partial}{\partial x'} [x^A \phi(x^A)^{-1}]_{x'=x^A(p)} \\ &= \iota_{\mathbb{R}^d} d[x^A \phi(x^A)^{-1}]_{x^A(p)} (\iota_{\mathbb{R}^d})^{-1} \\ &= \iota_{\mathbb{R}^d} dx_{\phi(p)}^A d\phi_p (dx_p^A)^{-1} (\iota_{\mathbb{R}^d})^{-1} \\ &= \hat{dx}_{\phi(p)}^A \phi_{*,TM} (\hat{dx}_p^A)^{-1} \\ &= g_\phi^{AA}(p) \end{aligned} \quad (15.26)$$

In the penultimate step we identified the differential $d\phi$ as an alternative notation for the pushforward $\phi_{*,TM}$ and identified the chart gradients $\hat{dx}^A := \iota_{\mathbb{R}^d} dx^A$ as defined in Eq. (C.30). The index p is dropped in the notation $g_\phi^{AA} = g_\phi^{AA}(p)$ due to its arbitrariness.

That $t_\phi^{AA} g_\phi^{AA}$ is not only an element element of $\text{Aff}(\text{GL}(d))$ but of its subgroup $\text{Aff}(G)$ is clear since Theorem 15.1.5 states that $g_\phi^{AA}(p) \in G$ for any $\phi \in \text{Aff}_{GM}$.

To prove that the coordinatization map $C^A : \text{Aff}_{GM} \rightarrow \text{Aff}(G)$, $\phi \mapsto x^A \phi(x^A)^{-1}$ is indeed a group isomorphism, we need to show that it is 1) a group homomorphism, 2) injective and 3) surjective. That C^A is a group homomorphism follows immediately from its definition since

$$C^A(\phi \tilde{\phi}) = x^A \phi \tilde{\phi}(x^A)^{-1} = x^A \phi(x^A)^{-1} x^A \tilde{\phi}(x^A)^{-1} = C^A(\phi) C^A(\tilde{\phi}) \quad (15.27)$$

holds for any $\phi, \tilde{\phi} \in \text{Aff}_{GM}$. The injectivity of C^A requires that, for any $\phi, \tilde{\phi} \in \text{Aff}_{GM}$, the equality $C^A(\phi) = C^A(\tilde{\phi})$ implies $\phi = \tilde{\phi}$. That this is the case is clear since $C^A(\phi) = C^A(\tilde{\phi})$ is equivalent to $x^A \phi(x^A)^{-1} = x^A \tilde{\phi}(x^A)^{-1}$, which implies the equality of ϕ and $\tilde{\phi}$ since x^A is an isomorphism. Lastly, C^A is surjective if and only if for any $tg \in \text{Aff}(G)$ there exists some $\phi \in \text{Aff}_{GM}$, such that $C^A(\phi) = tg$. As an ansatz, let $\phi = (x^A)^{-1} tg x^A$, such that $C^A(\phi) = tg$. What remains to be shown is that this construction of ϕ is indeed an element of Aff_{GM} . As one can easily check, $g_\phi^{AA} = g \in G$, such that $\phi \in \text{Aff}_{GM}$ follows from Theorem 15.1.5, with which surjectivity holds. Overall, this proves that $C^A : \text{Aff}_{GM} \rightarrow \text{Aff}(G)$ is a group isomorphism if Aff_{GM} is induced by an $\text{Aff}(G)$ -atlas. \square

The isomorphism between Aff_{GM} and $\text{Aff}(G)$ is not unique, as it depends on the particular chart considered. Different choices are related by the inner automorphisms of $\text{Aff}(G)$ since

$$\begin{aligned} C^B(\phi) &= x^B \phi(x^B)^{-1} \\ &= x^B (x^A)^{-1} x^A \phi(x^A)^{-1} x^A (x^B)^{-1} \\ &= (t^{BA} g^{BA}) C^A(\phi) (t^{BA} g^{BA})^{-1}. \end{aligned} \quad (15.28)$$

This concludes our analysis of the Euclidean geometry and $\text{Aff}(G)$ -invariant G -structures that are required for the definition of coordinate free Euclidean convolutions in the next section.

15.2 Affine group equivariant CNNs on Euclidean spaces \mathbb{E}_d

We now turn to investigate Euclidean GM -convolutions on the $\text{Aff}(G)$ -atlas induced G -structures. Section 15.2.1 shows hereby that these GM -convolutions boil down to classical G -steerable convolutions on \mathbb{R}^d when being expressed in a chart. The coordinate chart independence is thereby guaranteed by the steerable convolutions' $\text{Aff}(G)$ -equivariance, which is demonstrated in Section 15.2.2. Section 15.2.3 proves the GM -convolutions' Aff_{GM} -equivariance in the coordinate free setting.

15.2.1 Recovering steerable convolutions on \mathbb{R}^d

GM -convolutions rely crucially on the transporter pullback $\text{Exp}_p^* f$ of feature fields, which in turn depends on parallel transporters and the exponential map. On Euclidean spaces, these operations take a particularly simple form, which we discuss first.

As stated before, Levi-Civita transporters move tangent vectors such over Euclidean spaces that they remain parallel in the usual sense on Euclidean spaces; see Fig. 8.4a. Let $x^A : \mathbb{E}_d \rightarrow \mathbb{R}^d$ be any global chart of an $\text{Aff}(G)$ -atlas. As the induced frame field is “parallel”, the transporters along *any* path γ become trivial when being expressed relative to the induced gauges \hat{dx}_p^A :

$$g_\gamma^{AA} = e \quad \text{for any path } \gamma \quad (15.29)$$

This implies in particular that the feature vector transporters are in this gauge given by identity maps, i.e.

$$\rho(g_\gamma^{AA}) = \text{id}_{\mathbb{R}^c} \quad \text{for any path } \gamma. \quad (15.30)$$

When expressing the exponential map in a chart, it reduces to a summation of the point and vector coordinate expressions in \mathbb{R}^d :

$$x^A(\exp_p v) = x^A(p) + \hat{dx}_p^A(v) \quad (15.31)$$

We furthermore need to express feature fields in coordinates, that is, we pull them via the inverse global chart back from \mathbb{E}_d to \mathbb{R}^d ,

$$F^A := f^A \circ (x^A)^{-1} : \mathbb{R}^d \rightarrow \mathbb{R}^c, \quad (15.32)$$

which is visualized by the following commutative diagram:

$$\begin{array}{ccccc} & & F^A & & \\ & \swarrow & & \searrow & \\ \mathbb{R}^d & \xleftarrow{x^A} & \mathbb{E}_d & \xrightarrow{f^A} & \mathbb{R}^c \end{array} \quad (15.33)$$

With these ingredients at hand, the transporter pullback, Eq. (9.21), of feature fields on Euclidean spaces can in coordinates be expressed as

$$\begin{aligned} [\text{Exp}_p^* f]^A(v) &= \rho(g_{p \leftarrow \exp_p(\hat{dx}_p^A)^{-1}(v^A)}^{AA}) f^A \exp_p((\hat{dx}_p^A)^{-1}(v)) \\ &= f^A(x^A)^{-1} x^A \exp_p((\hat{dx}_p^A)^{-1}(v)) \\ &= F^A(x^A(p) + v). \end{aligned} \quad (15.34)$$

Expressing the GM -convolution on \mathbb{E}_d relative to a chart on \mathbb{R}^d , we find that it reduces to a *steerable correlation*. This correlation is furthermore equivalent to a *steerable convolution* with a *point-inverted kernel*. Note that the terms “convolution” and “correlation” are in deep learning commonly used interchangeably, and that CNNs are usually implemented using correlations instead of convolutions. As the kernels are learned, and their G -steerability is unaffected by point inversion, convolutions and correlations are in practice equivalent.

Theorem 15.2.1 (Euclidean GM -convolutions in coordinates).

Let GM be a G -structure induced by an $\text{Aff}(G)$ -atlas of charts as defined in Section 15.1. When being expressed relative to any global chart $x^A : \mathbb{E}_d \rightarrow \mathbb{R}^d$ of this $\text{Aff}(G)$ -atlas, the GM -convolution takes the form of a steerable correlation $\star_{\mathbb{R}^d}$ on \mathbb{R}^d :

$$F_{\text{out}}^A(x) = [K \star_{\mathbb{R}^d} F_{\text{in}}^A](x) := \int_{\mathbb{R}^d} K(\mathbf{v}) F_{\text{in}}^A(x + \mathbf{v}) d\mathbf{v} \quad (15.35)$$

Let \overleftrightarrow{K} denote a flipped kernel, defined by $\overleftrightarrow{K}(\mathbf{v}) := K(-\mathbf{v})$, which is G -steerable iff K is. The GM -convolutions with a kernel K is then furthermore equivalent to a steerable convolution $*$ with \overleftrightarrow{K} on \mathbb{R}^d :

$$F_{\text{out}}^A(x) = [\overleftrightarrow{K} * F_{\text{in}}^A](x) := \int_{\mathbb{R}^d} \overleftrightarrow{K}(x - y) F_{\text{in}}^A(y) dy \quad (15.36)$$

Proof: The Euclidean correlation is derived by expressing the GM -convolution \star_{GM} in coordinates, Eq. (9.39), and inserting the transporter pullback’s coordinate expression, Eq. (15.34):

$$\begin{aligned} F_{\text{out}}^A(x) &= f_{\text{out}}^A((x^A)^{-1}(x)) = [K \star_{GM} f_{\text{in}}]^A((x^A)^{-1}(x)) \\ &= \int_{\mathbb{R}^d} K(\mathbf{v}) [\text{Exp}_{(x^A)^{-1}(x)}^* f_{\text{in}}]^A(\mathbf{v}) d\mathbf{v} \\ &= \int_{\mathbb{R}^d} K(\mathbf{v}) F_{\text{in}}^A(x + \mathbf{v}) d\mathbf{v} \end{aligned} \quad (15.37)$$

That the steerability of K and \overleftrightarrow{K} imply each other is clear since the G -steerability constraint, Eq. (5.1), needs to hold for *any* $\mathbf{v} \in \mathbb{R}^d$, in particular also $-\mathbf{v}$, and thus point-inverted kernels.

The equivalence of correlations and convolutions up to a kernel reflection is well known and quickly shown:

$$\begin{aligned} [K \star_{\mathbb{R}^d} F_{\text{in}}](x) &= \int_{\mathbb{R}^d} K(\mathbf{v}) F_{\text{in}}(x + \mathbf{v}) d\mathbf{v} = \int_{\mathbb{R}^d} K(y - x) F_{\text{in}}(y) dy \\ &= \int_{\mathbb{R}^d} \overleftrightarrow{K}(x - y) F_{\text{in}}(y) dy = [\overleftrightarrow{K} * F_{\text{in}}](x), \end{aligned}$$

The second step substituted $y = x + \mathbf{v}$. □

That correlations with G -steerable kernels are $\text{Aff}(G)$ -equivariant follows directly from the corresponding steerable convolution’s equivariance. The following calculation asserts this

claim for completeness explicitly:

$$\begin{aligned}
[K \star_{\mathbb{R}^d} (tg \triangleright_{\rho_{\text{in}}} F_{\text{in}})](x) &= [K \star_{\mathbb{R}^d} (\rho_{\text{in}}(g) F_{\text{in}}(tg)^{-1})](x) \\
&= \int_{\mathbb{R}^d} K(v) \rho_{\text{in}}(g) F_{\text{in}}((tg)^{-1}(x + v)) dv \\
&= \int_{\mathbb{R}^d} K(v) \rho_{\text{in}}(g) F_{\text{in}}(g^{-1}(x + v - t)) dv \\
&= \int_{\mathbb{R}^d} K(g\tilde{v}) \rho_{\text{in}}(g) F_{\text{in}}(g^{-1}(x - t) + \tilde{v}) |\det g| d\tilde{v} \\
&= \int_{\mathbb{R}^d} \rho_{\text{out}}(g) K(\tilde{v}) F_{\text{in}}(g^{-1}(x - t) + \tilde{v}) d\tilde{v} \\
&= \rho_{\text{out}}(g) [K \star_{\mathbb{R}^d} F_{\text{in}}](g^{-1}(x - t)) \\
&= tg \triangleright_{\rho_{\text{out}}} [K \star_{\mathbb{R}^d} F_{\text{in}}](x)
\end{aligned} \tag{15.38}$$

$\triangleright_{\rho_{\text{in}}}$ and $\triangleright_{\rho_{\text{out}}}$ denote hereby the induced $\text{Aff}(G)$ -representation actions on feature fields as defined in Eq. (4.5).

15.2.2 Affine chart independence

Before proceeding to our proof of the Euclidean *GM*-convolutions' Aff_{GM} -equivariance in a *coordinate free* setting, we consider its $\text{Aff}(G)$ -chart independence – as we will see, both notions are closely related.

The transformation law of the feature field pullbacks to \mathbb{R}^d when switching between charts follows directly from the transition functions and can from the commutativity of the diagram

$$\begin{array}{ccccc}
\mathbb{R}^d & \xrightarrow{\quad F^A \quad} & & \mathbb{R}^c & \\
t^{BA} g^{BA} \downarrow & \swarrow x^A & \searrow f^A & \downarrow \rho(g^{BA}) & \\
& \mathbb{E}_d & & & \\
\downarrow & \swarrow x^B & \searrow f^B & \downarrow & \\
\mathbb{R}^d & \xrightarrow{\quad F^B \quad} & & \mathbb{R}^c &
\end{array} \tag{15.39}$$

be read off to be given by

$$F^B = \rho(g^{BA}) F^A (t^{BA} g^{BA})^{-1} = (t^{BA} g^{BA}) \triangleright_{\rho} F^A, \tag{15.40}$$

that is, by the induced representation action, Eq. (4.5), of the chart transition maps $t^{BA} g^{BA}$.

Coordinate chart independence means then that the *GM*-convolution's coordinate expressions in different affine charts x^A and x^B imply each other, i.e.

$$\begin{aligned}
K \star_{\mathbb{R}^d} F_{\text{in}}^B &= K \star_{\mathbb{R}^d} (t^{BA} g^{BA} \triangleright_{\rho_{\text{in}}} F_{\text{in}}^A) = t^{BA} g^{BA} \triangleright_{\rho_{\text{out}}} (K \star_{\mathbb{R}^d} F_{\text{in}}^A) \\
&= t^{BA} g^{BA} \triangleright_{\rho_{\text{out}}} F_{\text{out}}^A = F_{\text{out}}^B,
\end{aligned} \tag{15.41}$$

where we leveraged the $\text{Aff}(G)$ -equivariance of the steerable correlation (Eq. (15.38)). The *active* $\text{Aff}(G)$ -equivariance of classical G -steerable correlations or convolutions on \mathbb{R}^d is therefore seen to imply the *passive* $\text{Aff}(G)$ coordinate independence of Euclidean GM -convolutions and vice versa. The two are two sides of the same coin. In addition, one can prove the Aff_{GM} -equivariance of the GM -convolution in the coordinate free setting, which we will do next.

15.2.3 Affine group equivariance

To prove the Aff_{GM} -equivariance of Euclidean GM -convolutions, we first define the transformation law of coordinate free feature fields $f \in \Gamma(\mathcal{A})$ under affine transformations $\phi \in \text{Aff}_{GM}$, denoted as *pushforward*, as:^{2,3}

$$\phi \triangleright f = \phi_{*,\mathcal{A}} f \phi^{-1}. \quad (15.42)$$

The (Levi-Civita) transporter pullback of an affine transformed feature field $\phi \triangleright f$ is relative to an affine chart x^A given by:

$$\begin{aligned} & [\text{Exp}_p^*(\phi \triangleright f)]^A(\mathfrak{v}) \\ & \stackrel{(1)}{=} [\text{Exp}_p^*(\phi_{*,\mathcal{A}} f \phi^{-1})]^A(\mathfrak{v}) \\ & \stackrel{(2)}{=} \underbrace{\rho(g_p^{AA} \leftarrow \exp_p(\hat{dx}_p^A)^{-1}(\mathfrak{v}))}_{= \text{id}_{\mathbb{R}^c}} \psi_{\mathcal{A},p}^A (\phi_{*,\mathcal{A}} f \phi^{-1}) \exp_p((\hat{dx}_p^A)^{-1}(\mathfrak{v})) \\ & \stackrel{(3)}{=} \psi_{\mathcal{A},p}^A \phi_{*,\mathcal{A}} \left[(\psi_{\mathcal{A},\phi^{-1}(p)}^A)^{-1} \psi_{\mathcal{A},\phi^{-1}(p)}^A \right] f \left[(x^A)^{-1} x^A \right] \phi^{-1} \left[(x^A)^{-1} x^A \right] \exp_p((\hat{dx}_p^A)^{-1}(\mathfrak{v})) \\ & \stackrel{(4)}{=} [\psi_{\mathcal{A},p}^A \phi_{*,\mathcal{A}} (\psi_{\mathcal{A},\phi^{-1}(p)}^A)^{-1}] [\psi_{\mathcal{A},\phi^{-1}(p)}^A f (x^A)^{-1}] [x^A \phi^{-1}(x^A)^{-1}] [x^A \exp_p((\hat{dx}_p^A)^{-1}(\mathfrak{v}))] \\ & \stackrel{(5)}{=} \rho(g_\phi^{AA}) F^A (t_\phi^{AA} g_\phi^{AA})^{-1} (x^A(p) + \mathfrak{v}) \\ & \stackrel{(6)}{=} [(t_\phi^{AA} g_\phi^{AA}) \triangleright_\rho F^A] (x^A(p) + \mathfrak{v}) \end{aligned} \quad (15.43)$$

It relates to the transporter pullback of the untransformed field via the induced representation \triangleright_ρ , acting with the coordinate expression $t_\phi^{AA} g_\phi^{AA}$ of ϕ (Eq. (15.21)). The first two steps make use of Eq. (15.42) and the definition of the transporter pullback in coordinates, where $(\phi_{*,\mathcal{A}} f \phi^{-1})^A := \psi_{\mathcal{A},p}^A (\phi_{*,\mathcal{A}} f \phi^{-1})$. To translate all morphisms into the corresponding coordinate expressions, step three inserts identities $\text{id}_{\mathbb{R}^c} = (\psi_{\mathcal{A},\phi^{-1}(p)}^A)^{-1} \psi_{\mathcal{A},\phi^{-1}(p)}^A$ and $\text{id}_{\mathbb{R}^d} = (x^A)^{-1} x^A$, which are in step four rebracketed to clarify which combinations result in the coordinate expressions after step five. Recall for step 5 that, by Theorem 15.1.6, $g_\phi^{AA}(p) = g_\phi^{AA}$ for any p in \mathbb{E}_d . As stated above, the last step identifies the resulting transformation law in coordinates as the action of the induced representation.

With this result we can prove the Aff_{GM} -equivariance of Euclidean convolutions in the coordinate free setting. This generalizes Theorem 13.2.5, proving the isometry equivariance of GM -convolutions for the specific case of Euclidean spaces.

²This is the same definition of pushforwards as for isometries; see Def. 13.1.2.

³Since the feature vector bundle is defined as a G -bundle, i.e. associated to GM , pushforwards can only be defined for the G -structure preserving affine transformations in Aff_{GM} .

Theorem 15.2.2 (Affine equivariance of Euclidean GM-convolutions). *Let GM be a G -structure that is induced by some $\text{Aff}(G)$ -atlas of the Euclidean space $M = \mathbb{E}_d$ and assume feature vectors to be transported according to the Levi-Civita connection on \mathbb{E}_d . The corresponding GM-convolutions is then guaranteed to be equivariant under the action of G -structure preserving affine transformations $\text{Aff}_{GM} \cong \text{Aff}(G)$. In equations, we have for arbitrary feature fields $f_{\text{in}} \in \Gamma(\mathcal{A}_{\text{in}})$ and G -steerable kernels $K \in \mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^G$ that*

$$[K \star (\phi \triangleright f_{\text{in}})] = \phi \triangleright [K \star f_{\text{in}}] \quad \forall \phi \in \text{Aff}_{GM}, \quad (15.44)$$

i.e. that the following diagram commutes for any ϕ in Aff_{GM} :

$$\begin{array}{ccc} \Gamma(\mathcal{A}_{\text{in}}) & \xrightarrow{\phi \triangleright} & \Gamma(\mathcal{A}_{\text{in}}) \\ K \star \downarrow & & \downarrow K \star \\ \Gamma(\mathcal{A}_{\text{out}}) & \xrightarrow{\phi \triangleright} & \Gamma(\mathcal{A}_{\text{out}}) \end{array} \quad (15.45)$$

Proof: Let $x^A : \mathbb{E}_d \rightarrow \mathbb{R}^d$ be any global chart of the considered $\text{Aff}(G)$ -atlas and let $p \in \mathbb{E}_d$. Our proof of the Aff_{GM} -equivariance is then performed by expressing the convolution relative to these coordinates and making use of the $\text{Aff}(G)$ -equivariance of G -steerable correlations (and convolutions) on \mathbb{R}^d from Eq. (15.38):

$$\begin{aligned} & \psi_{\mathcal{A}_{\text{out}}, p}^A [K \star (\phi \triangleright f_{\text{in}})](p) & (15.46) \\ &= \int_{\mathbb{R}^d} K(v) [\text{Exp}_p^*(\phi \triangleright f_{\text{in}})]^A(v) dv & (\text{GM-conv. in coords., Eq. (9.39)}) \\ &= \int_{\mathbb{R}^d} K(v) [(t_\phi^{AA} g_\phi^{AA}) \triangleright_{\rho_{\text{in}}} F_{\text{in}}^A] (x^A(p) + v) dv & (\text{transformed pullback, Eq. (15.43)}) \\ &= [K \star_{\mathbb{R}^d} (t_\phi^{AA} g_\phi^{AA} \triangleright_{\rho_{\text{in}}} F_{\text{in}}^A)](x^A(p)) & (\text{identified correlation } \star_{\mathbb{R}^d} \text{ on } \mathbb{R}^d) \\ &= [t_\phi^{AA} g_\phi^{AA} \triangleright_{\rho_{\text{out}}} (K \star_{\mathbb{R}^d} F_{\text{in}}^A)](x^A(p)) & (\text{Aff}(G)\text{-equivariance of } \star_{\mathbb{R}^d}, \text{ Eq. (15.38)}) \\ &= \rho_{\text{out}}(g_\phi^{AA}) (K \star_{\mathbb{R}^d} F_{\text{in}}^A) ((t_\phi^{AA} g_\phi^{AA})^{-1} x^A(p)) & (\text{induced representation } \triangleright_{\rho_{\text{out}}}, \text{ Eq. (4.5)}) \\ &= \rho_{\text{out}}(g_\phi^{AA}) (K \star_{\mathbb{R}^d} F_{\text{in}}^A) (x^A(\phi^{-1}(p))) & (\text{coordinate expression of } \phi, \text{ Eq. (15.21)}) \\ &= \rho_{\text{out}}(g_\phi^{AA}) \int_{\mathbb{R}^d} K(v) F_{\text{in}}^A(x^A(\phi^{-1}(p)) + v) dv & (\text{expanded correlation } \star_{\mathbb{R}^d} \text{ on } \mathbb{R}^d) \\ &= \rho_{\text{out}}(g_\phi^{AA}) \int_{\mathbb{R}^d} K(v) [\text{Exp}_{\phi^{-1}(p)}^* f_{\text{in}}]^A(v) dv & (\text{Euclidean pullback, Eq. (15.34)}) \\ &= \rho_{\text{out}}(g_\phi^{AA}) \psi_{\mathcal{A}_{\text{out}}, \phi^{-1}(p)}^A [K \star f_{\text{in}}] \phi^{-1}(p) & (\text{GM-conv. in coords., Eq. (9.39)}) \\ &= \psi_{\mathcal{A}_{\text{out}}, p}^A \phi_{*, \mathcal{A}_{\text{out}}} [K \star f_{\text{in}}] \phi^{-1}(p) & (\text{pushforward in coordinates, Eq. (13.39)}) \\ &= \psi_{\mathcal{A}_{\text{out}}, p}^A [\phi \triangleright [K \star f_{\text{in}}]](p) & (\text{Aff}_{GM} \text{ action on fields, Eq. (15.42)}) \end{aligned}$$

The statement follows since $\psi_{\mathcal{A}_{\text{out}}, p}^A$ is an isomorphism. \square

In summary, Euclidean GM -convolutions with $\text{Aff}(G)$ -atlas induced G -structures have the following two properties:

- Aff(G)-coordinate independence:** They are guaranteed to produce equivalent results in any chart of the $\text{Aff}(G)$ -atlas $\mathcal{A}_{\mathbb{E}_d}^{\text{Aff}(G)}$. This property was shown in Eq. (15.41) and is in Fig. 15.2 visualized as the transformation *between charts*.
- Aff $_{GM}$ -equivariance:** As proven in Theorem 15.2.2 they are equivariant under active transformations of feature fields by $\text{Aff}_{GM} \cong \text{Aff}(G)$. In Fig. 15.2, this would correspond to a transformation of the signal on \mathbb{E}_d , which would reflect in an active transformation on its representation relative to the *same chart*.

The proofs of both properties rely ultimately on the active $\text{Aff}(G)$ -equivariance of classical G -steerable convolutions and correlations on \mathbb{R}^d .

15.3 Euclidean CNNs in the literature

All of the models in rows (1-30) of Table 14.1 are $\text{Aff}(G)$ -equivariant GM -convolutions (or steerable convolutions) on Euclidean spaces \mathbb{E}_d , as discussed in this chapter (or Chapter 4). They differ in the dimensionality d of the Euclidean space, the structure group G and thus global symmetry group $\text{Aff}(G)$, the group representations or field types ρ and choices of discretizations. This section discusses the models briefly by grouping them by their field types into irrep models, regular representation models (corresponding to group convolutions) and variations of them, quotient representation models and others. As these models are essentially steerable CNNs, some of them were already explained in the benchmark experiment Section 6.5. In contrast to Section 6.5, the models here are not restricted to $d = 2$ dimensions and cover non-compact structure groups, including scaling and the Lorentz group. The differential geometric formulation allows furthermore to interpret group and quotient convolutions from a different viewpoint, which generalizes to arbitrary manifolds.

Row (1) of Table 14.1 lists Euclidean GM -convolutions on translation invariant $\{e\}$ -structures as visualized in Fig. 15.3a. Due to the triviality of the structure group $G = \{e\}$, no (non-trivial) gauge transformations exist and the only possible choice of group representation is the trivial representation. The G -steerability constraint becomes therefore trivial, such that the space of admissible convolution kernels remains unrestricted. When being pulled back to \mathbb{R}^d via a chart, the GM -convolution becomes by Theorem 15.2.1 a conventional convolution (or correlation). Theorem 15.2.2 asserts its translational equivariance. The models are therefore seen to correspond to the conventional Euclidean CNNs by LeCun et al. [175].

All of the other Euclidean models in rows (2-30) consider non-trivial structure groups. They can be thought of as conventional convolutions on \mathbb{R}^d with the additional constraint on the kernels to be G -steerable, which guarantees their $\text{Aff}(G)$ -equivariance.

Irrep features: The networks in rows (4, 6, 11, 12, 20, 24, 27) and (30) operate on feature fields that transform according to *irreducible representations* (irreps) of G . For $G = \text{SO}(2)$, listed in row (6) and already covered in Table 6.6, this leads to so-called harmonic networks [335, 322]. This name is motivated by the fact that $\text{SO}(2)$ -irrep steerable convolution kernels are circular harmonics. The additional reflectional constraint for $G = \text{O}(2)$, listed in row (14), adds parity selection rules that fix the phase of the circular harmonics, suppressing half of the degrees of freedom as compared to the $G = \text{SO}(2)$ case. Both kernel spaces were derived in Section 5.3.4; see in particular Tables 5.2 and 5.3.

The models by [323, 301, 211, 161, 3, 184] in row (20) consider irreps of $G = \text{SO}(3)$ and can therefore be seen as the analog to the models in row (6) in three dimensions. The space of valid kernels to map between fields that transform according to irreps (Wigner D-matrices) of orders l and J is here spanned by all spherical harmonics of the $2(\min(l, J)+1)$ orders j with $|l - J| \leq j \leq l + J$; Fig. 5.4 visualizes these selection rules. Irrep models for $G = \text{O}(3)$, listed in row (24), are again adding parity selection rules, which enforce here that convolutions between fields of the same parity require even parity spherical harmonics, while transitions between field parities require odd parity kernels. A variation of this approach is listed in row (27) [233]. A convolution of an input scalar field, i.e. $l = 0$, with spherical harmonics of order j yields irrep fields of the corresponding order $J = j$. However, instead of processing these irrep features further via convolutions, the authors compute their norm. This results in scalar fields, which are in the next layer processed in the same manner.

The model from [274] in row (30) does not assume the standard Euclidean metric but the Minkowski metric. Its structure group is the Lorentz group $G = \text{SO}(d-1, 1)$ and the global symmetry group is the Poincaré group. In addition to building the equivariant network, the authors propose an algorithm to compute the irreps of Lie groups from the structure constants of their Lie algebra.

A special case of irreps are *trivial representations* (row 11), which describe G -invariant feature vectors (scalars). Due to their invariance, such features can not encode differences between any patterns in G -related poses. The constraint on kernels that map between scalar fields becomes $K(gv) = \frac{1}{|\det g|} K(v)$ for any $v \in \mathbb{R}^d$ and any $g \in G$, enforcing kernels that are (in every channel individually) invariant under the action of G . The first column of Fig. 5.2, the first row of Fig. 5.3 and the upper left entry of Table 5.1 show such kernels for $G = \text{O}(2)$, $G = \text{O}(3)$ and $G = \mathcal{R}$, respectively. Interpreting the pixel grid of an image as a graph and applying a standard graph convolution to it corresponds to a trivially steerable convolution with $\text{O}(2)$ -invariant kernels since standard graph convolutions apply isotropic kernels [144].

An advantage of irrep features from a practical viewpoint is their low dimensionality and thus memory consumption per feature field. However, empirical results show that irrep field based steerable convolutions usually achieve a lower performance than other field types, for instance those based on regular representations. This statement is reflected in our evaluation of Möbius convolutions in Section 10.5 and the benchmark of isometry equivariant Euclidean steerable CNNs in Table 6.6.

Regular features and group convolutions: The probably most prominent class of field types for equivariant CNNs are *regular representations* of the structure groups. As defined in Def. B.5.18, regular representations operate on functions $F : G \rightarrow \mathbb{R}$, which assign a feature response to every G -pose of a kernel. The regular representation acts thereby by G -translating these functions, that is, $[\rho_{\text{reg}}(\tilde{g})F](g) = F(\tilde{g}^{-1}g)$. Specifically for finite groups this implies feature fields with the number of channels $c = |G|$ given by the order of the group; see Remark B.5.19. As non-finite groups imply non-finite regular representations, the corresponding features are in practice discretized, which can for instance be done by considering a finite subgroup of the structure group or by Monte Carlo sampling. Since regular feature fields $f \in \Gamma(\mathcal{A})$ assign a function $f^A(p) : G \rightarrow \mathbb{R}$ to each point $p \in M$ (when being expressed relative to any gauge A at p), they are equivalent to real-valued functions $\tilde{f} : GM \rightarrow \mathbb{R}$ on the G -structure GM .⁴ For the case of GM being induced by an

⁴Theorem J.0.1 in Appendix J proves this isomorphism $C^\infty(GM) \cong \Gamma(\mathcal{A}_{\rho_{\text{reg}}})$ for the practically relevant case of G being a finite group.

$\text{Aff}(G)$ -atlas, this is furthermore equivalent to real-valued functions $\tilde{f} : \text{Aff}(G) \rightarrow \mathbb{R}$ on $\text{Aff}(G) \cong GM$ (along the isomorphism in Eq. (15.14)). Equivariant linear maps between functions on the group $\text{Aff}(G)$ are *group convolutions*, which means that affine group convolution based CNNs are covered by our framework – this relation was in the context of steerable CNNs already proven in Section 4.5 above. Regular representation based GM -convolutions are in this regard best thought of as generalizations of group convolutions to the differential geometric setting.

$\text{Aff}(G)$ -group convolutions are in Table 14.1 listed in rows (2,3,5,7,13,17,19,22,25,28) and (29). As these models typically process scalar field inputs, like grayscale images, they apply an initial convolution from scalar fields to regular fields, followed by group convolutions, i.e. convolutions from regular to regular fields. As regular representations are permutation representations, they typically apply pointwise nonlinearities like ReLUs to each of the field channels individually.

The reflection equivariant CNN on \mathbb{E}_2 from [322] in row (5) applies \mathcal{R} -steerable kernels as derived in Section 5.2 and visualized in the bottom right entry of Table 5.1. Since the reflection group is finite with order $|\mathcal{R}| = 2$, the regular feature fields have two channels, each of which is associated to one of the two frame orientations of the \mathcal{R} -structure in Fig. 15.3c. The resulting model is globally $(\mathbb{R}^2, +) \rtimes \mathcal{R} = \text{Aff}(\mathcal{R})$ equivariant.

To construct $\text{SE}(2) = \text{Aff}(\text{SO}(2))$ equivariant group convolutions one would in theory have to consider the $\text{SO}(2)$ -structure in Fig. 15.3b with feature fields transforming according to the regular representation of $\text{SO}(2)$. In practice, most of the models in row (7) of Table 14.1 approximate this via regular representations of the finite cyclic subgroups $C_N \leq \text{SO}(2)$, consisting of discrete rotations by multiples of $2\pi/N$, as described in Chapter 6. As the order of these groups is $|C_N| = N$, the corresponding feature fields are N -dimensional. While the model performance is initially significantly increasing with N , it is empirically found to saturate at approximately 8 to 12 sampled directions; see Fig. 6.4 or [324, 322, 10]. For an intuition on the spaces of C_N -steerable kernels we refer to the visualizations in Fig. 18.3 or [324, 12, 10].

The $\text{E}(2) = \text{Aff}(\text{O}(2))$ equivariant group convolutions in row (13) are similarly approximated via dihedral subgroups $D_N \leq \text{O}(2)$, which consist of N rotations, each in two reflections. The feature fields are in this case $|D_N| = 2N$ -dimensional.

Simultaneous equivariance under translations and scaling is achieved by the $(\mathbb{R}^d, +) \rtimes \mathcal{S} = \text{Aff}(\mathcal{S})$ group convolutions in rows (2) and (17). The scaling group is hereby commonly discretized. As this would still lead to a (countably) infinite group order, the implementations introduce cutoffs, i.e. minimal and maximal scales as shown by the frames in Fig. 15.3d. Note that this leads to similar boundary effects as for conventional convolutions at the border of an image. The model in row (19) combines rotation and scaling equivariance, i.e. $G = \text{SO}(2) \times \mathcal{S}$. As the steerability constraints for $G = \text{SO}(2)$ and $G = \mathcal{S}$ affect only the kernels' angular and radial parts, respectively, $G = (\text{SO}(2) \times \mathcal{S})$ -steerable kernels are easily constructed from these solutions.

The models in rows (22) and (25) are equivariant w.r.t. translations, rotations and, for the latter, reflections of three-dimensional Euclidean spaces \mathbb{E}_3 . While Finzi et al. [91] choose a Monte-Carlo discretization of the regular representation, the models in [40, 333, 329] are based on different discrete subgroups of $\text{SO}(3)$ or $\text{O}(3)$. A current limitation of group convolution based rotation and reflection equivariant models in three dimensions is their high memory and compute requirement. For instance, the symmetry group of the cube, which has still a quite coarse resolution of rotations by $\pi/2$, already consists of $|G| = 48$ group elements, implying 48-dimensional feature fields in three-dimensional space. On the

other hand, the large number of symmetries reflects the greatly enhanced data efficiency of such models: the authors of [329] report the same performance of an equivariant model in comparison to a non-equivariant ($\{e\}$ -steerable) network despite training on a 10 times smaller dataset.

The models in rows (28) and (29) convolve on \mathbb{E}_3 , however, they consider cyclic and dihedral structure groups C_4 and D_4 , i.e. planar rotations and reflections around the (thus defined) z -axis. Their steerable kernels are therefore similar to those of the models from rows (7) and (13) but extend additionally in a new z -direction.

Regular to scalar and vector pooling: A variation of group convolutional networks are the models in rows (9,10,15,18) and (23), which are labeled by regular $\xrightarrow{\text{pool}} \text{trivial}$ and regular $\xrightarrow{\text{pool}} \text{vector}$. After applying a convolution to regular feature fields, they perform a max-pooling operation over G -responses (Eq. (6.2)), which results in scalar (trivial) fields [52, 195, 322, 107, 4], or a max-pooling together with an argmax, from which vector fields can be computed [196, 322]. Subsequent convolutions map from the resulting scalar or vector fields back to regular feature fields. As the pooling operations reduce the number of channels significantly from $|G|$ to 1 or d , respectively, the models become more memory and compute efficient than conventional group convolutions. On the downside, the pooling is accompanied with a loss of information, which is empirically found to decrease the model performance [322].

Quotient features: Rows (8,14) and (26) list models whose feature fields transform according to *quotient representations* of the structure group, which are permutation representations that are similar to regular representations. Given a subgroup \widehat{G} of G , the corresponding quotient representation acts on scalar functions $F : G/\widehat{G} \rightarrow \mathbb{R}$ on the quotient space G/\widehat{G} via translation, that is, $[\rho_{\text{quot}}^{G/\widehat{G}}(\tilde{g}) F](g\widehat{G}) = F(\tilde{g}^{-1}g\widehat{G})$; see Def. B.5.20. The dimensionality of the feature fields is therefore given by the index $|G : \widehat{G}|$ of \widehat{G} in G , which is for finite groups equal to $|G|/|\widehat{G}|$. Feature fields that transform under quotient representations can be seen as symmetry-constrained regular feature fields that are forced to take the same value on all group elements in the same coset $g\widehat{G}$ of \widehat{G} in G . In the differential geometric setting, these models can be viewed as assigning kernel responses to equivalence classes of \widehat{G} -related frames – and therefore to different choices of \widehat{G} -substructures $\widehat{G}M$ in GM .

A specific example are the representations in row (26), which are associated with the quotient $O(3)/O(2) \cong S^2$. Instead of allowing for arbitrary convolution kernels, the kernel constraint leads here to kernels which are invariant under rotations around the z -axis; see the visualizations in [136]. More details and a graphical intuition on quotient representation based feature fields can be found in Appendix C of [322]. The theory proposed in [162] covers quotient fields from an alternative viewpoint of group convolutions on right quotient spaces.

Induced representations: A generalization of regular and quotient representations are induced representations like the *induced $SO(2)$ -irreps* in row (16) of Table 14.1. Given any $SO(2)$ -irrep $\rho : SO(2) \rightarrow GL(n)$, the induced representation $\text{Ind}_{SO(2)}^{O(2)} \rho : O(2) \rightarrow GL(c)$ of $O(2)$ with $c = n \cdot |O(2) : SO(2)| = 2n$ acts in the following way: reflections permute two n -dimensional, orthogonal subspaces of \mathbb{R}^{2n} which correspond to the two cosets in $O(2)/SO(2)$ while rotations act on the individual subspaces via ρ . For ρ being the trivial representation of $SO(2)$ this recovers quotient representations as discussed above. In

comparison to $O(2)$ -irrep feature fields, the induced $SO(2)$ -irrep fields show a significantly improved performance. A more detailed description and empirical evaluation of these field types can be found in [322].

Quaternion representation: The last type of representation listed in Table 14.1 is the quaternion representation of three-dimensional rotations in row (21) [345]. It makes use of the usual representation of rotations via quaternions, which relies the identification of unit quaternions with $SU(2)$ and the existence of a surjective group homomorphism from $SU(2)$ to $SO(3)$. Note that the quaternion representation is actually a projective representation of $SO(3)$.

Discretization: While our theory is formulated on continuous Euclidean spaces, implementations sample feature fields usually on discrete subsets. The most common discretization of \mathbb{E}_d is in terms of the pixel grid \mathbb{Z}^d . An alternative are hexagonal planar grids on \mathbb{E}_2 as investigated by Hoogeboom et al. [125]. If such regular pixel grids are chosen, a basis of G -steerable kernels can be precomputed and sampled on this grid. Data like events in spacetime [274] or molecules in \mathbb{R}^3 [301, 161, 3, 211] are instead usually represented by irregular point clouds. In this case the kernels need to be given analytically, which allows their online sampling during the forward pass.

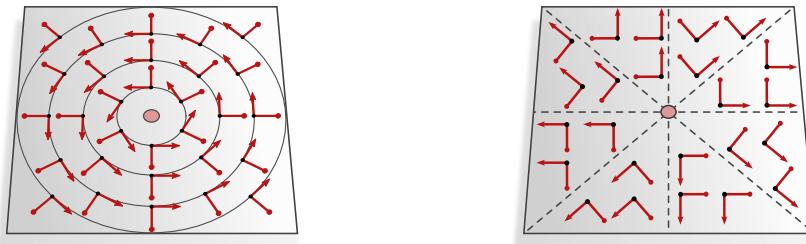
Finally, we want to mention that there exist *globally* $\text{Aff}(G)$ -equivariant models which are *not locally* G -equivariant. An example is TI-pooling (transformation-invariant pooling) [174], which feeds a set of globally transformed feature fields through a conventional Euclidean CNN and finally pools the resulting features over these transformations, which results in an invariant descriptor. While this model is not an $\text{Aff}(G)$ -equivariant GM -convolution as described in this Chapter, it is nonetheless constructed from an $\text{Aff}(\{e\}) = (\mathbb{R}^d, +)$ -equivariant (conventional) $\{e\}M$ -convolution.

Rotation equivariant CNNs on punctured Euclidean spaces

The models in rows (31-34) of Table 14.1 provide an interesting alternative for rotation equivariant convolutions on punctured Euclidean spaces $\mathbb{E}_d \setminus \{0\}$. They rely on *G-structures that are invariant under rotations around the chosen origin $\{0\}$* , as visualized for instance in Fig. 16.1. By specifying a preferred origin, these models lose the property to be translation equivariant.¹ However, if the *G*-structure is additionally invariant under scaling, which is for instance the case when it is induced by hyperspherical coordinates with a logarithmic radial component as shown in Fig. 16.3, the models become equivariant w.r.t. the direct product $\text{SO}(d) \times \mathcal{S}$ of the rotation and scaling group. Similarly, rotation and reflection invariant *G*-structures, visualized in Fig. 16.2, imply the $\text{O}(d)$ -equivariance of the corresponding *GM*-convolutions.

The models relate to spherical CNNs, discussed in Chapter 17 below, in two ways. Firstly, they assume rotationally invariant *G*-structures on $\mathbb{E}_d \setminus \{0\} \cong S^{d-1} \times \mathbb{R}_{>0}$, which can be seen as being composed of multiple rotationally invariant *G*-structures on $(d-1)$ -dimensional spherical shells S^{d-1} at different radii. The models can therefore be thought

¹This issue can be resolved by combining the network with a translation invariant origin predictor network [84]. Note that the rotation equivariance of the combined model is only preserved if this origin predictor is $\text{SE}(d)$ -equivariant.



(a) $\text{SO}(2)$ -invariant $\{e\}$ -structure as implicitly assumed by Finzi et al. [91].

(b) C_8 -invariant $\{e\}$ -structure as implicitly assumed by Chidester et al. [47].

Figure 16.1: Two examples of $\{e\}$ -structures on the punctured plane $\mathbb{E}_d \setminus \{0\}$ which 1) are invariant under rotations around the origin $\{0\}$ and 2) consist of orthonormal frames relative to the standard Euclidean metric. The corresponding *GM*-convolutions are rotation equivariant but not translation equivariant (in fact, $\mathbb{E}_d \setminus \{0\}$ does not even admit translations).

of as (hyper)spherical CNNs with an additional radial dimension $\mathbb{R}_{>0}$ [238], which is in Fig. 16.4 visualized for the case of $d = 3$ dimensions. Secondly, the polar coordinate systems of [84, 91, 47] (Figs. 16.1 and 16.3) induce G -structures that exhibit the same type of singularity at their origin like those of the punctured spherical CNNs in Fig. 17.2b at the poles. Note that the punctured Euclidean plane $\mathbb{E}_2 \setminus \{0\}$ and the punctured sphere $S^2 \setminus \{n, s\}$ (with north and south poles $\{n, s\}$ removed) are both topologically equivalent to a cylinder $S^1 \times \mathbb{R}_{>0} \cong S^1 \times \mathbb{R}$ and that the cylindrical $\{e\}$ -structures visualized in Figs. 16.1a, 16.3 (left) and 17.2b are diffeomorphic.

A major difference in comparison to the $SE(d)$ -equivariant networks from the previous chapter is that the models of the current chapter are only *globally* $SO(d)$ -equivariant around the origin instead of *locally* $SO(d)$ -equivariant ($SO(d)$ -steerable). While the globally equivariant models do not require $SO(d)$ -steerable kernels, they still require at least $SO(d-1)$ -steerable kernels. This is the case since $SO(d)$ is a $SO(d-1)$ -bundle over the spherical shells $S^{d-1} \cong SO(d)/SO(d-1)$ on which the G -structure is required to be $SO(d)$ rotation equivariant. For $d = 2$ this allows for $\{e\}$ -structures and non-steerable kernels since $SO(d-1) = SO(1) = \{e\}$; see Fig. 16.1 or 16.3. For $d = 3$ this requires at least a $SO(d-1) = SO(2)$ -structure on the individual spherical shells, which is visualized in Fig. 17.2a.

After these general remarks we will in the following briefly review the individual models on $\mathbb{E}_d \setminus \{0\}$ found in the literature from the viewpoint of coordinate independent CNNs. We start with the models in row (31) of Table 14.1, which are equivariant w.r.t. $SO(2)$ rotations around a chosen origin of \mathbb{E}_2 and proceed with the models in row (32), which are additionally scale equivariant. The network listed in row (33), which we discuss last, is globally $O(3)$ -equivariant around the origin of \mathbb{E}_3 .

16.1 Global rotation equivariance on $\mathbb{E}_2 \setminus \{0\}$

We start with the conceptually simplest models, which are globally rotation equivariant networks that rely on solely rotation invariant $\{e\}$ -structures on $\mathbb{E}_2 \setminus \{0\}$ [91, 47]. These models assume the standard Euclidean metric on $\mathbb{E}_2 \setminus \{0\}$, relative to which the frames are orthonormal. Together, these two requirements imply $\{e\}$ -structures as shown in Fig. 16.1.

In addition to the considered G -structures, the networks depend on the specific implementation of the transporter pullback and thus on the geodesics and parallel transporters. The geodesics are in both models assumed to be the standard geodesics on Euclidean spaces (i.e. straight lines), corresponding to the Levi-Civita connection of the Euclidean metric. As $\mathbb{E}_2 \setminus \{0\}$ is not geodesically complete, zero-padding has to be used for exponential maps that would end at the origin. Note that this does not have an impact on the final result as the lost geodesics are of measure zero.

The parallel transport of feature vectors, on the other hand, does *not* correspond to the Levi-Civita connection since the Levi-Civita connection is not compatible with the $\{e\}$ -structures. Instead, the models assume the unique $\{e\}$ -compatible *trivial connections* which are implied by the respective $\{e\}$ -structures.² According to the trivial connections, the numerical coefficients of feature vectors do not transform when being transported, despite the frames being rotated relative to the usual notion of parallelism on Euclidean spaces. In practice, this just

² An animation of the $\{e\}$ -compatible transport corresponding to Fig. 16.1a can be found on [Wikipedia](#).

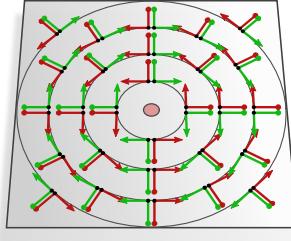


Figure 16.2: An $O(2)$ -invariant \mathcal{R} -structure on $\mathbb{E}_2 \setminus \{0\}$, which is constructed by adding a reflected versions to each frame of the $\{e\}$ -structure in Fig. 16.1a. The corresponding GM -convolution is simultaneously equivariant w.r.t. global rotations and reflections in $\text{Isom}_{RM} = O(2)$ around the origin.

means that the transporters $\rho(g_\gamma^{A\tilde{A}}) = \text{id}_{\mathbb{R}^c}$ can be ignored in the implementation – which is the reason that they are not being discussed in the original papers [91, 47].

As rotations leave the considered $\{e\}$ -structures invariant and are at the same time isometries, we have $\text{Isom}_{\{e\}M} = SO(2)$ for the model by Finzi et al. [91] (Fig. 16.1a) and $\text{Isom}_{\{e\}M} = C_8$ for the model by Chidester et al. [47] (Fig. 16.1b). Theorem 13.2.5 asserts then that the corresponding GM -convolutions are $\text{Isom}_{\{e\}M}$ -equivariant, which is in agreement with the statements made by the authors.

Before going on we want to mention that the C_8 -invariant $\{e\}$ -structure in Fig. 16.1b is not continuous and does therefore not guarantee a continuous (or smooth) inference. An advantage of this $\{e\}$ -structure from an engineering viewpoint is that it is locally isometric to the canonical $\{e\}$ -structure of \mathbb{R}^2 , which allows to run conventional Euclidean convolution routines on each octant. The authors discuss the generalization to C_N -invariant $\{e\}$ -structures, which become in the limit $N \rightarrow \infty$ equivalent to the $SO(2)$ -invariant $\{e\}$ -structure in Fig. 16.1a.

It is furthermore possible to make the models globally $O(2)$ -equivariant by using reflection steerable kernels instead of unconstrained kernels. From a theoretic viewpoint this corresponds to the $\text{Isom}_{RM} = O(2)$ -invariant \mathcal{R} -structure RM on $\mathbb{E}_2 \setminus \{0\}$ shown in Fig. 16.2. Note that RM is a \mathcal{R} -bundle over $\mathbb{E}_2 \setminus \{0\}$, whose restriction to circles of constant radius is as a principal bundle isomorphic to $O(2)$, interpreted as \mathcal{R} -bundle over the quotient space $O(2)/\mathcal{R} \cong S^1$.

16.2 Global rotation and scale equivariance on $\mathbb{E}_2 \setminus \{0\}$ via log-polar coordinates

By making the rotation invariant G -structures from the last section additionally scale invariant, the corresponding GM -convolutions become equivariant w.r.t. the direct product group $SO(2) \times \mathcal{S}$. Such G -structures are induced by *log-polar coordinates*, shown in Fig. 16.3, which allow for a convenient implementation of the GM -convolution in terms of conventional Euclidean convolutions on the coordinate representation \mathbb{R}^2 . The translation equivariance of convolutions on \mathbb{R}^2 corresponds then to the $SO(2) \times \mathcal{S}$ -equivariance on $\mathbb{E}_2 \setminus \{0\}$. For clarity, we start by describing the model in terms of log-polar coordinates as proposed by Esteves et al. [84].³ Subsequently, we investigate how this model and its properties are explained in our framework.

³The idea to implement rotation invariant correlations via log-polar transforms appeared already in the 80's [255, 34].

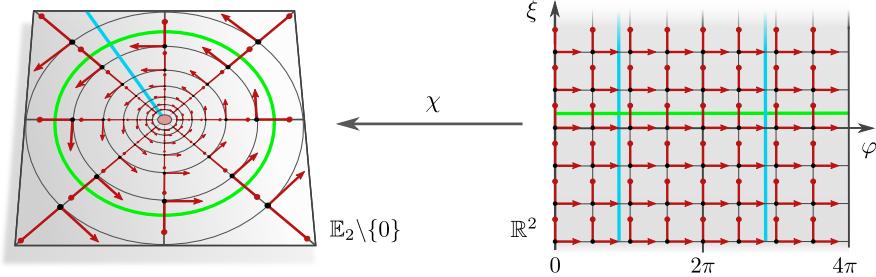


Figure 16.3: Log-polar coordinates $\chi : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \setminus \{0\} : (\varphi, \xi) \mapsto (e^\xi \cos(\varphi), e^\xi \sin(\varphi))$ map angles $\varphi \in \mathbb{R}$ and log-radii $\xi = \log\|p\| \in \mathbb{R}$ to points p in $\mathbb{R}^2 \setminus \{0\}$. After choosing Cartesian coordinates of $\mathbb{E}_2 \setminus \{0\} \cong \mathbb{R}^2 \setminus \{0\}$, this yields a coordinatization of $\mathbb{E}_2 \setminus \{0\}$ by \mathbb{R}^2 . The log-polar coordinates imply an $\{e\}$ -structure on $\mathbb{E}_2 \setminus \{0\}$, consisting of reference frames $[\frac{\partial}{\partial \varphi}, \frac{\partial}{\partial \xi}]$ that are aligned with the coordinate grid. They furthermore induce a Riemannian metric, which differs from the usual Euclidean metric and relative to which the induced frames are orthonormal. *GM*-convolutions on the $\{e\}$ -structure correspond to conventional Euclidean convolutions in the coordinates \mathbb{R}^2 . Translations $(\Delta\varphi, \Delta\xi) \in (\mathbb{R}^2, +)$ on \mathbb{R}^2 correspond via χ to rotations and rescalings of $\mathbb{E}_2 \setminus \{0\}$, where the rotation angles and rescaling factors are given by $\Delta\varphi$ and $e^{\Delta\xi}$, respectively. The translation equivariance of the convolution in coordinates \mathbb{R}^2 implies therefore the $\text{SO}(2) \times \mathcal{S}$ -equivariance of the *GM*-convolution on $\mathbb{E}_2 \setminus \{0\}$. This result is in agreement with the isometry equivariance of the *GM*-convolution since the transformations in $\text{Isom}_{GM} = \text{SO}(2) \times \mathcal{S}$ are isometries relative to the induced metric. Esteves et al. [84] implement such *GM*-convolutions in terms of conventional convolutions on \mathbb{R}^2 .

Log-polar coordinates of the punctured Euclidean vector space $\mathbb{R}^2 \setminus \{0\}$ are defined in terms of the smooth surjection

$$\chi : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \setminus \{0\}, \quad (\varphi, \xi) \mapsto (e^\xi \cos(\varphi), e^\xi \sin(\varphi)), \quad (16.1)$$

which assigns points $p = \chi(\varphi, \xi)$ in $\mathbb{R}^2 \setminus \{0\}$ to a given polar angle $\varphi \in \mathbb{R}$ and log-radius $\xi = \log\|p\| \in \mathbb{R}$. This map is 2π -periodic in the angular coordinate (note the repetition of the blue stripe on the right-hand side of Fig. 16.3) and is therefore in particular non-injective. A restriction to $[0, 2\pi) \times \mathbb{R}$ would be bijective and continuous, however, not homeomorphic – this will require us below to consider at least two charts to cover the punctured plane. Cartesian coordinates identify $\mathbb{R}^2 \setminus \{0\}$ with $\mathbb{E}_2 \setminus \{0\}$, and therefore allow to assign log-polar coordinates to the latter. As different (right-handed) Cartesian coordinate systems that are centered in the origin of $\mathbb{E}_2 \setminus \{0\}$ differ only by rotations, the assignment of log-polar coordinates is ambiguous by a shift in the angular component.

Given a feature map $f : \mathbb{R}^2 \setminus \{0\} \rightarrow \mathbb{R}^c$ (an $\{e\}M$ -associated feature field, as clarified below), Esteves et al. [84] consider its pullback $\tilde{f} := f \circ \chi : \mathbb{R}^2 \rightarrow \mathbb{R}^c$ via log-polar coordinates, defined by the commutativity of the following diagram:

$$\begin{array}{ccccc} \mathbb{R}^2 & \xrightarrow{\chi} & \mathbb{R}^2 \setminus \{0\} & \xrightarrow{f} & \mathbb{R}^c \\ & \downarrow & \underbrace{\hspace{10em}}_{\tilde{f}} & & \\ & & & & \end{array} \quad (16.2)$$

A rotation and scaling equivariant group convolution of the feature field f on $\mathbb{R}^2 \setminus \{0\}$ is then defined by 1) pulling it via χ back to coordinates \mathbb{R}^2 2) applying a conventional Euclidean convolution there and 3) mapping the result back to $\mathbb{R}^2 \setminus \{0\}$. This procedure is well defined since χ is smooth, such that smooth feature maps (feature fields) f result in smooth and

periodic pullbacks \tilde{f} . Since convolutions are position-independent, their output feature map will still be periodic and smooth, and corresponds therefore uniquely to a smooth feature map on $\mathbb{R}^2 \setminus \{0\}$.⁴

The rotation and scaling equivariance of the implied group convolution on $\mathbb{R}^2 \setminus \{0\}$ follows from the translation equivariance of the coordinate function χ .⁵ Let (φ, ξ) be any coordinates in \mathbb{R}^2 and let $(\Delta\varphi, \Delta\xi)$ be any translation in $(\mathbb{R}^2, +)$. The point of $\mathbb{R}^2 \setminus \{0\}$ that corresponds to translated coordinates $(\varphi + \Delta\varphi, \xi + \Delta\xi)$ relates then to the point corresponding to non-translated coordinates (φ, ξ) via a scaling by the factor $e^{\Delta\xi}$ and rotation by the angle $\Delta\varphi$:

$$\begin{aligned}\chi(\varphi + \Delta\varphi, \xi + \Delta\xi) &= e^{\xi + \Delta\xi} \begin{pmatrix} \cos(\varphi + \Delta\varphi) \\ \sin(\varphi + \Delta\varphi) \end{pmatrix} \\ &= e^{\Delta\xi} \begin{pmatrix} \cos(\Delta\varphi) & -\sin(\Delta\varphi) \\ \sin(\Delta\varphi) & \cos(\Delta\varphi) \end{pmatrix} e^\xi \begin{pmatrix} \cos(\varphi) \\ \sin(\varphi) \end{pmatrix} \\ &= e^{\Delta\xi} \begin{pmatrix} \cos(\Delta\varphi) & -\sin(\Delta\varphi) \\ \sin(\Delta\varphi) & \cos(\Delta\varphi) \end{pmatrix} \chi(\varphi, \xi) \\ &=: (\Delta\varphi, \Delta\xi) \triangleright \chi(\varphi, \xi)\end{aligned}\tag{16.3}$$

In terms of a diagram, this means that

$$\begin{array}{ccc} \mathbb{R}^2 & \xrightarrow{(\Delta\varphi, \Delta\xi)+} & \mathbb{R}^2 \\ \chi \downarrow & & \downarrow \chi \\ \mathbb{R}^2 \setminus \{0\} & \xrightarrow{(\Delta\varphi, \Delta\xi)\triangleright} & \mathbb{R}^2 \setminus \{0\} \end{array}\tag{16.4}$$

commutes for arbitrary translations. Together with the translation equivariance of conventional convolutions on \mathbb{R}^2 , this implies that rotated and scaled input feature maps on $\mathbb{R}^2 \setminus \{0\}$ will lead to rotated and scaled output feature maps on $\mathbb{R}^2 \setminus \{0\}$, i.e. the $\text{SO}(2) \times \mathcal{S}$ -equivariance of the convolution on $\mathbb{R}^2 \setminus \{0\}$. More details on this viewpoint are found in [84] and [15].

We will now revisit this convolution operation and its properties from the viewpoint of coordinate free *GM*-convolutions. To do so, we consider an atlas of charts that are consistent with the log-polar coordinates, and discuss the induced $\{e\}$ -structure, gauges, Riemannian metric, geodesics and parallel transport that it implies. The claimed $\text{SO}(2) \times \mathcal{S}$ -equivariance follows immediately from the $\text{Isom}_{\{e\}M}$ -equivariance of *GM*-convolutions. For notational convenience, we will again identify $\mathbb{E}_2 \setminus \{0\}$ via some choice of Cartesian coordinates with $\mathbb{R}^2 \setminus \{0\}$.

As the restriction $\tilde{\chi} : [0, 2\pi) \times \mathbb{R} \rightarrow \mathbb{R}^2 \setminus \{0\}$ of the log-polar coordinates χ to non-redundant angles is bijective and continuous, one might be tempted to take its inverse as a coordinate

⁴To see this, note that χ is a quotient map (since its angular part is a quotient map $\mathbb{R} \rightarrow S^1 \cong \mathbb{R}/2\pi\mathbb{Z}$). For continuous (instead of smooth) feature maps the statement follows from the universal property of quotient spaces; see e.g. Wikipedia. As the smoothness of a function is defined as its continuous differentiability, the universal property can be applied recursively to show that the statement holds for smooth feature maps as well.

⁵That this is possible relies on the fact that there is a group homomorphism $(\mathbb{R}^2, +) \rightarrow \text{SO}(2) \times \mathcal{S}$, $(\Delta\varphi, \Delta\xi) \mapsto (R_{\Delta\varphi}, e^{\Delta\xi})$, defined by the group isomorphism $\exp : (\mathbb{R}^1, +) \rightarrow \mathcal{S}$ on the second factor and the group homomorphism (quotient map) $R : (\mathbb{R}^1, +) \rightarrow \text{SO}(2) \cong (\mathbb{R}^1, +)/2\pi\mathbb{Z}$ on the second factor, where $R_{\Delta\varphi}$ denotes the rotation matrix by an angle of $\Delta\varphi$.

chart. This is, however, not possible, since $\tilde{\chi}$ is not a homeomorphism, as required for charts. Instead, we consider an atlas consisting of two charts that are defined in terms of restrictions of χ and that cover $\mathbb{R}^2 \setminus \{0\}$. One particular choice is to define chart codomains as open sets $V^A = (0, 2\pi) \times \mathbb{R}$ and $V^B = (-\epsilon, \epsilon) \times \mathbb{R}$ for some $0 < \epsilon < \pi$ and, for $X = A, B$, define charts on $U^X = \chi(V^X)$ as $x^X := (\chi|_{V^X})^{-1} : U^X \rightarrow V^X$. Intuitively, this atlas achieves the same as the naive attempt to define charts as the inverse $\tilde{\chi}$. The important difference is, however, that the charts are diffeomorphic, which is necessary to assure the smoothness of all operations.

As usual, these charts induce local frame fields and bundle trivializations on U^A and U^B , respectively. It is easy to see that the transition maps $g_p^{BA} = \frac{\partial x^B}{\partial x^A}|_{x^A(p)}$ on $U^A \cap U^B$ are trivial, which implies that the union of the frame fields defines a smooth $\{e\}$ -structure $\{e\}M$ on $\mathbb{R}^2 \setminus \{0\}$. These coordinate bases, which are in the literature often denoted as $[\frac{\partial}{\partial \varphi}, \frac{\partial}{\partial \xi}]$, are shown in Fig. 16.3 (left). Our calculation in Eq. (16.3) above implies that the induced $\{e\}$ -structure is $\text{SO}(2) \times \mathcal{S}$ -invariant.

The charts induce furthermore a Riemannian metric, which differs from the usual Euclidean metric on $\mathbb{R}^2 \setminus \{0\}$. It is defined as the pullback of the Euclidean metric $\langle \cdot, \cdot \rangle_{\mathbb{R}^2}$ in the charts' codomains, and is therefore pointwise given by

$$\eta_p(v, w) := \langle \hat{dx}_p^X(v), \hat{dx}_p^X(w) \rangle_{\mathbb{R}^2}, \quad (16.5)$$

where $v, w \in T_p M$ and X denotes either chart with $p \in U^X$. The chart induced $\{e\}$ -structure consists by construction of *frames that are orthonormal w.r.t. this chart induced metric*, even though these *frames grow with the radius when measured relative to the standard Euclidean metric*. The Levi-Civita connection of the induced metric differs from the usual Euclidean connection and implies therefore alternative geodesics and parallel transporters. As the metric is pulled back via the charts, the geodesics correspond to straight lines in the charts' codomains – an example are the coordinate lines on $\mathbb{R}^2 \setminus \{0\}$ in Fig. 16.3. The parallel transport corresponds to the usual transport in the charts' codomains as well, which implies that it keeps transported vectors in a fixed angle to the coordinate lines on $\mathbb{R}^2 \setminus \{0\}$; cf. footnote 2. Note that this is the same transport as already discussed above in the models corresponding to Figs. 16.1, where it was *not* the transport corresponding to the Levi-Civita connection since these models assumed the standard metric on $\mathbb{R}^2 \setminus \{0\}$ instead of the chart induced metric.

The $\{e\}$ -structure preserving isometries $\text{Isom}_{\{e\}M} \cong \text{SO}(2) \times \mathcal{S}$ *relative to the chart induced metric* are given by rotations and rescaling of the $\{e\}$ -structure *relative to the usual Euclidean metric*. Theorem 13.2.5 implies the $\text{SO}(2) \times \mathcal{S}$ -equivariance of the corresponding GM-convolution – which recovers the statement made by Esteves et al. [84] in our theory. As stated above, the fact that the metric is induced via the charts means that all operations reduce to the usual Euclidean operations when being expressed in the chart. The GM-convolution is therefore best implemented via a conventional convolution on the chart, as proposed by Esteves et al. [84].

Note that the $\text{SO}(2) \times \mathcal{S}$ -equivariance of the GM-convolution is easily extended to $\text{O}(2) \times \mathcal{S}$ -equivariance, which includes reflections. This is implemented by performing a reflection equivariant convolution in the chart, which corresponds to the \mathcal{R} -structure shown in Fig. 15.3c. On $\mathbb{R}^2 \setminus \{0\}$, this implies a \mathcal{R} -structure that looks similar to that in Fig. 16.2 above, with the difference that the \mathcal{R} -structure is additionally invariant under a global rescaling.

16.3 Global rotation equivariance on $\mathbb{E}_3 \setminus \{0\}$

The ideas presented above can be generalized to the three-dimensional setting, i.e. to the punctured Euclidean space $\mathbb{E}_3 \setminus \{0\}$. Globally rotation equivariant *GM*-convolutions correspond here to G -structures that are invariant under $\text{SO}(3)$ rotations around the origin. While the radial dependency of such G -structures is left unconstrained, the demand for rotational invariance imposes a constraint on their form over spherical shells at fixed radii, which are the orbits of the action of $\text{SO}(3)$ on $\mathbb{E}_3 \setminus \{0\}$. The fact that the sphere $S^2 = \text{SO}(3)/\text{SO}(2)$ is a homogeneous space of $\text{SO}(3)$ with stabilizer subgroups isomorphic to $\text{SO}(2)$ implies that the structure group of an $\text{SO}(3)$ -invariant G -structure can not be reduced further than $G = \text{SO}(2)$; see Fig. 17.2a. We are therefore essentially considering spherical CNNs with an additional radial dimension. For a review on spherical CNNs we refer the reader forward to Chapter 17.

Ramasinghe et al. [238] identified this situation and designed $\text{SO}(3)$ -equivariant convolutions on $\mathbb{E}_3 \setminus \{0\}$. Before coming to our classification as *GM*-convolution, listed in row (33) of Table 14.1, we briefly review the authors' formulation and implementation. Their implementation is based on spherical CNNs with the addition that 1) kernels extend in the radial direction and 2) are shared over shells at different radii; see Fig. 16.4 (left). As commonly done for spherical CNNs, the angular dependency of the kernels is encoded via their Fourier spectrum on S^2 , that is, in terms of spherical harmonics expansion coefficients. The sharing of these expansion coefficients implies that the shared kernels cover the same solid angle for all radii, implying that the *kernels dilate in angular direction linearly with the radius*.⁶ In the discretized implementation, the spherical shells are located at equidistant radii – which implies that the *kernels do not dilate in radial direction*. From these insights we infer the specific G -structure that the model assumes below. The kernels themselves are constrained such that they are invariant under $\text{SO}(2)$ rotations around the radial axis through their center, which is often referred to as *zonal kernels*; see Fig. 16.5 and [83]. As proven in [83] and [238], the convolution with such kernels is $\text{SO}(3)$ -equivariant. That this is the case is intuitively clear since rotations of the spherical shells have $\text{SO}(2)$ as stabilizer subgroup, w.r.t. which the zonal kernels are invariant. As we will argue below, the model is actually $\text{O}(3)$ -equivariant, that is, additionally equivariant under reflections.

To recover this model from the viewpoint of *GM*-convolutions, we need to determine the corresponding G -structure on $M = \mathbb{E}_3 \setminus \{0\}$. As stated above, the $\text{SO}(3)$ -equivariance of the model requires the G -structure to be invariant under the action of $\text{SO}(3)$ but does not constrain their radial variation. To infer this radial dependency of the G -structure, recall that we defined convolutional weight sharing at $p \in M$ as aligning the template kernel $K : \mathbb{R}^3 \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ relative to some (arbitrary) frame in $G_p M$ of the tangent spaces $T_p M$. The kernel sharing considered by Ramasinghe et al. [238] lets us therefore draw conclusions about the implicitly considered G -structure. The authors share kernels such that their area tangent to the spherical shells extends with growing distance from the origin (they cover the same solid angle at each radius) while their radial thickness remains constant. Fig. 16.4 (left) shows this radial variation of the shared kernels while Fig. 16.4 (middle) shows the corresponding scaling of exemplary reference frames. Together with the required $\text{SO}(3)$ -invariance of the G -structure, this implies (at least) an $\text{SO}(2)$ -structure, whose restriction to one spherical shell is visualized in Fig. 16.4

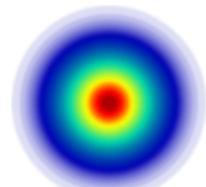


Figure 16.5: A *zonal* (isotropic) kernel is simultaneously $\text{SO}(2)$ - and $\text{O}(2)$ -steerable; cf. Eqs. (16.6) and (16.7).

⁶The dilation is here measured relative to the standard Euclidean metric of $\mathbb{E}_3 \setminus \{0\}$.

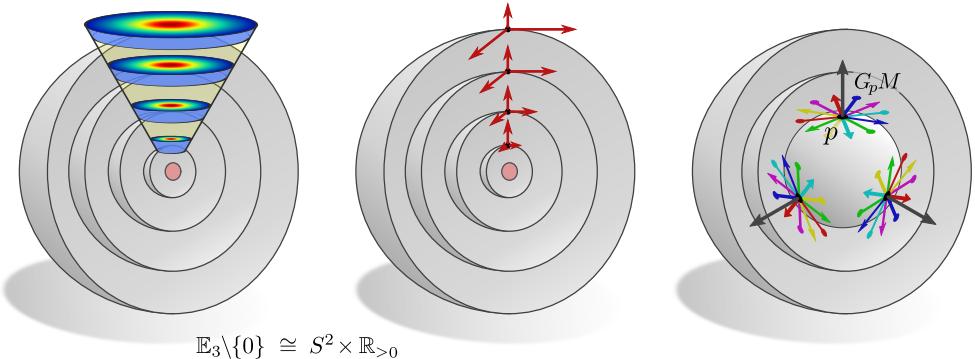


Figure 16.4: The G -structure that was implicitly assumed by Ramasinghe et al. [238] can be deduced from the weight sharing scheme. *Left:* Weight sharing of (isotropic) convolution kernels over $\mathbb{E}_3 \setminus \{0\} \cong S^2 \times \mathbb{R}_{>0}$ as proposed in [238]. The kernels are defined to cover the same solid angle, independent of the distance from the origin, such that their diameter grows linearly with this distance. The kernels' extent in radial direction is independent from the distance from the origin.. *Middle:* In our theory, kernels are shared relative to reference frames of the G -structure. To recover the proposed weight sharing scheme, GM needs to consist of frames whose axes in angular direction grow linearly with the radial distance from the origin, while the axes in radial directions need to keep their size fixed (both relative to the standard Euclidean metric). Such frames imply an alternative Riemannian metric on $\mathbb{E}_3 \setminus \{0\}$. *Right:* As the resulting GM -convolution should be $\text{SO}(3)$ -equivariant, the G -structure is required to be invariant under rotations around the origin. This requires (at least) an $\text{SO}(2)$ -structure, whose restriction to one spherical shell is shown in the right part of the figure. Compare this to the $\text{SO}(3)$ -invariant $\text{SO}(2)$ -structure of spherical CNNs in Fig. 17.2a.

(right).⁷ The considered metric follows from this G -structure, since its frames define the relevant notion of orthonormality. Note that this metric differs from the usual Euclidean metric.

By construction, we have rotations $\text{Isom}_{GM} = \text{SO}(3)$ as G -structure preserving isometries. GM -convolutions defined by this G -structure, which may differ in their input and output field type, will therefore (by Theorem 13.2.5) be rotation equivariant. The specific GM -convolution assumed by Ramasinghe et al. [238], i.e. the assumed field types, can be deduced from the fact that the authors assume zonal kernels: such kernels arise naturally when considering *scalar fields*, i.e. trivial field representations, since the kernel constraint, Eq. (9.37), becomes in this case

$$K(gv) = K(v) \quad \forall v \in \mathbb{R}^3, g \in \text{SO}(2), \quad (16.6)$$

enforcing isotropic (zonal) kernels.⁸ These kernels are listed in the upper left entry of the $\text{SO}(2)$ -steerable kernel space solution Table 5.2.

As a variation of the model, one could consider the $\text{O}(2)$ -structure that follows from the $\text{SO}(2)$ -structure by adding reflected reference frames (reflecting over an arbitrary axis within the planes tangent to the spherical shells, keeping the radial frame vectors still pointing outwards).⁹ In this case one has G -structure preserving isometries $\text{Isom}_{GM} = \text{O}(3)$ that consist

⁷The two-dimensional analog would look similar to the G -structure in Fig. 16.3 but with all frame vectors in radial direction having unit norm (relative to the Euclidean metric).

⁸Kernels which map between “scalar fields”, i.e. fields that transform according to the trivial representation of G , are always G -invariant. For $G = \text{SO}(2)$, this implies isotropic (zonal) kernels, while $G = \mathcal{R}$ implies the reflection invariant kernels in the upper left entry of Table 5.1.

⁹This $\text{O}(2)$ -structure is the analog of the $\text{O}(1) = \mathcal{R}$ -structure in Fig. 16.2 for $d=3$ instead of $d=2$.

of global rotations and reflections around the origin, and therefore $O(3)$ -equivariant GM -convolutions. An interesting special case in the current context is that of GM -convolutions that map between scalar fields, for which the kernel constraint reads

$$K(gv) = K(v) \quad \forall v \in \mathbb{R}^3, g \in O(2). \quad (16.7)$$

This seems like a stronger constraint than that in Eq. (16.6) above: instead of only demanding kernels to be rotationally invariant, it requires them additionally to be invariant under reflections. However, since rotation invariant kernels are already invariant under reflections, this leads again to zonal kernels, and therefore exactly the same kernel space as for $SO(2)$.¹⁰ This implies that the model by Ramasinghe et al. [238] is actually not only $SO(3)$ -equivariant, as claimed by the authors, but more generally $O(3)$ -equivariant, which justifies our classification in row (33) of Table 14.1. Note that this is a special case that applies only for scalar fields – the spaces of $SO(2)$ - and $O(2)$ -steerable kernels differ for general group representations; cf. the kernel space solution Tables 5.2 and 5.3.

How does the model by Ramasinghe et al. [238] relate to that of Esteves et al. [84], which relies on the G -structure shown in Fig. 16.3? A key difference between the two approaches is that the G -structure in Fig. 16.3 consists of frames whose outward pointing axes grow with the radial distance from the origin, which is not the case for the G -structure in Fig. 16.4. If we modify the latter to consist of frames whose radial axes grow linearly with the frames’ distance from the origin, one would have $\text{Isom}_{GM} = SO(3) \times \mathcal{S}$ (instead of $\text{Isom}_{GM} = SO(3)$). The corresponding GM -convolution would therefore additionally be scale equivariant. In an implementation, this could easily be realized by spacing the discrete spherical shells considered by Ramasinghe et al. [238] exponentially instead of uniformly (corresponding to a uniform spacing of the logarithmized radius).

Lastly, we briefly discuss the convolution by Boomsma and Frellsen [20] that is listed in row (34) of Table 14.1. It relies on a radial projection of the signal on spherical shells to a circumscribing cube. To define a convolution on the cube, the authors cut it open at some of its edges and flatten it out; see Fig. 2 in their work. Subsequently, they perform a conventional two-dimensional convolution on the flattened cube faces. Extending this operation with a third, radial dimension, defines a convolution on $\mathbb{E}_3 \setminus \{0\}$. As the radial shells are in the discretized implementation again spaced equidistantly, this operation corresponds to a GM -convolution on an $\{e\}$ -structure that varies radially as shown in Fig. 16.4. The projection from the spherical shells to the cube implies a distortion of the frames on each of the cubes faces, and thus to a distortion of the metric on the spherical shells. The $\{e\}$ -structure is discontinuous at most of the cuts and does therefore not allow the convolution to preserve the continuity of feature fields. Since S^2 is not parallelizable, this issue can not be resolved without assuming a non-trivial structure group G . The $\{e\}$ -structure as a whole is not preserved by any isometries, implying that the model’s global equivariance group $\text{Isom}_{\{e\}M} = \{e\}$ is trivial. However, as the restriction of the $\{e\}$ -structure to the four “vertical” faces of the cube is invariant under rotations by multiples of $\pi/2$, the model is in practice partially equivariant w.r.t. global C_4 -rotations around the vertical axis. For datasets whose samples are centered around the origin $\{0\}$ and are rotationally symmetric in distribution, this property is empirically shown to lead to an improved performance in comparison to conventional convolutions on \mathbb{E}_3 . The authors are furthermore investigating the effect of different weight sharing schemes over the radial dimension, finding that full weight sharing works best in practice.

¹⁰More formally, we are searching for kernels that satisfy $K(gv) = K(v) \quad \forall g \in G$, that is, which are invariant on the orbits $G.v = \{gv \mid g \in G\} \in G \backslash \mathbb{R}^d$ of points v in \mathbb{R}^d . As the orbits $O(2).v = SO(2).v$ agree for any $v \in \mathbb{R}^3$, the resulting kernel spaces are the same.

Spherical coordinate independent CNNs

Beyond convolutions on Euclidean spaces, convolutions on the 2-sphere S^2 are of great practical relevance. Applications include omnidirectional vision tasks, global weather forecasting, or the analysis of the cosmic microwave background. Instead of being translation equivariant, spherical convolutions are typically required to be rotation equivariant. The isometry group $\text{Isom}(S^2) = \text{O}(3)$ of the sphere and its decomposition in the most relevant subgroups is visualized in Fig. 17.1.

A major difference between Euclidean spaces \mathbb{E}_d and the sphere S^2 is that the latter is not parallelizable, i.e. does not allow for a global, continuous frame field. Reductions of the structure group beyond $G = \text{SO}(2)$ are topologically obstructed, which means that spherical convolutions require at least $\text{SO}(2)$ -steerable kernels if they should preserve the continuity of feature fields. The corresponding $\text{SO}(2)$ -structure, which is fully determined by the sphere's metric and orientation, is shown in Fig. 17.2a. *GM*-convolutions on this globally rotation invariant G -structure are guaranteed to be $\text{Isom}_{\text{SOM}} = \text{SO}(3)$ -equivariant.

Despite the unavoidable topological obstruction, many authors proposed spherical CNNs that do not apply $\text{SO}(2)$ -steerable kernels. The most prevalent choice of $\{e\}$ -structure corresponding to such convolutions is the frame field shown in Fig. 17.2b, whose orthonormal reference frames (Eq. (17.6)) are aligned with the coordinate grid of spherical coordinates. Note that this frame field comes with singularities at the poles, where the convolutions becomes discontinuous. To reconcile such models with our theory, in particular the smoothness assumption of the G -structures, they need to be described as convolutions on a topological cylinder with sphere-like metric. The isometry group of this punctured sphere $S^2 \setminus \{n, s\}$ without poles $n, s \in S^2$ is the subgroup $\text{O}(2)$ (Fig. 17.1 (middle and right)) of the sphere's full isometry group $\text{O}(3)$. The visualized $\{e\}$ -structure is preserved by azimuthal rotations in $\text{Isom}_{\{e\}M} = \text{SO}(2)$, i.e. rotations around the axis through the poles.

From an engineering perspective, both approaches have their justification: fully isotropic applications like the analysis of the cosmic microwave background in Fig. 1.4a require fully $\text{SO}(3)$ -equivariant models on S^2 . Learning tasks that come with a preferred rotation axis, which is for instance the case for the earth or panoramic images with a distinguished “up” and “down” direction, might benefit from the additional geometric information encoded in the $\{e\}$ -structure. Our empirical results form Section 6.3 suggest that it is in such cases often useful to work with a combination of both approaches: initial layers with fully equivariant convolutions can exploit local symmetries in the data, while subsequent “group restricted” layers with only azimuthal equivariance can learn to discriminate based on the preferred axis; see Section 4.4 and [323].

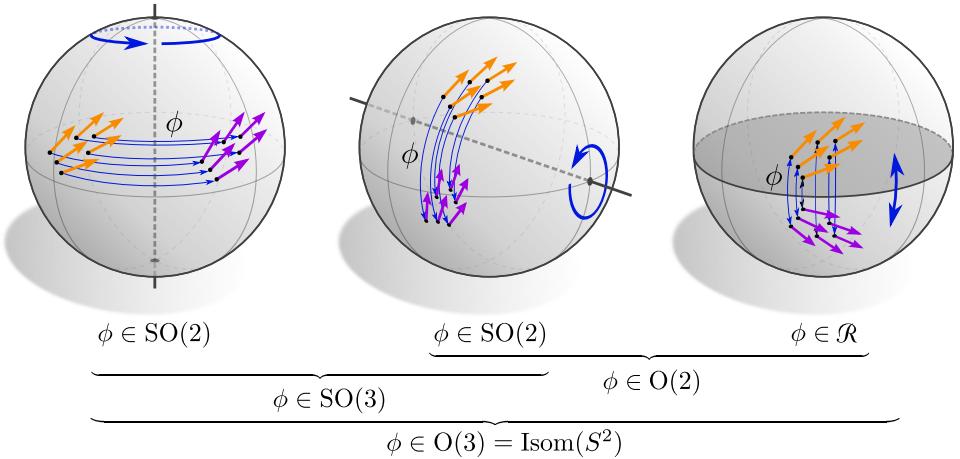


Figure 17.1: Visualization of the 2-sphere’s isometry group $\text{Isom}(S^2) = \text{O}(3)$ and its various subgroups. The isometry group can be thought of as being composed of the orientation preserving rotations in $\text{Isom}_+(S^2) = \text{SO}(3)$ and reflections \mathcal{R} via the direct product $\text{O}(3) = \text{SO}(3) \times \mathcal{R}$. $\text{SO}(3)$, in turn, is generated by $\text{SO}(2)$ rotations around any two non-parallel axes, which is used in the Euler angle parametrization. See the main text for more relevant subgroups and their relations.

We start by describing the sphere’s geometry in Section 17.1. Section 17.2 discusses fully $\text{SO}(3)$ and $\text{O}(3)$ -equivariant spherical GM-convolutions, which rely on $\text{SO}(2)$ or $\text{O}(2)$ -structures as shown in Fig. 17.2a. Globally $\text{SO}(2)$ and $\text{O}(2)$ -equivariant spherical CNNs, corresponding to the $\{e\}$ -structure in Fig. 17.2b or the corresponding \mathcal{R} -structure, respectively, are reviewed in Section 17.3. Section 17.4 focuses on icosahedral approximations of spherical convolutions, which allow for compute-efficient implementations since the icosahedron is piecewise flat and admits regular sampling grids; see Fig. 17.5. The $\text{SO}(2)$ -structure and $\{e\}$ -structure in Figs. 17.2a and 17.2b are hereby approximated by the C_6 -structure and the $\{e\}$ -structures in Figs. 17.6c and 17.6a or 17.6b, respectively.

17.1 Geometry of the 2-sphere S^2

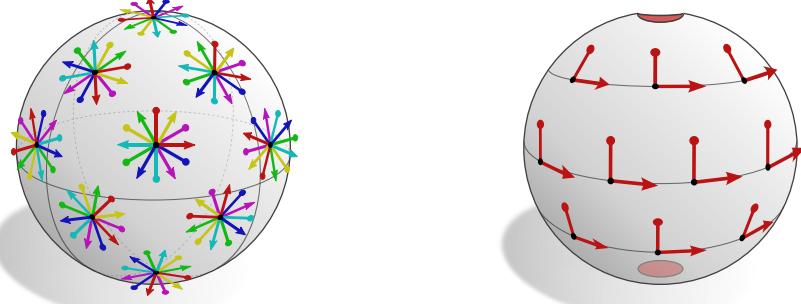
As a basis for our discussion of spherical CNNs, this section discusses the differential geometry of the (unit) sphere $M = S^2$. It is usually defined as the subset

$$S^2 := \{p \in \mathbb{R}^3 \mid \|p\| = 1\} \quad (17.1)$$

of those points in \mathbb{R}^3 that have unit distance from the origin. As an embedded surface, it inherits a Riemannian metric (first fundamental form) from the embedding space \mathbb{R}^3 . When interpreting the tangent spaces $T_p M$ literally as those two-dimensional subspaces of \mathbb{R}^3 that contain all tangent vectors at $p \in S^2$, then the metric, exponential maps, parallel transporters, frames and gauges can all be expressed in terms of usual vector space operations in \mathbb{R}^3 . Before coming to these concrete expressions, which come handy when implementing spherical CNNs, we discuss some properties of the sphere from a more abstract angle.

The isometry group of the sphere is given by

$$\text{Isom}(S^2) = \text{O}(3), \quad (17.2)$$



(a) $\text{SO}(2)$ -structure SOM on the 2-sphere $M = S^2$, preserved by general three-dimensional rotations in $\text{Isom}_{\text{SOM}} = \text{SO}(3)$.

(b) $\{e\}$ -structure $\{e\}M$ on a punctured 2-sphere $M = S^2 \setminus \{n, s\}$, preserved by azimuthal rotations in $\text{Isom}_{\{e\}M} = \text{SO}(2)$.

Figure 17.2: Common G -structures underlying spherical CNNs. Topological obstructions prevent a reduction of the 2-sphere’s structure group beyond $G = \text{SO}(2)$. Fig. 17.2a shows the standard $\text{SO}(2)$ -structure on S^2 , which is in agreement with the embedding metric (Eq. (17.9)) induced from the inner product of \mathbb{R}^3 . It is invariant under rotations in $\text{Isom}_{\text{SOM}} = \text{SO}(3)$, implying the rotation equivariance of the corresponding GM -convolution. Note that the fibers $G_p M$ and $G_q M$ at different points p and q are isomorphic but not canonically so – frame colors in the visualization seem to suggest such an isomorphism, however, they are randomly chosen and carry no meaning. Fig. 17.2b shows a sphere that is punctured at two antipodal poles. This turns the sphere into a topological cylinder $S^2 \setminus \{n, s\} \cong S^1 \times (-\frac{\pi}{2}, \frac{\pi}{2})$ with sphere like metric – which allows for a complete reduction to a trivial structure group. The figure shows the most prominent choice of $\{e\}$ -structure, which corresponds to orthonormal frames that are aligned with the coordinate grid of spherical coordinates; cf. Fig. 17.3. As this $\{e\}$ -structure is invariant under azimuthal rotations around the polar axis, the corresponding GM -convolutions are $\text{Isom}_{\{e\}M} = \text{SO}(2)$ -equivariant. Note that the puncturing of the sphere is just a means of hiding the models’ discontinuity at the poles.

i.e. three-dimensional rotations and reflections, which are visualized in Fig. 17.1. The action of any isometry $\phi \in \text{O}(3)$ coincides with its usual action on \mathbb{R}^3 via matrix multiplication, restricted to the embedded sphere $S^2 \subset \mathbb{R}^3$. Note that this yields indeed a well defined action on S^2 since $\text{O}(3)$ consists by definition of all distance and angle preserving linear maps, and thus preserves the sphere. As the sphere is orientable, it comes with a subgroup of orientation preserving isometries

$$\text{Isom}_+(S^2) = \text{SO}(3), \quad (17.3)$$

consisting of all three-dimensional rotations. Further subgroups that are relevant in the deep learning context are the following: any choice of rotation axis determines a subgroup of two-dimensional rotations, isomorphic to $\text{SO}(2)$, and all of these subgroups are conjugated to each other. Similarly, any choice of two-dimensional subspace of \mathbb{R}^3 corresponds to a subgroup of reflections over this plane, which is isomorphic to \mathcal{R} . The subgroups of two-dimensional rotations around two non-parallel rotation axes generate $\text{SO}(3)$, which relates to the Euler angle parametrization of $\text{SO}(3)$. A choice of reflection plane and any rotation axis within this plane generates the semidirect product subgroup $\text{O}(2) = \text{SO}(2) \rtimes \mathcal{R}$. If the rotation axis is instead chosen to be orthogonal to the reflection plane, the two-dimensional rotations and reflections commute, and generate therefore subgroups isomorphic to the direct product $\text{SO}(2) \times \mathcal{R}$. $\text{O}(3)$ has furthermore discrete subgroups, the most practically relevant

of which are the symmetry groups of the platonic solids, for instance of the icosahedron, shown in Fig. 17.5.¹

$O(3)$ acts transitively (Def. B.3.8) on the sphere, that is, for any two points p and q of S^2 , there exists at least one isometry $\phi \in O(3)$ such that $q = \phi(p)$. The actions of $O(3)$ on S^2 are not fixed point free (Def. B.3.10): any point $p \in S^2$ is stabilized (Def. B.3.6) by the subgroup $\text{Stab}_p \cong O(2) < O(3)$, consisting of rotations and reflections around the axis spanned by p in \mathbb{R}^3 . Together, these two properties imply that the sphere is a homogeneous space (Def. B.3.11) of $O(3)$ and is algebraically realized as the quotient space

$$O(3)/O(2) \cong S^2, \quad (17.4)$$

which consists of cosets of the form $\phi.O(2)$. A similar statement holds for $SO(3)$, which has stabilizer subgroups $\text{Stab}_p \cong SO(2) < SO(3)$ and thus

$$SO(3)/SO(2) \cong S^2. \quad (17.5)$$

With these relations, Theorem 13.3.3 proves that any $O(3)$ or $SO(3)$ -equivariant kernel field transform on S^2 is equivalent to a GM -convolution with G being $O(2)$ or $SO(2)$, respectively. This result is in line with the classical viewpoint of group equivariant CNNs on homogeneous spaces [56] – the precise relation between the two is clarified in Theorem 17.2.2 below. Recall that isometries preserve the Riemannian metric by definition. That $O(3)$ acts transitively on S^2 with stabilizer $O(2)$ implies therefore that the Riemannian geometry of S^2 “looks similar” at each point and in each direction and orientation – S^2 is a maximally symmetric space.

As a Riemannian manifold, S^2 comes by design with an $O(2)$ -structure. A restriction to right-handed frames, which is possible since the sphere is orientable, yields the $SO(2)$ -structure in Fig. 17.2a, which is preserved by rotations in $SO(3)$. One can show that these two G -structures OM and SOM are as principal bundles isomorphic to $O(3)$ and $SO(3)$, respectively. The specific isomorphism is hereby given by a choice of frame from the G -structure, that is to be identified with the identity group element.

The hairy ball theorem states that no continuous vector field exists on S^2 , which implies in particular that no (continuous) $\{e\}$ -structure can exist. A reduction of the structure group beyond $SO(2)$ requires therefore a change in the topology of the manifold. For example, puncturing the sphere at an arbitrary point $p \in S$ yields a surface that is homeomorphic to the Euclidean plane, and is therefore parallelizable.² Puncturing the sphere at two arbitrarily chosen antipodal points, as shown in Fig. 17.2b, turns the topology of the sphere into that of a cylinder and thus allows for $\{e\}$ -structures. The most common choice of $\{e\}$ -structure on the punctured sphere $S^2 \setminus \{n, s\}$ is the $SO(2)$ -invariant $\{e\}$ -structure in Figs. 17.2b and 17.3. Its frames

$$\left[\frac{\partial}{\partial \theta}, \frac{1}{\cos(\theta)} \frac{\partial}{\partial \varphi} \right] \quad (17.6)$$

are aligned with the usual spherical coordinates, which are in physics conventions (i.e. with φ and θ denoting the azimuthal angle and inclination against the xy -plane, respectively) given by the following surjective, 2π -periodic map:

$$\chi : \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \times \mathbb{R} \rightarrow S^2 \setminus \{n, s\}, \quad (\theta, \varphi) \mapsto \begin{pmatrix} \cos(\theta) \cos \varphi \\ \cos(\theta) \sin \varphi \\ \sin(\theta) \end{pmatrix} \quad (17.7)$$

¹An exhaustive list of all finite subgroups of $SO(3)$ can be found at nLab.

²This process corresponds for instance to the stereographic projection of the sphere.

Some $\{e\}$ -steerable CNNs are implemented by representing feature fields on $S^2 \setminus \{n, s\}$ in spherical coordinates; see Section 17.3 below. As the coordinate map χ is *not* isometric, these methods require an alternative metric (or $\{e\}$ -structure) on the coordinates $(-\pi/2, \pi/2) \times \mathbb{R} \subset \mathbb{R}^2$; see the stretched frames in Fig. 17.3 (right).

Since S^2 is compact, it is geodesically complete. The geodesics are given by the great circles of the sphere, i.e. those circles that correspond to the intersection of the sphere with a plane through the origin of \mathbb{R}^3 . The exponential maps $\exp_p(v)$ follow these great circles through p in direction of v for a distance of $\|v\|$. Logarithmic maps $\log_p(q)$ are therefore for all points $q \in S^2 \setminus p$, which are not antipodal to p , given by the unique vector in the shorter direction along the great circle through p and q , with $\|\log_p(q)\|$ given by the arc-length along this path. Geodesics between antipodal points p and $-p$ are not unique, such that the logarithmic map does not exist.

Explicit geometry of S^2 as embedded surface in \mathbb{R}^3

As stated above, the tangent spaces of $S^2 \subset \mathbb{R}^3$ are in the classical differential geometry of surfaces defined as two-dimensional subspaces of the embedding space \mathbb{R}^3 . A specific tangent space $T_p M$ at $p \in S^2$ is in this interpretation given by

$$T_p M = \{v \in \mathbb{R}^3 \mid \langle p, v \rangle = 0\} \subset \mathbb{R}^3, \quad (17.8)$$

i.e. the space of all vectors that are orthogonal to the surface normal at p , which coincides for the sphere with p itself. Note that, despite being expressed relative to the standard frame of \mathbb{R}^3 , these tangent vectors are coordinate free object in the sense that they are not described by 2-tuples of coefficients $v^A \in \mathbb{R}^2$ relative to some gauge $\psi_{TM,p}^A$ of $T_p M$. The identification of tangent spaces with subspaces of the embedding space allows to express many of the abstract algebraic relations in terms of vector space operations on \mathbb{R}^3 . In the remainder of this section we will state such expressions for the metric, exponential and logarithmic maps, frames, gauges, Levi-Civita transporters along geodesics and induced gauge transformations.

By definition, S^2 inherits its Riemannian metric from the embedding space. This induced metric is for any $v, w \in T_p M \subset \mathbb{R}^3$ given by

$$\eta_p(v, w) := \langle v, w \rangle_{\mathbb{R}^3}, \quad (17.9)$$

i.e. the standard inner product of \mathbb{R}^3 , restricted to $T_p M$. To reduce clutter, we drop the subscript \mathbb{R}^3 in the notation $\langle \cdot, \cdot \rangle_{\mathbb{R}^3}$ in the remainder of this section.

The exponential map \exp_p maps vectors $v \in T_p M$ to points $q = \exp_p(v) \in S^2$ at a distance of $\|v\|$ along the great circle in direction of v . Lying on the same great circle, p and q relate via a rotation by an angle of $\alpha = \|v\|/r = \|v\|$ around the rotation axis $a = \frac{p \times v}{\|p \times v\|} = \frac{p \times v}{\|v\|}$, where the equations simplify since the sphere has unit radius $r = \|p\| = 1$ and the vectors p and v are orthogonal in \mathbb{R}^3 . Using Rodrigues' rotation formula, $q = p \cos(\alpha) + (a \times p) \sin(\alpha) + a \langle a, p \rangle (1 - \cos(\alpha))$, together with the orthogonality $\langle a, p \rangle = 0$ and $a \times p = \frac{1}{\|v\|}(p \times v) \times p = \frac{1}{\|v\|}(\langle p, p \rangle v + \langle p, v \rangle p) = \frac{v}{\|v\|}$, this leads to the explicit expression

$$\exp_p : \mathbb{R}^3 \supset T_p M \rightarrow S^2 \subset \mathbb{R}^3, \quad v \mapsto \exp_p(v) = p \cos(\|v\|) + \frac{v}{\|v\|} \sin(\|v\|) \quad (17.10)$$

for the exponential map.

An explicit expression of the logarithmic map is found along the same line of reasoning: the norm of $\log_p(q)$, where $q \in S^2 \setminus -p$, coincides with the rotation angle $\alpha = \arccos(\langle p, q \rangle)$. Its direction is given by the direction tangent to the great circle, which may be expressed in terms of the normalized projection $\frac{v}{\|v\|} = \frac{q - \langle p, q \rangle p}{\|q - \langle p, q \rangle p\|}$ of q on $T_p M$. Overall, the logarithmic map is therefore instantiated as

$$\log_p : S^2 \setminus -p \rightarrow B_{T_p M}(0, \pi), \quad q \mapsto \log_p(q) = \arccos(\langle p, q \rangle) \frac{q - \langle p, q \rangle p}{\|q - \langle p, q \rangle p\|}, \quad (17.11)$$

where $B_{T_p M}(0, \pi) \subset T_p M \subset \mathbb{R}^3$ denotes the open ball of injectivity-radius π around the origin of $T_p M$.

Reference frames on S^2 are by definition just 2-tuples of linearly independent tangent vectors. When expressing the axes of a reference frame explicitly as vectors in the embedding space \mathbb{R}^3 , this frame can be identified with the 3×2 rank 2 matrix

$$[e_1^A, e_2^A] = \begin{bmatrix} e_{1,1}^A & e_{2,1}^A \\ e_{1,2}^A & e_{2,2}^A \\ e_{1,3}^A & e_{2,3}^A \end{bmatrix} =: E_p^A \in \mathbb{R}^{3 \times 2}. \quad (17.12)$$

It defines the vector space isomorphism

$$E_p^A = [e_1^A, e_2^A] : \mathbb{R}^2 \rightarrow T_p M, \quad v^A \mapsto E_p^A v^A = v_1^A e_1^A + v_2^A e_2^A \quad (17.13)$$

from vector coefficients to coordinate free tangent vectors. The tangent spaces $T_p M$ are therefore exactly the image of E_p^A .

The corresponding gauges $\psi_{TM,p}^A : T_p M \rightarrow \mathbb{R}^2$ are technically just the inverses of the frames, when being interpreted as maps $E_p^A : \mathbb{R}^2 \rightarrow T_p M$. In contrast, when being interpreted as 3×2 matrices that map \mathbb{R}^2 non-surjectively to \mathbb{R}^3 , E_p^A is non-invertible but only admits a pseudo-inverse

$$(E_p^A)^+ := ((E_p^A)^\top E_p^A)^{-1} (E_p^A)^\top \in \mathbb{R}^{2 \times 3}. \quad (17.14)$$

Geometrically, this matrix acts by 1) projecting vectors in \mathbb{R}^3 to the image of E_p^A , which is just $E_p^A(\mathbb{R}^2) = T_p M \subset \mathbb{R}^3$, and 2) applying the inverse of the isomorphism $E_p^A : \mathbb{R}^2 \rightarrow T_p M$ on this subspace. This means that the pseudo-inverse is indeed the inverse of E_p^A on the tangent space, implying that the gauge map is given by

$$\psi_{TM,p}^A : T_p M \rightarrow \mathbb{R}^2, \quad v \mapsto (E_p^A)^+ v. \quad (17.15)$$

Written out, the gauge map acts according to

$$\begin{aligned} \psi_{TM,p}^A(v) &= \begin{pmatrix} \langle e_1^A, e_1^A \rangle & \langle e_1^A, e_2^A \rangle \\ \langle e_2^A, e_1^A \rangle & \langle e_2^A, e_2^A \rangle \end{pmatrix}^{-1} \begin{pmatrix} \langle e_1^A, v \rangle \\ \langle e_2^A, v \rangle \end{pmatrix} \\ &= \frac{1}{\langle e_1^A, e_1^A \rangle \langle e_2^A, e_2^A \rangle - \langle e_1^A, e_2^A \rangle \langle e_2^A, e_1^A \rangle} \begin{pmatrix} \langle e_2^A, e_2^A \rangle & -\langle e_1^A, e_2^A \rangle \\ -\langle e_2^A, e_1^A \rangle & \langle e_1^A, e_1^A \rangle \end{pmatrix} \begin{pmatrix} \langle e_1^A, v \rangle \\ \langle e_2^A, v \rangle \end{pmatrix}. \end{aligned} \quad (17.16)$$

Note that, in general, $\langle e_i^A, v \rangle \neq v_i^A$. However, if (and only if) E_p^A is an orthonormal frame, i.e. for $G \leq O(2)$, the gauge map is simply given by the projection of the tangent vector on

the frame axes:

$$\psi_{TM,p}^A(v) = (E_p^A)^\top v = \begin{pmatrix} \langle e_1^A, v \rangle \\ \langle e_2^A, v \rangle \end{pmatrix} \quad \text{for orthonormal frames, i.e. } \langle e_i^A, e_j^A \rangle = \delta_{ij} \quad (17.17)$$

The explicit expression for the coordinate free Levi-Civita transporters *along geodesics* is similar to that of the exponential map, with the difference that Rodrigues' rotation formula is not applied to rotate the source to the target point but tangent vectors between source and target. Let γ be the shortest geodesic between $p \in S^2$ and $q \in S^2 \setminus p$. The rotation from p to q along this geodesic is then given by the axis $a = p \times q$ and angle $\alpha = \arccos(\langle p, q \rangle)$. In terms of these quantities, the Levi-Civita transport of an embedded tangent vector $v \in T_p M \subset \mathbb{R}^3$ along the geodesic γ is given by the rotated vector

$$\mathcal{P}_{TM,\gamma}(v) = v \cos(\alpha) + (a \times v) \sin(\alpha) + (a \langle a, v \rangle) (1 - \cos(\alpha)) \quad (17.18)$$

in $T_q M \subset \mathbb{R}^3$. Relative to gauges $\psi_{TM,p}^A$ and $\psi_{TM,q}^{\tilde{A}}$ at the start point p and end point q of the geodesic, this transporter is expressed by the group element

$$g_\gamma^{A\tilde{A}} = \psi_{TM,p}^A \circ \mathcal{P}_{TM,\gamma} \circ (\psi_{TM,q}^{\tilde{A}})^{-1} = (E_p^A)^+ \circ \mathcal{P}_{TM,\gamma} \circ E_q^{\tilde{A}}. \quad (17.19)$$

Isometry induced gauge transformations are relative to the explicit reference frames similarly given by the following matrix multiplication:

$$g_\phi^{A\tilde{A}}(p) = \psi_{TM,\phi(p)}^A \circ \phi \circ (\psi_{TM,p}^{\tilde{A}})^{-1} = (E_{\phi(p)}^A)^+ \phi E_p^{\tilde{A}} \quad (17.20)$$

17.2 Fully rotation equivariant spherical CNNs

This section discusses the fully SO(3) or O(3)-equivariant spherical convolutions that are listed in rows (35-37) of Table 14.1. They can all be understood as specific instances of *GM*-convolutions on either the SO(2)-structure in Fig. 17.2a or the corresponding O(2)-structure, which is additionally closed under frame reflections.

Instead of organizing this discussion in terms of the considered structure groups and group representations, we assort the models by the theoretical frameworks in which they are developed: Kicanaoglu et al. [148] define a pixel grid on the sphere and formulate the convolution directly as *GM*-convolution, that is, in terms of gauges, steerable kernels and feature vector transporters. An alternative framework is that of graph convolutions on spherical pixel meshes [228, 338]. Such graph convolutions correspond to *GM*-convolutions with isotropic kernels. They map therefore between (directionally insensitive) scalar fields. Lastly, we come to implementations that consider (steerable) convolution kernels on S^2 instead of our kernels on the tangent spaces [83, 54, 163, 86]. Theorem 17.2.1 proves that such spherical steerable kernels can be identified with *G*-steerable kernels on the tangent spaces, when being expressed in geodesic normal coordinates. Based on this result, we prove in Theorem 17.2.2 that convolutions with spherical kernels are equivalent to our *GM*-convolutions. For completeness, we need to mention that such models are typically implemented in the spectral domain. We do not focus on this viewpoint but refer the interested reader to the review by Esteves [82].

Spherical GM-convolutions: We start with the spherical CNN by Kicanaoglu et al. [148] since its formulation agrees precisely with our more general theory when being applied to the spherical geometry. The authors assume the $\text{SO}(2)$ -structure from Fig. 17.2a, and therefore $\text{SO}(2)$ -steerable feature fields and convolution kernels. Feature fields are discretized in terms of feature vectors that are assigned to a sampling grid on the sphere. While the method is in principle independent from the particular sampling scheme, the authors propose to discretize the spherical geometry by an icosahedron mesh. This mesh is constructed by taking an embedded icosahedron, repeatedly subdividing its faces as shown in Fig. 17.5, and finally projecting the grid vertices radially to the sphere, i.e. to unit norm. The sampled feature fields are numerically represented by a set of coefficient vectors $f^A(p) \in \mathbb{R}^c$ at the grid vertices p , which are expressed relative to some arbitrarily chosen right-handed orthonormal frames $[e_1^A, e_2^A]$ at the vertices.³ In practice, the frames are represented by a single tangent vector of unit norm, from which the second frame vector follows uniquely since the frames are right-handed.

To compute the coordinate independent convolution $[K \star f](p)$ from Eq. (9.39), Kicanaoglu et al. [148] need to contract the $\text{SO}(2)$ -steerable kernel K with the transporter pullback $[\text{Exp}_p^* f]^A$ of the feature field f , Eq. (9.21). As usual in deep learning, K is hereby assumed to be compactly supported, such that it covers only a few vertices in a one-ring or two-ring neighborhood \mathcal{N}_p around a center vertex p . In the continuous theory, the transporter pullback takes features from all points $\exp_p(\psi_{TM,p}^A)^{-1}(v)$ for $v \in \mathbb{R}^2$ and transports them back to p . In practice, the feature fields are only sampled at the grid vertices q , which correspond to the tangent vector coefficients $v_{pq}^A = \psi_{TM,p}^A \log_p(q) \in \mathbb{R}^2$ relative to gauge A at vertex p .⁴ The logarithmic maps $\log_p(q)$ are thereby computed as defined in Eq. (17.11). The Levi-Civita transporters $\rho(g_{p \leftarrow q}^{A\tilde{A}})$ along the geodesics from q to p are in principle given by Eq. (17.19). Since the frames are all right-handed and orthonormal, and since the transport corresponds to the Levi-Civita connection on S^2 , the group elements $g_{p \leftarrow q}^{A\tilde{A}}$ are $\text{SO}(2)$ -valued. They are therefore fully determined by the angle between the transported first frame axis $P_{TM,\gamma}(e_1^{\tilde{A}})$ from q and the first frame axis e_1^A at p . With these ingredients at hand, the authors propose to approximate the continuous convolution integral by the discrete sum

$$[K \star f]^A(p) = \int_{\mathbb{R}^2} K(v) [\text{Exp}_p^* f]^A(v) dv \approx \sum_{q \in \mathcal{N}_p} K(v_{pq}^A) \rho(g_{p \leftarrow q}^{A\tilde{A}}) f^{\tilde{A}}(q) \quad (17.21)$$

over neighboring mesh nodes. The missing normalization factor can be thought of as being absorbed in the learnable parameters $w_i \in \mathbb{R}$ of the $\text{SO}(2)$ -steerable convolution kernel $K = \sum_i w_i K_i$. As an alternative to this naive approximation, the authors propose an optimized quadrature integration scheme, which is empirically shown to improve the model's $\text{SO}(3)$ isometry equivariance.

The model is in Table 14.1 listed as processing feature fields that transform according to the regular representation of $\text{SO}(2)$. In their implementation, Kicanaoglu et al. [148] consider irrep fields of $\text{SO}(2)$ in the convolutions. A change of basis before and after the convolutions transforms these feature fields to regular feature fields, which are then acted on by pointwise nonlinearities like e.g. ReLU. The infinite-dimensional regular representation of

³This corresponds to an independent choice of gauge $\psi_{TM,p}^{Ap}$ on any open neighborhood U^{Ap} of each vertex p .

⁴If the exponential map is not restricted to the injectivity radius, each vertex q is represented by multiple tangent vectors. This is in practice no issue since the kernel is assumed to be locally supported within the injectivity radius.

$\text{SO}(2)$ is hereby approximated by regular representations of discrete cyclic subgroups C_N , whose irreps are just the irreps of $\text{SO}(2)$ up to a bandlimiting frequency of $\lfloor N/2 \rfloor$; see e.g. Appendix F.2 of [322]. The change of basis between the representations is in this specific case just the usual discrete Fourier transform.

Spherical graph convolutions: The spherical CNNs by Perraudin et al. [228] and Yang et al. [338], which are listed in row (37) of Table 14.1, are based on conventional graph convolutions [152]. Pixel meshes on the sphere are hereby interpreted as graphs. The graph convolutional networks process signals on the sphere by multiplying them with degree κ polynomials $\sum_{k=0}^{\kappa} w_k L^k$ of the graph's Laplacian matrix L , where $w_k \in \mathbb{R}$ are trainable parameters. Since the Laplacian matrix has non-zero entries only for adjacent nodes, the k -th order term affects only the k -hop neighborhood around each node. On a regular mesh with unweighted graph edges, the contribution of a neighboring node q to the accumulated feature at p depends only on their graph distance ("radius"), but not on the particular neighbor ("direction"). The graph convolution applies therefore in such cases *isotropic* kernels on the graph. The considered pixel graph on the sphere satisfy these properties approximately. As their *embedding* on the sphere is furthermore such, that the nodes are geodesically approximately equidistant, the topological isotropy of the graph convolution kernels corresponds to their metric isotropy on the sphere.

The isometry group $\text{O}(3)$ of the sphere induces $\text{O}(2)$ -valued gauge transformation, that is, it acts by moving patterns to a new location and in a new orientation. Due to the convolutional weight sharing and the isotropy of the kernels, the graph convolutions are trivially isometry equivariant. As already argued in Eq. (16.7), isotropic kernels are in our framework recovered as $\text{O}(2)$ -steerable kernels that map between *scalar fields*. The $\text{O}(3)$ -equivariance of the convolution is in our theory explained by the $\text{O}(3)$ -invariance of the sphere's $\text{O}(2)$ -structure.

Spherical convolutions with kernels on S^2 : As a homogeneous space, the sphere admits group (or quotient space) convolutions [162] and more general steerable convolutions on homogeneous spaces [56].⁵ Instead of defining the convolution kernels on the tangent spaces or on graph neighborhoods, these approaches define kernels immediately as matrix-valued function on the sphere, that is, as

$$\kappa : S^2 \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} . \quad (17.22)$$

Cohen et al. [56] showed that these kernels are required to satisfy a symmetry constraint in order to guarantee the equivariance of the convolution. We argue in the following that such kernels on S^2 are equivalent to G -steerable kernels on the tangent spaces (Theorem 17.2.1), which implies that the spherical CNNs covered in [56] and [162] can be viewed as *GM*-convolutions (Theorem 17.2.2). The identification between the two kinds of kernels is hereby made by pulling the spherical kernels via the exponential map back to the tangent spaces. Before explaining this operation, we briefly discuss the models proposed in [54, 83, 86, 163] as specific instances of spherical convolutions with spherical kernels. For a more details on these models, specifically on their formulation in Fourier space, we refer the reader to the comprehensive review by Esteves [82].

We start our discussion with the group convolutional spherical CNN by Cohen et al. [54], listed in row (36) of Table 14.1. This model processes stacks of c_{in} *scalar* fields

$$f : S^2 \rightarrow \mathbb{R}^{c_{\text{in}}} \quad (17.23)$$

⁵A more general review of convolutions on homogeneous spaces is found in Appendix F.

on the sphere by matching them with spherical kernels, Eq. (17.22), in any $\text{SO}(3)$ transformed pose. In equations, this operation is defined as

$$[\kappa \star_{S^2} f](\phi) := \int_{S^2} \kappa(\phi^{-1}(p)) f(p) dp \quad \phi \in \text{SO}(3). \quad (17.24)$$

Note that the resulting feature map is viewed as a stack of c_{out} scalar functions on the symmetry group $\text{SO}(3)$. Such feature maps of the form $f : \text{SO}(3) \rightarrow \mathbb{R}^{c_{\text{in}}}$ (with the new number of input channels corresponding to the previous layer's output channels) are processed further by group convolutions of the form

$$[\kappa \star_{\text{SO}(3)} f](\phi) := \int_{\text{SO}(3)} \kappa(\phi^{-1}\omega) f(\omega) d\omega \quad \phi \in \text{SO}(3), \quad (17.25)$$

where $\kappa : \text{SO}(3) \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ is now a matrix-valued function on $\text{SO}(3)$ and $d\omega$ is the Haar measure on $\text{SO}(3)$. From the viewpoint of steerable CNNs on homogeneous spaces [56] and GM -convolutions, scalar functions on $\text{SO}(3)$ are viewed as feature fields on $S^2 \cong \text{SO}(3)/\text{SO}(2)$, that transform according to the *regular representation* of the fibers (stabilizer subgroups) $\text{SO}(2)$; cf. Section 4.5 and Appendix J. The initial convolution in Eq. (17.24) applies in this interpretation $\text{SO}(2)$ -steerable kernels between scalar and regular fields, while the group convolution in Eq. (17.25) applies $\text{SO}(2)$ -steerable kernels between regular fields on S^2 .

Esteves et al. [83] apply spherical convolutions as in Eq. (17.24) with the additional assumption that the kernels are *zonal*, that is, invariant under $\text{SO}(2)$ rotations around the polar axis; cf. Fig. 16.5. While the integral technically still gives responses in $\text{SO}(3)$, the kernel symmetry implies that these responses are constant on the fibers $\text{SO}(2)$ of $\text{SO}(3)$, when being interpreted as bundle over S^2 . The resulting feature fields are therefore identified as scalar fields on S^2 , which allows for a repeated application of this type of convolution. Note that the zonal symmetry of the kernel is consistent with the steerability kernel constraint between scalar fields (trivial representations) that we encountered before in Eq. (16.6) and the upper left entry of Table 5.2. As already discussed in the previous Section 16.3, this constraint is equivalent to the $\text{O}(2)$ -steerability constraint between scalar fields in Eq. (16.7), which implies that the model of Esteves et al. [83] is actually $\text{O}(3)$ -equivariant. It is in spirit similar to the spherical graph convolutions discussed above, but is derived from a different viewpoint and is discretized differently in the implementation.

Esteves et al. [86] generalize this model from scalar fields to general *spin weighted spherical functions*. These functions depend not only on the position $p \in S^2$ on the sphere, but in addition on the particular choice of right-handed, orthonormal reference frame at that point. They are associated to the *irreps* ρ_s of $\text{SO}(2)$, where the integer $s \in \mathbb{Z}$ is denoted as the functions' spin weight.⁶ Their values for the different frames $\text{SO}_p M$ of the $\text{SO}(2)$ -structure SOM are constrained such that gauge transformations of the frame by $g \in \text{SO}(2)$ lead to a transformation of the function value by $\rho_s(g)$. In equations, they are therefore defined by⁷

$${}_s f : \text{SOM} \rightarrow \mathbb{C} \quad \text{such that} \quad {}_s f([e_1, e_2] \triangleleft g) = \rho_s(g) {}_s f([e_1, e_2]) \quad (17.26)$$

$$\forall [e_1, e_2] \in \text{SO}_p M, g \in \text{SO}(2);$$

see [24] for more details and alternative definitions. Note the similarity of this symmetry constraint to the equivalence relation

$$[[e_i]_{i=1}^2 \triangleleft g, f] \sim_{\rho_s} [[e_i]_{i=1}^2, \rho_s(g)f] \quad (17.27)$$

⁶One can generalize this concept to spin representations, labeled by half-integer spin weights.

⁷A real-valued implementation would instead consider spin weighted functions of the form ${}_s f : \text{SOM} \rightarrow \mathbb{R}^{\dim(\rho_s)}$, where ρ_s are the irreps of $\text{SO}(2)$ over the real numbers.

from Eq. (11.42), which is underlying the definition of associated bundles. Spin weighted spherical functions are indeed equivalent to sections of the associated bundles $(\text{SOM} \times \mathbb{C})/\sim_{\rho_s}$; see for instance Proposition 1.6.3 in [326]. They appear in our theory simply as $\text{SO}(2)$ -irrep fields, including scalar fields for $s = 0$ and vector fields for $s = 1$. The neural networks proposed by Esteves et al. [86] convolve spin weighted features with spin weighted kernels on the sphere. This operation corresponds to a convolution with $\text{SO}(2)$ -steerable kernels where ρ_{in} and ρ_{out} are irreps – these kernels were derived in Section 5.3 and are listed in Table 5.2.

The models in [54, 83, 86] are initially formulated in the spatial domain, i.e. as processing functions on S^2 as discussed above. They are, however, implemented in the spectral domain, which is possible thanks to generalized convolution theorems on S^2 and on $\text{SO}(3)$ [191, 162, 157]. Kondor et al. [163] generalize these approaches, proposing a model that is based on learned linear combinations of all feature fields' Fourier modes of the same frequency. The authors argue that this approach covers the full space of $\text{SO}(3)$ -equivariant linear maps between feature fields on the sphere. On the other hand, Cohen et al. [56] show that any such map can in the spatial domain be written as a convolution with $\text{SO}(2)$ -steerable spherical kernels. A notable property of the model proposed by Kondor et al. [163] is that it operates fully in Fourier space: instead of transforming back to the spatial domain and applying pointwise nonlinearities like ReLUs there, as done in the previous approaches, the authors compute the tensor product (Def. B.5.4) between all feature fields and decompose them subsequently via the Clebsch-Gordan decomposition (Def. B.5.17) back into irreducible features (Fourier modes). This is computationally beneficial, however, comes at the expense of losing the locality of the nonlinearities. Certain learning tasks, especially in the natural sciences, might benefit from such nonlinearities since physical interactions are often described by tensor products.

As argued in [56, 55], all of these models can be viewed as applying steerable kernels on S^2 that map between scalar fields [83], regular feature fields [54] or irrep fields [86, 163]. In the remainder of this section and Appendix L we show that they can as well be viewed as GM -convolutions. The claim that spherical convolutions with steerable kernels on S^2 are equivalent to GM -convolutions is thereby made precise in Theorem 17.2.2. This theorem relies crucially on Theorem 17.2.1, which establishes an isomorphism between the spherical steerable kernels and G -steerable kernels on the tangent spaces.

Let \mathcal{I} be any transitive isometry group of the sphere, i.e. $\mathcal{I} = \text{O}(3)$ or $\mathcal{I} = \text{SO}(3)$. Cohen et al. [56] describe \mathcal{I} -equivariant spherical convolutions in terms of Stab_n -steerable spherical kernels $\kappa : S^2 \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$, where $\text{Stab}_n < \mathcal{I}$ is the stabilizer subgroup of any point $n \in S^2$, e.g. the north pole. As these kernels are defined on the sphere, which is topologically distinct from \mathbb{R}^2 , it is not directly possible to define an isomorphism between them and G -steerable kernels. However, as the south pole $-n$ is a set of measure zero, we can replace the integration domain S^2 of the spherical convolutions with $S^2 \setminus -n$ without changing the result. With this adaptation, the spherical steerable kernels of Cohen et al. [56] are defined as

$$\mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^{\text{Stab}_n} := \left\{ \kappa : S^2 \setminus -n \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \mid \kappa(\xi(p)) = \rho_{\text{out}}(g_\xi^{NN}(n)) \cdot \kappa(p) \cdot \rho_{\text{in}}(g_\xi^{XP}(p))^{-1} \right. \\ \left. \forall p \in S^2 \setminus -n, \xi \in \text{Stab}_n \right\}, \quad (17.28)$$

when being translated to our notation. Since the kernels are globally defined on the sphere, their values in $\mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ are expressed relative to potentially different gauges N at n , where the kernel is centered, P at $p \in S^2$, where the kernel contracts a feature $f^P(p) \in \mathbb{R}^{c_{\text{in}}}$ and X at $\xi(p)$, where this feature moves under the action of $\xi \in \text{Stab}_n$. This kernel constraint

relates all kernel values that lie on the orbits $\text{Stab}_n.p = \{\xi(p) \mid \xi \in \text{Stab}_n\}$ via their isometry induced gauge transformations $g_\xi^{XP}(p)$ and $g_\xi^{NN}(n)$; see Eqs. (13.38) and (13.39).⁸ Our equivalent G -steerable kernels, where $G \cong \text{Stab}_n$, is given by

$$\mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^{G, B_{\mathbb{R}^2}(0, \pi)} := \left\{ K : B_{\mathbb{R}^2}(0, \pi) \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \mid K(gv) = \rho_{\text{out}}(g) \cdot K(v) \cdot \rho_{\text{in}}(g)^{-1} \right. \\ \left. \forall v \in B_{\mathbb{R}^2}(0, \pi), g \in G \right\}. \quad (17.29)$$

The kernel domain is hereby restricted from \mathbb{R}^2 to the open ball $B_{\mathbb{R}^2}(0, \pi) := \{v \in \mathbb{R}^2 \mid \|v\| < \pi\}$ of radius π around the origin of \mathbb{R}^2 – which can via the exponential map be identified with $S^2 \setminus -n$. Note that $\mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^{G, B_{\mathbb{R}^2}(0, \pi)}$ is well defined since $\text{Stab}_n \cong G$ contains isometries, implying $G = O(2)$ or $G = SO(2)$, under whose action $B_{\mathbb{R}^2}(0, \pi)$ is closed. We furthermore dropped the determinant factor from the more general G -steerability constraint in Eq. (5.4) since $|\det g| = 1$ for $G \leq O(2)$. Our kernel constraint is considerably simpler than that of Cohen et al. [56] since it describes the kernel locally relative to a single gauge, instead of globally relative to an atlas of gauges. Note further that we dropped the smoothness assumption on the kernels, since the smoothness or continuity of feature fields is not discussed by Cohen et al. [56]. This property could easily be added by demanding that the G -steerable kernels converge to the same value for $\|v\|$ going to π , corresponding via the exponential map to the south pole.

The spaces of Stab_n -steerable kernels on $S^2 \setminus -n$ and G -steerable kernels on $B_{\mathbb{R}^2}(0, \pi)$ are isomorphic, that is, their kernels are identified by an invertible map Ω that respects the kernel constraints:

$$\begin{array}{ccc} \mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^{G, B_{\mathbb{R}^2}(0, \pi)} & \xrightarrow{\Omega} & \mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^{\text{Stab}_n} \\ & \xleftarrow{\Omega^{-1}} & \end{array} \quad (17.30)$$

This isomorphism (or rather its inverse Ω^{-1}) can be viewed as the analog of the *transporter pullback* of feature fields: it pulls the kernel values from points $\exp_n(\psi_{TM,n}^N)^{-1}v$ in $S^2 \setminus -n$ back to *geodesic normal coordinates* $v \in B_{\mathbb{R}^2}(0, \pi)$. To express the kernel values from all points $p \in S^2 \setminus -n$ relative to the same gauge, it applies Levi-Civita transporters $\rho(g_{n \leftarrow p}^{NP})$ from p along the geodesics to the north pole n . In addition, it rescales the kernel values by the Riemannian volume element $\sqrt{|\eta_p^{\partial/\partial v}|} := \sqrt{|\det(\eta_p(\frac{\partial}{\partial v_i}|_p, \frac{\partial}{\partial v_j}|_p)_{ij})|}$ relative to the geodesic normal coordinate system (coordinate chart) $v : S^2 \setminus -n \rightarrow B_{\mathbb{R}^2}(0, \pi)$, $p \mapsto v(p) := \psi_{TM,n}^N \log_n p$.⁹ The following theorem defines and proves the kernel space isomorphism formally.

⁸Cohen et al. [56] denote the isometry induced gauge transformations by $h(p, \xi)$ instead of $g_\xi^{XP}(p)$, assuming that the gauges X at $\xi(p)$ and P at p are the same. Their definition of $h(p, \xi)$ is similar to our Eq. (13.32).

⁹Note that the coordinate bases $[\frac{\partial}{\partial v_1}|_p, \frac{\partial}{\partial v_2}|_p]$ that are induced by the geodesic normal coordinates $v : S^2 \setminus -n \rightarrow B_{\mathbb{R}^2}(0, \pi)$ are for $G \leq O(2)$ *not* contained in GM . These bases play no role for the GM -convolution but appear only to correct for the Riemannian volume when integrating in geodesic normal coordinates over the sphere.

Theorem 17.2.1 (Spherical steerable kernels in geodesic coordinates). Let \mathcal{I} be any transitive isometry group of S^2 and let Stab_n be its stabilizer subgroup at the north pole $n \in S^2$. Given any choice of gauge $\psi_{TM,n}^N$ at this pole, let $G \leq \text{GL}(2)$ be the isomorphic structure group that represents Stab_n in coordinates according to $\text{Stab}_n \xrightarrow{\sim} G$, $\xi \mapsto \psi_{TM,n}^N \circ \xi_{*,TM} \circ (\psi_{TM,n}^N)^{-1}$. The space $\mathcal{K}_{\rho_{\text{in}},\rho_{\text{out}}}^{\text{Stab}_n}$ of Stab_n -steerable kernels on $S^2 \setminus -n$ by Cohen et al. [56] (Eq. (17.28)) is then isomorphic to the space $\mathcal{K}_{\rho_{\text{in}},\rho_{\text{out}}}^{G,B_{\mathbb{R}^2}(0,\pi)}$ of G -steerable kernels on the open ball $B_{\mathbb{R}^2}(0,\pi)$ (Eq. (17.29)). The kernel space isomorphism

$$\Omega : \mathcal{K}_{\rho_{\text{in}},\rho_{\text{out}}}^{G,B_{\mathbb{R}^2}(0,\pi)} \xrightarrow{\sim} \mathcal{K}_{\rho_{\text{in}},\rho_{\text{out}}}^{\text{Stab}_n} \quad (17.31)$$

is given by

$$\begin{aligned} \Omega(K) : S^2 \setminus -n &\rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}, \\ p &\mapsto [\Omega(K)](p) := K(\psi_{TM,n}^N \log_n p) \rho_{\text{in}}(g_{n \leftarrow p}^{NP}) \sqrt{|\eta_p^{\partial/\partial v}|}^{-1} \end{aligned} \quad (17.32)$$

if the kernel is expressed relative to (potentially independent) gauges N at n and P at p . Its inverse is given by

$$\begin{aligned} \Omega^{-1}(\kappa) : B_{\mathbb{R}^2}(0,\pi) &\rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}, \\ v &\mapsto [\Omega^{-1}(\kappa)](v) := \kappa(\exp_n(\psi_{TM,n}^N)^{-1} v) \rho_{\text{in}}(g_{n \leftarrow p}^{NP})^{-1} \sqrt{|\eta_p^{\partial/\partial v}|}, \end{aligned} \quad (17.33)$$

where we abbreviated $p := \exp_n(\psi_{TM,n}^N)^{-1} v$.

Proof: By inserting the two expressions, one can easily see that Ω^{-1} is a well defined inverse of Ω since $\Omega \circ \Omega^{-1} = \text{id}_{\mathcal{K}_{\rho_{\text{in}},\rho_{\text{out}}}^{\text{Stab}_n}}$ and $\Omega^{-1} \circ \Omega = \text{id}_{\mathcal{K}_{\rho_{\text{in}},\rho_{\text{out}}}^{G,B_{\mathbb{R}^2}(0,\pi)}}$. The technical part of the proof is to show that the two kernel constraints imply each other, which is done in Appendix L.1. \square

Note that the volume scaling factor is not necessary to establish the isomorphism between the kernel spaces but is required to make the spherical convolution integral over $S^2 \setminus -n$ equivalent to the GM -convolution integral over $B_{\mathbb{R}^2}(0,\pi)$.

Cohen et al. [55] define the convolution $[\kappa *_{S^2} f]$ of a feature field $f \in \Gamma(\mathcal{A}_{\text{in}})$ with spherical steerable kernels $\kappa \in \mathcal{K}_{\rho_{\text{in}},\rho_{\text{out}}}^{\text{Stab}_n}$ in coordinates. Given gauges P at p and Q at Q , let $\phi_p \in \mathcal{I}$ be the unique isometry that moves the north pole to p , i.e. $\phi_p(n) = p$, and that maps the frame at n to the frame at p , that is, $(\phi_p)_{*,GM} \sigma^N(n) = \sigma^P(p)$ or, equivalently, $g_{\phi_p}^{PN}(n) = e$. Let furthermore X be the gauge at $\phi_p^{-1}(q)$. The spherical convolution is then in [55] relative to these gauges pointwise defined by

$$\begin{aligned} [\kappa *_{S^2} f]^P(p) &:= \int_{S^2} \kappa(\phi_p^{-1} q) \rho_{\text{in}}(g_{\phi_p^{-1}}^{XQ}(q)) f^Q(q) dq \\ &= \int_{S^2 \setminus -p} \kappa(\phi_p^{-1} q) \rho_{\text{in}}(g_{\phi_p^{-1}}^{XQ}(q)) f^Q(q) dq, \end{aligned} \quad (17.34)$$

where we removed the antipodal point $-p$ in the second step without changing the result.¹⁰ Intuitively, this operation computes an output feature at p by 1) taking both the kernel and

¹⁰This formulation is more general than that in Eq. (17.24). The latter is recovered for kernels that map scalar fields to regular feature fields.

the input field, 2) rotating them via ϕ_p^{-1} such that p moves to the north pole (via the induced gauge transformation for the feature vector) and 3) integrating their product over the sphere. Instead of sharing weights directly over the tangent spaces, as we do, this operation is therefore sharing weights via the isometry action. By the definition of ϕ_p , both definitions of weight sharing orient the kernel at the target location p such that it is aligned with the chosen frame $\sigma^P(p)$ at this location. The following theorem proves that the GM -convolution with a kernel $K \in \mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^{G, B_{\mathbb{R}^2}(0, \pi)}$ is equivalent to the spherical convolution with the corresponding spherical kernel $\Omega(K) \in \mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^{\text{Stab}_n}$.

Theorem 17.2.2 (Spherical steerable convolutions as GM -convolutions). *Let Stab_n be the stabilizer subgroup of any transitive isometry group \mathcal{I} of S^2 and let $G \leq \text{GL}(2)$ be any isomorphic structure group. Let furthermore $K \in \mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^{G, B_{\mathbb{R}^2}(0, \pi)}$ be any G -steerable kernel on the open ball $B_{\mathbb{R}^2}(0, \pi)$ of radius π (Eq. (17.29)) and let $\Omega(K) \in \mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^{\text{Stab}_n}$ be its corresponding Stab_n -steerable kernel on $S^2 \setminus -n$ (Eqs. (17.28) and (17.32)). The GM -convolution (here for clarity denoted by \star_{GM}) with K is then equivalent to the spherical convolution (\star_{S^2} , Eq. (17.34)) by Cohen et al. [55] with the spherical kernel $\Omega(K)$, that is,*

$$\Omega(K) \star_{S^2} f = K \star_{GM} f \quad (17.35)$$

holds for any spherical feature field $f \in \Gamma(\mathcal{A}_{\text{in}})$.

Proof: The proof is given in Appendix L.2. \square

This proof justifies our claim that the models from [54, 83, 86, 163], discussed in this section, are all special cases of GM -convolutions.

17.3 Azimuthal rotation equivariant spherical CNNs on cylindrical topologies

Besides fully $\text{SO}(3)$ or $\text{O}(3)$ -equivariant spherical convolutions, many spherical CNNs are designed to be equivariant w.r.t. azimuthal rotations around a specified polar axis. All of the models discussed in this section rely either on the $\text{SO}(2)$ -invariant $\{e\}$ -structure that is shown in Figs. 17.2b and 17.3 or, alternatively, that in Fig. 17.4. Due to the triviality of the structure group $G = \{e\}$, the kernel spaces remain unconstrained ($\{e\}$ -steerable). Features are transported according to the unique $\{e\}$ -compatible trivial connection which differs from the usual spherical Levi-Civita connection. With this information, and with the explicit exponential maps in Eq. (17.10), the spherical GM -convolutions in this section are in theory fully specified. In practice, the implementations, listed in row (38) of Table 14.1, differ in their numerical implementations, which we discuss in the following.

Concurrent with our definition of convolutional weight sharing, the models in [58, 351, 295, 77, 203] share a given template kernel over the tangent spaces by orienting it relative to the frames of the considered $\{e\}$ -structure in Fig. 17.2b. However, seemingly contrary to GM -convolutions, the matching of these kernels with the feature field is not done via exponential maps (or transporter pullbacks), but via the gnomonic projection. This gnomonic projection is at any point p defined by

$$\mathcal{G}_p : T_p M \rightarrow H_p^2 \subset S^2, \quad v \mapsto \frac{p + v}{\|p + v\|}, \quad (17.36)$$

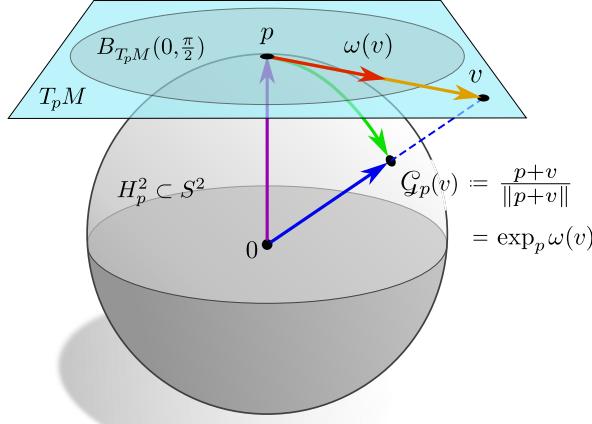


Figure 17.2: Gnomonic projection $\mathcal{G}_p: T_p M \rightarrow H_p^2$ of the tangent space at p to the upper hemisphere $H_p^2 \subset S^2$ around p . When interpreting the sphere as being embedded in \mathbb{R}^3 , the gnomonic projection $\mathcal{G}_p(v)$ (blue) is given by the sum of $p \in S^2 \subset \mathbb{R}^3$ (purple) and $v \in T_p M \subset \mathbb{R}^3$ (yellow) in ambient space, followed by a normalization back to the sphere. Theorem 17.3.1 proves that this operation is equivalent to a projection of a radially warped vector $\omega(v) = \arctan(\|v\|) \frac{v}{\|v\|} \in B_{T_p M}(0, \frac{\pi}{2})$ (red) via the exponential map (green). The gnomonic projection based spherical convolutions in

[58, 351, 295, 77, 203] are therefore special cases of spherical GM-convolutions with radially warped kernels. GM-convolutions are more general since they allow for kernel projections over the whole sphere instead of the upper hemisphere only.

which is visualized in Fig. 17.2. The summation of $p \in S^2 \subset \mathbb{R}^3$ with tangent vectors $v \in T_p M \subset \mathbb{R}^3$ is hereby performed in the embedding space \mathbb{R}^3 and the normalization projects the result back to the sphere. The codomain of the gnomonic projection is the “upper” hemisphere

$$H_p^2 := \{q \in S^2 \mid \langle p, q \rangle_{\mathbb{R}^3} > 0\} \subset S^2 \quad (17.37)$$

centered around p . Given this difference in the kernel projections, it might seem like the models in [58, 351, 295, 77, 203] are not (or only approximately) explained as GM-convolution. The following theorem proves, however, that the gnomonic projection is equivalent to a projection via the exponential map after applying a *radial warp*

$$\omega: T_p M \rightarrow B_{T_p M}(0, \frac{\pi}{2}), \quad v \mapsto \arctan(\|v\|) \frac{v}{\|v\|} \quad (17.38)$$

to the tangent spaces, which contracts tangent vectors to an open ball of radius $\pi/2$ around the origin:

Theorem 17.3.1 (Gnomonic projections as warped exponential maps). *The gnomonic projection \mathcal{G}_p of $T_p M$ to the upper hemisphere $H_p^2 \subset S^2$, defined in Eq. (17.36), is equivalent to a projection of its radial warp $\omega(T_p M) = B_{T_p M}(0, \frac{\pi}{2})$, Eq. (17.38), via the exponential map, that is, the following diagram commutes:*

$$\begin{array}{ccc} T_p M & \xrightarrow{\mathcal{G}_p} & H_p^2 \subset S^2 \\ \omega \downarrow & \nearrow \exp_p & \\ B_{T_p M}(0, \frac{\pi}{2}) & & \end{array} \quad (17.39)$$

In equations,

$$\mathcal{G}_p(v) = \exp_p \circ \omega(v) \quad (17.40)$$

holds for any $p \in S^2$ and any $v \in T_p M$.

Proof: The proof is given by the following simple calculation, which holds for any $p \in S^2$ and any $v \in T_p M$:

$$\begin{aligned}
\exp_p \circ \omega(v) &\stackrel{(1)}{=} p \cdot \cos(\|\omega(v)\|) + \frac{\omega(v)}{\|\omega(v)\|} \cdot \sin(\|\omega(v)\|) \\
&\stackrel{(2)}{=} p \cdot \cos(\arctan(\|v\|)) + \frac{v}{\|v\|} \cdot \sin(\arctan(\|v\|)) \\
&\stackrel{(3)}{=} \frac{p + v}{\sqrt{1 + \|v\|^2}} \\
&\stackrel{(4)}{=} \frac{p + v}{\|p + v\|} \\
&\stackrel{(5)}{=} \mathcal{G}_p(v)
\end{aligned} \tag{17.41}$$

The first two steps make use of the explicit definition of the embedded sphere's exponential map, Eq. (17.10), and the radial warp, Eq. (17.38). The third step follows since $\cos \circ \arctan(x) = \frac{1}{\sqrt{1+x^2}}$ and $\sin \circ \arctan(x) = \frac{x}{\sqrt{1+x^2}}$. In the fourth step we used that $\|p\| = 1$ and $\langle p, v \rangle_{\mathbb{R}^3} = 0$, while the last step identified the gnomonic projection, Eq. (17.36). \square

This theorem implies that the gnomonic projection based convolutions in [58, 351, 295, 77, 203] are indeed specific *GM*-convolutions after identifying the kernels via the radial warp ω .¹¹ Note that this identification holds not only for $\{e\}$ -steerable kernels but for any subgroup $G \leq O(2)$ since the corresponding *G*-steerability constraints affect only the kernels' angular parts but are independent from the warped radial parts. We furthermore want to mention that the exponential map based projection of *GM*-convolutions is insofar more general than the gnomonic kernel projection that it can describe kernels that extend beyond the upper hemisphere H_p^2 around p . Note that both kernel projections become in the practically relevant limit of small kernels even without the radial warp equivalent since $\arctan(\|v\|) = \|v\| + \mathcal{O}(\|v\|^3)$.

The implementations in [58, 351, 295, 77, 203] are in the continuum all equivalent to each other and to our *GM*-convolution, however, their numerical discretizations differ. Coors et al. [58], Eder and Frahm [77] and Martin et al. [203] discretize feature fields on (approximately) uniform sampling grids on the sphere. Specifically, Coors et al. [58] and Martin et al. [203] use the “generalized spiral set on S^2 ” from [254] as sampling points, while Eder and Frahm [77] use the vertices of an icosphere. Since the gnomonic projections of kernel sampling grids on the tangent spaces do not match the spherical sampling grid, the authors interpolate bilinearly between them. The kernel sampling coefficients can hereby be pre-computed in an offline step. The actual convolution computes then an output feature field by contracting the projected, interpolated kernels at each point with the input field.

Zhao et al. [351] and Tateno et al. [295] discretize their spherical feature fields $f : S^2 \setminus \{n, s\} \rightarrow \mathbb{R}^c$ instead in form of a regular pixel grid on an equirectangular projection of the sphere. Mathematically, the equirectangular projection, visualized in Fig. 17.3, is formalized as the pullback $\chi^* f = f \circ \chi : (-\frac{\pi}{2}, \frac{\pi}{2}) \times \mathbb{R} \rightarrow \mathbb{R}^c$, of the image by the spherical

¹¹Technically, the equivalence of both convolutions requires furthermore a radially dependent change of the kernel amplitude to account for the change in the volume measure when warping the kernel. As the kernels are anyways learned, this difference does not matter.

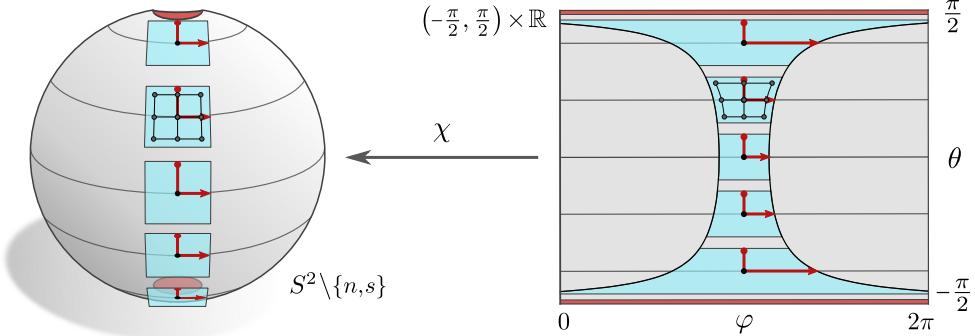


Figure 17.3: Visualization of the $\text{SO}(2)$ -invariant $\{e\}$ -structure that is considered by most models discussed in Section 17.3. All of the frames $[\frac{\partial}{\partial\theta}, \frac{1}{\cos(\theta)}\frac{\partial}{\partial\varphi}]$ are aligned towards the north pole and are orthonormal w.r.t. the embedding metric of the sphere in \mathbb{R}^3 . The spherical coordinate map $\chi : (-\frac{\pi}{2}, \frac{\pi}{2}) \times \mathbb{R} \rightarrow S^2 \setminus \{n, s\}$ from Eq. (17.7) allows to pull spherical feature fields back to feature fields on spherical angles $(-\frac{\pi}{2}, \frac{\pi}{2}) \times \mathbb{R}$, which is denoted as equirectangular projection. Since χ is non-isometric, the spherical $\{e\}$ -structure is in coordinates deformed by a latitude dependent factor of $1/\cos(\theta)$, which diverges towards the poles. The spherical convolution on this $\{e\}$ -structure is in [58, 77, 203] implemented by projecting and interpolating a kernel sampling grid on the tangent spaces to a feature field sampling grid on the sphere. If the feature fields are instead sampled on the equirectangular projection, the kernel sampling grid is in a second step mapped further from the sphere to a deformed sampling grid on $(-\frac{\pi}{2}, \frac{\pi}{2}) \times \mathbb{R}$ [351, 295]. Note that regular sampling grids on the equirectangular projection oversample the signal (relative to the spherical metric) towards the poles.

coordinate map χ from Eq. (17.7):

$$(-\frac{\pi}{2}, \frac{\pi}{2}) \times \mathbb{R} \xrightarrow{\chi} S^2 \setminus \{n, s\} \xrightarrow{f} \mathbb{R}^c \quad (17.42)$$

$\underbrace{\qquad\qquad\qquad}_{\chi^* f}$

As in the previous approaches, the authors project a kernel sampling grid via the gnomonic projection from the tangent spaces to the sphere. In an additional step, they map it via χ to the equirectangular projection where they compute interpolation coefficients between the projected kernel sampling grid and the feature field sampling grid. Since the deformation incurred by the equirectangular projection is independent from the longitude $\phi \in \mathbb{R}$, it is sufficient to compute it only once for each latitude $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$. The following diagram, which commutes by the definitions of K_p^{sphere} and K_p^{equirect} , gives an overview of the gnomonic projection of a kernel $K : \mathbb{R}^2 \rightarrow \mathbb{R}^{C_{\text{out}} \times C_{\text{in}}}$ to the sphere [58, 77, 203] and to its equirectangular projection [351, 295] (note that \mathcal{G}_p is invertible on H_p^2):

$$\begin{array}{ccccc} K & \longrightarrow & \mathbb{R}^{C_{\text{out}} \times C_{\text{in}}} & \xleftarrow{\quad} & K_p^{\text{equirect}} \\ \downarrow & & \uparrow K_p^{\text{sphere}} & & \downarrow \chi^{-1}(H_p^2) \\ \mathbb{R}^2 & \xleftarrow{\psi_{TM,p}^A} & T_p M & \xrightarrow{\mathcal{G}_p = \exp_p \circ \omega} & H_p^2 & \xleftarrow{\chi} & \underbrace{\quad\quad\quad}_{\subset (-\frac{\pi}{2}, \frac{\pi}{2}) \times \mathbb{R}} \end{array} \quad (17.43)$$

A major disadvantage of discretizing spherical feature fields via a regular pixel grid on the equirectangular projection is that this approach oversamples the signal towards the poles.

Further variants of spherical convolutions on the equirectangular projection were proposed by Su and Grauman [287, 288]. Instead of precomputing the deformed kernel sampling pattern, Su and Grauman [287] untie the weight sharing such that each latitude applies its own, independent kernel in a 1-dimensional Euclidean convolution. The network is then on each latitude being pretrained to recover the result that would be obtained when convolving with a kernel that is shared over the tangent spaces as discussed above. If convolutional weight sharing is a suitable inductive bias, this method should optimally converge to the geometry based methods by Zhao et al. [351], Tateno et al. [295]. Su and Grauman [288] develop this approach further and employ a meta-network that predicts a deformed kernel based on a shared input template kernel and the target latitude. Both of these approaches share weights over the circular orbits (lines of constant latitude) of the considered isometry group $\text{SO}(2)$ of $S^2 \setminus \{n, s\}$; cf. Fig. 13.6. They are therefore identified as kernel field transforms with $\text{SO}(2)$ -invariant kernel fields, which are by Theorem 13.2.4 $\text{SO}(2)$ -equivariant.

Given a spherical feature field in equirectangular projection, it might furthermore be tempting to process it directly with a conventional Euclidean CNN, skipping the kernel projection from the tangent spaces, as done for instance in [171, 128]. As discussed in Chapter 15, such Euclidean convolutions correspond to GM -convolutions on the canonical $\{e\}$ -structure of $(-\frac{\pi}{2}, \frac{\pi}{2}) \times \mathbb{R} \subset \mathbb{R}^2$, visualized in Figs. 15.3a and 17.4 (top right). This $\{e\}$ -structure consists of frames $[\frac{\partial}{\partial\theta}, \frac{\partial}{\partial\varphi}]$, which are orthonormal w.r.t. the *Euclidean metric* of $(-\frac{\pi}{2}, \frac{\pi}{2}) \times \mathbb{R}$. These frames are, however, not orthonormal w.r.t. the *spherical metric*, Eq. (17.9), which is in Fig. 17.4 (top left) reflected in the frame contraction by a factor of $\cos(\theta)$ in longitudinal direction. A GM -convolution on this $\{e\}$ -structure corresponds therefore geometrically *not* to a spherical convolution. It rather corresponds to a GM -convolution on a cylinder, which is via the *isometric* coordinate map

$$\tilde{\chi} : (-\frac{\pi}{2}, \frac{\pi}{2}) \times \mathbb{R} \rightarrow (-\frac{\pi}{2}, \frac{\pi}{2}) \times S^1, \quad (\theta, \phi) \mapsto \begin{pmatrix} \cos \phi \\ \sin \phi \\ \theta \end{pmatrix} \quad (17.44)$$

embedded in \mathbb{R}^3 . In contrast, the $\{e\}$ -structure that is shown in Figs. 17.2b and 17.3 consists of frames $[\frac{\partial}{\partial\theta}, \frac{1}{\cos(\theta)} \frac{\partial}{\partial\varphi}]$, which are orthonormal w.r.t. the spherical metric. Note that these frames and the spherical metric are stretched by a factor of $1/\cos(\theta)$ relative to their canonical Euclidean counterparts on $(-\frac{\pi}{2}, \frac{\pi}{2}) \times \mathbb{R}$.

Jiang et al. [139] propose an alternative approach for spherical convolutions on the $\{e\}$ -structure shown in Figs. 17.2b and 17.3. Instead of defining kernels on the tangent spaces, they process the signal via second order partial differential operators of the form $w_{\text{id}} + w_{e_1^A} \partial_1 + w_{e_2^A} \partial_2 + w_{\text{Laplace}} (\partial_1^2 + \partial_2^2)$, where ∂_i denotes the partial derivative in the direction of the i -th frame axis and the weights $w_{(\cdot)} \in \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ are optimized during training. That the weights are position-independent corresponds to our spatial weight sharing. Together with the $\text{SO}(2)$ -invariance of the $\{e\}$ -structure, along which the differential operators are aligned, this guarantees the $\text{SO}(2)$ -equivariance of the operation. In the continuous theory, this model corresponds to a $\{e\}M$ -convolution in the limit of infinitesimally small kernels or, equivalently, to a convolution with $\{e\}$ -steerable PDOs by Jenner and Weiler [137]. In practice, Jiang et al. [139] sample the feature field on an icosphere mesh and represent the differential operators in terms of spatially extended stencils on the mesh vertices. This makes the method equivalent to a GM -convolution with spatially extended kernels.

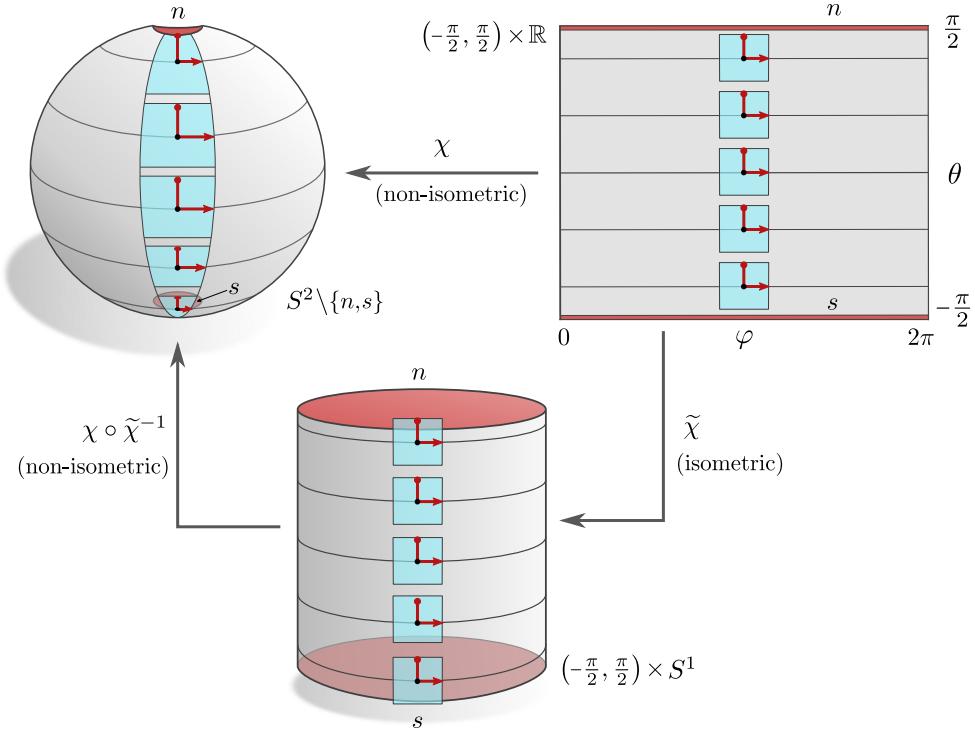


Figure 17.4: The spherical coordinate map $\chi : (-\frac{\pi}{2}, \frac{\pi}{2}) \times \mathbb{R} \rightarrow S^2 \setminus \{n, s\}$, Eq. (17.7), sends angles (θ, ϕ) to points on the sphere. It is non-isometric, which means that the pushforward of orthonormal frames $[\frac{\partial}{\partial \theta}, \frac{\partial}{\partial \phi}]$ w.r.t. the Euclidean metric on $(-\frac{\pi}{2}, \frac{\pi}{2}) \times \mathbb{R}$ does not yield frames that are orthonormal w.r.t. the spherical metric. A conventional Euclidean convolution in coordinates $(-\frac{\pi}{2}, \frac{\pi}{2}) \times \mathbb{R}$ does therefore not correspond to a spherical convolution – its kernels would be contracted by a factor of $\cos(\theta)$ in longitudinal direction. Since distances are measured in terms of angles, this operation corresponds rather to a convolution on a cylinder, which is via the isometric map $\tilde{\chi} : (-\frac{\pi}{2}, \frac{\pi}{2}) \times \mathbb{R} \rightarrow (-\frac{\pi}{2}, \frac{\pi}{2}) \times S^1$, Eq. (17.44), embedded in \mathbb{R}^3 . A spherical convolution requires the $\{e\}$ -structure that is shown in Fig. 17.3.

The model of Lee et al. [178] operates again on an icosphere, however, with a drastically changed (non-smooth) $\{e\}$ -structure: instead of aligning the reference frames such that they all point towards the north pole, the frames point alternatingly towards the north or south. This design is motivated by the pixelation of the icosphere, whose triangular faces are facing either north or southwards. Adjacent pixels can therefore be processed by kernels that are rotated by 180° relative to each other. The authors argue that the training process should make up for this rotation by learning accordingly steerable kernels. Despite the drastic kernel rotations, the $\{e\}$ -structure is invariant under those azimuthal rotations that map northwards pointing frames on themselves, resulting in an approximate $\text{SO}(2)$ -equivariance of the convolution.

The models discussed in this section are easily extended to other solids of revolution ($\text{SO}(2)$ -invariant manifolds) like the cylinder from Fig. 17.4 or the egg from Fig. 13.1. They are furthermore adapted to be $\text{O}(2)$ -equivariant when considering a lift of the $\{e\}$ -structures to \mathcal{R} -structures, which corresponds to using \mathcal{R} -steerable kernels as shown in Fig. 13.7.

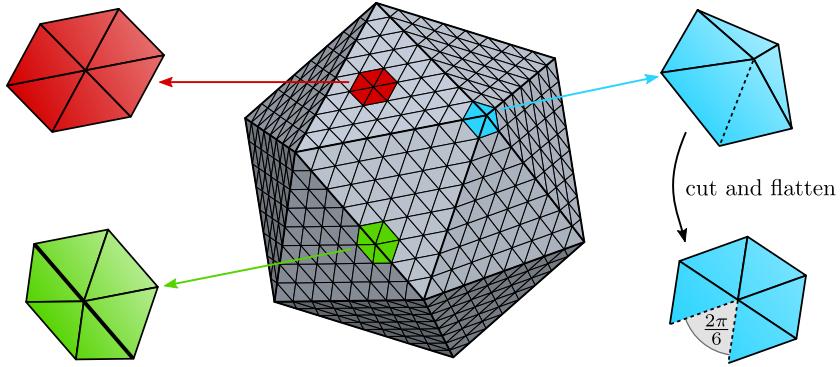


Figure 17.5: The icosahedron is a Platonic solid that is in [187, 57, 346] used as a piecewise flat approximation of the spherical geometry. It consists of 12 vertices, 20 equilateral triangular faces and 30 edges. It admits a regular sampling grid, which is constructed by iteratively subdividing each triangle into four smaller triangles. After r iterations, this procedure results in a grid of $5 \cdot 2^{r+1} + 2$ vertices. The three highlighted patches show the qualitatively different geometry of neighborhoods around vertices on faces (red), edges (green) and icosahedron vertices (blue). The red neighborhood is obviously flat. While the green neighborhood is bent in the embedding space, its intrinsic Gaussian curvature is again vanishing. That this is the case reflects in the facts that it can be flattened out isometrically (i.e. without being cut) and, equivalently, that the Levi-Civita transport along a closed path around the central node is the identity map. The blue neighborhood needs to be cut along one edge in order to be flattened out. The angle defect, i.e. the angle by which the cut is spread when flattening the cusp, equals $\frac{2\pi}{6}$. When parallel transporting a vector once around the central vertex of the neighborhood, it gets rotated by this angle defect. Instead of having constant positive Gaussian curvature like the sphere S^2 , the icosahedron's curvature is concentrated (singular) at its vertices and vanishes everywhere else.

17.4 Icosahedral approximations of spherical CNNs

The sphere S^2 is in computational sciences commonly approximated by Platonic solids, i.e. regular convex polyhedra. In the context of deep learning, interest has mostly been focused on the icosahedron, Fig. 17.5, which approximates the sphere most closely among the platonic solids [261]. While the Riemannian geometry of the sphere is only approximated, Platonic solids have the advantage to be piecewise flat and admit regular meshes. These properties allow for the use of planar convolution routines as described in Part I, which are computationally better optimized than the methods from the previous two sections. This section discusses the icosahedral CNNs from [187], [346] and [57], which rely on the G -structures that are shown in Figs. 17.6a, 17.6b and 17.6c, respectively. Before coming to their implementations in terms of the atlas of affine charts in Fig. 17.7, we give more details on the icosahedral geometry and the considered G -structures.

Icosahedral geometry: The icosahedron is a discrete two-dimensional manifold consisting of 20 equilateral triangular faces, 12 vertices and 30 edges. As done for the 2-sphere, we define the icosahedron as being embedded in \mathbb{R}^3 , from which it inherits the embedding metric in Eq. (17.9). The embedded tangent spaces $T_p M \subset \mathbb{R}^3$ on the faces are hereby defined such that their normals coincide with the face normals. Tangent spaces on the vertices and edges could be defined via the average of the adjacent faces' normals as discussed in the following Chapter 18. However, as we consider feature fields as being sampled on the icosa-

hedron faces (which is almost everywhere), we are independent from this choice. Assuming the Levi-Civita connection, the parallel transport of tangent vectors over faces acts such that it keeps them parallel in the embedding space \mathbb{R}^3 . When being parallel transported across an edge, tangent vectors keep the same angle relative to the edge on either side – this transport may intuitively be thought of as 1) flattening the two adjacent faces out 2) transporting the vector over the edge as usual on a two-dimensional Euclidean space, and 3) bending the two faces back to their original embedding; see Fig. 18.2 and [60]. Geodesics are therefore piecewise linear in \mathbb{R}^3 , crossing edges such that their angle of emanation equals their angle of incidence. Exponential maps $\exp_p(v)$ are thus easily computed by tracing out a piecewise constant path for a distance of $\|v\|$. In practice, the authors of [187, 57, 346] sample feature fields on a regular mesh and consider only those tangent vectors that map to the neighboring mesh vertices.

Fig. 17.5 shows disc-like neighborhoods around exemplary points on faces (red), edges (green) and vertices (blue) of the icosahedron. The red neighborhood is fully contained within a face, and is therefore flat. The green neighborhood is bent in the embedding space, however, its intrinsic (Riemannian or Gaussian) curvature is still zero since the Levi-Civita transport of vectors once around the central vertex preserves them as they are. That this is the case is equivalent to the fact that the green neighborhood can be flattened out isometrically, i.e. without stretching or cutting it. This isometric flattening is not possible for the blue type of neighborhoods around vertices, which have to be cut open at one of the edges in order to be flattened out. Being constructed from five equilateral triangles, the flattened cusp exhibits an angle defect of $\frac{2\pi}{6}$. The holonomy of any closed path around any (single) vertex, that is, the angle between an arbitrary vector and its transport once around the loop, is given exactly by this angle defect. Overall, these results imply that the (discrete) Gaussian curvature of the icosahedron is zero everywhere but at the vertices, where it is singular with holonomy $\frac{2\pi}{6}$. The simple geometry of the icosahedron allows for it to be cut open and globally flattened out as visualized in Fig. 17.7, which was in [187, 57, 346] used for an efficient implementation of icosahedral GM -convolutions.

The icosahedron's full isometry group $\text{Isom}(M) = I_h \leq O(3)$ is finite and consists of 120 elements. It can be thought of as being constructed as the direct product $I \times \mathcal{R}$ of the subgroup \mathcal{R} of reflections and the subgroup $\text{Isom}_+(M) = I \leq SO(3)$ of orientation preserving isometries, containing 60 rotations. Each vertex p is stabilized by five discrete rotations around the axis through p and its antipodal vertex, which form the cyclic group $C_5 \leq SO(2)$. The vertex p is furthermore stabilized by reflections over the plane defined by the rotation axis and any edge emanating from p , such that its full stabilizer subgroup is given by the dihedral group $\text{Stab}_p = D_5 \leq O(2)$. The equivariance of icosahedral GM -convolutions w.r.t. isometry groups I_h , I , D_5 or C_5 was in [57] shown to approximate the full $O(3)$, $SO(3)$, $O(2)$ or $SO(2)$ equivariance of spherical CNNs reasonably well when continuous rotational data augmentation is used.¹²

Icosahedral G -structures: The icosahedral GM -convolutions by Liu et al. [187] and Zhang et al. [346] (implicitly) assume $\{e\}$ -structures, while that by Cohen et al. [57] assumes a C_6 -structure. Fig. 17.6 visualizes the idea behind these G -structures, which we explain in the following three paragraphs in more detail.

¹²This was in [57] empirically shown for $I \leq SO(3)$. That this result generalizes to $I_h \leq O(3)$ is clear since the groups differ only by reflections, w.r.t. which icosahedral GM -convolutions can be made exactly equivariant. It holds furthermore for $D_5 \leq O(2)$ and $C_5 \leq SO(2)$, since these are subgroups of $I_h \leq O(3)$.

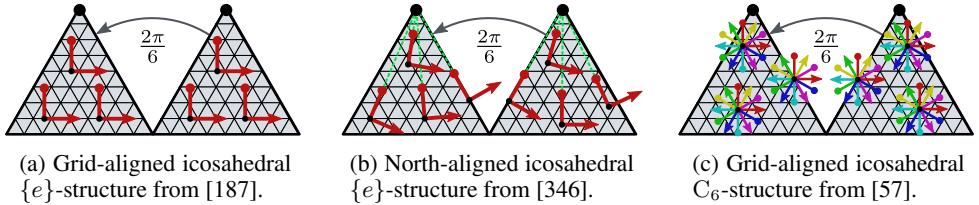


Figure 17.6: Conceptual idea of the G -structures assumed in [187, 346, 57]. For space reasons, only two adjacent faces next to the north pole of the flattened icosahedron (Fig. 17.7) are shown. The $\{e\}$ -structure in Fig. 17.6a is defined by aligning all frames along the “horizontal” edges of the faces (assuming the polar axis to be vertical). Fig. 17.6b shows an alternative $\{e\}$ -structure whose frames are aligned towards the north pole. It is in contrast to the previous $\{e\}$ -structure continuous since frames on the cut edges agree with each other when gluing the edges back together. The C_6 -structure in Fig. 17.6c is constructed by adding frames that are rotated by multiples of $\frac{2\pi}{6}$ to the $\{e\}$ -structure from Fig. 17.6a. Since this angle agrees with the angle defect at the cut edges, the thus defined C_6 -structure is smooth (continuous). Note that the two $\{e\}$ -structures are incompatible with (i.e. not closed under) the Levi-Civita transport but imply an alternative trivial connection. The C_6 -structure, in contrast, is compatible with the Levi-Civita transport.

The $\{e\}$ -structure by Liu et al. [187], shown in Fig. 17.6a, is defined by aligning the first frame axes with the “horizontal” edges of the corresponding triangular faces. When flattening the icosahedron into a plane as shown in Fig. 17.7, all frames of this $\{e\}$ -structure are parallel in this plane, which greatly simplifies the implementation of the corresponding GM -convolutions. As usual, the $\{e\}$ -structure specifies a unique trivial connection according to which features are transported. This trivial connection agrees within the faces, on edges which are not cut in Fig. 17.7 and on the magenta cut edge with the Levi-Civita connection. However, its transport over the remaining cut edges differs from the Levi-Civita transport since the frames of the $\{e\}$ -structure rotate there discontinuously by an angle of $\frac{2\pi}{6}$. As the $\{e\}$ -structure is preserved by rotations in C_5 around the polar axis, its GM -convolutions are approximately $SO(2)$ -equivariant, i.e. approximate the models from the previous Section 17.3. However, the $\{e\}$ -structure – and therefore the network inference – is *non-continuous* over the edges with non-zero angle defect. Furthermore, the reference frames do not point exactly towards the north pole, as it is the case for the spherical $\{e\}$ -structure from Section 17.3 and Fig. 17.2b.

Zhang et al. [346] propose to resolve the latter two issues by working with the $\{e\}$ -structure in Fig. 17.6b. It is defined such that the frames point exactly along the projection of the polar axis onto the faces, i.e. towards the north pole. This $\{e\}$ -structure is continuous everywhere except at the north and south poles.¹³ It is in this sense a better approximation of the spherical $\{e\}$ -structure from Fig. 17.2b. The $\{e\}$ -structure implies again a unique trivial connection. Its transport agrees with the Levi-Civita transport over edges, however, it differs from it when transporting over faces since it rotates vectors smoothly along with the frames. As the other $\{e\}$ -structure, this frame field is invariant under azimuthal rotations in C_5 , and approximates thus azimuthal rotation equivariant spherical CNNs.

The C_6 -structure in Fig. 17.6c by Cohen et al. [57] is defined by augmenting the frames of the $\{e\}$ -structure from Fig. 17.6a with those frames that are rotated by multiples of $\frac{2\pi}{6}$. It is clearly continuous since the angles between the set of preferred frames at each point equal exactly the angle defects at the cut edges. It is in contrast to the previous two $\{e\}$ -

¹³To see this, imagine to glue the cut edge in Fig. 17.6b back together: the frames on the left and right half of the edge are then being mapped together, which is not the case in Fig. 17.6a.

structures compatible with the Levi-Civita transport since the structure group C_6 agrees with the icosahedron's holonomy group. The C_6 -structure is furthermore preserved under the action of the icosahedron's orientation preserving isometries I. GM -convolutions on this C_6 -structure approximate therefore the fully $SO(3)$ rotation equivariant spherical CNNs from Section 17.2.

Implementations: To implement the GM -convolutions on the corresponding G -structures, Liu et al. [187], Zhang et al. [346] and Cohen et al. [57] assume a regular grid on the icosahedron's faces; see Fig. 17.5. This regular hexagonal grid is constructed by iteratively subdividing edges, replacing each triangle with four smaller ones. At resolution r , this yields a grid with $5 \cdot 2^{2r+1} + 2$ vertices. Note that this grid is by construction exactly symmetric under isometries of the icosahedron, which leads to an exact Isom_{GM} -equivariance of the discretized GM -convolutions.¹⁴ Liu et al. [187] proposed to represent icosahedral feature fields relative to the atlas of charts that is shown in Fig. 17.7. The charts have the advantage that they map the hexagonal grids on the icosahedron's faces to common square pixel grids. Note, however, that orthonormal frames on the icosahedron are in this representation deformed, such that they are not orthonormal relative to the canonical Euclidean metric. Hexagonal convolution kernels on the icosahedron are deformed accordingly and can be implemented in terms of square kernels which are masked such that two of their corners are filled with zeros.

The GM -convolution by Liu et al. [187] assumes frames that are all parallel and can therefore in the interior of the charts, where the kernel support does not extend over its boundaries, be implemented via a conventional Euclidean convolution; see Chapter 3. At points that are close to an edge between different charts, the kernel accumulates features from beyond the cut. As already discussed and visualized in Section 10.4 and Fig. 10.3, this is conveniently implemented via a transport padding operation which pads a border of parallel transported features around the array of square pixels before running the convolution operation. For the trivial transport implicitly assumed by Liu et al. [187], this padding operation just copies a row of features at each edge without transforming them. Since the authors assume the trivial structure group $G = \{e\}$, the hexagonal kernels remain unconstrained.

The implementation of Cohen et al. [57] is mostly similar, however, it differs crucially in that it uses Levi-Civita transporters and C_6 -steerable kernels. Instead of directly padding rows of pixels across edges, the Levi-Civita transport requires that the features are steered either by $g = e$ for all internal edges and the magenta edge or by an angle of $\pm \frac{2\pi}{6}$ over all edges with angle defect $\frac{2\pi}{6}$, with the sign depending on the transport direction. Cohen et al. [57] assume the regular representation of C_6 as field type and constrain the convolution kernels to satisfy the corresponding steerability constraint. After transport padding, their GM -convolution is implemented as a conventional Euclidean convolution with these steerable kernels, i.e. an $\text{Aff}(C_6)$ -steerable convolution from Chapter 4 on the chart codomains. Note that this GM -convolution is within the faces, i.e. except for the transport padding, similar to the HexaConv by Hoogeboom et al. [125].

Since the GM -convolution by Zhang et al. [346] assumes a trivial structure group $G = \{e\}$, the transport padding is again implemented as a trivial copy of pixels without steering and the kernels are again left unconstrained. However, as the frames of the $\{e\}$ -structure are aligned towards the north pole, they are no longer parallel in the rectangular square pixel representation, which prevents an immediate implementation in terms of conventional

¹⁴The icosphere grid, used by some of the models from Sections 17.2 and 17.3, is defined by projecting the nodes of this grid to unit radial distance from the origin, i.e. to S^2 . The models in this section do not assume this projection but convolve directly over the piecewise flat icosahedral geometry.

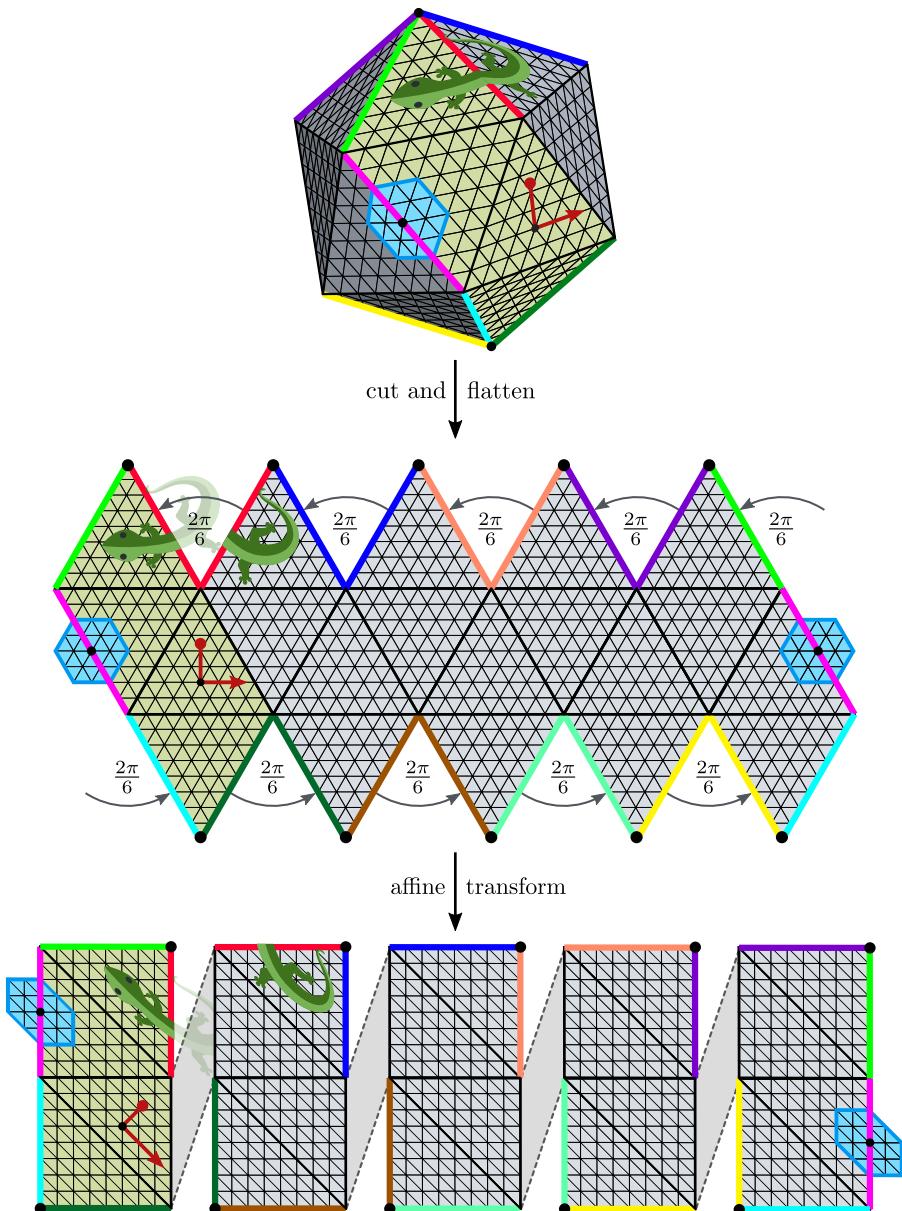


Figure 17.7: The implementations in [187, 57, 346] represent feature fields relative to an atlas that covers the icosahedron with five charts. To construct these charts, one cuts the icosahedron along the colored edges and flattens it out. Five regions, consisting of four triangles each, are then sheared to rectangular chart codomains. This operation maps the hexagonal grid to a grid of square pixels, such that icosahedral feature fields can be encoded by a set of five rectangular arrays. Note that reference frames and kernels are deformed accordingly on the chart codomains. The Levi-Civita transport over all colored edges but the magenta one picks up a rotation of $\pm \frac{2\pi}{6}$, with the sign depending on the transport direction. This is implemented by transport padding rows of pixels along the cut edges as previously described in Fig. 10.3.

(Lizards adapted under the Creative Commons Attribution 4.0 International license by courtesy of Twitter.)

convolutions. Instead, the kernels have to be applied in a different rotation at each grid point. As the hexagonal kernel can be rotated by $\frac{2\pi}{6}$ without using interpolation, and since the alignments towards the north pole differ at most by this angle from each other, the authors propose the following efficient approximating of this operation: they convolve twice on each face, once with the original kernel and once with its $\frac{2\pi}{6}$ rotated version. The two response fields are then linearly combined, with the precomputed interpolation weights depending on the angles of the north-aligned reference frames relative to the two kernel alignments (i.e. relative to the pixel grid). This implementation is therefore approximately twice as costly as those in [187, 57].

An alternative implementation of spherical convolutions on the icosahedron was proposed by Eder et al. [78]. The authors project the spherical signal on planes spanned by the 20 faces (denoted as tangent images) and subsequently run a conventional CNN on each of these images. We did not include this network in our list as it processes these representations independently from each other, that is, it does not transport or pad features between them, and is therefore not exactly described as *GM*-convolution.

As mentioned before, the empirical results by Kicanaoglu et al. [148] suggest that the icosahedral geometry approximates the spherical geometry reasonably well for deep learning applications. More specifically, the authors compare their spherical CNN on an icosphere grid with the piecewise flat icosahedral CNN by Cohen et al. [57] and find that both perform similarly despite the deformed geometry of the latter. The equivariance of icosahedral CNNs under continuous rotations in $\text{SO}(3)$ is found to be violated significantly, however, this seems to be a mere overfitting effect as it is easily and without loss of model performance counteracted by leveraging $\text{SO}(3)$ data augmentation.

As always, we want to mention that the C_5 -equivariant CNNs by Liu et al. [187] and Zhang et al. [346] can easily be made D_5 -equivariant by considering a \mathcal{R} -structure and thus reflection-steerable kernels. Similarly, the I -equivariant CNN by Cohen et al. [57] can be made equivariant under the full isometry group I_h of the icosahedron by making the kernels D_6 -steerable instead of C_6 -steerable.

Coordinate independent CNNs on general surfaces

Instead of operating on a fixed geometry, the GM -convolutions in the current chapter are defined on general manifolds. We restrict our review to surfaces ($d = 2$) since we are not aware of implementations on (general) higher-dimensional manifolds. The signals to be processed could either be directly given by the dataset or are computed from the surfaces' geometries. Examples for the former would be color textures or physical quantities like temperature fields or the wall stress of a pressurized container. The latter could for instance be Gaussian and principal curvatures, SHOT descriptors or wave kernel signatures. Most applications so far focus on classifying the surfaces [131, 140, 327], segmenting parts of them [232, 131, 327, 339] or finding correspondences between different surfaces [204, 22, 259, 327, 67]. Further applications are the prediction of physical quantities like mechanical stress [293] or the synthesis of color textures [304, 343] or geometric deformations [119].

The design of Euclidean and spherical CNNs is strongly guided by the requirement for global symmetry equivariance. Since general surfaces come usually with trivial isometry groups this guiding principle falls away, which leaves us with a large freedom in the choice of G -structures. The models that we review in this chapter can be classified into *rotation-steerable* and $\{e\}$ -steerable surface convolutions. Both approaches address the issue of a missing canonical direction on surfaces, however, they do it in a fundamentally different way. Rotation-steerable models account for the lack of reference direction by their equivariant design, treating all directions equivalently. Their underlying $SO(2)$ -structure is – up to a practically irrelevant choice of orientation¹ – fixed by the Riemannian metric. The rotation steerable models differ therefore mainly in their choice of field types. The $\{e\}$ -steerable models are non-equivariant and are therefore not associated to a (non-trivial) field type. However, they differ from each other by the specific choice of $\{e\}$ -structure that is used to determine the kernel alignments.

This chapter is organized as follows: we start in Section 18.1.1 with a brief overview of the classical differential geometry of surfaces, discussing in particular the difference between their intrinsic and extrinsic geometry. In practice, most implementations operate on discretized surfaces. Section 18.1.2 gives an overview of the geometry of triangular surface meshes, which are arguably the most common surface discretizations in the deep learning literature. In Section 18.2 we discuss rotation-steerable surface convolutions. Heuristics for fixing the frame fields that define $\{e\}$ -steerable surface convolutions are reviewed in Section 18.3.

¹The chosen orientation is on a (connected, orientable) manifold arbitrary when kernels are learned: if the opposite orientation was chosen, the training would just result in oppositely oriented kernels.

For completeness, we mention in the following paragraph a few alternative approaches to define surface convolutions before coming to the actual content of this chapter.

Surface CNNs beyond *GM*-convolutions

While quite some surface CNNs can be interpreted as *GM*-convolutions, many alternative network designs have been proposed. These methods rely for instance on graph convolutions on surface meshes, spectral approaches, multi view renderings of surface embeddings, volumetric methods in the embedding space, differential operators, or other operators which operate immediately on the mesh data structures. The following brief review is intended to give an overview of the different directions which have been explored.

One method to classify or segment embedded surfaces is to *render them from multiple viewpoints* and process the renderings with conventional Euclidean CNNs. The resulting features are then aggregated by pooling over the viewpoints [285, 236] or via a consensus method [224]. Esteves et al. [85] choose to place the camera viewpoints on a sphere according to a discrete subgroup of $\text{SO}(3)$, for instance the icosahedral group. The resulting features are then processed jointly via a discrete group convolution (not a surface convolution).

Instead of projecting the surface by rendering it, it can be projected to \mathbb{R}^2 by defining a *chart*. Sinha et al. [278] define approximately authalic (area preserving) global charts on spherical topologies. These charts are discontinuous and in general not conformal (angle preserving). A conventional Euclidean CNN is used to process the resulting images. The discontinuities can be circumvented by pulling the surface features back along toric [199] or more general [115, 13] covering maps. The subsequent Euclidean convolution on the pullback can not be interpreted as a *GM*-convolution since the sheets of the covering map induce different, incompatible $\{e\}$ -structures on the surface. Li et al. [180] use an atlas of (approximately) isometric charts – as discussed at the end of Section 18.3, this corresponds indeed to a *GM*-convolution.

Volumetric methods process embedded surfaces with conventional Euclidean CNNs in the embedding space \mathbb{R}^3 , for instance by interpreting the vertices of a surface mesh as a *point cloud* [234, 235, 300] or by *voxelizing* the input. Point cloud based methods are reviewed in [112]. Mescheder et al. [207] and Peng et al. [226] argue that an implicit surface parametrization is more economical and propose networks which model surfaces as decision boundaries.

Spectral approaches are inspired by the convolution theorem. The Fourier basis on a manifold is thereby given by the eigenfunctions of the Laplace-Beltrami operator. Spectral neural networks process feature maps by manipulating their Fourier spectrum with learned linear operators. As the Fourier basis is non-localized, Boscaini et al. [21] use instead a windowed Fourier transform; an alternative are the localized manifold harmonics of Melzi et al. [206]. Bruna et al. [28] interpret surface meshes as graphs. They are therefore applying graph Fourier transforms, which are based on the eigenfunctions of the graph Laplacian.

Sharp et al. [269] suggest a model which is based on *differential operators*. Scalar features are propagated via heat diffusion with a learnable diffusion time. As the Laplacian (occurring in the heat equation) is isotropic, it can not respond selectively to pattern in specific rotations. The authors are therefore additionally applying a gradient operator, followed by taking scalar products of the resulting tangent vector-valued features. Note that both operations are gauge invariant. The networks can be implemented on all data structures which admit partial differential operators, for instance point clouds or meshes.

Quite some networks do not operate on the *Riemannian manifold structure* but rather on the *data structure* which represents the surfaces numerically. An example are networks which interpret the nodes and edges of a surface mesh as forming a graph and consequently apply *graph networks*. The isometry equivariance of graph networks was investigated in [144, 126]. Verma et al. [309] proposed a graph network with dynamic filters, i.e. filters that are during the forward pass predicted from the features. The model of Milano et al. [210] operates on the primal and dual graphs of meshes and utilizes attention mechanisms.

Spiral nets process features on meshes via local spiral operators [183, 108]. These operators enumerate features by following a spiral path outwards from the central node. A response is computed by applying a LSTM to the resulting sequence of features or an MLP to their concatenation. The choice of first neighbor and spiraling direction corresponds to a choice of $\{e\}$ -structure. Hanocka et al. [117] and Hertz et al. [119] define convolutions on mesh faces and edges, respectively. Both models are made invariant to the arbitrariness in the mesh element ordering, which could be generalized to a permutation equivariant design.

For more in-depth reviews of such methods we point the reader to Bronstein et al. [26] and Guo et al. [112].

18.1 Geometry of embedded surfaces

This section gives a brief introduction to the geometry of surfaces. Some concepts of the differential geometry of *smooth* embedded surfaces are discussed in Section 18.1.1. Section 18.1.2 attempts to give an overview of possible ways to *discretize* differential quantities on surface meshes.

For a more in depth treatment of parameterized surfaces, we refer the reader to [99]. A concise and intuitive introduction to the topic and its relation to computational (discretized) geometry can be found in [60].

18.1.1 Classical differential geometry of embedded surfaces

Classically, surfaces have been described *extrinsically*, that is, as being immersed (or embedded) in an Euclidean ambient space \mathbb{R}^3 . This immersion can be defined in multiple equivalent ways, for instance local parametrizations, Monge patches or implicit functions. Local surface parametrizations are smooth maps

$$\chi : \mathbb{R}^2 \supset V \rightarrow M \subset \mathbb{R}^3 \quad (18.1)$$

which immerse open subsets V of \mathbb{R}^2 into the ambient space \mathbb{R}^3 . They are required to be regular, that is, their partial derivatives

$$e_i = \frac{\partial \chi}{\partial x_i}, \quad i = 1, 2 \quad (18.2)$$

are required to be linearly independent in \mathbb{R}^3 . The derivatives $e_1(x_1, x_2) \in \mathbb{R}^3$ and $e_2(x_1, x_2) \in \mathbb{R}^3$ span the embedded tangent spaces $T_p M \subset \mathbb{R}^3$ at $p = \chi(x_1, x_2)$.² Surface normals in the embedding space are therefore well defined and given by $n = \frac{e_1 \times e_2}{\|e_1 \times e_2\|}$.

²The derivative vectors e_i correspond in the intrinsic chart formalism to *coordinate bases*; see Appendix C.3.1.

An atlas of compatible local surface parametrizations allows to describe surfaces that differ topologically from the plane on a global level.

The Riemannian metric of the surface – in this context often denoted as its *first fundamental form* – is induced from the embedding space. In accordance with the analogous definition for the embedded sphere S^2 in Eq. (17.9) we have:

$$\eta_p(v, w) := \langle v, w \rangle_{\mathbb{R}^3} \quad \forall v, w \in T_p M \quad (18.3)$$

Let $v = \sum_i v_i e_i$ and $w = \sum_i w_i e_i$ be tangent vectors in $T_p M$ that are expressed in terms of their coefficient vectors $v, w \in \mathbb{R}^2$ relative to the coordinate basis. The metric is relative to this basis represented by a symmetric coefficient matrix

$$I = \begin{pmatrix} E & F \\ F & G \end{pmatrix} \quad (18.4)$$

with elements³ $E = \langle e_1, e_1 \rangle_{\mathbb{R}^3}$, $F = \langle e_1, e_2 \rangle_{\mathbb{R}^3} = \langle e_2, e_1 \rangle_{\mathbb{R}^3}$ and $G = \langle e_2, e_2 \rangle_{\mathbb{R}^3}$. This matrix acts on vector coefficients according to $\eta_p(v, w) = v^\top I w$. The first fundamental form encodes the *intrinsic* geometry of a surface as a two-dimensional Riemannian manifold, i.e. that part of the geometry which is independent of its immersion into the ambient space.

A surface's *extrinsic* geometry, i.e. details about its particular immersion into the ambient space, is captured by its *second fundamental form*. Relative to e_1 and e_2 this form is represented by the matrix

$$II = \begin{pmatrix} L & M \\ M & N \end{pmatrix} \quad (18.5)$$

with elements $L = \langle n, \frac{\partial^2 \chi}{\partial x_1^2} \rangle_{\mathbb{R}^3} = \langle n, \frac{\partial e_1}{\partial x_1} \rangle_{\mathbb{R}^3}$, $M = \langle n, \frac{\partial^2 \chi}{\partial x_1 \partial x_2} \rangle_{\mathbb{R}^3} = \langle n, \frac{\partial e_1}{\partial x_2} \rangle_{\mathbb{R}^3} = \langle n, \frac{\partial e_2}{\partial x_1} \rangle_{\mathbb{R}^3}$ and $N = \langle n, \frac{\partial^2 \chi}{\partial x_2^2} \rangle_{\mathbb{R}^3} = \langle n, \frac{\partial e_2}{\partial x_2} \rangle_{\mathbb{R}^3}$. These elements measure essentially how the coordinate bases – and thus tangent spaces – bend in ambient space (into the normal direction) when moving along the coordinate lines. It can for instance be used to determine the *normal curvature*

$$\kappa_n(v) = \frac{v^\top II v}{v^\top I v} \quad (18.6)$$

of the surface at p in direction of $v = \sum_i v_i e_i \in T_p M$. Intuitively, this normal curvature can be understood as the curvature of the curve defined by the intersection of the surface with the plane spanned by the direction v and the normal n at that point. This curvature agrees with the inverse radius $r = 1/\kappa_n(v)$ of the osculating circle to the curve at p , and therefore measures how the surface bends into the normal direction when moving in the direction of v ; see [60] for great visualizations of this situation. Other quantities of interest in the study of immersed surfaces are their principal, mean and Gaussian curvatures, which can be expressed in terms of the normal curvatures and are exemplified in Fig. 18.1. The directions (unit vectors in $T_p M$) v_{\max} and v_{\min} in which the normal curvature at a given point p are maximal or minimal are denoted as *principal directions* at p . The corresponding curvatures

$$\kappa_{\max} = \kappa_n(v_{\max}) \quad \text{and} \quad \kappa_{\min} = \kappa_n(v_{\min}) \quad (18.7)$$

are the *principal curvatures* at p . Their mean value

$$\kappa_{\text{mean}} = \frac{\kappa_{\max} + \kappa_{\min}}{2} \quad (18.8)$$

³In modern notation, the coefficients of a (coordinate free) metric g relative to a given basis are often denoted by $g_{\mu\nu}$.

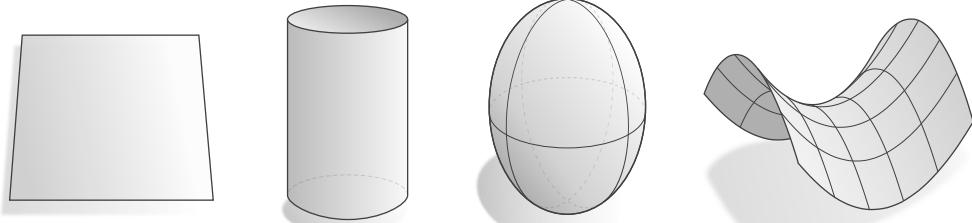


Figure 18.1: Embedded surfaces of qualitatively different extrinsic curvatures. *Left:* The plane is characterized by vanishing principal and Gaussian curvatures $\kappa_{\max} = \kappa_{\min} = \kappa_{\text{Gauss}} = 0$. *Middle left:* A cylinder has one direction of positive curvature and one of vanishing curvature, i.e. $\kappa_{\max} > 0$ and $\kappa_{\min} = 0$. Its Gaussian curvature $\kappa_{\text{Gauss}} = 0$ is therefore zero as well. The plane and the cylinder are locally isometric, that is, their intrinsic geometry is locally indistinguishable. Note that the plane can be rolled up (developed) to form a cylinder – the difference between the two is just the embedding into ambient space. *Middle right:* An ellipsoid is characterized by its positive principal and Gaussian curvatures $\kappa_{\max} > 0$, $\kappa_{\min} > 0$ and $\kappa_{\text{Gauss}} > 0$ at every point. *Right:* The surface of a saddle bends in opposite directions, implying opposite signs of the principal curvatures $\kappa_{\max} > 0$ and $\kappa_{\min} < 0$. As a result, the Gaussian curvature $\kappa_{\text{Gauss}} < 0$ is negative.

is known as *mean curvature*. The mean curvature is zero at “saddle-like” points where $\kappa_{\min} = -\kappa_{\max}$. Minimal surfaces have zero mean curvature at every point. The product

$$\kappa_{\text{Gauss}} = \kappa_{\max} \cdot \kappa_{\min} \quad (18.9)$$

of the principal curvatures is denoted as *Gaussian curvature*. This curvature is positive if the principal curvatures have the same sign, which is for instance the case for ellipsoids. In order for the Gaussian curvature to be negative, the signs of the principal curvatures need to differ, as around hyperbolic (saddle-like) regions. The Gaussian curvature is zero if either (or both) of the principal curvature values is (are) zero, i.e. if the surface has a flat direction. An example for a manifold with zero Gaussian curvature is the cylinder. Such surfaces are said to be *developable*, which means that they can be flattened out into a plane without being distorted, or, more rigorously formulated, they are locally isometric to the plane. Carl Friedrich Gauss proved in his *theorema egregium* that the Gaussian curvature of a surface is actually an intrinsic property, i.e. that it does not depend on how the surface is immersed into ambient space. It is in one-to-one correspondence with the (intrinsic) Riemannian curvature tensor of a surface (and thus also to its Ricci and scalar curvature). An important property of the Gaussian curvature is that its integral over a topological disk $D \subset M$ equals the holonomy $\delta_{\partial D}$, i.e. the angle by which a vector is rotated when (Levi-Civita) transporting it once around the disk boundary ∂D :

$$\int_D \kappa_{\text{Gauss}} dp = \delta_{\partial D} \quad (18.10)$$

As we will see below, this relation can be used to generalize the Gaussian curvature to meshes, where the holonomy $\delta_{\partial D}$ agrees with the angle defect of an unfolded loop of faces (like the blue neighborhood in Fig. 17.5).

Since *GM*-convolutions depend only on the intrinsic geometry of a surface, the reader might wonder why we are discussing their extrinsic properties like principal curvatures. The reason is that *GM*-convolutions may nonetheless be informed about a surface’s extrinsic geometry, for instance by encoding it in feature fields. The extrinsic geometry may furthermore be used to heuristically align the frames of an $\{e\}$ -structure and thus kernels. For example, Jin et al. [140] and Li et al. [180] align the frames along the z -axis of the ambient space \mathbb{R}^3 , while

Boscaini et al. [22] and Tatarchenko et al. [294] align the frames along the surface’s dominant principal curvature direction. Note that these heuristics are not always well defined: for instance, the projection of the z -axis on a “horizontal” tangent space (in the ambient space) is zero, and the dominant principal curvature direction might not be defined, as it is the case on the sphere.

18.1.2 Discretized geometry of surface meshes

In principle, it would be possible to describe GM -convolutions on local surface parametrizations as described in the last section. While this approach might be suitable for certain simple or symmetric geometries like ellipsoids, hyperboloids or tori, it seems impractical for more complex geometries. In practice, surfaces come mostly discretized, for instance in form of triangle meshes, quad meshes, halfedge meshes, subdivision surfaces or point clouds. Due to their widespread use – both in general and specifically in the description of the surface GM -convolutions that we review in the next two sections – we will in the following focus primarily on triangle meshes. Our goal for the remainder of the current section is therefore to take the quantities and definitions from the smooth theory and discuss their discrete counterparts on triangle meshes. Unfortunately, these discrete analogues are usually not unique, such that a plethora of inequivalent definitions exists.⁴ We will in the following try to give a general idea about some of the most common approaches of discretizing the smooth geometry of surfaces in terms of triangle meshes.

Topology, geometry and embedding of triangle meshes: Triangle meshes $(\mathcal{V}, \mathcal{F})$ are commonly encoded in terms of a set

$$\mathcal{V} \subseteq \mathbb{N} \tag{18.11}$$

of *vertices* and a set

$$\mathcal{F} \subseteq \{\{i, j, k\} \mid i \neq j \neq k \in \mathcal{V}\} \tag{18.12}$$

of triangular *faces*, satisfying that each vertex is contained in at least one of the faces.⁵ A set

$$\mathcal{E} = \{\{i, j\} \mid i \neq j \in \{i', j', k'\} \text{ for some } \{i', j', k'\} \in \mathcal{F}\} \tag{18.13}$$

of *edges*, bounding the faces, follows immediately. In practice, one is often given a set

$$P = \{p_i \in \mathbb{R}^3 \mid i \in \mathcal{V}\} \tag{18.14}$$

of vertex positions, specifying an *embedding* of the mesh in the ambient space \mathbb{R}^3 . This embedding implies lengths

$$l_{\{i, j\}} = \|p_j - p_i\| \tag{18.15}$$

⁴Meyer et al. [209] describe this situation as follows: “*Despite extensive use of triangle meshes in Computer Graphics, there is no consensus on the most appropriate way to estimate simple geometric attributes such as normal vectors and curvatures on discrete surfaces.*”. Similarly, Crane [60] claims: “*There is no one “right” way to discretize a given geometric quantity, but rather many different ways, each suited to a particular purpose.*”

⁵Faces may alternatively be defined as ordered 3-tuples of vertices. The ordering of the vertices (or rather the equivalence classes of orderings under an even number of permutations) may then be used to encode the faces’ orientations. We will instead encode face orientations as in our smooth theory by a choice of handedness of reference frames.

of edges $\{i, j\}$ and areas

$$A_{\{i,j,k\}} = \frac{1}{2} \| (p_j - p_i) \times (p_k - p_i) \| \quad (18.16)$$

of faces $\{i, j, k\}$.

We are specifically interested in *surface meshes*, which are required to satisfy additional conditions. In order to formulate these conditions, note that the mesh elements $\{i_0, \dots, i_n\}$ (where $n = 0, 1$ or 2 for vertices, edges or faces) imply *n-simplices*, defined as the convex hulls

$$\begin{aligned} \text{convex}(\{i_0, \dots, i_n\}) := \\ \left\{ \sum_{j=0}^n \alpha_j p_{i_j} \mid \sum_{j=0}^n \alpha_j = 1 \text{ and } \alpha_j \geq 0 \forall j = 0, \dots, n \right\} \subset \mathbb{R}^3. \end{aligned} \quad (18.17)$$

The set that comprises all of these simplices (mesh elements) forms a *pure 2-simplicial complex* [68, 60]. That the 2-simplicial complex is *pure* means that each 0-simplex (vertex) and 1-simplex (edge) is a subset of at least one 2-simplex (face). In other words, there are no disconnected vertices or edges in the mesh. The *underlying space*

$$\bigcup_{\{i_0, \dots, i_n\} \in \mathcal{V} \cup \mathcal{E} \cup \mathcal{F}} \text{convex}(\{i_0, \dots, i_n\}) \stackrel{\text{(pure)}}{=} \bigcup_{\{i, j, k\} \in \mathcal{F}} \text{convex}(\{i, j, k\}) \subset \mathbb{R}^3 \quad (18.18)$$

of the simplicial complex is defined as the union of all of its simplices, equipped with the usual topology as a subset of \mathbb{R}^3 . A mesh is then said to be a surface mesh (manifold mesh) if the underlying space is a topological surface (manifold), optionally with boundary. Intuitively, this requires 1) that each edge is adjacent to two faces (or one at boundaries) and 2) that the faces around each vertex form a topological disk (or a half-disk at boundaries).

Such defined surface meshes are discrete counterparts to *embedded* Riemannian surfaces. However, since *GM*-convolutions are independent from the *extrinsic* geometry of the underlying manifold, it is instructive to briefly discuss their *intrinsic* geometry. Take therefore the vertex, edge and face sets \mathcal{V} , \mathcal{E} and \mathcal{F} , but discard the embedding locations P of the vertices. Together, these sets form an *abstract 2-simplicial complex* $\mathcal{V} \cup \mathcal{E} \cup \mathcal{F}$, defined as a family of abstract simplices $\{i_0, \dots, i_n\}$ that is closed under taking subsets [60]. If this (now abstract) 2-simplicial complex is 1) pure and 2) such that the “Star” of every vertex (given by the simplices containing that vertex) form a combinatorial disk, it forms an *abstract simplicial surface* (which is exactly the case if the embedded mesh is a surface mesh). Abstract simplicial surfaces can be viewed as combinatorial counterparts of topological manifolds. They admit to compute topological invariants, for instance the Euler characteristic

$$\chi_{\text{Euler}} = |\mathcal{V}| - |\mathcal{E}| + |\mathcal{F}|. \quad (18.19)$$

As a topological invariant, the Euler characteristic agrees for any two homeomorphic spaces, in particular for a smooth manifold and any of its triangulations. For instance, the icosahedron from Section 17.4 has $\chi_{\text{Euler}}^{\text{ico}} = 12 - 30 + 20 = 2$, which agrees with $\chi_{\text{Euler}}^{S^2} = 2$ for the 2-sphere.

To arrive at an intrinsic description of a triangulated surface’s *geometry*, one assigns *edge lengths* $l_{ij} \in \mathbb{R}_{>0}$ to edges $\{i, j\} \in \mathcal{E}$. For consistency, these lengths are required to satisfy the triangle inequality $l_{\{i,j\}} + l_{\{j,k\}} > l_{\{k,i\}}$ for any face $\{i, j, k\} \in \mathcal{F}$. The edge lengths imply Euclidean metrics (distance functions) on the faces, and therefore a piecewise defined Euclidean metric on the whole surface. It corresponds to a Riemannian metric (or first fundamental form) which is Euclidean away from vertices and “cone-like” (singular) on a small neighborhood around the vertices [61, 68].

In order to close the circle to our initial, extrinsic definition of triangle meshes, one needs to embed the mesh into ambient space \mathbb{R}^3 . The necessary information on the extrinsic geometry is given by equipping the mesh with a *second fundamental form*. In the discrete setting, this form can be defined as a choice of dihedral angle (bending angle) between any two adjacent triangles, i.e. one angle per non-boundary edge of the mesh. Provided that this data is chosen consistently⁶, it is possible to reconstruct the embedding, i.e. vertex positions P , up to rigid motions in $E(3)$ [185, 320]. While an embedding of the surface is not necessary for the intrinsic *GM*-convolutions, all of the papers listed in rows (41-45) of Table 14.1 evaluate their models on embedded triangle meshes.

Tangent spaces and vector fields: To describe vector fields on meshes, and to equip the meshes with geometric structure like connections, it is necessary to define a notion of tangent spaces that are attached to them. Multiple incompatible definitions, tailored towards the specific application in mind, occur in the literature. Since vector fields are commonly sampled at discrete locations, the discrete tangent bundles are often only partially defined, for instance only on faces, edges or vertices. We briefly review some of these definitions, a more detailed survey can be found in [65].

Since the faces (2-simplices) of an embedded mesh are flat, one can naturally define their tangent spaces as those two-dimensional subspaces of \mathbb{R}^3 in which they are contained [62, 60, 320]. Specifically, given a face $\{i, j, k\} \in \mathcal{F}$, one may define the tangent spaces $T_p M = \text{span}(p_j - p_i, p_k - p_i) \subset \mathbb{R}^3$ for every $p \in \text{convex}(\{i, j, k\})$ as the *linear span of any two edge vectors*. The alignment of the tangent space in ambient space is often represented in terms of the face normal $n = (p_j - p_i) \times (p_k - p_i)$. Discrete tangent (or feature) vector fields can in face based representations be defined as being face-wise constant, i.e. represented by a single tangent (or feature) vector per face. Relative to a choice of reference frame on each of the faces, such tangent and feature vector fields are encoded by $2|\mathcal{F}|$ or $c|\mathcal{F}|$ vector coefficients, respectively. Note that such vector fields do not extend to vertices or edges. Due to their discontinuity, the notion of differential operators, acting on such fields, is quite limited [65] (which is irrelevant for our specific application). A linear interpolation scheme of face based vector fields was proposed in [181].

As there is no natural normal direction at the vertices of a mesh, there are multiple common definitions of vertex tangent spaces. *Vertex normals* can for instance be defined as an area weighted average of the adjacent faces' normals [185, 172, 67]. Besides area weighting, uniform weights or tip angle weights are sometimes used [60]. Another option is to define normal vectors via a mean curvature normal operator [209]. The resulting normal agrees with normals derived via area gradients, but differs from those that are derived via volume gradients or sphere inscribed normals; see [60].

Alternatively, one can define vertex tangent spaces in an *intrinsic* way, simply by defining them as two-dimensional vector spaces that are attached to the vertex. Their relation to the mesh geometry in a local neighborhood around the vertex is hereby encoded by representing the one-ring neighborhood in the tangent planes. The arguably most prominent of such approaches is based on a rescaling of the total angle

$$\Theta_i = \sum_{\{i,j,k\} \in \mathcal{F}} \theta_{\{i,j,k\}}^i, \quad (18.20)$$

which is summed from the tip angles $\theta_{\{i,j,k\}}^i = \arccos \left\langle \frac{p_j - p_i}{\|p_j - p_i\|} \frac{p_k - p_i}{\|p_k - p_i\|} \right\rangle$ of all the triangles $\{i, j, k\}$ adjacent to vertex $i \in \mathcal{V}$. If this angle is exactly 2π , the local neighborhood around

⁶The discrete first and second fundamental forms are required to satisfy an integrability condition, similar to the Gauss's equation and the Mainardi-Codazzi equations in the smooth setting [320].

the vertex is intrinsically flat; see for instance the red or green neighborhood in Fig. 17.5. An angle $\Theta_i < 2\pi$, as for the blue neighborhood, signals a positive discrete Gaussian curvature (properly defined below) $\kappa_{\text{Gauss},i} = 2\pi - \Theta_i$, i.e. a cone-like neighborhood. An angle $\Theta_i > 2\pi$ corresponds similarly to a saddle-like neighborhood with negative Gaussian curvature. The approach followed in [231, 347, 158, 268, 60] is then to *flatten* the one-ring neighborhood out by isotropically *rescaling polar angles* by a factor of $s_i = \frac{2\pi}{\Theta_i}$ to the total $s_i\Theta_i = 2\pi$ of a Euclidean (tangent) space. A vector field can be represented by one vector per vertex. A choice of gauge, which are often aligned with one of the edges, allows then to encode tangent or feature vector fields in terms $2|\mathcal{V}|$ or $c|\mathcal{V}|$ coefficients, respectively. Zhang et al. [347] proposed to interpolate the vectors with a piecewise linear hat function weighting from the vertices to the faces. The direction of the vectors is thereby determined by a usual Euclidean transport on the flattened tangent spaces. As pointed out by de Goes et al. [65], this interpolation is not continuous. To resolve this issue, the same authors propose in [186] to define a smooth structure on the triangle mesh and to represent the one-ring neighborhoods in *smooth charts*. A smooth interpolation is then performed by transporting vectors via a smooth simplicial connection on the mesh, which is optimized to be as close as possible to the original embedding space induced Levi-Civita connection. Note that both approaches are effectively flattening the geometry around the vertices, that is, they do not exactly operate on the triangle mesh.

Yet another approach, rooted in discrete exterior calculus [68, 81], is to define tangent vectors $v = \omega^\sharp \in T_p M$ in terms of *1-forms* $\omega \in T_p^* M$ by leveraging the (metric-dependent) musical isomorphism $\sharp^n : T^* M \rightarrow TM$ (“index raising”). Since simplicial 1-forms are naturally assigned to *edges* (1-simplices), this leads to vector fields which are parameterized in terms of one vector per edge and thus $2|\mathcal{E}|$ coefficients after choosing frames. However, as argued by de Goes et al. [65] a piecewise linear interpolation of the vectors over the faces will again lead to discontinuities. It is furthermore not clear to us how this approach could be generalized to general associated vector bundles and thus feature fields.

Given any of the above constructions of tangent spaces, *local reference frames* are readily defined as 2-tuples of linearly independent tangent vectors. A common choice is thereby to align the first frame axis with one of the adjacent edges of the current simplex (mesh element). Specifically for the case of orthonormal, right-handed frames, i.e. whenever $G \leq \text{SO}(2)$, a choice of (oriented) edge determines a frame completely. Tangent vectors are then often represented in polar coordinates, with the angle measured relative to the reference edge. If the tangent spaces are modeled extrinsically, that is, as two-dimensional subspaces of the ambient space \mathbb{R}^3 , it is most common to represent the frames explicitly as a 2-tuple of vectors in $T_p M \subset \mathbb{R}^3$. The definitions of frames and gauges are then fully equivalent to those in Eqs. (17.17) and (17.12) in Section 17.1.

G -structures are, as usual, defined as bundles of frames, which are in each tangent space related through G -valued gauge transformations. In the computer graphics community, there is a particular interest in *N-direction fields* (or unit N -RoSy fields), which are there defined as a collection of N unit vector fields, such that the N vectors in each tangent space are spaced by an angle of $2\pi/N$. Since any unit vector implies on an oriented manifold a corresponding right-handed, orthonormal frame, N -direction fields are seen to be equivalent to C_N -structures. An example is the C_6 -structure on the icosahedron in Fig. 17.6c, which effectively assigns 6 unit directions to each point, except for the poles, where it has singularities of index $\frac{1}{6}$ (or angle $\frac{2\pi}{6}$). The interactive design of smooth direction fields, with user defined singularities amongst other constraints, is an active field of research in the computer graphics community [181, 242, 172, 62, 158, 186, 268]. Some of the surface

GM-convolutions that we review in the following section use such algorithms to compute a C_N -structure [131, 339].

Riemannian metric and isometries: Having a mesh equipped with tangent spaces, one can define a *Riemannian metric* on it. The most common case that of isometrically embedded meshes with tangent spaces modeled as two-dimensional subspaces of the ambient space \mathbb{R}^3 . As described before in Eqs. (17.9) and (18.3), the metric is then induced by restricting the standard Euclidean inner product $\langle \cdot, \cdot \rangle_{\mathbb{R}^3}$ of the embedding space to the tangent spaces.

If the tangent spaces are modeled intrinsically, a metric can be fixed by choosing an $O(d)$ -structure, i.e. reference frames that are *defined* to be orthonormal. Somewhat less tautological, if one is given edge lengths $l_{\{i,j\}}$, and therefore a piecewise defined Euclidean distance function on the surface as discussed above, the choice of $O(d)$ -structure is required to be compatible with these lengths. Specifically, the logarithmic map should result in tangent vectors of (Riemannian) norm $|\log_p(q)| = d$ if the points p and q are separated by a Euclidean distance $d \in \mathbb{R}_{\geq 0}$. Note that this statement requires a consistent definition of Levi-Civita connection on the mesh, which we discuss further below.

Isometries are intrinsically defined as usual, that is, as those mappings of the mesh to itself, which preserve the metric. Extrinsicly, the isometry group is comprised of those isometries $\phi \in E(2)$ of the embedding space, which leave the mesh invariant. Most of the papers in rows (41-45) of Table 14.1 consider datasets whose meshes have a trivial isometry group. However, local neighborhoods of the meshes are often nonetheless isometric (or approximately isometric) to each other, which was exemplified in Fig. 13.2. As discussed in Section 13.1.1, the isometry equivariance of *GM*-convolutions will still hold locally if the kernels' field of view is sufficiently small.

Connections, transporters and geodesics on triangle meshes: The last ingredient that we need to implement *GM*-convolutions on meshes is the transporter pullback Exp_p^* of feature fields, Eq. (9.21). We are therefore required to know how to 1) parallel transport feature vectors over meshes and 2) compute geodesics on meshes, specifically the exponential map or, depending on the implementation, the logarithmic map. All of these mappings depend ultimately on a choice of *connection* on the mesh. In the smooth setting, a connection is essentially a collection of infinitesimal transporters between adjacent tangent spaces. One defines discretized connections on meshes therefore usually as transporters between adjacent mesh elements. The particular choice of tangent bundle discretization, options of which were discussed above, influences the particular definition of connection. In the following, we review some discretizations of connections found in the literature and explain how they can be used to compute transporters and geodesics.

The simplest case to consider is the transport or connection between two adjacent faces. Recall that the Levi-Civita transport on a flat plane is defined as shifting a vector such that it stays parallel in the usual Euclidean sense; see Fig. 8.4a. As connections are inherently intrinsic, they do not depend on the particular embedding of this plane into ambient space, which tells us how to transport on any developable surface. It tells us in particular how to transport between two adjacent triangles, since they can be unfolded (developed) into a plane as visualized in Fig. 18.2. The Levi-Civita connection between faces can therefore be thought of as 1) flattening the faces 2) transporting the vector as usual on the plane and 3) folding the faces back to their original embedding [62, 213, 60]. The resulting transporter $\mathcal{P}_{TM, \{i,j,k\} \rightarrow \{i,j,l\}}$ between the faces $\{i,j,k\}$ and $\{i,j,l\}$ can optionally be expressed in terms of a group element $g_{\{i,j,k\} \rightarrow \{i,j,l\}}^{A\tilde{A}}$ in $GL(2)$. Since the Levi-Civita connection is a

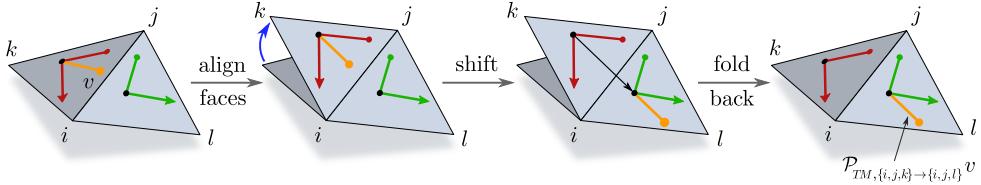


Figure 18.2: Parallel transport between mesh faces. The local geometry of two adjacent faces is developable, that is, it is intrinsically flat and can be unfolded into a plane. The Levi-Civita transport between the faces is therefore given by shifting a vector over the unfolded faces, followed by bending the faces back to their original embedding. This parallel transport between adjacent faces can be viewed as the discrete analog of the continuous Levi-Civita connection in the smooth setting [62]. Given any choice of reference frames, the transport $P_{TM, \{i,j,k\} \rightarrow \{i,j,l\}}$ is represented by a group element $g_{\{i,j,k\} \rightarrow \{i,j,l\}}^{A\bar{A}} \in GL(2)$ (or $SO(2)$ when considering right-handed, orthonormal frames). More general connections apply an additional linear transformation to the coordinate free vector when transitioning between the faces. Alternative definitions of discrete connections, for instance for the transport between vertices along edges, are discussed in the main text.

metric connection, it results for the specific case of orthonormal frames in group elements in $O(2)$, and for oriented orthonormal frames in $SO(2)$ -elements or rotation angles.

As proposed by Crane et al. [62], it is possible to generalize this construction beyond Levi-Civita connections: instead of merely shifting vectors between the flattened faces, more general connections apply an additional linear transformation, for instance an additional rotation. While this additional transformation will be reflected by a corresponding transformation of the transporter's coordinate expression $g_{\{i,j,k\} \rightarrow \{i,j,l\}}^{A\bar{A}}$, it is conceptually independent from it and can be defined in a purely coordinate free setting. The authors use this idea to construct smooth trivial connections, which are defined by having a transport of zero holonomy around any possible loop, and which are optimized to be as smooth as possible, except for at some singularities, which are topologically enforced [62]. They consider furthermore connections which apply (coordinate free) rotations by $\frac{2\pi}{N}$ and can be used to construct N -direction fields, corresponding to C_N -structures. In our applications, we will always consider Levi-Civita connections to compute geodesics. The models reviewed in the following Section 18.2 assume a structure group $G = SO(2)$ on oriented meshes and utilize Levi-Civita transporters for feature vectors. In contrast, the models in Section 18.3 assume a trivial structure group $G = \{e\}$ and therefore allow only for $\{e\}$ -structure compatible trivial connections. They transport features such that their coefficient vectors relative to the $\{e\}$ -structure frames remain invariant, i.e. merely copy their numerical values.

Given embedded tangent spaces $T_p M \subset \mathbb{R}^3$ at other mesh elements like vertices or edges, this approach is naturally generalized to transitions between arbitrary mesh elements [67]: instead of aligning the faces, one could e.g. align the vertex tangent space with the adjacent face before shifting the vector. Geometrically, this operation can be thought of as the transport over a mesh whose vertices and edges are cut off in an infinitesimal neighborhood, an are replaced with a polygonal face.

An alternative definition of discrete connections is given in [158] and [268]. The authors of both papers model tangent spaces only at the vertices, where they are defined in terms of the rescaling of the total incident angle, Eq. (18.20), to 2π , as discussed above. A connection on the mesh is then given by transporters over all edges $\{i,j\} \in \mathcal{E}$, which link the adjacent vertices' tangent spaces. Since the geometric notion of unfolding triangles is hereby missing, the edge transporters are encoded via group elements relative to a source and ref-

erence frame. Specifically for the Levi-Civita connection, and orthonormal, right-handed reference frames, these group elements lie in $\text{SO}(2)$. The utility of this construction for the direct transport along arbitrary paths over the manifold is unclear, however, it is useful to solve PDEs that depend on the covariant derivative. Sharp et al. [268] showed that a solution of the vector heat equation allows nonetheless to use such connections to (indirectly) compute the parallel transport between arbitrary points on a mesh. Liu et al. [186] propose yet another construction, namely smooth simplicial connections between and within all mesh elements. They discuss furthermore how such connections can be optimized to be as close to the (non-smooth) Levi-Civita connection as possible.

A given connection determines the *parallel transport* along a path. In the smooth setting, where connections are infinitesimal transporters, the finite transport is computed by integrating the connection along the path. In the discrete setting, the transport is accordingly given by composing the individual transformations that constitute the connection between the mesh elements that are crossed by the path. For the Levi-Civita connection, this process corresponds to a flattening of all the mesh elements along the path, followed by shifting the vector over it; see Fig. 7 in [172]. The vector heat equation based method by Sharp et al. [268] computes the transport of vectors specifically along geodesics. Since it solves for the transport from a source location to *any* other location on the manifold simultaneously, this approach can be more efficient than integrating the transport for every single path individually.

The *curvature of a connection* is in the smooth setting defined as the holonomy of its transport around an infinitesimally small disk. The curvature at a vertex is in the discrete setting similarly defined as the holonomy of the transport around this vertex. For the Levi-Civita connection, this is just the Gaussian curvature, which is given by the angle defect

$$\kappa_{\text{Gauss},i} = \delta_i = 2\pi - \Theta_i, \quad (18.21)$$

where Θ_i is the total tip angle from Eq. (18.20). We refer again to the icosahedron as example, which has vanishing curvature everywhere, except for its original twelve vertices, where the angle defect (curvature) equals $\frac{2\pi}{6}$. Trivial connections have by construction zero curvature.

Lastly, we need to discuss *geodesics*. In the smooth setting, geodesics are defined as *straightest paths*, which is formalized by the statement that the covariant derivatives of their tangent vectors along the curve vanish, that is, $\nabla_{\dot{\gamma}}\dot{\gamma} = 0$. This is equivalent to the requirement that the transport of a tangent vector $\dot{\gamma}(t_0)$ along the geodesic remains tangent to it, i.e. $\mathcal{P}_{TM,\gamma(t_1)\leftarrow\gamma(t_0)}\dot{\gamma}(t_0) = \dot{\gamma}(t_1)$ for arbitrary t_0 and t_1 . Furthermore, the *shortest path* between any two points on a connected manifold is given by a geodesic. As pointed out by Polthier and Schmies [231], this equivalence of shortest and straightest paths does not longer hold on meshes, such that one needs to distinguish between the two concepts.

Recall that the *exponential map* $\exp_p : T_p M \rightarrow M$ is defined as mapping vectors v to that point which is reached when walking for a distance of $\|v\|$ from p along the (unit speed) geodesic in direction of v . This concept is readily generalized to meshes, where one follows the *straightest geodesic* in the direction of v for distance $\|v\|$. As in the smooth setting, one may define such straightest geodesics on meshes as those curves that keep their tangent vector parallel to the curve. This property is naturally satisfied on the planar faces (or along edges), such that the resulting geodesic is *piecewise linear*, with the only nontrivial points being those where the geodesics transitions between adjacent mesh elements. The outgoing direction of the geodesic after such a transition is thereby determined by the connection, i.e. by the transport of the incoming tangent direction to the next mesh element. If one considers the Levi-Civita connection, which we always do to compute geodesics, this results

in an ordinary straight line after unfolding the mesh elements into a plane. To implement the discrete exponential map, it is sufficient to trace out such a straightest geodesic until reaching the distance $\|v\|$.

Logarithmic maps $\log_p : M \rightarrow T_p M$, on the other hand, can be thought of as computing the *shortest geodesics* between points p and q .⁷ They return that vector $\log_p(q)$ in $T_p M$ which is tangent to this geodesic at p and whose norm equals the geodesic distance between the points. A prominent way of computing geodesic distances from a source point (or set) p is to solve the eikonal equation

$$|\nabla \tau| = 1 \quad \text{subject to} \quad \tau(p) = 0, \quad (18.22)$$

where ∇ denotes the covariant derivative. The first part of this PDE enforces the natural requirement that the gradient of the distance function should be one, while the second part fixes the distance at the source to zero. A Fast Marching algorithm, which solves the eikonal equation on triangle meshes, was proposed by Kimmel and Sethian [150]. Given the distance function τ , the geodesic γ between p and any other point q can be traced back by following the distance gradient starting from q , i.e. by solving the ODE

$$\dot{\gamma} = -\nabla \tau. \quad (18.23)$$

With this information, we know that $\|\log_p(q)\| = \tau(q)$, with the direction of $\log_p(q)$ given by geodesic path at p . The solution by Mitchell et al. [213] generalizes the Dijkstra algorithm for computing distances along edges of a graph to a continuous version, which can cross faces and therefore operate on meshes. It computes a distance function by propagating a wavefront starting from p . The heat method by Crane et al. [63] computes geodesic distances by exploiting Varadhan's formula, which establishes a connection to the heat kernel. Their algorithm is essentially solving the heat equation $\dot{u} = \Delta u$ with initial condition $u_0 = \delta(p)$, i.e. it diffuses a “heat spike” from the source point p . For short diffusion times, the gradient ∇u points exactly in the opposite direction of the geodesic distances’ gradient. Since it is known that the geodesic distance gradient has unit magnitude (Eq. (18.22)), one can compute the distance field from this information. The method is substantially faster than previous algorithms. Sharp et al. [268] generalize this method to the vector heat equation, which allows to diffuse vector-valued quantities instead of scalar heat. The algorithm can be used to transport vectors from a source point (or set) over the whole manifold, but it also suitable for solving with high accuracy for logarithmic maps.

18.2 Rotation-steerable surface convolutions

In this section we review the $\mathrm{SO}(2)$, C_N and D_N -steerable surface convolutions that are listed in rows (41-44) of Table 14.1. All of these models have in common that they address the ambiguity of reference directions on general surfaces via a locally rotation equivariant (or invariant) design, which distinguishes them from the $\{e\}$ -steerable models discussed in the following Section 18.3. Before discussing the individual models in detail, we start with a higher level overview of common design choices and possible numerical discretizations.

General remarks and overview: All of the models that are reviewed in this section operate on *triangle surface meshes* and are rotation-steerable. The continuous structure group

⁷Strictly speaking, the logarithmic map \log_p can only be defined on that subset of M to which \exp_p maps injectively.

$G = \text{SO}(2)$ is for all models that assume regular field representations (rows (42) and (43)) discretized by cyclic groups C_N , i.e. N equally spaced directions. The model by Huang et al. [131] assumes a more specific structure group D_4 . Note that the purely rotation-steerable architectures operate only on *oriented surfaces* without violating the smoothness (continuity) of their inference. Non-oriented surfaces require additional reflection-steerability, i.e. structure groups $O(2)$ or D_N . This requirement is often easily satisfiable with minor adaptations, most importantly by using further restricted kernel spaces.

In accordance with the definition of *GM*-convolutions, the models parameterize features in the local neighborhood around each sampling point in terms of *geodesic normal coordinates*. Almost all of the models sample feature fields on the *mesh vertices*; only Huang et al. [131] samples features densely on the mesh faces. The continuous convolution integral in Eq. (9.39), which matches the features in geodesic normal coordinates with a steerable kernel, can be discretized in different ways. The majority of models discretize this integral at a vertex $p \in \mathcal{V}$ as a summation over its neighboring vertices $\mathcal{N}_p \subset \mathcal{V}$. Features from these vertices $q \in \mathcal{N}_p$ are then matched with the values of the continuous kernel at point $\psi_{TM,p} \log_p(q) \in \mathbb{R}^2$, where $\psi_{TM,p}^A$ is the gauge corresponding to the chosen reference frame at p . Together with the transport from q to p , this results in the discretization

$$f_{\text{out}}^A(p) = \sum_{q \in \mathcal{N}_p} A_q K(\psi_{TM,p}^A \log_p(q)) \rho(g_{p \leftarrow q}^{A \tilde{A}}) f_{\text{in}}^{\tilde{A}}(q), \quad (18.24)$$

where $A_q \in \mathbb{R}$ are suitably chosen area weights that sum to the total mesh area, $\sum_{q \in \mathcal{V}} A_q = \int_M 1 dp$. Common choices are barycentric area weights of the form

$$w_q = \frac{1}{3} \sum_{\{i,j,q\} \in \mathcal{F}} A_{\{i,j,q\}}, \quad (18.25)$$

with the sum running over all triangles that are adjacent to vertex q , or Voronoi areas [310]. Since the discretization in Eq. (18.24) sums over neighboring vertices, the algorithms compute log maps via *shortest geodesics* between q and p ; see Section 18.1.2 and [231].

Instead of computing logarithmic maps of neighboring vertices, one can alternatively discretize the convolution integral on the kernel domain \mathbb{R}^2 . The authors of [204] use an equiangular and equiradial binning of geodesic polar coordinates. They compute the exponential map for each sampling point (r, φ) , that is, they shoot a *straightest geodesic* ([231]) of length r in direction φ relative to the reference frame. As these geodesics end in general in a face, the feature vectors from adjacent vertices need to be interpolated, for instance based on barycentric coordinates. Yang et al. [339] approximate the geodesic neighborhood via a “parallel transport unfolding” algorithm [29].

Table 14.1 organizes the models by their respective *field types*, i.e. by the group representations ρ that specify their transformation laws under gauge transformations. The only non-trivial field types used so far are (complex) *irreducible representations* of $\text{SO}(2)$ [327] and *regular representations* of $\text{SO}(2)$, discretized by regular representations of a discrete subgroup C_N [232, 293, 67, 339]. Regular representations of $\text{SO}(2)$ act by definition on functions on $L^2(\text{SO}(2))$, that is, on features which assign “one value per direction”. In the discretized version, we have $L^2(C_N) \cong \mathbb{R}^{|C_N|} = \mathbb{R}^N$, where each of the N dimensions of a regular feature vector corresponds to one of the directions in $\{k \frac{2\pi}{N} \mid k = 0, \dots, N-1\}$. The correspondence to regular representations is in most of these papers implicit – the network architectures are rather derived from a more intuitive viewpoint. It turns out that the authors use only a subset of the complete space of steerable kernels that map between C_N -regular feature fields. We substantiate this claim further below when discussing the models in detail.

A construction of the complete kernel space is given in [322], a visualization can be found in Fig. 3 of [324]. The remaining models are based on *trivial representations*, i.e. *scalar fields*. One approach to compute scalar fields is to apply a kernel in N directions, resulting in an intermediate C_N -regular feature field, followed by a pooling operation over the N responses [204, 217, 293]. Since gauge transformations in C_N will lead to a mere cyclic shift (a permutation) of the feature’s direction channels, the pooling operations are *invariant* under gauge transformations, i.e. result in scalar fields. Huang et al. [131] uses immediately D_4 -invariant kernels; see Fig. 18.5. As gauge transformation leave such kernels invariant, the resulting feature fields are invariant as well, i.e. scalar fields.

Lastly, we can compare the models by the *feature transporters* that they assume. All of the convolutional networks in [327, 232, 293, 67] assume the canonical *Levi-Civita* transporters on the mesh. As all of the models in [204, 217, 293, 131] rely on *scalar fields* their parallel transport is trivial. An alternative approach was followed by Yang et al. [339] who compute a C_N -valued connection on the mesh. This connection is flat (trivial) everywhere except for at a few singularities with a holonomy of $k \frac{2\pi}{N}$ for some $k = 0, \dots, N - 1$ and fixed N . The authors optimize their C_N -valued connection such that it approximates the $SO(2)$ -valued Levi-Civita connection as close as possible; see also [62]. Note that this approach is similar to the local flattening of spherical CNNs into icosahedral CNNs ($N = 6$) from Section 17.4 but applies to general meshes.

With these general remarks in mind, we focus on some more specific design choices that are made in the models.

Harmonic Surface Networks: The *Harmonic Surface Networks* by Wiersma et al. [327], listed in row (41) of Table 14.1, are a prototypical example of *GM*-convolutions on meshes. They generalize Harmonic Networks [335] – whose features transform according to the complex irreps of $G = SO(2)$; see rows (35–38) of Table 6.6 in Chapter 6 – from the Euclidean plane to general curved spaces. The authors define their convolution as in Eq. (18.24), using the barycentric area weights from Eq. (18.25). Levi-Civita transporters and logarithmic maps are computed via the vector heat method [268], which is not restricted to triangle meshes but allows to apply the model to polygon meshes and point clouds. The $SO(2)$ -equivariant nonlinearities used by the models act only on the absolute value of the complex features but leave their argument invariant.

As proven in [173, 322], the $SO(2)$ -steerable kernel spaces that are used by the authors are complete over the complex field. However, if the complex feature fields are implemented in terms of two channels that contain their real and complex parts, they should rather be viewed as transforming according to the real irreps of $SO(2)$, as derived in Section 5.3.4. The kernel constraint allows in this case for additional steerable kernels; see Appendix F.5 of [322] for a detailed discussion. We furthermore want to mention that empirical evidence suggests that networks which are based on irrep fields perform significantly worse than those that are based on regular representations; see e.g. the benchmark in [322]. Note that Harmonic Surface Networks can easily be turned into networks that operate on regular feature fields by employing the “regular nonlinearity” from [67], which essentially applies a Fourier transformation of a stack of irrep fields to transform them into a regular feature field.

Multi Directional Geodesic CNNs: Poulenard and Ovsjanikov [232] proposed *Multi Directional Geodesic CNNs* (MDGCNNs) which operate on so-called *directional functions*. As we argue in the following, directional functions are equivalent to regular feature fields and MDGCNNs are specific *GM*-convolutions between such features. The authors define directional functions as real-valued function that depend on points $p \in M$ and unit directions

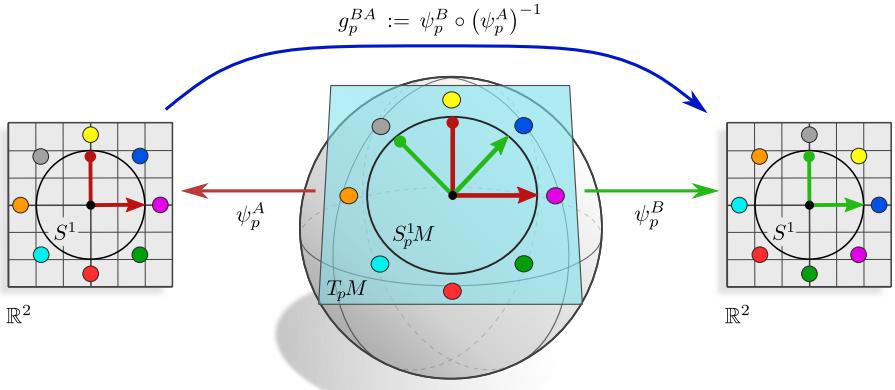


Figure 18.3: Visualization of the *directional functions* by Poulenard and Ovsjanikov [232]. Directional functions assign a real-valued response (colored dots) to each direction (unit vector) in $S_p^1 M \subset T_p M$ (black circle). When expressing these functions relative to right-handed, orthonormal reference frames or gauges $\psi_{TM,p}^X$, the coordinate representations assign real-valued responses to unit vectors in $S^1 \subset \mathbb{R}^2$. The transformation law between these coordinate representations is given by a rotation of the feature values on S^1 . Mathematically, this transformation law is identified as the action of the *regular representation* of $SO(2)$; see Eq. (18.28). Directional functions are therefore regular feature fields, and the surface CNN of Poulenard and Ovsjanikov [232] is based on GM-convolutions between such fields. A diagrammatic version of this figure is given in Eq. (18.29).

$v \in T_p M$, $\|v\| = 1$. Denoting the circle of unit directions in $T_p M$ by

$$S_p^1 M := \{v \in T_p M \mid \|v\| = 1\} \cong S^1, \quad (18.26)$$

a directional feature at p is defined as a map

$$F : S_p^1 M \rightarrow \mathbb{R} \quad (18.27)$$

from unit directions in the tangent plane to real-valued responses.⁸ A choice of right-handed, orthonormal reference frame fixes a reference direction relative to which the directional function can be expressed. Let $\psi_{TM,p}^A$ be the gauge corresponding to a chosen frame, which maps the unit directions in $S_p^1 M \subset T_p M$ to “coordinate unit directions” in $S^1 \subset \mathbb{R}^2$. The coordinate expression of the directional function is then given by

$$F_p^A := F \circ (\psi_{TM,p}^A|_{S_p^1 M})^{-1} : S^1 \rightarrow \mathbb{R}, \quad (18.28)$$

⁸The full directional function can then be defined as a map from $S^1 M$, the bundle with fibers $S_p^1 M$, to real values.

that is, it assigns real-valued responses to the unit coefficient vectors on \mathbb{R}^2 . From the commutativity of the diagram

$$\begin{array}{ccccc}
 & & g_p^{BA} & & \\
 & \swarrow & \downarrow & \searrow & \\
 \mathbb{R}^2 \supset S^1 & \xleftarrow{\psi_{TM,p}^A|_{S_p^1 M}} & S_p^1 M & \xrightarrow{\psi_{TM,p}^B|_{S_p^1 M}} & S^1 \subset \mathbb{R}^2 \\
 & \searrow F_p^A & \downarrow F & \swarrow F_p^B & \\
 & & \mathbb{R} & &
 \end{array} \tag{18.29}$$

one can read off that the coordinate expressions of directional functions obey the following transformation law:

$$F_p^B = F_p^A \circ (g_p^{BA})^{-1} =: \rho_{\text{reg}}(g_p^{BA}) F_p^A \tag{18.30}$$

The second equality identified the transformation law between the coordinate expressions as the action of the regular representation (Def. B.5.18), which justifies our statement that directional functions are just regular feature fields.⁹ Fig. 18.3 shows a directional function and its coordinate representations relative to different frames.

The multi directional geodesic convolutions by Poulenard and Ovsjanikov [232] map in a coordinate independent manner between directional functions by contracting them with equivariant kernels in a geodesic parametrization around each vertex. This observation implies that these convolutions are specific *GM*-convolutions between regular feature fields. A difference in the formulation of multi directional geodesic convolutions is that their transporter pullback does not transport the whole regular feature vector (directional function) back along the geodesics, but only that single response that corresponds to the tangent direction of the geodesic. Instead of matching the transported features with a matrix-valued kernel, multi directional convolutions match the single transported response with a scalar kernel. The equivalence of both operations is restored by imposing a corresponding sparsity pattern to our matrix-valued $\text{SO}(2)$ -steerable kernels, effectively zeroing out those responses that are not transported back by MDGCNNs. While multi directional geodesic convolutions are just *GM*-convolutions between regular feature fields, they do therefore not use the complete space of G -steerable kernels between regular feature fields. This sparsity makes MDGCNNs computationally efficient, however, the memory cost remains the same and it is unclear how severely this choice limits their expressional capacity.

The infinite number of directions in $\text{SO}(2)$ (or $S_p^1 M$ or S^1) is in practice discretized to the N equally spaced directions in the cyclic group C_N , e.g. the 8 directions that are visualized in Fig. 18.3. Since the Levi-Civita transport along features is in general $\text{SO}(2)$ -valued instead of C_N -valued, the authors use a linear interpolation between the N discrete directions.

As discussed above, MDGCNNs transport only those specific responses of the features back which correspond to the direction of the emanating geodesic relative to the local reference frame at p . This direction is undefined at the origin $v = 0 \in T_p M$, which prevents self-interactions of the vertices. The authors resolve this issue by applying an additional 1×1 -convolution which adds the missing self-interaction back. As derived in Section 9.1.1,

⁹Strictly speaking, the regular representation of $\text{SO}(2)$ acts on functions $\text{SO}(2) \rightarrow \mathbb{R}$. However, we can canonically identify such functions with functions on S^1 by identifying $(1, 0) \in S^1$ with $\{e\} \in \text{SO}(2)$.

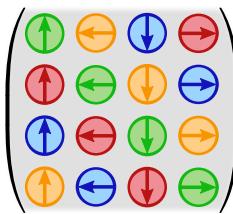


Figure 18.3: Degrees of freedom of a C_N -steerable kernel $K : \mathbb{R}^2 \rightarrow \mathbb{R}^{N \times N}$ which maps between feature fields that transform according to the regular representation $\rho_{\text{reg}} : C_N \rightarrow \text{GL}(N)$ for $N = 4$; see Table B.1 and [322]. The kernel constraint, Eq. (9.37), enforces the color-coded weight sharing pattern. The PFCNNs of Yang et al. [339] convolve features on each sheet of their N -direction field (C_N -structure) with rotated versions of a single scalar-valued kernel but do not include interaction between the different sheets. This shared kernel corresponds to the diagonal entries (green) of the complete kernel space. Off-diagonal entries, which are implicitly forced to zero, would correspond to interactions between the sheets.

the 1×1 -convolution kernels are required to be *intertwiners* in order to preserve the coordinate independence of the model. This requirement is indeed satisfied by MDGCNNs¹⁰ as the 1×1 -convolution matrix is constructed such that it mixes whole regular feature vectors with the same weight instead of linearly combining their channels independently. This is implemented by representing m_{in} regular C_N -features not as a $c = N \cdot m_{\text{in}}$ -dimensional feature vector but as an array of shape (N, m_{in}) , and then applying a (shared) matrix of shape $(m_{\text{out}}, m_{\text{in}})$ over the last axis which results in an output array of shape (N, m_{out}) .

Parallel Frame CNNs: The *Parallel Frame CNNs* (PFCNNs) by Yang et al. [339] rely on *N-direction frame fields*, which are just G -structures GM for cyclic structure groups $G = C_N$. Recall from our discussion above that these fields encode a connection which is trivial everywhere but at a few singularities and which is optimized to approximate the original Levi-Civita connection. As this G -structure is precomputed in an offline step, we take it in the following as given and focus on the actual PFCNN convolution. It turns out that this operation is equivalent to a GM -convolution between C_N -regular feature fields, however, again assuming specific sparsity pattern in the kernels that is implied by the particular network design.

The feature spaces of PFCNNs are the spaces $C^\infty(GM)$ of real-valued functions on GM . Since $GM \xrightarrow{\pi_{GM}} M$ is for $G = C_N$ a $|G| = N$ -fold cover of M , such feature fields can analogously be seen as assigning a tuple of N real numbers to each point $p \in M$. As the N sheets of the covering space are furthermore identified with N directions (given by the first frame axes), these features are equivalent to the (discretized) directional functions of Poulenard and Ovsjanikov [232]. Theorem J.0.1 in Appendix J proves furthermore that there is an isomorphism

$$C^\infty(GM) \cong \Gamma(\mathcal{A}_{\rho_{\text{reg}}}) \quad (18.31)$$

between the features of PFCNNs and our *regular feature fields*. PFCNNs are therefore performing coordinate independent convolutions between (an equivalent to) regular feature fields, and are thus identified as (specific) regular GM -convolutions.

The formulation of parallel frame convolutions seems at first glance to be quite different from ours: instead of convolving the full N -dimensional regular feature fields with a matrix-valued C_N -steerable kernel $K : \mathbb{R}^2 \rightarrow \mathbb{R}^{N \times N}$, PFCNNs convolve their scalar functions on each of the N sheets independently with a shared scalar-valued kernel which is aligned with the frame of the respective sheet. This operation is in our framework interpreted as a convolution with a matrix-valued C_N -steerable kernel whose only non-zero values are on its diagonal and are rotated relative to each other, which is visualized by the green entries in

¹⁰Personal correspondence with the author.

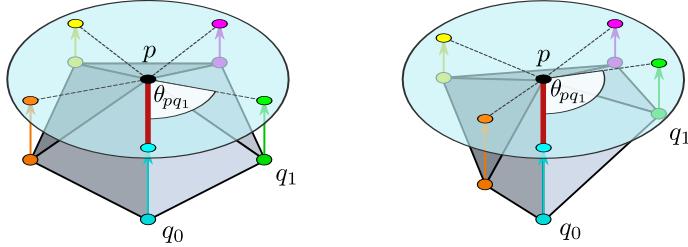


Figure 18.4: Two mesh regions which are topologically equivalent but geometrically distinct. One approach to define convolutions on meshes is to consider their underlying graph $(\mathcal{V}, \mathcal{E})$, which captures the mesh topology, and run a graph neural network on it. Lacking information about the mesh geometry, (conventional) graph neural networks can not distinguish between the two visualized neighborhoods. Geometrically, they apply *isotropic kernels*. The *Gauge Equivariant Mesh CNNs* by de Haan et al. [67] address this issue by projecting the neighboring vertices q_i on the tangent planes and assigning them angles θ_{pq_i} relative to some reference edge, i.e. gauge (red). Requiring the coordinate independence of the convolutions leads to G -steerable kernels. While the model can discriminate based on the *direction* of the neighboring nodes, it ignores their *distance*. It furthermore deviates from our geodesic parametrization in that its kernel support is based on the local edge connectivity instead of geodesic distances.

Fig. 18.3. The missing coupling between features on different sheets implies that the off-diagonal entries (yellow, blue and red) of the complete steerable kernel space are implicitly set to zero. As already stated for MDGCNNs, the sparsity pattern of this regular GM -convolution makes it computationally more efficient than a dense GM -convolution but is likely to affect its performance and does not save memory cost.

Gauge equivariant Mesh CNNs: The *Gauge Equivariant Mesh CNN* (GEMCNN) by de Haan et al. [67] is motivated by the shortcomings of conventional graph neural networks for the processing of feature fields on meshes. Specifically, vanilla graph neural networks could be used to process vertex-sampled feature fields on meshes by convolving over the graph $(\mathcal{V}, \mathcal{E})$ that is induced by the mesh. The issue with this approach is that the graph encodes only the mesh *topology*, but is not able to capture its *geometry*. Conventional graph convolutions do accordingly not distinguish between the ordering of edges, which corresponds on meshes to the use of *isotropic kernels* that map between scalar fields. Fig. 18.4 shows two regions of a mesh with distinct geometry but equivalent topology – for conventional graph convolutions both neighborhoods look the same. GEMCNNs address this issue by choosing a reference edge at each vertex $p \in \mathcal{V}$, relative to which the direction of all other edges $\{p, q_i\} \in \mathcal{E}$ to the one-ring of neighbors $q_i \in \mathcal{N}_p \subset \mathcal{V}$ is measured in terms of angles $\theta_{pq_i} \in [0, 2\pi)$. A choice of reference edge corresponds to a choice of orthonormal, right-handed frame. Different choices are related by gauge transformations in the structure group $G = \text{SO}(2)$.

As in our theory, the feature spaces of GEMCNNs are defined as sections of associated vector bundles, i.e. as spaces of c -dimensional feature fields whose coefficients transform under gauge transformations according to some group representation $\rho : \text{SO}(2) \rightarrow \text{GL}(c)$. Each edge is assigned an $\text{SO}(2)$ -valued Levi-Civita transporter. The convolution operation is demanded to be independent from the choice of reference edge, which leads to the requirement on the kernels to be G -steerable (gauge equivariant). In contrast to our formulation, the kernels are not directly applied in geodesic normal coordinates but pass messages only from the one-ring neighborhoods $\mathcal{N}_p := \{q \in \mathcal{V} \mid \{p, q\} \in \mathcal{E}\}$ to that node p around which the kernel

is centered.¹¹ The kernels are furthermore *radially insensitive* – in how far this affects the model performance remains an open question.

The authors decided for (real) *irreps* as field types for the convolution, however, they perform a change of basis to *regular representations* to apply ReLU nonlinearities, which is why we list them in row (42) of Table 14.1 instead of row (41).¹² Specifically, the authors use the change of basis $Q \in \mathbb{R}^{N \times N}$ that decomposes the regular representation $\rho_{\text{reg}} : C_N \rightarrow \text{GL}(N)$ of C_N into its irrep components to transform a stack of irrep fields into one regular feature field; cf. Section 5.3.2. For C_N , this matrix is just the discrete Fourier transform. After applying the ReLU nonlinearity to each of the N channels of the regular feature field individually – which is a C_N -equivariant operation since regular representations are permutation representations – the features are transformed back to a stack of irrep fields for the following convolution operation. This design has the advantage that the features can be transported exactly with $\text{SO}(2)$ -valued transporters, without having to fall back to an interpolation scheme, as done by Poulenard and Ovsjanikov [232]. Note, however, that the full network is due to the use of the regular nonlinearities only C_N -equivariant.

That the authors use the *real* irreps of $\text{SO}(2)$ means that their kernel spaces are approximately twice as large as those of the Harmonic Surface Networks by Wiersma et al. [327]; cf. the discussions in [322, 173].

Geodesic CNNs: The earliest work on geodesic convolutions that we are aware of is that of Masci et al. [204]. The authors identified the rotational ambiguity of geodesic polar coordinates on an oriented Riemannian manifold and address it via a rotation *invariant* architecture. Their *Geodesic convolutions* represent a *scalar* field relative to arbitrarily oriented geodesic polar coordinates. As the field type is trivial, the transporter pullback to geodesic coordinates does not require (non-trivial) transporters. The feature field in geodesic coordinates is then matched with a scalar kernel, which is applied in N equally spaced rotations by angles $\frac{2\pi}{N}k$ relative to the reference frame, where $k = 0, \dots, N - 1$. Since a gauge transformation by $\frac{2\pi}{N}l$ for some $l \in \{0, \dots, N - 1\}$ rotates all kernels accordingly, it result in a cyclic permutation of the responses by l steps. This operation corresponds therefore in our framework to a C_N -steerable convolution from scalar fields to C_N -regular feature fields. Instead of processing these fields further via regular group convolutions – as done in MDGCNNs [232], PFCNNs [339] and GEMCNNs [67] – the authors apply a max-pooling operation over the N responses. Since C_N -valued gauge transformation result in cyclic shifts of the intermediate regular feature fields, the pooling operation is gauge-invariant, i.e. produces scalar fields. While this networks design is simple to implement, it prevents features from encoding directional information. Further variations of this networks design can be found in [205, 217].

ZerNet: Next, we turn to *ZerNet* by Sun et al. [293]. To avoid confusion, we point out that the authors proposed two models, which we list in rows (42) and (43) of Table 14.1, respectively. We describe both models, starting with their common design choices.

The key concept underlying ZerNets is their parameterization of convolution kernels in terms of *Zernike polynomials*, which form an orthogonal basis of functions on the closed unit disk

¹¹For a (sufficiently) regular grid and compactly supported kernel in geodesic coordinates both approaches become equivalent.

¹²The equivalence of ρ -fields to their *irrep decomposition* (Theorem B.5.16) was discussed in Section 10.2.

$B_{\mathbb{R}^2}(0, 1)$ around the origin of \mathbb{R}^2 . In polar coordinates, Zernike polynomials are given by

$$\text{even: } Z_n^m : [0, 1] \times [0, 2\pi) \rightarrow [-1, 1], \quad (r, \varphi) \mapsto R_n^m(r) \cos(m\varphi) \quad n \in \mathbb{N}, \quad 0 \leq m \leq n \quad (18.32)$$

$$\text{odd: } Z_n^{-m} : [0, 1] \times [0, 2\pi) \rightarrow [-1, 1], \quad (r, \varphi) \mapsto R_n^m(r) \sin(m\varphi) \quad n \in \mathbb{N}, \quad 1 \leq m \leq n, \quad (18.33)$$

where R_n^m are the Zernike radial polynomials. That (suitably normalized) Zernike polynomials are orthonormal means that they satisfy the orthonormality relations

$$\langle Z_n^m, Z_k^l \rangle_{B_{\mathbb{R}^2}(0,1)} = \int_0^1 \int_0^{2\pi} Z_n^m(r, \varphi) Z_k^l(r, \varphi) r dr d\varphi = \delta_{nk} \delta_{ml}. \quad (18.34)$$

A function on the unit disk, for instance a scalar kernel $K : B_{\mathbb{R}^2}(0, 1) \rightarrow \mathbb{R}$, can be expanded in the Zernike polynomial basis:

$$K(r, \varphi) = \sum_{n \in \mathbb{N}} \sum_{m=-n}^n \hat{K}_n^m Z_n^m(r, \varphi) \quad (18.35)$$

To retrieve the expansion coefficients of a given function on the unit disk, one projects it on the Zernike basis:

$$\hat{K}_n^m = \langle K, Z_n^m \rangle_{B_{\mathbb{R}^2}(0,1)} = \int_0^1 \int_0^{2\pi} K(r, \varphi) Z_n^m(r, \varphi) r dr d\varphi \quad (18.36)$$

The inner product between two functions K and $\text{Exp}_p^* f^A$ on the unit disk can with these relations be expressed in terms of their expansion coefficients:

$$\begin{aligned} & \langle K, \text{Exp}_p^* f^A \rangle_{B_{\mathbb{R}^2}(0,1)} \\ &= \int_0^1 \int_0^{2\pi} K(r, \varphi) \text{Exp}_p^* f^A(r, \varphi) r dr d\varphi \\ &= \int_0^1 \int_0^{2\pi} \sum_{n \in \mathbb{N}} \sum_{m=-n}^n \hat{K}_n^m Z_n^m(r, \varphi) \sum_{k \in \mathbb{N}} \sum_{l=-k}^k [\widehat{\text{Exp}_p^* f^A}]_k^l Z_k^l(r, \varphi) r dr d\varphi \\ &= \sum_{n \in \mathbb{N}} \sum_{m=-n}^n \sum_{k \in \mathbb{N}} \sum_{l=-k}^k \underbrace{\int_0^1 \int_0^{2\pi} Z_n^m(r, \varphi) Z_k^l(r, \varphi) r dr d\varphi}_{\delta_{nk} \delta_{ml}} \hat{K}_n^m [\widehat{\text{Exp}_p^* f^A}]_k^l \\ &= \sum_{n \in \mathbb{N}} \sum_{m=-n}^n \hat{K}_n^m [\widehat{\text{Exp}_p^* f^A}]_n^m \end{aligned} \quad (18.37)$$

As suggested by the choices K and $\text{Exp}_p^* f^A$ for these functions, the authors use this property to match kernels with the pullback of the feature fields to geodesic polar coordinates. The kernel coefficients \hat{K}_n^m , which are set to zero beyond a user specified threshold, are optimized as learnable parameters of the network. The expansion coefficients $[\widehat{\text{Exp}_p^* f^A}]_n^m$ of the feature field's transporter pullback are computed by solving a linear system of equations.

An advantage of the kernel parameterization in terms of Zernike polynomials is that they are by definition SO(2)-steerable kernels. Specifically, the pairs $(Z_n^m, Z_n^{-m})^\top$ of kernels for a

given $n \in \mathbb{N}$ and $1 \leq m \leq n$ form a pair of kernels that are rotated by multiplying them with the m -th order real irrep of $\text{SO}(2)$,

$$\begin{pmatrix} Z_n^m \\ Z_n^{-m} \end{pmatrix} (r, \varphi + \Delta\varphi) = \begin{pmatrix} \cos(m\Delta\varphi) & -\sin(m\Delta\varphi) \\ \sin(m\Delta\varphi) & \cos(m\Delta\varphi) \end{pmatrix} \begin{pmatrix} Z_n^m \\ Z_n^{-m} \end{pmatrix} (r, \varphi), \quad (18.38)$$

while the kernels Z_n^0 , i.e. for $m = 0$, transform trivially (they are isotropic). Note that the *expansion coefficients* \hat{K}_n^m of a kernel K transform *inversely* to the basis. The authors use this transformation law to rotate kernels analytically in terms of their expansion coefficients. The rotation steerability of the Zernike polynomials' is independent from their radial parts but relies on the fact that their angular parts are *circular harmonics* (Fig. 5.2), which are the harmonic basis functions in the Peter-Weyl decomposition (Theorem B.5.22) of $L^2(\text{SO}(2))$. Due to their steerability properties, circular harmonic bases have been extensively used to parameterize real [324, 110] and complex [335, 327] convolution kernels since at least the '80s [127, 251, 95, 118]. In fact, as discussed in detail in Section 5.3 and [322, 173], circular harmonics are underlying *any* $\text{SO}(2)$ -steerable kernel.

The first and main model design described by Sun et al. [293] is similar to that by Masci et al. [204]. A scalar field is pulled back to geodesic normal coordinates, where it is matched with a scalar kernel that is applied in N discrete rotations, resulting in an intermediate C_N -regular feature field. A subsequent max-pooling operation over the N responses yields then a C_N -invariant output, i.e. an output scalar field. The difference to the implementation by Masci et al. [204] is that this operation is performed in the Zernike polynomial basis as specified in Eq. (18.37). This choice corresponds ultimately to an alternative interpolation scheme. The second model design, described Section 4.4 of [293], is a reimplementation of the MDGCNNs from Poulenard and Ovsjanikov [232] in the Zernike polynomial basis. The authors observe that this design leads to a significantly improved performance since the regular feature fields are able to encode directional information.

TextureNet: The last rotation steerable model that we discuss is the *TextureNet* by Huang et al. [131]. In contrast to the previous models, TextureNets assume a D_4 -structure, which could easily be generalized to a D_N -structure. This D_4 -structure is precomputed via QuadriFlow, a 3rd party software package which can be used to compute 4-RoSy fields that are optimized to be smooth and have few singularities [130]. As the name suggests, TextureNets process input feature fields that are represented as textures, and are of potentially higher resolution than the mesh. The convolution kernels are applied at a dense set of sampling locations, which are uniformly distributed over the mesh's faces. At each sampling point the scalar feature field is pulled back into geodesic normal coordinates and represented relative to an arbitrary frame of the D_4 -structure. It is then matched with a D_4 -invariant 3×3 kernel. As visualized in Fig. 18.5, the 9 pixels of such kernels are described by 3 degrees of freedom. The convolution is implemented in terms of three 1×1 -convolutions whose responses are subsequently binned and aggregated in each of the tangent spaces. The additional reflection steerability of the kernels implies that TextureNets are well defined on non-orientable surfaces. However, as the features of TextureNet are scalar fields they can neither encode directions nor orientations. To overcome this issue, it is necessary to use non-trivial D_N or $O(2)$ -steerable kernels.

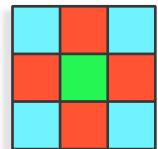


Figure 18.5: A D_4 -invariant kernel of 3×3 pixels is parameterized by three degrees of freedom.

18.3 $\{e\}$ -steerable surface convolutions

This section reviews the networks from [217, 140, 259, 294, 141, 180], which have in common that they rely on $\{e\}$ -structures on the surfaces. From the viewpoint of *GM*-convolutions, these architectures differ mainly in the specific choice of *heuristic that determines the $\{e\}$ -structure*.

Assuming a trivial structure group $G = \{e\}$, the models apply $\{e\}$ -steerable (i.e. unconstrained) kernels, which are aligned along the frames of the chosen $\{e\}$ -structure. The field types (group representations) are necessarily trivial. The same holds for all parallel transporters, which are necessarily $\{e\}$ -structure compatible. Transporter pullbacks $\text{Exp}_p^* f$ of feature fields f to the tangent spaces reduce therefore to pullbacks $\exp_p^* f$ by the usual exponential map, that is, they don't apply (non-trivial) transporters. Recall that continuous $\{e\}$ -structures exist only on parallelizable manifolds, implying that the networks' inference is inevitably discontinuous on non-parallelizable surfaces. The heuristics that determine the frame fields are furthermore not always well defined, or are unstable under deformations of the surfaces' geometry, as further discussed below.

The models of Monti et al. [217], Jin et al. [140] and Schonsheck et al. [259] operate on triangular meshes and process feature fields that are sampled at the vertices. Tatarchenko et al. [294] and Jin et al. [141] propose networks that operate on surface point clouds while the architecture of Li et al. [180] defines convolutions on texture atlases of meshes.

Geodesic MoNets: The first model family that we discuss are the *MoNets* by Monti et al. [217]. The authors discuss a variety of models on graphs and manifolds, most of which are not explained as *GM*-convolutions. These models have in common that they apply kernels relative to some choice of “pseudo-coordinates” on the manifold or graph – we are here only interested in those MoNets that rely on geodesic normal coordinates and are therefore identified as *GM*-convolutions.

As stated above, the main difference between $\{e\}$ -steerable surface convolutions is their particular choice of $\{e\}$ -structure. Inspired by previous work of Boscaini et al. [22]¹³, the authors choose to align the reference frames of the $\{e\}$ -structure with the *principal curvature direction* of the manifold. Note that this heuristic is not well defined when the principal curvatures $\kappa_{\max} = \kappa_{\min}$ agree, i.e. when the principal curvature direction is degenerate. An extreme example is the 2-sphere S^2 , where the principal curvature direction is nowhere well defined. Even when the principal curvatures are unequal, they determine only an undirected line, disambiguating reference frames up to a C_2 -structure (with the two constituent frames pointing along the two directions along the line). To make the network independent from the choice of frame, they should therefore actually apply C_2 -steerable kernels. Moreover, the principal curvature directions are unstable under deformations of the surface. As an example, imagine the principal curvature direction at the north pole (on the positive z -axis) of the 2-sphere S^2 : an infinitesimal squeezing of the sphere along the x -axis results in a principal curvature direction along the x -axis while an infinitesimal stretching along the x -axis results in a principal curvature direction along the y -axis. We furthermore want to mention that principal curvatures depend on the embedding of a manifold, that is, the approach is non-intrinsic.

¹³The *Anisotropic CNNs* by Boscaini et al. [22] assume the same principal curvature direction based $\{e\}$ -structure. However, their kernels are not defined in geodesic normal coordinates but are based on anisotropic heat kernels on the manifold. Monti et al. [217] claim that such heat kernels correspond to anisotropic Gaussian kernels in geodesic coordinates – if this statement is true, Anisotropic CNNs can be viewed as *GM*-convolutions.

3DMCNN: Jin et al. [140] proposed a *3D Mesh CNN* (3DMCNN) that convolves over the surfaces of scanned faces to recognize expressions like happiness, anger or surprise. As the face-masks are topologically planes (with holes at the eyes) they are parallelizable, which allows for smooth *GM*-convolutions for $G = \{e\}$.

The convolution kernel is discretized into one central sampling point and eight other points at a fixed radial distance R and angles $\varphi_k = k \frac{2\pi}{8}$, $k = 0, \dots, 7$ in polar coordinates. The kernels – and thus the frames that constitute the $\{e\}$ -structure – are rotated such that they are aligned with the z -axis of the embedding space \mathbb{R}^3 . This approach seems reasonable since the face masks are parallelizable and, more importantly, aligned upright. To match a such oriented kernel with a feature field, geodesics of length R are shot in the eight directions. Barycentric coordinates are used to interpolate the signal from the surrounding vertices to the end point of the geodesic.

Parallel Transport Convolutions: As a last mesh-based $\{e\}$ -steerable convolution we discuss the *Parallel Transport Convolutions* (PTCs) by Schonsheck et al. [259]. The key idea of PTCs is to define the convolution kernel at some “origin” $p_0 \in M$ and share it with any other location $p \in M$ by Levi-Civita transporting it along the shortest geodesics between p_0 and p . To formulate this weight sharing procedure in more detail, consider the closed disks $B_{T_p M}(0, R) \subset T_p M$ of radius R around the origins of the tangent spaces, where $R \in \mathbb{R}_{>0}$ is the injectivity radius of the manifold. Let furthermore $M_{p,R} := \exp_p(B_{T_p M}(0, R)) \subset M$ be the images of these disks under the exponential map, which include all points whose geodesic distance from p is smaller than or equal to R . Schonsheck et al. [259] define their (unconstrained) scalar convolution kernels than as real-valued functions

$$\hat{K}_{p_0} : M_{p_0,R} \rightarrow \mathbb{R} \quad (18.39)$$

on the neighborhood around the origin p_0 , i.e. directly on the manifold. To share the kernel with other locations $p \in M$, the authors compute the shortest geodesics between p_0 and the target locations p via Fast Marching. They parallel transport the kernel then along these geodesics, which is done by pulling them back to the tangent spaces. In equations, the kernel at p is defined as

$$\hat{K}_p : M_{p,R} \rightarrow \mathbb{R}, \quad q \mapsto \hat{K}_p(q) := \hat{K}_{p_0} \circ \exp_{p_0} \circ \mathcal{P}_{TM,p_0 \rightarrow p}^{-1} \circ \log_p(q), \quad (18.40)$$

which is visualized by the following commutative diagram:

$$\begin{array}{ccccc} M_{p_0,R} & \xleftarrow{\exp_{p_0}} & B_{T_{p_0} M}(0, R) & \xrightarrow{\mathcal{P}_{TM,p_0 \rightarrow p}^{-1}} & B_{T_p M}(0, R) & \xrightarrow{\exp_p} & M_{p,R} \\ & \text{---} \curvearrowright & & & \text{---} \curvearrowright & & \\ & & \hat{K}_{p_0} & & & & \hat{K}_p & \end{array} \quad (18.41)$$

The existence of the logarithmic map is guaranteed since the domain is restricted to points q within the injectivity radius. To compute the convolution response at p , the transported kernel is matched with the (scalar) feature field on $M_{p,R}$.

In order to describe PTCs as *GM*-convolutions, we need to identify the corresponding $\{e\}$ -structure and $\{e\}$ -steerable kernel on \mathbb{R}^2 . A compatible $\{e\}$ -structure is fixed by choosing an arbitrary frame $[e_i^A(p_0)]_{i=1}^d$ at the origin p_0 . The frames at any other location p are then determined by Levi-Civita transporting this initial frame along the shortest geodesics, that

is, they are defined as¹⁴

$$[e_i^A(p)]_{i=1}^d := \mathcal{P}_{FM, p_0 \rightarrow p} [e_i^A(p_0)]_{i=1}^d. \quad (18.42)$$

Note that this definition implies in particular the following equivalent relation for the corresponding gauges, which is easily seen by applying it to the frame field:

$$\psi_{GM,p}^A = \psi_{GM,p_0}^A \circ \mathcal{P}_{GM, p_0 \rightarrow p}^{-1} \quad (18.43)$$

Given the reference frame at p_0 , we can express \hat{K}_{p_0} in geodesic normal coordinates, which gives rise to our usual notion of template kernel on \mathbb{R}^2 :

$$K : B_{\mathbb{R}^2}(0, R) \rightarrow \mathbb{R}, \quad v \mapsto K(v) := \hat{K}_{p_0} \circ \exp_{p_0} \circ (\psi_{TM,p_0}^A)^{-1}(v) \quad (18.44)$$

To show that our weight sharing via the such constructed $\{e\}$ -structure is indeed consistent with that by Schonsheck et al. [259], we reproduce the kernels \hat{K}_p at p by mapping our template kernel K down to the manifold:

$$\begin{aligned} K \circ \psi_{TM,p}^A \circ \log_p &= \hat{K}_{p_0} \circ \exp_{p_0} \circ (\psi_{TM,p_0}^A)^{-1} \circ \psi_{TM,p}^A \circ \log_p \\ &= \hat{K}_{p_0} \circ \exp_{p_0} \circ \mathcal{P}_{TM, p_0 \rightarrow p}^{-1} \circ \log_p \\ &= \hat{K}_p \end{aligned} \quad (18.45)$$

The second step in this calculation used the equivalent to Eq. (18.43) for the tangent bundle transporter and gauges. All definitions, and their consistency, are concisely summarized by the statement that the following diagram commutes:

$$\begin{array}{ccccccc} M_{p_0,R} & \xleftarrow{\exp_{p_0}} & B_{T_{p_0}M}(0, R) & \xrightarrow{\mathcal{P}_{TM, p_0 \rightarrow p}} & B_{T_p M}(0, R) & \xrightarrow{\exp_p} & M_{p,R} \\ & \downarrow \hat{K}_{p_0} & \downarrow \psi_{TM,p_0}^A & \downarrow \mathcal{P}_{TM, p_0 \rightarrow p} & \downarrow \psi_{TM,p}^A & \downarrow \exp_p & \downarrow \hat{K}_p \\ & & B_{\mathbb{R}^2}(0, R) & & & & \\ & & \downarrow K & & & & \\ & & \mathbb{R} & & & & \end{array} \quad (18.46)$$

Since we constructed our $\{e\}$ -structure by choosing an initial frame at p_0 , the reader might wonder about the implications of this choice. A different choice of initial frame will result in a corresponding transformation of the geodesic normal coordinates at p_0 , and therefore of the template kernel K (Eq. (18.44)). However, since the $\{e\}$ -structure is constructed by transporting the initial frame, all of its reference frames will transform accordingly. The transformation of the template kernel will then cancel out with the transformation of the $\{e\}$ -structure such that all choices of initial frames are ultimately equivalent.

The $\{e\}$ -structures underlying PTCs depend crucially on the choice of origin p_0 from which the frame field is constructed – different choices of origins can lead to very different $\{e\}$ -structures. As most manifolds do not come with a canonical notion of origin, the proposed heuristic seems somewhat arbitrary. Transport based $\{e\}$ -structures, and thus PTCs, are

¹⁴Since this relation defines the $\{e\}$ -structure, we need to use the Levi-Civita transporters on the full frame bundle FM .

furthermore discontinuous at the cut locus. This implies in particular that they are close to the cut locus unstable under deformations of the surfaces' geometry since such deformations may shift the cut locus. In contrast to the heuristics of the previous models, the heuristic of PTCs depends solely on the intrinsic geometry of the surface, that is, it is not based on its embedding in ambient space.

To avoid confusion, we need to mention that Schonsheck et al. [259] construct in their implementation (Section 3.2) another frame field, which should not be confused the $\{e\}$ -structure that we described above. This frame field is required for the numerical computation of the Levi-Civita connection on the mesh, according to which the kernels are then transported. Our analysis above is purely based on their coordinate free definition of the model, most importantly the definition of weight sharing in (our) Eq. (18.40).

Note furthermore that the implicitly assumed feature vector transporters in the transporter pullback rely necessarily on the $\{e\}$ -compatible trivial connection that is implied by the $\{e\}$ -structure. The feature transport agrees along the geodesics emanating from p_0 , based on which the $\{e\}$ -structure was constructed, with Levi-Civita transporters. Transporters along any other path differ in general from the Levi-Civita transport.

Tangent convolutions: The *tangent convolutions* by Tatarchenko et al. [294] operate on point clouds $P \subset \mathbb{R}^3$ whose points are assumed to lie on a surface. Tangent spaces at the sampling points are computed via a *local principal component analysis* (LPCA). The LPCA at $p \in P$ is essentially computing the eigenvectors $e_i \in \mathbb{R}^3$, $i = 1, 2, 3$, of the covariance matrix of all points within a spherical neighborhood $\mathcal{N}_p = \{q \in P \mid \|q - p\| < R\}$ of radius R around p . As the point cloud is sampled from a surface, one of the eigenvalues should be close to zero. The corresponding eigenvector e_3 is taken as the normal vector of the embedded tangent plane $T_p M \subset \mathbb{R}^3$ at p . The two other eigenvectors span an orthonormal frame $[e_1, e_2]$ on the tangent plane, such that the collection of LPCA eigenvectors implies an $\{e\}$ -structure on the point cloud. Note that the eigenvector with the largest eigenvalue points in the direction of minimal principal curvature, that is, one has $\kappa_n(e_1) = \kappa_{\min}$ and $\kappa_n(e_2) = \kappa_{\max}$. The considered $\{e\}$ -structure is therefore similar to that of Boscaini et al. [21] and Monti et al. [217], however, the frames are rotated by $\pi/2$ since they are aligned with the minimal instead of maximal curvature direction.¹⁵ Since the sign of the eigenvectors is arbitrary, this heuristic fixes frames actually only up to rotations by π . To address this ambiguity, tangent convolutions would either have to disambiguate between the two directions or fall back to C_2 -steerable kernels.

Instead of representing the feature field in geodesic normal coordinates, tangent convolutions project the features along the normal direction on the tangent plane.¹⁶ They are then interpolated to a regular grid of $N \times N$ pixels. As this grid is aligned with the reference frame, it can be viewed as a discretization of the tangent space coordinatization $\psi_{TM,p}^A(T_p M) = \mathbb{R}^2$. The convolution computes features then by taking the inner product with a $N \times N$ pixel kernel.

NPTC-net: Jin et al. [141] proposed *NPTC-nets* on surface point clouds $P \subset \mathbb{R}^3$. Like tangent convolutions, NPTC-nets compute tangent planes via a local principal component analysis, however, their $\{e\}$ -structure is independent from the LPCA. The $\{e\}$ -structure that

¹⁵Since all reference frames are rotated by the same angle, this difference is irrelevant if the kernels are learned.

¹⁶This choice makes tangent convolutions (and NPTC-nets) different from *GM*-convolutions. In the limit of small kernels relative to the curvature of the surface both projections of feature fields to the tangent spaces become equivalent.

is underlying NPTC-nets is rather aligned with the gradient of the geodesic distance function from some initial point $p_0 \in P$. To solve for the distance function, Jin et al. [141] solve the Eikonal equation via a Fast Marching algorithm. Instead of operating directly on the point cloud as done for instance in [63], the authors propose to use a sparse voxel grid whose voxels lie in a narrow band around the point cloud. Having computed the distance function on the voxel grid, which should produce approximately geodesic distances, its gradient is computed and projected on the tangent planes. The projected vector determines the first frame axes of the $\{e\}$ -structure. Note that such defined frame fields are singular at p_0 .

Jin et al. [141] observe that this $\{e\}$ -structure implies a trivial connection on the surface (defined such that the frame field is closed under this transport). The frame field (or convolution kernels) can be understood as being transported according to this trivial connection, which motivates the “PTC” (parallel transport convolution) in the model name. Note, however, that NPTC-nets rely in contrast to the PTCs of Schonsheck et al. [259] not on the Levi-Civita transport. Moreover, this statement can be made for *any* $\{e\}$ -structure and corresponding trivial connection.

Like tangent convolutions, NPTC-nets project the features in the ambient space to the tangent plane. Instead of using a projection along the normal direction, the authors use a nearest neighbor interpolation with distances measured in ambient space. The convolution kernel is then oriented along the frames of the $\{e\}$ -structure and matched with the interpolated feature field. Given a convolution kernel $K : \mathbb{R}^2 \rightarrow \mathbb{R}$, the authors formulate its assignment to that tangent spaces as $K \circ \psi_{TM,p}^A : T_p M \rightarrow \mathbb{R}$ where $\psi_{TM,p}^A := (\langle e_1^A, v \rangle, \langle e_2^A, v \rangle)^\top$. This procedure matches our definition of weight sharing and gauges (Eq. (17.17)) exactly.

Cross-atlas convolutions: An entirely different approach was followed by Li et al. [180]. Their *cross-atlas convolutions* compute a texture atlas whose charts are optimized to be approximately isometric. The convolution operation is then performed on the texture atlas, with pixel offset maps modeling the transition maps between charts.

Before running the actual convolutions, an atlas of charts is computed. From an abstract viewpoint, the charts map patches of the surface to \mathbb{R}^2 , such that the whole surface is covered. Concretely, they map patches of a c -channel input feature field (texture) in a non-overlapping way to an array of dimensions (X, Y, c) . Since the patches in the array should approximately represent geodesic neighborhoods on the surface, the charts should be approximately isometric, i.e. minimize distortions. To satisfy this requirement, the surface is cut such into patches that the mutual angles between all triangle normals within a patch stay below a user specified threshold – note that this approach is based on the surfaces’ extrinsic geometry. After optimizing the patches on the surface, the feature field is on each patch projected along a dominant projection direction. A bin-packing algorithm packs the projected patches densely into the texture map of shape (X, Y, c) . To resolve the directional ambiguity of the patches they are required to be *rotation aligned*. This is achieved by demanding that the projections of the ambient space’s z -axis to each patch are all aligned in the texture map.

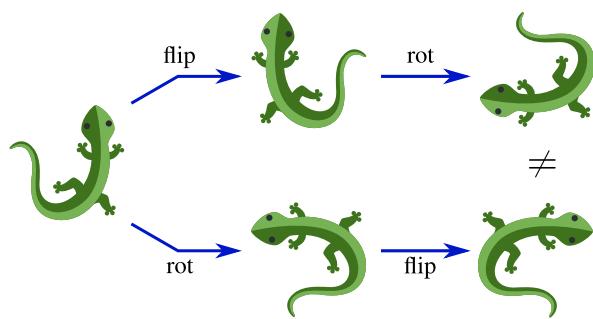
The convolution operates directly on the texture map. It groups the pixels into three different categories which are processed in a different manner. Pixels which are in the interior of a patch, such that the kernel does not range out of the patch, are convolved via conventional Euclidean convolutions. Since the charts are approximately isometric, this corresponds approximately to a geodesic convolution on the patch interior regions on the surface. Pixels that are outside of the patches are not processed, their value is fixed to zero. The interesting case is that of pixels which are close to the boundary of the patches. As the convolution kernel ranges for such pixels out of the current patch, it requires transition maps which query features from a neighboring patch on the surface. The query location is computed by

1) finding the original point on the surface that corresponds to the current kernel location, 2) shooting a geodesic to find the kernel sampling location on the surface and 3) mapping this location to the corresponding pixel in the texture map. Using these transition maps the patches are stitched together according to the surface geometry and the convolution on the texture map corresponds approximately to a geodesic convolution on the surface. In the limit of the normal angle threshold going to zero, the approximating converges to an exact geodesic convolution. However, the patches shrink then down to individual faces, leading to more non-trivial transition maps.

Cross-atlas convolutions correspond in this limit to *GM*-convolutions whose $\{e\}$ -structure is induced from the charts. The $\{e\}$ -structure is at the boundaries between adjacent patches discontinuous, however, the jumps should due to the rotation alignment of the patches in the texture map in most cases be minimized. The discontinuities are expected to be large at patches of the surface which are approximately horizontal.

For completeness, we want to point to the atlas based methods by Sinha et al. [278] and Maron et al. [199]. Both consider *non-isometric* projections of the surface to a planar domain, which implies that the subsequent Euclidean convolutions do not correspond to geodesic convolutions on the surface.

APPENDIX



List of theorems and definitions

Translation equivariant CNNs on Euclidean spaces

Definition 3.1.1	Euclidean feature maps as regular translation group representations	34
Theorem 3.2.1	Regular translation intertwiners are convolutions	36
Theorem 3.2.2	Translation equivariant bias summation	38
Theorem 3.2.3	Translation equivariant local nonlinearities	39
Theorem 3.2.4	Translation equivariance of local max pooling	39
Theorem 3.2.5	Translation equivariance of local average pooling	40
Theorem 3.2.6	Translation subgroup equivariance of subsampling	40
Theorem 3.2.7	Translation invariance of global max pooling	41
Theorem 3.2.8	Translation invariance of global average pooling	41

$\text{Aff}(G)$ -steerable CNNs on Euclidean spaces

Definition 4.2.1	Euclidean feature fields	46
Theorem 4.3.1	Steerable convolutions	51
Theorem 4.3.2	Affine equivariant bias summation	54
Theorem 4.3.3	Affine equivariant local nonlinearities	55
Theorem 4.3.4	Affine equivariance of local max pooling for permutation reps .	57
Theorem 4.3.5	Affine equivariance of local norm max pooling for unitary reps .	58
Theorem 4.3.6	Affine equivariance of local average pooling	59
Theorem 4.3.7	Affine equivariance of global max pooling for permutation reps .	60
Theorem 4.3.8	Affine equivariance of global norm max pooling for unitary reps	60
Theorem 4.3.9	Affine equivariance of global average pooling	61
Theorem 4.5.1	$\text{Aff}(G)$ group convolutions as G -regular steerable convolutions .	64

G-steerable kernels

Theorem 5.3.1	Wigner-Eckart theorem for G -steerable kernels	79
---------------	--	----

GM-coordinate independent neural networks

Definition 12.1.1	1×1 GM -convolution	210
Definition 12.2.1	Kernel field	214
Definition 12.2.2	G -steerable kernel	216
Definition 12.2.3	GM -convolutional kernel field	216
Definition 12.2.4	Transporter pullback of feature field to TM	218
Definition 12.2.5	Kernel field transform	219
Theorem 12.2.6	Kernel field transform existence for compactly supported kernels	220
Definition 12.2.7	GM -convolution	220
Theorem 12.2.8	Kernel field transform in coordinates	222
Theorem 12.2.9	GM -convolutions in coordinates	223

Isometry equivariance of coordinate independent CNNs

Definition 13.1.1	G -structure preserving isometries	230
Definition 13.1.2	Isometry pushforward of feature field:	233
Theorem 13.1.3	Isom_{GM} in local trivializations	237
Theorem 13.1.4	Isometry action on transporter pullbacks of feature fields	240
Definition 13.2.1	Isometry equivariant kernel field transform	241
Definition 13.2.2	Isometry action on kernel fields	242
Definition 13.2.3	Isometry invariant kernel fields	243
Theorem 13.2.4	Equivariant kernel field transform \Leftrightarrow invariant kernel field	244
Theorem 13.2.5	Isometry equivariance of GM -convolutions	247
Theorem 13.2.6	Full isometry equivariance of OM -convolutions	248
Theorem 13.2.7	$\text{Isom}_+(M)$ equivariance of SOM -convolutions	248
Theorem 13.3.1	Tangent quotient representative kernel fields	257
Theorem 13.3.2	Manifold quotient representative kernel fields	258
Theorem 13.3.3	Equivariance on homogeneous M implies convolution	260

Aff(G)-equivariant GM-convolutions on Euclidean spaces:

Definition 15.1.1	Aff(G)-atlas of Euclidean space	278
Theorem 15.1.2	Aff(G)-atlases of charts induce G -atlases of gauges	278
Theorem 15.1.3	Principal bundle isomorphism between Aff(G) and GM	280

Definition 15.1.4	G -structure preserving affine transformations	281
Theorem 15.1.5	Aff_{GM} in local trivializations	281
Theorem 15.1.6	Aff_{GM} in global affine charts	282
Theorem 15.2.1	Euclidean GM -convolutions in coordinates	285
Theorem 15.2.2	Affine equivariance of Euclidean GM -convolutions	288

Spherical GM -convolutions

Theorem 17.2.1	Spherical steerable kernels in geodesic coordinates	317
Theorem 17.2.2	Spherical steerable convolutions as GM -convolutions	318
Theorem 17.3.1	Gnomonic projections as warped exponential maps	319

Symmetry groups – basic definitions

Definition B.1.1	Group	367
Definition B.1.2	Abelian group	368
Definition B.1.3	Group homomorphism	369
Definition B.1.4	Group isomorphism	369

Subgroups and products of groups

Definition B.2.1	Subgroup	370
Definition B.2.2	Cosets	370
Definition B.2.3	Normal subgroup	370
Definition B.2.4	Direct product of groups	370
Definition B.2.5	Semidirect product of groups	371

Group actions, orbits and quotient spaces

Definition B.3.1	Left group action	372
Definition B.3.2	Right group action	372
Definition B.3.3	Group orbit	373
Definition B.3.4	Quotient set and quotient map	373
Definition B.3.5	Orbit representative	374
Definition B.3.6	Stabilizer subgroup	374
Theorem B.3.7	Orbit-stabilizer theorem	374
Definition B.3.8	Transitive action	374
Definition B.3.9	Faithful action	375
Definition B.3.10	Fixed-point free action	375
Definition B.3.11	Homogeneous space	375

Corollary B.3.12	Homogeneous space as group quotient	375
Definition B.3.13	Principal homogeneous space (torsor)	376

Invariant and equivariant maps

Definition B.4.1	Invariant map	376
Definition B.4.2	Equivariant map	377

Group representations and intertwiner maps

Definition B.5.1	Linear group representation	379
Definition B.5.2	Restricted representation	380
Definition B.5.3	Direct sum representation	380
Definition B.5.4	Tensor product representation	381
Definition B.5.5	Invariant subspace, subrepresentation	381
Definition B.5.6	Irreducible representation (irrep)	381
Definition B.5.7	Intertwiner	382
Definition B.5.8	Equivalent (isomorphic) representations	383
Definition B.5.9	Endomorphism	383
Lemma B.5.10	Schur's lemma	383
Definition B.5.11	Unitary transformation	383
Definition B.5.12	Unitary group	384
Definition B.5.13	Unitary representation	384
Theorem B.5.14	Compact groups & unitary representations	384
Definition B.5.15	Isomorphism of unitary representations	384
Theorem B.5.16	Complete reducibility	384
Definition B.5.17	Clebsch-Gordan decomposition and coefficients	384
Definition B.5.18	Regular representation	385
Definition B.5.20	Quotient representation	385
Theorem B.5.22	Peter-Weyl	386

Additional theorems & proofs

Theorem I.0.1	Differentiation lemma [93]	437
Theorem J.0.1	Regular feature fields as scalar functions on G -structure	440

Groups, representations and equivariant maps

This appendix gives a brief introduction to elementary group theory, the mathematics of symmetries. Since all definitions and theorems are well known and easy to find in the literature, we omit proofs. For a more in-depth discussion we point the reader to the literature, for instance [303, 212, 101, 33, 267].

After reviewing basic definitions in the following Appendix B.1, we discuss subgroups and products of groups in Appendix B.2, group actions, orbits, quotients and homogeneous spaces in Appendix B.3 and equivariant maps in Appendix B.4. Appendix B.5 gives a very brief introduction to the theory of group representations.

B.1 Symmetry groups – basic definitions

The symmetries of an object are the set of transformations that leave it invariant; see Fig. B.1. Simple examples are geometric transformations like translations, rotations or reflections. It is intuitively clear that symmetry operations can be composed with each other – for instance, two rotations that leave an object invariant can be composed to give 1) another rotation that 2) still leaves the object invariant. It is furthermore evident that the trivial (identity) transformation is a symmetry of each object and that every symmetry has an inverse. In the case of rotations, these would be the rotation by zero degrees and by the negative angle around the same axis. The set of all symmetries of an object and their composition forms a symmetry group, which is formalized as follows:

Definition B.1.1 (Group).

A group is a tuple (G, \cdot) , consisting of a set G and a binary operation

$$\cdot : G \times G \rightarrow G, \quad (g, h) \mapsto g \cdot h \tag{B.1}$$

satisfying the following three group axioms:

associativity: for all $g, h, k \in G$ one has $(g \cdot h) \cdot k = g \cdot (h \cdot k)$

identity element: $\exists e \in G$ such that $\forall g \in G$ one has $e \cdot g = g = g \cdot e$

inverse element: $\forall g \in G \exists g^{-1} \in G$ such that $g \cdot g^{-1} = e = g^{-1} \cdot g$

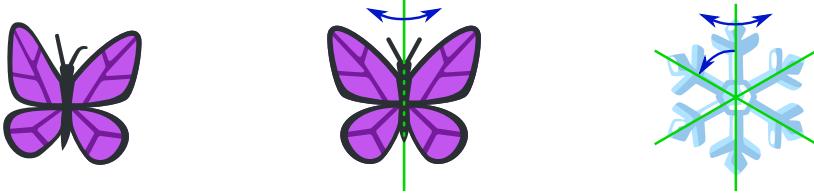
(a) Trivial symmetry group $\{e\}$ (b) Reflection symmetry group \mathcal{R} (c) Dihedral symmetry group D_6

Figure B.1: An object is said to be symmetric when it comes with a non-trivial set of symmetry transformations, leaving it invariant. Symmetries are mathematically modeled by *groups*, Def B.1.1. *Left*: The asymmetric butterfly in Fig. B.1a has no non-trivial symmetries. Its symmetry group is therefore trivial, i.e. contains only the identity element. *Middle*: The butterfly in Fig. B.1b has a bilateral symmetry and is therefore modeled by the reflection group \mathcal{R} with two elements (identity and reflection). *Right*: Fig. B.1c shows a snowflake with dihedral symmetry D_6 . It consists of rotations by multiples of $2\pi/6$ and reflections, making 12 group elements in total. The neural connectivity (synapse weights) of group equivariant neural networks is necessarily invariant (symmetric) under the action of the symmetry group. (Butterflies adapted under the Creative Commons Attribution 4.0 International license by courtesy of Twitter, snowflake adapted under the Apache license 2.0 by courtesy of Google.)

The identity element and the inverse of a group element can be shown to be unique. It is customary to abbreviate the group by its set G and to omit the binary operation, i.e. to write gh for $g \cdot h$. We will use these abbreviations in the following whenever the meaning is clear from the context.

Common examples of groups are the trivial group $\{e\}$, consisting of the identity element only, the d -dimensional continuous translation group $(\mathbb{R}^d, +)$, the special orthogonal rotation groups $\text{SO}(d) := \{A \in \mathbb{R}^{d \times d} \mid A^\top = A^{-1}, \det(A) = 1\}$, the general linear groups $\text{GL}(d) := \{A \in \mathbb{R}^{d \times d} \mid \det(A) \neq 0\}$, modeling transitions between arbitrary reference frames of \mathbb{R}^d (change of basis), or the Euclidean groups $E(d)$, which model the isometries of \mathbb{R}^d (combined translations, rotations and reflections).

Groups are often equipped with additional mathematical structure. For instance, *topological groups* are equipped with a topology w.r.t. which the group composition and inversion are required to be continuous, (*locally*) *compact topological groups* are in addition (locally) compact spaces or *Lie groups* are smooth manifolds, with smooth composition and inversion maps.

While the composition of group elements is associative, it is in general not commutative; see Fig. B.2.

Definition B.1.2 (Abelian group).

A group is called abelian iff all of its elements commute, i.e. iff:

$$gh = hg \quad \forall g, h \in G \tag{B.2}$$

Examples of abelian groups are translations or rotations in two dimensions. As one can easily check with any object in reach, rotations in three dimensions do in general not commute.¹

¹Rotate the object for instance by $\pi/2$ around the z -axis and then by $\pi/2$ around the x -axis. Rotating instead first around the x and then around the z -axis yields a different final rotation.

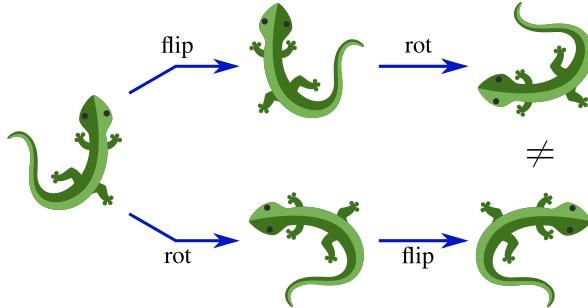


Figure B.2: The composition of group elements is in general *not commutative*, that is, $gh \neq hg$. This is exemplified by the action of reflections and rotations, whose order matters. Groups for which all elements commute are called *abelian*.
 (Lizards adapted under the Creative Commons Attribution 4.0 International license by courtesy of Twitter)

Group homomorphisms are structure preserving maps between groups in the sense that the composition of group elements in the domain is compatible with the composition of their images in the codomain.

Definition B.1.3 (Group homomorphism). A group homomorphism between two groups (G, \cdot) and (G', \star) is a map $\gamma : G \rightarrow G'$ such that

$$\gamma(g \cdot h) = \gamma(g) \star \gamma(h) \quad \forall g, h \in G, \quad (\text{B.3})$$

visualized by the following commutative diagram:

$$\begin{array}{ccc} G \times G & \xrightarrow{\cdot} & G \\ \downarrow \gamma \times \gamma & & \downarrow \gamma \\ G' \times G' & \xrightarrow{\star} & G' \end{array} \quad (\text{B.4})$$

This implies generally that $\gamma(g^{-1}) = \gamma(g)^{-1}$ and that $\gamma(e_G) = e_{G'}$. A simple example for a homomorphism is the map $\text{proj}_1 : (\mathbb{R}^2, +) \rightarrow (\mathbb{R}, +)$, $(x_1, x_2) \mapsto x_1$ since $\text{proj}_1((x_1, x_2) + (y_1, y_2)) = \text{proj}_1(x_1 + y_1, x_2 + y_2) = x_1 + y_1 = \text{proj}_1(x_1, x_2) + \text{proj}_1(y_1, y_2)$. As evident from this example, homomorphism may lose some of the structure (here the translation of the second axis).

If all group structure is preserved by a homomorphism, which is exactly the case when it is invertible, one speaks of an *isomorphism*:

Definition B.1.4 (Group isomorphism). A group homomorphism $\gamma : G \rightarrow G'$ is denoted as group isomorphism if it is invertible, that is, if there exists an inverse $\gamma^{-1} : G' \rightarrow G$ satisfying $\gamma^{-1} \circ \gamma = \text{id}_G$ and $\gamma \circ \gamma^{-1} = \text{id}_{G'}$. One writes $G \cong G'$ to state that G and G' are isomorphic.

Isomorphic groups are fully equivalent in their structure. An example of isomorphic groups are all subgroups of $\text{SO}(3)$ consisting of rotations around a fixed axis - all of these subgroups are planar rotations in the plane orthogonal to the respective axis, and therefore in particular isomorphic to $\text{SO}(2)$. One can show that they are furthermore all isomorphic to the unitary group $\text{U}(1) := \{e^{i\phi} \mid \phi \in [0, 2\pi)\}$, modeling rotations of the complex plane \mathbb{C} .

B.2 Subgroups and products of groups

One is often interested in a subset of symmetries that forms a group itself:

Definition B.2.1 (Subgroup). A subset $H \subseteq G$ of a group G forms a subgroup if it is closed under composition and taking inverses:

composition: for all $g, h \in H$ one has $gh \in H$

inversion: for all $g \in H$ one has $g^{-1} \in H$

As the name suggests, subgroups are themselves groups, that is, they satisfy the three group axioms. One writes $H \leq G$ to denote that H is not only a subset, but a subgroup of G .

Note that every group has itself as subgroup. Any subgroup $H \leq G$ different from the group G itself is denoted as *proper subgroup*, which may be symbolized by using a $<$ symbol instead of \leq . Any group has the trivial group $\{e\}$ as subgroup. Further examples are discrete translations $(\mathbb{Z}^d, +) < (\mathbb{R}^d, +)$ or discrete two-dimensional rotations $C_N < SO(2)$ by multiples of $2\pi/N$.

Definition B.2.2 (Cosets). Let G be a group and $H \leq G$ a subgroup. The subsets of G defined by left (right) translations of H by $g \in G$ are known as left (right) cosets of H in G :

left coset: $gH := \{gh \mid h \in H\}$

right coset: $Hg := \{hg \mid h \in H\}$

The (quotient) space of left cosets is denoted by G/H , while the (quotient) space of right cosets is written $H\backslash G$.

The spaces G/H and $H\backslash G$ do in general not carry a group structure. However, they can be shown to do, in the special case where H is a *normal subgroup* of G .

Definition B.2.3 (Normal subgroup). A subgroup $N \leq G$ of a group G is called *normal* iff its left and right cosets coincide, that is, iff

$$gN = Ng \quad \forall g \in G. \tag{B.5}$$

One usually writes $N \trianglelefteq G$ if N is a normal subgroup of G .

Two groups may be combined in different ways to form a new group. The simplest way of doing so is the (outer) *direct product* of groups, which combines its factors such that they transform independently:

Definition B.2.4 (Direct product of groups). Let (H, \cdot) and (K, \star) be arbitrary groups. Their (outer) direct product $(H, \cdot) \times (K, \star)$ is defined on the Cartesian product $H \times K$ of the underlying sets, equipped with the binary operation

$$H \times K \rightarrow H \times K, \quad ((\tilde{h}, \tilde{k}), (h, k)) \mapsto (\tilde{h} \cdot h, \tilde{k} \star k) \tag{B.6}$$

which composes the elements of the factors H and K independently from each other. One commonly abbreviates the direct product by $H \times K$.

The direct product is generalized to an arbitrary number of factors by taking the Cartesian product over all sets and defining the binary operation element wise. Given a direct product $H \times K$, one can recover subgroups $H' := \{(h, e_K) \mid h \in H\}$ and $K' := \{(e_H, k) \mid k \in K\}$, which are isomorphic to H and K , respectively. Note that these subgroups satisfy the following algebraic properties:

- the intersection $H' \cap K' = (e_H, e_K)$ is trivial
- there is a unique decomposition of elements in $H \times K$ into one element of H' and one of K'
- elements in H' and K' commute – they are both *normal subgroups* of $H \times K$

If a group contains subgroups that satisfy these properties, it is called *inner direct product* and is guaranteed to be isomorphic to the *outer direct product* of these subgroups.

An example of a direct product is the symmetry group $(\mathbb{R}, +) \times \text{SO}(2)$ of an infinitely extended cylinder, whose factors describe the independent translation along the cylinder and rotations around its axis.

A generalization of the direct product is the *semidirect product* of groups, in which the subgroups are not composed independently anymore but the second factor acts on the first one:

Definition B.2.5 (Semidirect product of groups). Assume arbitrary groups (N, \cdot) and (H, \star) to be given and let $\gamma : H \rightarrow \text{Aut}(N)$ be a group homomorphism from H into the automorphism group (symmetries) of N . The corresponding (outer) semidirect product $(N, \cdot) \rtimes_{\gamma} (H, \star)$ is defined by:

- the underlying set is the Cartesian product $N \times H$
 - a binary operation given by $(N \times H) \times (N \times H) \rightarrow N \times H$,
- $$((\tilde{n}, \tilde{h}), (n, h)) \mapsto (\tilde{n} \cdot \gamma(\tilde{h})n, \tilde{h} \star h)$$

This definition includes the *direct product* for the trivial homomorphism $\gamma(h) = \text{id}_N \in \text{Aut}(N)$ for any $h \in H$, which leaves N invariant. A non-trivial example is the *Euclidean group* $E(d) = (\mathbb{R}^d, +) \rtimes_{\gamma} O(d)$, for which $\gamma(h) = h \in O(d) \leq \text{Aut}(\mathbb{R}^d)$, such that $(t_2, h_2) \cdot (t_1, h_1) = (t_2 + h_2 t_1, h_2 h_1)$ for any translations $t_1, t_2 \in (\mathbb{R}^d, +)$ and orthogonal group elements $h_1, h_2 \in O(d)$. It is worth building an intuition by drawing a few examples of the composition of translations and rotations on a piece of paper – you will indeed find that rotations of the plane act on previous translations, while translations do not affect the overall rotation.

As for the direct product, the semidirect product $N \rtimes_{\gamma} H$ contains subgroups $N' := \{(n, e_H) \mid n \in N\}$ and $H' := \{(e_N, h) \mid h \in H\}$ that are isomorphic to N and H . In general, only N is a normal subgroup of the semidirect product. One can show that there exists again a unique decomposition $g = nh$ of an element $g \in N \rtimes_{\gamma} H$ into elements $n \in N'$ and $h \in H'$.

Without going into more detail, we want to mention that further products of groups, like the *Wreath product* or the *Zappa–Sz  p* product, exist, and are applied in equivariant deep learning [316].

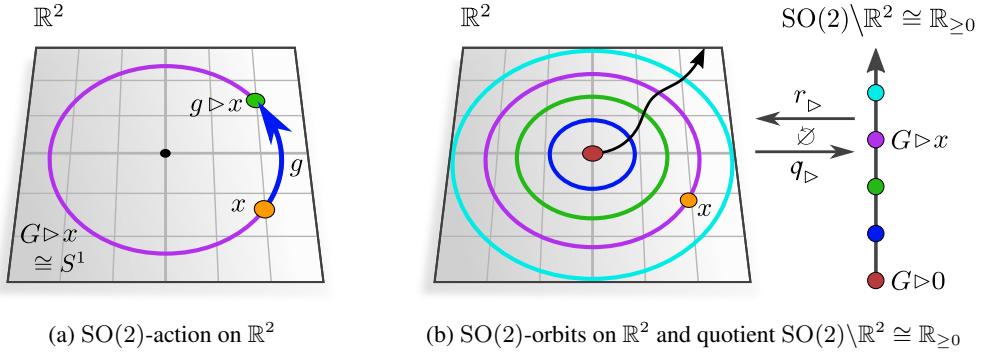


Figure B.3: Action, orbits and quotient of the 2-dimensional rotation group $G = \text{SO}(2)$ on the plane \mathbb{R}^2 . *Left:* A group element $g \in \text{SO}(2)$ (blue) acts on a point $x \in \mathbb{R}^2$ (orange) by rotating it to another location $g \triangleright x \in \mathbb{R}^2$ (green). The set of points $\text{SO}(2) \triangleright x := \{g \triangleright x \mid g \in \text{SO}(2)\}$ (purple) reached by acting with $\text{SO}(2)$ on x is denoted as orbit of x . *Right:* The group action partitions \mathbb{R}^2 in disjoint orbits. While the orbits for points different from the origin are circles, the origin's orbit is a single point. The space of orbits is the quotient space $\text{SO}(2) \setminus \mathbb{R}^2$ of the action. It is in this example isomorphic to the positive real line (radii) $\mathbb{R}_{\geq 0}$. The quotient map $q_{\triangleright} : \mathbb{R}^2 \rightarrow \text{SO}(2) \setminus \mathbb{R}^2$ collapses points in the space acted on to their orbits, i.e. to the quotient space. The wiggly arrow is a (non-canonical) choice of orbit representatives.

B.3 Group actions, orbits and quotient spaces

In our application to equivariant networks we are mainly interested in the *action* of groups on some object like a space or feature field.

Definition B.3.1 (Left group action).

Let G be a group and X be a set. A left group action is a map

$$\triangleright : G \times X \rightarrow X, \quad (g, x) \mapsto g \triangleright x \quad (\text{B.7})$$

that is compatible with the group composition and identity element:

associativity: $(gh) \triangleright x = g \triangleright (h \triangleright x)$ for any $g, h \in G, x \in X$

identity: $e \triangleright x = x$ for any $x \in X$

G is then said to act on X from the left and X said to be a left G -set.

Fig B.3a visualizes the natural action of the matrix group $\text{SO}(2) \subset \mathbb{R}^{2 \times 2}$ on \mathbb{R}^2 by matrix multiplication, thereby rotating points in the plane.

Besides left actions, there are right actions, which are similarly defined:

Definition B.3.2 (Right group action).

Let G be a group and X be a set. A map

$$\triangleleft : X \times G \rightarrow X, \quad (x, g) \mapsto x \triangleleft g \quad (\text{B.8})$$

is denoted as right group action iff it satisfies:

associativity: $x \triangleleft (gh) = (x \triangleleft g) \triangleleft h$ for any $g, h \in G, x \in X$

identity: $x \triangleleft e = x$ for any $x \in X$

The difference between left and right group actions lies in their associativity law, specifically in the order in which the composition gh of two group elements $g, h \in G$ acts. Left and right actions can be converted into each other by means of inversion of the acting group elements since this inversion $(gh)^{-1} = h^{-1}g^{-1}$ swaps their order. To make this explicit, let \triangleright be a left group action, then

$$\triangleleft_{\text{inv}} : X \times G \rightarrow G, \quad (x, g) \mapsto x \triangleleft_{\text{inv}} g := g^{-1} \triangleright x \quad (\text{B.9})$$

is a right action since, for arbitrary $g, h \in G$ and $x \in X$,

$$\begin{aligned} x \triangleleft_{\text{inv}} (gh) &= (gh)^{-1} \triangleright x = (h^{-1}g^{-1}) \triangleright x \\ &= h^{-1} \triangleright (g^{-1} \triangleright x) = (x \triangleleft_{\text{inv}} g) \triangleleft_{\text{inv}} h, \end{aligned} \quad (\text{B.10})$$

as required for right actions. An equivalent argument holds to convert right actions to left actions. We will make heavy use of both left and right group actions on fiber bundles, with left actions modeling their active transformations (diffeomorphism pushforwards) and right actions modeling passive gauge transformations. Since all of the following definitions can be made similarly for both types of actions, we will only give them for left group actions.

If G acts on X , one may ask where a point $x \in X$ may be moved by the G -action. The set of such points is known as *orbit* of x ; see Fig. B.3.

Definition B.3.3 (Group orbit). Let \triangleright be an action of G on X and consider any element $x \in X$. The subset

$$G \triangleright x := \{g \triangleright x \mid g \in G\} \quad (\text{B.11})$$

of X is then denoted as *orbit* of x .

Note that one can define an equivalence relation, defined in footnote 11, by identifying any two elements of X iff they are lying in the same orbit:

- reflexivity:* $x \sim_{\triangleright} x$, that is, x is contained in its own orbit $G \triangleright x$
- symmetry:* $x \sim_{\triangleright} y \Leftrightarrow y \sim_{\triangleright} x$, that is, if x is contained in y 's orbit, then y is contained in x 's orbit
- transitivity:* $x \sim_{\triangleright} y \wedge y \sim_{\triangleright} z \Rightarrow x \sim_{\triangleright} z$, that is, if x is contained in y 's orbit and if y is contained in z 's orbit, then x is contained in z 's orbit

The G -action thus partitions X into disjoint equivalence classes (the orbits). The set of all orbits is the corresponding quotient set:

Definition B.3.4 (Quotient set and quotient map). The quotient set induced by a G -action \triangleright on X is the set of all orbits:

$$G \setminus X := \{G \triangleright x \mid x \in X\} \quad (\text{B.12})$$

The corresponding quotient map collapses elements of X to their orbit:

$$q_{\triangleright} : X \rightarrow G \setminus X, \quad x \mapsto G \triangleright x \quad (\text{B.13})$$

We distinguish quotient spaces arising from right group actions in our notation by writing them as X/G instead of $G \setminus X$.

Since the quotient map is in general non-injective, it does not have an inverse. However, by the axiom of choice, one may make a non-canonical choice of *orbit representative* $r_{\triangleright}(x) \in G \triangleright x \subset X$ for each orbit $G \triangleright x$:

Definition B.3.5 (Orbit representative). Orbit representatives are specified by a map

$$r_{\triangleright} : G \setminus X \rightarrow X \quad \text{such that} \quad q_{\triangleright} \circ r_{\triangleright}(G \triangleright x) = G \triangleright x \quad \forall G \triangleright x \in G \setminus X, \quad (\text{B.14})$$

i.e. such that the following diagram commutes:

$$\begin{array}{ccccc} G \setminus X & \xrightarrow{r_{\triangleright}} & X & \xrightarrow{q_{\triangleright}} & G \setminus X \\ & & \underbrace{\hspace{10em}}_{\text{id}_{G \setminus X}} & & \uparrow \end{array} \quad (\text{B.15})$$

Orbit representatives can in general not be chosen continuously.

The additional requirement on r_{\triangleright} in Eq. (B.15) ensures that the orbit representative is indeed an element of its orbit as subset of X . Fig. B.3 visualizes the orbits, quotient space and some choice of representatives for the $\text{SO}(2)$ -action on \mathbb{R}^2 .

The orbit of a G -set element x is related to its *stabilizer subgroup*, which is that subgroup of G that leaves x invariant:

Definition B.3.6 (Stabilizer subgroup). Let X be a G -set, acted on by \triangleright . The stabilizer subgroup of G corresponding to some element $x \in X$ is defined as:

$$\text{Stab}_x := \{g \in G \mid g \triangleright x = x\} \leq G \quad (\text{B.16})$$

With this definition, we formulate the orbit stabilizer theorem:

Theorem B.3.7 (Orbit-stabilizer theorem). Let \triangleright be a continuous action of a topological group G on a G -space X and let $x \in X$. The orbit of x is then isomorphic (homeomorphic) to the quotient (coset space) of G w.r.t. the stabilizer subgroup:

$$G \triangleright x \cong_{\text{top}} G / \text{Stab}_x \quad (\text{B.17})$$

Note that the stabilizer subgroups of different elements of the same orbit are always isomorphic (even conjugate), that is, $\text{Stab}_x \cong \text{Stab}_y$ for any $x, y \in G \triangleright z$ and $z \in X$.

B.3.1 Properties of group actions

Group actions may have further properties of interest, like being *transitive*, *faithful* or *fixed-point free*. A transitive group action is distinguished by being able to move any element of the set it acts on to any other element – there is only one single orbit:

Definition B.3.8 (Transitive action). A G -action \triangleright on X is called transitive iff it satisfies:

$$\forall x, y \in X \implies \exists g \in G \text{ such that } y = g \triangleright x. \quad (\text{B.18})$$

Spaces with a transitive group action are known as *homogeneous spaces*; see Appendices B.3.2 and F for more details. Examples of transitive actions are $\text{SO}(3)$ -rotations of the sphere S^2 . The plane \mathbb{R}^2 is being acted on transitively by the standard actions of the two-dimensional translation group $(\mathbb{R}^2, +)$ or the Euclidean group $E(2)$, but not by the rotation group $\text{SO}(2)$ or by one-dimensional translations $(\mathbb{R}, +)$ along some axis.

Faithful group actions are actions for which any non-trivial group element moves at least one element of the space it acts on:

Definition B.3.9 (Faithful action).

Let \triangleright be a G -action on X . This action is called faithful if it satisfies

$$\forall g \in G \setminus \{e\} \implies \exists x \in X \text{ such that } g \triangleright x \neq x. \quad (\text{B.19})$$

$G \setminus \{e\} := \{g \in G \mid g \neq e\}$ refers hereby to all group elements except for the identity, not to a quotient.

To construct a counterexample, consider the two-dimensional translation group $(\mathbb{R}^2, +)$, acting on the plane \mathbb{R}^2 as defined by $(u, v) \triangleright (x, y) := (x + u, y)$. This is a well defined group action since $(0, 0) \triangleright (x, y) = (x, y)$ and $(u, v) \triangleright (r, s) \triangleright (x, y) = (x + u + r, y) = (u + r, v + s) \triangleright (x, y)$ for any $x, y, u, v, r, s \in \mathbb{R}$. However, it is not faithful since $(0, v) \triangleright (x, y) = (x, y)$ for any $v, x, y \in \mathbb{R}$. One can show that a non-faithful G -action corresponds always to a faithful action of some subgroup of G .

As the name suggests, a *fixed-point free* action leaves no single point of the space on which it acts invariant:

Definition B.3.10 (Fixed-point free action). A G -action \triangleright is fixed-point free iff all stabilizer subgroups are trivial, that is, iff $\text{Stab}_x = \{e\} \forall x \in X$.

Fixed-point free actions are always faithful if the set acted on is non-empty. The standard action of the translation group $(\mathbb{R}^d, +)$ on \mathbb{R}^d is fixed-point free, while the standard action of the Euclidean group $E(d)$ on \mathbb{R}^d is not – it has stabilizer subgroups $\text{Stab}_x \cong O(d)$, consisting of rotations and reflections around any point $x \in \mathbb{R}^d$. Another counterexample is the standard action of $O(d)$ on \mathbb{R}^d : while the stabilizers for all points $x \neq 0$ are trivial, the origin is fixed by all group elements, i.e. $\text{Stab}_0 = O(d)$.

B.3.2 Homogeneous spaces

Homogeneous spaces are of particular importance for the theory of equivariant networks since they admit a natural definition of convolution integrals. Since homogeneous spaces and their convolutions are reviewed in detail in Appendix F, we will here just give a brief overview.

Definition B.3.11 (Homogeneous space). A G -space is said to be homogeneous iff it is equipped with a transitive group action, Def. B.3.8.

The transitivity of the group action implies that the homogeneous space consists of a single orbit. As a corollary of the orbit-stabilizer theorem B.3.7, any homogeneous space arises as a group quotient, i.e. as a space of cosets:

Corollary B.3.12 (Homogeneous space as group quotient). Let X be a homogeneous G -space of a topological group G and $x \in X$ arbitrary, then

$$X \cong_{\text{top}} G / \text{Stab}_x. \quad (\text{B.20})$$

Since the stabilizer subgroups at all points of the homogeneous space are isomorphic to the same group $H \leq G$, the quotient map

$$G \rightarrow G/H, \quad g \mapsto gH \quad (\text{B.21})$$

makes G a *principal H -bundle* over the homogeneous space G/H ; see Section 11.1.5 for a definition of principal bundles.

An example of a homogeneous space is $\mathbb{R}^d \cong E(d)/O(d)$ since the action of the Euclidean group $E(d)$ on \mathbb{R}^d is transitive with stabilizer subgroups isomorphic to $O(d) < E(d)$. It arises further as a homogeneous space $\mathbb{R}^d \cong (\mathbb{R}^d,+)/\{e\}$ under the transitive action of the translation group $(\mathbb{R}^d, +)$ with trivial stabilizers $\{e\}$. Another example is the sphere $S^2 \cong SO(3)/SO(2) \cong O(3)/O(2)$ with transitive symmetries $SO(3)$ and $O(3)$, whose stabilizer subgroups are isomorphic to $SO(2)$ and $O(2)$, respectively.

One further distinguishes *principal homogeneous spaces*. They have trivial stabilizers and are thus isomorphic to the group as topological spaces, but don't carry a group structure.

Definition B.3.13 (Principal homogeneous space (torsor)). A principal homogeneous G -space has a transitive and fixed-point free group action.

Missing the group structure of G , a principal homogeneous G -space does not have a distinguished identity element. Principal G -bundles, defined in Section 11.1, have principal homogeneous G -spaces as fibers. In the case of G -structures, which are principal G -bundles of reference frames, the lack of identity element reflects the lack of a canonical reference frame.

A simple example of a principal homogeneous space is the circle $S^1 \cong_{\text{top}} SO(2)$, which is acted on transitively and fixed-point free by $SO(2)$.

B.4 Invariant and equivariant maps

We turn now to functions mapping between G -sets. Such functions are said to be *G -invariant* if their result does not change when acting on their input. Since we are in our application interested in invariant (or equivariant) neural network *layers*, we denote the maps in the following by L .

Definition B.4.1 (Invariant map). Let X be a G -set, acted on by some group action \triangleright_X . A function $L : X \rightarrow Y$ is called G -invariant, iff it satisfies

$$L(g \triangleright_X x) = L(x) \quad \forall g \in G, x \in X, \quad (\text{B.22})$$

visualized by the commutativity of the following diagram:

$$\begin{array}{ccc} X & & Y \\ g \triangleright_X \downarrow & \searrow L & \\ X & \nearrow L & \end{array} \quad (\text{B.23})$$

As an example, consider the map $|\cdot| : \mathbb{R}^d \rightarrow \mathbb{R}_{\geq 0}$, $x \mapsto |x|$, which maps any vector to its Euclidean norm. This map is invariant under rotations and reflections of the vector, i.e. under the standard action of the orthogonal group $O(d)$ on \mathbb{R}^d . Another example are neural networks for image classification, which should usually be invariant under translations of the input image.

Since G -invariant maps are constant on the group orbits, they imply a unique unconstrained map $\tilde{L} : G \setminus X \rightarrow Y$ on the quotient set such that $L = \tilde{L} \circ q_{\triangleright}$:

$$\begin{array}{ccc} X & \xrightarrow{L} & Y \\ q_{\triangleright} \downarrow & \nearrow \tilde{L} & \\ G \setminus X & & \end{array} \quad (\text{B.24})$$

While the result of an invariant map does not change when acting on their domain, the result of an *equivariant map* changes according to some group action on the function's codomain.

Definition B.4.2 (Equivariant map). Let X and Y be G -sets, acted on by group action \triangleright_X and \triangleright_Y , respectively. If a function $L : X \rightarrow Y$ commutes with these group actions,

$$L(g \triangleright_X x) = g \triangleright_Y L(x) \quad \forall g \in G, x \in X, \quad (\text{B.25})$$

it is said to be G -equivariant. This condition corresponds to the commutative diagram below:

$$\begin{array}{ccc} X & \xrightarrow{L} & Y \\ g \triangleright_X \downarrow & & \downarrow g \triangleright_Y \\ X & \xrightarrow{L} & Y \end{array} \quad (\text{B.26})$$

Note that invariant functions are a special case of equivariant functions for which the output group action $g \triangleright = \text{id}_Y$ is trivial:²

$$\begin{array}{ccc} X & \xrightarrow{L} & Y \\ g \triangleright_X \downarrow & \nearrow L & \curvearrowright \text{id}_Y \\ X & \xrightarrow{L} & Y \end{array} \iff \begin{array}{ccc} X & \xrightarrow{L} & Y \\ g \triangleright_X \downarrow & & \downarrow \text{id}_Y \\ X & \xrightarrow{L} & Y \end{array} \quad (\text{B.27})$$

On the other hand, any equivariant map w.r.t. group actions \triangleright_X and \triangleright_Y on its domain and codomain can itself be viewed as an invariant under the joint action on the function space, since for any $g \in G$:³

$$\begin{aligned} L \circ (g \triangleright_X (\cdot)) &= (g \triangleright_Y (\cdot)) \circ L \\ \iff (g^{-1} \triangleright_Y (\cdot)) \circ L \circ (g \triangleright_X (\cdot)) &= L \end{aligned} \quad (\text{B.28})$$

This insight has immediate consequences for equivariant neural networks: their neural connectivity (synapses) is necessarily invariant under the simultaneous group action on the neurons' input and output space, a property which is usually referred to as *weight sharing*. Our central Theorem 13.2.4 confirms indeed that equivariant kernel field transforms (network layers) on manifolds are specified by invariant kernel fields (neural connectivity).

²The trivial (invariant) action $\triangleright : G \times Y \rightarrow Y$, $(g, y) \mapsto g \triangleright y := y$ is indeed a well defined group action since $(gh) \triangleright y = y = g \triangleright y = g \triangleright (h \triangleright y)$ and $e \triangleright y = y$, as demanded in Def. B.3.1.

³One checks that this is indeed a well defined group action on the function space.

As an example for an equivariant map, consider the one-dimensional Euclidean convolution map

$$K * : L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R}), \quad f \mapsto K * f := \int_{\mathbb{R}} dy K(x - y) f(y), \quad (\text{B.29})$$

acting on square integrable functions on the real line (the kernel K is also an element of $L^2(\mathbb{R})$). Denote by \triangleright the regular translation group action on real valued functions, which is defined by $(g \triangleright f)(x) = f(x - g)$. The convolution is then easily shown to be equivariant w.r.t. this action on its domain and codomain since for any $g \in (\mathbb{R}, +)$, any $x \in \mathbb{R}$ and any $K, f \in L^2(\mathbb{R})$ one has:

$$\begin{aligned} (K * (g \triangleright f))(x) &= \int_{\mathbb{R}} dy K(x - y) (g \triangleright f)(y) \\ &= \int_{\mathbb{R}} dy K(x - y) f(y - g) \\ &= \int_{\mathbb{R}} dz K((x - g) - z) f(z) \\ &= (K * f)(x - g) \\ &= (g \triangleright (K * f))(x) \end{aligned} \quad (\text{B.30})$$

That convolutions are not only sufficient, but necessary for a linear and regular translation group action equivariant map is proven in Theorem 3.2.1.

B.5 Group representations and intertwiner maps

We now turn to the study of *group representations*, which are essentially linear group actions on vector spaces. A group representation assigns matrices (or linear operators) to group elements and models the group operation by matrix multiplications (composition of linear maps). Representation theory is of great practical relevance in physics and deep learning where the objects being acted on are typically elements of some vector space.

The definitions and results of representation theory depend heavily on the class of groups, representations and vector spaces under consideration. In the following Appendix B.5.1 we focus on *finite dimensional representations* of *locally compact groups*. Appendix B.5.2 studies *unitary representations* of *compact groups*, for which some additional properties can be proven. While one considers usually complex vector spaces, we allow for \mathbb{K} -vector spaces, where \mathbb{K} is either the field of *real* or *complex* numbers, \mathbb{R} or \mathbb{C} , respectively.

As we keep our introduction short and intuitive, it can hardly live up to the rich literature on representation theory. For a more rigorous introduction we point the reader to [102, 303]. The *representation theory of G-steerable kernels* for *compact groups* G is in depth developed in [173].⁴ We will not discuss *induced representations*, which are described in [35, 55].

⁴Compact groups include in particular (all subgroups of) the *unitary groups* $U(d)$ and the *orthogonal groups* $O(d)$, which model rotations and reflections of \mathbb{C}^d and \mathbb{R}^d , respectively. Further examples of compact groups are listed on [Wikipedia](#).

B.5.1 Finite dimensional representations of locally compact groups

This section introduces finite dimensional representations of locally compact groups. The finite dimensionality refers hereby to the vector space being acted on. Locally compact groups include in particular *finite groups*, *compact groups* and *Lie groups*.

Definition B.5.1 (Linear group representation). A linear group representation of a group G on a vector space V is a tuple (ρ, V) where

$$\rho : G \rightarrow \mathrm{GL}(V) \quad (\text{B.31})$$

is a group homomorphism from G to the general linear group $\mathrm{GL}(V)$ (invertible linear maps) of the vector space. V is called representation space.

One sometimes refers to only ρ or V as representation if the other constituent is clear from the context.

By the definition B.1.3 of homomorphisms, group representations satisfy

$$\text{composition: } \rho(gh) = \rho(g)\rho(h) \quad \forall g, h \in G \quad (\text{B.32})$$

$$\text{inverse: } \rho(g^{-1}) = \rho(g)^{-1} \quad \forall g \in G \quad (\text{B.33})$$

$$\text{identity: } \rho(e) = \mathrm{id}_V, \quad (\text{B.34})$$

which makes the composition of the linear maps in the image $\mathrm{Im}(\rho) \subseteq \mathrm{GL}(V)$ of ρ consistent with the composition of G -elements in its domain.

If $V = \mathbb{K}^c$, the representation assigns invertible matrices $\rho(g) \in \mathrm{GL}(\mathbb{K}^c)$ to group elements and acts via matrix multiplication. One can always choose such an explicit matrix representation of finite-dimensional representations by choosing some basis of V .

Group representations are in one-to-one relation with *linear* group actions as defined in Def. B.3.1 since any G -representation ρ defines an action

$$\triangleright_\rho : G \times V \rightarrow V, \quad (g, v) \mapsto g \triangleright_\rho v := \rho(g)v \quad (\text{B.35})$$

and any linear G -action \triangleright defines a representation

$$\rho_\triangleright : G \rightarrow \mathrm{GL}(V), \quad g \mapsto \rho_\triangleright(g) := g \triangleright (\cdot), \quad (\text{B.36})$$

satisfying their respective defining properties, as is easily checked.

If G is a *topological* (e.g. locally compact) group and V is a topological vector space, linear group representations (ρ, V) are required to be *continuous* in the sense that the associated group action \triangleright_ρ is a continuous map.

We list some of the representations commonly encountered in equivariant deep learning:

- the *trivial representation* of any group is given by $\rho(g) = \mathrm{id}_V$, where $V \cong \mathbb{K}^1$
- the *standard representation* (or defining representation) of any matrix group $G \leq \mathrm{GL}(\mathbb{K}^c)$ is given by the group element itself, i.e. by $\rho(g) = g$. It acts on \mathbb{K}^c by matrix multiplication.
- the *tensor representation* of rank (r, s) of any matrix group $G \leq \mathrm{GL}(\mathbb{K}^c)$ is given by tensor products of the group elements, i.e. by $\rho(g) = \otimes^s(g^{-1})^\top \otimes^r g$. It acts on $((\mathbb{K}^c)^*)^{\otimes s} \otimes (\mathbb{K}^c)^{\otimes r}$, where $((\mathbb{K}^c)^*)^*$ is the dual of \mathbb{K}^c .

ϕ	0	$\frac{\pi}{2}$	π	$\frac{3\pi}{2}$
$\rho_{\text{reg}}^{C_4}(\phi)$	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$

Table B.1: Visualization of the regular representation matrices of the cyclic group C_4 , consisting of rotations that are multiples of $\pi/2$. It is a permutation representation that shifts the four axes of \mathbb{K}^4 in a cyclic manner.

- the *regular representation* of a finite group G acts on $\mathbb{K}^{|G|}$ by permuting its $|G|$ elements according to the group composition law.⁵ Specifically, consider the standard basis $\{e_g \mid g \in G\}$ of $\mathbb{K}^{|G|}$, labeled by group elements. The regular representation acts then as $\rho(\tilde{g})e_g := e_{\tilde{g}g}$. Table B.1 visualizes the regular representation for the cyclic group C_4 . A more abstract definition, extending regular representations to infinite-dimensional vector spaces, is given in Def. B.5.18 below.

From the example of the trivial representation it is clear that group representations are, just as group actions, not necessarily *faithful* (Def. B.3.9).

A representation of a group implies representations of any of its subgroups via *restriction*:

Definition B.5.2 (Restricted representation). Let (ρ, V) be a G -representation and let $H \leq G$ be a subgroup. The restricted representation is the H -representation on V defined by restricting the domain of ρ from G to H :

$$\text{Res}_H^G \rho : H \rightarrow \text{GL}(V), \quad h \mapsto \rho(h) \quad (\text{B.37})$$

The direct sum of vector spaces extends naturally to representations:

Definition B.5.3 (Direct sum representation). Assume two G -representations (ρ_1, V_1) and (ρ_2, V_2) to be given. Their direct sum $(\rho_1 \oplus \rho_2, V_1 \oplus V_2)$ is a $\dim(V_1) + \dim(V_2)$ -dimensional representation, acting on the direct sum of vector spaces as defined by:

$$(\rho_1 \oplus \rho_2)(g)(v_1 \oplus v_2) := \rho_1(g)v_1 \oplus \rho_2(g)v_2 \quad (\text{B.38})$$

When a basis of V_1 and V_2 is chosen, the resulting matrix representation is the matrix direct sum of the matrix representations ρ_1 and ρ_2 :

$$(\rho_1 \oplus \rho_2)(g) = \begin{pmatrix} \rho_1(g) & 0 \\ 0 & \rho_2(g) \end{pmatrix} \quad (\text{B.39})$$

The two subspaces V_1 and V_2 of $V_1 \oplus V_2$ are transforming independently under this representation.

⁵ More abstractly, the regular representation is identified with the *group algebra* $\mathbb{K}[G]$ of G over \mathbb{K} . As a \mathbb{K} -vector space, it is the *free vector space* generated by G , consisting of formal linear combinations $\lambda g + \mu h$ of group elements where $g, h \in G$ and $\lambda, \mu \in \mathbb{K}$. The algebra structure of $\mathbb{K}[G]$ is inherited from the group multiplication, that is, it is defined by $k(\lambda g + \mu h) := \lambda(kg) + \mu(kh)$ for any $g, h, k \in G$ and $\lambda, \mu \in \mathbb{K}$.

It is furthermore possible to combine two representations by constructing their tensor product:

Definition B.5.4 (Tensor product representation).

Let (ρ_1, V_1) and (ρ_2, V_2) be two G -representations. Their tensor product representation $(\rho_1 \otimes \rho_2, V_1 \otimes V_2)$ is $\dim(V_1) \cdot \dim(V_2)$ -dimensional and acts on the tensor product of vector spaces as follows:

$$(\rho_1 \otimes \rho_2)(g)(v_1 \otimes v_2) := \rho_1(g)v_1 \otimes \rho_2(g)v_2 \quad (\text{B.40})$$

Not every element of $V_1 \otimes V_2$ can be represented as simple tensor $v_1 \otimes v_2$, but this definition extends to the general case by linearity.

In the case of matrix representations, the tensor product is simply given by the Kronecker product,

$$(\rho_1 \otimes \rho_2)(g) = \begin{pmatrix} \rho_1(g)_{11} \cdot \rho_2(g) & \cdots & \rho_1(g)_{1N} \cdot \rho_2(g) \\ \vdots & \ddots & \vdots \\ \rho_1(g)_{N1} \cdot \rho_2(g) & \cdots & \rho_1(g)_N \cdot \rho_2(g) \end{pmatrix}, \quad (\text{B.41})$$

where we abbreviated $\dim(V_1) =: \mathcal{V}$. Direct sum and tensor product representations can be extended to a finite number of factors, which are written $\bigoplus_{i=1}^N \rho_i$ and $\bigotimes_{i=1}^N \rho_i$, respectively.

Having seen how we can build new, larger representations from existing ones, we ask now whether we can split representations into separate constituent parts.

Definition B.5.5 (Invariant subspace, subrepresentation).

Let (ρ, V) be a G -representation and consider a vector subspace $W \subseteq V$. This subspace is called invariant if it is closed under the action of ρ , that is, if $\rho(g)w \in W$ for any $w \in W$, $g \in G$. This implies a homomorphism $\rho|_W : G \rightarrow \text{GL}(W)$, denotes as subrepresentation of ρ .

An obvious example are direct sum representations, which have their summands by construction as subrepresentations.

Representations of special interest are those which are not further reducible in (nontrivial) subspaces:

Definition B.5.6 (Irreducible representation (irrep)).

A representation (ρ, V) is called irreducible representation (irrep) if it has only the two trivial subrepresentations $W = V$ and $W = 0$.

Whether a representation is reducible or not may depend on the field \mathbb{K} under consideration. For instance, the *real valued* irreps of $\text{SO}(2)$ are the trivial representation and the frequency- k rotation matrices

$$\rho_k^{\text{SO}(2), \mathbb{R}}(\phi) := \begin{pmatrix} \cos(k\phi) & -\sin(k\phi) \\ \sin(k\phi) & \cos(k\phi) \end{pmatrix}, \quad k \in \mathbb{N}, \quad (\text{B.42})$$

while the *complex* irreps of $\text{SO}(2)$ are frequency- k complex exponentials $\rho_k^{\text{SO}(2), \mathbb{C}}(\phi) := e^{ik\phi}$, $k \in \mathbb{Z}$ (including the trivial irrep for $k = 0$). Over the complex field, the representation matrices in Eq.(B.42) are reducible. They decompose into a direct sum of two invariant subspaces, corresponding to complex valued irreps of opposite frequency, i.e. $\rho_k^{\text{SO}(2), \mathbb{R}} \cong_{\mathbb{C}} \rho_{-k}^{\text{SO}(2), \mathbb{C}} \oplus \rho_k^{\text{SO}(2), \mathbb{C}}$, $k \in \mathbb{N}$.

The restriction $\text{Res}_H^G(\rho)$ of a G -irrep ρ to a subgroup $H \leq G$ is in general not irreducible anymore. An obvious example is the restriction to the trivial group $H = \{e\}$, for which $\text{Res}_{\{e\}}^G \rho = \bigoplus_{i=1}^{\dim(V)} (1)$ decomposes into a (reducible) direct sum of $\dim(V)$ trivial irreps of $\{e\}$. As another example, consider the defining representation $\rho(g) = g$ of $\text{SO}(3)$, which is irreducible. If it is restricted to the subgroup $\text{SO}(2)_z$ of rotations around the z -axis, we end up with a direct sum of the defining representation of $\text{SO}(2)$, modeling the rotation of the xy -plane around the z -axis, and the trivial representation, ensuring that the z -axis is itself fixed:

$$[\text{Res}_{\text{SO}(2)_z}^{\text{SO}(3)} g](\phi) = \begin{pmatrix} \cos(\phi) & -\sin(\phi) & 0 \\ \sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix} \oplus (1) \quad (\text{B.43})$$

The central building blocks of any equivariant neural network are linear equivariant maps, which are known as *intertwiners*:

Definition B.5.7 (Intertwiner). Let (ρ_1, V_1) and (ρ_2, V_2) be two G -representations. An intertwiner between them is an equivariant linear map $L : V_1 \rightarrow V_2$. It satisfies

$$L \circ \rho_1(g) = \rho_2(g) \circ L \quad \forall g \in G, \quad (\text{B.44})$$

that is, it makes the following diagram commute:

$$\begin{array}{ccc} V_1 & \xrightarrow{L} & V_2 \\ \rho_1(g) \downarrow & & \downarrow \rho_2(g) \\ V_1 & \xrightarrow{L} & V_2 \end{array} \quad (\text{B.45})$$

The vector space of intertwiners is usually denoted as $\text{Hom}_G(V_1, V_2)$.

Recall our example of convolutions as translation equivariant maps in Appendix B.4. Since the convolution and the considered group action are linear, convolutions are identified as intertwiners.⁶

If V_1 and V_2 are finite dimensional and given some choice of bases for them, the intertwiner is represented by a matrix $L \in \mathbb{K}^{\dim(V_2) \times \dim(V_1)}$. Rewriting the equivariance constraint in Eq. (B.44), it needs to satisfy $\rho_2(g) L \rho_1(g)^{-1} = L$ for any group element g . Vectorizing this linear constraint leads to⁷

$$((\rho_1(g)^{-1})^\top \otimes \rho_2(g)) \text{ vec}(L) = \text{ vec}(L) \quad \forall g \in G, \quad (\text{B.46})$$

which shows that intertwiners are the invariants under the simultaneous action of ρ_1 and ρ_2 . This equation is particularly useful to solve numerically for intertwiners.

Working towards a better understanding of intertwiner spaces, specifically those between irreps, we need to introduce both *isomorphisms* and *endomorphisms* of representations:

⁶The (linear) group action on the convolution's domain and codomain is the translation of functions with domain \mathbb{R} , which is the (infinite dimensional) regular representation of $(\mathbb{R}, +)$; see Def. B.5.18.

⁷The *vectorization operator* $\text{vec} : \mathbb{K}^{m \times n} \rightarrow \mathbb{K}^{m \cdot n}$ acts on an $m \times n$ -matrix by stacking its columns into an $m \cdot n$ -vector. It satisfies $\text{vec}(AXB) = (B^\top \otimes A) \text{ vec}(X)$ for any triple A, X, B of dimensionally matching matrices [229].

Definition B.5.8 (Equivalent (isomorphic) representations).

Two G -representations (ρ_1, V_1) and (ρ_2, V_2) are said to be equivalent or isomorphic if there exists an invertible intertwiner, i.e. a vector space isomorphism $L : V_1 \xrightarrow{\sim} V_2$ satisfying $L \circ \rho_1(g) = \rho_2(g) \circ L \quad \forall g \in G$, between them.

In terms of matrix representations, this just means that there exists a change of basis (invertible intertwiner) Q such that the representations are similar as matrices:

$$\rho_2(g) = Q \rho_1(g) Q^{-1} \quad \forall g \in G. \quad (\text{B.47})$$

Definition B.5.9 (Endomorphism).

Consider a G -representation (ρ, V) . Intertwiners from (ρ, V) to itself, that is, linear maps $L : V \rightarrow V$ such that $L \circ \rho(g) = \rho(g) \circ L \quad \forall g \in G$, are called endomorphisms. The endomorphism space is written $\text{End}_G(V) = \text{Hom}_G(V, V)$.

With these preparations we can formulate *Schur's lemma*, which sheds light on the space of intertwiners between irreducible representations:

Lemma B.5.10 (Schur's lemma).

Let (ρ_1, V_1) and (ρ_2, V_2) be G -irreps over $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$, then:

- if (ρ_1, V_1) and (ρ_2, V_2) are not isomorphic, there exists no non-trivial⁸ intertwiner between them.
- if $(\rho_1, V_1) = (\rho_2, V_2) =: (\rho, V)$ are identical, any intertwiner is (by definition) an endomorphism and:
 - if $\mathbb{K} = \mathbb{C}$, it is given by scalar multiples λid_V of the identity, where $\lambda \in \mathbb{C}$.
 - if $\mathbb{K} = \mathbb{R}$, the endomorphism space $\text{End}_G(V)$ is either one-, two- or four-dimensional, depending on whether the representation is of real, complex or quaternionic type [173, 25].

The second statement of the theorem is usually only stated for complex representations (or representations over any other algebraically closed field), however, the case of real representations is of particular importance for the theory of steerable kernels [173].

B.5.2 Unitary representations of compact groups

In many applications the representation space is equipped with an *inner product*, making it an inner product space. It is in this case possible to define (not necessarily finite-dimensional) *unitary representations*, which assign unitary (norm-preserving) transformations to group elements. Specifically for unitary representations of *compact groups*, strong results regarding their decomposability into irreps can be proven. The Clebsch-Gordan decomposition and the Peter-Weyl theorem describe this decomposition for the case of tensor products of irreps and for quotient (or regular) representations, respectively.

Definition B.5.11 (Unitary transformation).

Let V_1 and V_2 be two inner product spaces. A unitary transformation from V_1 to V_2 is an isometric (norm preserving) bijective linear map:

$$U : V_1 \rightarrow V_2 \quad \text{such that} \quad \langle Uv, Uw \rangle_{V_2} = \langle v, w \rangle_{V_1} \quad \forall v, w \in V_1 \quad (\text{B.48})$$

Unitary transformations are the isomorphisms between inner product spaces.

⁸The zero-map $0 : V_1 \rightarrow \{0\} \subset V_2$ is a trivial intertwiner between any pair of representations (ρ_1, V_1) and (ρ_2, V_2) .

As the composition of unitary maps is again unitary, the unitary maps from a vector space to itself form a group:

Definition B.5.12 (Unitary group). *Let V be an inner product space. The unitary group*

$$\mathrm{U}(V) = \{g \in \mathrm{GL}(V) \mid \langle gv, gw \rangle_V = \langle v, w \rangle_V \quad \forall v, w \in V\} \leq \mathrm{GL}(V) \quad (\text{B.49})$$

is the group formed by all unitary transformations from V to itself.

If $V = \mathbb{C}^c$, the unitary group is concretely realized as the group of unitary matrices $\mathrm{U}(\mathbb{C}^c) = \{g \in \mathrm{GL}(\mathbb{C}^c) \mid gg^\dagger = g^\dagger g = \mathrm{id}_{\mathbb{C}^c}\}$, where \dagger is the Hermitian adjoint. In the case of real inner product spaces V , it is common to talk about *orthogonal groups* $\mathrm{O}(V)$ instead. Specifically for $V = \mathbb{R}^c$, the orthogonal group is realized by orthogonal matrices, i.e. $\mathrm{O}(c) = \{g \in \mathrm{GL}(c) \mid gg^\top = g^\top g = \mathrm{id}_{\mathbb{R}^c}\}$. We will in the following for simplicity refer to both as unitary groups.

Definition B.5.13 (Unitary representation).

A unitary representation of a locally compact topological group G on a (potentially infinite-dimensional) inner product space V is a tuple (ρ, V) where

$$\rho : G \rightarrow \mathrm{U}(V) \quad (\text{B.50})$$

is a continuous group homomorphism to the unitary group of V . The continuity requirement refers hereby to the continuity of the associated group action $G \times V \rightarrow V$, $(g, v) \mapsto \rho(g)v$.

The significance of unitary representations for the representation theory of compact groups becomes evident by the following theorem:

Theorem B.5.14 (Compact groups & unitary representations).

Every linear representation of a compact group on an inner product space is equivalent to a unitary representation.

This statement is certainly true if the inner product (\cdot, \cdot) on the inner product space is already G -invariant. If it is not, one can always define a group averaged inner product $\langle x, y \rangle := \int_G dg (gx, gy)$, with g being the Haar measure, which is by construction (left) G -invariant.

Definition B.5.15 (Isomorphism of unitary representations).

An isomorphism between unitary representations (ρ_1, V_1) and (ρ_2, V_2) is an invertible unitary intertwiner $V_1 \xrightarrow{\sim} V_2$.

A central result for finite unitary representations is their decomposability into irreps:

Theorem B.5.16 (Complete reducibility).

Let (ρ, V) be a finite dimensional unitary representation of any group G . It decomposes then into an orthogonal direct sum $\rho \cong \bigoplus_i \rho_i$ of irreducible unitary subrepresentations ρ_i [297].

An explicit application is the Clebsch-Gordan decomposition of tensor products of irreps into their irreducible subrepresentations.

Definition B.5.17 (Clebsch-Gordan decomposition and coefficients).

Let (ρ_l, V_l) and (ρ_k, V_k) be unitary irreducible representations of a compact group G . Their tensor product $\rho_l \otimes \rho_k$ is not necessarily irreducible, however, by Theorem B.5.16 there exists an isomorphism

$$\mathrm{CG}_{lk} : V_l \otimes V_k \rightarrow \bigoplus_{j \in \widehat{G}} \bigoplus_{s=1}^{m_{j,lk}} V_j, \quad (\text{B.51})$$

known as Clebsch-Gordan decomposition, which decouples the tensor product into a direct sum of irreps. \hat{G} denotes hereby the set of isomorphism classes of unitary irreps of G while $m_{j,lk} \in \mathbb{N}_0$ is the multiplicity of irrep j in the tensor product of irreps l and k .

To make the Clebsch-Gordan decomposition concrete, consider a choice of basis $\{e_j^\mu \mid \mu = 1, \dots, \dim(V_j)\}$ of irrep j , implying basis tensors $e_l^m \otimes e_k^n$ of $V_l \otimes V_k$ and basis elements e_{js}^M of $\bigoplus_{s=1}^{m_{j,lk}} V_j$ (additionally labeled by s). The Clebsch-Gordan coefficients are then the matrix elements of CG_{lk} in this basis:

$$\langle s, jM | lm; kn \rangle := \langle e_{js}^M | \text{CG}_{lk} | e_l^m \otimes e_k^n \rangle \quad (\text{B.52})$$

Harmonic analysis (Fourier transforms) on homogeneous spaces decompose a signal into a basis of harmonic functions. It is formalized in the Peter-Weyl theorem, which describes the decomposition of unitary *regular* or *quotient representations* into their irreducible subspaces, corresponding to the individual harmonics.

Definition B.5.18 (Regular representation).

The (left) regular representation $(\rho_{\text{reg}}^G, L^2_{\mathbb{K}}(G))$ of a group G acts on the space of square integrable functions on the group by left translation:

$$[\rho_{\text{reg}}^G(\tilde{g}) f](g) := f(\tilde{g}^{-1}g) \quad \forall g, \tilde{g} \in G, f \in L^2_{\mathbb{K}}(G) \quad (\text{B.53})$$

Remark B.5.19 (Regular representation for finite groups). Specifically for finite groups

G , functions $f : G \rightarrow \mathbb{K}$ can take $|G|$ independent values for the $|G|$ group elements, and are therefore in one-to-one correspondence to finite-dimensional vectors $\vec{f} \in \mathbb{K}^{|G|}$. To make this isomorphism explicit, one identifies the canonical basis of $\mathbb{K}^{|G|}$ with group elements, i.e. considers basis vectors e_g for any $g \in G$, and defines $\vec{f} = \sum_g f(g)e_g$. This implies the regular group action $\rho_{\text{reg}}^G(\tilde{g}) \vec{f} = \sum_g f(\tilde{g}^{-1}g)e_g = \sum_g f(g)e_{\tilde{g}g}$ on \vec{f} , first written as acting on the coefficients, then on the basis itself. More abstractly, $\mathbb{K}^{|G|}$ with the regular representation action forms the group algebra $\mathbb{K}[G]$ of G over \mathbb{K} ; see footnote 5 above.

Note that regular representations are *permutation representations*, i.e. act by permuting function (or vector) values. Table B.1 gives an explicit example of the regular representation of the cyclic group C_4 , instantiated by 4×4 permutation matrices.

Definition B.5.20 (Quotient representation). Let G be a group with subgroup $H \leq G$.

The corresponding quotient representation $(\rho_{\text{quot}}^{G/H}, L^2_{\mathbb{K}}(G/H))$ of G acts on the space of square integrable functions on the homogeneous space G/H by left translation:

$$[\rho_{\text{quot}}^{G/H}(\tilde{g}) f](gH) := f(\tilde{g}^{-1}gH) \quad \forall \tilde{g} \in G, gH \in G/H, f \in L^2_{\mathbb{K}}(G/H) \quad (\text{B.54})$$

Remark B.5.21 (Quotient representation for finite groups). For finite groups G , one may as in Remark B.5.19 identify functions $f : G/H \rightarrow \mathbb{K}$ with finite-dimensional vectors, here $\vec{f} \in \mathbb{K}^{|G|/|H|}$. The basis of $\mathbb{K}^{|G|/|H|}$ is naturally labeled by cosets gH in the quotient space G/H , and the quotient representation action permutes basis vectors via the group action on cosets, $\rho_{\text{quot}}^{G/H}(\tilde{g})e_{gH} = e_{\tilde{g}gH}$.

The regular representation is a special case of quotient representations for $H = \{e\}$. That these representations are well defined homomorphisms is easily checked by asserting that $[\rho_{\text{quot}}^{G/H}(\tilde{k}) \rho_{\text{quot}}^{G/H}(\tilde{g}) f](gH) = [\rho_{\text{quot}}^{G/H}(\tilde{g}) f](k^{-1}gH) = f(\tilde{g}^{-1}k^{-1}gH) =$

$f((k\tilde{g})^{-1}gH) = [\rho_{\text{quot}}^{G/H}(k\tilde{g}) f](gH)$, which holds for arbitrary $k, \tilde{g} \in G$, $gH \in G/H$ and $f \in L^2_{\mathbb{K}}(G/H)$.

Quotient (and thus regular) representations of compact groups are decomposed as follows:⁹

Theorem B.5.22 (Peter-Weyl). *The quotient representation $(\rho_{\text{quot}}^{G/H}, L^2_{\mathbb{K}}(G/H))$ of a compact group G decomposes into irreducible subrepresentations*

$$L^2_{\mathbb{K}}(G/H) \cong \widehat{\bigoplus}_{j \in \widehat{G}} \bigoplus_{i=1}^{m_j} V_j, \quad (\text{B.55})$$

where \widehat{G} is the set of isomorphism classes of G -irreps, $\widehat{(\cdot)}$ is a topological closure and the integer $m_j \leq \dim(V_j)$ is the multiplicity of irrep V_j in $L^2_{\mathbb{K}}(G/H)$.

For $\mathbb{K} = \mathbb{C}$ and $H = \{e\}$, i.e. complex regular representations, one has multiplicities $m_j = \dim(V_j)$.

Since quotient representations act on (square integrable) functions $f : G/H \rightarrow \mathbb{K}$, the Peter-Weyl theorem ensures in practice that there exist linear subspaces of such functions which transform according to irreducible representations, and that general functions may be expanded in terms of functions from these subspaces. Specifically, denote by

$$\{Y_{ji}^m : G/H \rightarrow \mathbb{K} \mid m = 1, \dots, \dim(V_j)\} \quad (\text{B.56})$$

an (orthonormal) basis of the i -th occurrence of V_j in the the Peter-Weyl decomposition, Eq. (B.55), satisfying by definition the *G-steerability condition*

$$\sum_{n=1}^{\dim V_j} \rho_j(\tilde{g})_{nm} Y_{ji}^n(gH) = Y_{ji}^m(\tilde{g}^{-1}gH) \quad (\text{B.57})$$

and the *orthonormality relation*

$$\int_{G/H} \overline{Y_{ji}^m(gH)} Y_{j'i'}^{m'}(gH) d(gH) = \delta_{jj'} \delta_{ii'} \delta_{mm'}. \quad (\text{B.58})$$

Then there exist expansion coefficients $\lambda_{ji}^m \in \mathbb{K}$ such that any $f \in L^2_{\mathbb{K}}(G/H)$ is given by $f = \sum_{mji} \lambda_{ji}^m Y_{ji}^m$. The Y_{ji}^m are called *harmonic basis functions*. Well-known examples are *circular harmonics* for $G = \text{SO}(2)$ and $H = \{e\}$, such that $G/H \cong_{\text{top}} S^1$ and *spherical harmonics* for $G = \text{SO}(3)$ and $H = \text{SO}(2)$, such that $G/H \cong_{\text{top}} S^2$; see Figs. 5.2 and 5.3 respectively.

⁹Regular and quotient representations are defined for arbitrary (not necessarily compact) groups, but the Peter-Weyl theorem applies only to compact groups.

Coordinate chart formalism of differential geometry

This appendix serves the purpose of drawing connections between the *fiber bundle formalism*, underlying the theory of coordinate independent CNNs, and the *coordinate chart formalism*, which one likely encounters in a first study of differential geometry. The main difference between both is that the bundle formalism refers to points p of the base space M in a *coordinate free* way. If required, coordinates are directly assigned to the fibers (e.g. tangent spaces) via local bundle trivializations. In contrast, the chart formalism relies on *coordinate charts* (diffeomorphisms)

$$x : M \supseteq U \rightarrow V \subseteq \mathbb{R}^d, \quad (\text{C.1})$$

which assign coordinates to local patches U of the manifold. Local bundle trivializations and gauge transformations between them are *induced as differentials of charts and chart transition functions*. In this appendix we work out the connection between both formalisms. An overview of the results is given in Table C.1.

We start in Appendix C.1 by briefly introducing tangent spaces $T_p M$ as spaces of directional derivative operators, from which the cotangent spaces $T_p^* M$ follow as dual spaces. Appendix C.2 defines general differentials and the more specific gradients and Jacobians. Based on these preparations, we will in Appendix C.3.1 define *coordinate bases* (holonomic bases) $\left[\frac{\partial}{\partial x_1} \Big|_p, \dots, \frac{\partial}{\partial x_d} \Big|_p \right] \in F_p M$ of the tangent spaces $T_p M$, which are spanned by directional derivative operators along the coordinate grid that is pulled by the chart from V to U . The dual bases $\left[\hat{dx}_\mu |_p, \dots, \hat{dx}_\mu |_p \right]$ of the cotangent spaces $T_p^* M$ are given by the gradients of the chart components x_μ . Transition maps between charts induce covariant and contravariant gauge transformations between the corresponding bases, which are derived in Appendix C.3.2. Appendix C.4 interprets the coordinate bases as local bundle trivializations and makes the connection between the bundle formalism and the chart formalism precise. The bases and trivializations induced from coordinate charts do not cover all possible trivializations, such that one distinguishes between coordinate bases and non-coordinate bases (the bundle formalism allows with general non-coordinate bases). In the physics literature, non-coordinate bases are usually introduced via *vielbein fields*. Appendix C.5 argues that these vielbein fields are just $GL(d)$ -valued gauge transformations from general frames in $F M$ into a given G -structure $G M$, within which one can subsequently apply G -valued gauge transformations that preserve the G -structure.

Comprehensive introductions to the chart formalism are given in [221, 262, 32]. A more rigorous exposition is found in [262].

We want to remind the reader that we are *not* making use of covariant and contravariant indices. Indices will always appear as subscripts, with Greek letters μ, ν, \dots signaling coordinate chart related indices and Latin letters i, j, \dots signaling indices of general gauges. Superscripts A, B, \dots are preserved for labeling different charts or gauges.

C.1 Tangent spaces, cotangent spaces and dual bases

C.1.1 Tangent spaces in terms of directional derivatives

common definition of the tangent spaces $T_p M$ of a manifold M is as vector spaces of directional derivative operators at $p \in M$, which we will briefly motivate here. Let $f \in C^\infty(M)$, that is, $f : M \rightarrow \mathbb{R}$ is a smooth map, and, for some interval $I \subseteq \mathbb{R}$ containing 0, let $\gamma : I \rightarrow M$ be a smooth curve which passes at time $t = 0$ through p , i.e. satisfies $\gamma(0) = p$. One then defines the *directional derivative operator* at p along γ as the linear operator

$$v_\gamma : C^\infty(M) \rightarrow \mathbb{R}, \quad f \mapsto (f \circ \gamma)'(0). \quad (\text{C.2})$$

As the derivative is taken along the direction of γ , that is, tangential to it, v_γ is called *tangent vector*. It can be thought of as the velocity of a particle with trajectory γ at time $t = 0$. For later reference we give the following simple commutative diagram, which shows the pullback $f \circ \gamma$ of f from M to \mathbb{R} via γ , in terms of which the directional derivative is defined:

$$\begin{array}{ccccc} \mathbb{R} & \supset & I & \xrightarrow{\gamma} & M & \xrightarrow{f} & \mathbb{R} \\ & & \downarrow & & \downarrow & & \uparrow \\ & & & & f \circ \gamma & & \end{array} \quad (\text{C.3})$$

One can show that the space of all tangent vectors to curves at p forms a d -dimensional vector space

$$T_p M := \{v_\gamma \mid \gamma \text{ is a smooth curve through } p\}, \quad (\text{C.4})$$

known as the tangent space at p . For more details on the definition of tangent vectors and the vector space structure of the tangent spaces we refer to [262].

Having defined the tangent spaces as vector spaces, one might choose to treat tangent vectors as abstract geometric vectors, thereby “forgetting” about their definition via directional derivatives (or any alternative definition made). We do this at most places, but refer back to the definition via directional derivatives in the following sections to derive differentials of smooth maps and coordinate bases.

C.1.2 Cotangent spaces

As real vector spaces, the tangent spaces $T_p M$ have corresponding *dual spaces* $T_p^* M := (T_p M)^*$, the *cotangent spaces*. By the definition of dual spaces, they consist of linear functionals

$$\omega : T_p M \rightarrow \mathbb{R}, \quad (\text{C.5})$$

which are in differential geometry usually called *covectors* or *1-forms*. Together with the (co)vector addition $(\omega + \tilde{\omega})(v) = \omega(v) + \tilde{\omega}(v)$ and scalar multiplication $(\lambda \cdot \omega)(v) = \lambda \cdot (\omega(v))$, the cotangent spaces are vector spaces themselves.

As finite-dimensional duals of each other, $T_p M$ and $T_p^* M$ are isomorphic and are thus in particular of the same dimensionality $d = \dim(M) = \dim(T_p M) = \dim(T_p^* M)$. The isomorphism between both is, however, not canonical. A vector space isomorphism can be specified via a (non-degenerate) bilinear form $\eta_p : T_p M \times T_p M \rightarrow \mathbb{R}$ on $T_p M$, for instance a Riemannian metric, via

$$\hat{\eta}_p : T_p M \rightarrow T_p^* M, \quad v \mapsto \eta_p(v, \cdot), \quad (\text{C.6})$$

which determines the linear functional $\hat{\eta}_p(v) : T_p M \rightarrow \mathbb{R}, \quad w \mapsto \eta_p(v, w)$.

C.1.3 Dual bases

Any basis $[e_i]_{i=1}^d$ of $T_p M$ canonically induces a *dual basis* $[e_i^*]_{i=1}^d$ of $T_p^* M$, defined to satisfy the relations

$$e_i^* e_j = \delta_{ij} \quad \text{for any } i, j \in 1, \dots, d. \quad (\text{C.7})$$

Let $[e_i^A]_{i=1}^d$ and $[e_i^B]_{i=1}^d = [e_i^A]_{i=1}^d \triangleleft (g^{BA})^{-1}$ be two bases of $T_p M$, which are related by the right action \triangleleft of the (inverse) structure group element $(g^{BA})^{-1} \in \text{GL}(d)$ in Eq. (7.10), that is, for $j = 1, \dots, d$:

$$e_j^B = \sum_l e_l^A (g^{BA})_{lj}^{-1} \quad (\text{C.8})$$

The dual basis $[e_i^{A,*}]_{i=1}^d$ transforms accordingly under that left action which sends $e_i^{A,*}$ to

$$e_i^{B,*} = \sum_k g_{ik}^{BA} e_k^{A,*}. \quad (\text{C.9})$$

This is affirmed by pairing

$$\begin{aligned} e_i^{B,*} e_j^B &= \sum_{k,l} g_{ik}^{BA} e_k^{A,*} e_l^A (g^{BA})_{lj}^{-1} \\ &= \sum_{k,l} g_{ik}^{BA} \delta_{kl} (g^{BA})_{lj}^{-1} \\ &= \sum_k g_{ik}^{BA} (g^{BA})_{kj}^{-1} \\ &= \delta_{ij}. \end{aligned} \quad (\text{C.10})$$

The inverse transformation behavior of bases and dual bases is usually referred to as *covariant* and *contravariant* transformation. Note the similarity of the dual basis transformation to the contravariant transformations $\psi^B = g^{BA} \psi^A$ of gauges in Eq. (7.7) and $v^B = g^{BA} v^A$ of vector components in (7.9). Indeed, gauges are just choices of a cotangent basis as further discussed below.

C.2 Differentials, gradients and Jacobians

In vector calculus one considers functions $\phi : \mathbb{R}^m \rightarrow \mathbb{R}^n$, which can at any point $p \in \mathbb{R}^m$ be linearly approximated by their Jacobian matrix (or total derivative or differential) $d\phi_p = (\frac{\partial \phi_i}{\partial x_j}|_p)_{ij}$. Here we introduce the generalization of this concept to differentials of smooth functions between smooth manifolds.

Differentials in general: Let $\phi : M \rightarrow N$ be a smooth map between smooth manifolds M and N . At any point $p \in M$, such a map induces a differential (or pushforward)

$$d\phi_p : T_p M \rightarrow T_{\phi(p)} N, \quad v \mapsto d\phi_p(v) \quad (\text{C.11})$$

which linearly maps tangent vectors at p to tangent vectors at $\phi(p)$. For the definition of tangent spaces in terms of directional derivatives in Eq. (C.2), the pushforward of $v \in T_p M$ along ϕ is explicitly given by

$$d\phi_p(v) : C^\infty(N) \rightarrow \mathbb{R}, \quad f \mapsto (d\phi_p(v))(f) := v(f \circ \phi), \quad (\text{C.12})$$

that is, by the application of v on the pullback $f \circ \phi : M \rightarrow \mathbb{R}$ of $f : N \rightarrow \mathbb{R}$ via ϕ . These definitions are clarified by the following two commutative diagrams:

$$\begin{array}{ccc} M & \xrightarrow{\phi} & N \\ & \searrow f \circ \phi & \downarrow f \\ & & \mathbb{R} \end{array} \qquad \begin{array}{ccc} C^\infty(M) & \xleftarrow{(\cdot) \circ \phi} & C^\infty(N) \\ v \downarrow & & \swarrow d\phi(v) \\ \mathbb{R} & & \end{array} \quad (\text{C.13})$$

From this definition it follows immediately that the differential of the composition of smooth maps equals the composition of their individual differentials, which is just the chain rule:

$$d(\phi \circ \psi)_p = d\phi_{\psi(p)} \circ d\psi_p \quad (\text{C.14})$$

If ϕ is invertible (a diffeomorphism) it furthermore follows that its differential is a vector space isomorphism whose inverse equals the differential of ϕ^{-1} , that is,

$$(d\phi_p)^{-1} = d(\phi^{-1})_{\phi(p)}. \quad (\text{C.15})$$

Together, the differentials $d\phi_p$ at individual points $p \in M$ imply a vector bundle morphism (a fiber-wise linear bundle map, see Sections 11.1) between the tangent bundles of M and N :

$$\begin{array}{ccc} TM & \xrightarrow{d\phi} & TN \\ \pi_{TM} \downarrow & & \downarrow \pi_{TN} \\ M & \xrightarrow{\phi} & N \end{array} \quad (\text{C.16})$$

Note that we are in this appendix using a different notation, namely $d\phi$, than in the main paper, where we instead write $\phi_{*,TM}$. We decided for the former to connect to the usual notation dx_μ for the chart induced bases of cotangent spaces. The latter is used in the main text to emphasize the similarity to the bundle maps $\phi_{*,FM}$, $\phi_{*,GM}$ and $\phi_{*,A}$, which are induced on the associated bundles FM , GM and \mathcal{A} .

Gradients: In the case of smooth real-valued functions $\phi : M \rightarrow \mathbb{R}$, i.e. $\phi \in C^\infty(M)$, the differential $d\phi_p : T_p M \rightarrow T_{\phi(p)} \mathbb{R}$ pushes vectors v in $T_p M$ to vectors $d\phi(v) : C^\infty(\mathbb{R}) \rightarrow \mathbb{R}$, $f \mapsto v(f \circ \phi)$ in $T_{\phi(p)} \mathbb{R}$. By leveraging the canonical isomorphism

$$\iota_{\mathbb{R}} : T_{\phi(p)} \mathbb{R} \xrightarrow{\sim} \mathbb{R}, \quad v \mapsto v(\text{id}_{\mathbb{R}}) \quad (\text{C.17})$$

one defines the *gradient operator*

$$\hat{d}_p : C^\infty(M) \rightarrow T_p^*M, \quad \phi \mapsto \hat{d}\phi_p := \iota_{\mathbb{R}} \circ d\phi_p = (d\phi_p(\cdot))(id_{\mathbb{R}}), \quad (\text{C.18})$$

which sends smooth functions ϕ to covectors¹ $\hat{d}\phi$, which in turn act on vectors as

$$\hat{d}\phi_p : T_p M \rightarrow \mathbb{R}, \quad v \mapsto \hat{d}\phi_p(v) = (d\phi_p(v))(id_{\mathbb{R}}) = v(id_{\mathbb{R}} \circ \phi) = v(\phi). \quad (\text{C.19})$$

By an *abuse of notation* one usually drops the “hat” on \hat{d} and immediately defines $d\phi_p(v) := v(\phi)$. While this notation is very common, we stick in the following with the “hat” to make the requirement for the canonical isomorphism $\iota_{\mathbb{R}}$ explicit.

In Appendix C.3.1 below we will see that the bases of T_p^*M which are dual to coordinate bases of $T_p M$ are given by the gradient 1-forms $\hat{d}x_\mu|_p$, where x_μ are the components of the coordinate chart.

Jacobians: Specifically for functions $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ between (subsets of) Euclidean spaces the differential $d\phi_{x_0} : T_{x_0}\mathbb{R}^n \rightarrow T_{\phi(x_0)}\mathbb{R}^m$ is easily seen to coincide with the *Jacobian* $\frac{\partial \phi}{\partial x}|_{x_0} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ after canonically identifying $T_p\mathbb{R}^k \cong \mathbb{R}^k$ in both the domain and codomain. The canonical isomorphism is here given by

$$\iota_{\mathbb{R}^k} : v \mapsto (v(\text{proj}_1), \dots, v(\text{proj}_k)), \quad (\text{C.20})$$

which generalizes $\iota_{\mathbb{R}}$ from Eq. (C.17) to multiple dimensions. As the calculation is mostly similar as in the case of gradients, we will not repeat it here but visualize the idea via a commutative diagram:

$$\begin{array}{ccccc} \mathbb{R}^n & \xleftarrow{\iota_{\mathbb{R}^n}} & T_{x_0}\mathbb{R}^n & \xrightarrow{d\phi|_{x_0}} & T_{\phi(x_0)}\mathbb{R}^m & \xrightarrow{\iota_{\mathbb{R}^m}} & \mathbb{R}^m \\ & & \downarrow & & & & \uparrow \\ & & \frac{\partial \phi}{\partial x} \Big|_{x_0} & & & & \end{array} \quad (\text{C.21})$$

If ϕ is invertible, the identity in Eq. (C.15) becomes

$$\frac{\partial \phi}{\partial x} \Big|_{x_0}^{-1} = \frac{\partial \phi^{-1}}{\partial x} \Big|_{\phi(x_0)}, \quad (\text{C.22})$$

which is just the inverse function theorem. We will use this identity later on to invert gauge transformations between different coordinate bases which are induced as Jacobians of chart transition maps.

C.3 Chart induced coordinate bases

In this section we consider *coordinate charts* of the form

$$x : U \rightarrow V, \quad (\text{C.23})$$

¹The gradient field is often defined as a *vector* field $\nabla f := (\hat{df})^{\sharp\eta}$ which is computed from the covector field \hat{df} via the musical isomorphism $\sharp^\eta : T^*M \rightarrow TM$ corresponding to the metric (“raising indices”).

which diffeomorphically assign coordinates $x(p) \in V \subseteq \mathbb{R}^d$ to each point $p \in U \subseteq M$. Any such chart induces a natural choice of bases for the tangent spaces $T_p M$ over U , known as *coordinate bases*. The dual spaces $T_p^* M$ of the tangent spaces over U are accordingly endowed with dual coordinate bases of cotangent vectors. Transition maps between the coordinates of two charts induce gauge transformations which translate between the corresponding coordinate bases. These gauge transformations are given by the Jacobians of the transition maps.

C.3.1 Charts and induced coordinate bases

Coordinate bases for $T_p M$: To motivate the definition of coordinate bases, observe that x implies a “coordinate grid” on U by pulling the canonical coordinate grid on V back to the manifold. The coordinate basis at a specific point $p \in U$ can then be thought of as consisting of those d many *directional derivative operators* which are going *along the coordinate grid lines of x on U* .

To make this more precise, consider first the curves

$$\tilde{\gamma}_\mu : I \rightarrow V, \quad t \mapsto x(p) + t\epsilon_\mu \quad \mu = 1, \dots, d \quad (\text{C.24})$$

which pass at time $t = 0$ with unit velocity in μ -direction through $x(p) \in V$. Mapping those $\tilde{\gamma}_\mu$ via the chart to U defines the above mentioned curves

$$\gamma_\mu : I \rightarrow U, \quad t \mapsto x^{-1} \circ \tilde{\gamma}_\mu(t) = x^{-1}(x(p) + t\epsilon_\mu) \quad (\text{C.25})$$

which pass at time $t = 0$ along the coordinate grid of x on U through p . The d -dimensional coordinate basis of $T_p M$ induced by x is then given by the directional derivative operators in Eq. (C.2) along the paths γ_μ . Denoting the μ -th basis vector by the usual abuse of notation as $\frac{\partial}{\partial x_\mu}|_p$ one therefore defines:

$$\begin{aligned} \left. \frac{\partial}{\partial x_\mu} \right|_p : f &\mapsto \left. \frac{\partial}{\partial x_\mu} \right|_p f := (f \circ \gamma_\mu)'(0) \\ &= (f \circ x^{-1} \circ \tilde{\gamma}_\mu)'(0) \\ &= (f \circ x^{-1}(x(p) + t\epsilon_\mu))'(0) \\ &= [\partial_\mu(f \circ x^{-1})](x(p)) \end{aligned} \quad (\text{C.26})$$

In the last step we identified the usual μ -th partial derivative of the pullback $f \circ x^{-1} : V \rightarrow \mathbb{R}$, which motivates the notation $\left. \frac{\partial}{\partial x_\mu} \right|_p$. These definitions are visualized in the following commutative diagram which extends the diagram in Eq. (C.3):

$$\begin{array}{ccccc} & & V & & \\ & \nearrow \tilde{\gamma}_\mu & \uparrow x & \searrow f \circ x^{-1} & \\ \mathbb{R} \ni I & \xrightarrow{\gamma_\mu} & U & \xrightarrow{f} & \mathbb{R} \\ & \underbrace{\hspace{10em}}_{f \circ \gamma_\mu} & & & \end{array} \quad (\text{C.27})$$

Dual coordinate bases for T_p^*M : As stated in Appendix C.1, any basis of $T_p M$ induces a *dual basis* of $T_p^* M$. Specifically for coordinate bases, spanned by vectors $\frac{\partial}{\partial x_\mu}|_p$, the dual basis elements are given by the *gradients* $\hat{dx}_\mu|_p = \hat{d}(x_\mu)_p \in T_p^* M$ of the chart components $x_\mu = \text{proj}_\mu \circ x : U \rightarrow \mathbb{R}$. That these gradients do indeed make up the dual basis, is easily seen by acting on the basis vectors as defined in Eq. (C.19):

$$\begin{aligned}\hat{dx}_\mu|_p \frac{\partial}{\partial x_\nu}|_p &= \frac{\partial}{\partial x_\nu}|_p x_\mu \\ &= [\partial_\nu(x_\mu \circ x^{-1})](x(p)) \\ &= [\partial_\nu(\text{proj}_\mu)](x(p)) \\ &= \delta_{\mu\nu}.\end{aligned}\tag{C.28}$$

Chart differentials as canonical local trivialization: Given that the chart maps from $U \subseteq M$ to $V \subseteq \mathbb{R}^d$, its differentials at $p \in U$ are maps of the form

$$dx_p : T_p M \rightarrow T_{x(p)} \mathbb{R}^d.\tag{C.29}$$

Employing the canonical isomorphism $\iota_{\mathbb{R}^d}$ from $T_{x(p)} \mathbb{R}^d$ to \mathbb{R}^d from Eq. (C.20) once again, we obtain a map

$$\begin{aligned}\hat{dx}_p : T_p M &\rightarrow \mathbb{R}^d, \\ v &\mapsto \hat{dx}_p(v) := \iota_{\mathbb{R}^d} \circ dx_p(v) \\ &= ((dx_p(v))(\text{proj}_1), \dots, (dx_p(v))(\text{proj}_d))^\top \\ &= (v(\text{proj}_1 \circ x \circ x^{-1})(x(p)), \dots, v(\text{proj}_d \circ x \circ x^{-1})(x(p)))^\top \\ &= (v(x_1(p)), \dots, v(x_d(p)))^\top \\ &= (\hat{dx}_1|_p(v), \dots, \hat{dx}_d|_p(v))^\top\end{aligned}\tag{C.30}$$

after identifying the individual chart component gradients in the last step. Note that the action of this chart differential on the μ -th coordinate basis yields

$$\begin{aligned}\hat{dx}_p \frac{\partial}{\partial x_\mu}|_p &= \left(\hat{dx}_1|_p \frac{\partial}{\partial x_\mu}|_p, \dots, \hat{dx}_d|_p \frac{\partial}{\partial x_\mu}|_p \right)^\top \\ &= (\delta_{\mu 1}, \dots, \delta_{\mu d})^\top \\ &= \epsilon_\mu,\end{aligned}\tag{C.31}$$

that is, the μ -th unit vector ϵ_μ of \mathbb{R}^d . This implies that $\hat{dx}_p : T_p M \rightarrow \mathbb{R}^d$ plays the role of a *gauge* ψ_p at p . One could therefore equally well have started by defining a cotangent basis and setting

$$\frac{\partial}{\partial x_\mu}|_{x(p)} = \hat{dx}_p^{-1}(\epsilon_\mu),\tag{C.32}$$

which is the analog of Eq. (7.4) in the chart formalism.

C.3.2 Chart transition maps and induced gauge transformations

Different charts induce different coordinate bases. Chart transitions therefore induce gauge transformations, i.e. transformations of bases and vector coefficients, which we derive in this section.

In the following we consider two arbitrary, overlapping charts $x^A : U^A \rightarrow V^A$ and $x^B : U^B \rightarrow V^B$. The different coordinates which they assign to the overlap $U^A \cap U^B \neq \emptyset$ are then related via *chart transition maps*

$$x^B \circ (x^A)^{-1} : x^A(U^A \cap U^B) \rightarrow x^B(U^A \cap U^B). \quad (\text{C.33})$$

Transformation of tangent coordinate bases: The coordinate bases of $T_p M$ which are induced by the two charts are according to the last line of Eq. (C.26) by their action on $f \in C^\infty(M)$ defined as

$$\frac{\partial}{\partial x_\mu^A} \Big|_p f = \left[\partial_\mu(f \circ (x^A)^{-1}) \right] (x^A(p)) \quad (\text{C.34})$$

and

$$\frac{\partial}{\partial x_\mu^B} \Big|_p f = \left[\partial_\mu(f \circ (x^B)^{-1}) \right] (x^B(p)), \quad (\text{C.35})$$

which is visualized by the following commutative diagram:

$$\begin{array}{ccc} & V^A \supset x^A(U^A \cap U^B) & \\ & \uparrow x^A & \searrow f \circ (x^A)^{-1} \\ x^B \circ (x^A)^{-1} & U^A \cap U^B & \xrightarrow{f} \mathbb{R} \\ & \downarrow x^B & \swarrow f \circ (x^B)^{-1} \\ & V^B \supset x^B(U^A \cap U^B) & \end{array} \quad (\text{C.36})$$

Via the chart transition maps, the different coordinate bases relate by

$$\begin{aligned} \frac{\partial}{\partial x_\mu^B} \Big|_p f &= \left[\partial_\mu(f \circ (x^B)^{-1}) \right] (x^B(p)) \\ &= \left[\partial_\mu(f \circ (x^A)^{-1} \circ x^A \circ (x^B)^{-1}) \right] (x^B(p)), \end{aligned} \quad (\text{C.37})$$

which, making use of the multivariate chain rule, further leads to:

$$\begin{aligned} \frac{\partial}{\partial x_\mu^B} \Big|_p f &= \sum_{\nu=1}^d \left[\partial_\nu(f \circ (x^A)^{-1}) \right] (x^A(p)) \cdot \left[\partial_\mu(x_\nu^A \circ (x^B)^{-1}) \right] (x^B(p)) \\ &= \sum_{\nu=1}^d \frac{\partial f}{\partial x_\nu^A} \Big|_p \frac{\partial x_\nu^A}{\partial x_\mu^B} \Big|_{x^B(p)} \end{aligned} \quad (\text{C.38})$$

In the last step we introduced the usual abuse of notation²

$$\frac{\partial x_\nu^A}{\partial x_\mu^B} \Big|_{x^B(p)} := \partial_\mu(x_\nu^A \circ (x^B)^{-1})(x^B(p)) \quad (\text{C.39})$$

for the components of the *Jacobian*

$$\frac{\partial x^A}{\partial x^B} \Big|_{x^B(p)} = \hat{dx}_p^A \circ \hat{d}(x_p^B)^{-1} \quad (\text{C.40})$$

of the transition maps. Dropping f from Eq. (C.37), we identify the transformation law

$$\frac{\partial}{\partial x_\mu^B} \Big|_p = \sum_{\nu=1}^d \frac{\partial}{\partial x_\nu^A} \Big|_p \frac{\partial x_\nu^A}{\partial x_\mu^B} \Big|_{x^B(p)} \quad (\text{C.41})$$

of tangent coordinate bases. We did hereby choose to write the Jacobian on the right of the basis vector to emphasize that the change of basis is to be understood as a *right action*. Doing so, we need to warn the reader that $\frac{\partial}{\partial x_\nu^A} \Big|_p$ is just an abuse of notation for the basis vector but does not imply an action of a differential operator on the Jacobian on the right.

Transformation of cotangent coordinate bases: The contravariant transformation law of cotangent space coordinate bases follows from the inverse transformation of dual bases in Eq. (C.9) relative to (C.8). To apply this relation, we first adapt Eq. (C.41) to our convention that bases transform according to a right action with an *inverse* group element. This is achieved by applying Eq. (C.22) to invert the Jacobian (remember the abuse of notation)

$$\frac{\partial x^A}{\partial x^B} \Big|_{x^B(p)} = \frac{\partial x^B}{\partial x^A} \Big|_{x^A(p)}^{-1} \quad (\text{C.42})$$

which implies:

$$\begin{aligned} \frac{\partial}{\partial x_\mu^B} \Big|_p &= \sum_{\nu=1}^d \frac{\partial}{\partial x_\nu^A} \Big|_p \frac{\partial x_\nu^A}{\partial x_\mu^B} \Big|_{x^B(p)} = \sum_{\nu=1}^d \frac{\partial}{\partial x_\nu^A} \Big|_p \left(\frac{\partial x^A}{\partial x^B} \Big|_{x^B(p)} \right)_{\nu\mu} \\ &= \sum_{\nu=1}^d \frac{\partial}{\partial x_\nu^A} \Big|_p \left(\frac{\partial x^B}{\partial x^A} \Big|_{x^A(p)}^{-1} \right)_{\nu\mu} \end{aligned} \quad (\text{C.43})$$

The cotangent basis elements therefore transform according to Eqs. (C.8) and (C.9) like

$$\hat{dx}_\mu^B|_p = \sum_{\nu=1}^d \frac{\partial x_\mu^B}{\partial x_\nu^A} \Big|_{x^A(p)} \hat{dx}_\nu^A|_p. \quad (\text{C.44})$$

Transformation of chart differentials: The expression of chart differentials $\hat{dx}^A|_p$ in terms of chart component gradients $\hat{dx}_\mu^A|_p$ in Eq. (C.30) allows to deduce their transformation law from that in Eq. (C.44). Alternatively, one obtains the transformation law by

²The “abuse” is that x^A is interpreted as a function of $x^B(p)$, and should therefore rather be written $x^A \circ (x^B)^{-1}$ as made precise on the right-hand side.

right multiplying with the identity in the form $\text{id}_{T_p M} = \hat{dx}^A|_p \circ (\hat{dx}^A|_p)^{-1}$ and identify a left multiplication with the Jacobian of the chart transition maps:

$$\begin{aligned}\hat{dx}^B|_p &= \hat{dx}^B|_p \circ (\hat{dx}^A|_p)^{-1} \circ \hat{dx}^A|_p \\ &= \left. \frac{\partial x^B}{\partial x^A} \right|_{x^A(p)} \hat{dx}^A|_p\end{aligned}\quad (\text{C.45})$$

Note that this result is simply the matrix expression of Eq. (C.44).

Transformation of vector coefficients: Vectors $v \in T_p M$ are relative to a coordinate basis $\left[\frac{\partial}{\partial x_\mu^B} \right]_{\mu=1}^d$ expressed by coefficients $v^A \in \mathbb{R}^d$:

$$v = \sum_{\mu=1}^d v_\mu^A \left. \frac{\partial}{\partial x_\mu^A} \right|_p \quad (\text{C.46})$$

The individual coefficients are recovered by the action of the cotangent basis:

$$\hat{dx}_\mu^A|_p(v) = \hat{dx}_\mu^A|_p \sum_{\nu=1}^d v_\nu^A \left. \frac{\partial}{\partial x_\nu^A} \right|_p = \sum_{\nu=1}^d v_\nu^A \delta_{\mu\nu} = v_\mu^A \quad (\text{C.47})$$

This implies that the coefficients transform contravariantly, just as the cotangent coordinate basis:

$$v_\mu^B = \hat{dx}_\mu^B|_p(v) = \sum_{\nu=1}^d \left. \frac{\partial x_\mu^B}{\partial x_\nu^A} \right|_{x^A(p)} \hat{dx}_\nu^A|_p(v) = \sum_{\nu=1}^d \left. \frac{\partial x_\mu^B}{\partial x_\nu^A} \right|_{x^A(p)} v_\nu^A \quad (\text{C.48})$$

It is easily asserted that this transformation law does indeed lead to a coordinate independent representation of coordinate free vectors $v \in T_p M$:

$$\begin{aligned}\sum_\mu \left. \frac{\partial}{\partial x_\mu^B} \right|_p v_\mu^B &= \sum_{\mu,\nu,\rho} \left. \frac{\partial}{\partial x_\nu^A} \right|_p \left. \frac{\partial x_\nu^A}{\partial x_\mu^B} \right|_{x^B(p)} \left. \frac{\partial x_\rho^B}{\partial x_\nu^A} \right|_{x^A(p)} v_\rho^A = \sum_{\nu,\rho} \left. \frac{\partial}{\partial x_\nu^A} \right|_p \delta_{\nu\rho} v_\rho^A \\ &= \sum_\nu \left. \frac{\partial}{\partial x_\nu^A} \right|_p v_\nu^A\end{aligned}\quad (\text{C.49})$$

C.4 Coordinate bases as local bundle trivializations

The chart transition map induced transformation laws in Appendix C.3.2 coincide the gauge transformations as formulated in Chapters 7 and 11 when identifying the Jacobians $\left. \frac{\partial x^B}{\partial x^A} \right|_{x^A(p)}$ with g_p^{BA} . In Appendix C.4.1 we make these connections precise by listing all correspondences. Appendix C.4.2 extends these results by deriving expressions for chart induced bundle trivializations on extended domains $U \subseteq M$ as introduced in Chapter 11. A dictionary which summarizes the correspondences is given in Table C.1.

C.4.1 Correspondences to pointwise trivializations of $T_p M$

Gauges and chart differentials: The bundle formalism relies on the definition of gauges (Eq. (7.1))

$$\psi_{TM,p}^A : T_p M \rightarrow \mathbb{R}^d, \quad (\text{C.50})$$

which are vector bundle isomorphisms, assigning coordinates to tangent spaces with $p \in U^A$. In the chart formalism, gauges over U^A are *induced* as chart differentials (Eq. (C.30)):

$$\hat{dx}_p^A : T_p M \rightarrow \mathbb{R}^d \quad (\text{C.51})$$

Different gauges are related by gauge transformations (Eq. (7.7))

$$\psi_{TM,p}^B = g_p^{BA} \psi_{TM,p}^A \quad \text{with} \quad g_p^{BA} := \psi_{TM,p}^B \circ (\psi_{TM,p}^A)^{-1} \in G. \quad (\text{C.52})$$

The same definition holds for the chart induced gauges, where gauge transformations turn out to coincide with the Jacobian of the chart transition maps (Eq. (C.45)):

$$\hat{dx}_p^B = \frac{\partial x^B}{\partial x^A} \Big|_{x^A(p)} \hat{dx}_p^A \quad \text{with} \quad \frac{\partial x^B}{\partial x^A} \Big|_{x^A(p)} = \hat{dx}_p^B \circ (\hat{dx}_p^A)^{-1} \in \text{GL}(d) \quad (\text{C.53})$$

Vector components: As vector components $v^A = \psi_{TM,p}^A(v)$ or $v^A = \hat{dx}_p^A|_p(v)$ are given by the action of gauges, they show the same covariant transformation behavior

$$v^B = g_p^{BA} v^A \quad \text{and} \quad v^B = \frac{\partial x^B}{\partial x^A} \Big|_{x^A(p)} v^A. \quad (\text{C.54})$$

In terms of components, these relations are written as

$$v_i^B = \sum_{j=1}^d (g_p^{BA})_{ij} v_j^A \quad \text{and} \quad v_\mu^B = \sum_{\nu=1}^d \frac{\partial x_\mu^B}{\partial x_\nu^A} \Big|_{x^A(p)} v_\nu^A. \quad (\text{C.55})$$

Induced reference frames: Reference frames are in the bundle formalism induced by mapping the vectors ϵ_i of the standard frame $e \in G$ of \mathbb{R}^d through the gauge map back to $T_p M$ (Eq. (7.4)).

$$[e_i^A]_{i=1}^d = \left[(\psi_{TM,p}^A)^{-1}(\epsilon_i) \right]_{i=1}^d \quad (\text{C.56})$$

The corresponding relation in the chart formalism is according to Eq. (C.32) given by

$$\left[\frac{\partial}{\partial x_\mu^A} \Big|_p \right]_{\mu=1}^d = \left[(\hat{dx}_p^A)^{-1}(\epsilon_\mu) \right]_{\mu=1}^d \quad (\text{C.57})$$

Eq. (7.10) shows that the transformation laws of reference frames is given by the right action

$$\begin{aligned} [e_i^B]_{i=1}^d &= [e_i^A]_{i=1}^d \lhd (g_p^{BA})^{-1} := \left[\sum_{j=1}^d e_j^A (g_p^{BA})_{ji}^{-1} \right]_{i=1}^d \\ &= \left[\sum_{j=1}^d e_j^A (g_p^{AB})_{ji} \right]_{i=1}^d. \end{aligned} \quad (\text{C.58})$$

In analogy, the transformation law of coordinate bases is from Eq. (C.43) seen to be given by

$$\begin{aligned} \left[\frac{\partial}{\partial x_\mu^B} \Big|_p \right]_{\mu=1}^d &= \left[\frac{\partial}{\partial x_\mu^A} \Big|_p \right]_{\mu=1}^d \triangleleft \left. \frac{\partial x^B}{\partial x^A} \right|_{x^A(p)}^{-1} = \left[\sum_{\nu=1}^d \left. \frac{\partial}{\partial x_\nu^A} \right|_p \left(\left. \frac{\partial x^B}{\partial x^A} \right|_{x^A(p)} \right)^{-1} \right]_{\mu=1}^d \\ &= \left[\sum_{\nu=1}^d \left. \frac{\partial}{\partial x_\nu^A} \right|_p \left. \frac{\partial x_\nu^A}{\partial x_\mu^B} \right|_{x^B(p)} \right]_{\mu=1}^d \end{aligned} \quad (\text{C.59})$$

C.4.2 Chart induced local trivializations of $\pi_{TM}^{-1}(U)$

The correspondences laid out in the last section were relating *pointwise* trivializations $\psi_{TM,p}$ of $T_p M$ to chart differentials \hat{dx}_p . In order to complete this picture, this section adds expressions for local trivializations

$$\Psi_{TM} : \pi_{TM}^{-1}(U) \rightarrow U \times \mathbb{R}^d \quad (\text{C.60})$$

which are induced by charts.

A good candidate to construct Ψ_{TM} from is the chart differential

$$dx : \pi_{TM}^{-1}(U) \rightarrow TV \quad (\text{C.61})$$

which is a vector bundle isomorphism that differs from the vector space isomorphisms dx_p by not being restricted to a single point $p \in U$. To proceed, we generalize the canonical isomorphism $\iota_{\mathbb{R}^d}$ in Eq. (C.20) from a single point to all the tangent spaces $T_x V \cong \mathbb{R}^d$ over $V \subseteq \mathbb{R}^d$, resulting in the following *canonical local trivialization* of TV :

$$\iota_{TV} : TV \rightarrow V \times \mathbb{R}^d, \quad v \mapsto (\pi_{TV}(v), \iota_{\mathbb{R}^d}(v)). \quad (\text{C.62})$$

This allows to generalize \hat{dx}_p from a single point to a map

$$\hat{dx} := \iota_{V \times \mathbb{R}^d} \circ dx : \pi_{TM}^{-1}(U) \rightarrow V \times \mathbb{R}^d, \quad (\text{C.63})$$

which is, however, still not the local trivialization sought for. By mapping the first factor via the inverse chart from V to U , we obtain the *chart induced local bundle trivialization*:

$$\Psi_{TM} := (x^{-1} \times \text{id}) \circ \hat{dx} \quad (\text{C.64})$$

As usual, we visualize the definitions made in a commutative diagram:

$$\begin{array}{ccccccc} & & (x^{-1} \times \text{id}) & & & & \\ & & \downarrow & & & & \downarrow \\ & & \hat{dx} & & & & \\ & & \curvearrowleft \iota_{V \times \mathbb{R}^d} & \longleftarrow dx & \longrightarrow \pi_{TM}^{-1}(U) & \xrightarrow{\Psi_{TM}} & U \times \mathbb{R}^d \\ & & \searrow \text{proj}_1 & \downarrow \pi_{TV} & \downarrow \pi_{TM} & \swarrow \text{proj}_1 & \\ V \times \mathbb{R}^d & \leftarrow & TV & \leftarrow & U & \leftarrow & x \\ & & & & & & \end{array} \quad (\text{C.65})$$

Considering two overlapping charts $x^A : U^A \rightarrow V^A$ and $x^B : U^B \rightarrow V^B$ and denoting $U^{AB} = U^A \cap U^B$, one obtains transition maps

$$\hat{dx}^B \circ (\hat{dx}^A)^{-1} = \left(x^B \circ (x^A)^{-1} \times \frac{\partial x^B}{\partial x^A} \right) : x^A(U^{AB}) \times \mathbb{R}^d \rightarrow x^A(U^{AB}) \times \mathbb{R}^d \quad (\text{C.66})$$

and

$$\Psi_{TM}^B \circ (\Psi_{TM}^A)^{-1} = \left(\text{id} \times \frac{\partial x^B}{\partial x^A} \right) : U^{AB} \times \mathbb{R}^d \rightarrow U^{AB} \times \mathbb{R}^d. \quad (\text{C.67})$$

These definitions and their mutual relation is shown in the following commutative diagram:

$$\begin{array}{ccc} x^B(U^{AB}) \times \mathbb{R}^d & \xrightarrow{((x^B)^{-1} \times \text{id})} & U^{AB} \times \mathbb{R}^d \\ \uparrow \scriptstyle \left(x^B \circ (x^A)^{-1} \times \frac{\partial x^B}{\partial x^A} \right) & \nearrow \scriptstyle \hat{dx}^B & \downarrow \scriptstyle \Psi_{TM}^B \\ \pi_{TM}^{-1}(U^{AB}) & & \\ \uparrow \scriptstyle \hat{dx}^A & \nearrow \scriptstyle \Psi_{TM}^A & \downarrow \scriptstyle \left(\text{id} \times g^{BA} \right) \\ x^A(U^{AB}) \times \mathbb{R}^d & \xrightarrow{((x^A)^{-1} \times \text{id})} & U^{AB} \times \mathbb{R}^d \end{array} \quad (\text{C.68})$$

C.5 G -structures and vielbein fields

As discussed in Sections 11.3 and 11.4, any G -atlas $\{(\Psi_{TM}^X, U^X)\}$ of local tangent bundle trivializations specifies a corresponding G -structure, that is, a subbundle GM of distinguished reference frames which respect (or define) some geometric structure on M . By definition, the transition maps g^{BA} of associated G -bundles take values in a reduced structure group $G \leq \text{GL}(d)$. This raises the question whether one can similarly find “ G -atlases of charts” $\{(x^X, U^X)\}$, whose Jacobians $\frac{\partial x^B}{\partial x^A}$ take values in a reduced structure group $G \leq \text{GL}(d)$ and therefore encode a G -structure. For some structure groups this is certainly possible; for instance, an orientation of an orientable manifold can always be fixed by specifying some $\text{GL}^+(d)$ -atlas of positively oriented charts, whose transition Jacobians take values in $\text{GL}^+(d)$. In general, it is, however, impossible to find coordinate charts which induce coordinate bases that lie in a given G -structure. One therefore resorts to *explicit gauge transformation from coordinate bases into the G -structure*, known as *vielbein fields* [342, 357, 221, 32]. After initially transforming from coordinate bases to the G -structure, the gauge freedom within the G -structure allows for further G -valued gauge transformations.

An important example in physics are $O(d)$ -structures (or $O(1, d - 1)$ -structures for space-times), which consist of orthonormal reference frames relative to the (pseudo) Riemannian metric η of M .³ Such orthonormal frames represent the possible laboratory frames of an

³The symbol η is in the physics literature commonly preserved for the Minkowski metric $\text{diag}(+1, -1, \dots, -1)$ while the (pseudo) Riemannian metric of M is denoted by g . In contrast, we are writing group elements in the structure group as $g \in G$ and thus use η for the (pseudo) Riemannian metric of M .

inertial observer. They are for instance used to formulate relativistic quantum field theories, specifically the Dirac equation, in curved spacetimes. Recall that a given G -structure is to be respected by local bundle trivializations, which means that the gauge maps $\psi_{CM,p}$ need to map the G -structure $G_p M$ at $p \in M$ to the canonical standard G -structure G of \mathbb{R}^d . For the specific case of $O(d)$ -structures this is equivalent to the requirement on bundle trivializations to preserve the metric, i.e. $\eta_p(v, w) = \langle \psi_{TM,p}(v), \psi_{TM,p}(w) \rangle$ for any $p \in M$ and $v, w \in T_p M$, which is accomplished without problems in the bundle formalism. Given a coordinate chart $x : U \rightarrow V$, the induced gauges on $p \in U$ were in the previous sections shown to be given by $\psi_{TM,p} = \hat{dx}_p : T_p M \rightarrow \mathbb{R}^d$. The requirement on them to preserve the metric therefore becomes

$$\eta_p(v, w) = \langle \hat{dx}_p(v), \hat{dx}_p(w) \rangle, \quad (\text{C.69})$$

which is exactly the defining property for x being an *isometry*. This result implies that *coordinate bases only define an $O(d)$ -structures if U and V are isometric* – which is only the case if M is locally flat on U . For any non-flat region of M it is therefore impossible to describe an $O(d)$ -structures via coordinate bases directly. This incompatibility expresses itself for instance in the fact that the components $\eta_{\mu\nu}$ of the Riemannian metric on M relative to the chosen coordinate basis differ from $\delta_{\mu\nu}$ (or $\text{diag}(+1, -1, \dots, -1)_{\mu\nu}$).

As mentioned before, the orthonormal frames of an $O(d)$ -structure OM are in the physics literature typically defined via a gauge transformation relative to some chart induced frame field $\left[\frac{\partial}{\partial x_\mu} \right]_{\mu=1}^d$. Denoting this gauge transformation, which is called *vielbein field*, by

$$\epsilon^A : U \rightarrow \text{GL}(d), \quad (\text{C.70})$$

the orthonormal frame field is defined by⁴

$$\left[e_i^A \right]_{i=1}^d := \left[\frac{\partial}{\partial x_\mu} \right]_{i=1}^d \lhd (\epsilon^A)^{-1} = \left[\sum_\mu \frac{\partial}{\partial x_\mu} (\epsilon^A)^{-1}_{\mu i} \right]_{i=1}^d \in \Gamma(U, OM). \quad (\text{C.71})$$

The orthonormality of the resulting frame field is usually expressed as⁵

$$\begin{aligned} \delta_{ij} &= \eta(e_i^A, e_j^A) \\ &= \eta\left(\sum_\mu \frac{\partial}{\partial x_\mu} (\epsilon^A)^{-1}_{\mu i}, \sum_\nu \frac{\partial}{\partial x_\nu} (\epsilon^A)^{-1}_{\nu j}\right) \\ &= \sum_{\mu\nu} \eta\left(\frac{\partial}{\partial x_\mu}, \frac{\partial}{\partial x_\nu}\right) (\epsilon^A)^{-1}_{\mu i} (\epsilon^A)^{-1}_{\nu j} \\ &= \sum_{\mu\nu} \eta_{\mu\nu} (\epsilon^A)^{-1}_{\mu i} (\epsilon^A)^{-1}_{\nu j}, \end{aligned} \quad (\text{C.72})$$

which explains why the vielbein field is sometimes called “square root of the metric”. As usual, vector components are translated via the non-inverted gauge transformation, that is:⁶

$$v_i^A = \sum_\mu \epsilon_{i\mu}^A v_\mu \quad (\text{C.73})$$

⁴In the physics literature this relation is expressed as $e_i^A = (\epsilon^A)_i^\mu \frac{\partial}{\partial x^\mu}$. The inverse is here merely signaled by the opposite position of the indices $(\epsilon^A)_i^\mu := (\epsilon^A)^{-1}_{\mu i}$ in comparison to $(\epsilon^A)_\mu^i := \epsilon_{\mu i}^A$.

⁵In the physics literature this relation is usually written $\eta_{\mu\nu} (\epsilon^A)_i^\mu (\epsilon^A)_j^\nu = \delta_{ij}$.

⁶Again, in the usual notation in physics this relation reads $(v^A)_i = (\epsilon^A)_\mu^i v^\mu$.

A simple dimension counting argument illustrates the gauge freedom in the $O(d)$ -structure:⁷ Being an element of the general linear group, a vielbein $e^A(p) \in GL(d)$ has d^2 degrees of freedom, while the metric η , as a symmetric, bilinear form, has $d(d+1)/2$ degrees of freedom. The missing $d(d-1)/2$ degrees of freedom correspond exactly to gauge transformations by structure group elements $g^{BA} \in O(d)$. Alternatively, from the viewpoint of G -structures, $F_p M \cong GL(d)$ has d^2 degrees of freedom while $O_p M \cong O(d)$ has $d(d-1)/2$ degrees of freedom, fixing $d(d+1)/2$ degrees of freedom which correspond to the choice of metric.

All constructions are obviously generalized to arbitrary G -structures with $GL(d)$ -valued vielbein fields mapping coordinate bases into GM and the freedom to apply G -valued gauge transformation afterwards.

⁷In physics, one rather considers local Lorentz transformations $\Lambda \in O(1, 3)$, which describe rotations and boosts of local reference frames.

	isomorphism	bundle formalism	chart formalism
chart	$x^A : U^A \xrightarrow{\sim} V^A$	—	any diffeomorphism
transition map	$x^B \circ (x^A)^{-1} : x^B(U^{AB}) \xrightarrow{\sim} x^A(U^{AB})$	—	implied by charts
pointwise trivialization	$\psi_{TM,p}^A : T_p M \xrightarrow{\sim} \mathbb{R}^d$	linear isomorphism from G -atlas	$\hat{dx}_p^A = (\hat{dx}_1^A _p, \dots, \hat{dx}_d^A _p)^\top$
transition map	$\psi_{TM,p}^B \circ (\psi_{TM,p}^A)^{-1} : \mathbb{R}^d \xrightarrow{\sim} \mathbb{R}^d$	structure group element $g_p^{BA} \in G$	$\hat{dx}_p^B \circ (\hat{dx}_p^A)^{-1} = \left. \frac{\partial x^B}{\partial x^A} \right _{x^A(p)}$
local trivialization	$\Psi_{TM}^A : \pi_{TM}^{-1}(U^A) \xrightarrow{\sim} U^A \times \mathbb{R}^d$	$v \mapsto (\pi_{TM}(v), \psi_{TM, \pi_{TM}(v)}(v))$	$((x^A)^{-1} \times \text{id}) \circ \hat{dx}^A$
transition map	$\Psi_{TM}^B \circ (\Psi_{TM}^A)^{-1} : U^{AB} \times \mathbb{R}^d \xrightarrow{\sim} U^{AB} \times \mathbb{R}^d$	$(\text{id} \times g^{BA})$	$\left(\text{id} \times \frac{\partial x^B}{\partial x^A} \right)$
general frame	$[e_i^A]_{i=1}^d \in F_p M$	$\left[(\psi_{TM,p}^A)^{-1}(\epsilon_i) \right]_{i=1}^d$ from $\text{GL}(d)$ -atlas	$\left[\frac{\partial}{\partial x_\mu^A} \Big _p \right]_{\mu=1}^d = \left[(\hat{dx}_p^A)^{-1}(\epsilon_i) \right]_{i=1}^d$
G -structure frame	$[e_i^A]_{i=1}^d \in G_p M$	$\left[(\psi_{TM,p}^A)^{-1}(\epsilon_i) \right]_{i=1}^d$ from G -atlas	$\left[\sum_\mu \frac{\partial}{\partial x_\mu^A} \Big _p (\epsilon^A)_{\mu i}^{-1} \right]_{i=1}^d$

Table C.1: An overview of different types of coordinatizations on manifolds. The bundle formalism (3rd column), which is used in this work, directly assigns coordinates to the tangent spaces, while referring to the points p of the base space M in a coordinate free fashion. In contrast, the chart formalism (4th column) assigns coordinates to local subsets $U^X \subseteq M$ of the manifold. Local trivializations of the tangent bundle and bundle transition maps between them are induced as differentials of the charts and their transition maps, the latter usually referred to as Jacobians. The second last row gives expressions for the reference frames which are induced as identity sections of local trivializations of TM (3rd column) or as chart induced coordinate bases (4th column). Similarly, the last row compares definitions of G -structures – for instance orthonormal frames – via an G -atlas for TM (3rd column) and via vielbein fields as gauge transformations relative to coordinate bases (4th column). As usual, we abbreviate $U^A \cap U^B$ by U^{AB} and assume $p \in U^{AB}$.

Integration over tangent spaces

On a Riemannian *manifold* (M, η) the volume density¹ dp on M is uniquely specified by demanding that *orthonormal frames* $[e_1^O, \dots, e_d^O]$ with respect to the metric η are assigned *unit volume*:

$$dp(e_1^O, \dots, e_d^O) = 1 \quad \text{for any orthonormal frame } [e_1^O, \dots, e_d^O] \text{ of } T_p M \quad (\text{D.1})$$

Similarly, a volume density dv on the *tangent spaces* $T_p M$ of a Riemannian manifold is uniquely defined by assigning unit volume to its orthonormal frames w.r.t. η_p :

$$dv(\epsilon_1^O, \dots, \epsilon_d^O) = 1 \quad \text{for any orthonormal frame } [\epsilon_1^O, \dots, \epsilon_d^O] \text{ of } T_v T_p M \quad (\text{D.2})$$

To avoid an unnecessarily complicated discussion of the double tangent bundle TTM , we define the integration over $T_p M$ equivalently by pulling it via some *isometric* (and thus volume preserving) gauge back to \mathbb{R}^d . Let $\psi_{TM,p}^O$ be such an isometric gauge from an $O(d)$ -atlas, which identifies orthonormal frames in $T_p M$ with orthonormal frames in \mathbb{R}^d . The integral of a function $f : T_p M \rightarrow \mathbb{R}$ is then defined via its pullback

$$\begin{aligned} \int_{T_p M} f(v) dv &:= \int_{\mathbb{R}^d} f \circ (\psi_{TM,p}^O)^{-1}(v^O) dv^O \\ &= \int_{\mathbb{R}^d} f^O(v^O) dv^O, \end{aligned} \quad (\text{D.3})$$

where we defined the coordinate expression $f^O := f \circ (\psi_{TM,p}^O)^{-1} : \mathbb{R}^d \rightarrow \mathbb{R}$ of f as usual. The fact that $\psi_{TM,p}^O$ is isometric ensures hereby that dv does indeed assign unit volume to orthonormal frames if dv^O does. Since the latter is just the standard Lebesgue measure on \mathbb{R}^d , this is the case.

Let now $\psi_{TM,p}^A$ be *any* gauge at p , relative to which one might want to express the integration. The transition map between both coordinatizations is simply given by the gauge transformation $v^O = \psi^O \circ (\psi^A)^{-1}(v^A) = g_p^{OA}(v^A)$. By the standard rules for changes of variables in multidimensional integrals, the differentials are required to transform according to the Jacobian determinant of this transformation in order for the volume to be preserved. As the

¹In contrast to a volume *form* ω , volume *densities* $|\omega|$ assign a positive volume to any frame. They exist both on oriented and non-oriented manifolds.

transformation is liner, the Jacobian is given by g_p^{OA} itself, such that we obtain

$$\int_{T_p M} f(v) dv = \int_{\mathbb{R}^d} f^A(v^A) |\det(g_p^{OA})| dv^A. \quad (\text{D.4})$$

Through the gauge transformation, this expression still depends on the arbitrary choice of isometric gauge $\psi_{TM,p}^O$. This dependency can be purged by expressing the integration measure directly in terms of the metric as

$$\int_{T_p M} f(v) dv = \int_{\mathbb{R}^d} f^A(v^A) \sqrt{|\eta_p^A|} dv^A, \quad (\text{D.5})$$

where the factor

$$\sqrt{|\eta_p^A|} := \sqrt{|\det([\eta_p(e_i^A, e_j^A)]_{ij})|} \quad (\text{D.6})$$

measures the (absolute) volume of the reference frame $[e_i^A]_{i=1}^d$ relative to the metric η . To assert the equality of the right-hand sides of Eqs. (D.4) and (D.5), we express the metric η_p of $T_p M$ in terms of the standard inner product $\langle \cdot, \cdot \rangle$ of \mathbb{R}^d , which is once again done by using the isometric gauge $\psi_{TM,p}^O$ from the $O(d)$ -atlas:

$$\begin{aligned} \eta_p(e_i^A, e_j^A) &= \langle \psi_{TM,p}^O(e_i^A), \psi_{TM,p}^O(e_j^A) \rangle \\ &= \langle \psi_{TM,p}^O \circ (\psi_{TM,p}^A)^{-1}(\epsilon_i), \psi_{TM,p}^O \circ (\psi_{TM,p}^A)^{-1}(\epsilon_j) \rangle \\ &= \langle g_p^{OA} \epsilon_i, g_p^{OA} \epsilon_j \rangle \\ &= \epsilon_i^\top (g_p^{OA})^\top g_p^{OA} \epsilon_j \\ &= \left((g_p^{OA})^\top g_p^{OA} \right)_{ij} \end{aligned} \quad (\text{D.7})$$

The absolute value of the determinant in Eq. (D.6) is therefore given by

$$\begin{aligned} |\det([\eta_p(e_i^A, e_j^A)]_{ij})| &= |\det((g_p^{OA})^\top g_p^{OA})| \\ &= |\det((g_p^{OA})^\top) \det(g_p^{OA})| \\ &= |\det(g_p^{OA})|^2, \end{aligned} \quad (\text{D.8})$$

from which the equality of the right-hand-sides of Eqs. (D.4) and (D.5) follows by taking the square root.

Since the factors $\sqrt{|\eta_p^A|}$ and $\sqrt{|\eta_p^B|}$ measure the volumes of their respective frames, one can easily show that they are related by the *inverse* change of volume $|\det g_p^{BA}|$:

$$\sqrt{|\eta_p^B|} = \frac{1}{|\det g_p^{BA}|} \sqrt{|\eta_p^A|} \quad (\Rightarrow -1\text{-density}) \quad (\text{D.9})$$

Together with the usual change of variables formula

$$dv^B = |\det g_p^{BA}| dv^A \quad (\Rightarrow +1\text{-density}), \quad (\text{D.10})$$

this implies that the coordinatizations of the Riemannian volume element dv are by design invariant under gauge transformations, that is,

$$\sqrt{|\eta_p^B|} dv^B = \sqrt{|\eta_p^A|} dv^A \quad (\Rightarrow 0\text{-density}). \quad (\text{D.11})$$

This relation assures that the integration in Eq. (D.5) is well defined, i.e. coordinate independent.

Equivariant MLPs

Multilayer perceptrons (MLPs) are the most basic neural network architectures. Their feature spaces are finite dimensional vector spaces \mathbb{K}^c over the field $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$. Basic MLPs are constructed as a sequence of blocks of layers, where each block

$$\text{block} : \mathbb{K}^{c_{\text{in}}} \rightarrow \mathbb{K}^{c_{\text{out}}}, \quad x_{\text{in}} \mapsto x_{\text{out}} := \sigma(Wx_{\text{in}} + b) \quad (\text{E.1})$$

consists of 1) a linear map (matrix multiplication) $W \in \mathbb{K}^{c_{\text{out}} \times c_{\text{in}}}$, 2) a bias summation layer where $b \in \mathbb{K}^{c_{\text{out}}}$ and 3) a nonlinearity $\sigma : \mathbb{K}^{c_{\text{out}}} \rightarrow \mathbb{K}^{c_{\text{out}}}$. Equivariant MLPs are consequently constructed from equivariant linear layers (intertwiners), equivariant bias summation operations and equivariant nonlinearities. As the latter two were already discussed in Sections 4.3.2 and 4.3.3, we focus here on equivariant matrix multiplications.¹ We assume that the feature spaces are finite dimensional unitary group representations (Def. B.5.13), which allows their complete reducibility into a direct sum of irreducible subspaces; see Theorem B.5.16.

Consider a linear layer $W : \mathbb{K}^{c_{\text{in}}} \rightarrow \mathbb{K}^{c_{\text{out}}}$, mapping between unitary representation spaces $(\rho_{\text{in}}, \mathbb{K}^{c_{\text{in}}})$ and $(\rho_{\text{out}}, \mathbb{K}^{c_{\text{out}}})$. We are interested in the subspace $\text{Hom}_G(\mathbb{K}^{c_{\text{in}}}, \mathbb{K}^{c_{\text{out}}}) \subseteq \mathbb{K}^{c_{\text{out}} \times c_{\text{in}}}$ of linear layers that are equivariant (intertwiners, Def. B.5.7), that is, those layers that satisfy the constraint

$$\rho_{\text{out}}(g) W = W \rho_{\text{in}}(g) \quad \forall g \in G. \quad (\text{E.2})$$

To break this constraint down, let $Q_{\text{in}} \in \text{GL}(\mathbb{K}^{c_{\text{in}}})$ and $Q_{\text{out}} \in \text{GL}(\mathbb{K}^{c_{\text{out}}})$ be the change of basis matrices which decompose the feature spaces into an orthogonal direct sum of irreducible subspaces, acted on by irreducible subrepresentations (irreps, Def. B.5.6). In equations, these irrep decompositions are defined by the relations

$$Q_{\text{in}} \rho_{\text{in}}(g) Q_{\text{in}}^{-1} = \bigoplus_{j \in \widehat{G}} \bigoplus_{i=1}^{m_j} \rho_j(g) \quad \text{and} \quad Q_{\text{out}} \rho_{\text{out}}(g) Q_{\text{out}}^{-1} = \bigoplus_{J \in \widehat{G}} \bigoplus_{I=1}^{m_J} \rho_J(g) \quad (\text{E.3})$$

for any $g \in G$, where \widehat{G} is the set of all *isomorphism classes* of irreps of G and $m_j, m_J \in \mathbb{N}$ are the unique (mostly zero) multiplicities of the corresponding irreps ρ_j and ρ_J in ρ_{in} and ρ_{out} , respectively. Introducing the linear map

$$\widetilde{W} := Q_{\text{out}} W Q_{\text{in}}^{-1} \quad (\text{E.4})$$

¹The results in these sections applied to feature fields. However, since the considered biases and nonlinearities were applied *pointwise*, i.e. individually to each feature vector, the results are exactly equivalent to the case of MLPs.

between the decoupled feature spaces, the intertwiner constraint on W in Eq. (E.2) can be rewritten as the equivalent constraint

$$\widetilde{W} = \left(\bigoplus_{J \in \widehat{G}} \bigoplus_{I=1}^{m_J} \rho_J(g) \right) \widetilde{W} \left(\bigoplus_{j \in \widehat{G}} \bigoplus_{i=1}^{m_j} \rho_j(g)^{-1} \right) \quad \forall g \in G. \quad (\text{E.5})$$

on \widetilde{W} . Due to the direct sum decomposition into G -independent subspaces, this constraint is equivalent to $(\sum_{J \in \widehat{G}} m_J) \cdot (\sum_{j \in \widehat{G}} m_j)$ independent *irrep constraints*

$$\widetilde{W}_{JI,ji} = \rho_J(g) \widetilde{W}_{JI,ji} \rho_j(g)^{-1} \quad \forall g \in G \quad (\text{E.6})$$

on blocks $\widetilde{W}_{JI,ji} \in \mathbb{K}^{\dim(\rho_J) \times \dim(\rho_j)}$ of \widetilde{W} which map between all pairs of invariant subspaces. The space of such irrep intertwiners is in the following denoted as $\text{Hom}_G(\rho_j, \rho_J)$.

To solve the irrep constraints in Eq. (E.6), recall *Schur's lemma* B.5.10, which states in particular that the irrep intertwiners $\widetilde{W}_{JI,ji}$ are zero for non-isomorphic irreps $J \neq j$. The only possibly non-zero components of \widetilde{W} are therefore the *endomorphisms*

$$\widetilde{W}_{JI,ji} \in \text{End}_G(\rho_j) = \text{Hom}_G(\rho_j, \rho_j) \quad (\text{E.7})$$

which map between isomorphic irreducible subrepresentations $j = J$.

Assume that we are given bases $\{\mathcal{E}_{j,\mu} \mid \mu = 1, \dots, \dim(\text{End}_G(\rho_j))\}$ of the endomorphism (vector) spaces $\text{End}_G(\rho_j)$. The blocks $\widetilde{W}_{JI,ji}$ may then be parameterized as

$$\widetilde{W}_{JI,ji} = \begin{cases} 0 & \text{if } J \neq j \\ \sum_{\mu} \lambda_{jIi,\mu} \mathcal{E}_{j,\mu} & \text{if } J = j, \end{cases} \quad (\text{E.8})$$

where $\lambda_{jIi,\mu} \in \mathbb{K}$ are $\dim(\text{End}_G(\rho_j))$ learnable parameters (for fixed j, J, i, I). For complex numbers $\mathbb{K} = \mathbb{C}$, the endomorphism spaces are one-dimensional and contain elements $\lambda_{jIi} \cdot \text{id}_{\mathbb{C}^{\dim(\rho_j)}}$ which are complex multiples of the identity. For $\mathbb{K} = \mathbb{R}$ one has $\dim(\text{End}_G(\rho_j)) = 1, 2$ or 4 when the real irrep is of real, complex or quaternionic type, respectively [173].

With these results we have all ingredients that are necessary to construct the most general linear equivariant network layer $\widetilde{W} \in \text{Hom}_G(\rho_{\text{in}}, \rho_{\text{out}})$. All that is required is to

- 1) parameterize the individual matrix blocks $\widetilde{W}_{JI,ji}$ according to Eq. (E.8),
- 2) fill them into \widetilde{W} and
- 3) undo the irrep decomposition in Eq. (E.4), i.e. set $W = Q_{\text{out}}^{-1} \widetilde{W} Q_{\text{in}}$.

As many of the matrix blocks are filled with zeros and since the dimensionality of the endomorphism spaces is usually lower than the dimensionality of an unconstrained block, equivariant linear layers are more parameter efficient than their non-equivariant counterparts.

The representation theoretic viewpoint discussed here was already proposed in the early 90's by Wood and Shawe-Taylor [332]. Nowadays, its main advocates are probably Kondor et al. [163][161, 162, 163, 164, 134, 3, 18, 299]. Finzi et al. [92] developed a general algorithm to solve automatically for the intertwiners between finite representation spaces of arbitrary matrix groups. Their approach is based on the observation that it is sufficient to solve the intertwiner constraint for the *generators* of the group only, i.e. for Lie algebra elements and/or

finite generators. Most other works describe linear equivariant layers from a group theoretic instead of representation theoretic viewpoint. Their results are equivalent to ours, however, they might be expressed quite differently. Specifically for finite groups with permutation actions, the intertwining matrices are usually formulated in terms of *weight sharing patterns* [241, 200, 109], where weights are shared over *invariant subspaces* of the permutation action $W \mapsto \rho_{\text{out}}(g) W \rho_{\text{in}}(g)^{-1}$ [2]. The universality of equivariant MLPs was investigated in [240].

Linear layers between non-finite representations can be significantly harder to characterize. An example are the steerable convolutions between induced affine group representations from Chapters 4 and 5.

Equivariant convolutions on homogeneous spaces

The works by Kondor and Trivedi [162], Cohen et al. [55][56] and Bekkers [10] are in spirit quite similar to ours in that they are defining group equivariant convolutions in a fairly general setting. These papers have in common that they operate on *feature maps on homogeneous spaces* \mathcal{I}/H of a global symmetry group \mathcal{I} , where $H \leq \mathcal{I}$.¹ They differ in the types of groups \mathcal{I} which they cover and in the definition of their feature spaces, specifically the linear group actions on them. The main theorems of the papers assert that *the most general equivariant linear maps between such feature spaces are convolutions* (or correlations) with *symmetry constrained kernels*; cf. Theorems 3.2.1 and 4.3.1. The specific details on these generalized convolutions depend on the particular feature spaces and group actions which the models consider.

This appendix examines these theories and their relation to our coordinate independent convolutions from Parts II and III. The most important similarities and differences are summarized in the following list:

- Not any *homogeneous space* is a *Riemannian manifold* and not any Riemannian manifold is a homogeneous space of its isometry group.² There is, however, a significant overlap, for instance for Euclidean steerable CNNs on $\mathbb{E}_d \cong \text{Aff}(G)/G$ from Chapter 4 or spherical CNNs on $S^2 \cong \text{O}(3)/\text{O}(2)$ from Chapter 17.
- The authors consider *compact* [162], *locally compact*, *unimodular* [55][56], and *Lie* groups [10], respectively. The global symmetry groups in our theory are *isometries* of M or, specifically for Euclidean spaces, *affine* groups $\text{Aff}(G)$. Note that affine groups are not compact and only for $G \leq \text{O}(d)$ unimodular – general affine groups are therefore not covered in the respective theories.
- Coordinate independent CNNs shift the focus from *global* to *local symmetries*. On homogeneous spaces \mathcal{I}/H these local symmetries correspond to the stabilizer subgroups $\text{Stab}_p \cong H$ of \mathcal{I} . Our Chapter 13 works out the relations between global and local symmetries in detail – the models' local equivariance induces their global equivariance.
- The models assume different *types* of feature fields and *group actions* on them: Kondor and Trivedi [162] and Bekkers [10] assume scalar fields on homogeneous spaces,

¹[162, 56, 10] use G instead of \mathcal{I} to refer to global symmetries. We use \mathcal{I} since we reserve G for the structure group. Note furthermore that we use \mathcal{I} here to denote arbitrary global symmetries, not necessarily isometries as in Parts II and III.

²For instance, $\mathcal{I}/H = \text{O}(2)/\text{SO}(2) \cong \mathcal{R}$ is a finite group (or set) but not a Riemannian manifold. Another example are $(\mathbb{Z}^d, +)$ group convolutions on the discrete pixel grid \mathbb{Z}^d .

i.e. real-valued functions $f : \mathcal{I}/H \rightarrow \mathbb{R}$ which transform according to $\phi.f(\zeta.H) = f(\phi^{-1}\zeta.H)$.³ Cohen et al. [55][56] consider feature fields of more general types ρ which are defined as sections of H -associated feature vector bundles. Their transformation laws are given by induced representations $\text{Ind}_H^{\mathcal{I}} \rho$. This setting covers the real-valued functions from [162, 10] as a special case when choosing trivial field representations (or, as made precise below, more general quotient representations $\rho_{\text{quot}}^{G/H}$ (Def. B.5.20) where $H \leq G \leq \mathcal{I}$). Our theory models feature fields as sections of associated bundles as well. Their transformation is given by pushforwards $\phi \triangleright f := \phi_{*,A} \circ f \circ \phi^{-1}$, which generalize induced representations.

- The works by Kondor and Trivedi [162], Cohen et al. [55][56] and Bekkers [10] derive *convolutional weight sharing* from the requirement on the models to be globally equivariant; just as we did in Part I. Our *GM*-convolutions, on the other hand, share weights by definition over the G -structure. We adopted the idea of deriving weight sharing over Riemannian manifolds from global symmetries (isometries) in Section 13.3. The requirement for isometry equivariance implies weight sharing over the isometry orbits and a stabilizer constraint on the kernels; see e.g. Fig. 13.7. Theorem 13.3.3 asserts that *isometry equivariant kernel field transforms on homogeneous spaces are GM-convolutions* – this result mirrors those of Kondor and Trivedi [162], Cohen et al. [55][56] and Bekkers [10] closely.
- All of the theories derive some *linear symmetry constraint on the kernel spaces*. In the case of Kondor and Trivedi [162] and Bekkers [10], the kernels are essentially scalar functions on double quotient spaces $H_{\text{out}} \backslash \mathcal{I}/H_{\text{in}}$ (assuming correlations, for convolutions H_{in} and H_{out} are swapped; see below). The kernels of Cohen et al. [55][56] and in our theory are satisfying a steerability constraint which depends on the particular choice of field types ρ_{in} and ρ_{out} . Note that the determinant factor is missing in the G -steerability constraint of Cohen et al. [55][56] since the authors restrict to unimodular groups. The factor does appear in the kernel constraint by Bekkers [10].
- While Kondor and Trivedi [162] and Cohen et al. [55][56] describe kernels immediately on the group or homogeneous space, Bekkers [10] and our *GM*-convolutions define kernels on the tangent spaces and project them subsequently via the exponential map. These approaches are in general inequivalent, for instance since the exponential map is on a non-connected manifold non-injective. On Euclidean spaces both approaches are obviously equivalent since the exponential map becomes trivial; see Section 15.2. Our Theorem 17.2.1 in Section 17.2 bridges this gap furthermore for spherical kernels by providing an isomorphism between kernels of the two approaches. In practice, the general incompatibility is irrelevant since kernels of convolutional networks are usually compactly supported within the injectivity radius of the exponential map.

We will in the following elaborate on the theories of Kondor and Trivedi [162], Bekkers [10] and Cohen et al. [55][56] in more detail. As a preparation, we will first discuss homogeneous spaces, group convolutions and group correlations. For alternative reviews of the topic we refer the reader to Esteves [82] and Gerken et al. [105]. After the original version of this appendix was published in [325], Aronsson [5] published another review of the topic, including in particular a formulation in terms of reproducing kernel Hilbert spaces and Xu et al. [337] presented the Fourier space analogue to Cohen et al. [55][56]. We furthermore want to point to the work by Chakraborty et al. [41], which also defines convolutions on homogeneous spaces but is not covered in more detail in this appendix since their models

³Multi-channel feature maps are constructed by stacking multiple such functions.

assume $H_{\text{out}} = \{e\}$, that is, their convolution kernels are unconstrained and always lift the input signal to a scalar field on \mathcal{I} .

F.1 Homogeneous spaces, group convolutions and group correlations

Homogeneous spaces: Let \mathcal{I} be some group which acts on some space X . The space is said to be *homogeneous* if the group action is *transitive* (Def. B.3.8), i.e. if any two points $p, q \in X$ are related by the \mathcal{I} -action. In equations, X is homogeneous if and only if for any $p, q \in X$ there exists an element $\phi \in \mathcal{I}$ such that $q = \phi(p)$. Note that the action on X is not required to be fixed point free (Def. B.3.10), that is, each point $p \in X$ has a potentially non-trivial stabilizer subgroup $\text{Stab}_p = \{\xi \in \mathcal{I} \mid \xi(p) = p\} \leq \mathcal{I}$ (Def. B.3.6). It can be shown that the homogeneous space can be identified with the quotient space \mathcal{I}/H where $H = \text{Stab}_p$ for some $p \in X$.⁴

Since any homogeneous space arises as a quotient, we consider in the following always some subgroup H of \mathcal{I} . This subgroup has *left cosets* (Def. B.2.2), i.e. subsets of the form

$$\phi.H = \{\phi h \mid h \in H\} \quad (\text{F.1})$$

which are elements of the (homogeneous) quotient space

$$\mathcal{I}/H = \{\phi.H \mid \phi \in \mathcal{I}\}. \quad (\text{F.2})$$

A natural left action of \mathcal{I} on \mathcal{I}/H is given by

$$\mathcal{I} \times \mathcal{I}/H \rightarrow \mathcal{I}/H, \quad (\tilde{\phi}, \phi.H) \mapsto \tilde{\phi}\phi.H. \quad (\text{F.3})$$

This action is easily seen to be transitive, making \mathcal{I}/H a homogeneous space of \mathcal{I} . The canonical quotient map

$$q_{\mathcal{I}/H}^{\mathcal{I}} : \mathcal{I} \rightarrow \mathcal{I}/H, \quad \phi \mapsto \phi.H \quad (\text{F.4})$$

turns \mathcal{I} into a principal H -bundle over \mathcal{I}/H . Analogous definitions can be made for *right cosets*

$$H.\phi \in H \backslash \mathcal{I}. \quad (\text{F.5})$$

and *double cosets*

$$\tilde{H}.\phi.H \in \tilde{H} \backslash \mathcal{I}/H \quad (\text{F.6})$$

and their respective quotient spaces.

An universal property of the quotient maps $q_{\mathcal{I}/H}^{\mathcal{I}}$, which will become important in our discussion below, is the following. Let $f^{\uparrow} : \mathcal{I} \rightarrow \mathbb{R}$ be a continuous, right H -invariant function, i.e. a function which satisfies $f^{\uparrow}(\phi h) = f^{\uparrow}(\phi)$ for any $\phi \in \mathcal{I}$ and $h \in H$. Then there exists a unique continuous function $f : \mathcal{I}/H \rightarrow \mathbb{R}$ such that $f^{\uparrow} = f \circ q_{\mathcal{I}/H}^{\mathcal{I}}$. Conversely, one may lift any continuous map $f : \mathcal{I}/H \rightarrow \mathbb{R}$ uniquely to a right H -invariant map $f^{\uparrow} : \mathcal{I} \rightarrow \mathbb{R}$, which is used by Kondor and Trivedi [162] to generalize group convolutions to homogeneous

⁴Other choices of points yield other realizations of the non-canonical isomorphism $\mathcal{I}/H \cong X$. Any choice is equally valid since $\text{Stab}_p \cong \text{Stab}_q$ for homogeneous spaces.

spaces. The relation between both functions is visualized in the following commutative diagram:

$$\begin{array}{ccc} \mathcal{I} & & \\ q_{\mathcal{I}/H}^{\mathcal{I}} \downarrow & \searrow f^{\uparrow} := f \circ q_{\mathcal{I}/H}^{\mathcal{I}} & \\ \mathcal{I}/H & \xrightarrow{f} & \mathbb{R} \end{array} \quad (\text{F.7})$$

An analogous construction can obviously be made for right quotient spaces $H \backslash \mathcal{I}$ and left H -invariant maps. The following commutative diagram visualizes the case of double quotient spaces $\tilde{H} \backslash \mathcal{I}/H$ and maps f^{\uparrow} which are simultaneously left \tilde{H} -invariant and right H -invariant, i.e. which satisfy $f^{\uparrow}(\tilde{h}\phi h) = f^{\uparrow}(\phi)$ for any $\phi \in \mathcal{I}$, $\tilde{h} \in \tilde{H}$, and $h \in H$:

$$\begin{array}{ccc} \mathcal{I} & & \\ q_{\tilde{H} \backslash \mathcal{I}/H}^{\mathcal{I}} \downarrow & \searrow f^{\uparrow} := f \circ q_{\tilde{H} \backslash \mathcal{I}/H}^{\mathcal{I}} & \\ \tilde{H} \backslash \mathcal{I}/H & \xrightarrow{f} & \mathbb{R} \end{array} \quad (\text{F.8})$$

Group convolutions and group correlations: Convolutions are naturally generalized from Euclidean spaces (or translation groups) to arbitrary locally compact groups. Let \mathcal{I} be a locally compact group and let $d\zeta$ be a left Haar measure on \mathcal{I} . The *group convolution* $(f *_{\mathcal{I}} \kappa) : \mathcal{I} \rightarrow \mathbb{R}$ of two integrable functions $f : \mathcal{I} \rightarrow \mathbb{R}$ and $\kappa : \mathcal{I} \rightarrow \mathbb{R}$ is then defined by the following equivalent expressions, taken from [102]:

$$\begin{aligned} (f *_{\mathcal{I}} \kappa)(\phi) &:= \int_{\mathcal{I}} f(\zeta) \kappa(\zeta^{-1} \phi) d\zeta \\ &= \int_{\mathcal{I}} f(\phi \zeta) \kappa(\zeta^{-1}) d\zeta \\ &= \int_{\mathcal{I}} f(\zeta^{-1}) \kappa(\zeta \phi) \Delta(\zeta^{-1}) d\zeta \\ &= \int_{\mathcal{I}} f(\phi \zeta^{-1}) \kappa(\zeta) \Delta(\zeta^{-1}) d\zeta, \end{aligned} \quad (\text{F.9})$$

The group homomorphism $\Delta : \mathcal{I} \rightarrow (\mathbb{R}_{>0}, *)$, appearing in the last two expressions, is the modular function of \mathcal{I} . Kondor and Trivedi [162] define group convolutions as in the last line, however, without the modular function. This is valid since the authors assume compact groups, which are unimodular, i.e. satisfy $\Delta(\phi) = 1$ for any $\phi \in \mathcal{I}$.

Closely related to group convolutions are *group correlations*

$$(f \star_{\mathcal{I}} \kappa)(\phi) := \langle f, \phi \cdot \kappa \rangle_{L^1(\mathcal{I})} = \int_{\mathcal{I}} f(\zeta) \kappa(\phi^{-1} \zeta) d\zeta, \quad (\text{F.10})$$

which are defined as the inner product of a function f with a shifted kernel $\phi \cdot \kappa$. A comparison with Eq. (F.9) reveals that group convolutions and group correlations are equivalent up to an inversion of the kernel argument, that is,

$$(f \star_{\mathcal{I}} \kappa) = (f *_{\mathcal{I}} [\kappa \circ (\cdot)^{-1}]). \quad (\text{F.11})$$

While Kondor and Trivedi [162] consider (generalized) group convolutions, Bekkers [10] and Cohen et al. [56] assume correlations – to reconcile the theories one has to invert the kernel arguments.

Group convolutions and group correlations are by definition *equivariant* w.r.t. left actions $\alpha.f(\phi) = f(\alpha^{-1}\phi)$ of group elements $\alpha \in \mathcal{I}$ on the first factor. For the case of convolutions, this is shown by

$$\begin{aligned} ([\alpha.f] *_{\mathcal{I}} \kappa)(\phi) &= \int_{\mathcal{I}} [\alpha.f](\zeta) \kappa(\zeta^{-1}\phi) d\zeta \\ &= \int_{\mathcal{I}} f(\alpha^{-1}\zeta) \kappa(\zeta^{-1}\phi) d\zeta \\ &= \int_{\mathcal{I}} f(\tilde{\zeta}) \kappa(\tilde{\zeta}^{-1}\alpha^{-1}\phi) d(\alpha\tilde{\zeta}) \\ &= (f *_{\mathcal{I}} \kappa)(\alpha^{-1}\phi) \\ &= [\alpha.(f *_{\mathcal{I}} \kappa)](\phi), \end{aligned} \quad (\text{F.12})$$

where we substituted $\tilde{\zeta} = \alpha^{-1}\zeta$ in the third step and made use of the fact that $d\tilde{\zeta}$ is a left Haar measure, i.e. satisfies $d(\alpha\tilde{\zeta}) = d\tilde{\zeta}$. The case of correlations follows trivially by Eq. (F.11).

The majority of equivariant CNNs rely on group convolutions or group correlations. In particular, the models in rows (1-3), (5), (7), (13), (17), (22), (25), (28), (29) and (36) of Table 14.1, all of which operate on homogeneous spaces and are (or could equivalently be) labeled by regular representations, are group convolutional CNNs. Prior to their use in equivariant CNNs, group convolutions have been widely applied in robotics [49] or for image analysis [192, 275, 276, 27, 277, 223]. Cohen and Welling [52] showed that group convolutions (or rather correlations) naturally generalize conventional CNNs. Since the feature maps of convolutional networks comprise multiple channels, they are not given by real-valued functions on \mathcal{I} but by vector-valued functions $f : \mathcal{I} \rightarrow \mathbb{R}^c$. Kernels are accordingly defined to be (unconstrained) matrix-valued functions on the group, i.e. $\kappa : \mathcal{I} \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$. The works of Kondor and Trivedi [162], Bekkers [10] and Cohen et al. [55][56], which we review in the following, generalize such group convolutional networks to arbitrary homogeneous spaces.

F.2 Scalar field convolutions on homogeneous spaces

We start with the \mathcal{I} -equivariant convolutional (or correlational) networks on homogeneous spaces by Kondor and Trivedi [162] and Bekkers [10]. Both theories define feature maps as *scalar fields on homogeneous spaces*, that is, each channel is given by a real-valued function

$$f : \mathcal{I}/H \rightarrow \mathbb{R}. \quad (\text{F.13})$$

Individual channels transform independently under the action of the global symmetry group \mathcal{I} as specified by

$$[\tilde{\phi}.f](\phi.H) := f(\tilde{\phi}^{-1}\phi.H) \quad \tilde{\phi} \in H, \quad \phi.H \in \mathcal{I}/H. \quad (\text{F.14})$$

Each layer $l = 1, \dots, L$ may be assigned a different subgroup $H_l \leq \mathcal{I}$ and thus homogeneous space \mathcal{I}/H_l on which its feature maps live. This allows for instance to model lifting

convolutions from the sphere $S^2 \cong \mathrm{SO}(3)/\mathrm{SO}(2)$ to the group $\mathrm{SO}(3) \cong \mathrm{SO}(3)/\{e\}$ when choosing subgroups $\mathrm{SO}(2)$ and $\{e\}$, respectively. The choices of subgroups correspond in some sense to the choices of group representations in our theory, which we will explain further below.

The results of the two papers are to large parts equivalent, however, Kondor and Trivedi [162] consider compact groups \mathcal{I} and convolutions while Bekkers [10] assume \mathcal{I} to be a Lie group and use correlations.

Kondor and Trivedi [162] : In a nutshell, Kondor and Trivedi [162] investigate the most general \mathcal{I} -equivariant linear maps between scalar field features on homogeneous spaces $\mathcal{I}/H_{\text{in}}$ and $\mathcal{I}/H_{\text{out}}$, assuming the transformation law in Eq. (F.14). They prove that this operation is given by a generalized group convolution with a kernel

$$\kappa : H_{\text{in}} \setminus \mathcal{I}/H_{\text{out}} \rightarrow \mathbb{R} \quad (\text{F.15})$$

on the double quotient space specified by H_{in} and H_{out} . Formulated for finite groups, as done by the authors, this generalized convolution operation is shown to be given by

$$(f *_{\mathcal{I}/H_{\text{in}}} \kappa)(\phi \cdot H_{\text{out}}) := |H_{\text{in}}| \sum_{H_{\text{in}} \cdot \zeta \in H_{\text{in}} \setminus \mathcal{I}} f(\phi \zeta^{-1} \cdot H_{\text{in}}) \kappa(H_{\text{in}} \cdot \zeta \cdot H_{\text{out}}). \quad (\text{F.16})$$

A comparison with the last line of Eq. (F.9) suggests that this operation is indeed closely related to group convolutions – the modular function Δ drops out since \mathcal{I} is compact and therefore unimodular. The generalized convolution is in fact equivalent to a group convolution

$$(f *_{\mathcal{I}/H_{\text{in}}} \kappa)(\phi \cdot H_{\text{out}}) = (f^\dagger *_{\mathcal{I}} \kappa^\dagger)(\phi) \quad (\text{F.17})$$

with features and kernels that are lifted according to the diagrams in Eqs. (F.7) and (F.8). Note that the convolution kernel on $H_{\text{in}} \setminus \mathcal{I}/H_{\text{out}}$ corresponds to a correlation kernel on $H_{\text{out}} \setminus \mathcal{I}/H_{\text{in}}$ since convolutions and correlations are according to Eq. (F.11) related by an inversion of the kernel argument. One could therefore view the kernels by Kondor and Trivedi [162] as left H_{out} -invariant correlation kernels on the input space $\mathcal{I}/H_{\text{in}}$.

To give an intuition on these results, we come back to our spherical CNN example from above. Let therefore $\mathcal{I} = \mathrm{SO}(3)$, $H_{\text{in}} = \mathrm{SO}(2)$ and, for now, $H_{\text{out}} = \{e\}$. This setting describes lifting convolutions from the 2-sphere $\mathcal{I}/H_{\text{in}} = \mathrm{SO}(3)/\mathrm{SO}(2) \cong S^2$ to the rotation group manifold $\mathcal{I}/H_{\text{out}} = \mathrm{SO}(3)/\{e\} \cong \mathrm{SO}(3)$. Considering correlations instead of convolutions, the kernels are real-valued functions on $H_{\text{out}} \setminus \mathcal{I}/H_{\text{in}} = \{e\} \setminus \mathrm{SO}(3)/\mathrm{SO}(2) \cong S^2$. If we let instead $H_{\text{out}} = \mathrm{SO}(2)$, the convolution maps from scalar fields on the 2-sphere to scalar fields on the 2-sphere $\mathcal{I}/H_{\text{out}} = \mathrm{SO}(3)/\mathrm{SO}(2)$. In this case the correlation kernels are given by real-valued functions on $\mathrm{SO}(2) \setminus \mathrm{SO}(3)/\mathrm{SO}(2)$. Equivalently, the correlation kernels are given by left $\mathrm{SO}(2)$ -invariant functions on S^2 , i.e. zonal kernels as visualized in Fig. 16.5. When assuming $H_{\text{in}} = H_{\text{out}} = \{e\}$, one has $\mathcal{I}/H_{\text{in}} = \mathcal{I}/H_{\text{out}} \cong \mathrm{SO}(3)$ and unconstrained kernels on $H_{\text{out}} \setminus \mathcal{I}/H_{\text{in}} \cong \mathrm{SO}(3)$, corresponding to conventional group convolutions (or correlations). These results are in line with our discussion in Section 17.2.

For completeness, we mention that Kondor and Trivedi [162] explain their results additionally from a representation theoretic perspective, i.e. with features and kernels in Fourier space. The fact that features and kernels live on quotient spaces is in this formulation reflected in sparsity patterns of the Fourier coefficients.

Bekkers [10] : Instead of considering compact groups, Bekkers [10] assumes \mathcal{I} to be a general Lie group. The feature maps of layer l are defined as real-valued square integrable functions in $L^2(\mathcal{I}/H_l)$ which transform according to Eq. (F.14) when being acted on by \mathcal{I} .

Bekkers [10] models the layers of his convolutional (or rather correlational) networks as linear bounded operators

$$\mathfrak{K} : L^2(\mathcal{I}/H_{\text{in}}) \rightarrow L^2(\mathcal{I}/H_{\text{out}}) \quad (\text{F.18})$$

between feature maps on homogeneous spaces $\mathcal{I}/H_{\text{in}}$ and $\mathcal{I}/H_{\text{out}}$. Such operators are in general given by integral operators of the form

$$[\mathfrak{K}f](\phi.H_{\text{out}}) = \int_{\mathcal{I}/H_{\text{in}}} \widehat{\kappa}(\phi.H_{\text{out}}, \zeta.H_{\text{in}}) f(\zeta.H_{\text{in}}) d\mu_{\mathcal{I}/H_{\text{in}}}, \quad (\text{F.19})$$

where $d\mu_{\mathcal{I}/H_{\text{in}}}$ is some Radon measure on $\mathcal{I}/H_{\text{in}}$ and

$$\widehat{\kappa} : \mathcal{I}/H_{\text{out}} \times \mathcal{I}/H_{\text{in}} \rightarrow \mathbb{R} \quad (\text{F.20})$$

is an integrable 2-argument kernel.

The requirement on the operator to be equivariant, that is,

$$\mathfrak{K}(\phi.f) = \phi.\mathfrak{K}(f) \quad \forall \phi \in \mathcal{I}, f \in L^2(\mathcal{I}/H_{\text{in}}), \quad (\text{F.21})$$

is shown to imply that the 2-argument kernel reduces to a single argument kernel

$$\widehat{\kappa}(\phi.H_{\text{out}}, \zeta.H_{\text{in}}) = \frac{d\mu_{\mathcal{I}/H_{\text{in}}}(\phi^{-1}\zeta.H_{\text{in}})}{d\mu_{\mathcal{I}/H_{\text{in}}}(\zeta.H_{\text{in}})} \kappa(\phi^{-1}\zeta.H_{\text{in}}). \quad (\text{F.22})$$

The group element $\phi \in \phi.H_{\text{out}} \subset \mathcal{I}$ is hereby an arbitrary representative of the coset in which it is contained. This 1-argument kernel is – up to a measure dependent scale factor – constrained to be left H_{out} -invariant:

$$\kappa(\zeta.H_{\text{in}}) = \frac{d\mu_{\mathcal{I}/H_{\text{in}}}(\xi^{-1}\zeta.H_{\text{in}})}{d\mu_{\mathcal{I}/H_{\text{in}}}(\zeta.H_{\text{in}})} \kappa(\xi^{-1}\zeta.H_{\text{in}}) \quad \forall \zeta.H_{\text{in}} \in \mathcal{I}/H_{\text{in}}, \xi \in H_{\text{out}} \quad (\text{F.23})$$

Note that this result is very similar to that of Kondor and Trivedi [162] since a left H_{out} -invariant kernel on $\mathcal{I}/H_{\text{in}}$ is equivalent to an element of $H_{\text{out}} \backslash \mathcal{I}/H_{\text{in}}$ (again assuming correlation kernels instead of convolution kernels). The main difference is the additional scale factor, which appears since the Radon measure $d\mu_{\mathcal{I}/H_{\text{in}}}$ is not necessarily left \mathcal{I} -invariant.

One of the practically relevant cases is that of group correlations, for which $H_{\text{in}} = \{e\}$ and $\mathcal{I}/\{e\} = \mathcal{I}$. In this case $d\mu_{\mathcal{I}}$ is a left (invariant) Haar measure on \mathcal{I} , such that the scale factor drops out. A second relevant case is that of affine equivariant convolutions on Euclidean spaces, i.e. the choices $\mathcal{I} = \text{Aff}(G)$ and $H_{\text{in}} = G$, for which $\mathcal{I}/H_{\text{in}} \cong \mathbb{R}^d$. Assuming $d\mu_{\mathcal{I}/H_{\text{in}}}$ to be the Lebesgue measure on \mathbb{R}^d and denoting $\phi = tg \in \mathcal{I}$, Bekkers [10] prove that the scale factor is in this case given by:

$$\frac{d\mu_{\mathcal{I}/H_{\text{in}}}((tg)^{-1}x)}{d\mu_{\mathcal{I}/H_{\text{in}}}(x)} = \frac{1}{|\det g|} \quad \forall x \in \mathbb{R}^d \quad (\text{F.24})$$

This is exactly the determinant factor which appears in our G -steerability kernel constraint, Eq. (9.37), as well.

Since $\text{SO}(3)$ is a Lie group, the spherical CNN examples that we gave after discussing the theory by Kondor and Trivedi [162] apply without changes (assuming the standard left-invariant measure on S^2).

Bekkers [10] defines kernels in close analogy to our GM -convolutions on the tangent spaces and projects them via exponential maps to the homogeneous spaces. The kernels on the tangent spaces are hereby modeled via B-splines. A difference is that Bekkers [10] does not need to consider parallel transporters since he is assuming scalar feature maps on the homogeneous spaces.

Relation to GM -convolutions: Due to the quite different formulation it is not immediately obvious how the results of Kondor and Trivedi [162] and Bekkers [10] relate to our theory. Instead of considering different quotient spaces \mathcal{I}/H_l in each layer l , we consider a fixed manifold M . To see how both approaches connect, assume another subgroup G to be given such that $H_l \leq G \leq \mathcal{I}$ for all layers $l = 1, \dots, L$ and satisfying that $M := \mathcal{I}/G$ is a manifold. The scalar features on \mathcal{I}/H_l can in this case be viewed as G -associated feature fields on M which transform according to *quotient representations* $\rho_{\text{quot}}^{G/H_l}$. To see this, note that the group action in Eq. (F.14) is nothing but the induced representation $\text{Ind}_{H_l}^{\mathcal{I}} \rho_{\text{triv}}^{H_l} = \rho_{\text{quot}}^{G/H_l}$ from the trivial representation of H_l , which describes the transformation law of scalar fields on \mathcal{I}/H_l . This representation can via induction in stages (see [35]) be decomposed into

$$\text{Ind}_{H_l}^{\mathcal{I}} \rho_{\text{triv}}^{H_l} = \text{Ind}_G^{\mathcal{I}} \text{Ind}_{H_l}^G \rho_{\text{triv}}^{H_l} = \text{Ind}_G^{\mathcal{I}} \rho_{\text{quot}}^{G/H_l}, \quad (\text{F.25})$$

that is, into the induction of the quotient representation $\rho_{\text{quot}}^{G/H_l}$ from G to \mathcal{I} . The real-valued functions on \mathcal{I}/H_l are therefore equivalent to $\rho_{\text{quot}}^{G/H_l}$ -fields on $M = \mathcal{I}/G$. This result was for Euclidean spaces already shown in Section 4.5.

Interesting special cases are $G = H_l$ and $G = \{e\}$. For the former one has $\rho_{\text{quot}}^{G/H_l} = \rho_{\text{triv}}^G$, describing scalar fields on $M = \mathcal{I}/G = \mathcal{I}/H_l$. For the latter, $\rho_{\text{quot}}^{G/H_l} = \rho_{\text{reg}}^G$ is the regular representation, corresponding to conventional group convolutions.

These insights imply that the theory of Kondor and Trivedi [162] explains all models in Table 14.1 which operate on homogeneous spaces of compact groups \mathcal{I} and are labeled by either trivial, regular or more general quotient representations – these are essentially the spherical CNNs in rows (36) and (37). A minor generalization of the theory to locally compact, unimodular groups would additionally describe some of the isometry equivariant Euclidean CNNs. As Bekkers [10] is assuming arbitrary Lie groups, his models additionally describe those $\text{Aff}(G)$ -equivariant CNNs in Table 14.1 which are labeled by trivial, regular or more general quotient representations. They cover in particular scale equivariant Euclidean CNNs ($G = \mathcal{S}$) for which the determinant factor $|\det g|$ is non-trivial.

Other types of feature fields and non-homogeneous spaces like punctured Euclidean spaces $\mathbb{E}_d \setminus \{0\}$ and spheres $S^2 \setminus \{n, s\}$, the icosahedron, general surfaces and the Möbius strip are not covered.

F.3 Steerable CNNs on homogeneous spaces

Motivated by Kondor and Trivedi's [162] generalization of group convolutions to homogeneous spaces, Cohen et al. [55][56] generalized steerable CNNs to homogeneous spaces

of locally compact unimodular groups.⁵ Instead of restricting to scalar fields, Cohen et al. [55][56] assume more general H_l -associated feature fields on \mathcal{I}/H_l which transform according to *induced representations* $\text{Ind}_{H_l}^{\mathcal{I}} \rho_l$ of \mathcal{I} . The network layers implement linear equivariant maps between such fields, i.e. they are *intertwiners between induced representations*. As expected, these layers are parameterized by – and are thus isomorphic to – spaces of steerable kernels. Cohen et al. [55][56] show that these kernels can be described on \mathcal{I} , on $\mathcal{I}/H_{\text{in}}$ or on $H_{\text{out}} \setminus \mathcal{I}/H_{\text{in}}$, in each case still satisfying a linear steerability constraint.⁶ The following three paragraphs will 1) introduce feature fields and their transformation laws on a global and local level, 2) review the spaces of intertwiners and steerable kernels which map between such fields, and 3) discuss how these results relate to ours.

Our formulation and notation in this section is adapted to be more similar to that which was chosen to develop our theory. It differs therefore slightly from that of Cohen et al. [55][56]. Most notably, we do not assume a single local trivialization (section) which is defined almost everywhere on \mathcal{I}/H_l but consider an atlas of local trivializations which cover the homogeneous space.⁷ The notation of local, coordinatized quantities is therefore augmented with gauge labels A, B, \dots .

Feature fields and induced representations: Let \mathcal{I} be a locally compact unimodular group and let $H_l \leq \mathcal{I}$ be any subgroup of it. As stated above, the quotient map

$$q_{\mathcal{I}/H_l}^{\mathcal{I}} : \mathcal{I} \rightarrow \mathcal{I}/H_l, \quad \phi \mapsto \phi.H_l \quad (\text{F.26})$$

implies a *principal H_l -bundle*; see Section 11.1.5. The right H_l -action on the total space \mathcal{I} is given by the usual right multiplication

$$\mathcal{I} \times \mathcal{I}/H_l \rightarrow \mathcal{I}/H_l, \quad (\phi, h) \mapsto \phi h \quad (\text{F.27})$$

of group elements. It preserves the fibers $\mathcal{I}_{\phi.H_l} = (q_{\mathcal{I}/H_l}^{\mathcal{I}})^{-1}(\phi.H_l) \subset \mathcal{I}$ since it satisfies

$$q_{\mathcal{I}/H_l}^{\mathcal{I}}(\phi h) = \phi h.H_l = \phi.H_l = q_{\mathcal{I}/H_l}^{\mathcal{I}}(\phi) \quad (\text{F.28})$$

for any $\phi \in \mathcal{I}$ and $h \in H_l$ and is easily seen to be both transitive and free. Abbreviating $U := U^A \cap U^B$, local trivializations $\Psi_{\mathcal{I}}^A$, $\Psi_{\mathcal{I}}^B$ of this bundle and the transition maps h^{BA} between them are defined via the following commutative diagram:

$$\begin{array}{ccc} & U \times H_l & \\ & \nearrow \Psi_{\mathcal{I}}^B & \uparrow (\text{id} \times h^{BA}) \\ \mathcal{I} \supseteq (q_{\mathcal{I}/H_l}^{\mathcal{I}})^{-1}(U) & \xrightarrow{\Psi_{\mathcal{I}}^A} & U \times H_l \\ q_{\mathcal{I}/H_l}^{\mathcal{I}} \downarrow & & \searrow \text{proj}_1 \\ \mathcal{I}/H_l \supseteq U & & \end{array} \quad (\text{F.29})$$

⁵Note that there are a preprint version [55] and a conference version [56] of this paper.

⁶As we will argue below, the constructions on $\mathcal{I}/H_{\text{in}}$ and $H_{\text{out}} \setminus \mathcal{I}/H_{\text{in}}$ depend on *local* sections and are therefore only possible for trivial bundles. We adapt the former to nontrivial bundles by defining kernels on an open cover of $\mathcal{I}/H_{\text{in}}$.

⁷This is only necessary if the homogeneous space is a (non-trivial) manifold. If it is discrete, one may always choose a global section $\mathcal{I}/H \rightarrow \mathcal{I}$ which selects coset representatives. One would in this case usually not talk about “atlases” and “local trivializations”, however, we will do so for simplicity.

As usual, the principal bundle trivializations imply local identity sections

$$\sigma^A : U^A \rightarrow (q_{\mathcal{I}/H_l}^{\mathcal{I}})^{-1}(U^A), \quad \phi.H_l \mapsto \sigma^A(\phi.H_l) := (\Psi_{\mathcal{I}}^A)^{-1}(\phi.H_l, e), \quad (\text{F.30})$$

which were introduced in Section 11.4. The identity sections labeled by \tilde{A} at $\zeta.H_l$ and A at $\phi\zeta.H_l$ are related by

$$\phi \sigma^{\tilde{A}}(\zeta.H_l) = \sigma^A(\phi\zeta.H_l) h_{\phi}^{A\tilde{A}}(\zeta.H_l), \quad (\text{F.31})$$

which defines the \mathcal{I} -induced gauge transformations $h_{\phi}^{A\tilde{A}}(\zeta.H_l) \in H_l$; see Eq. (13.32).⁸

The feature fields of steerable CNNs on homogeneous spaces are defined as sections $f \in \Gamma(\mathcal{A}_l)$ of associated H_l -bundles

$$\mathcal{A}_l := (\mathcal{I} \times \mathbb{R}^{c_l}) / \sim_{\rho_l}, \quad (\text{F.32})$$

which were introduced in Section 11.3. The equivalence relation

$$(\phi, f) \sim_{\rho_l} (\phi h^{-1}, \rho_l(h)f) \quad (\text{F.33})$$

is determined by a choice of field representation

$$\rho_l : H_l \rightarrow \mathbb{R}^{c_l} \quad (\text{F.34})$$

of the layer's subgroup; compare this to our analogous definition in Eq. (11.42). Being an associated H_l -bundle, the local feature vector bundle trivializations transform covariantly with those of the corresponding principal bundle:

$$\begin{array}{ccc} & U \times \mathbb{R}^{c_l} & \\ \Psi_{\mathcal{A}_l}^B & \nearrow & \uparrow (\text{id} \times \rho_l(h^{BA}) \cdot) \\ \pi_{\mathcal{A}_l}^{-1}(U) & \xrightarrow{\Psi_{\mathcal{A}_l}^A} & U \times \mathbb{R}^{c_l} \\ \pi_{\mathcal{A}_l} \downarrow & & \searrow \text{proj}_1 \\ U & & \end{array} \quad (\text{F.35})$$

The precise construction of associated bundle trivializations from principal bundle trivializations was given in Eq. (11.65).

Cohen et al. [55][56] use two different approaches to describe feature fields. Globally, feature fields are represented as functions

$$F : \mathcal{I} \rightarrow \mathbb{R}^{c_l} \quad \text{such that} \quad F(\phi h^{-1}) = \rho_l(h)F(\phi) \quad \forall \phi \in \mathcal{I}, h \in H_l, \quad (\text{F.36})$$

whose definition is consistent with the equivalence relation from Eq. (F.32). On trivializing neighborhoods $U^A \subseteq \mathcal{I}/H_l$, the fields are furthermore given by feature vector coefficient fields

$$f^A : U^A \rightarrow \mathbb{R}^c \quad (\text{F.37})$$

⁸To avoid confusion, note that Cohen et al. [55][56] denote $h_{\phi}^{A\tilde{A}}(\zeta.H_l)$ by $h(\zeta.H_l, \phi)$, omitting the gauge labels.

relative to some gauge $\Psi_{A_l}^A$. While the former is more convenient for algebraic manipulations, the latter is non-redundant, and therefore more suitable for numerical implementations. The local field representation may at any time be computed from the global one by setting

$$f^A(\phi.H_l) = F(\sigma^A(\phi.H_l)) \quad \text{for } \phi.H_l \in U^A. \quad (\text{F.38})$$

Here $\sigma^A : U^A \rightarrow \mathcal{I}$ is that local section of the principal H_l bundle which corresponds to the chosen trivialization $\Psi_{\mathcal{I}}^A$ (“identity section”) and is analogously defined to Eq. (11.61). Note that the global field representation can in general not be recovered from a (single) local one. It is, however, locally over U^A given by

$$F(\phi) = \rho_l(\psi_{\mathcal{I}, \phi.H_l}^A(\phi))^{-1} f^A(\phi.H_l) \quad \text{for } \phi \in (q_{\mathcal{I}/H_l}^{\mathcal{I}})^{-1}(U^A) \subseteq \mathcal{I}, \quad (\text{F.39})$$

which is closely related to Eq. (11.66).

The global, active transformations of feature fields are formalized by *induced representations* $\text{Ind}_{H_l}^{\mathcal{I}} \rho_l$ of \mathcal{I} , which are conceptually similar to our isometry pushforwards from Def. 13.1.2. For the global field representations, this action is simply defined as a shift on \mathcal{I} :

$$[\text{Ind}_{H_l}^{\mathcal{I}} \rho_l](\zeta) F(\phi) = F(\zeta^{-1}\phi) \quad (\text{F.40})$$

Since the \mathcal{I} -action is global, it is more difficult to describe for local field representations. Let U^A be a trivializing neighborhood around $\phi.H_l$ and $U^{\tilde{A}}$ around $\zeta^{-1}\phi.H_l$. The action of the induced representation is relative to gauges on these neighborhoods given by

$$[\text{Ind}_{H_l}^{\mathcal{I}} \rho_l](\zeta) f^A(\phi.H_l) = \rho_l(h_{\zeta}^{A\tilde{A}}) f^{\tilde{A}}(\zeta^{-1}\phi.H_l), \quad (\text{F.41})$$

where $h_{\zeta}^{A\tilde{A}}$ is the ζ -induced gauge transformation, which is analogously defined to that in Eq. (13.38). Note the similarity of this definition to our isometry pushforward of feature fields in coordinates from Eq. (8.19). We furthermore identify the transformation law of scalar fields on homogeneous spaces from Eq. (F.14) as a special case for trivial representations ρ_l . Steerable CNNs on homogeneous spaces cover therefore the homogeneous scalar field convolutions of Kondor and Trivedi [162] and Bekkers [10] as a special case (ignoring the different assumptions made on the type of group \mathcal{I}).

Intertwiners between induced representations and steerable kernels: The main endeavor of Cohen et al. [55][56] is to characterize the space

$$\text{Hom}_{\mathcal{I}}(\Gamma(\mathcal{A}_{\text{in}}), \Gamma(\mathcal{A}_{\text{out}})) := \quad (\text{F.42})$$

$$\{\mathfrak{K} : \Gamma(\mathcal{A}_{\text{in}}) \rightarrow \Gamma(\mathcal{A}_{\text{out}}) \text{ linear} \mid \mathfrak{K} \circ \text{Ind}_{H_{\text{in}}}^{\mathcal{I}}(\phi) = \text{Ind}_{H_{\text{out}}}^{\mathcal{I}}(\phi) \circ \mathfrak{K} \quad \forall \phi \in \mathcal{I}\}$$

of intertwiners between induced representations, i.e. the space of linear equivariant maps between feature fields. Diagrammatically, this space consists of those linear maps \mathfrak{K} which let the following diagram commute for any $\phi \in \mathcal{I}$:

$$\begin{array}{ccc} \Gamma(\mathcal{A}_{\text{in}}) & \xrightarrow{\mathfrak{K}} & \Gamma(\mathcal{A}_{\text{out}}) \\ \text{Ind}_{H_{\text{in}}}^{\mathcal{I}} \rho_{\text{in}}(\phi) \downarrow & & \downarrow \text{Ind}_{H_{\text{out}}}^{\mathcal{I}} \rho_{\text{out}}(\phi) \\ \Gamma(\mathcal{A}_{\text{in}}) & \xrightarrow{\mathfrak{K}} & \Gamma(\mathcal{A}_{\text{out}}) \end{array} \quad (\text{F.43})$$

These maps are the analog to our isometry equivariant kernel field transforms, which were defined in Def. 13.2.1. Cohen et al. [55][56] prove that these maps are given by correlations with steerable kernels. We will in the following briefly review these results for both global and local field representations.

When working with the global field representation from Eq. (F.36), Cohen et al. [55][56] start with a general bounded linear operator \mathfrak{K} of the form

$$[\mathfrak{K}F](\phi) = \int_{\mathcal{I}} \hat{\kappa}(\phi, \zeta) F(\zeta) d\zeta \quad (\text{F.44})$$

where $d\zeta$ is a left Haar measure on \mathcal{I} and

$$\hat{\kappa} : \mathcal{I} \times \mathcal{I} \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \quad (\text{F.45})$$

is a matrix-valued 2-argument kernel. The equivariance constraint is shown to require the kernels to satisfy the relation $\hat{\kappa}(\tilde{\phi}\phi, \tilde{\phi}\zeta) = \hat{\kappa}(\phi, \zeta)$ for any choice of group elements $\tilde{\phi}, \phi, \zeta \in \mathcal{I}$. This result is resembling our Theorem 13.2.4, which states that isometry equivariant kernel field transforms imply kernel fields which are invariant under the action of isometries, Def. 13.2.3. Given this constraint, the 2-argument kernel can be replaced by a 1-argument kernel which is defined as

$$\kappa : \mathcal{I} \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}, \quad \phi \mapsto \kappa(\phi) := \hat{\kappa}(e, \phi). \quad (\text{F.46})$$

We therefore have $\hat{\kappa}(\phi, \zeta) = \hat{\kappa}(\phi^{-1}\phi, \phi^{-1}\zeta) = \kappa(\phi^{-1}\zeta)$, implying that the linear operator is given by a *group correlation* (Eq. (F.10)), that is:

$$[\mathfrak{K}F](\phi) = \int_{\mathcal{I}} \kappa(\phi^{-1}\zeta) F(\zeta) d\zeta = (F \star_{\mathcal{I}} \kappa)(\phi) \quad (\text{F.47})$$

The correlation kernel is furthermore required to satisfy the linear $H_{\text{out}}\text{-}H_{\text{in}}$ -steerability constraint

$$\kappa(h_{\text{out}} \phi h_{\text{in}}) = \rho_{\text{out}}(h_{\text{out}}) \kappa(\phi) \rho_{\text{in}}(h_{\text{in}}) \quad \forall \phi \in \mathcal{I}, h_{\text{in}} \in H_{\text{in}}, h_{\text{out}} \in H_{\text{out}}. \quad (\text{F.48})$$

This constraint is reminiscent of that found by Kondor and Trivedi [162] and Bekkers [10]. Instead of enforcing kernels to be left H_{out} - and right H_{in} -*invariant*, which would correspond to trivial representations $\rho_{\text{out}} = \rho_{\text{triv}}^{H_{\text{out}}}$ and $\rho_{\text{in}} = \rho_{\text{triv}}^{H_{\text{in}}}$, the constraint of Cohen et al. [55][56] allows for more general steerable kernels. The vector space $\mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^{\mathcal{I}}$ of such steerable correlation kernels is argued to be isomorphic to the intertwiner space $\text{Hom}_{\mathcal{I}}(\Gamma(\mathcal{A}_{\text{in}}), \Gamma(\mathcal{A}_{\text{out}}))$.

Since the global field representations F on \mathcal{I} are redundant they are not the best choice for numerical implementations. Cohen et al. [55][56] are therefore additionally investigating intertwiners which operate on local field representations. The authors approach this problem by assuming one *single local trivialization* to be given, which is *defined almost everywhere* on the homogeneous space $\mathcal{I}/H_{\text{in}}$. They are therefore effectively operating on a trivial bundle. Our following review adapts their results slightly to the more general case of a set of field representations relative to an *atlas of local trivializations*. The formulation of Cohen et al. [55][56] is retrieved by restricting the integration to one single trivialization. We explicitly write out all gauge labels to make the coordinate dependencies transparent. To give an overview of the local trivializations that will play a role in the following, we mention that we will need to consider trivializing neighborhoods $U^A, U^{\tilde{A}}, U^H \subseteq \mathcal{I}/H_{\text{in}}$ such that

$$\zeta.H_{\text{in}} \in U^A, \quad \tilde{\phi}\zeta.H_{\text{in}} \in U^{\tilde{A}} \quad \text{and} \quad h\zeta.H_{\text{in}} \in U^H \quad (\text{F.49})$$

and trivializing neighborhoods $U^P, U^{\tilde{P}}, U^E \subseteq \mathcal{I}/H_{\text{out}}$ such that

$$\phi.H_{\text{out}} \in U^P, \quad \tilde{\phi}\phi.H_{\text{out}} \in U^{\tilde{P}} \quad \text{and} \quad e.H_{\text{out}} \in U^E. \quad (\text{F.50})$$

We will furthermore assume any partition of unity $\{\mathcal{P}_{U^X}\}_{X \in \mathfrak{X}}$ subordinate to the open cover underlying the atlas $\mathcal{A}_{\text{in}} = \{(U^X, \Psi^X)\}_{X \in \mathfrak{X}}$ of local trivializations on $\mathcal{I}/H_{\text{in}}$. This means that we are given maps $\mathcal{P}_{U^X} : \mathcal{I}/H_{\text{in}} \rightarrow [0, 1]$ with the properties

$$\text{supp}(\mathcal{P}_{U^X}) \subseteq U^X \quad \text{and} \quad \sum_{U^X \in \mathcal{A}_{\text{in}}} \mathcal{P}_{U^X}(\phi.H_{\text{in}}) = 1 \quad \forall \phi.H_{\text{in}} \in \mathcal{I}/H_{\text{in}}. \quad (\text{F.51})$$

Eq. (F.44) stated the general form of a bounded linear operator between global field representations F . Its local analog, which makes use of the partition of unity, is given by

$$[\widehat{\kappa}f]^P(\phi.H_{\text{out}}) = \sum_{U^A \in \mathcal{A}_{\text{in}}} \int_{U^A} \mathcal{P}_{U^A}(\zeta.H_{\text{in}}) \widehat{\kappa}^{PA}(\phi.H_{\text{out}}, \zeta.H_{\text{in}}) f^A(\zeta.H_{\text{in}}) d(\zeta.H_{\text{in}}), \quad (\text{F.52})$$

where P and A label local trivializations as stated above and $d(\zeta.H_{\text{in}})$ is a measure on $\mathcal{I}/H_{\text{in}}$. We furthermore have 2-argument kernels

$$\widehat{\kappa}^{PA} : U^P \times U^A \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}, \quad (\phi.H_{\text{out}}, \zeta.H_{\text{in}}) \mapsto \widehat{\kappa}(\sigma^P(\phi.H_{\text{out}}), \sigma^A(\zeta.H_{\text{in}})) \quad (\text{F.53})$$

which are inherently *locally defined* on $U^P \times U^A \subseteq \mathcal{I}/H_{\text{out}} \times \mathcal{I}/H_{\text{in}}$. The global 2-argument kernel can be recovered from a set of local kernels on the open covers. Cohen et al. [55][56] prove that these local kernels are required to satisfy

$$\widehat{\kappa}^{PA}(\phi.H_{\text{out}}, \zeta.H_{\text{in}}) = \rho_{\text{out}}(h_{\tilde{\phi}}^{\tilde{P}P}(\phi.H_{\text{out}}))^{-1} \widehat{\kappa}^{\tilde{P}\tilde{A}}(\tilde{\phi}\phi.H_{\text{out}}, \tilde{\phi}\zeta.H_{\text{in}}) \rho_{\text{in}}(h_{\tilde{\phi}}^{\tilde{A}A}(\zeta.H_{\text{in}})) \quad (\text{F.54})$$

for any $\tilde{\phi} \in \mathcal{I}$. Note that $h_{\tilde{\phi}}^{\tilde{P}P}(\phi.H_{\text{out}})$ is hereby an induced gauge transformation on $\mathcal{I}/H_{\text{out}}$ while $h_{\tilde{\phi}}^{\tilde{A}A}(\zeta.H_{\text{in}})$ is an induced gauge transformation on $\mathcal{I}/H_{\text{in}}$. In order to reduce these local 2-argument kernels to local 1-argument kernels, Cohen et al. [55][56] consider the unique group element $\tilde{\phi} \in \mathcal{I}$ which satisfies 1) $\tilde{\phi}\phi.H_{\text{out}} = e.H_{\text{out}}$ and 2) $\tilde{\phi}\sigma^P(\phi.H_{\text{out}}) = \sigma^E(e.H_{\text{out}}) = e$, where the last equality fixes a specific gauge at the “origin” $e.H_{\text{out}}$, which is always possible. The first point allows us to identify the gauges \tilde{P} and E without loss of generality. The relations imply furthermore $\tilde{\phi} = \sigma^P(e.H_{\text{out}})^{-1}$ and, by Eq. (F.31), $h_{\tilde{\phi}}^{EP}(\phi.H_{\text{out}}) = e$. Plugging these choices into Eq. (F.54) yields

$$\begin{aligned} & \widehat{\kappa}^{PA}(\phi.H_{\text{out}}, \zeta.H_{\text{in}}) \\ &= \underbrace{\text{id}_{\mathbb{R}^{c_{\text{out}}}} \widehat{\kappa}^{E\tilde{A}}(e.H_{\text{out}}, \sigma^P(\phi.H_{\text{out}})^{-1}\zeta.H_{\text{in}})}_{=: \widehat{\kappa}^{E\tilde{A}}(\sigma^P(\phi.H_{\text{out}})^{-1}\zeta.H_{\text{in}})}, \end{aligned} \quad (\text{F.55})$$

where the identity map is kept explicit to explain the gauge labels. We furthermore introduced the local 1-argument kernels

$$\overleftarrow{\kappa}^{E\tilde{A}} : U^{\tilde{A}} \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}, \quad (\text{F.56})$$

whose responses are always given in the specific gauge E at $e.H_{\text{out}}$. These kernels are still required to satisfy the H_{out} -steerability constraints

$$\overleftarrow{\kappa}^{EH}(h_{\text{out}}\zeta.H_{\text{in}}) = \rho_{\text{out}}(h_{\text{out}}) \overleftarrow{\kappa}(\zeta.H_{\text{in}})^{EA} \rho_{\text{in}}(h_{h_{\text{out}}}^{HA}(\zeta.H_{\text{in}}))^{-1} \quad (\text{F.57})$$

for any $\zeta.H_{\text{in}} \in \mathcal{I}/H_{\text{in}}$ and $h_{\text{out}} \in H_{\text{out}}$. Putting everything together, the equivariant correlation becomes

$$\begin{aligned} [\mathfrak{K}f]^P(\phi.H_{\text{out}}) &= \text{id}_{\mathbb{R}^{c_{\text{out}}}}^{PE} \sum_{U^A \in \mathcal{A}_{\text{in}}} \int_{U^A} \mathcal{P}_{U^A}(\zeta.H_{\text{in}}) \overleftarrow{\kappa}^{E\tilde{A}}(\sigma^P(\phi.H_{\text{out}})^{-1}\zeta.H_{\text{in}}) \\ &\quad \cdot \rho_{\text{in}}(h_{\sigma^P(\phi.H_{\text{out}})^{-1}}^{\tilde{A}A}(\zeta.H_{\text{in}})) f^A(\zeta.H_{\text{in}}) d(\zeta.H_{\text{in}}). \end{aligned} \quad (\text{F.58})$$

Adding the assumption that a single gauge $A = \tilde{A}$ covers $\mathcal{I}/H_{\text{in}}$ almost everywhere, we can drop the partition of unity and retrieve the formulation of Cohen et al. [55][56]:

$$\begin{aligned} [\mathfrak{K}f]^P(\phi.H_{\text{out}}) &= \text{id}_{\mathbb{R}^{c_{\text{out}}}}^{PE} \int_{U^A} \overleftarrow{\kappa}^{EA}(\sigma^P(\phi.H_{\text{out}})^{-1}\zeta.H_{\text{in}}) \\ &\quad \cdot \rho_{\text{in}}(h_{\sigma^P(\phi.H_{\text{out}})^{-1}}^{AA}(\zeta.H_{\text{in}})) f^A(\zeta.H_{\text{in}}) d(\zeta.H_{\text{in}}) \end{aligned} \quad (\text{F.59})$$

We comment on the relation of this operation to our *GM*-convolutions further below.

Instead of defining the *local* 1-argument kernels in coordinates from Eq. (F.56) on local subsets $U^{\tilde{A}}$, Cohen et al. [55][56] define them *globally* on $\mathcal{I}/H_{\text{in}}$. Since their construction relies on a continuous section, this is only possible if the bundles are trivial. Our adaptation to local kernel representations on an open covering is bridging this gap.

Cohen et al. [55][56] claim an isomorphism between the global kernels on \mathcal{I} , satisfying the steerability constraint in Eq. (F.48), and their kernels on $\mathcal{I}/H_{\text{in}}$, satisfying the steerability constraint in Eq. (F.57). Note that this isomorphism can only hold if either the bundle is trivial or the continuity assumption on the sections (and therefore network inference) is dropped. It should, however, be possible to prove an isomorphism between the global kernel and a collection of local kernels on a covering of $\mathcal{I}/H_{\text{in}}$, satisfying the relations in Eq. (F.57).

The authors furthermore claim that the steerable kernels can be described on the double quotient space $H_{\text{out}} \setminus \mathcal{I}/H_{\text{in}}$, still satisfying a steerability constraint.

Relation to GM-convolutions: The steerable CNNs on homogeneous spaces by Cohen et al. [55][56] are conceptually quite similar to our *GM*-convolutions on Riemannian manifolds, however, there are some important differences which we discuss in the following. Most importantly, the theories differ in 1) being based on different spaces \mathcal{I}/H_l in each layer l vs. assuming a fixed manifold M , 2) modeling kernels on the space $\mathcal{I}/H_{\text{in}}$ itself or on tangent spaces $T_p M$ of it, 3) the way of how weights are shared, and 4) the types of global symmetry group \mathcal{I} and spaces \mathcal{I}/H_l or M , which they cover. Despite these differences, many of the results of Cohen et al. [55][56] have analogs in our theory.

Both theories share the idea to define feature fields as sections of associated vector bundles. While Cohen et al. [55][56] consider a global symmetry group \mathcal{I} as a set of multiple principal H_l -bundles over homogeneous spaces \mathcal{I}/H_l , we work with some G -structure *GM* over a fixed Riemannian manifold M . All of our feature vector bundles are defined as G -bundles and are associated to each other, while the feature bundles of Cohen et al. [55][56] may not be associated to each other if their structure groups H_l do not agree. As already claimed at the end of the last Appendix F.2, these differences can be mitigated if a structure group G can be chosen such that $H_l \leq G \leq \mathcal{I}$ for every layer l and $M := \mathcal{I}/G$ is a Riemannian manifold.

One can then replace all homogeneous spaces \mathcal{I}/H_l with M and all H_l -representations ρ_l with induced G -representations

$$\rho_l^G := \text{Ind}_{H_l}^G \rho_l. \quad (\text{F.60})$$

The global field transformation laws are preserved by this reinterpretation since

$$\text{Ind}_{H_l}^{\mathcal{I}} \rho_l = \text{Ind}_G^{\mathcal{I}} \text{Ind}_{H_l}^G \rho_l = \text{Ind}_G^{\mathcal{I}} \rho_l^G \quad (\text{F.61})$$

holds by induction in stages [35].

Another main difference lies in the definition of convolution kernels and weight sharing. On the global, coordinate free level and prior to the isometry assumption, Cohen et al. [55][56] start in Eq. (F.44) with a bounded linear operator which is parameterized by an unconstrained kernel

$$\hat{\kappa} : \mathcal{I} \times \mathcal{I} \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}. \quad (\text{F.62})$$

This operator corresponds in our theory to a general kernel field transform, Def. 12.2.5, which is parameterized by an unconstrained kernel field

$$\mathcal{K} : TM \rightarrow \text{Hom}(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}}), \quad (\text{F.63})$$

see Def. (12.2.1). The 2-argument kernels $\hat{\kappa}$ can be thought of as representing a kernel field as well. Their two arguments are thereby thought of as addressing 1) a specific (1-argument) kernel, yielding a response at the corresponding point in the output bundle $\mathcal{I} \rightarrow \mathcal{I}/H_{\text{out}}$ and 2) the spatial dependency of this 1-argument kernel on the input bundle $\mathcal{I} \rightarrow \mathcal{I}/H_{\text{in}}$. The analog in our kernel fields \mathcal{K} is that elements $v \in TM$ encode 1) the location $p = \pi_{TM}(v)$ of the kernel and 2) its spatial dependency via $v \in T_p M$.

When requiring the bounded linear operator to be \mathcal{I} -equivariant, the 2-argument kernel $\hat{\kappa}$ becomes constrained to satisfy

$$\hat{\kappa}(\tilde{\phi}\phi, \tilde{\phi}\zeta) = \hat{\kappa}(\phi, \zeta) \quad \forall \tilde{\phi} \in \mathcal{I}. \quad (\text{F.64})$$

Isometry equivariant kernel field transforms were in Theorem 13.2.4 shown to require the isometry invariance of the kernel field, i.e.

$$\tilde{\phi}_{*,\mathcal{K}} \mathcal{K} = \mathcal{K} \quad \forall \tilde{\phi} \in \mathcal{I}; \quad (\text{F.65})$$

see Def. 13.2.3 and Fig. 13.6.

The invariance constraint on 2-argument kernels $\hat{\kappa}$ allows to replace them with 1-argument kernels

$$\kappa : \mathcal{I} \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}, \quad (\text{F.66})$$

defined in Eq. (F.46). They are still required to satisfy the steerability constraint in Eq. (F.48). Our isometry invariant kernel fields were in Theorem 13.3.2 shown to be equivalent to a field of kernels

$$\widehat{\mathcal{Q}} : \pi_{TM}^{-1}(r_M(\mathcal{I} \setminus M)) \rightarrow \pi_{\text{Hom}}^{-1}(r_M(\mathcal{I} \setminus M)) \quad (\text{F.67})$$

whose support is restricted to the tangent spaces over representatives $r_M(\mathcal{I} \setminus M) \subseteq M$ of the quotient $\mathcal{I} \setminus M$.⁹ These kernels are required to satisfy a stabilizer subgroup steerability

⁹Theorem 13.3.1 proves another isomorphism to a space of kernels $\mathcal{Q} : r_{TM}(\mathcal{I} \setminus TM) \rightarrow r_{\text{Hom}}(\mathcal{I} \setminus \text{Hom})$ whose support is even further restricted to representatives of the tangent bundle quotient $\mathcal{I} \setminus TM$.

constraint as well. For the specific case that M is a homogeneous space of its isometry group, the quotient $\mathcal{I}\backslash M$ reduces to a single element. Theorem 13.3.2 implies in this case a single (1-argument) kernel

$$\widehat{\mathcal{Q}} : T_p M \rightarrow \text{Hom}(\mathcal{A}_{\text{in},p}, \mathcal{A}_{\text{out},p}) \quad (\text{F.68})$$

at $p = r_M(\mathcal{I}\backslash M)$, which is the direct analog to the 1-argument kernel of Cohen et al. [55][56].

Note that the full kernel fields can via the action of \mathcal{I} be reconstructed from the single 1-argument kernels. The theories derive therefore both a form of convolutional weight sharing from the requirement of global symmetry equivariance. While kernels can for transitive symmetries be shared over the whole homogeneous space, they can in general only be shared over the orbits of the symmetry group. If the manifold is asymmetric in such a way that the orbits are single points no weights can be shared with this definition. As this is the default case for Riemannian manifolds, GM -convolutions resort to the sharing of G -steerable kernels by placing them relative to frames of the G -structure. This definition does not have a counterpart in steerable CNNs on homogeneous spaces. Our Theorem 13.3.3 shows, however, that the global symmetry induced weight sharing is for the specific case of homogeneous spaces equivalent to our process of sharing G -steerable kernels. In other words, isometry equivariant kernel field transforms on homogeneous spaces are necessarily convolutions – this mirrors the central results of Kondor and Trivedi [162], Bekkers [10] and Cohen et al. [55][56].

After investigating the analogies for the global, coordinate free kernels of both theories, we compare the definition of their coordinate representations relative to local trivializations. Given some choice of trivializing neighborhoods $U^P \subseteq \mathcal{I}/H_{\text{out}}$ and $U^A \subseteq \mathcal{I}/H_{\text{in}}$, the unconstrained global 2-argument kernels of Cohen et al. [55][56] are locally represented by unconstrained functions

$$\widehat{\kappa} : U^P \times U^A \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} . \quad (\text{F.69})$$

In our theory, we instead have a single trivializing neighborhood $U^P = U^A \subseteq M$ relative to which a kernel field is given by an unconstrained map

$$\mathcal{K}^A : U^A \times \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} . \quad (\text{F.70})$$

Investigating the *global* \mathcal{I} -equivariance of the operator \mathfrak{K} based on *local* kernels is on non-trivial bundles necessarily difficult as it involves multiple trivializations. The equivariance requirement implies for steerable CNNs on homogeneous spaces the constraints between different local kernels in Eq. (F.54). They leads to the 1-argument kernels

$$\overleftarrow{\kappa}^E : U^A \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} , \quad (\text{F.71})$$

from Eq. (F.56), which are still subject to the steerability constraint in Eq. (F.57). Single kernels $\mathcal{K}_p : T_p M \rightarrow \text{Hom}(\mathcal{A}_{\text{in},p}, \mathcal{A}_{\text{out},p})$ (like e.g. $\widehat{\mathcal{Q}}$ from Eq. (F.68)) are according to Eq. (12.24) in coordinates given by functions

$$\mathcal{K}_p^A : \mathbb{R}^d \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} , \quad (\text{F.72})$$

whose domains are tangent space coordinates \mathbb{R}^d instead of of a open subset U^A of the manifold. A particular important example are G -steerable kernels, which correspond to the GM -convolutional kernel fields from Def. 12.2.3.

While our kernels are globally defined in a single gauge $\psi_{TM,p}^A$ of $T_p M$, the local 1-argument kernels of Cohen et al. [55][56] need to be defined on an open cover of $\mathcal{I}/H_{\text{in}}$. As this is

significantly more complicated, they propose therefore to represent the kernels on a single gauge which is defined almost everywhere.¹⁰ Note that this still requires that this single trivializing neighborhood is closed under the left action of H_{out} in order for the constraint in Eq. (F.57) to make sense. We investigated this approach in Section 17.2 for the specific example of spherical CNNs, defining kernels on the trivializing neighborhood $U^A = S^2 \setminus -n$. Theorem 17.2.1 proved that Cohen et al.’s [55][56] steerable kernels on $S^2 \setminus -n \subset S^2$ are in this case isomorphic to our G -steerable kernels on $B_{\mathbb{R}^2}(0, \pi) \subset \mathbb{R}^2$. The equivalence of the corresponding convolutions was established in Theorem 17.2.2.

Finally, we discuss which class of models steerable CNNs on homogeneous spaces cover. Obviously, the theory does not describe convolutions on non-homogeneous spaces like punctured Euclidean spaces $\mathbb{E}_d \setminus \{0\}$, the sphere without poles $S^2 \setminus \{n, s\}$, whose isometries $O(2)$ are non-transitive, the icosahedron, general surfaces or the Möbius strip. However, in contrast to GM -convolutions, the base spaces \mathcal{I}/H_l are not required to be Riemannian manifolds. While Kondor and Trivedi [162] and Bekkers [10] cover only those convolutions whose feature fields transform according to scalar fields on \mathcal{I}/H_l , the associated bundle formulation of Cohen et al. [55][56] allows for general field representations ρ_l . Restricting to unimodular groups, steerable CNNs on homogeneous spaces do, however, only include those $\text{Aff}(G)$ -equivariant Euclidean convolutions for which the structure groups are subgroups of $O(d)$. This reflects in the fact that the steerability constraints of Cohen et al. [55][56] do not include the determinant factor in the constraint of Bekkers [10] and of our G -steerable kernels.

¹⁰In practice, one might anyways work with compactly supported kernels on a single trivializing neighborhood, which would render this choice unproblematic.

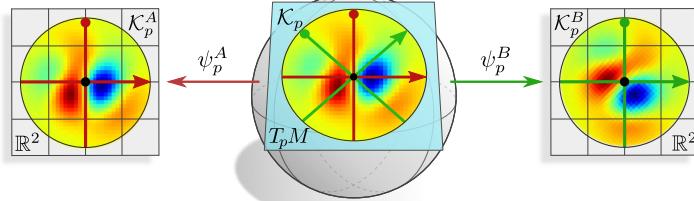


Figure F.1: A *given* coordinate free kernel \mathcal{K}_p on the tangent space $T_p M$ may be represented in arbitrary gauges ψ_p^A or ψ_p^B . Its coordinate expressions \mathcal{K}_p^A and \mathcal{K}_p^B on \mathbb{R}^d differ in general from each other. G -steerable kernels have the property to take exactly the same form in all gauges, that is, they satisfy $\mathcal{K}_p^A = \mathcal{K}_p^B = K$ (not visualized).

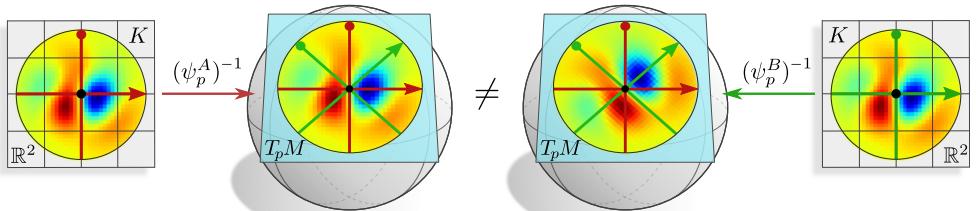


Figure F.2: A coordinate free kernel may be *defined by sharing* a given kernel K on \mathbb{R}^d relative to some reference frame. Different choices of frames result in a different coordinate free kernel. G -steerable kernels have the property to produce exactly the same coordinate free kernel, independent from the chosen reference frame along which they are shared (not visualized). This allows for a coordinate independent weight sharing.

Figure F.3: Visualizations of the concepts of 1) the coordinatization of a given coordinate free kernel on $T_p M$ in Fig. F.1 and 2) the sharing of a kernel on \mathbb{R}^d along different frames of $T_p M$ in Fig. F.2. Depending on the direction (coordinatization or sharing), the resulting kernels differ either on \mathbb{R}^d or on $T_p M$. G -steerable kernels are exactly those kernels that can be shared in a coordinate independent manner, i.e. for which the resulting kernels are equivalent in arbitrary gauges.

Coordinate independent weight sharing

A fundamental assumption in the design of GM -convolutions is that kernels K on \mathbb{R}^d are shared relative to some choice of reference frame as visualized in Fig. F.2. For general kernels different choices of frames will lead to different alignments of the resulting coordinate free kernel on the tangent space $T_p M$ – the weight sharing process is therefore not coordinate independent. Fig. F.1 shows a different situation: here we assume a coordinate free kernel \mathcal{K}_p that is already given on $T_p M$ and express it in different gauges on \mathbb{R}^d . The coordinate representations \mathcal{K}_p^A and \mathcal{K}_p^B do in general not agree with each other but the construction is coordinate independent.

G -steerable kernels are constrained exactly such that they guarantee the coordinate independence of the weight sharing process. Sharing them relative to different frames results in the same coordinate independent kernel on the tangent space, that is, there will be no difference between the two kernels in the middle of Fig. F.2. Equivalently, the resulting coordinate free kernel on $T_p M$ will take the same form $\mathcal{K}_p^A = \mathcal{K}_p^B = K$ when being expressed in different gauges, that is, the left and the right kernel in Fig. F.1 would agree.

Note that this does not necessarily require the kernel to be invariant in the sense that $K(gv) = K(v)$ for any $g \in G$ and any $v \in \mathbb{R}^d$, as the visual intuition might suggest. This is indeed a special case for kernels that map between scalar fields, i.e. for which both ρ_{in} and ρ_{out} are trivial representations (see e.g. Fig. 1.3b (left) for $G = \text{SO}(2)$ or $\text{O}(2)$ or Fig. 1.3b (left) for $G = \mathcal{R}$). For more general field types the kernels need to be gauge equivariant, i.e. need to satisfy the G -steerability constraint $K(gv) = |\det g|^{-1} \rho_{\text{out}}(g) K(v) \rho_{\text{in}}(g)^{-1}$ which allows for a steering of the $c_{\text{out}} \times c_{\text{in}}$ kernel channels (not visualized). Of course, G -steerable kernels can be interpreted as being gauge *invariant* in the sense that $K(v) = |\det g| \rho_{\text{out}}(g)^{-1} K(gv) \rho_{\text{in}}(g)$ for any $g \in G$ and any $v \in \mathbb{R}^d$. This notion of gauge invariance allows the coordinate independent sharing of G -steerable kernels.

A more detailed discussion of coordinate free kernels and their coordinate expressions is found in Section 9.2.2. The G -steerability constraint is in different settings derived in Sections 4.3.1 and 9.2.3. Chapter 5 gives more details on steerable kernels in general.

An intuition for the Wigner-Eckart theorem for steerable kernels

The Wigner-Eckart theorem 5.3.1 describes the construction of complete G -steerable kernel bases from 1) harmonics on G -orbits, 2) irrep endomorphisms and 3) Clebsch-Gordan coefficients. As the formal proofs in the original publications by Lang and Weiler [173] and Cesa et al. [40] are rather technical, we aim in this appendix to motivate these three ingredients step by step:

harmonics: We start in Appendix H.1 by considering the specific case of convolutions mapping from scalar inputs to general irrep output fields, observing that their kernel constraints are generally solved by harmonic basis functions.

endomorphisms: Appendix H.2 shows that these solutions may additionally be composed with irrep endomorphisms without violating the steerability constraint.

CG-coefficients: Appendix H.3 turns to the general constraint with arbitrary irreducible input and output field types V_I and V_J , and shows how it may be rewritten in terms of a tensor product $V_I \otimes V_J$. A Clebsch-Gordan decomposition allows to reduce this constraint into individual constraints on irrep subspaces. Each of those constraints corresponds to the ones considered in the previous two sections, which implies that their solutions in terms of harmonics and irrep endomorphisms can be combined via Clebsch-Gordan coefficients to give the general solution – this recovers the statement of the Wigner-Eckart theorem.

Note that this approach is constructivistic – it only shows the sufficiency of and intuition behind the constructions, but not their necessity, which is proven in [173] and [40]. To clarify the abstract constructions, we will give examples for different use cases throughout the appendix.

As in Section 5.3.3, we assume G throughout this appendix to be compact, which allows for the complete reducibility of finite dimensional G -representation into irreps and guarantees the existence of harmonic basis functions. X denotes again some G -orbit as described in Section 5.3.1. (ρ_j, V_j) denotes the real, unitary G -irrep of order j and $\dim_j := \dim V_j$ its dimensionality.

H.1 Harmonic kernels

Before turning to general G -steerable kernels that map between arbitrary irreducible input and output feature fields, i.e. satisfy Eq. (5.27), we consider the simpler case where only the output type (V_J, ρ_J) is an arbitrary irrep, while the input type (ρ_0, V_0) is assumed to be trivial for now. Appendix H.3 below will discuss the general case with arbitrary irrep input fields, reducing it to the current case via a Clebsch-Gordan decomposition.

Given a scalar input field, i.e. $\rho_{in} = \rho_0$, the kernel is a map $K^{J0} : X \rightarrow \mathbb{R}^{\dim_J}$, satisfying the constraint $K^{J0}(gx) = \rho_J(g) K^{J0}(x)$. Using the assumed unitarity (orthogonality) $\rho_J(g)^{-1} = \rho_J(g)^\top$, this constraint is equivalent to $K^{J0}(g^{-1}x) = \rho_J(g)^\top K^{J0}(x)$, or, in components,

$$K_m^{J0}(g^{-1}x) = \sum_n \rho_J(g)_{nm} K_n^{J0}(x) \quad \forall x \in X, g \in G, m = 1, \dots, \dim_J, \quad (\text{H.1})$$

where $K_m^{J0} : X \rightarrow \mathbb{R}$ are scalar functions on the orbit (homogeneous space) X .

We know from the Peter-Weyl theorem B.5.22 that square integrable functions on homogeneous spaces of compact groups G decompose into a direct sum of irreducible subspace components

$$L^2(X, \mathbb{R}) \cong \widehat{\bigoplus}_J \bigoplus_{i=1}^{m_J} V_{Ji}, \quad (\text{H.2})$$

where m_J is the multiplicity of the irreducible subspaces $V_{Ji} \cong \mathbb{R}^{\dim_J}$ that are acted on by irrep ρ_J . Each subspace V_{Ji} has a basis of *harmonic functions* $\{Y_{Ji}^m \mid m = 1, \dots, \dim_J\}$, transforming according to irrep ρ_J as $Y_{Ji}^m(g^{-1}x) = \sum_n \rho_J(g)_{nm} Y_{Ji}^n(x)$. Noting that this defining equation of the harmonic basis agrees exactly with the kernel constraint in Eq. (H.1), we see that the vector valued functions

$$\vec{Y}_{Ji} := (Y_{Ji}^1, \dots, Y_{Ji}^{\dim_J})^\top : X \rightarrow \mathbb{R}^{\dim_J} \quad (\text{H.3})$$

constitute m_J solutions. These solutions do, however, not span the full G -steerable kernel space yet, but may additionally be modulated via irrep endomorphisms, as discussed in the next Appendix H.2.

Examples: Fig. 4.1 visualizes a mapping from a scalar field ($l = 0$) to a frequency $J = 3$ irrep field, as discussed in this paragraph, for $G = SO(2)$. It requires a G -steerable kernel whose angular part is given by frequency $j = 3$ circular harmonics, for instance

$$\vec{Y}_3^{SO(2)}(\phi) = (\cos(3\phi), \sin(3\phi))^\top. \quad (\text{H.4})$$

Another example are steerable convolutions between input and output scalar fields, requiring G -invariant (constant) kernels, i.e. $K^{00}(gx) = K^{00}(x)$. This result is expected, since scalar fields cannot encode any information about the G -pose of patterns of features, and invariant kernels cannot detect such poses. Both examples have $m_J = 1$, i.e. there is only a single harmonic subspace associated to irrep ρ_J .

H.2 Irrep endomorphisms

While the harmonic basis functions in Eq. (H.3) are solutions of the kernel constraint, they are not necessarily spanning the complete solution space yet. As proven by Lang and Weiler

[173], the complete solution space allows for a postcomposition with *irrep endomorphisms* (Def. B.5.9), that is, linear maps $c_J \in \mathbb{R}^{\dim J \times \dim J}$ that commute with the irrep action:

$$c_J \cdot \rho_J(g) = \rho_J(g) \cdot c_J \quad \forall g \in G \quad (\text{H.5})$$

To make this result plausible, observe that

$$c_J \cdot K^{J0}(gx) = c_J \cdot \rho_J(g) K^{J0}(x) = \rho_J(g) \cdot c_J \cdot K^{J0}(x) \quad (\text{H.6})$$

holds for arbitrary points $x \in X$, group elements $g \in G$, endomorphism $c_J \in \text{End}(V_J)$ and G -steerable kernels K^{J0} , implying that $c_J \cdot K^{J0}$ is G -steerable if K^{J0} is. The Wigner-Eckart theorem 5.3.1 proves that such endomorphism modulated harmonics are indeed making up the *complete* space of steerable kernels K^{J0} .¹ More specifically, let

$$\{c_{Jr} \mid r = 1, \dots, \dim \text{End}(V_J)\} \quad (\text{H.7})$$

be a basis of the irrep endomorphism space $\text{End}(V_J)$, then a steerable basis is given by

$$\begin{aligned} \mathcal{K}_{\rho_0, \rho_J}^{G, X} &:= \{K^{J0}: X \rightarrow \mathbb{R}^{\dim J} \mid K^{J0}(gx) = \rho_J(g) \cdot K^{J0}(x) \quad \forall x \in X, g \in G\} \\ &= \text{span} \{c_{Jr} \cdot \vec{Y}_i \mid i = 1, \dots, m_J, r = 1, \dots, \dim \text{End}(V_J)\}, \end{aligned} \quad (\text{H.8})$$

i.e. all m_J harmonics of order J , modulated by the $\dim \text{End}(V_J)$ basis endomorphisms.

The dimensionality of $\text{End}(V_J)$ is for real irreps V_J either 1, 2 or 4, depending on whether the irrep is of real, complex or quaternionic type; see [25].

Example: To give an example, consider $G = \text{SO}(2)$, whose real irreps are the one-dimensional trivial representation $\rho_0(\phi) = (1)$ and the two-dimensional frequency J rotation matrices

$$\rho_J(\phi) = \begin{pmatrix} \cos(J\phi) & -\sin(J\phi) \\ \sin(J\phi) & \cos(J\phi) \end{pmatrix} \quad J \in \mathbb{N}. \quad (\text{H.9})$$

The latter are irreps of complex type and have a two-dimensional endomorphism space spanned by

$$\text{End}(V_J) = \text{span} \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \right) \quad J \in \mathbb{N}, \quad (\text{H.10})$$

as is easily checked. $\text{SE}(2)$ -equivariant linear maps from scalar fields to order $J \geq 1$ irrep fields are therefore convolutions with $\text{SO}(2)$ -steerable kernels, whose angular parts are given by a linear combination of

$$\begin{pmatrix} \cos(J\phi) \\ \sin(J\phi) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(J\phi) \\ \sin(J\phi) \end{pmatrix} \quad \text{e.g., for } J = 2, \quad \left(\begin{array}{c} \text{color wheel} \\ \text{color wheel} \end{array} \right)^\top$$

and

$$\begin{pmatrix} -\sin(J\phi) \\ \cos(J\phi) \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \cos(J\phi) \\ \sin(J\phi) \end{pmatrix} \quad \text{e.g., for } J = 2, \quad \left(\begin{array}{c} \text{color wheel} \\ \text{color wheel} \end{array} \right)^\top.$$

¹Here we are still assuming scalar input fields and general irrep output fields, i.e. $\rho_{\text{in}} = \rho_0$ and $\rho_{\text{out}} = \rho_J$.

These are a pair of circular harmonics, and another pair that is phase-shifted by $\pi/2$ (or a spatial angle of $\pi/(2J)$) relative to the first one. A linear combination of these basis elements allows to express any phase shifted pair of circular harmonics.

If we would have considered $G = O(2)$ instead of $SO(2)$, the endomorphism space would have been one-dimensional, thus resulting in a one-dimensional steerable kernel basis. These solutions for $SO(2)$ and $O(2)$ are shown in the bottom left entries of Tables 5.2 and 5.3, respectively.

H.3 General irrep steerable kernels and Clebsch Gordan coefficients

The case of general steerable kernels between irreducible input and output feature fields is via a *Clebsch Gordan decomposition* (Def. B.5.17) reduced to the previous case. To see this, we vectorize the irrep kernel constraint, Eq. (5.27), which yields²

$$\text{vec } K^{Jl}(gx) = (\rho_l^{-\top} \otimes \rho_J)(g) \text{vec } K^{Jl}(x) \quad \forall x \in X, g \in G, \quad (\text{H.11})$$

Since ρ_l is unitary, we have $\rho_l^{-1} = \rho_l^\top$, resulting in a constraint involving the *irrep tensor product* $\rho_l \otimes \rho_J$. Such tensor products of irreducible representations are in general not irreducible. Their reduction to irreducible subspaces is known as Clebsch-Gordan decomposition, which is an isomorphism

$$\text{CG}_{lJ} : V_l \otimes V_J \rightarrow \bigoplus_{j \in \widehat{G}} \bigoplus_{s=1}^{m_{j,lJ}} V_j, \quad (\text{H.12})$$

where the first summand is over all (isomorphism classes of) irreps, while the second sum determines the (potentially zero) multiplicity $m_{j,lJ} \in \mathbb{N}_0$ with which irrep ρ_j occurs in the tensor product $\rho_l \otimes \rho_J$. Using it to decompose $\text{CG}_{lJ}(\rho_l \otimes \rho_J) \text{CG}_{lJ}^{-1} = \bigoplus_{j \in \widehat{G}} \bigoplus_{s=1}^{m_{j,lJ}} \rho_j$, the general irrep kernel constraint in Eq. (H.11) implies

$$\begin{aligned} \text{CG}_{lJ} \text{vec } K^{Jl}(gx) &= \text{CG}_{lJ}(\rho_l^{-\top} \otimes \rho_J)(g) [\text{CG}_{lJ}^{-1} \text{CG}_{lJ}] \text{vec } K^{Jl}(x) \\ \iff K^{Jl}(gx) &= \bigoplus_{j \in \widehat{G}} \bigoplus_{s=1}^{m_{j,lJ}} \rho_j(g) \bar{K}^{Jl}(x) \end{aligned} \quad (\text{H.13})$$

for any $x \in X$ and $g \in G$, where $\bar{K}^{Jl} := \text{CG}_{lJ} \text{vec } K^{Jl}$. The steerable kernel consists therefore of $m_{j,lJ}$ subspaces that transform according to ρ_j . As discussed in the previous two sections, each of these subspaces has a basis described by harmonics of order j and their irrep endomorphisms c_{jr} , $r = 1, \dots, \dim \text{End}(V_j)$. The complete basis of steerable kernels is therefore spanned by

$$K_{srji}^{Jl} := \text{unvec } \text{CG}_{lJ,js}^+ c_{jr} \vec{Y}_{ji} \quad j \in \widehat{G}, s \leq m_{j,lJ}, i \leq m_j, r \leq \dim \text{End}(V_j) \quad (\text{H.14})$$

where $\text{CG}_{lJ,js} := \text{proj}_{js} \circ \text{CG}_{lJ}$ is the Clebsch-Gordan decomposition in Eq. (H.12), followed by a projection on the s -th irreducible subspace V_j , such that the pseudoinverse

²The *vectorization operator* $\text{vec} : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \cdot n}$ acts on an $m \times n$ -matrix by stacking its columns into an $m \cdot n$ -vector. It satisfies $\text{vec}(AXB) = (B^\top \otimes A) \text{vec}(X)$ for any triple A, X, B of dimensionally matching matrices [229].

$\text{CG}_{lJ,js}^+$ is the embedding of this subspace V_j into $V_l \otimes V_J$. This is exactly the statement of the Wigner-Eckart theorem 5.3.1 for G -steerable kernels.

Given the harmonics, irrep endomorphisms and Clebsch-Gordan coefficients of the structure group G , this result allows us to construct the most general G -steerable kernels for mappings between irrep fields. An additional change of basis by Q_{in} and Q_{out} (Section 5.3.2) turns this into steerable kernel bases for arbitrary field types ρ_{in} and ρ_{out} .

Example: As a simple example, let us consider the case of irrep fields for $G = \text{SO}(3)$. The irreps of $\text{SO}(3)$ are Wigner D-matrices $D_j : \text{SO}(3) \rightarrow \text{GL}(V_j)$, acting on $V_j \cong \mathbb{R}^{2j+1}$, where $j \in \mathbb{N}_0$. They are of real type, implying that the endomorphism spaces are one-dimensional, and can therefore be ignored. The (non-trivial) orbits are 2-spheres, such that $\vec{Y}_j : S^2 \rightarrow \mathbb{R}^{2j+1}$ is a vector of $2j + 1$ spherical harmonics as shown in the rows of Fig. 5.3. Each harmonic subspace in the Peter-Weyl decomposition of $L^2(S^2, \mathbb{R})$ appears with multiplicity $m_J = 1$, such that we may ignore the index i . The Clebsch-Gordan decomposition of $V_l \otimes V_J$ contains all of the $2 \min(l, J) + 1$ irreps V_j with indices $|l - J| \leq j \leq l + J$ with multiplicity $m_{j,l,J} = 1$ and all other irreps with multiplicity zero – we can therefore restrict to this range of indices and drop the index s as well. What remains is the basis

$$\mathcal{K}_{\rho_l, \rho_J}^{\text{SO}(3)} = \text{span} \{ K_j := \text{unvec } \text{CG}_{lJ,j}^+ \vec{Y}_j \mid |l - J| \leq j \leq l + J \}, \quad (\text{H.15})$$

of (angular parts of) $\text{SO}(3)$ -irrep steerable kernels; as originally derived by Weiler et al. [323]. A general $\text{SO}(3)$ -irrep steerable kernel is expanded in this basis, with one learnable parameter per harmonic component j and radial shell (or alternative parametrization of the radial part). Fig. 5.4 visualizes different input and output irrep orders l and J and the valid harmonic components j mapping between them. The similarity to the *selection rules* for quantum state transitions in the hydrogen atom are no coincidence, since the hydrogen atom's potential is $\text{SO}(3)$ -invariant, implying that its quantum states are irreducible representations of $\text{SO}(3)$.

Existence and smoothness of kernel field transforms

In Def. 12.2.5 we proposed *kernel field transforms* $\mathcal{T}_{\mathcal{K}}$ as smooth integral transforms

$$\mathcal{T}_{\mathcal{K}} : \Gamma(\mathcal{A}_{\text{in}}) \rightarrow \Gamma(\mathcal{A}_{\text{out}}) \quad (\text{I.1})$$

which are parameterized by some kernel field \mathcal{K} (Def. 12.2.1) and are pointwise given by

$$[\mathcal{T}_{\mathcal{K}}(f)](p) := \int_{T_p M} \mathcal{K}(v) \text{Exp}_p^* f(v) dv = \int_{T_p M} \mathcal{K}(v) \mathcal{P}_{\mathcal{A}_{\text{in}}, p \leftarrow \exp_p v} f(\exp_p v) dv. \quad (\text{I.2})$$

Kernel field transforms include *GM-convolutions* from Def. 12.2.7 as a special cases for *GM-convolutional kernel fields*.

Here we briefly discuss the well-definedness of kernel field transforms. It is clear that the integrand of Eq. (I.2) lies for any $p \in M$ and $v \in T_p M$ in $\mathcal{A}_{\text{out}, p}$. What remains to be shown is the *existence* of the integral and the *smoothness* of the resulting feature field. In the following we will first give some general remarks on how to approach these questions. We will then prove Theorem 12.2.6, i.e. the well-definedness of kernel field transforms for the specific case of fields of kernels which are compactly supported on a ball of fixed radius around the origin.

Existence: The existence and smoothness of kernel field transforms requires a suitable choice of kernel field \mathcal{K} . Similar to the case of conventional convolutions on $M = \mathbb{R}$, the requirements on \mathcal{K} in order for the kernel field transform to exist depend on the specific properties of the input feature field $f \in \Gamma(\mathcal{A}_{\text{in}})$.¹ In general, \mathcal{K} needs to decay sufficiently rapidly in order to make the integrand in Eq. (I.2) integrable.

A special case of great practical importance is that of kernels $\mathcal{K}_p : T_p M \rightarrow \text{Hom}(\mathcal{A}_{\text{in}, p}, \mathcal{A}_{\text{out}, p})$ which are at any $p \in M$ *compactly supported*. In this case the integral is always guaranteed to exist. To see this, note that (input) feature fields and kernel fields are defined to be smooth. The smoothness of the metric further implies that the Riemannian volume density, the exponential map and the parallel transport are smooth [100]. In combination, the whole integrand in Eq. (I.2) is seen to be a smooth and thus continuous function from $T M$ to \mathcal{A}_{out} . If \mathcal{K}_p is in addition compactly supported, the integrand becomes continuous and compactly supported on $T_p M$, which, by a generalization of the extreme

¹See the discussion at https://en.wikipedia.org/wiki/Convolution#Domain_of_definition.

value theorem, implies that its image is compact (and in a local trivialization \mathbb{R}^c of $\mathcal{A}_{\text{out},p}$ bounded) [252]. This guarantees the existence of the integral [93, 282].

Depending on the application, the requirement on the support of \mathcal{K} might be relaxed. For instance, images on $M = \mathbb{R}^d$ are usually compactly supported themselves, such that no additional properties of \mathcal{K} except for its smoothness are required.

Smoothness: We turn to discuss the smoothness of kernel field transforms, that is, their property to map smooth input fields $f_{\text{in}} \in \Gamma(\mathcal{A}_{\text{in}})$ to smooth output fields $f_{\text{out}} := \mathcal{T}_{\mathcal{K}}(f_{\text{in}}) \in \Gamma(\mathcal{A}_{\text{out}})$. By definition, a map $f_{\text{out}} : M \rightarrow \mathcal{A}_{\text{out}}$ between manifolds M and \mathcal{A}_{out} is said to be smooth if its coordinate representations are smooth. In equations, f_{out} is smooth if for any $p \in M$ there exist smooth charts (U, ϕ) about p in M and $(\tilde{U}, \tilde{\phi})$ about $f_{\text{out}}(p)$ in \mathcal{A}_{out} with $f_{\text{out}}(U) \subseteq \tilde{U}$ such that $\tilde{\phi} \circ f_{\text{out}} \circ \phi^{-1} : \phi(U) \rightarrow \tilde{\phi}(\tilde{U})$ is smooth as a map between (subsets of) Euclidean spaces. Given (U, ϕ) , a convenient choice for $(\tilde{U}, \tilde{\phi})$ would be $\tilde{\phi} := (\phi \times \text{id}) \circ \Psi_{\mathcal{A}_{\text{out}}} : \pi_{\mathcal{A}_{\text{out}}}^{-1}(U) \mapsto \phi(U) \times \mathbb{R}^c \subseteq \mathbb{R}^d \times \mathbb{R}^c$, however, the following discussion is independent from this choice.² A map between (subsets of) Euclidean spaces is smooth if it is smooth in each component of its image, here in each of the $d + c$ dimensions of $\phi(U) \times \mathbb{R}^c$. We are therefore interested in the smoothness of the maps

$$F_i : \phi(U) \rightarrow \mathbb{R}, \quad x \mapsto \left[\tilde{\phi} \circ f_{\text{out}} \circ \phi^{-1} \right]_i \quad (\text{I.3})$$

for any $i = 1, \dots, d + c$. By writing out f_{out} and expressing the integral over $T_p M$ by an integral over \mathbb{R}^d as discussed in Appendix C.4, the F_i are seen to be of the form

$$F_i(x) = \int_{\mathbb{R}^d} I_i(v, x) \, dv. \quad (\text{I.4})$$

The coordinate expressions of the integrands I_i are hereby for any $i = 1, \dots, d + c$ given by

$$\begin{aligned} I_i &: \mathbb{R}^d \times \phi(U) \rightarrow \mathbb{R}, \quad (v, x) \mapsto \\ &\left[\tilde{\phi} \circ \mathcal{K}(\psi_{TM, \phi^{-1}(x)}^{-1}(v)) \circ \mathcal{P}_{\mathcal{A}_{\text{in}}, \phi^{-1}(x)} \leftarrow \exp \circ \psi_{TM, \phi^{-1}(x)}(v) \circ f_{\text{in}} \circ \exp \circ \psi_{TM, \phi^{-1}(x)}^{-1}(v) \right]_i, \end{aligned} \quad (\text{I.5})$$

where we assumed, for convenience and without loss of generality, that $\psi_{TM, \phi^{-1}(x)}$ is an *isometric* gauge of $T_{\phi^{-1}(x)} M$, such that the volume scaling factor $\sqrt{|\det \eta_p|} = 1$ drops out. Note that the integrands I_i are composed of smooth maps and are therefore smooth as well.

From the previous discussion it is clear that the smoothness of f_{out} holds if all F_i are smooth, i.e. infinitely often partially differentiable. To prove the smoothness of the F_i , it is sufficient to show that the partial differentiations and the integration in Eq. (I.4) commute – which is not always the case. If they do commute, partial derivatives of arbitrary orders $(n_1, \dots, n_d) \in \mathbb{N}^d$ are given by

$$\left[\partial_{x_1}^{n_1} \dots \partial_{x_d}^{n_d} F_i \right](x) = \int_{\mathbb{R}^d} \left[\partial_{x_1}^{n_1} \dots \partial_{x_d}^{n_d} I_i \right](v, x) \, dv \quad (\text{I.6})$$

²Note that $\pi_{\mathcal{A}_{\text{out}}}^{-1}(U)$ is guaranteed to be trivializable given that (ϕ, U) is a chart of M . This is clear since the coordinate bases $\left[\frac{\partial}{\partial \phi_\mu} \right]_{\mu=1}^d$ of (ϕ, U) yields a trivialization of $\pi_{TM}^{-1}(U)$ (see Appendix C) and since the local trivializations of TM and \mathcal{A} were in Section 11.4 induced from those of M .

where 1) the partial derivatives $\left[\partial_{x_1}^{n_1} \dots \partial_{x_d}^{n_d} I_i \right]$ of the integrand exist (due to the smoothness of I_i their existence is guaranteed) and 2) their integral exists. Whether or not the differentiations commute with the integral can be investigated by making use of the following lemma from [93],³ which is a consequence of the dominated convergence theorem.

Theorem I.0.1 (Differentiation lemma [93]). *Let \mathcal{V} be a measure space, let $T \subset \mathbb{R}$ be a non-degenerate interval and let $I : \mathcal{V} \times T \rightarrow \mathbb{R}$ be a map with the following properties:*

- (i) *For any fixed $t \in T$ the map $v \mapsto I(v, t)$ is Lebesgue integrable on \mathcal{V}*
- (ii) *For any fixed $v \in \mathcal{V}$ the map $t \mapsto I(v, t)$ is differentiable in T*
- (iii) *There exists a Lebesgue integrable function $\mathcal{B} : \mathcal{V} \rightarrow \mathbb{R}$ such that $\left| \frac{\partial}{\partial t} I(v, t) \right| \leq \mathcal{B}(v)$ for any $(v, t) \in \mathcal{V} \times T$*

Then the function $F : T \rightarrow \mathbb{R}$, $t \mapsto \int_{\mathcal{V}} I(v, t) dv$ is differentiable with derivative

$$\frac{\partial}{\partial t} F(t) = \int_{\mathcal{V}} \frac{\partial}{\partial t} I(v, t) dv.$$

The applicability of this lemma (repeatedly for every single partial differentiation) depends on the properties of the integrand, which in turn depends on the specific properties of the kernel field \mathcal{K} and the input feature field f_{in} . For the case of a kernel field which is compactly supported on balls of fixed radius around the origin of each tangent space, the lemma applies. Based on this, we give a proof of Theorem 12.2.6 in the remainder of this appendix.

Proof of Theorem 12.2.6 – Kernel field transform existence for compactly supported kernels

Denote by $B_{T_p M}^{\text{closed}}(0, R) := \{v \in T_p M \mid \|v\| \leq R\}$ the closed ball of radius $R > 0$ around the origin of $T_p M$ and by $B_{\mathbb{R}^d}^{\text{closed}}(0, R) := \{v \in \mathbb{R}^d \mid \|v\| \leq R\}$ the corresponding ball around the origin of \mathbb{R}^d . Note that any isometric gauge satisfies $\psi_{TM,p}(B_{T_p M}^{\text{closed}}(0, R)) = B_{\mathbb{R}^d}^{\text{closed}}(0, R)$. Let $\tilde{\mathcal{K}}$ be a kernel field whose support falls within balls of the same radius R in each tangent space, i.e. which satisfies

$$\text{supp}(\tilde{\mathcal{K}}_p) \subseteq B_{T_p M}^{\text{closed}}(0, R) \quad \forall p \in M \tag{I.7}$$

and thus, for any isometric gauge $\psi_{TM,p}$:

$$\text{supp}(\tilde{\mathcal{K}}_p \circ (\psi_{TM,p})^{-1}) \subseteq B_{\mathbb{R}^d}^{\text{closed}}(0, R) \quad \forall p \in M \tag{I.8}$$

According to Theorem 12.2.6 this property is sufficient to guarantee that the corresponding kernel field transform $\mathcal{T}_{\mathcal{K}_R}$ is well defined. A proof of this statement is given in the following.

³Similar versions of this lemma in English language can be found in [156] or at https://en.wikipedia.org/wiki/Leibniz_integral_rule#Measure_theory_statement. In contrast to those versions, the version from [93] allows for T being any non-degenerate interval, including closed intervals, which saves us some additional steps below.

Proof: As already stated in the beginning of this appendix, the existence of the integral is guaranteed given that the kernel supports are compact: The compactness of the kernels carries over to the integrands of the kernel field transform. Their smoothness further implies their continuity and integrals of compactly supported continuous functions always exists.

To prove the smoothness of the resulting output feature field f_{out} , we proceed with the discussion earlier in this section. We aim to apply the differentiation lemma I.0.1 to swap partial derivatives $\frac{\partial}{\partial x_\mu}$ for any $\mu = 1, \dots, d$ in Eq. (I.6) at any $x_0 \in \phi(U)$ with the integration over \mathbb{R}^d . For this purpose, we introduce the auxiliary functions

$$I_{i,x_0,\mu} : \mathbb{R}^d \times [-\varepsilon, \varepsilon] \rightarrow \mathbb{R}, \quad (\vartheta, t) \mapsto I_i(\vartheta, x_0 + t\epsilon_\mu) \quad (\text{I.9})$$

and

$$F_{i,x_0,\mu} : [-\varepsilon, \varepsilon] \rightarrow \mathbb{R}, \quad t \mapsto F(x_0 + t\epsilon_\mu) = \int_{\mathbb{R}^d} I_{i,x_0,\mu}(\vartheta, t) d\vartheta, \quad (\text{I.10})$$

where $\epsilon_\mu \in \mathbb{R}^d$ is the unit vector in μ -direction and $\varepsilon > 0$ is chosen such that $\{x_0 + t\epsilon_\mu \mid t \in [-\varepsilon, \varepsilon]\} \subset \phi(U)$, which is always possible since $\phi(U)$ is open. Then $I_{i,x_0,\mu}$ is with the identifications $\mathcal{V} = \mathbb{R}^d$ and $T = [-\varepsilon, \varepsilon]$ of the form required by lemma I.0.1. It satisfies property (i) by the assumption that the kernel field transform exists as discussed earlier. Property (ii) holds due to the smoothness of the full integrand in Eq. (I.5). For property (iii), observe that both $I_{i,x_0,\mu}$ and its derivative are smooth such that the absolute value $|\frac{\partial}{\partial t} I_{i,x_0,\mu}|$ is continuous. Since it is in addition compactly supported on $B_{\mathbb{R}^d}^{\text{closed}}(0, R) \times [-\varepsilon, \varepsilon]$, it is by (a generalization of) the extreme value theorem bounded by some number $b \geq 0$. We therefore set $\mathcal{B}(\vartheta) = b \cdot \mathbb{I}_{B_{\mathbb{R}^d}^{\text{closed}}(0, R) \times [-\varepsilon, \varepsilon]}$ where \mathbb{I} is the indicator function. This choice satisfies $|\frac{\partial}{\partial t} I_i(\vartheta, t)| \leq \mathcal{B}(\vartheta)$ for any $(\vartheta, t) \in \mathcal{V} \times T$ and is integrable such that property (iii) is fulfilled as well. We can therefore swap the order of differentiation and integration for arbitrary choices of x_0 and μ , which we use to pull arbitrary partial derivatives into the integral:

$$\begin{aligned} \left[\frac{\partial}{\partial x_\mu} F_i \right](x_0) &= \left[\frac{\partial}{\partial t} I_{i,x_0,\mu} \right](0) = \int_{\mathbb{R}^d} \frac{\partial}{\partial t} I_{i,x_0,\mu}(\vartheta, t) \Big|_{t=0} d\vartheta \\ &= \int_{\mathbb{R}^d} \frac{\partial}{\partial x_\mu} I_i(\vartheta, x) \Big|_{x=x_0} d\vartheta \end{aligned} \quad (\text{I.11})$$

Due to the smoothness and compact support of the integrand I_i , its partial derivatives $\frac{\partial}{\partial x_\mu} I$ are smooth and compactly supported as well. They do therefore satisfy properties (i), (ii) and (iii) as well (with a potentially adapted bound b). It is thus possible to repeat the partial differentiation of F_i infinitely often, which proves its smoothness. Since the derivations were independent from the particular choices for the point $p \in M$, charts (U, ϕ) and $(\tilde{U}, \tilde{\phi})$, points $x_0 \in \phi(U)$ and indices i and μ , this result proves the smoothness of the whole output feature field $f_{\text{out}} = \mathcal{T}_K(f_{\text{in}})$. □

Regular feature fields as scalar functions on G -structure

Real-valued functions $F : GM \rightarrow \mathbb{R}$ on the G -structure are equivalent to regular feature fields $f : M \rightarrow \mathcal{A}_{\text{reg}}$ on the manifold, that is, that there is an isomorphism

$$C^\infty(GM) \cong \Gamma(\mathcal{A}_{\text{reg}}). \quad (\text{J.1})$$

This appendix presents a proof of this claim for the case of *finite structure groups* G . We start with the usual definition of (real) regular representations of finite structure groups, which act on the (free) vector spaces $\mathbb{R}^{|G|}$. One defines a basis $\{\epsilon_g \in \mathbb{R}^{|G|} \mid g \in G\}$ of $\mathbb{R}^{|G|}$, which is labeled by the group elements $g \in G$. The (left) regular representation's action on $\mathbb{R}^{|G|}$ is then defined in terms of its action on these basis vectors, which is given by left translation. Specifically, for any $h, g \in G$, the regular representation acts as follows:

$$\rho_{\text{reg}}(h) \epsilon_g := \epsilon_{hg}. \quad (\text{J.2})$$

Note that the action on *coefficients* of a vector is inverse

$$\rho_{\text{reg}}(h) \sum_{g \in G} f_g \epsilon_g = \sum_{g \in G} f_g \epsilon_{hg} = \sum_{\tilde{g} \in G} f_{h^{-1}\tilde{g}} \epsilon_{\tilde{g}}, \quad (\text{J.3})$$

which is useful to know, however, we won't need this property in the following. As the regular representation permutes the basis vectors of $\mathbb{R}^{|G|}$, it is a *permutation representation*. Some visualizations for the cyclic group $G = C_4$ are found in Appendix B of [322]. Regular feature fields are defined as smooth sections of the associated G -bundle

$$\mathcal{A}_{\text{reg}} = (GM \times \mathbb{R}^{|G|}) / \sim_{\rho_{\text{reg}}}, \quad (\text{J.4})$$

as defined in Section 11.3.3.

The isomorphism $C^\infty(GM) \cong \Gamma(\mathcal{A}_{\text{reg}})$ substantiates our claim in Chapter 18 that the *Parallel Frame CNNs* by Yang et al. [339] are specific GM -convolutions between regular feature fields. It furthermore establishes the link between *group convolutions* (see Appendix F.1) and *regular GM-convolutions* that was claimed in Section 15.3 and [322]. A related result, stating that regular steerable convolutions on Euclidean spaces are group convolutions, was discussed in Section 4.5.

With these preparations and remarks we are ready to formulate and prove the theorem:

Theorem J.0.1 (Regular feature fields as scalar functions on G -structure).

Let $G \leq \mathrm{GL}(d)$ be a finite structure group, let GM be a G -structure over M and let $\mathcal{A}_{\mathrm{reg}}$ be the bundle that is associated by the action of the regular representation ρ_{reg} of G . Regular feature fields are then identical to smooth, real-valued functions on the G -structure, that is, there is an isomorphism

$$\Lambda : C^\infty(GM) \xrightarrow{\sim} \Gamma(\mathcal{A}_{\mathrm{reg}}). \quad (\text{J.5})$$

This isomorphism is defined by

$$[\Lambda F](p) = \left[[e_i]_{i=1}^d, \sum_g F([e_i]_{i=1}^d \triangleleft g) \epsilon_g \right], \quad (\text{J.6})$$

where $[e_i]_{i=1}^d \in G_p M$ is an arbitrarily chosen representative frame at p . Its inverse is given by

$$[\Lambda^{-1} f]([e_i]_{i=1}^d) = \left\langle \epsilon_e, \psi_{\mathcal{A}, p}^{[e_i]_{i=1}^d} f(p) \right\rangle, \quad (\text{J.7})$$

where we abbreviated $p = \pi_{GM}(E)$ and denote by $\psi_{\mathcal{A}, p}^{[e_i]_{i=1}^d}$ that (unique) gauge that corresponds to the frame $[e_i]_{i=1}^d$, i.e. which satisfies $\psi_{\mathcal{A}, p}^{[e_i]_{i=1}^d}([e_i]_{i=1}^d) = e$.

Proof: To prove this statement, we need to show that 1) the isomorphism preserves the smoothness of the maps, 2) that the choice of representative frame $[e_i]_{i=1}^d \in G_p M$ in the definition of Λ is indeed arbitrary and 3) that Λ^{-1} is indeed a left and right inverse of Λ .

1) smoothness :

That the isomorphism preserves the smoothness of the equivalent field representations is clear since all involved morphisms (right action, gauge map, inner product) are smooth.

2) independence of the definition of Λ , Eq. (J.6), from the choice of representative frame $[e_i]_{i=1}^d \in G_p M$:

Suppose that we used *any* other frame $[e_i]_{i=1}^d \triangleleft h$ for an arbitrary $h \in G$. This arbitrary gauge transformation drops then out by making use of the equivalence relation $\sim_{\rho_{\mathrm{reg}}}$ that is underlying the associated bundle construction, Eq. (J.5):

$$\begin{aligned} & [\Lambda F](p) \\ &= \left[[e_i]_{i=1}^d \triangleleft h, \sum_g F([e_i]_{i=1}^d \triangleleft hg) \epsilon_g \right] && (\text{Def. of } \Lambda, \text{Eq. (J.6)}) \\ &= \left[[e_i]_{i=1}^d, \rho_{\mathrm{reg}}(h) \sum_g F([e_i]_{i=1}^d \triangleleft hg) \epsilon_g \right] && (\text{equiv. relation } \sim_{\rho_{\mathrm{reg}}}, \text{Eq. (11.42)}) \\ &= \left[[e_i]_{i=1}^d, \sum_g F([e_i]_{i=1}^d \triangleleft hg) \epsilon_{hg} \right] && (\rho_{\mathrm{reg}} \text{ action on basis } \epsilon_g, \text{Eq. (J.2)}) \\ &= \left[[e_i]_{i=1}^d, \sum_{\tilde{g}} F([e_i]_{i=1}^d \triangleleft \tilde{g}) \epsilon_{\tilde{g}} \right] && (\text{substitution } \tilde{g} = hg) \end{aligned} \quad (\text{J.8})$$

3) Λ^{-1} in Eq. (J.7) is a well defined inverse of Λ in Eq. (J.6) :

3a) $\Lambda^{-1} \circ \Lambda = \text{id}_{C^\infty(GM)}$, that is, Λ^{-1} is a left inverse of Λ :

For any $F \in C^\infty(GM)$ and any $[e_i]_{i=1}^d$ this is shown as follows:

$$\begin{aligned}
& [\Lambda^{-1} \circ \Lambda F]([e_i]_{i=1}^d) \\
&= \left\langle \epsilon_e, \psi_{A,p}^{[e_i]_{i=1}^d} [\Lambda F](p) \right\rangle && (\text{Def. of } \Lambda^{-1}, \text{Eq. (J.7)}) \\
&= \left\langle \epsilon_e, \psi_{A,p}^{[e_i]_{i=1}^d} [[e_i]_{i=1}^d, \sum_g F([e_i]_{i=1}^d \triangleleft g) \epsilon_g] \right\rangle && (\text{Def. of } \Lambda, \text{Eq. (J.6)}) \\
&= \left\langle \epsilon_e, \sum_g F([e_i]_{i=1}^d \triangleleft g) \epsilon_g \right\rangle && (\text{Def. of } \psi_{A,p}, \text{Eq. (11.66)}) \\
&= \sum_g F([e_i]_{i=1}^d \triangleleft g) \langle \epsilon_e, \epsilon_g \rangle && (\text{pull inner product into sum}) \\
&= F([e_i]_{i=1}^d) && (\text{Kronecker delta } \delta_{e,g} = \langle \epsilon_e, \epsilon_g \rangle)
\end{aligned} \tag{J.9}$$

3b) $\Lambda \circ \Lambda^{-1} = \text{id}_{\Gamma(\mathcal{A}_{\text{reg}})}$, that is, Λ^{-1} is a right inverse of Λ :

Let $f \in \Gamma(\mathcal{A}_{\text{reg}})$ and $p \in M$, then:

$$\begin{aligned}
& [\Lambda \circ \Lambda^{-1} f](p) \\
&= \left[[e_i]_{i=1}^d, \sum_g [\Lambda^{-1} f]([e_i]_{i=1}^d \triangleleft g) \epsilon_g \right] && (\text{Def. of } \Lambda, \text{Eq. (J.6)}) \\
&= \left[[e_i]_{i=1}^d, \sum_g \left\langle \epsilon_e, \psi_{A,p}^{[e_i]_{i=1}^d \triangleleft g} f(p) \right\rangle \epsilon_g \right] && (\text{Def. of } \Lambda^{-1}, \text{Eq. (J.7)}) \\
&= \left[[e_i]_{i=1}^d, \sum_g \left\langle \epsilon_e, \rho_{\text{reg}}(g)^{-1} \psi_{A,p}^{[e_i]_{i=1}^d} f(p) \right\rangle \epsilon_g \right] && (\text{gauge trafo, Eq. (11.69)}) \\
&= \left[[e_i]_{i=1}^d, \sum_g \left\langle \rho_{\text{reg}}(g) \epsilon_e, \psi_{A,p}^{[e_i]_{i=1}^d} f(p) \right\rangle \epsilon_g \right] && (\text{unitarity of } \rho_{\text{reg}}) \\
&= \left[[e_i]_{i=1}^d, \sum_g \left\langle \epsilon_g, \psi_{A,p}^{[e_i]_{i=1}^d} f(p) \right\rangle \epsilon_g \right] && (\rho_{\text{reg}} \text{ action on basis } \epsilon_e, \text{Eq. (J.2)}) \\
&= \left[[e_i]_{i=1}^d, \psi_{A,p}^{[e_i]_{i=1}^d} f(p) \right] && (\text{remove expansion in basis } \epsilon_g) \\
&= f(p) && (\text{Def. of } \psi_{A,p}, \text{Eq. (11.66)})
\end{aligned} \tag{J.10}$$

This concludes our prove of the equivalence of $C^\infty(GM)$ and $\Gamma(\mathcal{A}_{\text{reg}})$. \square

Quotient representative kernel fields – proofs

In this appendix we give proofs for Theorems 13.3.1 and 13.3.3.

K.1 Proof of Theorem 13.3.1 – Isomorphism between isometry invariant and quotient representative kernel fields

Theorem 13.3.1 claims that the spaces $\mathcal{K}_{\text{invar}}^{\mathcal{I}}$ of isometry invariant kernel fields in Eq. (13.53) and $\mathcal{K}_{\text{quot}}^{\mathcal{I}}$ of quotient representative kernel fields in Eq. (13.93) are isomorphic to each other and that the isomorphism is given by the lift Λ whose inverse Λ^{-1} is the restriction to $r_{TM}(\mathcal{I}\setminus TM)$. Here we present a proof for this statement which consists of showing that 1) Λ^{-1} is indeed an inverse of Λ , 2) the defining properties of $\mathcal{K}_{\text{invar}}^{\mathcal{I}}$ and $\mathcal{K}_{\text{quot}}^{\mathcal{I}}$ are satisfied after lifting and restricting and 3) the constructions do not depend on arbitrary choices.

1) Λ^{-1} in Eq. (13.95) is a well defined inverse of Λ in Eq. (13.94) :

1a) $\Lambda \circ \Lambda^{-1} = \text{id}_{\mathcal{K}_{\text{invar}}^{\mathcal{I}}}$, that is, Λ^{-1} is a right inverse of Λ :

This claim follows for any $\mathcal{K} \in \mathcal{K}_{\text{invar}}^{\mathcal{I}}$ and any $v \in TM$ from

$$\begin{aligned} [\Lambda \circ \Lambda^{-1}(\mathcal{K})](v) &= [\Lambda(\mathcal{K}|_{r_{TM}(\mathcal{I}\setminus TM)})](v) \\ &= \Phi_{r_{TM}}(v)_{*,\text{Hom}} \mathcal{K}|_{r_{TM}(\mathcal{I}\setminus TM)} r_{TM} Q_{TM}(v) \\ &= \Phi_{r_{TM}}(v)_{*,\text{Hom}} \mathcal{K} r_{TM} Q_{TM}(v) \\ &= \mathcal{K} \Phi_{r_{TM}}(v)_{*,TM} r_{TM} Q_{TM}(v) \\ &= \mathcal{K}(v), \end{aligned} \tag{K.1}$$

where the invariance (equivariance) of the kernel field in Eq. (13.54) allowed to swap the order of the isometry action and the evaluation of the kernel field in the penultimate step.

1b) $\Lambda^{-1} \circ \Lambda = \text{id}_{\mathcal{K}_{\text{quot}}^{\mathcal{I}}}$, that is, Λ^{-1} is a left inverse of Λ :

Let $\mathcal{Q} \in \mathcal{K}_{\text{quot}}^{\mathcal{I}}$ and $w \in r_{TM}(\mathcal{I}\setminus TM)$. Note that $r_{TM} Q_{TM}(w) = w$ since w is an orbit representative. Furthermore, since $w = \Phi_{r_{TM}}(w)_{*,TM} r_{TM} Q_{TM}(w) = \Phi_{r_{TM}}(w)_{*,TM}(w)$

it follows that $\Phi_{r_{TM}}(w) \in \text{Stab}_w$ such that, by the constraint in Eq. (13.93), $\Phi_{r_{TM}}(w)_{*,\text{Hom}} \mathcal{Q}(w) = \mathcal{Q}(w)$. Together, this proves the claim:

$$\begin{aligned} [\Lambda^{-1} \circ \Lambda(\mathcal{Q})](w) &= \Lambda(\mathcal{Q})|_{r_{TM}(\mathcal{I} \setminus TM)}(w) \\ &= \Lambda(\mathcal{Q})(w) \\ &= \Phi_{r_{TM}}(w)_{*,\text{Hom}} \mathcal{Q} r_{TM} Q_{TM}(w) \\ &= \Phi_{r_{TM}}(w)_{*,\text{Hom}} \mathcal{Q}(w) \\ &= \mathcal{Q}(w) \end{aligned} \quad (\text{K.2})$$

2) The defining properties of $\mathcal{K}_{\text{invar}}^{\mathcal{I}}$ and $\mathcal{K}_{\text{quot}}^{\mathcal{I}}$ are satisfied after lifting and restricting :

- 2a) $\pi_{\text{Hom}} \circ \Lambda(\mathcal{Q}) = \pi_{TM}$ for any $\mathcal{Q} \in \mathcal{K}_{\text{quot}}^{\mathcal{I}}$, that is, the lift $\Lambda(\mathcal{Q})$ is a bundle M -morphism:
For any $\mathcal{Q} \in \mathcal{K}_{\text{quot}}^{\mathcal{I}}$ and for any $v \in TM$ this claim follows from

$$\begin{aligned} [\pi_{\text{Hom}} \Lambda(\mathcal{Q})](v) &= \pi_{\text{Hom}} \Phi_{r_{TM}}(v)_{*,\text{Hom}} \mathcal{Q} r_{TM} Q_{TM}(v) \\ &= \Phi_{r_{TM}}(v) \pi_{\text{Hom}} \mathcal{Q} r_{TM} Q_{TM}(v) \\ &= \Phi_{r_{TM}}(v) \pi_{TM} r_{TM} Q_{TM}(v) \\ &= \Phi_{r_{TM}}(v) r_M \pi_{\mathcal{I} \setminus TM} Q_{TM}(v) \\ &= \Phi_{r_{TM}}(v) r_M Q_M \pi_{TM}(v) \\ &= \pi_{TM}(v), \end{aligned} \quad (\text{K.3})$$

where the last step made use of Eq. (13.87).

- 2b) $\pi_{\text{Hom}} \circ \Lambda^{-1}(\mathcal{K}) = \pi_{TM}$ for any $\mathcal{K} \in \mathcal{K}_{\text{invar}}^{\mathcal{I}}$, that is, $\Lambda^{-1}(\mathcal{K})$ is a bundle $r_M(\mathcal{I} \setminus M)$ -morphism :

This property follows immediately from the corresponding property of \mathcal{K} after restricting to $r_{TM}(\mathcal{I} \setminus TM) \subseteq \pi_{TM}^{-1}(r_M(\mathcal{I} \setminus M))$. For any $w \in r_{TM}(\mathcal{I} \setminus TM)$:

$$\begin{aligned} \pi_{\text{Hom}} [\Lambda^{-1}(\mathcal{K})](w) &= \pi_{\text{Hom}} \mathcal{K}|_{r_{TM}(\mathcal{I} \setminus TM)}(w) \\ &= \pi_{\text{Hom}} \mathcal{K}(w) \\ &= \pi_{TM}(w) \end{aligned} \quad (\text{K.4})$$

(K.5)

- 2c) $\phi_{*,\text{Hom}} \Lambda(\mathcal{Q}) \phi_{*,TM}^{-1} = \Lambda(\mathcal{Q}) \quad \forall \phi \in \mathcal{I}$, that is, $\Lambda(\mathcal{Q})$ satisfies the full isometry invariance constraint :

Let $v \in TM$ and $\phi \in \mathcal{I}$. Due to the invariance of the quotient map Q_{TM} under isometries we have $Q_{TM}(\phi_{*,TM}^{-1} v) = Q_{TM}(v)$. Note further that

$$\begin{aligned} &[\Phi_{r_{TM}}(v)^{-1} \phi \Phi_{r_{TM}}(\phi_{*,TM}^{-1} v)]_{*,TM} r_{TM} Q_{TM}(v) \\ &= [\Phi_{r_{TM}}(v)^{-1} \phi \Phi_{r_{TM}}(\phi_{*,TM}^{-1} v)]_{*,TM} r_{TM} Q_{TM}(\phi_{*,TM}^{-1} v) \\ &= [\Phi_{r_{TM}}(v)^{-1} \phi]_{*,TM} \phi_{*,TM}^{-1} v \\ &= \Phi_{r_{TM}}(v)_{*,TM}^{-1} v \\ &= r_{TM} Q_{TM}(v) \end{aligned} \quad (\text{K.6})$$

implies

$$[\Phi_{r_{TM}}(v)^{-1} \phi \Phi_{r_{TM}}(\phi_{*,TM}^{-1} v)] \in \text{Stab}_{r_{TM} Q_{TM}}(v), \quad (\text{K.7})$$

which, via the stabilizer constraint in Eq. (13.93), leads to

$$[\Phi_{r_{TM}}(v)^{-1} \phi \Phi_{r_{TM}}(\phi_{*,TM}^{-1} v)]_{*,\text{Hom}} \mathcal{Q} r_{TM} Q_{TM}(v) = \mathcal{Q} r_{TM} Q_{TM}(v). \quad (\text{K.8})$$

Putting these observations together proves the claim:

$$\begin{aligned} & \phi_{*,\text{Hom}} \Lambda(\mathcal{Q}) \phi_{*,TM}^{-1}(v) \\ &= \phi_{*,\text{Hom}} \Phi_{r_{TM}}(\phi_{*,TM}^{-1} v)_{*,\text{Hom}} \mathcal{Q} r_{TM} Q_{TM}(\phi_{*,TM}^{-1} v) \\ &= \phi_{*,\text{Hom}} \Phi_{r_{TM}}(\phi_{*,TM}^{-1} v)_{*,\text{Hom}} \mathcal{Q} r_{TM} Q_{TM}(v) \\ &= [\Phi_{r_{TM}}(v) \Phi_{r_{TM}}(v)^{-1}]_{*,\text{Hom}} \phi_{*,\text{Hom}} \Phi_{r_{TM}}(\phi_{*,TM}^{-1} v)_{*,\text{Hom}} \mathcal{Q} r_{TM} Q_{TM}(v) \\ &= \Phi_{r_{TM}}(v)_{*,\text{Hom}} [\Phi_{r_{TM}}(v)^{-1} \phi \Phi_{r_{TM}}(\phi_{*,TM}^{-1} v)]_{*,\text{Hom}} \mathcal{Q} r_{TM} Q_{TM}(v) \\ &= \Phi_{r_{TM}}(v)_{*,\text{Hom}} \mathcal{Q} r_{TM} Q_{TM}(v) \\ &= \Lambda(\mathcal{Q}) \end{aligned} \quad (\text{K.9})$$

2d) $\xi_{*,\text{Hom}} [\Lambda^{-1}(\mathcal{K})](w) = [\Lambda^{-1}(\mathcal{K})](w) \quad \forall w \in r_{TM}(\mathcal{I} \setminus TM)$, $\xi \in \text{Stab}_w$, that is, $\Lambda^{-1}(\mathcal{K})$ satisfies the stabilizer constraint :

This statement is easily proven since the invariance (equivariance) properties of \mathcal{K} carry over to its restriction $\Lambda^{-1}(\mathcal{K})$. We obtain for arbitrary $w \in r_{TM}(\mathcal{I} \setminus TM)$ and $\xi \in \text{Stab}_w$, that:

$$\begin{aligned} \xi_{*,\text{Hom}} [\Lambda^{-1}(\mathcal{K})](w) &= \xi_{*,\text{Hom}} \mathcal{K}|_{r_{TM}(\mathcal{I} \setminus TM)}(w) \\ &= \xi_{*,\text{Hom}} \mathcal{K}(w) \\ &= \mathcal{K}(\xi_{*,TM} w) \\ &= \mathcal{K}(w) \\ &= \mathcal{K}|_{r_{TM}(\mathcal{I} \setminus TM)}(w) \\ &= [\Lambda^{-1}(\mathcal{K})](w) \end{aligned} \quad (\text{K.10})$$

3) All constructions and proofs are independent from the particular choice of $\Phi_{r_{TM}}$:

The definition

$$\Phi_{r_{TM}} : TM \rightarrow \mathcal{I} \quad \text{such that} \quad \Phi_{r_{TM}}(v)_{*,TM} r_{TM} Q_{TM}(v) = v \quad (\text{K.11})$$

from Eq. (13.86) is unique up to right multiplication of $\Phi_{r_{TM}}$ with *any*

$$\xi_{r_{TM}} : TM \rightarrow \mathcal{I} \quad \text{such that} \quad \xi_{r_{TM}}(v) \in \text{Stab}_{r_{TM} Q_{TM}}(v) \quad (\text{K.12})$$

since, obviously, $\Phi_{r_{TM}}(v)_{*,TM} \xi_{r_{TM}}(v)_{*,TM} r_{TM} Q_{TM}(v) = \Phi_{r_{TM}}(v)_{*,TM} r_{TM} Q_{TM}(v) = v$ for any $v \in TM$. As argued in footnote 10, this covers all degrees of freedom in the definition of reconstruction isometries. From the stabilizer constraint in Eq. (13.93) it follows that $\xi_{r_{TM}}(v)_{*,\text{Hom}} \mathcal{Q} r_{TM} Q_{TM}(v) = \mathcal{Q} r_{TM} Q_{TM}(v)$ such that the lift Λ is seen to be invariant w.r.t. the ambiguity of $\Phi_{r_{TM}}$:

$$\begin{aligned} \Lambda(\mathcal{Q}) &= \Phi_{r_{TM}}(v)_{*,\text{Hom}} \mathcal{Q} r_{TM} Q_{TM}(v) \\ &= \Phi_{r_{TM}}(v)_{*,\text{Hom}} \xi_{r_{TM}}(v)_{*,\text{Hom}} \mathcal{Q} r_{TM} Q_{TM}(v) \end{aligned} \quad (\text{K.13})$$

Except from the definition of the lifting isomorphism, $\Phi_{r_{TM}}$ is only used (in a slightly different context) in step 2 c), where the ambiguity is seen to drop out by similar arguments.

Together, these steps prove that $\Lambda : \mathcal{K}_{\text{quot}}^{\mathcal{I}} \rightarrow \mathcal{K}_{\text{invvar}}^{\mathcal{I}}$ is an isomorphism. \square

K.2 Proof of Theorem 13.3.3 – Equivalence of equivariant kernel field transforms and convolutions on homogeneous spaces

To keep a better overview, we split the proof in two parts, proving the claims made in the first and second statement of Theorem 13.3.3, respectively.

Part 1) – Constructing H , HM and Isom_{HM} : Let $r \in M$ be any representative point and, without loss of generality, let $\psi_{GM,r}^{\tilde{A}}$ be any isometric gauge at r . We set

$$H := \psi_{GM,r}^{\tilde{A}} \text{Stab}_r (\psi_{GM,r}^{\tilde{A}})^{-1}, \quad (\text{K.14})$$

which is just a particular representation of Stab_r relative to the chosen coordinatization. Since the gauge maps are isomorphisms, we get an isomorphism between the two groups:

$$\alpha : \text{Stab}_r \rightarrow H, \quad \xi \rightarrow \psi_{GM,r}^{\tilde{A}} \xi_{*,GM} (\psi_{GM,r}^{\tilde{A}})^{-1} =: h_{\xi}^{\tilde{A}\tilde{A}}(r) \quad (\text{K.15})$$

Since $\text{Stab}_r \leq \mathcal{I} \leq \text{Isom}_{GM}$, Theorem 13.1.3 assures that $h_{\xi}^{\tilde{A}\tilde{A}}(r)$ is for any $\xi \in \text{Stab}_p$ an element of G and thus that $H \leq G$. We furthermore have that $H \leq O(d)$, which is seen by the following calculation, which holds for any $v, w \in \mathbb{R}^d$:

$$\begin{aligned} & \langle h_{\xi}^{\tilde{A}\tilde{A}}(r) \cdot v, h_{\xi}^{\tilde{A}\tilde{A}}(r) \cdot w \rangle \\ & \stackrel{(1)}{=} \left\langle \left(\psi_{GM,r}^{\tilde{A}} \xi_{*,GM} (\psi_{GM,r}^{\tilde{A}})^{-1} \right) \cdot v, \left(\psi_{GM,r}^{\tilde{A}} \xi_{*,GM} (\psi_{GM,r}^{\tilde{A}})^{-1} \right) \cdot w \right\rangle \\ & \stackrel{(2)}{=} \left\langle \psi_{TM,r}^{\tilde{A}} \xi_{*,TM} (\psi_{TM,r}^{\tilde{A}})^{-1} v, \psi_{TM,r}^{\tilde{A}} \xi_{*,TM} (\psi_{TM,r}^{\tilde{A}})^{-1} w \right\rangle \\ & \stackrel{(3)}{=} \eta_r \left(\xi_{*,TM} (\psi_{TM,r}^{\tilde{A}})^{-1} v, \xi_{*,TM} (\psi_{TM,r}^{\tilde{A}})^{-1} w \right) \\ & \stackrel{(4)}{=} \eta_r \left((\psi_{TM,r}^{\tilde{A}})^{-1} v, (\psi_{TM,r}^{\tilde{A}})^{-1} w \right) \\ & \stackrel{(5)}{=} \langle v, w \rangle \end{aligned} \quad (\text{K.16})$$

Step (1) made use of Eq. (K.15). In step (2) we identified the expression of $h_{\xi}^{\tilde{A}\tilde{A}}(r)$ via $\psi_{GM,r}^{\tilde{A}}$ with its expression via $\psi_{TM,r}^{\tilde{A}}$, which is justified by the commutativity of the diagrams in Eqs. (13.38) and (13.24). As we assumed $\psi_{TM,r}^{\tilde{A}}$ w.l.o.g. to be isometric, we can identify the inner product $\langle \cdot, \cdot \rangle$ on \mathbb{R}^d in step (3) with the Riemannian metric η_r . Step (4) uses that $\xi \in \text{Stab}_r \leq \mathcal{I}$ is an isometry, which preserves the metric by definition; see Eq. (13.1). Lastly, we pull the metric in step (5) via the isometric gauge back to the inner product on \mathbb{R}^d . The equality of the initial and final expression shows that $h_{\xi}^{\tilde{A}\tilde{A}}(r)$ preserves the inner product on \mathbb{R}^d – this is exactly the requirement that *defines* the orthogonal group. We therefore have that $H \leq O(d)$, and, together with $H \leq G$, that

$$H \leq G \cap O(d). \quad (\text{K.17})$$

This proves the first statement of part 1) of Theorem 13.3.3. We move on to the second statement of part 1), the construction of HM and Isom_{HM} .

Given that Stab_r is a subgroup of \mathcal{I} , we have the canonical quotient map

$$q : \mathcal{I} \rightarrow \mathcal{I} / \text{Stab}_r, \quad \phi \rightarrow \phi \cdot \text{Stab}_r \quad (\text{K.18})$$

which sends group elements $\phi \in \mathcal{I}$ to the left coset $\phi \cdot \text{Stab}_r := \{\phi \xi \mid \xi \in \text{Stab}_r\}$ of Stab_r . It is well known that this quotient map makes \mathcal{I} to a principal Stab_r -bundle over the base space $\mathcal{I}/\text{Stab}_r$, with the right action given by the right multiplication $\blacktriangleleft : \mathcal{I} \times \text{Stab}_r \rightarrow \mathcal{I}$, $(\phi, \xi) \mapsto \phi \xi$ with stabilizer elements [101, 222]. Furthermore, $\mathcal{I}/\text{Stab}_r$ is isomorphic to the homogeneous space M . The isomorphism is given by

$$\beta : \mathcal{I}/\text{Stab}_r \rightarrow M, \quad \phi \cdot \text{Stab}_r \mapsto \phi(r), \quad (\text{K.19})$$

which is obviously independent of the choice of coset representative since different representatives differ by group elements that stabilize r . Note that we could equally well view $q : \mathcal{I} \rightarrow \mathcal{I}/\text{Stab}_r$ as a principal H -bundle since the typical fiber is only defined up to isomorphism.

With these preparations we define the H -structure HM as an embedding of the principal H -bundle \mathcal{I} into GM (and therefore into FM). We define the embedding map as

$$\mathcal{E} : \mathcal{I} \rightarrow GM, \quad \phi \mapsto \phi_{*,GM} \sigma^{\tilde{A}}(r), \quad (\text{K.20})$$

which depends once again on our choice of gauge since $\sigma^{\tilde{A}}(r) = (\psi_{GM,r}^{\tilde{A}})^{-1}(e)$. It can be thought of as tracing out an embedded copy of \mathcal{I} in GM by pushing around the frame $\sigma^{\tilde{A}}(r) \in G_r M$. That this gives indeed a valid embedding is guaranteed since the action of \mathcal{I} on frames is fixed point free. The embedding \mathcal{E} is a bundle map over β , that is, $\beta \circ q = \pi_{GM} \circ \mathcal{E}$. To show this, it is sufficient to apply both sides on an arbitrary element $\phi \in \mathcal{I}$, which gives the same result: $\beta \circ q(\phi) = \beta(\phi \cdot \text{Stab}_r) = \phi(r)$ and $\pi_{GM} \circ \mathcal{E}(\phi) = \pi_{GM} \phi_{*,GM} \sigma^{\tilde{A}}(r) = \phi \pi_{GM} \sigma^{\tilde{A}}(r) = \phi(r)$. The embedding map is furthermore right equivariant: For any $\xi \in \text{Stab}_r$ and any $\phi \in \mathcal{I}$ one has

$$\begin{aligned} \mathcal{E}(\phi \xi) &= \phi_{*,GM} \xi_{*,GM} \sigma^{\tilde{A}}(r) \\ &= \phi_{*,GM} \xi_{*,GM} (\psi_{GM,r}^{\tilde{A}})^{-1}(e) \\ &= \phi_{*,GM} (\psi_{GM,r}^{\tilde{A}})^{-1} \psi_{GM,r}^{\tilde{A}} \xi_{*,GM} (\psi_{GM,r}^{\tilde{A}})^{-1}(e) \\ &= \phi_{*,GM} (\psi_{GM,r}^{\tilde{A}})^{-1} (h_{\xi}^{\tilde{A}\tilde{A}}(r)) \\ &= \phi_{*,GM} (\psi_{GM,r}^{\tilde{A}})^{-1}(e) \lhd h_{\xi}^{\tilde{A}\tilde{A}}(r) \\ &= \mathcal{E}(\phi) \lhd h_{\xi}^{\tilde{A}\tilde{A}}(r), \end{aligned} \quad (\text{K.21})$$

where we used the right G (and thus H) equivariance of $\psi_{GM,r}^{\tilde{A}}$ (and thus $(\psi_{GM,r}^{\tilde{A}})^{-1}$) in the penultimate step. Together, these properties show that \mathcal{E} is a principal bundle map that makes the following diagram commutative:

$$\begin{array}{ccc} \mathcal{I} \times \text{Stab}_r & \xrightarrow{\mathcal{E} \times \alpha} & GM \times H \\ \blacktriangleleft \downarrow & & \downarrow \lhd \\ \mathcal{I} & \xrightarrow{\mathcal{E}} & GM \\ q \downarrow & & \downarrow \pi_{GM} \\ \mathcal{I}/\text{Stab}_r & \xrightarrow{\beta} & M \end{array} \quad (\text{K.22})$$

The claimed H -structure is then defined as the image

$$HM := \mathcal{E}(\mathcal{I}) = \{\phi_{*,GM} \sigma^{\tilde{A}}(r) \mid \phi \in \mathcal{I}\} \quad (\text{K.23})$$

of \mathcal{E} together with the restricted right action and projection map of GM . Since embeddings are necessarily injective, we have in particular that \mathcal{I} and HM are isomorphic as principal bundles.

As a last point we argue that \mathcal{I} and $\text{Isom}_{HM} = \{\theta \in \text{Isom}(M) \mid \theta_{*,GM} HM = HM\}$ coincide. The equality $\theta_{*,GM} HM = HM$ holds for a given $\theta \in \text{Isom}(M)$ if $\theta_{*,GM} HM$ is at the same time a subset and a superset of HM . The first case, $\theta_{*,GM} HM \subseteq HM$, requires that for any element $\theta_{*,GM} \phi_{*,GM} \sigma^{\tilde{A}}(r) \in \theta_{*,GM} HM$, there exists some $\phi'_{*,GM} \sigma^{\tilde{A}}(r) \in HM$ such that $\theta_{*,GM} \phi_{*,GM} \sigma^{\tilde{A}}(r) = \phi'_{*,GM} \sigma^{\tilde{A}}(r)$. Since the action of isometries on the frame bundle is free, this requires $\theta_{*,GM} = \phi'_{*,GM} \phi_{*,GM}^{-1}$, which in turn implies $\theta = \phi' \phi^{-1}$. As one can easily check, the second case results in the same requirement. Both ϕ' and ϕ are elements of \mathcal{I} such that θ is required to be an element of \mathcal{I} . This proves the claim

$$\text{Isom}_{HM} = \mathcal{I}. \quad (\text{K.24})$$

Part 2) – Equivalence of \mathcal{I} -equivariant kernel field transforms and HM -convolutions: To prove the second statement of the theorem, we construct an \mathcal{I} -equivariant kernel field transform on M and show that it is equivalent to a HM -convolution. Theorem 13.2.4 proved that \mathcal{I} -equivariant kernel field transforms require \mathcal{I} -invariant kernel fields, which can, according to Theorem 13.3.2, be equivalently encoded in terms of a field of representative kernels $\widehat{\mathcal{Q}} : \pi_{TM}^{-1}(r_M(\mathcal{I} \setminus M)) \rightarrow \pi_{Hom}^{-1}(r_M(\mathcal{I} \setminus M))$. For the case of a homogeneous space M , the quotient space $\mathcal{I} \setminus M$ consists of a single element, which we represent by $r = r_M(\mathcal{I} \setminus M) \in M$. The full invariant kernel field is therefore described by a single kernel $\widehat{\mathcal{Q}}|_r = \widehat{\mathcal{Q}} : T_r M \rightarrow \text{Hom}(\mathcal{A}_{in,r}, \mathcal{A}_{out,r})$. This kernel is required to satisfy the stabilizer constraint $\xi_{*,Hom} \widehat{\mathcal{Q}} \xi_{*,TM}^{-1} = \widehat{\mathcal{Q}} \quad \forall \xi \in \text{Stab}_r$ and is shared over M via the lifting isomorphism $\widehat{\Lambda}(\widehat{\mathcal{Q}})(v) = \Phi_{r_{TM}}(v)_{*,Hom} \widehat{\mathcal{Q}} r_{TM} Q_{TM}(v) = \Phi_{r_{TM}}(v)_{*,Hom} \widehat{\mathcal{Q}} \Phi_{r_{TM}}(v)_{*,TM}^{-1}(v)$. As shown below, the single Stab_r -constrained representative kernel corresponds exactly to an H -steerable template kernel, while the weight sharing via the lifting isomorphism $\widehat{\Lambda}$ from Theorem 13.3.2 corresponds exactly to the convolutional weight sharing in Def. 12.2.3.

To make the equivalence of the kernel constraints explicit, we express the kernel $\widehat{\mathcal{Q}}$ via Eq. (12.31) relative to the same gauge \tilde{A} as considered before as $K := \psi_{Hom,r}^{\tilde{A}} \widehat{\mathcal{Q}} (\psi_{TM,r}^{\tilde{A}})^{-1}$.

The frame volume factor $\sqrt{|\eta_r^{\tilde{A}}|}$ drops hereby out since we assumed the gauge w.l.o.g. to be isometric. The stabilizer constraint relative to this gauge then leads to

$$\begin{aligned} K &= \psi_{Hom,r}^{\tilde{A}} \widehat{\mathcal{Q}} (\psi_{TM,r}^{\tilde{A}})^{-1} \\ &= \psi_{Hom,r}^{\tilde{A}} \xi_{*,Hom} \widehat{\mathcal{Q}} \xi_{*,TM}^{-1} (\psi_{TM,r}^{\tilde{A}})^{-1} \\ &= \psi_{Hom,r}^{\tilde{A}} \xi_{*,Hom} (\psi_{Hom,r}^{\tilde{A}})^{-1} K \psi_{TM,r}^{\tilde{A}} \xi_{*,TM}^{-1} (\psi_{TM,r}^{\tilde{A}})^{-1} \\ &= \rho_{Hom}(h_{\xi}^{\tilde{A}\tilde{A}}(r)) K (h_{\xi}^{\tilde{A}\tilde{A}}(r))^{-1} \\ &= \frac{1}{|\det h_{\xi}^{\tilde{A}\tilde{A}}(r)|} \rho_{Hom}(h_{\xi}^{\tilde{A}\tilde{A}}(r)) K (h_{\xi}^{\tilde{A}\tilde{A}}(r))^{-1} \end{aligned} \quad (\text{K.25})$$

for any ξ in Stab_r . Note that we can include the determinant factor in the last step since $h_\xi^{\tilde{A}\tilde{A}}(r) \in O(d)$, as shown above. The isomorphism between Stab_r and H in Eq. (K.15) thus allows us to rewrite the stabilizer constraint as the H -steerability constraint

$$K = \frac{1}{|\det h|} \rho_{\text{Hom}}(h) \circ K \circ h^{-1} \quad \forall h \in H. \quad (\text{K.26})$$

on template kernels K of a HM -convolution.¹

What remains to be shown is the equivalence of the two ways of sharing weights. The weight sharing via $\hat{\Lambda}$, expressed via gauge \tilde{A} in terms of K , reads

$$\begin{aligned} \hat{\Lambda}(\hat{\mathcal{Q}})(v) &= \Phi_{r_{TM}}(v)_{*,\text{Hom}} \hat{\mathcal{Q}} r_{TM} Q_{TM}(v) \\ &= \Phi_{r_{TM}}(v)_{*,\text{Hom}} \hat{\mathcal{Q}} \Phi_{r_{TM}}(v)_{*,TM}^{-1}(v) \\ &= \Phi_{r_{TM}}(v)_{*,\text{Hom}} (\psi_{\text{Hom},r}^{\tilde{A}})^{-1} \psi_{\text{Hom},r}^{\tilde{A}} \hat{\mathcal{Q}} (\psi_{TM,r}^{\tilde{A}})^{-1} \psi_{TM,r}^{\tilde{A}} \Phi_{r_{TM}}(v)_{*,TM}^{-1}(v) \\ &= \left(\psi_{\text{Hom},r}^{\tilde{A}} \Phi_{r_{TM}}(v)_{*,\text{Hom}}^{-1} \right)^{-1} K \left(\psi_{TM,r}^{\tilde{A}} \Phi_{r_{TM}}(v)_{*,TM}^{-1} \right)(v). \end{aligned} \quad (\text{K.27})$$

The last line already looks quite similar to the definition of HM -convolutional kernel fields in Def. 12.2.3. To prove their equivalence, we need to show 1) that the isometry induced gauges $\psi_{TM,r}^{\tilde{A}} \Phi_{r_{TM}}(v)_{*,TM}^{-1}$ and $\psi_{\text{Hom},r}^{\tilde{A}} \Phi_{r_{TM}}(v)_{*,\text{Hom}}^{-1}$ at $\pi_{TM}(v)$ are H -compatible with the original gauges $\psi_{TM,r}^{\tilde{A}}$ and $\psi_{\text{Hom},r}^{\tilde{A}}$ and 2) that the induced gauges correspond to reference frames of unit volume (to explain the missing frame volume factor in Eq. (K.27)). For the first point, note that the codomain of the reconstruction isometry $\Phi_{r_{TM}} : TM \rightarrow \mathcal{I}$ coincides by Eq. (K.24) with Isom_{HM} . Theorem 13.1.3 therefore asserts that these induced gauges are compatible with any H -atlas of HM . The second point follows immediately since $H \leq O(d)$ (or since $\Phi_{r_{TM}}(v)$ is an isometry and \tilde{A} is isometric). The weight sharing of $\hat{\mathcal{Q}}$ via the lifting isomorphism in Eq. (K.27) is therefore seen to coincide with the HM -convolutional weight sharing of the H -steerable kernel K in Def. 12.2.3. Together with the result that the stabilizer kernel constraint results in the H -steerability constraint, this implies that the lifted kernel field is equivalent to a HM -convolutional kernel field, which proves part 2) of the theorem.

A different choice of gauge \tilde{A} might for $G < O(d)$ result in a conjugate subgroup \bar{H} to H and an embedding \bar{HM} of \mathcal{I} that differs from HM . As one can easily check, the \bar{H} -steerability constraint allows to describe the same kernel relative to \bar{HM} like the H -steerability constraint in relation to HM , since the transformation falls out.

¹Since $h \in H \leq G \cap O(d)$, the determinant factor always drops out and could therefore be omitted.

Spherical convolutions as *GM*-convolutions – proofs

This appendix presents the proofs of Theorems 17.2.1 and 17.2.2 from Section 17.2. Together, these theorems assert that the Stab_n -steerable spherical convolution kernels by Cohen et al. [56] are equivalent to certain $\text{Stab}_n \cong G$ -steerable kernels, and that the \mathcal{I} -equivariant spherical convolutions with these kernels are equivalent to our corresponding *GM*-convolutions.

L.1 Proof of Theorem 17.2.1 – Kernel space isomorphism

Theorem 17.2.1 establishes an isomorphism

$$\Omega : \mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^{G, B_{\mathbb{R}^2}(0, \pi)} \xrightarrow{\sim} \mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^{\text{Stab}_n} \quad (\text{L.1})$$

between the space $\mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^{G, B_{\mathbb{R}^2}(0, \pi)}$ of G -steerable kernels on the open ball $B_{\mathbb{R}^2}(0, \pi) \subset \mathbb{R}^2$ and the space $\mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^{\text{Stab}_n}$ of $G \cong \text{Stab}_n$ -steerable kernels on $S^2 \setminus n$, which are defined in Eqs. (17.29) and (17.28). Given arbitrary gauges N at the north pole n , around which the kernel is centered, and gauges P at any other point p , this isomorphism is given by

$$\Omega(K) : S^2 \setminus n \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}, \quad (\text{L.2})$$

$$p \mapsto [\Omega(K)](p) := K(\psi_{TM,n}^N \log_n p) \rho_{\text{in}}(g_{n \leftarrow p}^{NP}) \sqrt{|\eta_p^{\partial/\partial v}|}^{-1}.$$

Abbreviating $p := \exp_n(\psi_{TM,n}^N)^{-1}v$, its inverse is given by

$$\Omega^{-1}(\kappa) : B_{\mathbb{R}^2}(0, \pi) \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}, \quad (\text{L.3})$$

$$v \mapsto [\Omega^{-1}(\kappa)](v) := \kappa(\exp_n(\psi_{TM,n}^N)^{-1}v) \rho_{\text{in}}(g_{n \leftarrow p}^{NP})^{-1} \sqrt{|\eta_p^{\partial/\partial v}|}.$$

Proof: That Ω^{-1} is a well defined inverse of Ω is easily shown by inserting their expressions and verifying that

$$\Omega \circ \Omega^{-1} = \text{id}_{\mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^{\text{Stab}_n}} \quad \text{and} \quad \Omega^{-1} \circ \Omega = \text{id}_{\mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^{G, B_{\mathbb{R}^2}(0, \pi)}} \quad (\text{L.4})$$

hold. To see this, note that gauges $\psi_{TM,n}^N$, the transporters $\rho_{in}(g_{n \leftarrow p}^{NP})$ and the (non-zero) volume factor $\sqrt{|\eta_p^{\partial/\partial v}|}$ are always invertible and the latter two commute since the volume scaling factor is a scalar. The exponential map $\exp_n : B_{\mathbb{R}^2}(0, \pi) \rightarrow S^2 \setminus -n$ on $B_{\mathbb{R}^2}(0, \pi)$ is inverted by $\log_n : S^2 \setminus -n \rightarrow B_{\mathbb{R}^2}(0, \pi)$.

The kernel constraints of the two kernel spaces furthermore imply each other. Given any G -steerable kernel $K \in \mathcal{K}_{\rho_{in}, \rho_{out}}^{G, B_{\mathbb{R}^2}(0, \pi)}$, the kernel $\Omega(K) \in \mathcal{K}_{\rho_{in}, \rho_{out}}^{\text{Stab}_n}$ satisfies the Stab_n -steerability constraint from Eq. (17.28). This is for any $p \in S^2 \setminus -n$, any $\xi \in \text{Stab}_n$ and any gauge X at $\xi(p)$ shown by:

$$\begin{aligned} & [\Omega(K)](\xi(p)) \\ & \stackrel{(1)}{=} K(\psi_{TM,n}^N \log_n \xi(p)) \cdot \rho_{in}(g_{n \leftarrow \xi(p)}^{NX}) \sqrt{|\eta_{\xi(p)}^{\partial/\partial v}|}^{-1} \\ & \stackrel{(2)}{=} K(\psi_{TM,n}^N \xi_{*,TM} \log_n p) \cdot \rho_{in}(g_{n \leftarrow \xi(p)}^{NX}) \sqrt{|\eta_{\xi(p)}^{\partial/\partial v}|}^{-1} \\ & \stackrel{(3)}{=} K(g_{\xi}^{NN}(n) \psi_{TM,n}^N \log_n p) \cdot \rho_{in}(g_{n \leftarrow \xi(p)}^{NX}) \sqrt{|\eta_{\xi(p)}^{\partial/\partial v}|}^{-1} \\ & \stackrel{(4)}{=} \rho_{out}(g_{\xi}^{NN}(n)) \cdot K(\psi_{TM,n}^N \log_n p) \cdot \rho_{in}(g_{\xi}^{NN}(n))^{-1} \rho_{in}(g_{n \leftarrow \xi(p)}^{NX}) \sqrt{|\eta_{\xi(p)}^{\partial/\partial v}|}^{-1} \\ & \stackrel{(5)}{=} \rho_{out}(g_{\xi}^{NN}(n)) \cdot K(\psi_{TM,n}^N \log_n p) \cdot \rho_{in}(g_{\xi}^{NN}(n))^{-1} \rho_{in}(g_{n \leftarrow \xi(p)}^{NX}) \sqrt{|\eta_p^{\partial/\partial v}|}^{-1} \\ & \stackrel{(6)}{=} \rho_{out}(g_{\xi}^{NN}(n)) \cdot K(\psi_{TM,n}^N \log_n p) \cdot \rho_{in}(g_{n \leftarrow p}^{NP}) \rho_{in}(g_{\xi}^{XP}(p))^{-1} \sqrt{|\eta_p^{\partial/\partial v}|}^{-1} \\ & \stackrel{(7)}{=} \rho_{out}(g_{\xi}^{NN}(n)) \cdot [\Omega(K)](p) \cdot \rho_{in}(g_{\xi}^{XP}(p))^{-1} \end{aligned} \tag{L.5}$$

The first step just expanded $\Omega(K)$, while the second step used $\log_n \xi(p) = \xi_{*,TM} \log_{\xi^{-1}(n)} p$, which follows from Eq. (13.42), together with $\xi^{-1}(n) = n$ since $\xi \in \text{Stab}_n$. In the third step, we used the definition of isometry induced gauge transformations in Eq. (13.23). Step four used the G -steerability constraint from Eq. (17.29). The fifth step replaced the volume element $\sqrt{|\eta_{\xi(p)}^{\partial/\partial v}|}$ with that at $\sqrt{|\eta_p^{\partial/\partial v}|}$, which is possible since the whole Riemannian geometry of the sphere, including the metric and exponential map and therefore the volume factors of the geodesic normal coordinates, is invariant under the action of Stab_n . Before identifying $\Omega(K)$ in the last step, step six used the identity

$$\begin{aligned} & \rho_{in}(g_{\xi}^{NN}(n))^{-1} \rho_{in}(g_{n \leftarrow \xi(p)}^{NX}) \\ &= [\psi_{A_{in}, n}^N \xi_{*, A_{in}}^{-1} (\psi_{A_{in}, n}^N)^{-1}] [\psi_{A_{in}, n}^N \mathcal{P}_{A_{in}, n \leftarrow \xi(p)} (\psi_{A_{in}, \xi(p)}^X)^{-1}] \\ &= \psi_{A_{in}, n}^N \xi_{*, A_{in}}^{-1} \mathcal{P}_{A_{in}, n \leftarrow \xi(p)} (\psi_{A_{in}, \xi(p)}^X)^{-1} \\ &= \psi_{A_{in}, n}^N \mathcal{P}_{A_{in}, n \leftarrow p} \xi_{*, A_{in}}^{-1} (\psi_{A_{in}, \xi(p)}^X)^{-1} \\ &= [\psi_{A_{in}, n}^N \mathcal{P}_{A_{in}, n \leftarrow p} (\psi_{A_{in}, p}^P)^{-1}] [\psi_{A_{in}, p}^P \xi_{*, A_{in}}^{-1} (\psi_{A_{in}, \xi(p)}^X)^{-1}] \\ &= \rho_{in}(g_{n \leftarrow p}^{NP}) \rho_{in}(g_{\xi}^{XP}(p))^{-1}, \end{aligned} \tag{L.6}$$

which relies crucially on the commutativity of transporters and isometry pushforwards from Eq. (13.44).

For the opposite direction, assume a Stab_n -steerable kernel $\kappa \in \mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^{\text{Stab}_n}$ to be given. The corresponding kernel $\Omega^{-1}(\kappa)$ satisfies then the G -steerability constraint from Eq. (17.29). To show this, let $v \in B_{\mathbb{R}^2}(0, \pi)$, let $g \in G$ and let $\xi \in \text{Stab}_n$ be the unique stabilizer element such that $g_\xi^{NN}(n) = \psi_{TM,n}^N \xi_{*,TM} (\psi_{TM,n}^N)^{-1} = g$. For brevity, we abbreviate $p := \exp_n (\psi_{TM,n}^N)^{-1} v$ and thus $\xi(p) = \exp_n (\psi_{TM,n}^N)^{-1} g v$, which is as justified by steps 1-3 below. We then find:

$$\begin{aligned}
& [\Omega^{-1}(\kappa)](gv) \\
& \stackrel{(1)}{=} \kappa(\exp_n (\psi_{TM,n}^N)^{-1}(gv)) \cdot \rho_{\text{in}}(g_{n \leftarrow \xi(p)}^{NX})^{-1} \sqrt{|\eta_{\xi(p)}^{\partial/\partial v}|} \\
& \stackrel{(2)}{=} \kappa(\exp_n \xi_{*,TM} (\psi_{TM,n}^N)^{-1} v) \cdot \rho_{\text{in}}(g_{n \leftarrow \xi(p)}^{NX})^{-1} \sqrt{|\eta_{\xi(p)}^{\partial/\partial v}|} \\
& \stackrel{(3)}{=} \kappa(\xi \exp_n (\psi_{TM,n}^N)^{-1} v) \cdot \rho_{\text{in}}(g_{n \leftarrow \xi(p)}^{NX})^{-1} \sqrt{|\eta_{\xi(p)}^{\partial/\partial v}|} \\
& \stackrel{(4)}{=} \rho_{\text{out}}(g_\xi^{NN}(n)) \cdot \kappa(\exp_n (\psi_{TM,n}^N)^{-1} v) \cdot \rho_{\text{in}}(g_\xi^{XP}(p))^{-1} \rho_{\text{in}}(g_{n \leftarrow \xi(p)}^{NX})^{-1} \sqrt{|\eta_{\xi(p)}^{\partial/\partial v}|} \\
& \stackrel{(5)}{=} \rho_{\text{out}}(g_\xi^{NN}(n)) \cdot \kappa(\exp_n (\psi_{TM,n}^N)^{-1} v) \cdot \rho_{\text{in}}(g_\xi^{XP}(p))^{-1} \rho_{\text{in}}(g_{n \leftarrow \xi(p)}^{NX})^{-1} \sqrt{|\eta_p^{\partial/\partial v}|} \\
& \stackrel{(6)}{=} \rho_{\text{out}}(g_\xi^{NN}(n)) \cdot \kappa(\exp_n (\psi_{TM,n}^N)^{-1} v) \cdot \rho_{\text{in}}(g_{n \leftarrow p}^{NP})^{-1} \rho_{\text{in}}(g_\xi^{NN}(n))^{-1} \sqrt{|\eta_p^{\partial/\partial v}|} \\
& \stackrel{(7)}{=} \rho_{\text{out}}(g_\xi^{NN}(n)) \cdot [\Omega^{-1}(\kappa)](v) \cdot \rho_{\text{in}}(g_\xi^{NN}(n))^{-1} \\
& \stackrel{(8)}{=} \rho_{\text{out}}(g) \cdot [\Omega^{-1}(\kappa)](v) \cdot \rho_{\text{in}}(g)^{-1}
\end{aligned} \tag{L.7}$$

The first three steps expanded $\Omega^{-1}(\kappa)$, used the definition of ξ in terms of g and the commutativity of exponential maps with isometry pushforwards, Eq. (13.42). In the fourth step, the Stab_n -steerability constraint of κ from Eq. (17.28) is used. Step five replaced again the Riemannian volume element at $\xi(p)$ with that at p since they are equal. The sixth step used the relation

$$\begin{aligned}
& \rho_{\text{in}}(g_\xi^{XP}(p))^{-1} \rho_{\text{in}}(g_{n \leftarrow \xi(p)}^{NX})^{-1} \\
& = \rho_{\text{in}}(g_{n \leftarrow \xi(p)}^{NX} g_\xi^{XP}(p))^{-1} \\
& = \left([\psi_{A_{\text{in}}, n}^N \mathcal{P}_{A_{\text{in}}, n \leftarrow \xi(p)} (\psi_{A_{\text{in}}, \xi(p)}^X)^{-1}] [\psi_{A_{\text{in}}, \xi(p)}^X \xi_{*, A_{\text{in}}} (\psi_{A_{\text{in}}, p}^P)^{-1}] \right)^{-1} \\
& = \left(\psi_{A_{\text{in}}, n}^N \mathcal{P}_{A_{\text{in}}, n \leftarrow \xi(p)} \xi_{*, A_{\text{in}}} (\psi_{A_{\text{in}}, p}^P)^{-1} \right)^{-1} \\
& = \left(\psi_{A_{\text{in}}, n}^N \xi_{*, A_{\text{in}}} \mathcal{P}_{A_{\text{in}}, n \leftarrow p} (\psi_{A_{\text{in}}, p}^P)^{-1} \right)^{-1} \\
& = \left(\psi_{A_{\text{in}}, n}^N \xi_{*, A_{\text{in}}} (\psi_{TM,n}^N)^{-1} \psi_{TM,n}^N \mathcal{P}_{A_{\text{in}}, n \leftarrow p} (\psi_{A_{\text{in}}, p}^P)^{-1} \right)^{-1} \\
& = \rho_{\text{in}}(g_{n \leftarrow p}^{NP})^{-1} \rho_{\text{in}}(g_\xi^{NN}(n))^{-1},
\end{aligned} \tag{L.8}$$

which relies again on the commutativity of transporters and isometry pushforwards from Eq. (13.44). The last two steps identify $\Omega^{-1}(\kappa)$ and, by definition of ξ , that $g_\xi^{NN}(n) = g$.

Together, these arguments that Ω is indeed an isomorphism between the kernel spaces. \square

L.2 Proof of Theorem 17.2.2 – Equivalence of steerable spherical and *GM*-convolutions

Theorem 17.2.2 claims that *GM*-convolutions with a G -steerable kernel $K \in \mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^{G, B_{\mathbb{R}^2}(0, \pi)}$ are equivalent to the spherical convolution with the Stab_n -steerable kernel $\Omega(K) \in \mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^{\text{Stab}_n}$. The spherical convolution with a Stab_n -steerable kernel $\kappa \in \mathcal{K}_{\rho_{\text{in}}, \rho_{\text{out}}}^{\text{Stab}_n}$ from Cohen et al. [55, 56] was hereby in Eq. (17.34) pointwise defined as

$$[\kappa \star_{S^2} f]^P(p) = \int_{S^2 \setminus -p} \kappa(\phi_p^{-1}q) \rho_{\text{in}}(g_{\phi_p^{-1}}^{XQ}(q)) f^Q(q) dq, \quad (\text{L.9})$$

where P, Q and X denote arbitrary gauges at p, q and $\phi_p^{-1}(q)$, respectively. The isometry $\phi_p \in \mathcal{I}$ is uniquely specified by demanding that $(\phi_p)_{*, GM} \sigma^N(n) = \sigma^P(p)$. Note that this implies in particular that

$$\phi_p(n) = p \quad (\text{L.10})$$

and, using the definition of sections of *GM* (frame fields) in terms of inverse gauges from Eq. (11.61), that

$$\psi_{TM, n}^N \circ (\phi_p)_{*, GM}^{-1} = \psi_{TM, p}^P, \quad (\text{L.11})$$

both of which we will use below. With these preparations, we turn to the proof of Theorem 17.2.2, i.e. the equivalence

$$\Omega(K) \star_{S^2} f = K \star_{GM} f \quad (\text{L.12})$$

of the convolutions.

Proof: Since $\Omega(\kappa)$ is defined on $S^2 \setminus -n$, the transformed kernel $\Omega(\kappa) \circ \phi_p^{-1}$ is defined on $S^2 \setminus -p$. Inserting $\Omega(\kappa)$ in the pointwise definition of the spherical convolution in Eq. (17.34) leads therefore to

$$\begin{aligned} & [\Omega(K) \star_{S^2} f]^P(p) \\ &= \int_{S^2 \setminus -p} [\Omega(\kappa)](\phi_p^{-1}q) \rho_{\text{in}}(g_{\phi_p^{-1}}^{XQ}(q)) f^Q(q) dq \\ &= \int_{S^2 \setminus -p} K(\psi_{TM, n}^N \log_n \phi_p^{-1}q) \rho_{\text{in}}(g_{n \leftarrow \phi_p^{-1}(q)}^{NX}) \rho_{\text{in}}(g_{\phi_p^{-1}}^{XQ}(q)) f^Q(q) \sqrt{|\eta_{\phi_p^{-1}(q)}^{\partial/\partial v}|}^{-1} dq, \end{aligned} \quad (\text{L.13})$$

where the second step follows by expanding $\Omega(K)$ as defined in Eq. (17.32). To simplify this expression, note that

$$\psi_{TM, n}^N \log_n \phi_p^{-1}(q) = \psi_{TM, n}^N (\phi_p)_{*, TM}^{-1} \log_{\phi_p(n)}(q) = \psi_{TM, p}^P \log_p(q), \quad (\text{L.14})$$

which follows from Eq. (13.42) in the first step and Eqs. (L.10) and (L.11) in the second step. Note furthermore, that

$$\begin{aligned}
& \rho(g_{n \leftarrow \phi_p^{-1}(q)}^{NX}) \rho(g_{\phi_p^{-1}}^{XQ}(q)) \\
&= [\psi_{\mathcal{A},n}^N \mathcal{P}_{\mathcal{A},n \leftarrow \phi_p^{-1}(q)} (\psi_{\mathcal{A},\phi_p^{-1}(q)}^X)^{-1}] [\psi_{\mathcal{A},\phi_p^{-1}(q)}^X (\phi_p)_{*,\mathcal{A}}^{-1} (\psi_{\mathcal{A},q}^Q)^{-1}] \quad (\text{Eqs. (11.85) and (13.39)}) \\
&= \psi_{\mathcal{A},n}^N \mathcal{P}_{\mathcal{A},n \leftarrow \phi_p^{-1}(q)} (\phi_p)_{*,\mathcal{A}}^{-1} (\psi_{\mathcal{A},q}^Q)^{-1} \quad (\text{canceled inverse gauges}) \\
&= \psi_{\mathcal{A},n}^N (\phi_p)_{*,\mathcal{A}}^{-1} \mathcal{P}_{\mathcal{A},\phi_p(n) \leftarrow q} (\psi_{\mathcal{A},q}^Q)^{-1} \quad (\text{Eq. (13.44)}) \\
&= \psi_{\mathcal{A},p}^P \mathcal{P}_{\mathcal{A},p \leftarrow q} (\psi_{\mathcal{A},q}^Q)^{-1} \quad (\text{Eq. (L.11)}) \\
&= \rho(g_{p \leftarrow q}^{PQ}). \quad (\text{Eq. (11.85)})
\end{aligned}$$

Inserting these two identities, we obtain

$$[\Omega(K) \star_{S^2} f]^P(p) = \int_{S^2 \setminus -p} K(\psi_{TM,p}^P \log_p q) \rho_{in}(g_{p \leftarrow q}^{PQ}) f^Q(q) \sqrt{|\eta_{\phi_p^{-1}(q)}^{\partial/\partial v}|}^{-1} dq, \quad (\text{L.15})$$

To proceed, we express the integral in geodesic normal coordinates $v : S^2 \setminus -p \rightarrow B_{\mathbb{R}^2}(0, \pi)$, $q \mapsto v(q) := \psi_{TM,p}^P \log_p q$ of $S^2 \setminus -p$, which are centered at point p . This cancels the Riemannian volume factor $\sqrt{|\eta_{\phi_p^{-1}(q)}^{\partial/\partial v}|}$ (and thus justifies its appearance in the definition of Ω), such that the spherical convolution becomes

$$\begin{aligned}
[\Omega(K) \star_{S^2} f]^P(p) &= \int_{B_{\mathbb{R}^2}(0, \pi)} K(v) \rho_{in}(g_{p \leftarrow \exp_p(\psi_{TM,p}^P)^{-1}v}^{PQ}) f^Q(\exp_p(\psi_{TM,p}^P)^{-1}v) dv, \\
&= \int_{B_{\mathbb{R}^2}(0, \pi)} K(v) [\text{Exp}_p^* f]^P(v) dv, \\
&= [K \star_{GM} f]^P(p).
\end{aligned} \quad (\text{L.16})$$

Since all arguments are independent form the chosen point p and the chosen gauges, this implies

$$\Omega(K) \star_{S^2} f = K \star_{GM} f, \quad (\text{L.17})$$

in a coordinate free setting, which proves the theorem. \square

Research questions & conclusions

This work is based on the main author’s doctoral dissertation at the University of Amsterdam (UvA). The Doctorate Regulations of UvA require a separate listing of the research questions addressed and answers found in the doctoral thesis, and an overview of the articles published during the doctoral studies. This appendix addresses these requirements, draws conclusions, and discusses limitations and directions of future research.

M.1 Research questions & contributions

The overarching goal of this work is to *develop a unified theory of equivariant convolutional neural networks*. Having access to such a theory is of great relevance to the research community since:

1. It brings order into the model zoo of equivariant CNNs, and clarifies how different approaches relate to each other.
2. It does not only map out the space of existing models, but also facilitates the design of novel CNN architectures.
3. It implies a unified implementation, which allows to construct arbitrary equivariant CNNs, and to benchmark these against each other.
4. It is a prime example of how prior knowledge and mathematical structure can be modeled directly into the networks’ architecture. This is not only relevant for equivariant CNNs, but also, for instance, for graph neural networks or networks for simulating PDEs or quantum systems.
5. It is necessarily more abstract and reduced to the essential mathematical structure, which helps to reveal connections to other sciences beyond deep learning.

The following research questions are targeted towards finding such a general theory of convolutional networks.

As a first step, we need to understand and formalize conventional CNNs and their equivariance properties.

Research question 1:

▷ Chapter 4

What is a convolutional network, how can it be formalized mathematically, and which role does equivariance play?

Convolutional networks are often described as variants of fully connected networks that operate on spatial signals and share neural weights between different spatial locations.¹ As a consequence, any shift of a layer's input feature map leads to a corresponding shift of its output feature map, that is, the layers are translation equivariant.

To formalize such conventional CNNs, we define their feature spaces as *regular translation group representations* (Def. 3.1.1), i.e. vector spaces of feature maps that are equipped with a translation group action which acts by moving signals spatially. Starting from generic fully connected network operations, we prove that spatial weight sharing is strictly necessary when demanding the layers' equivariance. Specifically for linear maps, Theorem 3.2.1 shows that equivariance implies *convolutions*. Theorems 3.2.2–3.2.8 prove similar results for bias summations, nonlinearities, and various pooling operations.

The crucial insight is that we do not only have the usual implication

$$\begin{array}{lll} \text{weight sharing} & \Rightarrow & \text{equivariance} & (\text{sufficiency}), \\ \text{but also} & & \text{weight sharing} & \Leftarrow \text{equivariance} & (\text{necessity}). \end{array}$$

As a consequence, CNNs are simply translation equivariant neural networks that operate on feature maps.

Viewing CNNs from this angle suggests that they can be generalized by

1. extending the symmetry groups w.r.t. which they are equivariant, and
2. defining feature maps on more general spaces, for instance, manifolds.

We tackle our overarching research goal by first developing a general theory of equivariant CNNs on Euclidean spaces; described in Part I and our publications [323, 322, 1, 40, 324, 137]. This formalism is subsequently generalized to a gauge theoretic description of CNNs on homogeneous spaces and Riemannian manifolds; found in Parts II and III and in [325, 57, 67, 55, 56, 46]. Finally, we investigate the generality of our theory in Part IV and [325]. The following research questions are structured accordingly.

M.1.1 Equivariant CNNs on Euclidean spaces

The context of our initial research question on equivariant Euclidean CNNs was set by several articles that have been published prior to the beginning of our studies. Most of these works were based on variants of *group convolutions* [223, 52, 324]. Cohen and Welling [53] proposed an alternative representation theoretic formulation of so-called (discrete) *steerable CNNs*, which allowed for more general group actions on the feature spaces. However, their formulation was limited to discrete pixel grids and finite symmetry groups. It was therefore unable to describe models like continuous group convolutions [324], vector field net-

¹In addition, neurons are often required to have a local receptive field, however, this is not strictly necessary for the network to be convolutional.

works [196], or harmonic networks [335]. Our mission was hence to formulate a general framework, which comprises all of the above mentioned models as special cases.

Research question 2:

▷ Chapter 4 and Weiler et al. [323]

How should equivariant CNNs on Euclidean spaces be defined? Is there a general formulation, which covers all of the abovementioned models?

A first sub-question to be addressed is *which symmetry groups* we should consider. The mutual implication of weight sharing and translation equivariance found above suggests that, given that we are interested in convolutional networks, translations should be contained as subgroup. We are furthermore interested in transformations like rotations, reflections, scaling, or shearing, which are modeled by matrix groups $G \leq \mathrm{GL}(d)$. To cover all of these settings, we consider a family of *affine groups* $\mathrm{Aff}(G) = (\mathbb{R}^d, +) \rtimes G$.

The next sub-question is *which feature spaces* and *which $\mathrm{Aff}(G)$ -actions* on them should be considered. In order for everything to reduce to conventional Euclidean CNNs when restricting to translations, it is necessary to define the feature spaces as *induced affine group representations* $\mathrm{Ind}_G^{\mathrm{Aff}(G)} \rho$; see Def. 4.2.1.² The feature fields are thereby characterized by their field type ρ , which is a G -representation. For any ρ of finite groups G , this reduces to the assumptions of Cohen and Welling [53], for $\mathrm{SO}(2)$ -irreps to the harmonics networks of Worrall et al. [335], for the defining representation of $\mathrm{SO}(2)$ to the vector fields of Marcos et al. [196], and for regular representations of $G \leq \mathrm{GL}(d)$ to group convolutions [223, 52, 324] (Theorem 4.5.1). Many other examples are found in Table 14.1 on page 273.

As done for conventional CNNs, we define generalized CNN layers as equivariant maps between the feature spaces. Similar to before, we find a mutual implication

$$\mathrm{Aff}(G)\text{-equivariant layer} \iff \mathrm{Aff}(G)\text{-invariant neural connectivity},$$

where the right-hand side is a generalized form of *weight sharing over affine transformations*. This generalized weight sharing is shown to split in 1) spatial weight sharing, and 2) novel G -steerability (equivariance) constraints on the shared neural connectivity; see Theorems 4.3.1-4.3.9. For instance, $\mathrm{Aff}(G)$ -equivariant *linear* maps are necessarily *convolutions with G -steerable kernels*.

The definition of equivariant CNNs in terms of induced representations was not novel to our work, but has previously been proposed by Cohen and Welling [53] – we denote our networks therefore, as in the original publication, as *steerable CNNs*. Our main contribution is rather that we extended the framework from finite groups and discrete pixel grids to non-finite groups and continuous space. This extension is non-trivial since it prevents the *numerical* solution of equivariance constraints proposed by Cohen and Welling [53]. Instead, we had to solve the constraints *analytically*, the first step of which leads to the statements of our Theorems 4.3.1-4.3.9, which require G -steerability constraints. The main difficulty in constructing steerable CNNs is to solve these G -steerability constraints, which is the goal of our next research question.

²The restriction $\mathrm{Res}_{(\mathbb{R}^d, +)}^{\mathrm{Aff}(G)} \mathrm{Ind}_G^{\mathrm{Aff}(G)} \rho$ of induced representations to translations results in the regular translation group representation that is underlying conventional Euclidean CNNs.

Research question 3:

▷ Chapter 5, Weiler et al. [323], Weiler and Cesa [322], Lang and Weiler [173], and Cesa et al. [40]

How can the G -steerability constraints on convolution kernels be solved, and how can steerable kernels be parameterized for learning?

To give some context, we note that we formalize convolution kernels as matrix-valued (non-linear) maps $K : \mathbb{R}^d \rightarrow \mathbb{R}^{C_{\text{out}} \times C_{\text{in}}} \cong \mathbb{R}^{C_{\text{out}}} \otimes (\mathbb{R}^{C_{\text{in}}})^*$. Given some $G \leq \text{GL}(d)$, we have the standard G -action on \mathbb{R}^d , and assume some G -representations on the feature vector spaces $\mathbb{R}^{C_{\text{in}}}$ and $\mathbb{R}^{C_{\text{out}}}$ to be given. The G -steerability constraint is then a linear symmetry constraint under these actions, which has to hold for any $g \in G$.

In general, we observed that the space of kernels is a vector space, and the constraint is linear, implying that we only need to solve for a *basis of steerable kernels*, in terms of which any steerable kernel can be expanded. The expansion coefficients constitute thereby the learnable parameters. Furthermore, the constraint decomposes into independent constraints on individual G -orbits. Beyond these statements, we have to consider specific families of groups and representations:

- If \mathbb{R}^d is sampled on a grid (e.g. \mathbb{Z}^d), and G is a *finite symmetry group* of this grid, this constraint can be solved numerically [53]. Otherwise it has to be solved analytically.
- Weiler et al. [323] considered $d = 3$ dimensions, rotations $G = \text{SO}(3)$, and $\text{SO}(3)$ -*irreps* acting on the feature vectors. We proved that the steerable basis is given by spherical harmonics, whose orders are determined by the Clebsch-Gordan decomposition of the irrep's tensor product.
- Weiler and Cesa [322] assumed $d = 2$, but arbitrary *finite G -representations* for any $G \leq \text{O}(2)$. The key observation here is that the constraint for general representations can be decomposed into irrep constraints. Furthermore, we proposed a Fourier expansion of the kernels.
- Lang and Weiler [173] proved a generalized Wigner-Eckart Theorem 5.3.1, which extends these results to arbitrary *compact groups* G (and hence any dimensionality d). The theorem describes how steerable kernel bases can be constructed from Clebsch-Gordan coefficients, harmonics on G -orbits, and irrep endomorphisms.
- Cesa et al. [40] reformulated this solution such that it becomes easier implementable and generalized it such that the smoothness of the kernels can be controlled.

Implementations of steerable kernel spaces for arbitrary field types, and many other equivariant network operations, are available in our `esccnn` library [38, 39].

A crucial difference between our approach and related work is that our theorems guarantee the *completeness* of the kernel spaces and hence equivariant maps. We found for instance that the original kernel spaces of harmonic networks [335] are incomplete, i.e. that there exist further steerable basis kernels beyond those found by the authors; see Appendix F.5 in [322]. Our experiments in Table 6.6 show that the complete kernel basis leads to improved results. Another example are the tensor field networks by Thomas et al. [301] (published simultaneously with our publication [323]), which proposed the same spherical harmonics basis. However, the authors only showed that convolutions with spherical harmonics imply equivariance, but not that they form a complete basis.

In our original derivation of steerable convolutions as equivariant linear maps, we assumed the *integral transform* ansatz in Eq. (3.8) for the linear maps [323]. This results in convolutions with spatially extended kernels, which is the standard practice in deep learning. However, such convolutions are not the only linear $\text{Aff}(G)$ -equivariant maps, since they miss, for instance, *partial differential operators* (PDOs). This shortcoming is addressed by the following question.

Research question 4:

▷ Jenner and Weiler [137]

How can partial differential operators be included into the framework of steerable CNNs? What are the most general linear equivariant maps between induced representations?

To address the first sub-question, we assumed the most general $c_{\text{out}} \times c_{\text{in}}$ -matrix of linear *partial differential operators*, mapping c_{in} -dimensional feature fields to c_{out} -dimensional feature fields. Demanding $\text{Aff}(G)$ -equivariance requires 1) that the coefficients of this PDO matrix are spatially constant (weight sharing), and 2) that it satisfies a G -steerability constraint. These steerability constraints for PDOs could, via the usual isomorphism between PDOs and polynomials and assuming compact G , be linked to our original steerability constraints for spatially extended kernels. This allowed us to express complete steerable PDO bases in terms of the steerable kernel bases found in the previous research question. Steerable PDOs are implemented as part of `escnn` [38, 39].

For the second sub-question, we assumed general *continuous linear functionals*. Once again, we find that $\text{Aff}(G)$ -equivariant maps correspond to steerable convolutions, however, now in the distributional sense, and requiring a G -steerability constraint on Schwartz distributions. This setting includes both PDOs and convolutions with classical kernels, and describes, in fact, the most general (continuous) linear equivariant maps between (Euclidean) induced representations.

The previous research questions resulted in a quite general formulation of equivariant Euclidean CNNs, complete solutions of their equivariant maps, and an implementation in form of the PyTorch extension `escnn` [39] (and its predecessor `e2cnn` [38]). Having access to this unified implementation enabled for the first time to conduct a large-scale comparative study of different equivariant models.

Research question 5:

▷ Chapter 6, Weiler and Cesa [322], and Cesa et al. [40]

How do different equivariant Euclidean CNNs compare relative to each other and to conventional CNNs in an empirical benchmark study?

The design space of equivariant Euclidean CNNs is determined by many novel hyperparameters. First and foremost, the models differ in their choices of symmetry groups G and the feature field types ρ , which determine the spaces of steerable kernels and biases. In addition, there is a multitude of equivariant nonlinearities, pooling operations, and, for classification tasks, final mappings to invariant predictions.

Table 6.6 shows the result of our benchmarking of 57 different equivariant CNNs in $d = 2$ spatial dimensions, where each model is evaluated on datasets with three different inherent symmetries [322]. We presented similar results in $d = 3$ dimensions in [40]. The main insight is that larger symmetry groups are preferable, and that regular and quotient representations usually perform best. For more details we refer to Section 6.5 and the original publications.

Furthermore, we examine in how far the models' equivariance properties hold in practice, and investigate their effect on the models' performance.

Research question 6: \triangleright Chapter 6, Weiler and Cesa [322], Weiler et al. [323], Weiler et al. [324], Jenner and Weiler [137] and Cesa et al. [40]

In how far do the theoretical properties of equivariant Euclidean CNNs hold in practice? What are their implications for the learning dynamics and model performance?

While our theory of Euclidean CNNs was developed in continuous space, implementations are usually discretized. If the symmetry is respected by the discretization, e.g. reflections or rotations by right angles on pixel grids \mathbb{Z}^d , equivariance is found to hold perfectly. Otherwise, e.g. for continuous rotations, there are discretization artifacts. Equivariance is then still found to hold quite well, and can furthermore be stabilized via data augmentation. With and without augmentation, such approximately equivariant implementations outperform non-equivariant CNNs significantly. They require furthermore less training data and converge faster. A more detailed overview and additional results are summarized at the beginning of Chapter 6.

Equivariant convolutions apply some G -steerable (i.e. G -equivariant) kernels at each point of space. It is hence intuitively obvious that steerable CNNs are not only equivariant under global $\text{Aff}(G)$ -actions, but also under independent local G -transformations of each kernel's field of view, as visualized in Fig. 4.7. This intuition brings us to our next bundle of research questions:

Research question 7: \triangleright Sections 4.4 and 6.3 and Weiler and Cesa [322]

How can the “local gauge equivariance” of steerable CNNs be formalized? How can local G -symmetries in signals be exploited when the learning task should not be G -equivariant on the global scale? Is this design useful in practice?

We start with the second question, assuming that local G -equivariance holds. A globally $\text{Aff}(H)$ -equivariant network for $H < G$ can then be constructed by using an $\text{Aff}(G)$ -equivariant network, whose features are at some layer restricted from $\text{Aff}(G)$ -representations to $\text{Aff}(H)$ -representations. This is mathematically described by the restriction functor Res_G^H on field types (Eq. (4.59)), which does in practice not change the actual data, but implies a decomposition of the feature fields in subrepresentations, which are by the next layer treated independently. Intuitively, the network is after restriction allowed to break $\text{Aff}(G)$ -equivariance, and maintain $\text{Aff}(H)$ -equivariance only.

Our experiments show that local rotation and reflection equivariance ($G = \text{O}(d)$) is usually always improving results by quite a margin. This holds specifically for natural image datasets (“upright” photos), which exhibit a preferred direction on their global scale ($H = \{e\}$ or reflections).

The local gauge equivariance of steerable CNNs can not be proven in the standard framework, however, it is shown to hold in their differential geometric generalization to Riemannian manifolds. Theorem 15.2.1 proves that this generalization includes Euclidean steerable CNNs as a special case, hence proving their conjectured gauge equivariance.

This concludes our main research questions and findings for *Euclidean* equivariant CNNs. In a parallel line of research, we investigated their generalization to non-Euclidean spaces.

M.1.2 Equivariant CNNs on homogeneous spaces and Riemannian manifolds

On Euclidean spaces, we defined CNNs as networks that are equivariant w.r.t. translations or more general affine group actions, from which the requirement for weight sharing followed. The reason for the neural connectivity to be shared across the whole space is that the considered group actions are *transitive*, that is, are able to move any point to any other location. From this viewpoint, it is obvious that our definition of CNNs as equivariant networks generalizes to any other *homogeneous space*, i.e. space that is equipped with some transitive action. For instance, spherical CNNs can be defined as $\text{SO}(3)$ or $\text{O}(3)$ -equivariant networks on the sphere S^2 [54, 83]. Kondor and Trivedi [162] used this insight to generalize *group convolution* based CNNs to arbitrary compact groups and homogeneous spaces thereof. Motivated by their work, and the insights from our initial paper on Euclidean steerable CNNs [323], we investigated how *steerable CNNs* can be generalized to homogeneous spaces.

Research question 8: ▷ Appendix F and Cohen et al. [56]

How can our Euclidean steerable CNNs from [323] be generalized to homogeneous spaces?

Euclidean steerable convolutions are intertwiners (equivariant linear maps) between induced $\text{Aff}(G)$ -representations $\text{Ind}_G^{\text{Aff}(G)} \rho$. The feature spaces contain feature fields of type ρ on the Euclidean homogeneous space $\text{Aff}(G)/G \cong \mathbb{R}^d$.

Any homogeneous space under some H -action arises as a quotient space H/G . Induced representations $\text{Ind}_G^H \rho$ can be defined for any choice of H and $G < H$ and describe feature fields of G -type ρ on H/G . We define steerable convolutions on homogeneous spaces – just as in the Euclidean setting – as intertwiners between induced representations. The space of intertwiners is shown to be given by convolution integrals with G -steerable kernels, which are now defined on H/G .

The group and homogeneous space convolutions of Kondor and Trivedi [162] are a special case of our steerable CNNs for field types ρ that are regular, or more general quotient representations, respectively. This claim is proven in Theorem 4.5.1 (there we are considering Euclidean spaces, however, the proof is purely algebraic, and applies to general homogeneous spaces as well). Our homogeneous steerable CNNs are more general since they allow to describe more general feature fields, for instance vector or other tensor fields.

Note that Cohen et al. [56] only derived the G -steerability constraint on kernels on homogeneous spaces, but did not provide a solution for their bases. This shortcoming is addressed in the following research question.

Research question 9: ▷ Lang and Weiler [173]

How can the steerability constraint for kernels on homogeneous spaces be solved?

The constraints on Euclidean steerable kernels decompose generally into independent constraints on G -orbits, which are by definition homogeneous spaces. Our Wigner-Eckart theorem does therefore actually describe steerable kernels on homogeneous spaces, from which the complete Euclidean kernels are assembled. However, the Wigner-Eckart theorem does not directly apply to Cohen et al.’s [56] definition of steerable kernels on homogeneous spaces, since their steerability constraint differs in general from that on Euclidean spaces. We could show that the solutions can nonetheless be mapped to each other, such that our Wigner-Eckart theorem describes complete bases for Cohen et al.’s [56] notion of steerable kernels as well.

Besides Euclidean and homogeneous spaces, there is a great interest in defining CNNs on Riemannian manifolds. Our previous approach – defining convolutions by demanding the networks’ equivariance under transitive global symmetries of space – does unfortunately not apply to Riemannian manifolds, since their isometry groups act in general non-transitively. We hence need to find another way to define convolutions on manifolds, which brings us to a gauge theoretic formulation of CNNs.

Research question 10: \triangleright Parts II and III, Weiler et al. [325] and Cohen et al. [57]

How can feature fields and convolutional networks be defined on Riemannian manifolds?

Since Riemannian manifolds are in general asymmetric (non-homogeneous), we revert to the alternative definition of CNNs as networks which share weights (kernels) between spatial locations, without deriving this property. The crucial difference to Euclidean spaces \mathbb{R}^d is that the *kernel alignment on manifolds is inherently ambiguous* – for instance, there may not be any preferred rotation or reflection of the kernel. From the viewpoint of steerable CNNs, this issue is naturally addressed by using steerable kernels, whose responses in different alignments (gauges) are due to their equivariance related by a predictable transformations, and encode therefore equivalent information.

How can this intuition be formalized? We identify kernel alignments with choices of reference frames (gauges), such that responses for different alignments can be viewed as merely being different coordinate representations of the same abstract feature vector. Mathematically, G -ambiguities in kernel alignments are identified with a G -structure, and feature fields are sections of G -associated vector bundles. The feature vectors, and any other G -associated quantities, are equipped with parallel transporters and, for G -structure compatible isometries, with pushforward actions. While our formulation in Cohen et al. [57] was still restricted to $G = \mathrm{O}(d)$ (or subgroups, if they are compatible with the Levi Civita connection’s holonomy), Weiler et al. [325] lifted this requirement and allowed for general $G \leq \mathrm{GL}(d)$.

Generic network layers are defined as any operations that map between associated bundle sections. Convolutional network layers are additionally required to share weights over the manifold. While generic operations are entirely unconstrained, we show that, in order for the weight sharing process to be coordinate independent, the shared connectivity is, once again, required to be G -steerable:

$$\left. \begin{array}{l} \text{coordinate independence} \\ \text{weight sharing} \end{array} \right\} \longrightarrow G\text{-steerability (gauge equivariance)}$$

This holds not only for convolution kernels, but for any other shared local operation, like bias summation, nonlinearities, 1×1 -convolutions (M -morphisms), and so on, as well.

Convolution kernels are defined on \mathbb{R}^d and shared over the tangent spaces $T_p M \cong \mathbb{R}^d$. They are via the exponential map matched with feature vectors, which are additionally parallel transported along the exponential maps’ geodesics. Theorem 12.2.6 proves the existence (well-definedness) of such defined convolutions provided that the kernels are compactly supported, and shows that smooth kernels map smooth input fields to smooth output fields. Besides providing local coordinate expressions of the network operations, we formulate them globally in a coordinate free language, which comes in handy when investigating the networks’ global isometry equivariance.

Coordinate independent CNNs emphasize local G -valued gauge transformations instead of global group actions. However, for manifolds with non-trivial isometries, we can ask for the models' isometry equivariance.

Research question 11: ▷ Chapter 13 and Weiler et al. [325]

Are coordinate independent CNNs isometry equivariant? Which requirement does isometry equivariance impose on the network structure?

As a preliminary point, we note that we need to consider subgroups Isom_{GM} of the manifold's full isometry group which respect its additional G -structure GM ; see Def. 13.1.1. This allows us to control the exact level of symmetries we are interested in, but is without loss of generality since we can always choose $G = O(d)$, for which we get general isometries.

We start with the second question about the constraints on the network connectivity that are imposed by isometry equivariance. To this end, we need to consider the unconstrained counterpart of convolutions, without the requirement for weight sharing. We call these operations (general) *kernel field transforms* (Def. 12.2.5), since they are parameterized by a field of (generally independent and non-steerable) kernels (Def. 12.2.1). Theorem 13.2.4 proves that a kernel field transform is exactly then isometry equivariant (Def. 13.2.1) when the kernel field is invariant under isometry actions (Def. 13.2.3):

$$\begin{array}{ccc} \text{isometry equivariant} & \iff & \text{isometry invariant} \\ \text{kernel field transform} & & \text{kernel field} \end{array}$$

This result enforces *weight sharing over isometry orbits* and requires, in addition, the shared kernels' *steerability w.r.t. their respective orbit's stabilizer subgroup*. Note the analogy to Euclidean and homogeneous spaces, where we got global weight sharing and a single steerability constraint.

Convolutions are specific instances of kernel field transforms with *convolutional kernel fields* (Def. 12.2.3), which are defined as sharing a single G -steerable kernel according to some G -structure GM over the whole manifold; see Def. 12.2.7. Theorem 13.2.5 proves that these kernel fields are by construction Isom_{GM} -invariant, which is visualized in Fig. 1.12. This implies, by the result above, the *Isom_{GM} -equivariance of the convolution operation*. Specifically for $G = O(d)$, the convolutions are guaranteed to be equivariant w.r.t. their manifold's full isometry group.

For Euclidean spaces, Theorem 15.2.2 proves the stronger result that coordinate independent CNNs are not only isometry equivariant, but equivariant under the action of more general affine groups $\text{Aff}(G)$; see research question 13 below.

Finally, Theorems 13.3.2 and 13.3.1 show that the space of isometry invariant kernel fields is isomorphic to reduced kernel fields on quotient spaces. Theorem 13.3.3 asserts that any equivariant kernel field transform on a homogeneous Riemannian manifold is necessarily a convolution.

We want to emphasize the result that both the convolutions' local gauge equivariance and global isometry equivariance properties are fully determined by the symmetries of the G -structure; more specifically, by their local G -symmetries, and global isometric symmetries Isom_{GM} , respectively. This is an entirely new way of thinking about equivariant CNNs, which greatly simplifies their analysis and design, as we will see in research questions 13-16 below.

With these results, we developed a fiber bundle formulation of CNNs on Riemannian manifolds with G -structure. The G -steerability constraints are already solved by our previous investigations in research question 3. We turn therefore to numerical implementations of coordinate independent CNNs.

Research question 12: \triangleright Chapters 10, 14, 15, 16, 17 and 18, Weiler et al. [325], de Haan et al. [67] and Cohen et al. [57]

How can coordinate independent CNNs on Riemannian manifolds be implemented? Do their theoretical properties hold in practice?

Chapter 14 gives a general overview of the design choices provided and implementation questions coming along with coordinate independent CNNs. The details of an implementation depend heavily on the manifold under consideration:

- For *Euclidean spaces*, the models reduce to our Euclidean steerable CNNs from above; see research question 13. Our library `escnn` implements them on pixel grids or sampled on a point cloud.
- If the manifold is *locally flat*, the convolution can be stitched together from local Euclidean convolutions on isometric chart codomains. The stitching of charts is formally described by transition maps, and is in practice implemented in form of parallel transporters along geodesics that transition between charts. We used this strategy to implement coordinate independent convolutions on the *icosahedron* in [57] and Section 17.4, and on the *Möbius strip* in [325] and Chapter 10.
- On curved manifolds, we need to implement custom convolution operations, whose details depend on the specific geometry. A practically relevant geometry is the 2-sphere S^2 . Chapter 17 derives analytical expressions for all necessary quantities, like exponential and logarithmic maps, transporters, or isometries. Kicanaoglu et al. [148] and implemented this model for regular feature fields. Theorem 17.2.2 suggests furthermore that spectral implementations of convolutions are suitable to realize our formulation of spherical CNNs.
- *General manifolds* are often modeled by meshes. In the computer graphics community, there is a specific interest in convolutions on *embedded surfaces*. Chapter 18 gives details on the classical differential geometry of embedded surfaces, on their discrete differential geometry counterparts, and discusses algorithms to compute geodesics and parallel transporters on meshes. We implemented gauge equivariant mesh CNNs on 2-dimensional meshes in [67]. Geodesics were computed using Danil Kirsanov's implementation [154] of the exact algorithm by Mitchell et al. [213]. We extended this code to support frame fields (gauges), and to compute (coordinate expressions of) parallel transporters.

All of our implementations – on Euclidean spaces, the icosahedron and Möbius strip, and on general meshes – are tested for their isometry equivariance. As already found in research question 6, equivariance holds perfectly whenever the discretization (e.g. pixel grid) is respected by the isometries. Otherwise, it holds approximately, and can be stabilized by using data augmentation. Our models outperform naive coordinate dependent (non-equivariant) baseline models on all benchmarks.

M.1.3 Generality of coordinate independent CNNs & literature review:

Since our overarching research goal was to develop a *unified* theory of convolutional networks, we need to assess in how far our coordinate independent CNNs are able to describe other models and theories that were proposed in the literature. Table 14.1 gives an overview of more than 100 models which we investigated and showed to be explained exactly as coordinate independent CNNs when assuming the correct manifolds, G -structures, connections, and field types.

The first class of models that should be covered by our framework are the steerable Euclidean CNNs devised in our first research questions.

Research question 13: ▷ Chapter 15 and Weiler et al. [325]

In how far do coordinate independent CNNs describe affine equivariant Euclidean CNNs?

In short, coordinate independent CNNs on Euclidean spaces recover our Euclidean steerable CNNs from above when considering suitable G -structures. Specifically, if we are interested in an $\text{Aff}(G)$ -equivariant steerable CNN, we need to consider an $\text{Aff}(G)$ -invariant G -structure, as shown in the left column of Fig. 1.6. We define these G -structures as being induced by $\text{Aff}(G)$ -atlases of global charts; see Defs. 15.1.1 and 15.1.2. Theorem 15.2.1 shows that the corresponding coordinate independent convolutions reduce to Euclidean steerable convolutions when being expressed relative to such charts. Their affine group equivariance is proven in Theorem 15.2.2. Conventional Euclidean CNNs are covered for the trivial structure group $G = \{e\}$.

Another interesting class of models are CNNs on Euclidean vector spaces \mathbb{R}^d which are rotation equivariant around the origin, but are *not* translation equivariant. They can be viewed as convolving on (hyper)spherical shells of different radii around the origin.

Research question 14: ▷ Chapter 16 and Weiler et al. [325]

In how far do coordinate independent CNNs describe such hyperspherical Euclidean CNNs?

The models' rotational equivariance corresponds in our framework to the rotationally invariant G -structures that are visualized in Chapter 16. Note that the canonical Levi-Civita connection of \mathbb{R}^d is not compatible with these G -structures, and the networks are indeed implicitly transporting features according to what we identify as alternative G -compatible connections.

The log-polar coordinate system induced G -structure in Fig. 16.3 is furthermore implying an alternative metric, whose distances in radial direction are shrunk logarithmically. Besides rotations, the isometries relative to this metric comprise “translations” along the radial direction. Relative to the canonical metric of \mathbb{R}^2 , these radial transformations correspond to a rescaling of the signal, such that the model is overall rotation and scale equivariant.

The model in Fig. 16.4 is identified as corresponding to an $\text{SO}(2)$ or $\text{O}(2)$ -structure, whose individual fibers are aligned in radial direction. The authors' assumptions correspond again to an alternative metric, whose radial dependence is this time unaltered, but whose distances in angular direction are radially independent.

The insight from studying these models is that our framework is quite flexible – we just need to assume the correct G -structures, metrics, connections, and field types to explain non-standard network architectures.

Another active area of research investigates spherical CNNs, which are used to process signals like omnidirectional images, weather data on the globe, or the cosmic microwave background.

Research question 15:

▷ Chapter 17 and Weiler et al. [325]

In how far do coordinate independent CNNs describe spherical CNNs?

The spherical CNNs in the literature can be categorized in fully $\text{SO}(3)$ or $\text{O}(3)$ -equivariant models, in $\text{SO}(2)$ or $\text{O}(2)$ -equivariant models around some distinguished rotation axis, and icosahedral approximations of spherical convolutions.

The first category corresponds to the $\text{SO}(3)$ or $\text{O}(3)$ -invariant $\text{SO}(2)$ or $\text{O}(2)$ -structures on S^2 shown in Fig. 17.2a. The models in the literature are in our framework interpreted as processing scalar fields, regular feature fields, or irrep fields, and as transporting features according to the sphere's Levi-Civita connection. A seeming difference is that most of these models define (steerable) kernels directly on the manifold S^2 instead of on its tangent spaces, as we do. However, our Theorems 17.2.1 and 17.2.2 prove an isomorphism between the kernel spaces and the equivalence of the resulting convolutions. These models are therefore exactly described by coordinate independent CNNs.

The second category of models corresponds to the $\text{SO}(2)$ -invariant $\{e\}$ -structure (frame field) in Fig. 17.2b or the $\text{O}(2)$ -invariant reflection group structure which results when adding reflected frames. These G -structures have singularities at the sphere's poles, and the convolutions are effectively operating on a cylindrical topology. Some of the models match the kernel on the tangent spaces via a gnomonic projection, instead of our projection via the exponential map, which again makes the models seemingly different. Theorem 17.3.1 proves that this is not the case since our projection can model theirs after a radial warp of the kernel (which does not interfere with isometric steerability constraints).

The locally flat icosahedral approximations are implemented in charts, as discussed in research question 12 above. Fig. 17.6 shows the G -structures that allow to explain the authors' models in our theory.

Next, we have convolutions on general surface meshes.

Research question 16:

▷ Chapter 18 and Weiler et al. [325]

In how far do coordinate independent CNNs describe CNNs on general surfaces?

CNNs on general surfaces, which are reviewed in Chapter 18, can be categorized in two broad classes.

The first group of models, described in Section 18.2, addresses the rotational ambiguity of kernel alignments on oriented surface meshes explicitly by applying some kind of $\text{SO}(2)$ -steerable kernels. Some of these models apply only rotation invariant (isotropic) kernels, which we interpret as steerable kernels that map between scalar fields. The other models apply non-trivially rotation steerable kernels, which map between non-trivial field types. We find that some of these models did not use the complete kernel basis for the field types considered. All of these models transport feature vectors according to the Levi-Civita connection.

The second group of models, explained in Section 18.3, uses non-steerable kernels, and assumes hence a trivial structure group $G = \{e\}$. Since no canonical kernel alignment is

given on general surfaces, the authors propose different heuristics to fix the alignments – in our framework, these are interpreted as heuristics to fix frame fields ($\{e\}$ -structures). Note that this leads on topologically non-trivial surfaces necessarily to singularities, and hence non-continuous convolution operations. The feature vectors are not transported according to the surfaces' Levi-Civita connections, but rather to the unique trivial connection corresponding to the heuristically fixed frame field.

At the end of Chapter 18's introduction, we furthermore describe some models that make assumptions which are incompatible with our framework of coordinate independent CNNs. Reasons for these incompatibilities are, for instance, that they use alternative projections of the kernels from the tangent spaces to the manifold, disregard the metric structure of the manifold, or are operating entirely differently, e.g. by applying Euclidean CNNs to renderings of the surface or by interpreting the mesh as a graph and applying a graph neural network. Coordinate independent CNNs are therefore not explaining all of the model applied to surfaces, but rather those which map between feature fields (associated bundle sections), and which project kernels via the exponential map.

Finally, we investigate how our coordinate independent CNNs on Riemannian manifolds from research questions 10-16 relate to our steerable CNNs on homogeneous spaces from questions 8 and 9.

Research question 17: \triangleright Chapter 13, Appendix F and Weiler et al. [325]

In how far do coordinate independent CNNs describe CNNs on homogeneous spaces?

We begin with the similarities of the theories. Both encode the models' symmetries via principal G -bundles, which are for coordinate independent CNNs given by G -structures $GM \xrightarrow{\pi} M \cong GM/G$, and for homogeneous steerable CNNs by the global symmetry groups H with projection maps $H \xrightarrow{\pi} H/G$. The feature spaces are in both cases sections of G -associated feature vector bundles. The main operations to map between these feature spaces are in both cases convolutions with G -steerable kernels.

On closer inspection, it becomes clear that neither of the theories could possibly be a strict generalization of the other: there exist homogeneous spaces which are not manifolds,³ and there exist manifolds which are not homogeneous spaces. Furthermore, homogeneous steerable CNNs define the kernels directly on the underlying space, while coordinate independent CNNs define them on the tangent spaces.

The practically most relevant spaces are Euclidean spaces and the 2-sphere, both of which are homogeneous spaces of their isometry groups. As already mentioned above, our Theorems 13.3.3, 17.2.1 and 17.2.2 assert that there exist G -structures on these spaces for which coordinate independent CNNs are equivalent to the homogeneous steerable CNNs. However, the G -structures of coordinate independent CNNs are less restrictive than the principal G -bundles assumed by steerable CNNs, allowing the former to describe more general Euclidean and spherical convolutions, not covered by the latter.

More details on the similarities of and differences between the two theories are discussed in Appendix F.

³A simple example are discrete pixel grids \mathbb{Z}^d , which are homogeneous spaces for e.g. discrete translations or more general grid-symmetries.

M.2 Conclusions & future research

This work proposed a unified theory of equivariant convolutional neural networks – first on Euclidean spaces, and then generalized to arbitrary homogeneous spaces and Riemannian manifolds. The fact that our theory is able to describe hundreds of seemingly independent models asserts that we indeed managed to identify and formalize the relevant geometric structure that is underlying convolutional neural networks. Interestingly, this geometric structure, given by a G -structure (principal G -bundle) and its associated bundles on the manifold, is exactly the same as that underlying fundamental theories in physics.

A central result of our theory is that convolutions on manifolds with a G -structure necessarily have to apply G -steerable (gauge equivariant) kernels. The majority of CNNs in the literature that are designed to be equivariant w.r.t. global symmetries of the underlying manifold are implicitly using some kind of steerable kernels. Our framework proves the previously unsolved conjecture that these models are, in fact, not only globally equivariant, but more generally equivariant under local gauge transformations. This insight explains the remarkable performance of such models on signals where G -symmetries are only present on a local, but not on a global scale.

Not only did we identify the requirement for steerable convolution kernels, but we also developed a full representation theoretic characterization of their (complete) solution spaces, without which the theory of equivariant CNNs would have been incomplete. As the implementation of general steerable kernels is quite technical, we provide a general and easy-to-use software library `escnn` [38, 39], which has been well adopted by the research community. It has in the meantime been used for a long list of projects, including applications of high societal impact, like medical imaging, or applications in the environmental, chemical, engineering, or material sciences. Plenty of other research projects relying on our implementation were listed at the end of page 12.

While we set out to develop a theory for *equivariant* CNNs, the formulation that we found is so general that it even encompasses the entirely *non-equivariant* CNNs in Section 18.3, which correspond to $\{e\}$ -structures and trivial connections. That this is possible is remarkable, since these models were designed without any equivariance in mind, or with the authors arguing that equivariance would not be desirable.

At the beginning of this appendix, we claimed five ways, in which a unified theory of equivariant CNNs would propel the field of deep learning. Having developed such a theory, we revisit these points, and see in how far they are addressed:

1. The research community proposed many different equivariant CNN models, each with their own very specific assumptions of spaces, symmetries, or group actions, and each using their own formulation and notation. Part IV showed that many of these models fit in our framework, and correspond merely to different choices of G -structures and field types. We hence developed a taxonomy of equivariant CNNs, which explains how different models relate to each other. In addition, we could show multiple times that the models do not make use of the complete kernel space corresponding to the implicitly assumed field types, or have a larger equivariance group than claimed by the authors.
2. Our findings allowed us furthermore to develop many novel equivariant network layers and architectures; including, for instance, all those models that are listed in rows of Table 6.6 without a citation. Some of the models found this way achieved new state-of-the-art results in various applications; see Chapter 6. Further examples from the community that build on our insights are conditional steerable neural processes [122],

- gauge equivariant spherical CNNs [148], or PDO-based steerable 3d CNNs [271], to name but a few.
3. Having access to our PyTorch extension `esCNN` [38, 39] enabled us – for the first time – to compare and classify different equivariant models on common ground; see our benchmark study in Section 6.5. This point is of particular importance to the community, since the results from different publications are usually not directly comparable.
 4. We claimed furthermore that insights from our work could be helpful for other focus areas in deep learning. One line of work that was directly motivated by our publications on “locally gauge equivariant” networks are graph neural networks which focus on local sub-graph automorphisms instead of global graph automorphisms [66, 298]. We believe that the insights and techniques of our work can furthermore be useful for *structure preserving deep learning* at large [36].
 5. Our abstract representation theoretic and differential geometric formulation did indeed highlight intriguing connections to other sciences beyond deep learning. Most notably, the formulation is in many ways similar to constructions found in physics, as discussed in more detail at the end of the previous Chapter 1. Another interesting connection exists to biological neural networks in neuroscience, since the cortical surface is itself a curved manifold. Specifically the visual cortex, which is the biological analog of CNNs, is found to be described by a fiber bundle with “contact structure” [121, 257].

Limitations & future research

Our theory is tailored to explain the current practices of convolutional networks, however there are several extensions that could be considered in future research.

One such extension would be to develop a partial differential operator (PDO) based counterpart of coordinate independent CNNs on manifolds, which are currently limited to spatially extended kernels. We already investigated such models on Euclidean spaces, finding a requirement for *steerable* PDOs. It is obvious that these networks generalize to Riemannian manifolds in the same way as Euclidean steerable CNNs with extended kernels did. We suppose that such models could be formulated by 1) prolonging the associated bundle sections (feature fields) to their jet bundles, 2) defining PDOs as linear maps from these jet prolongations back to an output field, and 3) deriving a steerability constraint from the requirement for coordinate independence. An advantage of such a formulation would be that these models could achieve full *diffeomorphism equivariance*, while spatially extended kernels allow in general only for isometry equivariance. Furthermore, this formulation would be even closer to the differential gauge theories in physics than our current non-differential version of it. Note that the models in the literature and in Part IV would *not* be explained by PDO based networks, since they are explicitly assuming spatially extended kernels.

A related extension would be to develop CNNs with continuous depth instead of discrete layers. Such models exist already for non-convolutional (fully connected) networks, where the neural activations’ dynamics in model depth is driven by a “neural ODE” [45]. The convolutional counterpart would therefore be (steerable) neural PDEs, describing the evolution of feature fields with continuous model depth.

Sticking with spatially extended kernels, one could consider alternative kernel projections from the tangent spaces to the manifold. Our projection via the exponential map and transporters along geodesics is a natural choice, since it corresponds on Euclidean spaces and the sphere to the usual parametrization of kernels. However, we found in [67] that the “gravitational lensing” effect when following geodesics leads to a deteriorated performance on

rugged manifolds. This issue is eased by reducing the kernel size (or using PDOs), however, there are alternative approaches, which rely on the insight that the networks do not necessarily require transporters along geodesics, but that any path (or isomorphism) will do. One could for instance consider a *path integral formulation*, which accumulates the transporters along all possible paths via a functional integral. Another related approach, proposed by Sommer and Bronstein [280], is to accumulate features via *diffusion* over their associated bundle.

There is furthermore a close relation between convolutions and spectral approaches, which is captured by the well-known convolution theorem. This relation motivated many generalized CNN models, and Xu et al. [337] have developed a spectral formulation of our steerable CNNs on homogeneous spaces [56]. Future research could investigate whether it is possible to reformulate coordinate independent CNNs on manifolds in the same manner, for instance by defining convolutions via point-wise multiplications of the spectrum of some generalized Laplacian which acts on bundle sections (e.g. the connection or the Bochner Laplacian).

A current limitation of our Wigner-Eckart theorem is that it applies only to compact groups, since it makes use of the Peter-Weyl theorem and the complete reducibility of the field types. Leveraging Pontryagin duality [243], it should be possible to prove a similar theorem for locally compact abelian groups. The obvious application would be to describe scale steerable kernels, however, the requirement for *abelian* structure groups is quite limiting. Sellaroli [265] present a Wigner-Eckart theorem for the Lorentz group, which shows that extensions beyond compact and abelian groups are in principle possible.

Further extensions of the theory could investigate structure groups that are not subgroups of $\mathrm{GL}(d)$. An interesting application from quantum field theory would be networks that rely on *spin structures* ($G = \mathrm{Spin}(d)$) and operate on spinor bundles. One could moreover try to *learn a G-structure* on a manifold instead of fixing it.

Besides equivariant *linear operations* (convolutions) and *bias summation*, which we characterized completely, there are many potentially interesting *nonlinear equivariant maps* between associated bundle sections. While we proposed, implemented and evaluated many novel nonlinear operations, it remains unclear how nonlinear layers could be investigated in a systematic fashion. A promising approach for future research in this direction is to develop equivariant counterparts of common non-equivariant network operations. For instance, the community already proposed several equivariant *attention mechanisms* [133, 247, 245, 246, 97, 98, 42].

Our equivariant and coordinate independent CNNs could furthermore be combined with orthogonal advances in equivariant deep learning, like equivariant *capsule networks* [179, 352, 308] or *probabilistic* equivariant models [16], including in particular equivariant *flows* [160, 244, 182] or equivariant *neural processes* [90, 122, 142]. Another topic of interest is the *universality* of equivariant networks [340, 201, 256, 143, 264, 240, 167, 75, 356, 355, 88, 194].

In the long term, it remains to be seen whether either our approach of *enforcing* equivariance and coordinate independence, or the antithetical philosophy of using unconstrained models and *learning* equivariance purely from data, will prevail. In the *infinite data limit*, learning would certainly yield optimal results, and might in addition yield more efficient data representations (specifically optimal symmetry groups and their actions). However, this is a somewhat trivial statement, as even a simple nearest neighbor classifier would perform optimally in this limit. Furthermore, we saw in Chapter 6 that, for *finite datasets*, equivariant models have an improved data efficiency, convergence rate, and final performance in comparison to non-equivariant models. The question is therefore rather how close we will in practice get to the infinite data limit and how costly training on such data will be. Models

like Vision Transformers [73] and MLP-Mixer [302] showed that, leveraging huge datasets like JFT-300M [292] with 300 million labeled images, non-equivariant models can achieve a comparable performance to CNNs. However, these models still exploit the images' spatial structure in one way or the other and contain operations that share weights, such that learning equivariance is encouraged. We therefore believe that hardcoding geometric cues will remain relevant, even though probably in a somewhat altered form. This holds specifically in areas with little data, whenever interpretability is required, or in scientific applications like computational chemistry, where exact equivariance is required.

M.3 List of publications

The Doctorate Regulations of the University of Amsterdam require a reference list of all articles contained in the thesis, stating the relative contributions of each (co-)author. The content of this thesis is not a concatenation of my original publications, but is entirely rewritten from scratch.⁴ However, as the insights from these publications did in one form or another flow into this work, they are nonetheless listed below. Authors that are not further mentioned in the listed contributions took an advisory role. A superscript * marks a shared first authorship (equal contributions).

Part I, *Equivariant Convolutional Networks on Euclidean Spaces*, contains and extends on insights from the following publications:

- 3D steerable CNNs [323]:

Maurice Weiler*, Mario Geiger*, Max Welling, Wouter Boomsma, and Taco Cohen
3D steerable CNNs: Learning rotationally equivariant features in volumetric data
 Conference on Neural Information Processing Systems (NeurIPS), 2018

Personal contributions: analytical solution of the kernel constraint in terms of spherical harmonics and Clebsch-Gordan coefficients, implementation (with Mario), experiments (with Mario and Wouter), and writing (with Mario, Wouter and Taco). In addition, Mario derived the kernel constraint and Wouter created the protein datasets.

- E(2)-steerable CNNs [322]:

Maurice Weiler* and Gabriele Cesa*
General E(2)-equivariant steerable CNNs
 Conference on Neural Information Processing Systems (NeurIPS), 2019

Personal contributions: idea, irrep decomposition of the kernel constraint, kernel Fourier expansion, e2cnn library design (with Gabriele), and writing (with Gabriele). In addition, Gabriele derived the irrep kernel space solutions, implemented and maintains the library and ran experiments.

- Steerable filter CNNs [321]:

Maurice Weiler, Fred Hamprecht, and Martin Storath
Learning steerable filters for rotation equivariant CNNs
 Conference on Computer Vision and Pattern Recognition (CVPR), 2018

Personal contributions: theory, implementation, experiments and writing.

⁴Parts II, III and IV were previously published as preprint on arXiv [325].

Section 5.3 on G -steerable kernel bases and the generalized Wigner-Eckart theorem contains insights from the previous three publications, and, in addition, from:

- Wigner-Eckart theorem [173]:

Leon Lang and Maurice Weiler
A Wigner-Eckart Theorem for Group Equivariant Convolution Kernels
 International Conference on Learning Representations (ICLR), 2020

Personal contributions: idea, writing (with Leon) and advisory role. Leon worked out all of the representation theory and proofs. This paper is based on Leon's Master's thesis.

- $E(N)$ -steerable CNNs [40]:

Gabriele Cesa, Leon Lang, and Maurice Weiler
A program to build $E(N)$ -equivariant steerable CNNs
 International Conference on Learning Representations (ICLR), 2022

Personal contributions: advisory role. Gabriele worked out the theory, implemented it, ran experiments and wrote the paper. Leon helped out with the theory and proofs.

Parts II, III and IV are based on the preprint *Coordinate Independent Convolutional Networks* [325]. Sections 17.4 and 18.2 describe (among other networks) models that were originally published in [57] and [67], respectively:

- Coordinate independent CNNs [325]:

Maurice Weiler, Patrick Forré, Erik Verlinde, and Max Welling
*Coordinate independent convolutional networks –
 isometry and gauge equivariant convolutions on Riemannian manifolds*
 arXiv preprint arXiv:2106.06020, 2021

Personal contributions: theory, implementation, experiments and writing.

- Gauge equivariant & icosahedral CNNs [57]:

Taco Cohen*, Maurice Weiler*, Berkay Kicanaoglu*, and Max Welling
Gauge equivariant convolutional networks and the Icosahedral CNN
 International Conference on Machine Learning (ICML), 2019

Personal contributions: original idea, formulation of the gauge equivariant convolution equation, derivation of the kernel constraint, implementation (with Berkay and Taco), and experiments (with Berkay). In addition, Taco formalized the bundle formulation and the convolution on the icosahedron in terms of an atlas of charts.

- Gauge equivariant mesh CNNs [67]:

Pim de Haan*, Maurice Weiler*, Taco Cohen, and Max Welling
Gauge equivariant mesh CNNs: Anisotropic convolutions on geometric graphs
 International Conference on Learning Representations (ICLR), 2021

Personal contributions: theory (with Pim), implementation of parallel transporters and logarithmic maps on meshes, experiments (with Pim), and writing (with Pim and Taco). In addition, Pim implemented the convolution operation given transporters and logarithmic maps.

Further publications, which are not part of this thesis are:

- Intertwiners between induced representations [55, 56]:

Taco Cohen, Mario Geiger, and Maurice Weiler

A general theory of equivariant CNNs on homogeneous spaces

Conference on Neural Information Processing Systems (NeurIPS), 2019

- Steerable PDOs [137]:

Erik Jenner and Maurice Weiler

Steerable Partial Differential Operators for Equivariant Neural Networks

International Conference on Learning Representations (ICLR), 2022

- Homeomorphic VAEs [87]:

Luca Falorsi, Pim de Haan, Tim R Davidson, Nicola De Cao,

Maurice Weiler, Patrick Forré, and Taco Cohen

Explorations in homeomorphic variational auto-encoding

arXiv preprint arXiv:1807.04689, 2018

- Covariance in convolutional networks [46]:

Miranda Cheng, Vassilis Anagiannis, Maurice Weiler,

Pim de Haan, Taco Cohen, and Max Welling

Covariance in physics and convolutional neural networks

ICML Workshop on Theoretical Physics for Deep Learning, 2019

Bibliography

- [1] Vishnu Agrawala. Wigner-Eckart theorem for an arbitrary group or Lie algebra. *Journal of Mathematical Physics*, 21, 1980. doi: 10.1063/1.524639.
- [2] Marjan Albooyeh, Daniele Bertolini, and Siamak Ravanbakhsh. Incidence networks for geometric deep learning. *International Conference on Machine Learning (ICML)*, 2020.
- [3] Brandon Anderson, Truong-Son Hy, and Risi Kondor. Cormorant: Covariant molecular neural networks. *arXiv preprint arXiv:1906.04015*, 2019.
- [4] Vincent Andereczyk, Julien Fageot, Valentin Oreiller, Xavier Montet, and Adrien Depeursinge. Exploring local rotation invariance in 3d cnns with steerable filters. *International Conference on Medical Imaging with Deep Learning*, pages 15–26, 2019.
- [5] Jimmy Aronsson. Homogeneous vector bundles and G -equivariant convolutional neural networks. *Sampling Theory, Signal Processing, and Data Analysis*, 20, 2022.
- [6] Aharon Azulay and Yair Weiss. Why do deep convolutional networks generalize so poorly to small image transformations? *arXiv preprint arXiv:1805.12177*, 2018.
- [7] Piyush Bagad, Floor Eijkelboom, Mark Fokkema, Danilo de Goede, Paul Hilders, and Miltiadis Kofinas. C-3po: Towards rotation equivariant feature detection and description. In *3rd Visual Inductive Priors for Data-Efficient Deep Learning Workshop*.
- [8] Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature communications*, 13(1):1–11, 2022.
- [9] Arash Behboodi, Gabriele Cesa, and Taco Cohen. A pac-bayesian generalization bound for equivariant networks. *arXiv preprint arXiv:2210.13150*, 2022.
- [10] Erik Bekkers. B-spline CNNs on Lie groups. *International Conference on Learning Representations (ICLR)*, 2020.
- [11] Erik Bekkers. Group equivariant deep learning lecture 1.7 - group convolutions are all you need!, 2022. URL https://uvadl2c.github.io/lectures/Geometric%20deep%20learning/Lecture_1_7_GConvsAreAllYouNeed.pdf.
- [12] Erik Bekkers, Maxime W Lafarge, Mitko Veta, Koen A.J. Eppenhof, Josien P.W. Pluim, and Remco Duits. Roto-translation covariant convolutional networks for medical image analysis. *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2018.

- [13] Heli Ben-Hamu, Haggai Maron, Itay Kezurer, Gal Avineri, and Yaron Lipman. Multi-chart generative surface modeling. 2018.
- [14] Siddharth Bhadra-Lobo, Georgy Derevyanko, and Guillaume Lamoureux. Dock2d: Synthetic data for the molecular recognition problem. *arXiv preprint arXiv:2212.03456*, 2022.
- [15] Naomi Blatt and Jacob Rubinstein. The canonical coordinates method for pattern recognition - II. isomorphisms with affine transformations. *Pattern recognition*, 27(1):99–107, 1994.
- [16] Benjamin Bloem-Reddy and Yee Whye Teh. Probabilistic symmetry and invariant neural networks. *arXiv preprint arXiv:1901.06082*, 2019.
- [17] Cristian Bodnar, Francesco Di Giovanni, Benjamin Paul Chamberlain, Pietro Liò, and Michael M Bronstein. Neural sheaf diffusion: A topological perspective on heterophily and oversmoothing in gnns. *arXiv preprint arXiv:2202.04579*, 2022.
- [18] Alexander Bogatskiy, Brandon Anderson, Jan Offermann, Marwah Roussi, David Miller, and Risi Kondor. Lorentz group equivariant neural network for particle physics. *arxiv:2006.04780*, 2020.
- [19] Georg Bökman and Fredrik Kahl. A case for using rotation invariant features in state of the art feature matchers. *CVPR 2022 Image Matching Workshop*, 2022.
- [20] Wouter Boomsma and Jes Frellsen. Spherical convolutions and their application in molecular modelling. *Conference on Neural Information Processing Systems (NeurIPS)*, 2017.
- [21] Davide Boscaini, Jonathan Masci, Simone Melzi, Michael M Bronstein, Umberto Castellani, and Pierre Vandergheynst. Learning class-specific descriptors for deformable shapes using localized spectral convolutional networks. *Computer Graphics Forum*, 34(5):13–23, 2015.
- [22] Davide Boscaini, Jonathan Masci, Emanuele Rodolà, and Michael Bronstein. Learning shape correspondence with anisotropic convolutional neural networks. *Conference on Neural Information Processing Systems (NeurIPS)*, pages 3189–3197, 2016.
- [23] Avishek Joey Bose and Ivan Kobyzev. Equivariant discrete normalizing flows. *arXiv preprint arXiv:2110.08649*, 2021.
- [24] Michael Boyle. How should spin-weighted spherical functions be defined? *Journal of Mathematical Physics*, 57(9):092504, 2016.
- [25] Theodor Bröcker and Tammo Tom Dieck. *Representations of compact Lie groups*, volume 98. Springer Science & Business Media, 2013.
- [26] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.
- [27] Joan Bruna and Stéphane Mallat. Invariant scattering convolution networks. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1872–1886, 2013.
- [28] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. *International Conference on Learning Representations (ICLR)*, 2013.
- [29] Max Budninskiy, Glorian Yin, Leman Feng, Yiying Tong, and Mathieu Desbrun. Parallel transport unfolding: A connection-based manifold learning approach. 2018.

- [30] James Burgess, Jeffrey J Nirschl, Maria Clara Zanellati, Sarah Cohen, and Serena Yeung. Learning orientation-invariant representations enables accurate and robust morphologic profiling of cells and organelles. *bioRxiv*, 2022.
- [31] Chen Cai and Yusu Wang. A note on over-smoothing for graph neural networks. *arXiv preprint arXiv:2006.13318*, 2020.
- [32] Sean M Carroll. *Spacetime and geometry. An introduction to general relativity*. Addison Wesley, 2004.
- [33] Nathan Carter. *Visual group theory*, volume 32. American Mathematical Soc., 2021.
- [34] David Casasent, Shao-Feng Xia, Andrew J Lee, and Jian-Zhong Song. Real-time deformation invariant optical pattern recognition using coordinate transformations. *Applied optics*, 26(5):938–942, 1987.
- [35] Tullio Ceccherini-Silberstein, A Machì, Fabio Scarabotti, and Filippo Tolli. Induced representations and mackey theory. *Journal of Mathematical Sciences*, 156(1):11–28, 2009.
- [36] Elena Celledoni, Matthias J. Ehrhardt, Christian Etmann, Robert I McLachlan, Brynjulf Owren, Carola-Bibiane Schönlieb, and Ferdia Sherry. Structure preserving deep learning. 2020.
- [37] Elena Celledoni, Matthias J Ehrhardt, Christian Etmann, Brynjulf Owren, Carola-Bibiane Schönlieb, and Ferdia Sherry. Equivariant neural networks for inverse problems. *Inverse problems*, 37(8):085006, 2021.
- [38] Gabriele Cesa, Erik Jenner, and Maurice Weiler. e2cnn PyTorch extension for E(2)-steerable CNNs, 2019. URL <https://github.com/QUVA-Lab/e2cnn>.
- [39] Gabriele Cesa, , Leon Lang, and Maurice Weiler. escnn PyTorch extension for E(d)-steerable CNNs, 2022. URL <https://github.com/QUVA-Lab/escnn>.
- [40] Gabriele Cesa, Leon Lang, and Maurice Weiler. A program to build E(N)-equivariant steerable CNNs. *International Conference on Learning Representations (ICLR)*, 2022. URL <https://openreview.net/pdf?id=WE4qe9xlnQw>.
- [41] Rudrasis Chakraborty, Monami Banerjee, and Baba Vemuri. H-CNNs: convolutional neural networks for Riemannian homogeneous spaces. *arXiv preprint arXiv:1805.05487*, 2018.
- [42] Evangelos Chatzipantazis, Stefanos Pertigkiozoglou, Edgar Dobriban, and Kostas Daniilidis. SE(3)-equivariant attention networks for shape reconstruction in function space. *arXiv preprint arXiv:2204.02394*, 2022.
- [43] Arnav Chavan, Udbhav Bamba, Rishabh Tiwari, and Deepak Gupta. Rescaling cnn through learnable repetition of network parameters. *arXiv preprint arXiv:2101.05650*, 2021.
- [44] Deli Chen, Yankai Lin, Wei Li, Peng Li, Jie Zhou, and Xu Sun. Measuring and relieving the over-smoothing problem for graph neural networks from the topological view. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3438–3445, 2020.
- [45] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Conference on Neural Information Processing Systems (NeurIPS)*, 2018.
- [46] Miranda Cheng, Vassilis Anagiannis, Maurice Weiler, Pim de Haan, Taco Cohen, and Max Welling. Covariance in physics and convolutional neural networks. *ICML 2019 Workshop on Theoretical Physics for Deep Learning*, 2019.

- [47] Benjamin Chidester, Tianming Zhou, Minh N Do, and Jian Ma. Rotation equivariant and invariant neural networks for microscopy image analysis. *Bioinformatics*, 35(14):i530–i537, 2019.
- [48] Gregory Chirikjian and Alexander Kyatkin. *Engineering applications of noncommutative harmonic analysis with emphasis on rotation and motion groups*. CRC press, 2001.
- [49] Gregory S Chirikjian and Imme Ebert-Uphoff. Numerical convolution on the euclidean group with applications to workspace generation. *IEEE Transactions on Robotics and Automation*, 14(1):123–136, 1998.
- [50] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*, 2015.
- [51] Adam Coates, Andrew Ng, and Honglak Lee. An analysis of single-layer networks in unsupervised feature learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 215–223. JMLR Workshop and Conference Proceedings, 2011.
- [52] Taco S. Cohen and Max Welling. Group equivariant convolutional networks. *International Conference on Machine Learning (ICML)*, 2016.
- [53] Taco S. Cohen and Max Welling. Steerable CNNs. *International Conference on Learning Representations (ICLR)*, 2017.
- [54] Taco S. Cohen, Mario Geiger, Jonas Köhler, and Max Welling. Spherical CNNs. *International Conference on Learning Representations (ICLR)*, 2018.
- [55] Taco S. Cohen, Mario Geiger, and Maurice Weiler. Intertwiners between induced representations (with applications to the theory of equivariant neural networks). *arXiv preprint arXiv:1803.10743*, 2018.
- [56] Taco S. Cohen, Mario Geiger, and Maurice Weiler. A general theory of equivariant CNNs on homogeneous spaces. *Conference on Neural Information Processing Systems (NeurIPS)*, 2019.
- [57] Taco S. Cohen, Maurice Weiler, Berkay Kicanaoglu, and Max Welling. Gauge equivariant convolutional networks and the Icosahedral CNN. *International Conference on Machine Learning (ICML)*, 2019.
- [58] Benjamin Coors, Alexandru Paul, and Andreas Geiger. Spherenet: Learning spherical representations for detection and classification in omnidirectional images. *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 518–533, 2018.
- [59] Marius Crainic and Federica Pasquotto. Linear G-structures by examples, 2013. URL <http://www.few.vu.nl/~pasquott/course16.pdf>.
- [60] Keenan Crane. Discrete Differential Geometry: An Applied Introduction. 30664:1–6.
- [61] Keenan Crane. Discrete Conformal Geometry. American Mathematical Society, 2020.
- [62] Keenan Crane, Mathieu Desbrun, and Peter Schröder. Trivial connections on discrete surfaces. 29(5):1525–1533. ISSN 1727-8384. doi: 10.1111/j.1467-8659.2010.01761.x. URL <http://dx.doi.org/10.1111/j.1467-8659.2010.01761.x>.
- [63] Keenan Crane, Clarisse Weischedel, and Max Wardetzky. The heat method for distance computation. *Commun. ACM*, 60(11), 2017.
- [64] Ekin D Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V Le. Autoaugment: Learning augmentation strategies from data. In *Proceedings of the*

- IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 113–123, 2019.
- [65] Fernando de Goes, Mathieu Desbrun, and Yiyi Tong. Vector field processing on triangle meshes. pages 1–49. 2016.
 - [66] Pim de Haan, Taco Cohen, and Max Welling. Natural graph networks. *Conference on Neural Information Processing Systems (NeurIPS)*, 2020.
 - [67] Pim de Haan, Maurice Weiler, Taco Cohen, and Max Welling. Gauge equivariant mesh CNNs: Anisotropic convolutions on geometric graphs. *International Conference on Learning Representations (ICLR)*, 2021.
 - [68] Mathieu Desbrun, Anil N Hirani, Melvin Leok, and Jerrold E Marsden. Discrete exterior calculus. *arXiv preprint math/0508341*, 2005.
 - [69] Terrance DeVries and Graham W Taylor. Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552*, 2017.
 - [70] Neel Dey, Antong Chen, and Soheil Ghafurian. Group equivariant generative adversarial networks. *arXiv preprint arXiv:2005.01683*, 2020.
 - [71] Sander Dieleman, Jeffrey De Fauw, and Koray Kavukcuoglu. Exploiting cyclic symmetry in convolutional neural networks. *International Conference on Machine Learning (ICML)*, 2016.
 - [72] Alexander Yom Din. Representation theory of compact groups - lecture notes. 2017. URL <http://www.math.caltech.edu/~2016-17/2term/ma145b/notes/RepCptGroups.pdf>.
 - [73] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations (ICLR)*, 2021.
 - [74] Bjørn Ian Dundas. *A Short Course in Differential Topology*. Cambridge University Press, 2018.
 - [75] Nadav Dym and Haggai Maron. On the universality of rotation equivariant point cloud networks. *arXiv preprint arXiv:2010.02449*, 2020.
 - [76] Alexander S. Ecker, Fabian H. Sinz, Emmanouil Froudarakis, Paul G. Fahey, Santiago A. Cadena, Edgar Y. Walker, Erick Cobos, Jacob Reimer, Andreas S. Tolias, and Matthias Bethge. A rotation-equivariant convolutional neural network model of primary visual cortex. *International Conference on Learning Representations (ICLR)*, 2019.
 - [77] Marc Eder and Jan-Michael Frahm. Convolutions on spherical images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019.
 - [78] Marc Eder, Mykhailo Shvets, John Lim, and Jan-Michael Frahm. Tangent images for mitigating spherical distortion. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
 - [79] Albert Einstein. The Foundation of the General Theory of Relativity (English translation). *Annalen der Physik*, 354(7):769–822, 1916.
 - [80] Albert Einstein. Die Grundlage der allgemeinen Relativitätstheorie. *Annalen der Physik*, 354(7):769–822, 1916. doi: <https://doi.org/10.1002/andp.19163540702>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19163540702>.
 - [81] Sharif Elcott and Peter Schröder. Building your own dec at home. pages 8–es. 2005.

- [82] Carlos Esteves. Theoretical aspects of group equivariant neural networks. *arXiv preprint arXiv:2004.05154*, 2020.
- [83] Carlos Esteves, Christine Allen-Blanchette, Ameesh Makadia, and Kostas Daniilidis. Learning SO(3) equivariant representations with spherical CNNs. *European Conference on Computer Vision (ECCV)*, 2018.
- [84] Carlos Esteves, Christine Allen-Blanchette, Xiaowei Zhou, and Kostas Daniilidis. Polar transformer networks. *International Conference on Learning Representations (ICLR)*, 2018.
- [85] Carlos Esteves, Yinshuang Xu, Christine Allen-Blanchette, and Kostas Daniilidis. Equivariant multi-view networks. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1568–1577, 2019.
- [86] Carlos Esteves, Ameesh Makadia, and Kostas Daniilidis. Spin-weighted spherical CNNs. 2020.
- [87] Luca Falorsi, Pim de Haan, Tim R Davidson, Nicola De Cao, Maurice Weiler, Patrick Forré, and Taco S Cohen. Explorations in homeomorphic variational auto-encoding. *arXiv preprint arXiv:1807.04689*, 2018.
- [88] Zhiying Fang, Han Feng, Shuo Huang, and Ding-Xuan Zhou. Theory of deep convolutional neural networks ii: Spherical analysis. *Neural Networks*, 131:154–162, 2020.
- [89] Damien Ferbach, Christos Tsirigotis, Gauthier Gidel, et al. A general framework for proving the equivariant strong lottery ticket hypothesis. *arXiv preprint arXiv:2206.04270*, 2022.
- [90] Marc Finzi, Roberto Bondesan, and Max Welling. Probabilistic numeric convolutional neural networks. *arXiv preprint arXiv:2010.10876*, 2020.
- [91] Marc Finzi, Samuel Stanton, Pavel Izmailov, and Andrew Gordon Wilson. Generalizing convolutional neural networks for equivariance to Lie groups on arbitrary continuous data. 2020.
- [92] Marc Finzi, Max Welling, and Andrew Gordon Wilson. A practical method for constructing equivariant multilayer perceptrons for arbitrary matrix groups. *Arxiv*, 2021.
- [93] Otto Forster. *Analysis 3 - Maß- und Integrationstheorie, Integralsätze im \mathbb{R}^n und Anwendungen*. Springer, 2012.
- [94] Daniel Franzen and Michael Wand. General nonlinearities in SO(2)-equivariant CNNs. *Conference on Neural Information Processing Systems (NeurIPS)*, 2021.
- [95] William Freeman and Edward Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):891–906, 1991.
- [96] Lan Fu, Hongkai Yu, Xiaoguang Li, Craig P Przybyla, and Song Wang. Deep learning for object detection in materials-science images: A tutorial. *IEEE Signal Processing Magazine*, 39(1):78–88, 2021.
- [97] Fabian B. Fuchs, Daniel E. Worrall, Volker Fischer, and Max Welling. SE(3)-transformers: 3d roto-translation equivariant attention networks. *Conference on Neural Information Processing Systems (NeurIPS)*, 2020.
- [98] Fabian B Fuchs, Edward Wagstaff, Justas Dauparas, and Ingmar Posner. Iterative SE(3)-transformers. *arXiv preprint arXiv:2102.13419*, 2021.
- [99] Jean Gallier. *Geometric Methods and Applications for Computer Science and Engineering*. Springer, 2011.

- [100] Jean Gallier and Jocelyn Quaintance. *Differential Geometry and Lie Groups - A Computational Perspective*. 2019.
- [101] Jean Gallier and Jocelyn Quaintance. *Differential Geometry and Lie Groups: A Second Course*. 2019.
- [102] Jean Gallier and Jocelyn Quaintance. Aspects of harmonic analysis and representation theory. 2019.
- [103] Kanchana Vaishnavi Gandikota, Jonas Geiping, et al. A simple strategy to provable invariance via orbit mapping. In *Proceedings of the Asian Conference on Computer Vision*, pages 3500–3518, 2022.
- [104] Jonas Geiping, Gowthami Somepalli, Ravid Shwartz-Ziv, Andrew Gordon Wilson, Tom Goldstein, and Micah Goldblum. How much data is augmentation worth? In *ICML 2022: Workshop on Spurious Correlations, Invariance and Stability*.
- [105] Jan E Gerken, Jimmy Aronsson, Oscar Carlsson, Hampus Linander, Fredrik Ohlsson, Christoffer Petersson, and Daniel Persson. Geometric deep learning and equivariant neural networks. *arXiv preprint arXiv:2105.13926*, 2021.
- [106] Jan E Gerken, Oscar Carlsson, Hampus Linander, Fredrik Ohlsson, Christoffer Petersson, and Daniel Persson. Equivariance versus augmentation for spherical images. *arXiv preprint arXiv:2202.03990*, 2022.
- [107] Rohan Ghosh and Anupam K Gupta. Scale steerable filters for locally scale-invariant convolutional neural networks. *arXiv preprint arXiv:1906.03861*, 2019.
- [108] Shunwang Gong, Lei Chen, Michael Bronstein, and Stefanos Zafeiriou. Spiralnet++: A fast and highly efficient mesh convolution operator. *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0, 2019.
- [109] Devon Graham, Junhao Wang, and Siamak Ravanbakhsh. Equivariant entity relationship networks. 2019.
- [110] Simon Graham, David Epstein, and Nasir Rajpoot. Dense steerable filter CNNs for exploiting rotational symmetry in histology images. 2020.
- [111] Nate Gruver, Marc Finzi, Micah Goldblum, and Andrew Gordon Wilson. The lie derivative for measuring learned equivariance. *arXiv preprint arXiv:2210.02984*, 2022.
- [112] Yulan Guo, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, and Mohammed Benamoun. Deep learning for 3d point clouds: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 2020.
- [113] Deepak K. Gupta, Devanshu Arya, and Efstratios Gavves. Rotation equivariant siamese networks for tracking. 2020.
- [114] David Gurarie. *Symmetries and Laplacians: introduction to harmonic analysis, group representations and applications*. Elsevier science publishers B.V., 1992.
- [115] Niv Haim, Nimrod Segol, Heli Ben-Hamu, Haggai Maron, and Yaron Lipman. Surface networks via general covers. 2018.
- [116] Jiaming Han, Jian Ding, Nan Xue, and Gui-Song Xia. ReDet: A Rotation-equivariant Detector for Aerial Object Detection. *arXiv preprint arXiv:2103.07733*, 2021.
- [117] Rana Hanocka, Amir Hertz, Noa Fish, Raja Giryes, Shachar Fleishman, and Daniel Cohen-Or. MeshCNN: a network with an edge. *ACM Transactions on Graphics (TOG)*, 38(4):1–12, 2019.

- [118] Yacov Hel-Or and Patrick C. Teo. Canonical decomposition of steerable functions. *Journal of Mathematical Imaging and Vision*, 9(1):83–95, 1998.
- [119] Amir Hertz, Rana Hanocka, Raja Giryes, and Daniel Cohen-Or. Deep geometric texture synthesis. *ACM Transactions on Graphics (TOG)*, 39(4):108–1, 2020.
- [120] Joel Hestness, Sharan Narang, Newsha Ardalani, Gregory Diamos, Heewoo Jun, Hassan Kianinejad, Md Patwary, Mostofa Ali, Yang Yang, and Yanqi Zhou. Deep learning scaling is predictable, empirically. *arXiv preprint arXiv:1712.00409*, 2017.
- [121] William C Hoffman. The visual cortex is a contact bundle. *Applied Mathematics and Computation*, 32(2-3):137–167, 1989.
- [122] Peter Holderrieth, Michael Hutchinson, and Yee Whye Teh. Equivariant Learning of Stochastic Fields: Gaussian Processes and Steerable Conditional Neural Processes. *arXiv preprint arXiv:2011.12916*, 2020.
- [123] Florian Hödlz, Daniel Rueckert, and Georgios Kaassis. Equivariant differentially private deep learning. *arXiv preprint arXiv:2301.13104*, 2023.
- [124] Florian A Hödlz, Daniel Rueckert, and Georgios Kaassis. Bridging the gap: Differentially private equivariant deep learning for medical image analysis. *arXiv preprint arXiv:2209.04338*, 2022.
- [125] Emiel Hoogeboom, Jorn W. T. Peters, Taco S. Cohen, and Max Welling. HexaConv. *International Conference on Learning Representations (ICLR)*, 2018.
- [126] Masanobu Horie, Naoki Morita, Yu Ihara, and Naoto Mitsume. Isometric transformation invariant and equivariant graph convolutional networks. *arXiv preprint arXiv:2005.06316*, 2020.
- [127] Yuan-Neng Hsu and H. Arsenault. Optical pattern recognition using circular harmonic expansion. *Applied Optics*, 21(22):4016–4019, 1982.
- [128] Hou-Ning Hu, Yen-Chen Lin, Ming-Yu Liu, Hsien-Tzu Cheng, Yung-Ju Chang, and Min Sun. Deep 360 pilot: Learning a deep agent for piloting through 360° sports videos. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [129] Haojie Huang, Dian Wang, Robin Walter, and Robert Platt. Equivariant transporter network. *arXiv preprint arXiv:2202.09400*, 2022.
- [130] Jingwei Huang, Yichao Zhou, Matthias Niessner, Jonathan Richard Shewchuk, and Leonidas J. Guibas. QuadriFlow: A Scalable and Robust Method for Quadrangulation. *Computer Graphics Forum*, 2018. ISSN 1467-8659. doi: 10.1111/cgf.13498.
- [131] Jingwei Huang, Haotian Zhang, Li Yi, Thomas Funkhouser, Matthias Nießner, and Leonidas J Guibas. Texturenet: Consistent local parametrizations for learning from high-resolution signals on meshes. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [132] Dale Husemöller. *Fibre Bundles*. Number 20 in Graduate Texts in Mathematics. Springer-Verlag, New York, 3rd ed edition, 1994. ISBN 978-0-387-94087-8.
- [133] Michael Hutchinson, Charline Le Lan, Sheheryar Zaidi, Emilien Dupont, Yee Whye Teh, and Hyunjik Kim. Lietransformer: Equivariant self-attention for lie groups. *arXiv preprint arXiv:2012.10885*, 2020.
- [134] Truong Son Hy, Shubhendu Trivedi, Horace Pan, Brandon M. Anderson, and Risi Kondor. Predicting molecular properties with covariant compositional networks. *The Journal of Chemical Physics*, 2018.

- [135] Priyank Jaini, Lars Holdijk, and Max Welling. Learning equivariant energy based models with equivariant stein variational gradient descent. *Advances in Neural Information Processing Systems*, 34, 2021.
- [136] Michiel Janssen, Augustus Janssen, Erik Bekkers, Javier Oliván Bescós, and Remco Duits. Design and processing of invertible orientation scores of 3D images. *Journal of Mathematical Imaging and Vision*, 2018.
- [137] Erik Jenner and Maurice Weiler. Steerable Partial Differential Operators for Equivariant Neural Networks. *International Conference on Learning Representations (ICLR)*, 2022. URL <https://arxiv.org/abs/2106.10163>.
- [138] Mingxi Jia, Dian Wang, Guanang Su, David Klee, Xupeng Zhu, Robin Walters, and Robert Platt. Seil: Simulation-augmented equivariant imitation learning. *arXiv preprint arXiv:2211.00194*, 2022.
- [139] Chiyu Jiang, Jingwei Huang, Karthik Kashinath, Prabhat, Philip Marcus, and Matthias Niessner. Spherical CNNs on unstructured grids. *International Conference on Learning Representations (ICLR)*, 2019.
- [140] Hai Jin, Yuanfeng Lian, and Jing Hua. Learning facial expressions with 3d mesh convolutional neural network. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(1):1–22, 2018.
- [141] Pengfei Jin, Tianhao Lai, Rongjie Lai, and Bin Dong. NPTC-net: Narrow-Band Parallel Transport Convolutional Neural Network on Point Clouds. *arXiv preprint arXiv:1905.12218*, 2019.
- [142] Makoto Kawano, Wataru Kumagai, Akiyoshi Sannai, Yusuke Iwasawa, and Yutaka Matsuo. Group Equivariant Conditional Neural Processes. *International Conference on Learning Representations (ICLR)*, 2021.
- [143] Nicolas Keriven and Gabriel Peyré. Universal invariant and equivariant graph neural networks. *Conference on Neural Information Processing Systems (NeurIPS)*, 2019.
- [144] Renata Khasanova and Pascal Frossard. Isometric transformation invariant graph-based deep neural network. *arXiv preprint arXiv:1808.07366*, 2018.
- [145] Naman Khetan, Tushar Arora, Samee Ur Rehman, and Deepak K Gupta. Implicit equivariance in convolutional networks. *arXiv preprint arXiv:2111.14157*, 2021.
- [146] Meenakshi Khosla and Leila Wehbe. High-level visual areas act like domain-general filters with strong selectivity and functional specialization. *bioRxiv*, 2022.
- [147] Meenakshi Khosla, Keith Jamison, Amy Kuceyeski, and Mert R Sabuncu. Characterizing the ventral visual stream with response-optimized neural encoding models. In *Advances in Neural Information Processing Systems*.
- [148] Berkay Kicanaoglu, Pim de Haan, and Taco Cohen. Gauge equivariant spherical CNNs. 2019.
- [149] Seung Bin Kim, Sung Goo Yoon, Jonghyun Tae, Jae Yoon Kim, Ji Sung Shim, Sung Gu Kang, Jun Cheon, Jeong Gu Lee, Je Jong Kim, and Seok Ho Kang. Detection and recurrence rate of transurethral resection of bladder tumors by narrow-band imaging: prospective, randomized comparison with white light cystoscopy. *Investigative and clinical urology*, 59(2):98–105, 2018.
- [150] Ron Kimmel and James A Sethian. Computing geodesic paths on manifolds. *Proceedings of the national academy of Sciences*, 95(15):8431–8435, 1998.
- [151] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations (ICLR)*, 2015.

- [152] Thomas Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *International Conference on Learning Representations (ICLR)*, 2017.
- [153] Thomas Kipf, Ethan Fetaya, Kuan-Chieh Wang, Max Welling, and Richard Zemel. Neural relational inference for interacting systems. *International Conference on Machine Learning (ICML)*, 2018.
- [154] Danil Kirisanov. Exact geodesic for triangular meshes – C++ code. URL <https://code.google.com/archive/p/geodesic/>.
- [155] David Klee, Ondrej Biza, Robert Platt, and Robin Walters. Image to icosahedral projection for so (3) object reasoning from single-view images. In *NeurIPS Workshop on Symmetry and Geometry in Neural Representations*, pages 64–80. PMLR, 2023.
- [156] Achim Klenke. *Probability theory: a comprehensive course*. Springer Science & Business Media, 2006.
- [157] Anatolii Ulianovich Klimyk and Ja Vilenkin. *Representation of Lie groups and special functions: Volume 3: Classical and quantum groups and special functions*, volume 75. Springer Science & Business Media, 2013.
- [158] Felix Knöppel, Keenan Crane, Ulrich Pinkall, and Peter Schröder. Globally optimal direction fields. *ACM Trans. Graph.*, 32(4), 2013.
- [159] S. Kobayashi. *Transformation Groups in Differential Geometry*. Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer-Verlag, 1972. ISBN 9783540058489.
- [160] Jonas Köhler, Leon Klein, and Frank Noé. Equivariant flows: exact likelihood generative learning for symmetric densities. *International Conference on Machine Learning (ICML)*, 2020.
- [161] Risi Kondor. N-body networks: a covariant hierarchical neural network architecture for learning atomic potentials. *arXiv preprint arXiv:1803.01588*, 2018.
- [162] Risi Kondor and Shubhendu Trivedi. On the generalization of equivariance and convolution in neural networks to the action of compact groups. *International Conference on Machine Learning (ICML)*, 2018.
- [163] Risi Kondor, Zhen Lin, and Shubhendu Trivedi. Clebsch–Gordan Nets: A Fully Fourier Space Spherical Convolutional Neural Network. *Conference on Neural Information Processing Systems (NeurIPS)*, 2018.
- [164] Risi Kondor, Hy Truong Son, Horace Pan, Brandon Anderson, and Shubhendu Trivedi. Covariant compositional networks for learning graphs. *International Conference on Learning Representations (ICLR)*, 2018.
- [165] Emmanuel Kowalski. *An introduction to the representation theory of groups*, volume 155. American Mathematical Society, 2010.
- [166] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Conference on Neural Information Processing Systems (NeurIPS)*, 2012.
- [167] Wataru Kumagai and Akiyoshi Sannai. Universal approximation theorem for equivariant maps by group cnns. 2020.
- [168] Anna Kuzina, Kumar Pratik, Fabio Valerio Massoli, and Arash Behboodi. Equivariant priors for compressed sensing with unknown orientation. *International Conference on Machine Learning (ICML)*, 2022.

- [169] Henry Kvinge, Tegan H Emerson, Grayson Jorgenson, Scott Vasquez, Timothy Doster, and Jesse D Lew. In what ways are deep neural networks invariant and how should we measure this? *arXiv preprint arXiv:2210.03773*, 2022.
- [170] Maxime Lafarge, Erik Bekkers, Josien Pluim, Remco Duits, and Mitko Veta. Roto-translation equivariant convolutional networks: Application to histopathology image analysis. 2020.
- [171] Wei-Sheng Lai, Yujia Huang, Neel Joshi, Christopher Buehler, Ming-Hsuan Yang, and Sing Bing Kang. Semantic-driven generation of hyperlapse from 360 degree video. *IEEE transactions on visualization and computer graphics*, 24(9):2610–2621, 2017.
- [172] Yu-Kun Lai, Miao Jin, Xuexiang Xie, Ying He, Jonathan Palacios, Eugene Zhang, Shi-Min Hu, and Xianfeng Gu. Metric-driven RoSy fields design and remeshing. *IEEE Transactions on Visualization and Computer Graphics*, 16(1):95–108, 2010.
- [173] Leon Lang and Maurice Weiler. A Wigner-Eckart Theorem for Group Equivariant Convolution Kernels. *International Conference on Learning Representations (ICLR)*, 2020. URL <https://arxiv.org/abs/2010.10952>.
- [174] Dmitry Laptev, Nikolay Savinov, Joachim M. Buhmann, and Marc Pollefeys. Ti-pooling: Transformation-invariant pooling for feature learning in convolutional neural networks. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [175] Yann LeCun, Bernhard E. Boser, John S. Denker, Donnie Henderson, R. E. Howard, Wayne E. Hubbard, and Lawrence D. Jackel. Handwritten digit recognition with a back-propagation network. 1990.
- [176] Jongmin Lee, Byungjin Kim, Seungwook Kim, and Minsu Cho. Learning rotation-equivariant features for visual correspondence.
- [177] Jongmin Lee, Byungjin Kim, and Minsu Cho. Self-supervised equivariant learning for oriented keypoint detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4847–4857, 2022.
- [178] Yeonkun Lee, Jaeseok Jeong, Jongseob Yun, Wonjune Cho, and Kuk-Jin Yoon. Spheredphd: Applying cnns on a spherical polyhedron representation of 360deg images. *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9181–9189, 2019.
- [179] Jan Eric Lenssen, Matthias Fey, and Pascal Libuschewski. Group equivariant capsule networks. *Conference on Neural Information Processing Systems (NeurIPS)*, 2018.
- [180] Shiwei Li, Zixin Luo, Mingmin Zhen, Yao Yao, Tianwei Shen, Tian Fang, and Long Quan. Cross-atlas convolution for parameterization invariant learning on textured mesh surface. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [181] Wan-Chiu Li, Bruno Vallet, Nicolas Ray, and Bruno Levy. Representing higher-order singularities in vector fields on piecewise linear surfaces. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):1315–1322, 2006.
- [182] Yang Li, Haidong Yi, Christopher M Bender, Siyuan Shan, and Junier B Oliva. Exchangeable neural ODE for set modeling. *arXiv preprint arXiv:2008.02676*, 2020.
- [183] Isaak Lim, Alexander Dielen, Marcel Campen, and Leif Kobbelt. A simple approach to intrinsic correspondence learning on unstructured 3d meshes. *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0, 2018.

- [184] Jiehong Lin, Hongyang Li, Ke Chen, Jiangbo Lu, and Kui Jia. Sparse steerable convolutions: An efficient learning of se (3)-equivariant features for estimation and tracking of object poses in 3d space. *Conference on Neural Information Processing Systems (NeurIPS)*, 2021.
- [185] Yaron Lipman, Olga Sorkine, David Levin, and Daniel Cohen-Or. Linear rotation-invariant coordinates for meshes. *ACM Transactions on Graphics (TOG)*, 24(3):479–487, 2005.
- [186] Beibei Liu, Yiyi Tong, Fernando De Goes, and Mathieu Desbrun. Discrete connection and covariant derivative for vector field analysis and design. *ACM Transactions on Graphics (TOG)*, 35(3):1–17, 2016.
- [187] Min Liu, Fupin Yao, Chiho Choi, Sinha Ayan, and Karthik Ramani. Deep learning 3d shapes using alt-az anisotropic 2-sphere convolution. *International Conference on Learning Representations (ICLR)*, 2019.
- [188] Shiqi Liu, Mengdi Xu, Piede Huang, Yongkang Liu, Kentaro Oguchi, and Ding Zhao. Continual reinforcement learning with group symmetries. *arXiv preprint arXiv:2210.12301*, 2022.
- [189] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [190] Sanae Lotfi, Marc Finzi, Sanyam Kapoor, Andres Potapczynski, Micah Goldblum, and Andrew Gordon Wilson. Pac-bayes compression bounds so tight that they can explain generalization. *arXiv preprint arXiv:2211.13609*, 2022.
- [191] Ameesh Makadia and Kostas Daniilidis. Rotation recovery from spherical images without correspondences. *IEEE transactions on pattern analysis and machine intelligence*, 28(7):1170–1175, 2006.
- [192] Stéphane Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, 2012.
- [193] Vincent Mallet and Jean-Philippe Vert. Reverse-complement equivariant networks for DNA sequences. *Conference on Neural Information Processing Systems (NeurIPS)*, 2021.
- [194] Tong Mao, Zhongjie Shi, and Ding-Xuan Zhou. Theory of deep convolutional neural networks iii: Approximating radial functions. *Neural Networks*, 144:778–790, 2021.
- [195] Diego Marcos, Michele Volpi, and Devis Tuia. Learning rotation invariant convolutional filters for texture classification. *International Conference on Pattern Recognition (ICPR)*, 2016.
- [196] Diego Marcos, Michele Volpi, Nikos Komodakis, and Devis Tuia. Rotation equivariant vector field networks. *International Conference on Computer Vision (ICCV)*, 2017.
- [197] Diego Marcos, Benjamin Kellenberger, Sylvain Lobry, and Devis Tuia. Scale equivariance in CNNs with vector fields. *arXiv preprint arXiv:1807.11783*, 2018.
- [198] Diego Marcos, Michele Volpi, Benjamin Kellenberger, and Devis Tuia. Land cover mapping at very high resolution with rotation equivariant cnns: Towards small yet accurate models. *ISPRS journal of photogrammetry and remote sensing*, 145:96–107, 2018.
- [199] Haggai Maron, Meirav Galun, Noam Aigerman, Miri Trope, Nadav Dym, Ersin Yumer, Vladimir G Kim, and Yaron Lipman. Convolutional neural networks on surfaces via seamless toric covers. *ACM Trans. Graph.*, 36(4):71–1, 2017.

- [200] Haggai Maron, Heli Ben-Hamu, Nadav Shamir, and Yaron Lipman. Invariant and equivariant graph networks. *International Conference on Learning Representations (ICLR)*, 2018.
- [201] Haggai Maron, Ethan Fetaya, Nimrod Segol, and Yaron Lipman. On the universality of invariant networks. *International Conference on Machine Learning (ICML)*, 2019.
- [202] Adam Marsh. Gauge theories and fiber bundles: Definitions, pictures, and results, 2016. URL <http://arxiv.org/abs/1607.03089>.
- [203] Daniel Martin, Ana Serrano, and Belen Masia. Panoramic convolutions for 360° single-image saliency prediction. 2020.
- [204] Jonathan Masci, Davide Boscaini, Michael Bronstein, and Pierre Vandergheynst. Geodesic convolutional neural networks on Riemannian manifolds. *Proceedings of the IEEE international conference on computer vision workshops*, 2015.
- [205] Jonathan Masci, Davide Boscaini, Michael Bronstein, and Pierre Vandergheynst. Shapenet: Convolutional neural networks on non-euclidean manifolds. Technical report, 2015.
- [206] Simone Melzi, Emanuele Rodolà, Umberto Castellani, and Michael M Bronstein. Localized manifold harmonics for spectral shape analysis. *Computer Graphics Forum*, 37(6):20–34, 2018.
- [207] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [208] Nicola Messina, Giuseppe Amato, Fabio Carrara, Claudio Gennaro, and Fabrizio Falchi. Recurrent vision transformer for solving visual reasoning problems. *arXiv preprint arXiv:2111.14576*, 2021.
- [209] Mark Meyer, Mathieu Desbrun, Peter Schröder, and Alan H Barr. Discrete differential-geometry operators for triangulated 2-manifolds. pages 35–57. Springer, 2003.
- [210] Francesco Milano, Antonio Loquercio, Antoni Rosinol, Davide Scaramuzza, and Luca Carlone. Primal-dual mesh convolutional neural networks. 2020.
- [211] Benjamin Kurt Miller, Mario Geiger, Tess E Smidt, and Frank Noé. Relevance of rotationally equivariant convolutions for predicting molecular properties. *arXiv preprint arXiv:2008.08461*, 2020.
- [212] James S Milne. Group theory. *Course Notes*, 1996.
- [213] Joseph SB Mitchell, David M Mount, and Christos H Papadimitriou. The discrete geodesic problem. *SIAM Journal on Computing*, 16(4):647–668, 1987.
- [214] Joshua Mitton and Roderick Murray-Smith. Rotation equivariant deforestation segmentation and driver classification. *arXiv preprint arXiv:2110.13097*, 2021.
- [215] Mirgahney Mohamed, Gabriele Cesa, Taco Cohen, and Max Welling. A data and compute efficient design for limited-resources deep learning. *International Conference on Learning Representations (ICLR)*, 2020.
- [216] Arnab Kumar Mondal, Pratheeeksha Nair, and Kaleem Siddiqi. Group equivariant deep reinforcement learning. *arXiv preprint arXiv:2007.03437*, 2020.
- [217] Federico Monti, Davide Boscaini, Jonathan Masci, Emanuele Rodola, Jan Svoboda, and Michael Bronstein. Geometric deep learning on graphs and manifolds using mixture model CNNs. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

- [218] Yusuke Mukuta and Tatsuya Harada. Self-supervised learning for group equivariant neural networks. *arXiv preprint arXiv:2303.04427*, 2023.
- [219] Mohamed Adel Musallam, Vincent Gaudillière, Miguel Ortiz del Castillo, Kassem Al Ismaeil, and Djamil Aouada. Leveraging equivariant features for absolute pose regression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6876–6886, 2022.
- [220] H. Naderi, L. Goli, and S. Kasaei. Scale equivariant CNNs with scale steerable filters. *2020 International Conference on Machine Vision and Image Processing (MVIP)*, 2020.
- [221] Mikio Nakahara. *Geometry, topology and physics*. CRC Press, 2003.
- [222] Karl-Hermann Neeb. Differential topology of fiber bundles. *FAU Erlangen-Nuernberg*, 2010.
- [223] Edouard Oyallon and Stéphane Mallat. Deep roto-translation scattering for object classification. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [224] Rasmus R Paulsen, Kristine Aavild Juhl, Thilde Marie Haspang, Thomas Hansen, Melanie Ganz, and Gudmundur Einarsson. Multi-view consensus cnn for 3d facial landmark placement. *Asian Conference on Computer Vision*, pages 706–719, 2018.
- [225] Suraj Pawar, Omer San, Adil Rasheed, and Prakash Vedula. Frame invariant neural network closures for kraichnan turbulence. *Physica A: Statistical Mechanics and its Applications*, 609:128327, 2023.
- [226] Songyou Peng, Michael Niemeyer, Lars Mescheder, Marc Pollefeys, and Andreas Geiger. Convolutional occupancy networks. *arXiv preprint arXiv:2003.04618*, 2020.
- [227] Abhishek Peri, Kinal Mehta, Avneesh Mishra, Michael Milford, Sourav Garg, and K Madhava Krishna. ReF–rotation equivariant features for local feature matching. *arXiv preprint arXiv:2203.05206*, 2022.
- [228] Nathanaël Perraudin, Michaël Defferrard, Tomasz Kacprzak, and Raphael Sgier. DeepSphere: Efficient spherical Convolutional Neural Network with HEALPix sampling for cosmological applications. *arXiv:1810.12186 [astro-ph]*, 2018.
- [229] Kaare Brandt Petersen, Michael Syskind Pedersen, et al. The matrix cookbook. *Technical University of Denmark*, 7(15):510, 2008.
- [230] Paolo Piccione and Daniel V Tausk. *The theory of connections and G-structures: applications to affine and isometric immersions*. IMPA, 2006.
- [231] Konrad Polthier and Markus Schmies. Straightest geodesics on polyhedral surfaces. *Mathematical visualization. Heidelberg: Springer Verlag*, pages 135–50, 1998.
- [232] Adrien Poulenard and Maks Ovsjanikov. Multi-directional geodesic neural networks via equivariant convolution. *SIGGRAPH Asia 2018 Technical Papers*, page 236, 2018.
- [233] Adrien Poulenard, Marie-Julie Rakotosaona, Yann Ponty, and Maks Ovsjanikov. Effective rotation-invariant point CNN with spherical harmonics kernels. *2019 International Conference on 3D Vision (3DV)*, pages 47–56, 2019.
- [234] Charles Qi, Hao Su, Kaichun Mo, and Leonidas Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [235] Charles Qi, Li Yi, Hao Su, and Leonidas Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Conference on Neural Information Processing Systems (NeurIPS)*, 2017.

- [236] Charles R Qi, Hao Su, Matthias Nißner, Angela Dai, Mengyuan Yan, and Leonidas J Guibas. Volumetric and multi-view cnns for object classification on 3d data. *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5648–5656, 2016.
- [237] Dietmar Rabich. Go (13×13) – 6737, CC BY-SA 4.0 (adapted). [https://commons.wikimedia.org/wiki/File:Go_\(13x13\)_-_2021_-_6737.jpg](https://commons.wikimedia.org/wiki/File:Go_(13x13)_-_2021_-_6737.jpg), 2021.
- [238] Sameera Ramasinghe, Salman Khan, Nick Barnes, and Stephen Gould. Representation learning on unit ball with 3d roto-translational equivariance. *International Journal of Computer Vision*, pages 1–23, 2019.
- [239] Matthias Rath and Alexandru Paul Condurache. Deep neural networks with efficient guaranteed invariances. *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2023.
- [240] Siamak Ravanbakhsh. Universal equivariant multilayer perceptrons. *International Conference on Machine Learning (ICML)*, 2020.
- [241] Siamak Ravanbakhsh, Jeff Schneider, and Barnabas Poczos. Equivariance through parameter-sharing. *International Conference on Machine Learning (ICML)*, 2017.
- [242] Nicolas Ray, Bruno Vallet, Wan Chiu Li, and Bruno Lévy. N-symmetry direction field design. *ACM Transactions on Graphics (TOG)*, 27(2):1–13, 2008.
- [243] Hans Reiter, Jan Derk Stegeman, et al. *Classical harmonic analysis and locally compact groups*. Number 22. Courier Corporation, 2000.
- [244] Danilo Jimenez Rezende, Sébastien Racanière, Irina Higgins, and Peter Toth. Equivariant Hamiltonian flows. *arXiv preprint arXiv:1909.13739*, 2019.
- [245] David Romero and Jean-Baptiste Cordonnier. Group equivariant stand-alone self-attention for vision. 2020.
- [246] David Romero and Mark Hoogendoorn. Co-attentive equivariant neural networks: Focusing equivariance on transformations co-occurring in data. *International Conference on Learning Representations (ICLR)*, 2020.
- [247] David Romero, Erik Bekkers, Jakub Tomczak, and Mark Hoogendoorn. Attentive group equivariant convolutional networks. 2020.
- [248] David Romero, Erik Bekkers, Jakub Tomczak, and Mark Hoogendoorn. Wavelet networks: Scale equivariant learning from raw waveforms. *arXiv preprint arXiv:2006.05259*, 2020.
- [249] David W Romero and Suhas Lohit. Learning equivariances and partial equivariances from data. *arXiv preprint arXiv:2110.10211*, 2021.
- [250] David W Romero, Anna Kuzina, Erik J Bekkers, Jakub M Tomczak, and Mark Hoogendoorn. CKConv: continuous kernel convolution for sequential data. *International Conference on Learning Representations (ICLR)*, 2022.
- [251] Joseph Rosen and Joseph Shamir. Circular harmonic phase filters for efficient rotation-invariant pattern recognition. *Applied Optics*, 27(14):2895–2899, 1988.
- [252] Walter Rudin. *Principles of mathematical analysis*. McGraw-hill, Inc. New York, 1976.
- [253] Sara Sabour, Nicholas Frosst, and Geoffrey E. Hinton. Dynamic routing between capsules. In *Conference on Neural Information Processing Systems (NIPS)*, 2017.
- [254] Edward Saff and Arno Kuijlaars. Distributing many points on a sphere. *The mathematical intelligencer*, 19(1):5–11, 1997.

- [255] Yoshiharu Saito, Shin-ichi Komatsu, and Hitoshi Ohzu. Scale and rotation invariant real time optical correlator using computer generated hologram. *Optics Communications*, 47(1):8–11, 1983.
- [256] Akiyoshi Sannai, Yuuki Takai, and Matthieu Cordonnier. Universal approximations of permutation invariant/equivariant functions by deep neural networks. *arXiv preprint arXiv:1903.01939*, 2019.
- [257] Alessandro Sarti, Giovanna Citti, and Jean Petitot. Functional geometry of the horizontal connectivity in the primary visual cortex. *Journal of Physiology-Paris*, 103 (1-2):37–45, 2009.
- [258] Anna M. M. Scaife and Fiona Porter. Fanaroff-riley classification of radio galaxies using group-equivariant convolutional neural networks. 2021.
- [259] Stefan C Schonsheck, Bin Dong, and Rongjie Lai. Parallel transport convolution: A new tool for convolutional neural networks on manifolds. *arXiv preprint arXiv:1805.07857*, 2018.
- [260] Lukas Schott, Julius von Kügelgen, Frederik Träuble, Peter Gehler, Chris Russell, Matthias Bethge, Bernhard Schölkopf, Francesco Locatello, and Wieland Brendel. Visual representation learning does not generalize strongly within the same domain. *arXiv preprint arXiv:2107.08221*, 2021.
- [261] Peter Schröder and Wim Sweldens. Spherical wavelets: Efficiently representing functions on the sphere. *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 161–172, 1995.
- [262] Frederic Schuller. Lectures on the Geometrical Anatomy of Theoretical Physics, 2016.
- [263] Kristof T Schütt, Huziel E Sauceda, P-J Kindermans, Alexandre Tkatchenko, and K-R Müller. Schnet – a deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148(24):241722, 2018.
- [264] Nimrod Segol and Yaron Lipman. On universal equivariant set networks. *International Conference on Learning Representations (ICLR)*, 2020.
- [265] Giuseppe Sellaroli. Wigner-eckart theorem and jordan-schwinger representation for infinite-dimensional representations of the lorentz group. *arXiv preprint arXiv:1509.05633*, 2015.
- [266] Ahyun Seo, Byungjin Kim, Suha Kwak, and Minsu Cho. Reflection and rotation symmetry detection via equivariant learning. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
- [267] Jean-Pierre Serre. Linear representations of finite groups. 1977.
- [268] Nicholas Sharp, Yousuf Soliman, and Keenan Crane. The vector heat method. *ACM Trans. Graph.*, 38(3), 2019.
- [269] Nicholas Sharp, Souhaib Attaiki, Keenan Crane, and Maks Ovsjanikov. Diffusion is all you need for learning on surfaces. *arXiv preprint arXiv:2012.00888*, 2020.
- [270] Zhengyang Shen, Lingshen He, Zhouchen Lin, and Jinwen Ma. PDO-eConvs: Partial differential operator based equivariant convolutions. *International Conference on Machine Learning (ICML)*, 2020.
- [271] Zhengyang Shen, Tao Hong, Qi She, Jinwen Ma, and Zhouchen Lin. PDO-s3dCNNs: Partial differential operator based steerable 3d CNNs. *International Conference on Machine Learning (ICML)*, 2022.

-
- [272] Masato Shiba, Fujimaro Ishida, Kazuhiro Furukawa, Hiroshi Tanemura, Masanori Tsuji, Shinichi Shimosaka, and Hidenori Suzuki. Relationships of morphologic parameters and hemodynamic parameters determined by computational fluid dynamics analysis with the severity of subarachnoid hemorrhage. *Journal of Neuroendovascular Therapy*, 06 2017. doi: 10.5797/jnet.oa.2016-0099.
 - [273] Katsumi Nomizu Shoshichi Kobayashi. *Foundations of Differential Geometry (Volume 1)*. 1963.
 - [274] Noah Shutty and Casimir Wierzynski. Learning irreducible representations of non-commutative lie groups. *arXiv preprint arXiv:2006.00724*, 2020.
 - [275] Laurent Sifre and Stéphane Mallat. Combined scattering for rotation invariant texture analysis. *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, 44:68–81, 2012.
 - [276] Laurent Sifre and Stéphane Mallat. Rotation, scaling and deformation invariant scattering for texture discrimination. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.
 - [277] Laurent Sifre and Stéphane Mallat. Rigid-motion scattering for texture classification. *arXiv preprint arXiv:1403.1687*, 2014.
 - [278] Ayan Sinha, Jing Bai, and Karthik Ramani. Deep learning 3d shape surfaces using geometry images. *European Conference on Computer Vision (ECCV)*, 2016.
 - [279] Bart Smets, Jim Portegies, Erik Bekkers, and Remco Duits. Pde-based group equivariant convolutional neural networks. *arXiv preprint arXiv:2001.09046*, 2020.
 - [280] Stefan Sommer and Alex Bronstein. Horizontal flows and manifold stochastics in geometric deep learning. 2019.
 - [281] Ivan Sosnovik, Michał Szmaja, and Arnold Smeulders. Scale-equivariant steerable networks. *International Conference on Learning Representations (ICLR)*, 2020.
 - [282] Michael Spivak. *Calculus*. Reverté, 2019.
 - [283] Norman Steenrod. *The Topology of Fibre Bundles*. 1951.
 - [284] Shlomo Sternberg. *Lectures on differential geometry*, volume 316. American Mathematical Soc., 1999.
 - [285] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. *Proceedings of the IEEE international conference on computer vision*, pages 945–953, 2015.
 - [286] Shuai Su, Zhongkai Zhao, Yixin Fei, Shuda Li, Qijun Chen, and Rui Fan. Sim2e: Benchmarking the group equivariant capability of correspondence matching algorithms. *arXiv preprint arXiv:2208.09896*, 2022.
 - [287] Yu-Chuan Su and Kristen Grauman. Learning spherical convolution for fast features from 360 imagery. *Conference on Neural Information Processing Systems (NeurIPS)*, 2017.
 - [288] Yu-Chuan Su and Kristen Grauman. Kernel transformer networks for compact spherical convolution. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
 - [289] Zhaoyu Su, Pin Siang Tan, Junkang Chow, Jimmy Wu, Yehur Cheong, and Yu-Hsing Wang. DV-ConvNet: Fully convolutional deep learning on point clouds with dynamic voxelization and 3d group convolution. *arXiv preprint arXiv:2009.02918*, 2020.

- [290] Zhuo Su, Linpu Fang, Wenxiong Kang, Dewen Hu, Matti Pietikäinen, and Li Liu. Dynamic group convolution for accelerating convolutional neural networks. *European Conference on Computer Vision (ECCV)*, 2020.
- [291] Julian Suk, Pim de Haan, Phillip Lippe, Christoph Brune, and Jelmer M. Wolterink. Mesh convolutional neural networks for wall shear stress estimation in 3d artery models. *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 93–102, 2021.
- [292] Chen Sun, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta. Revisiting unreasonable effectiveness of data in deep learning era. 2017.
- [293] Zhiyu Sun, Ethan Rooke, Jerome Charton, Yusen He, Jia Lu, and Stephen Baek. Zernet: Convolutional neural networks on arbitrary surfaces via zernike local tangent space estimation. 2018.
- [294] Maxim Tatarchenko, Jaesik Park, Vladlen Koltun, and Qian-Yi Zhou. Tangent convolutions for dense prediction in 3d. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [295] Keisuke Tateno, Nassir Navab, and Federico Tombari. Distortion-aware convolutional filters for dense prediction in panoramic images. *European Conference on Computer Vision (ECCV)*, 2018.
- [296] Max Tegmark, Angelica de Oliveira-Costa, and Andrew JS Hamilton. High resolution foreground cleaned CMB map from WMAP. *Physical Review D*, 68(12):123523, 2003.
- [297] Constantin Teleman. Representation theory, 2005. URL <https://math.berkeley.edu/~teleman/math/RepThry.pdf>.
- [298] Erik Thiede, Wenda Zhou, and Risi Kondor. Autobahn: Automorphism-based graph neural nets. *Conference on Neural Information Processing Systems (NeurIPS)*, 2021.
- [299] Erik Henning Thiede, Truong Son Hy, and Risi Kondor. The general theory of permutation equivariant neural networks and higher order graph variational encoders. 2020.
- [300] Hugues Thomas, Charles R Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J Guibas. Kpconv: Flexible and deformable convolution for point clouds. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6411–6420, 2019.
- [301] Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation-and translation-equivariant neural networks for 3D point clouds. *arXiv preprint arXiv:1802.08219*, 2018.
- [302] Ilya O Tolstikhin, Neil Houlsby, Alexander Kolesnikov, Lucas Beyer, Xiaohua Zhai, Thomas Unterthiner, Jessica Yung, Andreas Steiner, Daniel Keysers, Jakob Uszkoreit, et al. MLP-Mixer: An all-MLP architecture for vision. *Conference on Neural Information Processing Systems (NeurIPS)*, 2021.
- [303] Wu-Ki Tung. *Group theory in physics*, volume 1. World Scientific, 1985.
- [304] Greg Turk. Texture synthesis on surfaces. *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 347–354, 2001.
- [305] Sharvaree Vadgama, Jakub Mikolaj Tomczak, and Erik J Bekkers. Kendall shape-vae: Learning shapes in a generative framework. In *NeurIPS 2022 Workshop on Symmetry and Geometry in Neural Representations*, 2022.

- [306] Elise van der Pol, Daniel E. Worrall, Herke van Hoof, Frans A Oliehoek, and Max Welling. MDP homomorphic networks: Group symmetries in reinforcement learning. *arXiv preprint arXiv:2006.16908*, 2020.
- [307] Bastiaan S Veeling, Jasper Linmans, Jim Winkens, Taco Cohen, and Max Welling. Rotation equivariant cnns for digital pathology. In *International Conference on Medical image computing and computer-assisted intervention*, pages 210–218. Springer, 2018.
- [308] Sai Raam Venkataraman, S. Balasubramanian, and R. Raghunatha Sarma. Building deep equivariant capsule networks. *International Conference on Learning Representations (ICLR)*, 2020.
- [309] Nitika Verma, Edmond Boyer, and Jakob Verbeek. Feastnet: Feature-steered graph convolutions for 3d shape analysis. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [310] Etienne Vouga. Lectures in discrete differential geometry 3 – discrete surfaces, 2014.
- [311] Robin Walters, Jinxi Li, and Rose Yu. Trajectory prediction using equivariant continuous convolution. *arXiv preprint arXiv:2010.11344*, 2020.
- [312] Dian Wang, Mingxi Jia, Xupeng Zhu, Robin Walters, and Robert Platt. On-robot learning with equivariant models. 2022.
- [313] Dian Wang, Jung Yeon Park, Neel Sortur, Lawson LS Wong, Robin Walters, and Robert Platt. The surprising effectiveness of equivariant models in domains with latent symmetry. *arXiv preprint arXiv:2211.09231*, 2022.
- [314] Dian Wang, Robin Walters, and Robert Platt. SO(2)-equivariant reinforcement learning. *International Conference on Learning Representations (ICLR)*, 2022.
- [315] Dian Wang, Robin Walters, Xupeng Zhu, and Robert Platt. Equivariant q learning in spatial action spaces. In *Conference on Robot Learning*, pages 1713–1723. PMLR, 2022.
- [316] Renhao Wang, Marjan Albooyeh, and Siamak Ravanchkhsh. Equivariant networks for hierarchical structures. *Advances in Neural Information Processing Systems*, 33: 13806–13817, 2020.
- [317] Rui Wang, Robin Walters, and Rose Yu. Incorporating symmetry into deep dynamics models for improved generalization. *arXiv:2002.03061*, 2020.
- [318] Rui Wang, Robin Walters, and Rose Yu. Data augmentation vs. equivariant networks: A theory of generalization on dynamics forecasting. *arXiv preprint arXiv:2206.09450*, 2022.
- [319] Rui Wang, Robin Walters, and Rose Yu. Physics-guided deep learning for spatiotemporal forecasting. In *Knowledge-Guided Machine Learning*, pages 179–210. Chapman and Hall/CRC, 2023.
- [320] Yuanzhen Wang, Beibei Liu, and Yiying Tong. Linear surface reconstruction from discrete fundamental forms on triangle meshes. *Computer Graphics Forum*, 31(8): 2277–2287, 2012.
- [321] Maurice Weiler. Learning steerable filters for rotation equivariant convolutional neural networks. Master’s thesis, University of Heidelberg, 2017.
- [322] Maurice Weiler and Gabriele Cesa. General E(2)-equivariant steerable CNNs. *Conference on Neural Information Processing Systems (NeurIPS)*, 2019. URL <https://arxiv.org/abs/1911.08251>.

- [323] Maurice Weiler, Mario Geiger, Max Welling, Wouter Boomsma, and Taco S. Cohen. 3D steerable CNNs: Learning rotationally equivariant features in volumetric data. *Conference on Neural Information Processing Systems (NeurIPS)*, 2018. URL <https://arxiv.org/abs/1807.02547>.
- [324] Maurice Weiler, Fred A. Hamprecht, and Martin Storath. Learning steerable filters for rotation equivariant CNNs. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. URL <https://arxiv.org/abs/1711.07289>.
- [325] Maurice Weiler, Patrick Forré, Erik Verlinde, and Max Welling. Coordinate independent convolutional networks – isometry and gauge equivariant convolutions on Riemannian manifolds. *arXiv preprint arXiv:2106.06020*, 2021. URL <https://arxiv.org/abs/2106.06020>.
- [326] Chris Wendl. *Lecture Notes on Bundles and Connections*. 2008. URL <https://www.mathematik.hu-berlin.de/~wendl/connections.html>.
- [327] Ruben Wiersma, Elmar Eisemann, and Klaus Hildebrandt. CNNs on surfaces using rotation-equivariant features. *Transactions on Graphics*, 39(4), July 2020. doi: 10.1145/3386569.3392437.
- [328] Eugene P Wigner. The unreasonable effectiveness of mathematics in the natural sciences. pages 291–306. World Scientific, 1990.
- [329] Marysia Winkels and Taco S. Cohen. 3D G-CNNs for pulmonary nodule detection. *Conference on Medical Imaging with Deep Learning (MIDL)*, 2018.
- [330] Anderson M Winkler, Peter Kochunov, John Blangero, Laura Almasy, Karl Zilles, Peter T Fox, Ravindranath Duggirala, and David C Glahn. Cortical thickness or grey matter volume? the importance of selecting the phenotype for imaging genetics studies. *Neuroimage*, 53(3):1135–1146, 2010.
- [331] Robin Winter, Marco Bertolini, Tuan Le, Frank Noé, and Djork-Arné Clevert. Unsupervised learning of group invariant and equivariant representations. *arXiv preprint arXiv:2202.07559*, 2022.
- [332] Jeffrey Wood and John Shawe-Taylor. Representation theory and invariant neural networks. *Discrete applied mathematics*, 69(1-2):33–60, 1996.
- [333] Daniel E. Worrall and Gabriel J. Brostow. Cubenet: Equivariance to 3D rotation and translation. *European Conference on Computer Vision (ECCV)*, 2018.
- [334] Daniel E. Worrall and Max Welling. Deep scale-spaces: Equivariance over scale. *Conference on Neural Information Processing Systems (NeurIPS)*, 2019.
- [335] Daniel E. Worrall, Stephan J. Garbin, Daniyar Turmukhambetov, and Gabriel J. Brostow. Harmonic networks: Deep translation and rotation equivariance. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [336] Jin Xu, Hyunjik Kim, Thomas Rainforth, and Yee Teh. Group equivariant subsampling. *Conference on Neural Information Processing Systems (NeurIPS)*, 2021.
- [337] Yinshuang Xu, Jiahui Lei, Edgar Dobriban, and Kostas Daniilidis. Unified Fourier-based kernel and nonlinearity design for equivariant networks on homogeneous spaces. *International Conference on Machine Learning (ICML)*, 2022.
- [338] Qin Yang, Chenglin Li, Wenrui Dai, Junni Zou, Guo-Jun Qi, and Hongkai Xiong. Rotation equivariant graph convolutional network for spherical image classification. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

- [339] Yuqi Yang, Shilin Liu, Hao Pan, Yang Liu, and Xin Tong. PFCNN: Convolutional neural networks on 3d surfaces using parallel frames. *Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [340] Dmitry Yarotsky. Universal approximations of invariant maps by neural networks. *Constructive Approximation*, 55(1):407–474, 2022.
- [341] Yuki Yasuda. Roto-translation equivariant super-resolution of two-dimensional flows using convolutional neural networks. *arXiv preprint arXiv:2202.11099*, 2022.
- [342] Jeffrey Yepez. Einstein’s vierbein field theory of curved space. *arXiv preprint arXiv:1106.2037*, 2011.
- [343] Lexing Ying, Aaron Hertzmann, Henning Biermann, and Denis Zorin. Texture and shape synthesis on surfaces. *Eurographics Workshop on Rendering Techniques*, pages 301–312, 2001.
- [344] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. *arXiv preprint arXiv:1605.07146*, 2016.
- [345] Binbin Zhang, Wen Shen, Shikun Huang, Zhihua Wei, and Quanshi Zhang. 3d-rotation-equivariant quaternion neural networks. *arXiv preprint arXiv:1911.09040*, 2019.
- [346] Chao Zhang, Stephan Liwicki, William Smith, and Roberto Cipolla. Orientation-aware semantic segmentation on icosahedron spheres. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [347] Eugene Zhang, Konstantin Mischaikow, and Greg Turk. Vector field design on surfaces. *ACM Transactions on Graphics (ToG)*, 25(4):1294–1326, 2006.
- [348] Richard Zhang. Making convolutional networks shift-invariant again. *International Conference on Machine Learning (ICML)*, 2019.
- [349] Xinhua Zhang and Lance R Williams. Similarity equivariant linear transformation of joint orientation-scale space representations. *arXiv preprint arXiv:2203.06786*, 2022.
- [350] Linfeng Zhao, Xupeng Zhu, Lingzhi Kong, Robin Walters, and Lawson LS Wong. Integrating symmetry into differentiable planning with steerable convolutions. 2023.
- [351] Qiang Zhao, Chen Zhu, Feng Dai, Yike Ma, Guoqing Jin, and Yongdong Zhang. Distortion-aware CNNs for spherical images. *IJCAI*, 2018.
- [352] Yongheng Zhao, Tolga Birdal, Jan Eric Lenssen, Emanuele Menegatti, Leonidas Guibas, and Federico Tombari. Quaternion equivariant capsule networks for 3d point clouds. *arXiv preprint arXiv:1912.12098*, 2019.
- [353] Maksim Zhdanov, Nico Hoffmann, and Gabriele Cesa. Implicit neural convolutional kernels for steerable cnns. *arXiv preprint arXiv:2212.06096*, 2022.
- [354] Allan Zhou, Tom Knowles, and Chelsea Finn. Meta-learning symmetries by reparameterization. *arXiv preprint arXiv:2007.02933*, 2020.
- [355] Ding-Xuan Zhou. Theory of deep convolutional neural networks: Downampling. *Neural Networks*, 124:319–327, 2020.
- [356] Ding-Xuan Zhou. Universality of deep convolutional neural networks. *Applied and computational harmonic analysis*, 48(2):787–794, 2020.
- [357] Tao Zhou. Gauge aspect of tetrad field in gravity. *arXiv preprint arXiv:1603.07571*, 2016.
- [358] Yanzhao Zhou, Qixiang Ye, Qiang Qiu, and Jianbin Jiao. Oriented response networks. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

- [359] Wei Zhu, Qiang Qiu, Robert Calderbank, Guillermo Sapiro, and Xiuyuan Cheng. Scale-equivariant neural networks with decomposed convolutional filters. *arXiv preprint arXiv:1909.11193*, 2019.
- [360] Xupeng Zhu, Dian Wang, Ondrej Biza, Guanang Su, Robin Walters, and Robert Platt. Sample efficient grasp learning using equivariant models. *arXiv preprint arXiv:2202.09468*, 2022.