Eindhoven University of Technology

MASTER

Axiomatic PDE-G-CNN on SE(2)

Sakata, Sei

Award date:
2023

Technische Universiteit
Eindhoven
University of Technology

Axiomatic PDE-G-CNN on SE(2)

Sei Sakata

Master Thesis in Applied Mathematics
November 2023

Supervisor: Remco Duits
Co-supervisor: Gijs Bellaard

Contents

Chapter 1
Introduction

The field of machine learning is currently undergoing rapid development, with applications in various areas, including image processing. In particular, Convolutional Neural Networks (CNNs) are widely used for image processing, demonstrating state-of-the-art performance in various tasks, such as segmentation. CNNs differ from traditional fully connected neural networks, as they utilize convolutions with kernels, which provide shift-invariance. This shift-invariance, indicating that the output remains unaffected by translating the image, embodies translational symmetry, which effectively conserves network capacity. The ability to recognize patterns regardless of their position in the input is one of the key factors contributing to the success and widespread use of CNNs as this feature of CNNs more closely mimics the way humans recognise patterns [XV21].

As a result, there has been great attention on generalising this concept and designing neural networks that are equivariant (meaning that the output of the transformed image is the same as the transformed version of the output) with respect to a group operation that represents geometric transformations which are not restricted to translations. This type of neural network is called a group equivariant convolutional neural network (G-CNN) [CW16] and CNNs are a particular kind of G-CNN where the group operation is a translation on $\mathbb{R}^2$.

However, despite their effectiveness, these network structures lack mathematical interpretability as their network weights and kernels have limited geometric meaning, making them challenging to comprehend. Additionally, CNNs and G-CNNs require a significant amount of training data to train due to the large amount of trainable parameters. This is due to the fact that the kernels that are used are unrestricted and hence for each kernel, there are as many parameters as the size of the kernel. One approach to resolving this issue is restricting the space of allowed kernels by a priori designating what sort of shape the kernels can take. This way the parameters that control the shape of the kernels now become the new network parameters and these can be much smaller in quantity.

This idea along with the other shortcomings of CNN was one of the reasons that led to the development of PDE-G-CNNs [Sme+22] in which PDEs are used to guide the convolutional operations, allowing the network to learn and extract features from the input data in a more efficient and structured way. It essentially places geometric meaning on the kernels used in layers by stipulating that they obey evolution PDEs. These PDEs are not arbitrary and have been investigated in the context of geometric image analysis. These equations produce theoretically meaningful solutions as the training is done on sparse sets of association fields derived from neurogeometry.

In the PDE-G-CNN framework, the usual non-linear operations such as ReLU activation functions or min-max pooling present in CNNs are replaced by solvers for the evolution PDEs that establish specific image operations in each layer. In Figure 1.1 a traditional CNN and G-CNN layer is compared with a PDE-G-CNN layer. The PDEs that govern the evolution are defined on the underlying space that data lives in. Therefore, it is not restricted to $\mathbb{R}^2$ and in fact, they can be defined on so-called *homogeneous spaces* where symmetries that are not restricted to rotation or translation can arise (where the symmetries act on the homogeneous space as group action). As a generalisation of G-CNNs, PDE-G-CNNs too are equivariant with respect to group operations that represent geometric transformations on the underlying homogeneous space.

The homogeneous space that is of main interest in this project is the space of positions and orientations. Namely, an image is lifted to the space in which the newly lifted image contains not only information about the position of points belonging to the image but also how much the image is locally aligned with a specific orientation at each point. Topologically, this space is given by the product $\mathbb{R}^2 \times S^1$ where the $\mathbb{R}^2$ part represents the positional component and S^1 the directional. This lifting of an image to incorporate information about the

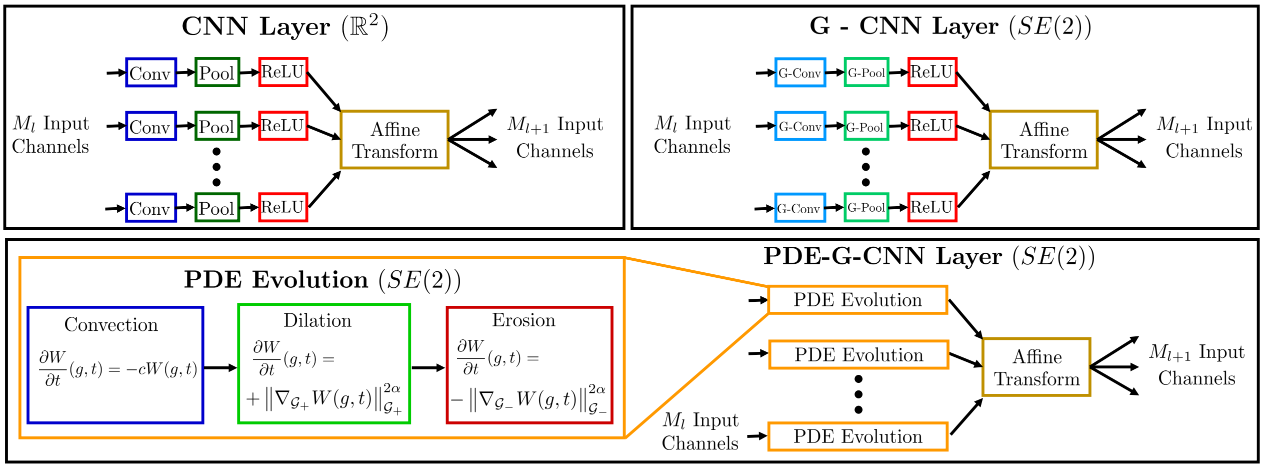

Figure 1.1: A schematic for a CNN, G-CNN and the PDE-G-CNN Layer in a deep neural network. CNNs are typically processed with $\mathbb{R}^2$ convolutions, whereas G-CNNs work with linear G-convolutions in the Lie group $G = SE(2)$. Elements in $SE(2)$ are denoted by $g = (x, y, \theta)$. In PDE-G-CNNs only the convection vector c and the metric parameters $\mathcal{G}_+, \mathcal{G}_-$ are *learned* and they lead to kernels that are used for non-linear morphological convolutions that solve the respective Erosion and Dilation PDEs in the Lie group $G \equiv SE(2)$.

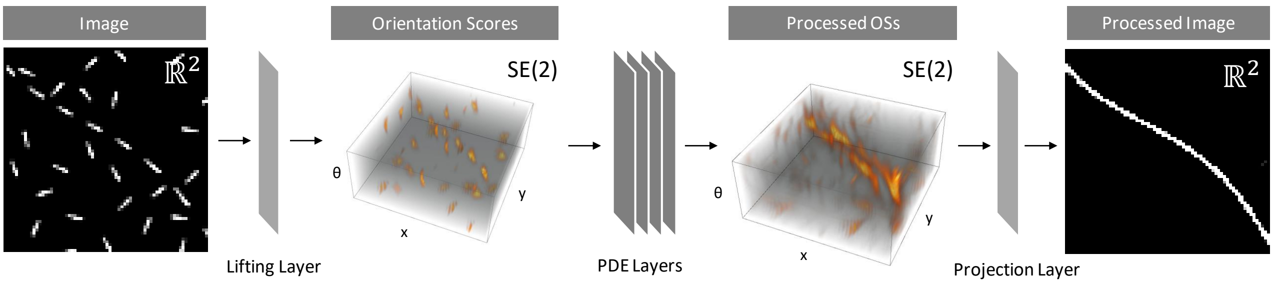

Figure 1.2: An overview of a PDE-G-CNN performing line completion on the Lines dataset [Bel+23]. First, the image is lifted to an orientation score on $G = SE(2)$, then multiple PDE-G-CNN layers are applied (Fig. 1.1), after which the result is projected down to $\mathbb{R}^2$.

directional features of images is called *orientation score* of the image. See Figure 1.3 for a visualization of a lifted image. It has been shown for example by [XV21] that performing image processing on lifted images via the orientation score improves performance.

PDE-based Convolutional Neural Networks (PDE-CNNs) are a particular case of PDE-G-CNNs where we take the underlying group G to be $\mathbb{R}^n$.In other words, in PDE-CNN, the image processing is only performed on the unlifted original image on $\mathbb{R}^2$. PDE-based Convolutional Neural Networks use only the PDE-based structure for each layer, resulting in a model that can be geometrically interpreted based on these PDEs.

The four primary types of PDEs utilized in a PDE-G-CNN are diffusion, convection, dilation, and erosion. In general, these PDEs correspond to mathematical image processing operations called scale space operations. Namely, these PDEs are associated with the scale-space operations of smoothing(regularization), shifting, max pooling, and min pooling respectively and there are many families of evolution PDEs to choose from, for each of the types.

In order to systematically narrow down the scale space operations of interest, an axiomatic approach for $\mathbb{R}^2$ case has been made by Duits [Dui+04] by establishing scale space axioms which are architectural properties considered desirable in geometric image analysis, that the scale space operations need to satisfy. It also was shown by Pauwels [Pau+95] that for the smoothing operation (or more precisely, scale space operation that is given by an integral operator with a kernel), a number of basic assumptions like semi-group property, isotropy

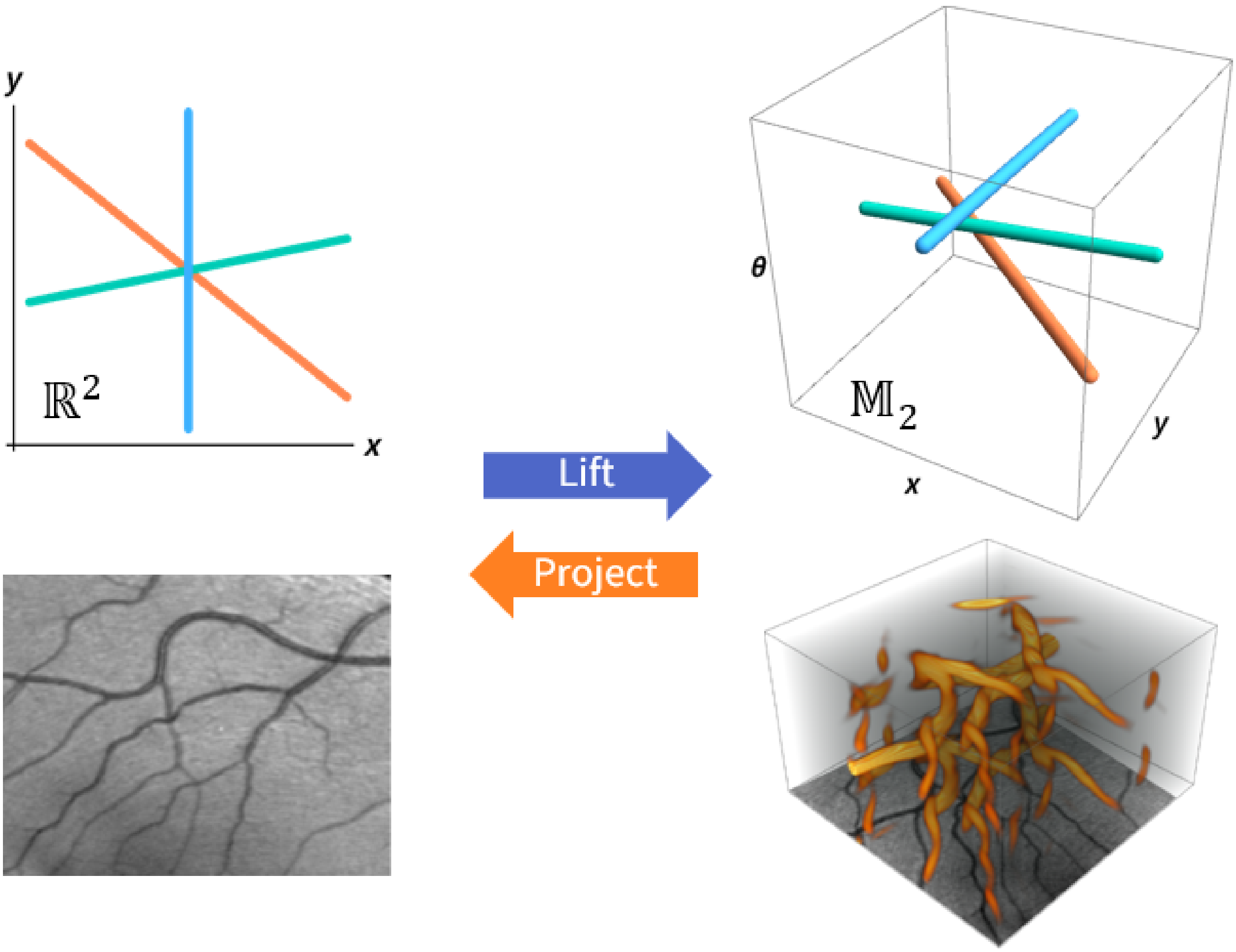

Figure 1.3: An instance of an image paired with its orientation score reveals a transformation from a 2D image to a 3D object, known as the orientation score. The core focus of PDE-G-CNNs involves processing these orientation scores. It's worth noting that the intersecting lines visible in the original image become untangled within the orientation score representation.

and scale-equivariance significantly narrow down the kernel used for the scale space operation to a single family which essentially only depends on one parameter. In Chapter 2 we explain this in more detail.

According to [Pai+23], the PDE-G-CNNs have greatly reduced network complexity, which results in reduced training data requirements. Even with the reduced network complexity, it was shown that PDE-G-CNNs can compete or even outperform both G-CNNs and CNNs. While PDE-CNNs do not utilize the lifted space of scales and orientations like their parent PDE-G-CNNs, it was shown by Castella [Cas21] that this technique also significantly reduces the number of trainable parameters required compared to conventional CNNs. Therefore, it seems that in terms of the reduction in network complexity, the "PDE" benefit seems larger than the "group-equivariant" benefit while the latter also contributes.

1.1 Research Objectives

Despite the key role PDEs play in the success of PDE-G-CNNs and PDE-CNNs, there has not been a systematic approach to pin down the relevant PDEs in a manner done for the scale space axioms. We, therefore, have the following research goal:

1. We want to underpin the PDEs used in PDE-G-CNNs[Cas21]. We want to derive them from geometrically motivated geometric learning axioms that we will set up in a similar way as scale space theory followed from their axioms [Dui+04].

2. In order to investigate the importance of the 'PDE' aspect of PDE-G-CNN, experiment if the PDE-CNNs have similar benefits as PDE-G-CNNs when it comes to training data reduction. It was already reported by [Cas21] that PDE-CNNs yield a reduction in network complexity compared to CNNs with similar performance.

For our research goal, we will establish this through a general framework. In the previous theoretical framework, although some theoretical work to relate diffusion, and dilation/erosion was established by means of the Cramer transform by Schmidt and Weickert [SW16], they were considered to arise from entirely different operations (namely, linear convolution, morphological convolution and convection). We will show in Chapter 3 upon the introduction of the notion of semiring that these different operations can be put under the same framework at least for the case $G = \mathbb{R}^2$ by stipulating that they all arise from the general notion of semiring integral operator and that different operations correspond to different underlying semirings. The semirings that we will work with are the linear semiring R_L which is $\mathbb{R}$ with ordinary addition and multiplication operations and the tropical semirings T_- and T_+. For these semirings, the addition operation is replaced with min/max operations and the multiplication operation with the usual addition respectively. We will give a formal treatment in Chapter 3.

1.2 Research Contributions

Our contribution can be divided into two parts, theoretical achievements, and experimental findings. Namely, we have

1. Fully axiomatic derivation of PDE-G-CNN for Lie group G and semiring R for the cases:

   (a) $G = \mathbb{R}^2$, $R = R_L$.
   Section 4.3, Theorem 4.2

   (b) $G = \mathbb{R}^2$, $R = T_-$, $R = T_+$.
   Section 4.3, Theorem 4.2

   (c) $G = SE(2)$, $R = R_L$.
   5.2, Theorem 5.3

2. Experiment showing the data efficiency of PDE-CNNs and PDE-G-CNNs. Namely, the reduced network complexity and training data.

   Chapter 6, Table 6.4 and Table 6.1.

Chapter 2
Preliminary Theory

While the theory of PDE-CNNs requires relatively simple mathematical technology, the case $G = SE(2)$ relies heavily on the theory of Lie groups and Riemannian geometry. We assume that the reader is familiar with basic concepts from these subjects and we therefore omit a thorough treatment and only state definitions and results that are central to this thesis. In this section, we start by briefly introducing homogeneous spaces and how they can arise from Lie groups. We then introduce the notion of group equivariance and then we will define the main Lie group of interest $SE(2)$. We will then introduce the notion of operator-valued Fourier transform which will be needed for defining the Fourier transform on $SE(2)$. We will conclude this section by having a close look at the theory of scale-spaces which will be central to this thesis and our theory will be built upon this axiomatically.

2.1 Homogeneous Spaces

The notion of "symmetries" in a space can be captured by the notion of homogeneous space and a group G which can be thought of as consisting of spacial transformations on the space acting on it. This is a general notion which is not restricted to the smooth setting. In this thesis, we will only concern ourselves with $\mathbb{R}^2$ and $\mathbb{R}^2 \times S^1$, however, we would still like to make the distinction of space in which symmetries act on with the group of symmetries. For the case of $SE(2)$, topologically, the space in which images are defined is identical to the space of transformations on the underlying space. We however make a distinction as in general, the roles that they play are very different. We therefore introduce the more general notion of homogeneous spaces and groups acting on it.

We say that a group action on a set is transitive if each orbit equals the entire set.

Definition 2.1. *Let G be a group acting on a topological space M. Then if G acts transitively on M, then we call M a homogeneous G space.*

In this thesis, we always assume every space involved is smooth and thus the group G is assumed to be a Lie group.

Lemma 2.1. *If a Lie group G acts transitively on a smooth manifold M, then M is diffeomorphic to the quotient of G by the isotropy group (i.e. the stabilizer subgroup) G_x.*

Note that we can speak of "the" isotropy group as transitive group action implies that isotropy groups are conjugates and thus $G_x \cong G_y$ for all $x, y \in M$. We, therefore, get that

$$M \cong G/G_x$$

where x is any element of M. Intuitively speaking, we may consider the group G to be consisting of transformations on a space M and homogeneity imposes that for any two elements x and y of M, there is a transformation that maps one to the other or vice versa.

A notable example of a homogeneous G space that we are concerned with in this thesis is G itself: As the isotropy groups only consist of the identity element, we get that $G \cong G/\{e\}$.

2.2 Group Equivariance

A central notion that our theory will rely heavily on is that of group equivariance.

Definition 2.2. *Let group G act on a set M and N. If a map $f : M \to N$ satisfies $f(x \cdot g) = g \cdot f(x)$ for all $g \in G$ and $x \in M$, then we say that f is an G-equivariant map.*

In this thesis, we discuss the concept of group equivariance in image processing operations. These image processing operations can be seen as an operator between the space of images which are realised as functions on a homogeneous space. We must therefore first make sense of what group equivariance means for operators between function spaces.

Given a space M, consider a set of real(or complex)-valued functions on M denoted by $\mathscr{F}(M)$. Then a G action on M canonically induces a (right) G action on $\mathscr{F}(M)$ via $\varphi \cdot g = (x \mapsto \varphi(g^{-1}x))$. We call this left-regular representation of G on $\mathscr{F}(M)$. Now the notion of group equivariance also becomes natural in this setting:

Definition 2.3. *Let group G act on spaces M and N and let $\mathscr{F}(M)$ and $\mathscr{F}(N)$ be sets containing functions acting on X and Y respectively. Let $\Phi : \mathscr{F}(M) \to \mathscr{F}(N)$ be a map. Then we say that Φ is equivariant if $\Phi(\varphi \cdot g) = \Phi(\varphi) \cdot g$ for all $g \in G$.*

2.3 Left Invariant Frames and Left Invariant Metric

For each $g \in G$, the G action on M defines a smooth map $L_g : M \to M,\ x \to x \cdot g$ and we call L_g left translation by g. In this section, we always assume that M is a homogeneous G space. Let M be a smooth manifold and by $\Gamma(M, E)$, we denote the set of smooth section on a vector bundle $E \to M$.

Definition 2.4. *Let $\mathcal{T} \in \Gamma(M, E)$ a smooth section of the vector bundle $E \to X$ where the fibre $E_x = (T_x^* M)^{\otimes n}$. In other words, $\mathcal{T}$ is a tensor field of type $(n, 0)$. Then we say that $\mathcal{T}$ is G-invariant if*

$$\mathcal{T}(X_1, \ldots, X_n)(x) = \mathcal{T}((L_g)_* X_1, \ldots, (L_g)_* X_n)(x \cdot g)$$

for all $X_i \in T_x M$ for $i = 1, \ldots n$.

A G left-invariant tensor field we are particularly interested in are left-invariant tensor fields that define Riemannian metric i.e $g \in \Gamma(M, TM^* \otimes TM^*)$ so that $g_x(X_1, X_2) = g_{hx}((L_h)_* X_1, (L_h)_* X_2)$ for all $h \in G$ and $X_i \in T_x M$.

The standard result that L_g is an isometry yields the following that will be used frequently and tacitly.

Lemma 2.2. *For a G-invariant metric tensor field $\mathcal{G}$ defined on G, we have for any $g, h \in G$,*

$$d(g, h) = d(h^{-1}g, e) = d(e, g^{-1}h),$$

where the induced Riemannian distance is given by $d(g, h) = d_{\mathcal{G}}(g, h)$ given by

$$d_{\mathcal{G}}(g, h) = \inf_{\substack{\gamma(0) = g,\ \gamma(1) = h, \\ \gamma \in PC([0,1], G)}} Length_{\mathcal{G}}(\gamma), \tag{2.1}$$

where $Length_{\mathcal{G}}(\gamma) := \int_0^1 \sqrt{\mathcal{G}_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))}\,\mathrm{d}t$, *and where the set of curves* $PC([0,1], G)$ *over which is optimised is the set consisting of piecewise continuously differentiable curves* $\gamma : [0,1] \to G$.

2.4 The Lie Group $SE(2)$

As we have noted earlier, the main homogeneous space of our interest in this thesis apart from the standard $\mathbb{R}^2$ case is the space of positions and orientations given topologically by the product $\mathbb{R}^2 \times S^1$. We realise it as a homogeneous space in the following way. We first endow this space $\mathbb{R}^2 \times S^1$ with a Lie group structure: We first identify $S^1 \cong SO(2)$, the special orthogonal group and for $(x_1, R_2), (x_2, R_2) \in \mathbb{R}^2 \times SO(2)$, we defined

$$(x_1, R_1)(x_2, R_2) = (x_1 + R_1 x_2, R_1 R_2)$$

where the identity element is given by $e = (0, I)$ where I is the identity in $SO(2)$. It follows that for (x, R), the inverse is given by

$$(x, R)^{-1} = (-R^{-1}x, R^{-1}).$$

We call this Lie group $SE(2)$. From this, it is evident that $\mathbb{R}^2 \times S^1$ can be realised as a homogeneous space where the Lie group $SE(2)$ acts canonically.

As $S^1 \cong \mathbb{R}/2\pi\mathbb{Z}$ we have a chart on $SE(2)$ with image $\mathbb{R}^2 \times (0, 2\pi)$ and $\{\partial_x, \partial_y, \partial_\theta\}$ is a frame on this chart domain. We can naturally ask what the left-invariant vector fields are under this chart. They are given by

$$\mathcal{A}_1 = \cos\theta\,\partial_x + \sin\theta\,\partial_y,\;\; \mathcal{A}_2 = -\sin\theta\,\partial_x + \cos\theta\,\partial_y,\;\; \mathcal{A}_3 = \partial_\theta, \tag{2.2}$$

with the corresponding dual frame,

$$\omega^1 = \cos\theta\,\mathrm{d}x + \sin\theta\mathrm{d}y,\;\; \omega^2 = -\sin\theta\,\mathrm{d}x + \cos\theta\,\mathrm{d}y,\;\; \omega^3 = \mathrm{d}\theta \tag{2.3}$$

that is defined by $\langle \omega^i, \mathcal{A}_j \rangle = \delta^i_j$.

2.5 Operator-Valued Fourier Transform

In order to formally introduce the Fourier transform on $SE(2)$ as a special case of Fourier transforms on locally compact unimodular Lie groups in their full generality, one needs to be equipped with a solid background in representation theory and abstract harmonic analysis. In this section, we only present relevant definitions and lemmas leading up to the definition of the operator-valued Fourier transform and the Plancherel theorem which essentially states that under appropriate conditions, the operator-valued transform is a unitary map. This section follows [FF05, Chapter 3] and [Fol95, Chapter 7] closely.

The first ingredient we need is the notion of direct integrals. The direct integral of family Hilbert spaces essentially generalises the notion of direct sums.

Definition 2.5 (Measurable field of Hilbert spaces)**.** *Let* X *be a Borel measurable set and let* $(\mathcal{H}_x)_{x\in X}$ *be a family of separable Hilbert spaces then we call a section* $\eta : X \to \bigcup_{x\in X} \mathcal{H}_x$ *a vector field on* $(\mathcal{H}_x)_{x\in X}$. *We write* $(\eta_x)_{x\in X}$ *for vector field* η.

Now a measurable structure on $(\mathcal{H}_x)_{x\in X}$ *is a* $\mathbb{N}$*-indexed family of vector fields (in other words* $((e^n_x)_{x\in X})_{n\in\mathbb{N}}$ *on* $(\mathcal{H}_x)_{x\in X}$ *such that* $e^n_x \in \mathcal{H}_x$ *for all* $x \in X$ *and* $n \in \mathbb{N}$*) that satisfy*

- *For all* $m, n \in \mathbb{N}$, $x \mapsto \langle e^m_x, e^n_x \rangle$ *is Borel measurable.*

- *For each $x \in X$, the span of $(e_x^n)_{n\in\mathbb{N}}$ is dense in $\mathcal{H}_x$.*

When a measurable structure exists, we say that $(\mathcal{H}_x)_{x\in X}$ is Borel measurable. Now given $(\mathcal{H}_x)_{x\in X}$ that is measurable, a vector field $(\eta_x)_{x\in X}$ is measurable if $x \mapsto \langle \eta_x, e_x^n \rangle$ is a measurable for all $x \in X$ and $n \in \mathbb{N}$.

Now given a measurable structure on $(\mathcal{H}_x)_{x\in X}$, it is natural to speak of "measurable structure" on $(L^2(\mathcal{H}_x))_{x\in X}$ where $L^2(\mathcal{H}_x)$ denotes the space of Hilbert-Schmidt operators on $\mathcal{H}_x$. Namely, we have the following lemma:

Lemma 2.3. *A measurable structure on $(\mathcal{H}_x)_{x\in X}$ induces a system of vector fields $((e_x^n \otimes e_x^m)_{x\in X})_{m,n\in\mathbb{N}}$ on $(L^2(\mathcal{H}_x))_{x\in X}$ which is total in $L^2(\mathcal{H}_x)$ for each $x \in X$.*

Proof. To see that $(e_x^n \otimes e_x^m)_{n,m\in\mathbb{N}}$ is total in $L^2(\mathcal{H}_x)$ for each $x \in X$, we note that $L^2(\mathcal{H}_x) \subset K(\mathcal{H}_x)$, where $K(\mathcal{H}_x)$ is the space of compact operators. Now we have that $K(\mathcal{H}_x)$ is the norm closure of finite rank operators on $\mathcal{H}_x$ and the result immediately follows. □

Now given a measure ν on X, we immediately see a canonical way to define what it means for $(\mathcal{H}_x)_{x\in X}$ to be ν-measurable. Namely we say that $(\mathcal{H}_x)_{x\in X}$ is a ν-measurable field of Hilbert spaces whenever $(\mathcal{H}_x)_{x\in X}$ is measurable on some conull subset of X.

Now we fix a Borel measure ν and a ν-measurable field of Hilbert spaces $(\mathcal{H}_x)_{x\in X}$. We define the direct integral space $\mathcal{H}$ of $(\mathcal{H}_x)_{x\in X}$ to be the set of all measurable vector fields $(\eta_x)_{x\in X}$ on $(\mathcal{H}_x)_{x\in X}$ satisfying

$$\int_X ||\eta_x||^2 \, \mathrm{d}\nu(\mathrm{x}) < \infty$$

where ν almost everywhere agreeing vector fields are identified in the usual manner and we denote

$$\mathcal{H} = \int_X^{\oplus} \mathcal{H}_x \, \mathrm{d}\nu(\mathrm{x}).$$

Now we immediately see that $\mathcal{H}$ inherits a vector space structure from $\mathcal{H}_x$'s. Now, it can be shown that by defining

$$\langle \eta, \varphi \rangle = \int_X \langle \eta_x, \varphi_x \rangle \, \mathrm{d}\nu(\mathrm{x}),$$

$\mathcal{H}$ becomes a Hilbert space. When $X = \mathbb{N}$ with the discrete topology and with counting measure the direct integral becomes

$$\int_{\mathbb{N}}^{\oplus} \mathcal{H}_k \, d\nu(k) \cong \bigoplus_{k\in\mathbb{N}} \mathcal{H}_k.$$

If $\mathcal{H}_x = \mathbb{C}$ for all $x \in X$, then, we have

$$\int_X^{\oplus} \mathcal{H}_x \, \mathrm{d}\nu(\mathrm{x}) \cong \mathrm{L}^2(\mathrm{X}, \nu; \mathbb{C}).$$

Definition 2.6. *Given $(\mathcal{H}_x)_{x\in X}$ that is measurable and a family of bounded operators $(T_x)_{x\in X}$ where $T_x : \mathcal{H}_x \to \mathcal{H}_x$, we say that $(T_x)_{x\in X}$ is a measurable operator field if for any measurable vector field η and φ on $(\mathcal{H}_x)_{x\in X}$, the map $x \mapsto \langle \eta_x, T_x\varphi_x \rangle$ is Borel measurable.*

When the operator norms of T_x's are essentially bounded, we can define the direct integral of measurable operator fields $(T_x)_{x\in X}$ via

$$\left(\int_X^{\oplus} T_x \, \mathrm{d}\nu(\mathrm{x}) \right) (\eta) = (T_x \eta_x)_{x\in X}$$

which is a bounded operator on $\mathcal{H} = \int_X^{\oplus} \mathcal{H}_x \, \mathrm{d}\nu(\mathrm{x})$.

Now we can extend this analogously to fields of measurable representations of a group G. Namely we call the family $(\sigma_x)_{x\in X}$ as= measurable field of representations if $(\sigma_x(g))_{x\in X}$ is a measurable field of operators for each $g \in G$. Now the direct integral of $(\sigma_x)_{x\in X}$ is defined by

$$\left(\int_X^{\oplus} \sigma_x \, \mathrm{d}\nu(\mathrm{x})\right)(g) = \int_X^{\oplus} \sigma_x(g) \, \mathrm{d}\nu(\mathrm{x})$$

which is a well-defined unitary representation of G, and unique up to unitary equivalence.

In order to integrate fields of representations indexed by the dual group $\widehat{G}$, we of course need to deal with measure theory on $\widehat{G}$. We let for each positive integer n $\mathcal{H}_n$ to be a Hilbert space of dimension n, and $\mathcal{H}_\infty$ to be an infinite dimensional separable Hilbert space. For each $n = 1, 2, \ldots \infty$ let $\mathrm{Irr}_n(G)$ denote the set of irreducible representation of G on $\mathcal{H}_n$ and let $\mathrm{Irr}(G) := \bigcup_n \mathrm{Irr}_n(G)$. For each n, we define Σ_n to be the a σ-algebra generated by the family of functions $(\mathrm{Irr}_n(G) \to \mathbb{C},\ \pi \mapsto \langle \pi(g)u, v\rangle)_{g\in G, u,v\in\mathcal{H}_n}$. Now, we define the σ-alegbra Σ on $\mathrm{Irr}(G)$ to be the σ-algebra generated by all the inclusions $\mathrm{Irr}_n(G) \hookrightarrow \mathrm{Irr}(G)$. Now the canonical projection $\mathrm{Irr}(G) \to \widehat{G}$, $\pi \mapsto [\pi]$, is a surjection, and hence we can define a quotient σ-algebra structure on $\widehat{G}$ induced by Σ. We call the resulting σ-algebra the Mackey Borel structure on $\widehat{G}$.

For unitary representation π and by $\mathrm{Hom}(\pi, \pi)$ we denote the set of all intertwining operators for π. If the centre of $\mathrm{Hom}(\pi, \pi)$ is $\mathbb{C} \cdot \mathrm{Id}$, then we say that π is primary. The group G is said to be type I if every primary representation of G is a direct sum of copies of some irreducible representation.

Definition 2.7. *Let G be a second countable locally compact type I Lie group. Let λ_G denote the left regular representation. A measure ν_G on $\widehat{G}$ is called Plancherel measure if there the decomposition*

$$\lambda_G \simeq \int_{\widehat{G}}^{\oplus} m(\sigma) \cdot \sigma \, \mathrm{d}\nu_{\mathrm{G}}(\sigma)$$

exists for some $m : \widehat{G} \to \mathbb{N}_0 \cup \{\infty\}$.

Note that this direct integral is the direct integral of the field of representations $(m(\sigma) \cdot \sigma)_{\sigma\in\widehat{G}}$ over the dual group $\widehat{G}$.

Theorem 2.1. *For a type I Lie group and a representation π on G, there exists dual measure ν_G on $\widehat{G}$ and a map $m_\pi : \widehat{G} \to \mathbb{N}_0 \cup \{\infty\}$ such that we have a decomposition*

$$\pi \simeq \int_{\widehat{G}}^{\oplus} m_\pi(\sigma) \cdot \sigma \, \mathrm{d}\nu_{\mathrm{G}}(\sigma).$$

For the detail confer [FF05, Theorem 3.24.]. From this, we immediately see that for type I Lie groups, a Plancherel measure always exists.

The Operator-valued Fourier Transform:

The operator valued Fourier transform of an element $f \in L_1(G)$ is the family $(\hat{f}(\sigma))_{\sigma\in\hat{G}}$ where each $\sigma(f)$ is given by

$$\hat{f}(\sigma) = \int_G f(g) \, \sigma_{g^{-1}} \, \mathrm{d}g$$

defined point-wise by

$$\hat{f}(\sigma)(\eta) = \int_G f(g) \, \sigma_{g^{-1}}(\eta) \, \mathrm{d}g.$$

for $\eta \in \mathcal{H}_\sigma$. Here the integral of $f(g)\sigma_g(\phi)$ is defined to be the unique element such that

$$\langle \hat{f}(\sigma)(\eta), \varphi \rangle = \int_G f(g) \langle \sigma_g^{-1}(\eta), \varphi \rangle \, \mathrm{d}g.$$

where $\varphi \in \mathcal{H}_\sigma$. We are now ready to introduce the main result known as the Plancherel theorem [FF05, Theorem 3.31.] for locally compact unimodular groups of type I.

Theorem 2.2. *Let G be a second countable unimodular, type I Lie group. Then $\hat{f}(\sigma) \in L_2(\mathcal{H}_\sigma)$ and there is a Plancherel measure ν_G on $\hat{G}$ such that the following holds:*

1. *The Fourier transform maps $L_1(G) \cap L_2(G)$ into the direct integral space $\int_{\widehat{G}}^{\oplus} L_2(\mathcal{H}_\sigma) \, \mathrm{d}\nu_G(\sigma)$ and extends to a map*
$$\mathcal{F} : L^2(G) \to L^2(\widehat{G}) := \int_{\widehat{G}}^{\oplus} L_2(\mathcal{H}_\sigma) \, \mathrm{d}\nu_G(\sigma)$$
which is unitary.

2. *The inverse of $\mathcal{F}$ is given by the inversion formula*
$$f(x) = \int_{\widehat{G}} \mathit{trace}(\hat{f}(\sigma)\sigma(x)^*) \, \mathrm{d}\nu_G(\sigma)$$
for all $f \in \mathit{span}\{g * h : g, h \in L^1(G) \cap L^2(G)\}$.

2.6 Scale Space Theory

In mathematical image analysis, a (grey-scale) *image* is a bounded function $f \colon \mathbb{R}^2 \to \mathbb{R}$. In our case, since we perform image analysis on lifted images, we generalise this by setting $f \colon G \to \mathbb{R}$ where G is the homogeneous space the images are lifted to. In addition, we assume that G is equipped with a left-invariant Haar measure in order to make sense of integration on the space. In this context, we may view the processing of an image at some scale $t \geq 0$ as something that inputs the original image f and outputs a new image say f_t. To be more mathematically precise, A *scale-space representation* of an image f is a map

$$\Phi f \colon G \times [0, \infty) \to G \;\; \text{s.t} \;\; \Phi_0 f := \Phi f|_{G \times \{0\}} = f$$

where the scale parameter $t \in [0, \infty)$ in the second argument determines the amount of smoothing or image simplification. Note that the condition $\Phi f|_{G \times \{0\}} = f$ says that initially (at $t = 0$), Φf is the original image f. In this framework, we can think of a certain image processing as an indexed family of maps Φ_t from the space of images to itself. In general, the space of all possible scale-spaces is far too large and therefore, one of the main goals of mathematical image analysis is to find conditions that impose reasonable restrictions on the allowable set of scale-spaces. While these conditions vary depending on the purpose and the application, a consensus has been established on certain properties that most interesting examples of scale-space should have. One of the main properties is the 'evolution' aspect of scale space. Namely, as the time scale t increases, the image also gets increasingly processed and the processing is the same at every time. This is commonly formulated in terms of the semi-group property of scale space

$$\Phi_t \circ \Phi_s = \Phi_{t+s}$$

for all $t, s \geq 0$. While this condition greatly reduces the set of allowable scale-spaces, it is still not entirely clear what sort of operation is performed on an image. One of the main applications of mathematical image analysis lies in convolutional neural networks. A convolutional neural network uses three basic image processing operations and they are regularization and min/max pooling. We can model these operations with their smooth analogues which are the image-processing operations of diffusion and dilation/erosion respectively. Now, these four operations are of main interest in the field of mathematical image analysis for neural networks. Diffusion constitutes so-called linear scale space. Linear scale spaces are a class of scale spaces in which the image processing operation is given by an integral operator with a given kernel. More precisely, we have that $\Phi_t f$ is given by

$$(\Phi_t(f))(x) = \int_G k_t(x,y)f(y)\ \mathrm{d}y$$

for some kernel k_t for each $t \geq 0$. Note that this implies that Φ_t is fully determined by the choice of kernel k_t. Now there are a number of desired additional architectural properties of linear scale-spaces that further narrow the set of allowed kernels that make up the scale space. We take a look at the case for $G = \mathbb{R}^2$ which is well-studied. For example, it can easily be verified that shift invariance forces the integral operator to be a convolution. Pauwels [Pau+95] showed that for linear, isotropic, shift-invariant convolutional scale-spaces, then two conditions namely scale equivariance and semi-group property forces a one-parameter family of kernels which are determined fully by the parameter only. These scale-spaces are called α scale-spaces. The well-known Gaussian kernel is precisely one of the kernels that belong in this family where the parameter $\alpha = 1$. Now, $\alpha = 1/2$, corresponds to the so-called Poisson scale space and it was shown by Duits [Dui+04] that the scale-space defined by this kernel is associated with a first-order pseudo partial differential equation equivalent to the Laplace equation on the upper half-plane. Namely, $\Phi_t f$ solves the PDE

$$\begin{cases} \partial_t u = -\sqrt{-\Delta} u \\ u(\cdot, 0) = f. \end{cases}$$

It is also well-known that the Gaussian scale space also solves an evolution PDE, namely the heat equation. Inspired by these examples, we can stipulate scale-spaces to be governed by PDEs. For differential operator

$$P(\nabla) = \sum_{|\alpha| \leq k} c_\alpha \nabla^\alpha,$$

where $\alpha \in \mathbb{N}_0^n$ multi-index, we call the class of shift-invariant scale-spaces that obey the evolution equation

$$\begin{cases} \partial_t u = P(\nabla) u \\ u(\cdot, 0) = f \end{cases} \tag{2.4}$$

linear shift-invariant scale-spaces(LSI scale-spaces). Erosion and dilation constitute so-called *morphological scale-spaces*. Similar to linear scale-spaces, there are a number of desirable architectural properties that morphological scale-spaces of our interest ought to satisfy and many of them are identical to those that we also wish to impose on linear scale-spaces. The main difference is that for morphological scale-spaces, the image processing is given by *morphological convolution* instead of linear integral convolution. The morphological convolution is divided into two and they are dilation operation and erosion operation. Dilation and erosion are defined by

$$(f \ \ominus\ g)(x) = \inf_{y \in \mathbb{R}^2} \{f(y) - g(x-y)\},$$

and

$$(f \ \oplus\ g)(x) = \sup_{y \in \mathbb{R}^2} \{f(y) - g(x-y)\}$$

for all $x \in \mathbb{R}^2$ respectively and the morphological analogue of the kernel is called the structuring function. Now as presented in [SW16], some morphological scale-spaces also have associated evolution equations. Namely, we consider Hamilton-Jacobi equations of the type

$$\begin{cases} \partial_t v = -H(\mathrm{d}v) \\ v(\cdot, 0) = f. \end{cases}$$

Then under certain assumptions (H is convex and coercive function and that f is bounded and lower semi-continuous), this PDE has a unique viscosity solution which defines a scale-space. Under these assumptions, the solution is given by the Lax-Oleinik formula

$$v(x,t) = (f \,\square\, s_t)(x), \qquad x \in \mathbb{R}^d, t \geq 0,$$

where the structuring function s_t is given by

$$s_t(x) = -\mathfrak{f}(tH)(x).$$

where $\mathfrak{f}$ denotes the Frenchel transform and $\square$ the morphological convolution. For more, see [Eva98] In the paper by a connection between LSI scale-spaces and morphological scale-spaces. The paper introduces a new transformation called Cramer-Fourier transform

$$\mathcal{C}f := \mathfrak{f} \circ (-\log) \circ \mathcal{F}f$$

and shows that each LSI scale space corresponding to a PDE given by (2.4) with kernel k_t defines its morphological counterpart with the structuring function

$$s_t = \mathcal{C}k_t.$$

The striking result in the paper is that the morphological scale-space defined by the structuring function s_t, under the assumption of P being proper, lower semi-continuous, and convex, is the unique viscosity solution of the PDE

$$\begin{cases} \partial_t v = -P(\nabla v) \\ v(\cdot, 0) = f. \end{cases}$$

Chapter 3

Semiring Structures and Corresponding Scale Space Theory

In conventional mathematical image analysis, an image is a function with $\mathbb{R}$ as the codomain. Now the choice of $\mathbb{R}$ is due to the ordered-ness of $\mathbb{R}$ and the intuition on how "bright" a point on an image (represented by an element of $\mathbb{R}$) is unaffected by the underlying ring structure of $\mathbb{R}$. Taking this into account, we extend the codomain of our functions, by endowing $\mathbb{R}$ with a semiring structure where $R = (\mathbb{R}, \mathfrak{o}, \mathfrak{1}, \otimes, \oplus)$ where $\otimes$ is a commutative binary multiplication operator and $\oplus$ is a commutative additional operator, such that

- $a \otimes (b \oplus c) = a \otimes b \oplus a \otimes b$ for all $a, b, c \in R$
- $a \otimes b = b \otimes a$ and $a \oplus b = b \oplus a$ for all $a, b \in R$,
- $a \otimes \mathfrak{1} = a$ for all $a \in R$ and some $\mathfrak{1} \in R$,
- $a \oplus \mathfrak{o} = a$ for all $a \in R$ and some $\mathfrak{o} \in R$,
- $a \otimes a \geq \mathfrak{o}$,

Furthermore, we request a partial ordering $\leq$ on R that complies with the semiring structure such that

$$a \leq c \text{ and } b \leq d \Rightarrow a \oplus b \leq c \oplus d \text{ for all } a, b, c, d \in R.$$

Definition 3.1. *The main semiring examples of interest are:*

1. *the linear semiring $R_L = \mathbb{R}$ with*

$$a \oplus b = a + b, \quad a \otimes b = a\ b, \mathfrak{1} = 1, \mathfrak{o} = 0,$$

2. *the min-tropical semiring $T_{-} = \mathbb{R} \cup \{\infty\}$ with*

$$a \oplus b = \min\{a, b\}, \quad a \otimes b = a + b, \mathfrak{1} = 0, \mathfrak{o} = \infty,$$

3. *the max-tropical semiring $T_{+} \cup \{-\infty\}$ with*

$$a \oplus b = \max\{a, b\}, \quad a \otimes b = a + b, \mathfrak{1} = 0, \mathfrak{o} = -\infty,$$

In the theory of equivariant PDE-G-CNNs [Sme+20] they play an important role for respectively:

1. linear convection (for transport) and fractional diffusion modules (for regularisation) in PDE-G-CNNs.

$$\begin{aligned} &\partial_t W = -c \cdot \nabla W - (-\Delta)^{2\alpha} W \\ &W(\cdot, 0) = W_0 \end{aligned}$$

which are solved with linear convolution $*$ on G over the linear semiring with fractional heat-kernel K_t^α

$$\begin{aligned} W(g,t) &= \int\limits_{h \in G} K_t^\alpha(h^{-1}g) W_0(h)\, \mathrm{d}h \\ &=: (K_t^\alpha * W_0)(g). \end{aligned} \tag{3.1}$$

2. erosion modules (for sharpening, min-pooling over balls, and inhibition) in PDE-G-CNNs, steered by PDE:

$$\begin{aligned} &\partial_t W = -H(\mathrm{d}W) = -\tfrac{1}{2\alpha}\|\nabla W\|^{2\alpha} \\ &W(\cdot,0) = W_0 \end{aligned}$$

which are solved with morphological convolution $\square$ on G over the linear semiring with positive fractional heat-kernel k_t^α

$$\begin{aligned} W(g,t) &= \inf_{h \in G} k_t^\alpha(h^{-1}g) + W_0(h) \\ &=: (k_t^\alpha \square W_0)(g). \end{aligned} \tag{3.2}$$

Let ρ be a metric on the semiring R (in the do-domain of our data functions). In Appendix A.1 we set up some left-invariant measure theory (simple functions, measurability) to define and justify

$$I_G^R(\psi) := \bigoplus_{x \in G} \psi(x)$$

For details see Appendix A.1.

Special cases:

- $R = R_L$ then
$$I_G^{R_L}(\psi) = \int\limits_G \psi(g)\, \mathrm{d}\mu_G(g)$$
for Lebesque measurable functions $\psi : G \to \mathbb{R}$, where we use the usual left-invariant Haar measure μ_G on G (that is uniquely determined up to a constant).

- $R = T_-$ then
$$I_G^{T_-}(\psi) = \inf_{g \in G} \psi(g)$$
for ψ measurable, which in the tropical semiring T_- amounts to lower-semi continuous (l.s.c.), see Appendix A.1.

- $R = T_+$ then
$$I_G^{R_T^{max}}(\psi) = \sup_{g \in G} \psi(g)$$
for ψ measurable, which in the tropical semiring T_- amounts to upper-semi continuous (u.s.c.), see Appendix A.1.

Remark 3.1. *Here we adhere to terminilogy in previous works [Pai+23] and [KM97] and talk about 'semirings', where the wording 'semifield' is actually more appropriate as it involves a multiplication operator that is distributive of the addition operator, (and the addition need not admit an inverse, like in the tropical settings T_+ and T_- above).*

3.1 The function space $\mathcal{H}_R^G$ associated to Lie group G and semiring R

Let ρ be a distance on the semiring R (in the do-domain of our data functions). After the introduction of the integral and the distance ρ in the semiring R, we define a (pseudo)metric δ_R on R-valued functions on G using this metric the function space (a semi-module):

$$H_\rho = \{f : G \to R \mid \ f \in S_R \text{ and } \ \delta_R(f, \mathfrak{o}) < \infty\} \tag{3.3}$$

where S_R is the set of sum approachable functions from G to $\mathbb{R}$ (relative to the semiring R). These sum approachable functions are special cases of measurable functions, see Definition A.3 in Appendix A.2. In the linear semiring case, sum-approachable is the same as measurable. In the tropical semiring case T_+ sum-approachable functions are lower-semicontinuous, see Lemma A.3 and Lemma A.5.

We define the partition $H_R^G = H_\rho/\sim$ in H_ρ with the equivalence relation

$$f \sim g \Leftrightarrow \delta_R(f, g) = 0 \tag{3.4}$$

Then denote $\mathcal{H}_R^G$ as the closure/completion $\overline{H}_R^G$ of the space H_R^G with respect to δ_R.

Definition 3.2 (Semimodule)**.** *With semi-module V over a semiring R we mean a 'R-linear vector space'. I.e. V is a semi-module over semiring R if $\alpha \otimes f \oplus \beta \otimes g \in V$. for all $\alpha, \beta \in R$ and $f, g \in V$.*

Special cases:

1. **On the linear semiring** $R = R_L$ we set $\rho(a, b) = |a - b|^2$ by default and we get the square integrability constraint

$$\delta_L(f, 0) := \int_{x\in G} |f(x)|^2 dx \leq \infty \tag{3.5}$$

w.r.t. left-invariant Haar measure in (3.3) and equivalence relation (3.4) becomes the usual equivalence relation underlying the function classes in L^2-spaces and we obtain

$$\mathcal{H}_L^G = L^2(G),$$

2. **On the max-tropical semiring** R_+ we have $\mathfrak{o} = -\infty$ and set $\rho(a, b) = |e^a - e^b|$ from which we get the bounded from above constraint

$$\delta_{R_+}(f, -\infty) = \sup_{x\in G} e^{f(x)} \leq \infty. \tag{3.6}$$

In the tropical setting R_+ the equivalence relation (3.4) becomes trivial: two functions are equivalent if they take the same values on G. Thereby we obtain

$$\mathcal{H}_{R_+}^G = \overline{\{f : G \to R \,|\, f \text{ is bounded above \& l.s.c.}\}}. \tag{3.7}$$

By Lemma A.5 in Appendix A.1: R_+-integrable functions are upper semicontinuous.

3. **On the min-tropical semiring** R_-, we have $\mathfrak{o} = \infty$ and set $\rho(a, b) = |e^{-a} - e^{-b}|$ from which we get the bounded from below constraint:

$$\delta_{R_-}(f, \infty) = -\inf_{x\in G} -e^{-f(x)} = \sup_{x\in G} e^{-f(x)} \leq \infty \tag{3.8}$$

In the tropical setting R_- the equivalence relation (3.4) becomes trivial: two functions are equivalent if they take the same values on G. Thereby we obtain:

$$\mathcal{H}^G_{R_-} = \overline{\{f : G \to R | \, f \text{ is bounded below \& u.s.c.}\}}. \tag{3.9}$$

Lemma 3.1 gives $L^\infty(G) \cap C(G) = \mathcal{H}^G_{R_+} \cap \mathcal{H}^G_{R_-}$. Regarding the distance δ_R on the semi-modules (*not linear vector spaces in the tropical setting!*) one has

$$\begin{aligned} 0 \leq \psi \in C(G) &\Rightarrow \delta_{R_+}(\psi, \mathfrak{o}) = e^{2\|\psi\|_{L^\infty(G)}}, \\ 0 \geq \psi \in C(G) &\Rightarrow \delta_{R_-}(\psi, \mathfrak{o}) = e^{2\|\psi\|_{L^\infty(G)}}, \end{aligned} \tag{3.10}$$

Lemma 3.1. *The following spaces are equal*

$$L^\infty(G) \cap C(G) = \mathcal{H}^G_{R_+} \cap \mathcal{H}^G_{R_-} = \mathcal{H}^G_{R_+} \cap (-\mathcal{H}^G_{R_+}).$$

Proof. A function $f : G \to R$ is continuous if and only if it is both l.s.c. and u.s.c. and for $f \in C(G)$ the essential supremum equals the normal supremum. Furthermore, we have

$$\begin{gathered} f \text{ is bounded from below \& u.s.c.} \\ \Leftrightarrow \\ -f \text{ is bounded from above \& l.s.c.} \end{gathered}$$

from which the result follows. □

3.2 Convolutions and Fourier transforms associated with Lie group G and semiring R

From now on we assume $G \in \{\mathbb{R}^2, SE(2)\}$ and $R \in \{R_L, T_-, T_+\}$.

Definition 3.3 (General Quasi-Linear Group)**.** *We define $GQL(V)$ as the group of invertible R-linear maps on a (possibly infinite dimensional) R-semimodule V.*

In the same spirit, we generalise the notion of representation.

Definition 3.4 (Semiring Representation)**.** *A semiring R representation $\sigma : G \to GQL(V)$ of a group G is a group homomorphism from G to the group $GQL(V)$, where V is a (possibly infinite dimensional) R-semimodule, and is endowed with a metric.*

Furthermore we require them to be irreducible, meaning that there exist no invariant closed sub-semi-modules in the space $\mathcal{H}^G_R$.

Definition 3.5 ((Ir)reducible Representation)**.**
A representation $\sigma : G \to GQL(V)$ is called reducible if there exists a non-trivial sub-semimodule $\{\mathfrak{o}\} \neq W \subsetneq V$ of V that is closed under σ:

$$\sigma(g)W \subseteq W, \quad \text{for all } g \in G \tag{3.11}$$

If such a closed non-trivial sub-semimodule does not exist the representation is called irreducible.

In our generalisation, one issue arises: the notion of unitarity needs to be relaxed. In the linear semiring setting we have $\mathcal{H}_L^G = L^2(G)$. In the tropical semiring setting, however, we have a relation to $L^\infty(G)$ via (3.10), and we are not at all in a Hilbert space setting, so we cannot speak of adjoint or unitary operators, and just restrict to *quasi-isometry* (i.e. isometry on $\mathcal{H}_R$). Inspired by the book of Kolokoltsov & Maslov [KM97] we define the Semiring R group G Fourier transform $\mathcal{F}_R^G$ as follows.

Definition 3.6 (Dual Group)**.** *The dual group $\hat{G}$ is the group of all quasi-linear, quasi-isometric, irreducible group representations $\sigma_R^{\hat{g}} : G \to GQL(V)$ indexed by $\hat{g}$.*

Definition 3.7 (Semiring Fourier Transform)**.**

$$(\mathcal{F}_R^G f)(\hat{g}) = \bigoplus_{g \in G} f(g) \otimes (\sigma_R^{\hat{g}})_{g^{-1}} \tag{3.12}$$

where $\hat{g} \in \hat{G}$ an element of the dual group.

We also introduce semiring convolutions.

Definition 3.8 (Semiring Group Convolution)**.**

$$(K \circledast f)(g) = \bigoplus_{h \in G} K(h^{-1}g) \otimes f(h)\ . \tag{3.13}$$

where we assume $K, f \in \mathcal{H}_R^G$ are chosen such that $K \circledast f \in \mathcal{H}_R^G$. The latter is for example the case if $I_R(K) < \infty$.

In the special case L such a convolution is an ordinary convolution $*$ as seen in (3.1). In the special case R_- is an infimal convolution $\square$ as seen in (3.2).

Lemma 3.2 (Integral Preservation)**.** *Note that if a kernel K is normalized w.r.t the semiring integration I_R:*

$$I_R(K) = \mathbb{1} \tag{3.14}$$

Then the semiring convolution is integral preserving in the sense that:

$$I_R(f) = I_R(K \circledast f). \tag{3.15}$$

In the linear case R_l this corresponds to mass preservation of convolution with normalized kernels. In the tropical cases this corresponds to infimum and supremum preservation of infimal/supremal convolution using normalized kernels.

semiring Fourier transforms relate to semiring convolutions, the way one expects:

Lemma 3.3 (Convolution Theorem)**.** *For all $K, f \in \mathcal{H}_R^G$ such that $K \circledast f \in \mathcal{H}_R^G$ we have*

$$\mathcal{F}_R(K \circledast f)(p) = \mathcal{F}_R(K)(p) \circ \mathcal{F}_R(f)(p) \tag{3.16}$$

Proof. By direct computation we have (using definition of convolution, associativity, representation property, switching integration order, integral invariance after substitution)

$$\begin{aligned}
\mathcal{F}_R(h_1 *_R^G h_2)(p) &= \bigoplus_{g\in G} \sigma_R^p(g^{-1}) \otimes (h_1 *_R h_2)(g) \\
&= \bigoplus_{g\in G} \sigma_R^p(g^{-1}) \otimes \bigoplus_{h\in G} h_1(h^{-1}g) \otimes h_2(h) \\
&= \bigoplus_{g\in G} \bigoplus_{h\in G} \sigma_R^p(g^{-1}) \otimes h_1(h^{-1}g) \otimes h_2(h) \\
&= \bigoplus_{g\in G} \bigoplus_{h\in G} \sigma_R^p((h^{-1}g)^{-1}) \circ \sigma_R^p(h^{-1}) \otimes\; h_1(h^{-1}g) \otimes h_2(h) \\
&= \bigoplus_{h\in G} \left(\bigoplus_{v\in G} h_1(v) \otimes \sigma_R^p(v^{-1}) \right) \circ h_2(h) \otimes \sigma_R^p(h^{-1}) \\
&= \left(\bigoplus_{v\in G} h_1(v) \otimes \sigma_R^p(v^{-1}) \right) \circ \left(\bigoplus_{h\in G} h_1(h) \otimes \sigma_R^p(h^{-1}) \right) \\
&= \mathcal{F}_R h_1(p) \circ \mathcal{F}_R h_2(p) \ ,
\end{aligned}$$

where we set $h_1 = K$ and $h_2 = f$. □

In the linear semiring setting (i.e. $R = R_L$) as we have seen in Section 2.5, the Fourier transform $\mathcal{F}_G^R$ is invertible for type-1 Lie groups such as $G = \mathbb{R}^d \rtimes SO(J)$ including our primary case of interest $G = SE(2) = \mathbb{R}^2 \rtimes SO(2)$. The inversion formula is given by (5.21) and it comes from a Hilbert space structure and Schur's lemma where $\mathcal{F}_{R_L}^* \mathcal{F}_{R_L}$ commutes with every UIR being a multiple of the identity and indeed

$$\|f\|^2_{L^2(G)} = \|\hat{f}\|^2_{L^2(\hat{G})} = \int_{\hat{G}} \mathrm{trace}\{\hat{f}(p)^* \hat{f}(p)\} \mathrm{d}\nu_G(p).$$

where $\hat{f} = \mathcal{F}_{R_L} f$ and ν_G is such that $\mathcal{F}_{R_L}^* \mathcal{F}_{R_L} f = f$.

In the tropical semiring setting (i.e. $R = R_L$) such inversions have to be adopted. Even in the case $G = \mathbb{R}^2$ the tropical semiring Fourier transform is minus the Fenchel transform, and for invertibility (where the Fenchel transform its own inverse) one *must constrain the Fourier transform* in $\mathcal{H}_R$ to the semi-module of lower-semi continuous and convex functions, as we will see in Section 4.2.

In the general semiring setting it is not easy to analyze to what semi-module in $\mathcal{H}_R^G$ one must constrain to ensure invertibility. Now in our generalisation to other semirings R (like the tropical semirings) we need quasi-linear group representations.

Chapter 4
Axiomatic PDE-CNN on $\mathbb{R}^2$

4.1 Axioms for Geometric Deep Learning PDEs on $\mathbb{R}^2$

The evolution that determines one connection between two nodes in a PDE-G-CNN network on $G = \mathbb{R}^2$ consists of modules each acting over a different semiring structure R imposed on the set $\mathbb{R}$. In this section we constrain ourselves to $\mathbb{R}^2$ as the generalisation to $\mathbb{R}^d$, $d \geq 2$ is straightforward. We will impose a list of 'axioms' for the evolutions:

$$\begin{cases} \Phi_t : \mathcal{H}_R^G \to \mathcal{H}_R^G. \\ \Phi_0 = \mathrm{id}_{\mathcal{H}_R^G}, \end{cases}$$

where $t \geq 0$.

Before, we continue to list the axioms, let us mention that our evolutions will be assumed to be parameterised by one equivariant metric tensor field $\mathcal{G}$ on $\mathbb{R}^2$ or a transport vector $\boldsymbol{c}$. By equivariance such metric tensor field is parameterised by a constant 2×2 matrix $[g_{ij}]$ and constant vector (c^1, c^2) relative to the left-invariant vector fields and co-vector fields which in case of Lie group $(\mathbb{R}^2, +)$ are just given by $\{\partial_x, \partial_y\}$ and $\{\mathrm{d}x, \mathrm{d}y\}$. The parameters $[g_{ij}]$ and (c^1, c^2) are the trainable parameters in a PDE-G-CNN on $\mathbb{R}^2$. For vessel classification experiments with such PDE-G-CNNs see [Cas21].

In the axioms we will use on (right-)shift translation operators given by

$$\mathcal{R}_{\boldsymbol{c}} f(\boldsymbol{x}) := f(\boldsymbol{x} + \boldsymbol{c}), f \in \mathcal{H}_R^G, \boldsymbol{c}, \boldsymbol{x} \in \mathbb{R}^2,$$

for localisation in scale equivariance and rotation equivariance.

We impose the following requirements for our PDE-G-CNNs on $G = \mathbb{R}^2$:

1. **Semigroup property:**
   For all $s, t \geq 0$ we require
   $$\Phi_t \circ \Phi_s = \Phi_{t+s}.$$

2. **Quasilinearity and Kernel operator:**
   For all $t \geq 0$ we must have for all $\alpha_1, \alpha_2 \in R$ constant and all $f_1, f_2 \in \mathcal{H}_R^G$ that quasi-linearity holds
   $$\Phi_t(\alpha_1 \otimes f_1 \oplus \alpha_2 \otimes f_2) = \alpha_1 \otimes \Phi_t(f_1) \oplus \alpha_2 \otimes \Phi_t(f_2).$$
   More specifically, we will assume that the evolution operator Φ_t can for each $t > 0$ be written as
   $$\Phi_t f(\boldsymbol{x}) = \bigoplus_{\mathbf{y} \in \mathbb{R}^2} k_t(\boldsymbol{x}, \mathbf{y}) \otimes f(\mathbf{y})$$
   for every $\boldsymbol{x} \in \mathbb{R}^2$ for some kernel $k_t : \mathbb{R}^2 \times \mathbb{R}^2 \to R$ such that $k_t(\cdot, \mathbf{y}) \in \mathcal{H}_R^G$ for every $\mathbf{y} \in \mathbb{R}^2$.

3. **Equivariance:**
   For all $\boldsymbol{x} \in \mathbb{R}^2$ and all $t \geq 0$ we require
   $$\Phi_t \circ \mathcal{L}_{\boldsymbol{x}} = \mathcal{L}_{\boldsymbol{x}} \circ \Phi_t,$$

with shift-operator given by (left)-action $(\mathcal{L}_{\boldsymbol{x}} f)(\mathbf{y}) = f(-\boldsymbol{x} + \mathbf{y})$, for all $\boldsymbol{x}, \mathbf{y} \in \mathbb{R}^2$ and all $f \in \mathcal{H}_R^G$.

4. **Localised Time-scale equivariance:**
   For all $t > 0$ there exists a $\psi(t)$ and $\boldsymbol{c}(t)$ such that

$$\mathcal{S}_{\psi(t)} \circ \Phi_1 \circ \mathcal{R}_{\boldsymbol{c}(1)} \circ \mathcal{S}^{-1}_{\psi(t)} = \Phi_t \circ \mathcal{R}_{\boldsymbol{c}(t)} \tag{4.1}$$

   in such a way $t \mapsto \boldsymbol{c}(t)$ is continuous and such that Ψ is continuously differentiable and monotonic, where $\mathcal{S} : \mathcal{H}_R^G \to \mathcal{H}_R^G$ is the (total integral I_R preserving) scaling operator given by

$$(\mathcal{S}_a f)(\boldsymbol{x}) = C_{a,R}\; f(\frac{\boldsymbol{x}}{a}). \tag{4.2}$$

   for all $a > 0$. If $R = R_L$ then $C_{a,R} = \frac{1}{a^2}$. If $R = T_-$ then $C_{a,R} = 1$.

5. **Localised Isotropy w.r.t. trained (left)-invariant metric tensor field $\mathcal{G}$:**
   For all $t \geq 0$ there exists a $\boldsymbol{c}(t) \in \mathbb{R}^2$ s.t. and for all $Q \in SO_{[\mathcal{G}]}$:

$$(\Phi_t \circ \mathcal{R}_{\boldsymbol{c}(t)}) \circ \mathcal{U}_Q = \mathcal{U}_Q \circ (\Phi_t \circ \mathcal{R}_{\boldsymbol{c}(t)})$$

   with $\mathcal{R}_{\boldsymbol{c}} f(\boldsymbol{x}) = f(\boldsymbol{x} + \boldsymbol{c})$, and with '$\mathcal{G}$-rotations' given by $\mathcal{U}_Q f(\boldsymbol{x}) := f(Q^{-1}\boldsymbol{x})$, for all $\boldsymbol{x} \in \mathbb{R}^2$ where

$$SO_{[\mathcal{G}]} := \{Q \in \mathbb{R}^{2\times 2} \mid \det(Q) = 1, Q^T[\mathcal{G}]Q = [\mathcal{G}]\}$$

   and where we assume $t \mapsto \boldsymbol{c}(t)$ is continuous.

6. **Optional Axiom: Isometry Constraint:**

$$\forall_{f \in \mathcal{H}_G^R, f \geq 0} \forall_{t \geq 0} \;:\; \delta_R(\epsilon_R f, \mathfrak{o}) = \delta_R(\epsilon_R \Phi_t(f), \mathfrak{o}),$$

   with $\epsilon_R = \mathrm{sign}(\chi'_R) \in \{-1, 1\}$.

7. **Optional Axiom: Positivity:**
   We require the kernel k_t, see Axiom 2, to be positive.

Now let us return to our first case of interest, namely $G = \mathbb{R}^2$, write $p = \omega \in \hat{\mathbb{R}}^2 \equiv \mathbb{R}^2$ and study our semirings of primary interest R_L, R_- and R_+.

Lemma 4.1 (Irreducible Semiring Representations)**.** *Linear irreducible unitary representations of $\mathbb{R}^2$ are defined on 1D vector spaces over $\mathbb{C}$. Likewise, quasi-linear irreducible representations of $\mathbb{R}^2$, with an eigenvalue[a], are defined on 1D semi-modules.*

[a] This assumption simplifies the proof but we would conjecture it is like the R_L-case also redundant for tropical semiring cases.

Proof. $G = \mathbb{R}^2$ is commutative, we can apply the standard Schur's lemma (over the algebraically complete field $\mathbb{C}$) and standard reasoning such as in Folland's book [Fol73, cor.3.6] in the linear semiring setting.
Now let us consider the tropical case, where we apply an extra assumption of having an eigenvalue. Let $R : G \mapsto C(\mathcal{X})$ be a quasi-linear group representation on a semi-module $\mathcal{X}$ that is irreducible (there no closed

semi-modules other than the full space and the origin $\equiv \mathbf{o}$). Then $R_{\boldsymbol{x}} f = \lambda(\boldsymbol{x}) \otimes f$ for some specific $f \neq 0$. Then for all $\mathbf{y} \in \mathbb{R}^2$ operator $R_{\mathbf{y}}$ maps the (closed and non-trivial) eigenspace $E_\lambda(R_{\boldsymbol{x}})$ onto itself as

$$\mathcal{R}_{\boldsymbol{x}} \circ \mathcal{R}_{\mathbf{y}} f = \mathcal{R}_{\mathbf{y}} \circ \mathcal{R}_{\boldsymbol{x}} f = \lambda(\boldsymbol{x}) \otimes \mathcal{R}_{\mathbf{y}} f.$$

By the irreducibility assumption of R this eigenspace is the full space from which we deduce that $\mathcal{R}_{\boldsymbol{x}} = \lambda(\boldsymbol{x}) \otimes \mathrm{id}_{\mathcal{X}}$. $\square$

Corollary 4.1. *When identifying the quasi-linear operators $(\omega_R)^\omega_{\boldsymbol{x}}$ acting on 1D semi-modules with their semiring-multipliers in the semiring Fourier transform $\mathcal{F}^R_{G=\mathbb{R}^2}$ given by (5.20) one can write for the* <u>*commutative*</u> *group $G = \mathbb{R}^2$:*

$$\mathcal{F}_R(h_1 \circledast h_2)(p) = \mathcal{F}_R(h_1)(p) \otimes \mathcal{F}_R(h_2)(p) \tag{4.3}$$

Remark 4.1. *In case the Lie group G is not commutative one must work with (3.16) rather than (4.3).*

Lemma 4.2 (Cauchy's Additive Functional Equation)**.** *Let $l : \mathbb{R} \to \mathbb{R}^n$ be continuous with $l(x) + l(y) = l(x + y)$ for all $x, y \in \mathbb{R}$ then $l(x) = \boldsymbol{c} \cdot x$ for some $\boldsymbol{c} \in \mathbb{R}^n$.*

Proof. This is a standard result in real analysis. We include the proof for completeness. We show that additivity implies linearity. We have that for an arbitrary $x \in \mathbb{R}$, we can approximate by a sequence of rational numbers $p_n/q_n \to x$. Then we have by the additivity that for every $n \in \mathbb{N}$, $l(p_n/q_n) = p_n \cdot l(1/q_n)$. Now $q_n \cdot l(1/q_n) = l(q_n/q_n) = l(1)$, so we get $l(1/q_n) = l(1)/q_n$. From this, we have $l(p_n/q_n) = p_n/q_n \cdot l(1)$. Now continuity of l implies that $l(x) = \lim_{n\to\infty} p_n/q_n \cdot l(1) = x \cdot l(1)$. $\square$

Corollary 4.2 (Cauchy's Multiplicative Functional Equation)**.** *Let $g : \mathbb{R}^+ \to \mathbb{R}^+$ be continuous with $g(s)g(\rho) = T\, g(\rho \cdot s)$ for all $\rho, s > 0$ then $g(\rho) = T\, \rho^\alpha$ for some $\alpha > 0$*

Proof. We first take $T = 1$. By taking logarithms we have

$$\begin{aligned} \log g(e^{\log(\rho \cdot s)}) &= \log g(e^{\log(\rho)+\log(s)}) \\ &= \log(g(e^{\log(\rho)})) + \log(g(e^{\log s})) \end{aligned}$$

so set $x = \log \rho$ and $y = \log s$ and apply the previous lemma with $l(\cdot) = \log g(e^{\cdot})$. Then $\log g(\rho) = \log g(e^x) = \alpha x \Rightarrow g(\rho) = \rho^\alpha > 0$ with $\alpha > 0$. The general case follows from $\rho \mapsto T\rho$. $\square$

Definition 4.1. *Let us write $* = *^{R_L}$ for linear convolution and $\square = *^{T_-}$ for morphological convolution.*

- For $R = R_L$ the 1D unitary irreducible representations of $\mathbb{R}^2$, indexed by ω, are given by

$$\sigma^\omega_{R_L} : \mathbb{R}^2 \to \mathbb{C}, \quad \left(\sigma^\omega_{R_L}\right)_{\boldsymbol{x}} = e^{i\omega \cdot \boldsymbol{x}}. \tag{4.4}$$

  Thereby the semiring Fourier transform is the ordinary unitary Fourier transform $\mathcal{F} : L^2(\mathbb{R}^2, \mathbb{C}) \to L^2(\mathbb{R}^2, \mathbb{C})$ and we have

$$\mathcal{F}(f_1 * f_2) = \mathcal{F} f_1 \cdot \mathcal{F} f_2$$

- For $R = R_T^{\min}$ the semiring Fourier transform is minus the Fenchel transform.

$$\mathfrak{f}(f_1 \Box f_2) = \mathfrak{f}(f_1) + \mathfrak{f}(f_2)$$

as the quasilinear, irreducible, quasi-isometric representations are given by (Recall Lemma 4.1)

$$\left(\sigma^{\omega}_{R_T^{min}}\right)_{\boldsymbol{x}} = (\omega \cdot \boldsymbol{x}) \tag{4.5}$$

and thereby

$$\mathcal{F}_{R_T^{min}} f(\omega) = \inf_{\boldsymbol{x} \in \mathbb{R}^2} \{f(\boldsymbol{x}) - \omega \cdot \boldsymbol{x}\} = -\mathfrak{f} f(\omega)$$

Remark 4.2. *Note that the fenchel transform of a function is always a l.s.c. function. A function f is lower semicontinuous if and only if $-f$ is upper semicontinuous. Thereby the output of $\mathcal{F}_{R_T^{min}}$ is always u.s.c..*

Remark 4.3. *The standard form (4.5) for quasilinear representations is a consequence of Lemma 4.1: σ^{ω}_R is a quasilinear representation of $\mathbb{R}^2$, with an eigenvalue, defined on a 1D semi-module $R = R_T^{min}$ (recall Lemma 4.1). Then $\sigma^{\omega}_R(x\mathbf{a}) \circ \sigma^{\omega}_R(y\mathbf{a}) = \sigma^{\omega}_R(x\mathbf{a}) \otimes_{R_T^{min}} \sigma^{\omega}_R(y\mathbf{a}) = \sigma^{\omega}_R((x+y)\mathbf{a})$ for all $\mathbf{a} \in S^1$ and $x, y \in \mathbb{R}$. By Lemma 4.2 $\sigma^{\omega}_R(\boldsymbol{x}) = \alpha[\omega] \cdot \boldsymbol{x}$ for some $\alpha[\omega] \in \mathbb{R}^2$. Now the dual group $\hat{\mathbb{R}}^2$ of $\mathbb{R}^2$ equals $\mathbb{R}^2$ and $\sigma^{\omega_1}_R \circ \sigma^{\omega_2}_R = \sigma^{\omega_1}_R \otimes \sigma^{\omega_2}_R = \sigma^{\omega_1+\omega_2}_R$. Thereby $\sigma^{\omega}_R(\boldsymbol{x}) \in R$ is linear in $\boldsymbol{x}$* and ω *and $\alpha[\omega] = \omega$ and $\sigma^{\omega}_R(\boldsymbol{x}) = \boldsymbol{x} \cdot \omega$.*

Remark 4.4. *The stand form (4.4) for UIR of $\mathbb{R}^2$ is well-known. Note $\sigma^{\omega}_{R_L}$ is a linear representation of $\mathbb{R}^2$ defined on a 1D semi-module $\equiv R_L + iR_L$. Then with similar arguments (again using Lemma 4.1 and Lemma 4.2) as in the above remark but now with $\sigma^{\omega}_{R_L} \circ \sigma^{\omega}_{R_L} = \sigma^{\omega}_{R_L} \otimes_{R_L} \sigma^{\omega}_{R_L}$ and one has $\sigma^{\omega}_{R_L}(\boldsymbol{x}) = 1 \cdot e^{i\lambda\omega\cdot x}$, for some $\lambda \in \mathbb{R}$.*

Note that in the above cases we see that the Fourier transform $\mathcal{F}_R$ commutes with rotations and intertwines integral preserving scaling.

Lemma 4.3. *Let $R \in \{R_L, R_T^{min}, R_T^{max}\}$. Let $\mathcal{F}_R$ be the Fourier transform on commutative group $\mathbb{R}^2$ given by (5.20). Then we have*

$$\begin{aligned} &\mathcal{F}_R \circ \mathcal{R}_Q = \mathcal{R}_Q \circ \mathcal{F}_R \text{ for all } Q \in SO(2) \\ &(\mathcal{F}_R \circ \mathcal{S}_a f)(\omega) = (\mathcal{F}_R f)(a\omega) \text{ for all } a > 0, \end{aligned}$$

where $\mathcal{R}_Q f(\boldsymbol{x}) = f(Q^{-1}\boldsymbol{x})$ and where the integral preserving scaling $\mathcal{S}_a$ is defined by (4.2).

Proof. In both semiring cases we have $\sigma^{\omega}_R(\boldsymbol{x}) = \sigma^{Q\omega}_R(Q\boldsymbol{x})$ for all $Q \in SO(2)$. As a result by (5.20) with $p = \omega$ we have

$$\mathcal{R}_Q \mathcal{F}_R f(\omega) = \mathcal{F}_R(Q^{-1}\omega) = \bigoplus_{\boldsymbol{x} \in \mathbb{R}^2} f(\boldsymbol{x}) \otimes \sigma^{Q^{-1}\omega}_{-\boldsymbol{x}} = \bigoplus_{\boldsymbol{x} \in \mathbb{R}^2} f(\boldsymbol{x}) \otimes \sigma^{\omega}_{-Q\boldsymbol{x}} = \bigoplus_{\boldsymbol{x} \in \mathbb{R}^2} f(Q^{-1}\boldsymbol{x}) \otimes \sigma^{\omega}_{\boldsymbol{x}} = \mathcal{F}_R \mathcal{R}_Q f(\omega)$$

for all $f \in \mathcal{H}_R$ and all $\sigma^{\omega} \in \hat{\mathbb{R}}^2$ and all $Q \in SO(2)$.

Regarding the integral preserving scaling we have

$$\mathcal{F}_R f(a\omega) = \bigoplus_{\boldsymbol{x} \in \mathbb{R}^2} f(\boldsymbol{x}) \otimes \sigma^{a^{-1}\omega}_{-\boldsymbol{x}} = \bigoplus_{\boldsymbol{x} \in \mathbb{R}^2} f(\boldsymbol{x}) \otimes \sigma^{\omega}_{-a\boldsymbol{x}} = C_{a,R} \bigoplus_{\boldsymbol{x} \in \mathbb{R}^2} f(a^{-1}\boldsymbol{x}) \otimes \sigma^{\omega}_{\boldsymbol{x}} = \mathcal{F}_R \mathcal{S}_a f(\omega).$$

Note that $\mathcal{F}_R f(\mathbf{0}) = I_R(f) = I_R(\mathcal{S}_a f)$, and the result follows. □

Lemma 4.4. *Let G be a unimodular Lie group. In particular for $G = \mathbb{R}^d \rtimes SO(J)$, and right action $\mathcal{R}_g : \mathcal{H}_G^R \to \mathcal{H}_G^R$. We have for $\mathcal{R}_q f(g) := f(gq)$ for all $q, g \in G$*

$$\mathcal{F}_R(\mathcal{R}_q f)(\omega) = (\sigma_R)_q^\omega \circ \mathcal{F}_R f(\omega)$$

and in the case that G is commutative, we get that

$$\mathcal{F}_R(\mathcal{R}_q f)(\omega) = (\sigma_R)_q^\omega \otimes \mathcal{F}_R f(\omega)$$

Proof. We directly compute: We get

$$\mathcal{F}_R(\mathcal{R}_q f)(\omega) = \bigoplus_{g\in G} \mathcal{R}_q f(g) \otimes (\sigma_R)_{g^{-1}}^\omega = \bigoplus_{h\in G} f(h) \otimes (\sigma_R)_{qh^{-1}}^\omega = \bigoplus_{h\in G} f(h) \otimes (\sigma_R)_q^\omega \circ (\sigma_R)_{h^{-1}}^\omega = (\sigma_R)_q^\omega \circ \mathcal{F}_R f(\omega),$$

where the last equality holds because of quasilinearity of σ_R. In the 2nd equality we apply $h = gq$ and in the linear semiring setting the Fourier transform uses a left-invariant measure for which we now applied a right-shift invariance, and in the unimodular Lie groups $G = \mathbb{R}^d \rtimes SO(J)$ where indeed elements $q = (\boldsymbol{x}, A)$ satisfy $\det(A) = +1 = \det(q)$ we indeed have $d\mu_G(g) = d\mu_G(hq) = d\mu_G(h) = 1 \cdot \mathrm{d}\boldsymbol{x}\mathrm{d}A.$. Now in the case that G is commutative, the result readily follows from Lemma 4.1. □

4.2 The Axiomatic Solutions to linear and morphological scale spaces in $\mathbb{R}^2$

Invertible semiring Fourier transform on $\mathbb{R}^2$

The linear semiring Fourier transform is invertible, in fact there one has a unitary operator $\mathcal{F}^{-1} = \mathcal{F} = \mathcal{F}_{R_L}$ from $L^2(\mathbb{R}^2) \to L^2(\mathbb{R}^2)$. In the tropical semiring one has that

$$(\mathfrak{f} \circ \mathfrak{f})f = f$$

iff f is lower-semi continuous and convex by Fenchel-Moreau's theorem. One may force injectivity on equivalence classes of the semi-module $\mathcal{H}_R$ by identifying u.s.c. functions:

$$f_1 \sim_R f_2 \Leftrightarrow \mathcal{F}_R f_1 = \mathcal{F}_R f_2.$$

For $R = R_L$ we have $f_1 \sim_{R_L} f_2 \Leftrightarrow f_1 = f_2 \in L^2(\mathbb{R}^2)$, but now let us see what happens in the tropical case.

The tropical case The Fenchel transform $\mathfrak{f} = -\mathcal{F}_{T_-}$ is its own inverse on the semi-module:

$$V := \{f \in \mathcal{H}_{R_{T_{min}}}^{\mathbb{R}^2} \mid f \text{ is convex}\}. \tag{4.6}$$

Regarding injectivity in the tropical sense we note

$$\mathcal{F}_{T_-} f_1 = \mathcal{F}_{T_-} f_2 \Leftrightarrow \mathfrak{f} f_1 = \mathfrak{f} f_2 = \mathfrak{f} f_1 \square \mathfrak{f}(f_2 - f_1)$$

using $\mathfrak{f}(a + b) = \mathfrak{f} a \,\square\, \mathfrak{f} b$ which is special case of Lemma 3.3.

The assumption of convexity in (4.6) is a restriction that can be dropped as we explain next. For a convenient construction of the Fenchel transform to a larger space we assume f to be absolute continuous rather than just lower-semicontinuous. I.e. we consider

$$\tilde{V} := \{f \in \mathcal{H}^{\mathbb{R}^2}_{R_{T_{min}}} \mid f \text{ is absolutely continuous}\}. \tag{4.7}$$

Now every function $f \in C^2(\mathbb{R}^2, R)$ is a difference $f = f^A - f^B$ of two C^2 convex functions f^A and f^B. If we assume f^A has the property that $f^A(\mathbf{0}) = 0$ and (generalised) derivative $Df^A(\mathbf{0}) = \mathbf{0}$ then the decomposition is unique.

This assumption avoids us to transfer linear terms (generalised eigenfunctions of $\mathfrak{f}$) from f^B to f^A and f^A to f^B. Splitting the positive eigenvalues of the Hessian Hf from the negative ones, provides the Hessians $Hf = Hf^A - Hf^B$.
In fact f^A and f^B can be explicitly obtained like this. Simply by integrating the positive (respectively negative) part of the Hessian and then do one more integration.

This explains why in $\tilde{V}$ we assume absolute continuity (rather than just lower-semicontinuity) since we need generalised derivatives in the explicit construction of f^A and f^B while imposing vanishing generalised derivative $Df^A(\mathbf{0}) = \mathbf{0}$ and vanishing function value $f^A(\mathbf{0}) = 0$. One may think of f^A as the 'convex part' of f and $-f^B$ as the 'concave part' of f, and we have

$$f_1 \sim_{T_-} f_2 \Leftrightarrow \mathfrak{f}f_1 = \mathfrak{f}f_2 \Leftrightarrow \mathfrak{f}^2 f_1^A = \mathfrak{f}^2 f_2^A \Leftrightarrow f_1^A = f_2^A$$

Remark 4.5. *(illustration 1D case)*
For absolutely continuous functions $f : \mathbb{R} \to \mathbb{R}$ we have

$$\begin{aligned} f(x) &= f^A(x) - f^B(x) = \int_0^x \left(\int_0^t (f''(q))_+ \mathrm{d}q \right) \mathrm{d}t \\ &- \left(-f(0) - xf'(0) - \int_0^x \left(\int_0^t (f''(q))_- \mathrm{d}q \right) \mathrm{d}t \right) \end{aligned}$$

with $b_+ = \max\{b, 0\}$ *and* $b_- = \min\{b, 0\}$.

On the relatively large space $\tilde{V}$ we can define the Fourier transform as follows

$$\mathcal{F}_{T_-}(f) = \mathcal{F}_{T_-}(f_A - f_B) = \mathcal{F}_{T_-}(f_A) - \mathcal{F}_{T_-}(f_B).$$

Time-Scale Equivariance on $\mathbb{R}^2$

We have formulate the time-scale equivariance, Axiom 4, for the case $G = \mathbb{R}^2$. We did this to be consistent with the axioms of scale space theory [Pau+95][Dui+04].

An equivalent coordinate formulation, as we will prove in Corollary 4.3, that better generalizes to other Lie groups is to index the scale space evolution $\Phi_t^{\mathcal{G}}$ with a left-invariant (LI) metric tensor field $\mathcal{G}$ and convection vector $\boldsymbol{c}$ (for centering) and to require

$$\exists_{\psi \in C^1(\mathbb{R}^+), \psi' > 0} \forall_{\mathcal{G}, \boldsymbol{c} \text{ LI}} \, \forall_{t \geq 0} \, : \, \Phi_t^{\mathcal{G}, \boldsymbol{c}} = \Phi_1^{|\psi(t)|^{-2}\mathcal{G}, t\boldsymbol{c}} \tag{4.8}$$

This replaces the scaling on the group $\mathbb{R}^2$ in (4.1) to a scaling of the left/shift invariant metric tensor field $\mathcal{G}$ in (4.8).

Remark 4.6. *For general Lie groups we will rather use the coordinate free symmetry (4.8) than (the possibly only locally defined) scaling in Riemannian normalised coordinates (4.1) as they break-down at the (left-invariant) cut-locus.*

Lemma 4.5. *Fix the semiring $R \in \{R_L, R_T^{min}, T_+\}$. In the setting $G = \mathbb{R}^2$, Axioms 1,2,3 together with the localised isotropy, Axiom 5, implies that the kernel $\hat{k}_t := \mathcal{F}_G^R k_t$ in the (semiring) Fourier domain is of the form*[a]

$$\hat{k}_t(\omega) = (\sigma_R)^{\omega}_{-t\boldsymbol{c}} \otimes g_t^R(||\omega||_{[\mathcal{G}]^{-T}}) \tag{4.9}$$

for every $t \geq 0$ for some $\boldsymbol{c} \in \mathbb{R}^2$.

[a] Here $\equiv$ represents equality of functions up to constant that depends on the choice of semiring R and $\det(Q)$, likewise (4.2). See also first paragraph of page 32 where we recognize that due to shift equivariance the kernel actually depends on 1 variable instead of 2.

Proof. We have that for all $t \geq 0$ there exists a $\boldsymbol{c}(t) \in \mathbb{R}^2$ s.t. for all $Q \in SO_{\mathcal{G}}$:

$$(\Phi_t \circ \mathcal{R}_{\boldsymbol{c}(t)}) \circ \mathcal{U}_Q = \mathcal{U}_Q \circ (\Phi_t \circ \mathcal{R}_{\boldsymbol{c}(t)})$$

with $\mathcal{R}_{\boldsymbol{c}} f(\boldsymbol{x}) = f(\boldsymbol{x} + \boldsymbol{c})$, and with G-rotations given by $\mathcal{U}_Q f(\boldsymbol{x}) := f(Q^{-1}\boldsymbol{x})$. Now, we have for any $f \in \mathcal{H}_R^G$, we have from Lemma 3.3, Lemma 4.3, Lemma 4.4 and Lemma 4.1:

$$\begin{aligned}
\mathcal{F}_R(\Phi_t \mathcal{R}_{\boldsymbol{c}(t)} \mathcal{U}_Q f) &= \mathcal{F}_R\left(k_t *_G^R \mathcal{R}_{\boldsymbol{c}(t)} \mathcal{U}_Q f\right) \\
&= \hat{k}_t \otimes \mathcal{F}_R\left(\mathcal{R}_{\boldsymbol{c}(t)} \mathcal{U}_Q f\right) \\
&= \hat{k}_t \otimes (\sigma_R)_{\boldsymbol{c}(t)} \otimes \mathcal{F}_R\left(\mathcal{U}_Q f\right) \\
&= \hat{k}_t \otimes (\sigma_R)_{\boldsymbol{c}(t)} \otimes \mathcal{U}_{Q^{-T}} \mathcal{F}_R\left(f\right)
\end{aligned} \tag{4.10}$$

as we have that $Q \in SO([\mathcal{G}])$ implies $\det(Q) = 1$. On the other hand, we have from respectively, Lemma 4.3, Corollary4.1 and Lemma 4.4 that

$$\begin{aligned}
\mathcal{F}_R(\mathcal{U}_Q \Phi_t \mathcal{R}_{\boldsymbol{c}(t)} f) &= \mathcal{U}_{Q^{-T}} \mathcal{F}_R(k_t *_G^R \mathcal{R}_{\boldsymbol{c}(t)} f) \\
&= \mathcal{U}_{Q^{-T}}(\hat{k}_t \otimes \mathcal{F}_R(\mathcal{R}_{\boldsymbol{c}(t)} f)) \\
&= \mathcal{U}_{Q^{-T}}(\hat{k}_t \otimes (\sigma_R)_{\boldsymbol{c}(t)} \otimes \mathcal{F}_R(f))
\end{aligned} \tag{4.11}$$

As a result from Axiom 5 which tells us to equate (4.10) and (4.11) yielding:

$$\hat{k}_t(\omega) \otimes (\sigma_R)^{\omega}_{\boldsymbol{c}(t)} \otimes \hat{f}(Q^T \omega) = \hat{k}_t(Q^T \omega) \otimes (\sigma_R)^{Q^T \omega}_{\boldsymbol{c}(t)} \otimes \hat{f}(Q^T \omega).$$

So we get that

$$\hat{k}_t(\omega) \otimes (\sigma_R)^{\omega}_{\boldsymbol{c}(t)} = \hat{k}_t(Q^T \omega) \otimes (\sigma_R)^{Q^T \omega}_{\boldsymbol{c}(t)}$$

But then we get that $\hat{k}_t \otimes (\sigma_R)_{\boldsymbol{c}(t)}$ only depends on the radial component $||\omega||_{[\mathcal{G}]^{-T}} = \sqrt{\omega \cdot [\mathcal{G}]^{-T} \omega}$ as

$$Q^T [\mathcal{G}] Q = [\mathcal{G}] \Leftrightarrow Q^{-1} [\mathcal{G}]^{-T} Q^{-T} = [\mathcal{G}]^{-T},$$

i.e.

$$Q \in SO_{[\mathcal{G}]} \Leftrightarrow Q^{-T} \in SO_{[\mathcal{G}]^{-T}} \Leftrightarrow Q^T \in SO_{[\mathcal{G}]^{-T}}$$

So if we write $\hat{k}_t(\omega) \otimes (\sigma_R)^{\omega}_{\boldsymbol{c}(t)} \equiv g^R_t(||\omega||_{[\mathcal{G}]^{-T}})$, then we get that

$$\hat{k}_t(\omega) = g^R_t(||\omega||_{[\mathcal{G}]^{-T}}) \otimes (\sigma_R)^{\omega}_{-\boldsymbol{c}(t)}.$$

Finally, we conclude from the semigroup property, Axiom 1, and the group representation property of $(\sigma_R)^{\omega}$ that $\boldsymbol{c}(t+s) = \boldsymbol{c}(t) + \boldsymbol{c}(s)$, and then Lemma 4.2 and we see that $\boldsymbol{c}(t) = t\boldsymbol{c} = t\boldsymbol{c}(1)$, which proves the result. □

In addition to the above, the time scale equivariance axiom implies the following:

Lemma 4.6. *For semiring $R \in \{R_L, R_T^{min}, T_+\}$, in the setting $G = \mathbb{R}^2$, the localised time-scale equivariance, Axiom 4, along with Axioms 1,2,3,5 implies that the radial component of the kernel in the Fourier domain satisfies*

$$g^R_t(||\omega||_{[\mathcal{G}]^{-T}}) = g^R_1(\psi(t)||\omega||_{[\mathcal{G}]^{-T}}), \text{ for all } \omega \in \mathbb{R}^2, t \geq 0.$$

and it holds that $\psi(1) = 1$

Proof. By Axiom 4, have that for all $t > 0$, $\psi(t) > 0$ satisfies

$$\mathcal{S}_{\psi(t)} \circ \Phi_1 \circ \mathcal{R}_{\boldsymbol{c}(1)} \circ \mathcal{S}^{-1}_{\psi(t)} = \Phi_t \circ \mathcal{R}_{t\boldsymbol{c}(1)}.$$

Now for any $f \in \mathcal{H}^G_R$, we have from Lemma 4.3 that

$$\begin{aligned}
&\mathcal{F}_R(\mathcal{S}_{\psi(t)}\Phi_1\mathcal{R}_{\boldsymbol{c}(1)}\mathcal{S}^{-1}_{\psi(t)}f)(\omega)\\
&\qquad = \mathcal{F}_R(\Phi_1\mathcal{R}_{\boldsymbol{c}(1)}\mathcal{S}^{-1}_{\psi(t)}f)(\psi(t)\omega)\\
&\qquad = \mathcal{F}_R(k_1 *^R_G \mathcal{R}_{\boldsymbol{c}(1)}\mathcal{S}^{-1}_{\psi(t)}f)(\psi(t)\omega)\\
&\qquad = \hat{k}_1(\psi(t)\omega) \otimes (\sigma_R)^{\psi(t)\omega}_{\boldsymbol{c}(1)} \otimes \mathcal{F}_R(\mathcal{S}_{\psi(t)^{-1}}f)(\psi(t)\omega)\\
&\qquad = g^R_1(\psi(t)||\omega||_{[\mathcal{G}]^{-T}}) \otimes \mathcal{F}_R(f)(\omega)
\end{aligned}$$

since we have that $\hat{k}_t(\omega) = g^R_t(||\omega||_{[\mathcal{G}]^{-T}}) \otimes (\sigma_R)^{\omega}_{-t\boldsymbol{c}(1)}$. On the other hand, we have that

$$\mathcal{F}_R(\Phi_t\mathcal{R}_{\boldsymbol{c}(t)}f)(\omega) = g^R_t(||\omega||_{[\mathcal{G}]^{-T}}) \otimes \mathcal{F}_R(f)(\omega)$$

and since f was arbitrary, we get that

$$g^R_t(||\omega||_{[\mathcal{G}]^{-T}}) = g^R_1(\psi(t)||\omega||_{[\mathcal{G}]^{-T}}),$$

and the result follows. Now taking $t = 1$, it immediately follows that $\psi(1) = 1$.

□

Corollary 4.3. *Fix the semiring $R \in \{R_L, R_T^{min}, T_+\}$ and the Lie group $G = \mathbb{R}^2$. Then the two formulations (4.8) and (4.1) of the time-scale equivariance axiom are equivalent if combined with Axiom 1,2,3,5.*

Proof. From Axiom 1,2,3,5 and Lemma 4.5 it follows that

$$\begin{aligned}
\Phi_t(f) &= k_t *^R_G f \text{ with } k_{t+s} = k_t *^R_G k_s\\
\hat{k}_t \otimes \hat{k}_s &= \hat{k}_{t+s} \text{ for all } t, s \geq 0,\\
\hat{k}^{\mathcal{G},\boldsymbol{c}}_t(\omega) &= (\sigma_R)^{\omega}_{t\boldsymbol{c}} \otimes g_t(\|\omega\|_{[\mathcal{G}]^{-T}}).
\end{aligned}$$

for some function $g_t^R : R \to R$. Now by Lemma 4.6 one has

$$\begin{aligned}
&\Phi_t^{\mathcal{G},\mathbf{c}} = \Phi_1^{|\psi(t)|^{-2}\mathcal{G},t\mathbf{c}} \Leftrightarrow \\
&\Phi_t^{\mathcal{G},\mathbf{c}} \circ \mathcal{R}_{t\mathbf{c}} = \Phi_1^{|\psi(t)|^{-2}\mathcal{G},t\mathbf{c}} \circ \mathcal{R}_{\mathbf{c}} \Leftrightarrow \\
&\Phi_t^{\mathcal{G},\mathbf{0}} = \Phi_1^{|\psi(t)|^{-2}\mathcal{G},\mathbf{0}} \Leftrightarrow \\
&\forall_{\omega\in\mathbb{R}^2} \ : \ g_t(\|\omega\|_{[\mathcal{G}]^{-T}}) \overset{lemma\ 4.6}{=} g_1(\psi(t)\|\omega\|_{[\mathcal{G}]^{-T}}) \\
&\qquad\qquad\qquad\qquad\qquad \overset{(4.8)}{=} g_1(\|\omega\|_{[|\psi(t)|^{-2}\mathcal{G}]^{-T}})
\end{aligned}$$

so that the result follows.

□

The precise formula for $g_t^R = g_t^R$ depending on R and $\psi(t)$ is not relevant for the proof above but from Lemma 4.2 and the technique by Pauwels [Pau+95] it follows that:

Lemma 4.7. *Assume Axioms 1,2,3,4,5 hold. Then the scale-rescaling function Ψ in Axiom 4 is independent on the choice of semiring $R \in \{R_L, T_-, T_+\}$ and the radial (or isotropic) part g_t^R (4.9) of the semiring Fourier transformed kernel is given by*

$$\boxed{\begin{aligned}
&\psi(t) = t^{\frac{1}{\alpha}} \text{ for some } \alpha > 0, \\
&\text{and } g_1^{R_L}(\rho) = e^{-T\rho^\alpha} \text{ and } (\sigma^{R_L})^{\omega}_{\mathbf{c}} = e^{i\omega\cdot\mathbf{c}} \\
&\text{and } g_1^{T_-}(\rho) = \tfrac{T}{\alpha}\rho^\alpha \text{ and } (\sigma^{T_-})^{\omega}_{\mathbf{c}} = \omega\cdot\mathbf{c}.
\end{aligned}}$$

for some rescaling constant $T > 0$.

Before we begin with the proof we first briefly introduce the notion of semiring power operation. If t is an integer(or even a fraction), the expression $a^{\otimes t}$ makes immediate sense as a repeated t fold semiring multiplication. For the general case $t > 0$ it is less obvious. We have the following: $f(t) = a^{\otimes t}$ is the unique (see Remark 4.7 below) continuous solution of

$$\begin{cases}
f(t) \otimes f(s) = f(t+s) \text{ for all } t, s > 0 \\
f(0) = \mathbb{1}, \\
f(1) = a, \\
f(t) \neq a \text{ for } t < 1.
\end{cases} \tag{4.12}$$

As the underlying semiring is assumed to be R_L or $R_\pm$ and commutative, we get that there is always a multiplicative inverse and hence the first equality implies the second.

To show that this indeed yields a unique solution, note that $f_1(0) = f_2(0)$ and $f_1(1) = f_2(1)$ but this implies that by the first equality that $f_1(q) = f_2(q)$ for all $q \in \mathbb{Q}^+$ but by using the density of $\mathbb{Q}$ in $\mathbb{R}$ and the continuity of f we get a uniqueness of f. Here we recall that the ring is embedded in $\mathbb{R}$ or $\mathbb{R} \cup \{\pm\infty\}$ and not per se part of $\mathbb{C}$ as complex square roots are not single-valued (and require branch-cuts for the logarithm to make them single-valued in view of the Residue theorem).

Remark 4.7. *The uniqueness of f is not immediately clear in the general setting where we have a semiring R that is not necessarily embedded in $\mathbb{R}$. However, we can still point towards a specific f as we explain next. The subgroup $Imf \subseteq (R, \otimes)$ must be commutative as $t + s = s + t$, so this is a commutative Lie group which is always isomorphic to direct products of $\mathbb{R}$ and tori $\mathbb{T} = \{z \in \mathbb{C} \mid |z| = 1\}$ [Ban10]. By choosing* *appropriate branch-cuts we can define a single valued (Lie group) logarithm and define $f(t) = \exp(t \log z)$ for all $t > 0$, where* exp *is the Lie group exponential.*

Proof.
First of all set $\rho := \sqrt{\omega^T \, [\mathcal{G}]^{-T} \omega}$.
Map Ψ is invertible, as $\Psi' > 0$, and we may write $s = \psi(t)$ and $t = \psi^{-1}(s)$. Enforcing the time-scale equivariance, Axiom 4, onto the form (4.9) that was obtained in Lemma 4.5, one finds:

$$g_1^R(s\rho) = g_t^R(\rho) \tag{4.13}$$

and the semigroup property, Axiom 1, tells us that:

$$g_t^R(\rho) = g_1^R(\rho)^{\otimes t} \tag{4.14}$$

and thus

$$g_1^R(\rho s) = g_t^R(\rho) = g_1^R(\rho)^{\otimes \psi^{-1}(s)} \tag{4.15}$$

for all $\rho, s > 0$.

- In the morphological semiring setting ($R = T_{\_}$ where $\otimes = +$) (4.15) gives:

$$\begin{aligned} g_1^{R_T^{min}}(\rho s) &= \psi^{-1}(s) \cdot g_1^{R_T^{min}}(\rho) \\ &= \frac{g_1^{R_T^{min}}(s)}{g_1^{R_T^{min}}(1)} g_1^{R_T^{min}}(\rho), \end{aligned} \tag{4.16}$$

  where the final equality follows by the first equality by setting $\rho = 1$. Because we do not care about the constant of rescaling, letting $\frac{T}{\alpha} = g_1^{T_-}(1)$ we find by (4.16) and Corollary 4.2 that $g_1^{R_T^{min}}(\rho) = \frac{T}{\alpha}\rho^\alpha$ and

$$t = \psi^{-1}(s) = s^\alpha \Leftrightarrow s = \psi(t) = t^{\frac{1}{\alpha}}.$$

- In the linear semiring setting ($R = R_L$ where $\otimes = \cdot$) the kernels k_t are in L^1 so their Fourier transforms $\hat{k}_t := \mathcal{F}_G^{R_L} k_t$ are continuous and in $\mathbb{L}_\infty$ [Rud87] and we follow Pauwels [Pau+95] and write $g_1^{R_L}(\rho) = e^{A(\rho)} > 0$. Now (4.15) again gives:

$$\begin{aligned} A(\rho s) &= \psi^{-1}(s) \cdot A(\rho) \\ &= \frac{A(s)}{A(1)} A(\rho), \end{aligned} \tag{4.17}$$

  where again the final equality follows by the first equality by setting $\rho = 1$. We find by (4.17) and Corollary 4.2 and $A(1) = T$ that $g_{R_L}(\rho) = e^{-T\rho^\alpha}$ and again

$$t = \psi^{-1}(s) = s^\alpha \Leftrightarrow s = \psi(t) = t^{\frac{1}{\alpha}}.$$

$\square$

There is a fundamental relation between linear and morphological convolution units in PDE-CNNs:

Lemma 4.8. *[SW16] The Cramer transform applied to functions $f_k \in L^2(\mathbb{R}^2) \cap L^1(\mathbb{R}^2)$, $k = 1, 2$ whose Fourier transform $\mathcal{F} f_k$ is positive in such a way that $-\log \mathcal{F} f_k$ is convex, has the isomorphic property:*

$$\mathcal{C}(f_1 * f_2) = \mathcal{C}(f_1) \,\square\, \mathcal{C}(f_2)$$

Proof. Then:

$$\begin{aligned}\mathcal{C}(f*g) &= \mathfrak{f}\circ -\log\circ\mathcal{F}(f*g) = \mathfrak{f}\circ -\log\mathcal{F}f\cdot\mathcal{F}g \\ &= \mathfrak{f}\circ(-\log\mathcal{F}f + -\log\mathcal{F}g) = \mathcal{C}f\,\square\,\mathcal{C}g.\end{aligned}$$

where in the final equality we used the standard identity $\mathfrak{f}(a+b) = \mathfrak{f}a\,\square\,\mathfrak{f}b$ which completes the proof. $\square$

Now let us derive our PDEs in PDE-G-CNNs from the axioms in the isotropic setting with positive kernels.

Theorem 4.1. *Let $R \in \{R_L, T_-, T_+\}$. Consider the isotropic case with $[\mathcal{G}] = I$. Then the PDE-CNN Axioms (1,2,3,4,5,7) force solutions*

$$\Phi_t f = k_t^{R,\alpha} *_{\mathbb{R}^2}^{R} f$$

where the kernels satisfy

$$k_t^{R,\alpha} *_{\mathbb{R}^2}^{R} k_s^{R,\alpha} = k_{s+t}^{R,\alpha} \geq 0$$

and where their Fourier transforms $\hat{k}_t^{R,\alpha}$ must satisfy

$$\begin{aligned}\hat{k}_t^{R,\alpha} \otimes \hat{k}_s^{R,\alpha} &= \hat{k}_{t+s}^{R,\alpha}, \text{ with} \\ \hat{k}_t^{R,\alpha}(\omega) &= (\sigma_R)_{\boldsymbol{c}t}^{\omega} \otimes g^{R,\alpha}(\psi(t)\|\omega\|) \geq 0\end{aligned}$$

for all $s, t \geq 0$ and all $\omega \in \mathbb{R}^2$, and where $\psi(t) = t^{\frac{1}{\alpha}}$ with $0 < \alpha \leq 2$.

- *For the linear semiring $R = R_L$ we have $g^{R_L,\alpha}(\rho) = e^{-\rho^\alpha}$ we get $0 < \alpha \leq 2$ and kernels*

$$\hat{k}_t^{R_L,\alpha}(\omega) = \mathcal{F}k_t^{R_L,\alpha}(\omega) = e^{-t\|\omega\|^\alpha} e^{-i\omega\cdot\boldsymbol{c}t} \tag{4.18}$$

  that connect the well-known Cauchy kernel $\alpha = 1$ to the Gaussian kernel $\alpha = 2$ and then Φ_t solves the PDE system

$$\begin{cases} \partial_t W = -|\Delta|^{\alpha/2} W - \boldsymbol{c}\cdot\nabla W \\ W(\cdot, 0) = f \end{cases} \tag{4.19}$$

  all up to a linear constant rescaling of time $t \mapsto T(t) = Tt$ for some constant $T > 0$.

- *For the semiring $R = T_-$ one has $g^{T_-,\alpha}(\rho) = \frac{1}{\alpha}\rho^\alpha$ and for $\alpha \geq 1$ the we have*

$$\begin{aligned}&\hat{k}_t^{T_-,\alpha}(\omega) = \mathfrak{f}(k_t^{T_-})(\omega) = \tfrac{t}{\alpha}\|\omega\|^\alpha - t\boldsymbol{c}\cdot\omega \\ &\Leftrightarrow (k_t^{T_-})(\boldsymbol{x}) = \tfrac{t}{\beta}\left\|\tfrac{\boldsymbol{x}}{t} - \boldsymbol{c}\right\|^\beta\end{aligned} \tag{4.20}$$

  with $\frac{1}{\beta} + \frac{1}{\alpha} = 1$ and $\beta \geq 2$, and Φ_t solves the (viscosity solution of the) PDE system

$$\begin{cases} \partial_t W = -\frac{1}{\alpha}\|\nabla W\|^\alpha - \boldsymbol{c}\cdot\nabla W \\ W(\cdot, 0) = f \end{cases} \tag{4.21}$$

  Also up to a linear constant rescaling of time $t \mapsto T(t) = Tt$ for some constant $T > 0$. For $\boldsymbol{c} = \boldsymbol{0}$ the erosion kernels are convex and positive and relate to the α-stable linear kernels via the Cramer-Fourier transform:

$$-\log\mathcal{F}k_t^{R_L,\alpha} = \mathfrak{f}k_t^{T_-,\alpha} \Leftrightarrow \mathcal{C}k_t^{R_L,\alpha} = k_t^{R_M^{min},\alpha} \tag{4.22}$$

  where $\mathcal{C} = \mathfrak{f}\circ(-\log)\circ\mathcal{F}$.

Proof. Fix the semiring $R \in \{R_L, T_-, T_+\}$.

By Axiom 2, we have that

$$\Phi_t f(\boldsymbol{x}) = \bigoplus_{\mathbf{y}\in\mathbb{R}^2} k_t(\boldsymbol{x},\mathbf{y}) \otimes f(\mathbf{y})$$

and by Axiom 3 one has $\Phi_t \circ \mathcal{L}_{\boldsymbol{x}} = \mathcal{L}_{\boldsymbol{x}} \circ \Phi_t$, we get

$$\begin{aligned}
\Phi_t(\mathcal{L}_{\boldsymbol{x}} f)(\boldsymbol{x}) &= \bigoplus_{\mathbf{y}\in\mathbb{R}^2} k_t(\boldsymbol{x},\mathbf{y}) \otimes f(-\boldsymbol{x}+\mathbf{y}) \\
&= \bigoplus_{\mathbf{u}\in\mathbb{R}^2} k_t(0,\mathbf{u}) \otimes f(\mathbf{u}) \\
&= \mathcal{L}_{\boldsymbol{x}}(\Phi_t f)(\boldsymbol{x})
\end{aligned}$$

Now, writing $\mathbf{u} = -\boldsymbol{x} + \mathbf{y}$ gives

$$\bigoplus_{\mathbf{y}\in\mathbb{R}^2} k_t(\boldsymbol{x},\mathbf{y}) \otimes f(-\boldsymbol{x}+\mathbf{y}) = \bigoplus_{\mathbf{y}\in\mathbb{R}^2} k_t(0,-\boldsymbol{x}+\mathbf{y}) \otimes f(-\boldsymbol{x}+\mathbf{y})$$

The above equality implies that $k_t(\boldsymbol{x},\mathbf{y}) = k_t(0,-\boldsymbol{x}+\mathbf{y}) =: k_t^R(\boldsymbol{x}-\mathbf{y})$. Thereby Φ_t is given by the semiring convolution

$$\Phi_t f(\boldsymbol{x}) = \bigoplus_{\mathbf{y}\in\mathbb{R}^2} k_t^R(\boldsymbol{x}-\mathbf{y}) \otimes f(\mathbf{y}) = (k_t^R *^R f)(\boldsymbol{x}).$$

using short-hand notation $*^R = *^R_{\mathbb{R}^2}$ for convolutions. Now, by Lemma 3.3 and taking the semiring Fourier transform $\mathcal{F}_R$ yields

$$\mathcal{F}_R(k_t^R * f)(\omega) = \mathcal{F}_R(k_t^R)(\omega) \otimes \mathcal{F}_R(f)(\omega).$$

Now, the Axiom 1 $\Phi_t \circ \Phi_s = \Phi_{t+s}$ implies that

$$\Phi_t \circ \Phi_s f = k_t *^R (k_s^R *^R f) = (k_t^R *^R k_s^R) *^R f = \Phi_{t+s} f$$

so if Φ is a linear or morphological scale space, then we have

$$k_t^R *^R k_s^R = k_{t+s}^R \text{ and } \widehat{k}_{t+s}^R = \widehat{k}_t^R \otimes \widehat{k}_s^R.$$

Now transferring the isotropy, Axiom 5, together with Lemma 4.5 noting $\mathcal{G} = I$ implies that

$$\hat{k}_t^R(\omega) = (\sigma_R)^{\omega}_{\boldsymbol{tc}} \otimes g_t^R(\|\omega\|).$$

Now by the time-scaling equivariance, Axiom 4, we have from Lemma 4.6 that

$$g_t^R(||\omega||) = g^R(\psi(t)||\omega||)$$

with $g^R = g^R_{t=1}$.

Now apply Lemma 4.7 and we deduce (4.18) and the first part of (4.20). The second part of (4.20) follows by the first and the first part of Lemma 4.3. This sets the convolutions with the prescribed kernels.

Finally, we investigate constraints on the α parameter and build upon standard theory that tells us that the convolution with these kernels indeed solve the mentioned PDEs in the appropriate manner. We do that final part separately for $R = R_L$ and $R = T_-$.

In the linear semiring setting $R = R_L$ it is well-known (see e.g. [Dui+04]) that the convolutions solve α-Stable Levy process PDEs (4.19). Here we stress that (apart from excluding $\alpha = 0$ in view of Axiom 2) we must exclude that cases $\alpha > 2$. To this end we note that for $\alpha > 2$ the kernel in the Fourier domain is twice differentiable in the Fourier domain with 0 second order derivative which would entail a vanishing second order moment which cannot happen for a positive kernel (Axiom 7):

$$\alpha > 2 \Rightarrow \Delta_\omega \hat{k}_t^{R_L}(\mathbf{0}) = 0 \Rightarrow \int_{\mathbb{R}^2} \|\boldsymbol{x}\|^2 k_t(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = 0.$$

In the morphological setting we observe that the spatial kernels in (4.20) are indeed positive and convex for $\alpha > 1$ and we can apply Lemma 4.8 and the the results by Schmidt-Weickert [SW16] to obtain (4.22). It also justifies the morphological PDE system (4.22), where we recall from the book of Evans [Eva98, ch:3.4.2, Thm.1 & 2, ch:10.3.4] Lax-Oleinik solutions are indeed the unique viscosity solutions of (4.22). □

Remark 4.8. *The fact that Lax-Oleinik solutions are viscosity solutions for time dependent HJB-PDEs (or 'morphological scale space PDEs') can be generalized to Riemannian manifolds on Lie groups like $(\mathbb{R}^2, \mathcal{G})$ in general [Bel+23, Prop.1] and this is already needed for the upcoming Theorem 4.2.*

4.3 The Axiomatic Solutions for Geometric Learning on $\mathbb{R}^2$: PDE-CNNs on $\mathbb{R}^2$

Now let us derive the PDEs used in PDE-G-CNNs on $G = \mathbb{R}^2$, for the general case where $[\mathcal{G}]^T = [\mathcal{G}] > 0$ and where we drop the positivity constraint (i.e. optional Axiom 7) and replace it by the isometry constraint (optional Axiom 6).

Theorem 4.2. *Let $R \in \{R_L, T_-, T_+\}$, and consider the general case where $[\mathcal{G}]$ is Symmetric and positive definite. The PDE-CNN Axioms (1,2,3,4,5,6) force solutions*

$$\Phi_t f = k_t^{R,\alpha} *_{\mathbb{R}^2}^R f$$

where the possibly complex-valued kernels satisfy

$$k_t^{R,\alpha} *_{\mathbb{R}^2}^R k_s^{R,\alpha} = k_{s+t}^{R,\alpha}$$

and where their Fourier transforms $\hat{k}_t^{R,\alpha}$ must satisfy

$$\begin{array}{rcl} \hat{k}_t^{R,\alpha} \otimes \hat{k}_s^{R,\alpha} & = & \hat{k}_{t+s}^{R,\alpha}, \text{ with} \\ \hat{k}_t^{R,\alpha}(\omega) & \equiv & (\sigma_R)^{\omega}_{\boldsymbol{c}t} \otimes g^{R,\alpha}(\psi(t)\|\omega\|_{[\mathcal{G}]^{-T}}), \end{array}$$

for all $s, t \geq 0$ and all $\omega \in \mathbb{R}^2$, and with $\psi(t) = t^{\frac{1}{\alpha}}, \alpha > 0$.

- *For the linear semiring $R = R_L$ we have $g^{R_L,\alpha}(\rho) = e^{-i\rho^\alpha}$ and kernels*

$$\hat{k}_t^{R_L,\alpha}(\omega) = \mathcal{F}k_t^{R_L,\alpha}(\omega) \equiv e^{-it\boldsymbol{c}\cdot\omega} e^{-it|\omega\cdot[\mathcal{G}]^{-T}\omega|^{\frac{\alpha}{2}}} \tag{4.23}$$

  that boil down to convection and quantum mechanical wave propagation:

$$\left\{ \begin{array}{rcl} \partial_t W & = & -c \cdot \nabla W - i|\Delta_{\mathcal{G}}|^\alpha W, \\ W(\cdot, 0) & = & f \end{array} \right.$$

  with $W(\cdot, t) = \Phi_t(f)$ up to a linear constant rescaling of time $t \mapsto T(t) = Tt$ for some constant $T > 0$.

- *For the semiring $R = T_-$ one has $g^{T_-,\alpha}(\rho) = \frac{1}{\alpha}\rho^{\alpha}$ and for $\alpha > 1$ one has kernels:*

$$\begin{aligned} \mathfrak{f}(k_t^{T_-,\alpha})(\omega) &= \tfrac{t}{\alpha}\left|\sqrt{\omega \cdot [\mathcal{G}]^{-T}\omega}\right|^{\alpha} - t\boldsymbol{c}\cdot\omega \\ \Leftrightarrow (k_t^{T_-})(\boldsymbol{x}) &= \tfrac{t}{\beta}\left|\frac{\sqrt{(\boldsymbol{x}-t\boldsymbol{c})\cdot[\mathcal{G}](\boldsymbol{x}-t\boldsymbol{c})}}{t}\right|^{\beta} \end{aligned} \tag{4.24}$$

with $\frac{1}{\beta}+\frac{1}{\alpha}=1$ and $\beta > 1$, and Φ_t solves the (viscosity solution of the) time dependent HJB-PDE system

$$\begin{cases} \partial_t W &= -\frac{1}{\alpha}\|\nabla_{\mathcal{G}} W\|^{\alpha} - \boldsymbol{c}\cdot\nabla W \\ W(\cdot,0) &= f \end{cases} \tag{4.25}$$

where $W(\cdot,t) = \Phi_t(f)$ all up to a linear constant rescaling of time $t \mapsto T(t) = Tt$ for some constant $T > 0$.

Proof. The proof of Theorem 4.2 is to a very large part analogous to the proof of Theorem 4.1, where we subsequently i) combine Axiom 2 and Axiom 3 to arrive at convolution kernel operators, ii) Lemma 3.3, iii) Lemma 4.5, iv) Lemma 4.6, v) Lemma 4.7, vi) standard results on linear and morphological PDEs [Eva98; SW16; BPD23; AQV94].

Therefore let us concentrate on the three main things that become different and address the adaptation of the proof:

1.) we no longer impose the positivity, Axiom 7 at imposed $0 \leq \alpha \leq 2$ in the linear semiring case, and as kernels can now be complex-valued the Schrödinger equation comes allowed as follows. We follows the procedure for Theorem up to Lemma 4.7. Now since g_t^R now becomes complex-valued, (4.17) implies that we have $|A(\rho)|_{\mathbb{C}} = \rho^{2\alpha}$. This implies that $A(\rho) = \rho^{2\alpha}e^{i\beta(\rho)}$. But then again by (4.17), we have that

$$e^{i\beta(\rho s)} = \frac{1}{e^{i\beta(1)}}e^{i\beta(\rho)}e^{i\beta(s)} \Leftrightarrow \beta(\rho s) = \beta(s) + \beta(\rho) - \beta(1)$$

which implies that $\beta(\rho) = \beta$ for some cosntant β. So we have $A(\rho) = \rho^{2\alpha}e^{i\beta}$ and hence $g_1^{R_L}(\rho) = e^{\rho^{2\alpha}e^{i\beta}}$.

2.) the newly imposed isometry, Axiom 6, excludes the fractional diffusions in the linear semiring setting, as we will see next. From item 1, above we have

$$\begin{aligned} \|k_t *^{R_L} f\|^2_{L^2(\mathbb{R}^2)} &= \int_{\mathbb{R}^2}\left|e^{-t\|\omega\|^{2\alpha}_{[\mathcal{G}]^{-T}}e^{i\beta}}\right||\hat{f}(\omega)|^2 \mathrm{d}\omega \\ &= \int_{\mathbb{R}^2}|\hat{f}(\omega)|^2\,\mathrm{d}\omega = \|f\|^2_{L^2(\mathbb{R}^2)} \Leftrightarrow e^{i\beta} = \pm i \end{aligned}$$

for $0 \neq f \in L^2(\mathbb{R}^2), t > 0$. When extending to complex numbers one has that the only surviving linear PDE is the Schrödinger PDE where the skew-adjoint generator provides a unitary solution operator:

$$\forall_{t\geq 0}: \|\Phi_t(f)\|^2_{L^2(\mathbb{R}^2)} = \|k_{it} *^{R_L} f\|^2_{L^2(\mathbb{R}^2)} = \|f\|^2_{L^2(\mathbb{R}^2)}.$$

In the tropical semiring setting the isometry property does not exclude the morphological kernel solutions, neither for $R = R_-$ where we solve a erosion PDE by infimal convolution, and neither for $R = R_+$ where we solve a dilation PDE for supremal convolution. This follows directly from Lemma 3.2 together with (3.6) and (3.8), for example in the $R = R_+$ case we have:

$$\delta_{R_+}(f, \mathfrak{o}) = e^{I_{R_+}(f)} \text{ and } I_{R_+}(k_t \circledast f) = I_{R_+}(f) \tag{4.26}$$

because indeed $I_{R_+}(k_t) = 0$.

Conclusion: the isometric property (6) only puts constraints in the linear semiring setting, where it forces Schrödinger PDEs.

3.) the shift-invariant metric tensor field is no longer the identity and Axiom 6 forces us to include anisotropy via $[\mathcal{G}]$. Corollary 4.3 and Lemma 4.7 now yield (4.23) and (4.24) by direct computation. Finally, in the tropical semiring, the solution operator still provides the Lax-Oleinik viscosity of (4.25) on the general Riemannian setting $(\mathbb{R}^2, \mathcal{G})$ by Remark 4.8 $\square$

Chapter 5
Axiomatic PDE-G-CNN on $SE(2)$

In this section we constrain ourselves to the case $G = SE(2) = \mathbb{R}^2 \rtimes SO(2)$. We do this to keep the formulas concise. Already the case $SE(3)$ where linear Fourier transform can be used to find exact solutions to (fractional) heat processes on $SE(3)$ [DBM19, Thm.1] [PD17] involves a lot of technical indexing that only clutters the algebraic structure. In principle, it can be expected that many results can be transferred to the general case of $G = \mathbb{R}^d \rtimes SO(J)$.

5.1 Formulation of the Axioms

We index our equivariant operator $\Phi_t = \Phi_t^{[\mathcal{G}],\boldsymbol{c}}$ in the network by 1) a left-invariant metric tensor field $\mathcal{G}$ and a left-invariant vector field c. By left-invariance, they are uniquely determined by

$$c = \sum_{i=1}^{3} c^i \mathcal{A}_i$$

with $\boldsymbol{c} = (c^1, c^2, c^3) \in \mathbb{R}^3$, and with

$$\mathcal{G} = \sum_{i,j=1}^{\dim(G)=3} g_{ij}\ \omega^i \otimes \omega^j$$

and $0 < [\mathcal{G}] = [g_{ij}] = [g_{ji}] \in \mathbb{R}^{3\times3}$ the constant matrix coefficients relative to the left-invariant dual frame. Let $d_\mathcal{G}$ denote the corresponding Riemannian distance and $\text{cut}_e(G)$ the cut-locus relative to unity element $e \in G$.

Definition 5.1. ***(local radial isometries)***
We call $\varphi : \Omega \to \mathbb{R}$ a local radial isometry (w.r.t. the unity element e) if

$$d_\mathcal{G}(\varphi(x), \varphi(e)) = d_\mathcal{G}(x, e), \text{ and } \varphi(e) = e, \tag{5.1}$$

for all x in an open set Ω around e outside the cutlocus.

Definition 5.2. ***(special local radial isometries)***
Within the set of local radial isometries we consider the special local radial isometries given by

$$\varphi_Q(g) = \exp_\mathcal{G} \circ Q \circ \log_\mathcal{G}(g),\ \ \varphi_Q : \Omega \to \Omega \tag{5.2}$$

indexed by orthogonal linear maps $Q : T_e(G) \to T_e(G)$, where $\exp_\mathcal{G} : T_e(G) \to G$ and $\log_\mathcal{G} : G \to T_e(G)$.

We impose the following requirements for our PDE-G-CNNs on G:

1. **Semigroup property:** For all $s, t \geq 0$ we require

$$\Phi_t \circ \Phi_s = \Phi_{t+s},$$

and we require strong continuity: for all $f \in \mathcal{H}_R^G$ we have that $\lim_{t\downarrow 0} \Phi_t f = f$.

2. **Quasilinearity and Kernel operator:** For all $t \geq 0$ we must have for all $\alpha_1, \alpha_2 \in R$ constant and all $f_1, f_2 \in \mathcal{H}_R^G$ that quasi-linearity holds

$$\Phi_t(\alpha_1 \otimes f_1 \oplus \alpha_2 \otimes f_2) = \alpha_1 \otimes \Phi_t(f_1) \oplus \alpha_2 \otimes \Phi_t(f_2).$$

More specifically, we will assume that the evolution operator Φ_t can for each $t > 0$ be written as

$$\Phi_t f(g) = \bigoplus_{h \in SE(2)} \kappa_t(g,h) \otimes f(h) \ ,$$

for every $h \in SE(2)$ for some kernel $\kappa_t : SE(2) \times SE(2) \to R$ s.t. $k_t(\cdot, h) \in \mathcal{H}_R^G$ for every $h \in SE(2)$.

3. **Equivariance:** For all $(g,t) \in SE(2) \times \mathbb{R}^+$ we require

$$\Phi_t \circ \mathcal{L}_g = \mathcal{L}_g \circ \Phi_t, \tag{5.3}$$

with the shift-operator given by left-action $(\mathcal{L}_g f)(h) = f(g^{-1}h)$, for all $g, h \in G = SE(2)$ and all $f \in \mathcal{H}_R^G$.

4. **Time-scale equivariance: Scaling equivariance w.r.t. Riemannian metric $\mathcal{G}$ and convection vector $\boldsymbol{c}$**

$$\exists_{\psi \in C^1(\mathbb{R}^+), \psi' > 0} \forall_{\mathcal{G}, \boldsymbol{c} \text{ LI}} \ \forall_{t \geq 0} \ : \Phi_t^{\mathcal{G}, c} = \Phi_1^{|\psi(t)|^{-2}\mathcal{G}, tc} \tag{5.4}$$

5. **Localised Isotropy w.r.t. the Riemannian metric:** We require that there exists a $\boldsymbol{c} \in T_e G$ such that the generator

$$\Psi := \left.\frac{d}{dt}\right|_{t=0} (\Phi_t \circ \mathcal{R}_{\exp t\boldsymbol{c}}) \tag{5.5}$$

commutes at unity element e with all special local radial isometries $\varphi = \varphi_Q : \Omega \to \Omega$:

$$\begin{array}{rcl} ((\Psi \circ \varphi_*)f)(e) & = & ((\varphi_* \circ \Psi)f)(e), \iff \\ \Psi(f \circ \varphi)(e) & = & (\Psi f)(e) \end{array} \tag{5.6}$$

where $\varphi_* : \mathcal{H}_R(\Omega) \to \mathcal{H}_R(\Omega)$ is the push-forward of functions associated with φ, i.e. $\varphi_* f := f \circ \varphi^{-1}$.

6. **Optional Axiom; Isometry Constraint:**

$$\forall_{f \in \mathcal{H}_R, t \geq 0} \ : \ \delta_R(f, \mathbb{o}) = \delta_R(\Phi_t f, \mathbb{o})$$

7. **Optional Axiom; Positivity:**
We require the kernel κ_t at Axiom 2 to be positive.

Remark 5.1. ***(Explanation on the generator and its domain)*** *The strong continuity with semi-group property provide strong continuity for all $t > 0$ and the generator*

$$\Psi := \left.\frac{d}{dt}\Phi_t\right|_{t=0} = \lim_{t \downarrow 0} t^{-1}(\Phi_t - I) \tag{5.7}$$

is a closed operator that is well-defined on a dense subset $\mathcal{D}_L(\Psi)$ of $\mathcal{H}_R^G$, cf. [Yos68], in the linear semiring setting. In the tropical setting the dense domain $\mathcal{D}_L(\Psi)$ consists of all elements in $\mathcal{H}_{T_\pm}^G$ that are absolutely continuous. Such elements have generalized derivatives and are differentiable almost everywhere. Recall that the constraint to absolutely continuous functions is also important in view of the Fenchel transform (i.e. the semiring Fourier transform for $R = T_-$), cf. (4.7).

Remark 5.2. ***(Explanation on Local Isotropy Axiom)***
Recall that in the Lie group case $\mathbb{R}^d$ there was no difference between global and local isometries, but in $G = SE(2)$ there is a substantial difference. We choose the for the evolution operators more restrictive local isotropy. Local isotropy in combination with equivariance yields global isotropy. In Axiom 5 we constrain ourselves to local radial *isometries. Inversion $g \mapsto g^{-1}$ is an example of a local radial isometry that is not a local isometry. Indeed $d_\mathcal{G}(g,e) = d_\mathcal{G}(e,g^{-1}) = d_\mathcal{G}(g^{-1},e)$ works for the unity element e but there exist many g_1, g_2 locally around e such that $d_\mathcal{G}(g_1,g_2) \neq d_\mathcal{G}(g_1^{-1},g_2^{-1})$.*

Remark 5.3. ***Explicit reformulation of Axiom 5***
We restrict ourselves in (5.6) to specific *local radial isometries that map the unity element e to itself. Namely to the ones given by (5.2) indexed by isometric linear maps $Q : T_e(G) \to T_e(G)$, The next lemma's explain the restriction and consequences.*

Lemma 5.1. *For all $g \in G$ the map $L_g : G \to G$ given by $L_g(q) = gq$ is an element in the group of global/local isometries of $d_\mathcal{G}$ with inverse $(L_g)^{-1} = L_{g^{-1}}$.*

Proof. The metric tensor field $\mathcal{G}$ was assumed to be left-invariant, thereby $d_\mathcal{G}$ is left-invariant. Concatenating global isometries gives again a global isometry. □

Lemma 5.2. *Any isometry $\varphi : \Omega \to \Omega$ with $\varphi(h) = h$ may be restricted to a small enough open ball B_h around h such that $\varphi|_{B_h} = L_h \circ \varphi_Q \circ L_h^{-1}|_{B_h}$, where $Q = (L_h^{-1})_* \circ \varphi_*(h) \circ (L_h)_*$ is an isometry on $T_e(G)$.*

Proof. For any isometry $f : G \to G$, the naturality of the Riemannian exponential and logarithm make the diagrams

$$\begin{tikzcd} T_hG \arrow[r, "f_*"] \arrow[d, "\exp_\mathcal{G}"] & T_{f(h)}G \arrow[d, "\exp_\mathcal{G}"] \\ G \arrow[r, "f"] & G \end{tikzcd} \qquad \begin{tikzcd} \Omega \subset G \arrow[r, "f"] \arrow[d, "\log_\mathcal{G}"] & \Omega \subset G \arrow[d, "\log_\mathcal{G}"] \\ T_hG \arrow[r, "f_*"] & T_{f(h)}G \end{tikzcd}$$

commute where we choose the radius of $B_h \subset \Omega$ small enough to stay away from the cut locus of h, and by left-invariance $L_{h^{-1}}(\Omega) = h^{-1}\Omega$ is then disjoint from the cut locus of e, as left translations are isometries by the previous lemma. Now both f and f_* are isometries.

First, we consider the case $\varphi(e) = e$. When g is away from the cut-locus $\mathrm{cut}(e)$ a geodesic γ connecting $\gamma(0) = e$ and $\gamma(1) = g$ is uniquely determined by $\gamma'(0) = v = \log_\mathcal{G}(g)$ and we write $\gamma(t) = \exp_\mathcal{G}(tv)$, $t \in [0,1]$.

Now consider $\tilde{\gamma} = \varphi \circ \gamma$ then since φ is an isometry and γ is a minimizing geodesic also $\tilde{\gamma}$ is a minimizing geodesic passing through e and $\tilde{\gamma}(1) = \varphi(g)$. Now, observe that we have $\tilde{\gamma}'(0) = \varphi_*(\gamma'(0)) = \varphi_*(\log_\mathcal{G}(g)) = \log_\mathcal{G}(\varphi(g))$ where φ_* is taken at e and the last equality follows from the naturality of $\log_\mathcal{G}$. Therefore, we evidently get that $\exp_\mathcal{G}(\tilde{\gamma}'(0)) = \varphi(g)$. Then by the uniqueness of minimising geodesic, we get that

$$\tilde{\gamma}(t) = \exp_\mathcal{G}(t\tilde{\gamma}'(0)) = \exp_\mathcal{G}(t(\varphi)_*(\log_\mathcal{G}(g))), \quad t \in [0,1]$$

Now, setting $t = 1$ yields

$$\varphi(g) = \exp_{\mathcal{G}} \circ \varphi_* \circ \log_{\mathcal{G}}(g).$$

Now for arbitrary $\varphi(h) = h$, conjugating by L_h:

$$\tilde{\varphi} = L_h^{-1} \circ \varphi \circ L_h$$

yields a local isometry with $\tilde{\varphi}(e) = e$. But then we get

$$\tilde{\varphi} = \exp_{\mathcal{G}} \circ \tilde{\varphi}_* \circ \log_{\mathcal{G}} = \exp_{\mathcal{G}} \circ Q \circ \log_{\mathcal{G}} = \tilde{\varphi}_Q$$

where $\tilde{\varphi}_* = (L_h^{-1})_* \circ \varphi_*(h) \circ (L_h)_* = Q$. But then we get locally, $\varphi = L_h \circ \tilde{\varphi}_Q \circ L_h^{-1}$ which was to be shown $\square$

Definition 5.3 (Laplacian)**.** *The Laplacian $\Delta_{\mathcal{G}} = div \circ grad$ at a point g can be written by introducing Riemannian normal coordinates y^i w.r.t $\mathcal{G}$ at g:*

$$(\Delta_{\mathcal{G}} f)(g) := \sum_{i=1}^{3} \partial_{y^i}\big|_g (\partial_{y^i} f). \tag{5.8}$$

As G is unimodular one can also express the Laplacian in the left-invariant frame of vector fields[a]:

$$\Delta_{\mathcal{G}} f = div \circ gradf = div \circ \mathcal{G}^{-1} \mathrm{d}f = \sum_{i,j=1}^{3} g^{ij} \mathcal{A}_i \mathcal{A}_j f. \tag{5.9}$$

where $[g^{ij}]$ is the inverse matrix of the matrix $[g_{ij}]$ with $g_{ij} = \mathcal{G}(\mathcal{A}_i, \mathcal{A}_j)$.

[a]The left-invariant frame is not induced by a local coordinate frame.

Lemma 5.3. *The local maps $\varphi_Q : \Omega \to \Omega$ given by (5.2) are indeed smooth local radial isometries (satisfying (5.1)), and they satisfy the following properties:*

$$\begin{array}{rcl} (\varphi_Q)_*\big|_e &=& Q, \\ (\Delta_{\mathcal{G}} \circ (\varphi_Q)_*)\big|_e &=& (\varphi_Q)_* \circ \Delta_{\mathcal{G}}\big|_e, \\ \|\nabla_{\mathcal{G}}((\varphi_Q)_* f)(e)\| &=& \|\nabla_{\mathcal{G}} f(e)\|, \end{array} \tag{5.10}$$

for all smooth $f : \Omega \to \Omega$.

Proof. Let $\gamma : [0, 1] \to G$ be the geodesic starting from $\gamma(0) = e$ and ending at $\gamma(1) = g \in \Omega_g$ then

$$\log_{\mathcal{G}} g = \gamma'(0) \text{ and } \exp_{\mathcal{G}} \gamma'(0) = g \text{ and } \|\gamma'(0)\| = d_{\mathcal{G}}(g, e).$$

As a result, since $\varphi_Q(e) = e$, the isometry of Q carries over to the isometry of φ_Q as we have

$$\begin{aligned} \forall_{g \in \Omega_Q} \ : \ d_{\mathcal{G}}(\varphi_Q(g), \varphi_Q(e)) &= \|Q\gamma'(0)\| = \|\gamma'(0)\| \\ &= d_{\mathcal{G}}(g, e) \end{aligned}$$

from which the first statement follows.

Now the remaining identities (5.10) follow by introduction of normal coordinates around $g = e$ which gives

$$
\begin{aligned}
y^j(\varphi_Q(g)) &= \sum_{i=1}^{3} Q_i^j y^i(g), \\
d_{\mathcal{G}}(e, g) &= \|\mathbf{y}(g)\|, \\
\Delta_{\mathcal{G}} f(e) &= \sum_{i=1}^{3} \partial_{y^i}^2 f(e)
\end{aligned}
$$

Now, we compute the LHS: for $f \in C^\infty(\Omega)$ we have

$$
\begin{aligned}
(\Delta_{\mathcal{G}} \circ (\varphi_Q)_*)_e(f) &= \Delta_{\mathcal{G}}(f \circ \varphi_Q) \\
&= \sum_{i=1}^{3} \partial_{y^i}^2 (f \circ \varphi_Q)(e) \\
&= \sum_{i=1}^{3} \partial_{y^i}|_e \partial_{y^i}(f \circ \varphi_Q) \\
&= \sum_{i=1}^{3} \partial_{y^i}|_e \left((\varphi_Q)_*(\partial_{y^i}) f\right)
\end{aligned}
$$

Now, evidently, under the normal coordinate chart, $(\varphi_Q)_*$ is linear by construction, we, therefore, have that $(\varphi_Q)_* = Q$ w.r.t the frame induced by the normal coordinate not only at e but also in the neighbourhood of e. We thus get

$$
\begin{aligned}
\sum_{i=1}^{3} \partial_{y^i}|_e \left((\varphi_Q)_*(\partial_{y^i}) f\right) &= \sum_{i=1}^{3} \partial_{y^i}|_e \left(\sum_{j}^{3} Q_i^j (\partial_{y^j} f) \circ \varphi_Q \right) \\
&= \sum_{i=1}^{3} \sum_{j=1}^{3} Q_i^j (\varphi_Q)_*(\partial_{y^i}|_e)(\partial_{y^j} f) \\
&= \sum_{i=1}^{3} \sum_{j=1}^{3} \sum_{k=1}^{3} Q_i^j Q_i^k \partial_{y^k}|_e(\partial_{y^j} f)
\end{aligned}
$$

where in the last equality, we used that $\varphi_Q(e) = e$. Now observe that we have $\sum_{i=1}^{3} Q_i^j Q_i^k = \sum_{i=1}^{3} (Q^T)_j^i Q_i^k = \delta_j^k$ as under the normal coordinate, the matrix representation of $\mathcal{G}$ equals the identity. We therefore get

$$
(\Delta_{\mathcal{G}} \circ (\varphi_Q)_*)_e(f) = \sum_{i=1}^{3} \partial_{y^i}|_e(\partial_{y^i} f) = (\varphi_Q)_* \circ \Delta_{\mathcal{G}}\big|_e (f)
$$

Now for the last property, we have

$$
||\nabla_{\mathcal{G}}((\varphi_Q)_* f)(e)||^2 = \sum_{i=1}^{n} | \sum_{j=1}^{n} (Q_i^j \partial_{y^j} f(e)|^2
$$

But since $Q \in SO(n)$, we have

$$\begin{aligned}\sum_{i=1}^{n} | \sum_{j=1}^{n} (Q_i^j \partial_{y^j} f(e)|^2 &= \sum_{i=1}^{n} |(\partial_{y^i} f)(e)|^2 \\ &= ||\nabla_{\mathcal{G}} f(e)||^2\end{aligned}$$

□

Remark 5.4. ***(Axioms 1–5 restrict the generator)***
Axiom 1, 3 and 5 provide that

$$\begin{aligned}\Psi \circ \varphi_*|_e &= \varphi_* \circ \Psi|_e\,, \\ \Psi \circ (L_g)_* &= (L_g)_* \circ \Psi\end{aligned} \tag{5.11}$$

for all $\varphi = \varphi_Q$ with $Q \in SO_{[\mathcal{G}]}(G)$, for all $g \in G$. This puts a major *restrictions in constructing the quasi-linear (Axiom 2) strongly continuous semigroups (Axiom 1) Φ_t!*

Definition 5.4. *Denote the set of all smooth and compactly supported elements in $\mathcal{H}_G^R$ by $\mathcal{D}(G)$.*

Operator $\Psi : \mathcal{H}_G^R \to \mathcal{H}_G^R$ is a local differential operator if ψ maps $\mathcal{D}(G)$ into itself and if moreover

$$\text{supp}(\Psi f) \subset \text{supp}(f)$$

for all $f \in \mathcal{D}(G) \subset \mathcal{H}_G^R$. If moreover Ψ is linear then by the Peetre theorem Ψ is a differential operator of some order $n \geq 0$ and we may write

$$\Psi f = \sum_{|\alpha| \leq n} \alpha^i(\cdot)\partial_{x^i} f$$

for some smooth coefficient functions $\alpha^i : G \to \mathbb{R}$ and all $f \in \mathcal{D}(G)$.

If operator Ψ is the generator (5.7) of a tropical semigroup (quasi-linear w.r.t. semiring $T_{\pm}$) that satisfies the isotropy axiom then by quasilinearity of the equivariant evolution Φ_t we have that $\Psi(f + C) = \Psi(f)$ *only* holds for locally constant $C \in \mathcal{H}_G^R$, and by isotropy we have

$$\Psi f = F_1(\nabla_{\mathcal{G}} f) = F_2(\|\nabla_{\mathcal{G}} f\|) \text{ for all } f \in \mathcal{D}(G)$$

where F_1 is a smooth mapping of the space of vector fields on G towards $\mathcal{D}(G)$ and where $F_2 : \mathbb{R}^+ \to \mathbb{R}^+$ is smooth.

5.2 Axiomatic Solutions: PDE-G-CNNs on $SE(2)$ & R_L

Now we constrain ourselves to the Lie group $G = SE(2)$ and the linear semi-ring setting.

For multi-index $\alpha \in \mathbb{N}^n$ denote $D^\alpha = \partial_1^{\alpha_1} \ldots \partial_n^{\alpha_n}$.

Lemma 5.4. *For each point $g \in G$, in the normal corrdinate centered at g, if for $\Psi_n := \sum_{|\alpha|=n} c_\alpha D^\alpha$ such that $(\Psi_n \circ (\varphi_Q)_*)f(g) = \Psi_n f(g)$ for any local radial isometry φ_Q centred at g, then we get that $\Psi_n = \Delta^{|\alpha|/2}$ if $|\alpha| = n$ is even. If n is odd, then $\Psi_n = 0$.*

Proof. Note that by observing that $\Psi_n = T_{(i_1,\ldots,i_n)}\partial_{i_1}\ldots\partial_{i_n}$, we can regard $T_{(i_1,\ldots,i_n)}$ as tensor coefficients and hence regard T as rank n tensor. Then we get that the condition $(\Psi_n \circ (\varphi_Q)_*)f(g) = \Psi_n f(g)$ implies that the tensor T is a $O(n)$ invariant under the normal coordinate chart. Now it is a standard result that $O(n)$ invariant tensors can be expressed as linear combinations of tensor products of knocker delta tensors. This implies that if n is odd then as δ_{ij} is rank 2, we get that if T is non-zero, we have that n is even. So if n is even, we get that

$$T_{(i_1,\ldots,i_n)}\partial_{i_1}\ldots\partial_{i_n} = \delta_{i_1i_2}\delta_{i_2i_3}\ldots\delta_{i_{n-1}i_n}\partial_{i_1}\ldots\partial_{i_n}$$

But this is precisely what we needed as

$$\begin{aligned}\delta_{i_1i_2}\delta_{i_2i_3}\ldots\delta_{i_{n-1}i_n}\partial_{i_1}\ldots\partial_{i_n} &= (\delta_{i_1i_2}\partial_{i_1}\partial_{i_2})\ldots(\delta_{i_{n-1}i_n}\partial_{i_{n-1}}\partial_{i_n}) \\ &= \Delta^{n/2}\end{aligned}$$

□

Corollary 5.1. *Let Ψ be the generator defined in (5.5), then the time-scale equivariance axiom together with the localized isometry axiom, force Ψ to be of the form*

$$\Psi = \Delta^{\alpha/2} \tag{5.12}$$

for some even α, up to a scalar multiple.

Proof. In general, the localized isometry axiom implies that Ψ ((5.5)) is a linear combination of powers of the Laplacian. Because these terms scale differently w.r.t. metric, the time-scale equivariance axiom forces us to only include one homogeneous power of the Laplacian. □

Remark 5.5. *The domain $\mathcal{D}_L(\Psi)$ of Ψ in the linear semiring setting is given by the isotropic Sobolev space on G of order α. I.e. $f \in \mathcal{D}_L(\Psi)$ if f is absolutely continuous and both f and the generalized Laplacian $|\Delta|^{\alpha/2}$ applied to f s again in $L^2(G)$.*

Theorem 5.1. *Constraint (5.11) with Axiom 2 & 4 again will yield us* $\psi(t) = t^{\frac{1}{\alpha}}$ *along with*

$$\left\{\begin{array}{l}\Phi_t \textit{ quasi-linear } \forall t \geq 0 \textit{ satisfying Axioms 1–5} \\ \textit{ and } \Psi \textit{ is a local differential operator}\end{array}\right.$$

$$\Rightarrow$$

$$\left\{\begin{array}{ll} R{=}R_L \Rightarrow & \Psi \textit{ is linear and for } W(\cdot,t) \in \mathcal{D}_L(\Psi) \textit{ one has} \\ & \Psi W = \Delta_{\mathcal{G}}^{\alpha}(W) = (div_{\mathcal{G}} \circ \nabla_{\mathcal{G}})^{\alpha}(W) \\ & \textit{ with } \alpha \in \mathbb{N} \textit{ yielding an} \\ & \textit{homogeneous generator} \\ & \Psi \textit{commuting with all } (\varphi_Q)_*, \\ R{=}T_{\pm} \Rightarrow & \textit{ If } \Psi W = \pm H_{\mathcal{G}}(\mathrm{d}W) \textit{ for a convex, superlinear} \\ & \textit{ } H_{\mathcal{G}} : T^*(G) \to \mathbb{R}^+, \textit{ and } W(\cdot,t) \in \mathcal{D}_{T^{\pm}}(\Psi) \\ & \textit{then} \\ & \textit{(Hamiltonian) } H_{\mathcal{G}} \textit{ must be left-invariant} \\ & \textit{and } \alpha\textit{-homogeneous in the dual norm:} \\ & \mathcal{H}_{\mathcal{G}}(p) = \frac{1}{\alpha}\|p\|_*^{\alpha} = \frac{1}{\alpha}\|p\|_{\mathcal{G}^{-1}}^{\alpha} \textit{yielding a} \\ & \textit{generator } \Psi \textit{ commuting with all } (\varphi_Q)_*, \end{array}\right. \tag{5.13}$$

Proof. We first prove the case $R = R_T$. Given that Φ is linear, it immediately follows that Ψ is also linear and therefore, Ψ is a local linear operator on $\mathcal{H}_G$. But then by Peetre's theorem, we get that Ψ is a differential operator, but then by Lemma 5.4 and Corollary 5.1, we get that $\Psi = \Delta^{\alpha}$ for some $\alpha \in \mathbb{N}$.

Now for the case $R = T_+$. By the equivariance axiom 2 and the local isotropy axiom 5, we have that

$$H_{\mathcal{G}}(\mathrm{d}W) = H^{1D}(||\mathrm{d}W||_{\mathcal{G}^{-1}}) = H^{1D}(\|\nabla_{\mathcal{G}} W\|), \tag{5.14}$$

where the second equality follows from the Riesz representation theorem and the fact that the gradient is the Riesz representative of the derivative.

Then as $H_{\mathcal{G}}$ was assumed to be convex, super-linear, and the semi-group property and quasi-linearity holds we know from general results on weak-QAM theory Fahti [FM07] and Balogh et al. [Bal+12] that

$$\begin{array}{l}\Phi_t f = k_t \circledast f \text{ and} \\ k_t(g) = \kappa_t^{1D}(d_{\mathcal{G}}(g,e)) = t\kappa_1^{1D}(d_{\mathcal{G}}(g,e)/t).\end{array} \tag{5.15}$$

Furthermore one has by Lemma A.6 in Appendix A.3

$$\begin{array}{rcl} k_{t+s}(g) & = & (k_t \circledast_R^G k_s)(g) = \\ \kappa_{t+s}(d_{\mathcal{G}}(g,e)) & = & (\kappa_t^{1D} *_R^{\mathbb{R}} \kappa_s^{1D})(d_{\mathcal{G}}(g,e)) \end{array} \tag{5.16}$$

for all $t, s \geq 0, g \in G$. By Item 4 we have

$$\Phi_t^{\mathcal{G}} = \Phi_1^{\psi(t)\mathcal{G}}$$

which via (5.15) gives

$$\kappa_1(\psi(t) d_{\mathcal{G}}(g,e)) = \kappa_t(d_{\mathcal{G}}(g,e)), \tag{5.17}$$

so then (5.16) and (5.17) with the same technique (via the Fenchel transform on the $\mathbb{R}^d$-setting, but now for

$d = 1$) as in Lemma 4.7 we get

$$\begin{aligned} &\psi(t) = t^{\frac{1}{\alpha}} \text{ and } \kappa_{1D}(x) = \tfrac{t}{\beta}\left(\tfrac{x}{t}\right)^{\beta}, \text{ with } {}^{1}\!/\!_{\alpha} + {}^{1}\!/\!_{\beta} = 1, \\ &\text{and associated 1D Hamiltonian} \\ &H^{1D}(p) = \tfrac{1}{\alpha}|p|^{\alpha}. \end{aligned} \tag{5.18}$$

Thereby the result follows from (5.18) and (5.14):

$$H_{\mathcal{G}^{-1}}(\mathrm{d}W) = H^{1D}(||\mathrm{d}W||_{\mathcal{G}^{-1}}) = \frac{1}{\alpha}||\mathrm{d}W||^{\alpha}_{\mathcal{G}^{-1}}$$

and the result follows

□

Remark 5.6. *In the linear semiring case $R = R_L$ the first implication in Theorem 5.1 could be replaced by an equivalence, as then the reverse implication '$\Leftarrow$' is true in view of [Bel+23, Prop.1], [Pai+23] (for $R = T_{\pm}$) and Lemma 5.3 and the fact that non-integer fractional powers of $|\Delta_{\mathcal{G}}|$ involve non-local operators [Yos68].*

As the axiomatic approach for PDE-G-CNNs on $G = \mathbb{R}^2$ primarily went via the semiring Fourier transform, we must investigate semiring Fourier transforms (linear and tropical) on the non-commutative group $G = SE(2)$.

This will be the topic of the next two paragraphs.

5.3 The Linear Fourier Transform on $SE(2)$

For $R = L$ the Fourier transform on $G = SE(2)$ is the usual linear Fourier transform given by (5.20).

By Mackey's imprimitivity theorem [Mac76], the dual group $\hat{G}$ is isomorphic to the set of compact dual orbits

$$\mathcal{A}_p^J = \{A^T\omega \mid A \in SO(J)\} = \{A\omega \mid A \in SO(J^{-T})\}$$

where $p = \sqrt{\omega J^{-T}\omega} \in \mathbb{R}$. Consequently, as commonly done in abstract harmonic analysis [Fol95; FF05], we can identify the irreducible representation $\sigma^p \in \hat{G}$ with $p \in \mathbb{R}^+$:

$$p \leftrightarrow \sigma^p. \tag{5.19}$$

Now in view of Theorem 2.2, the dual measure is given by $d\nu_G(\sigma) = pdp$. Using the identification (5.19) and writing $\sigma^p(g) = \sigma^p_g$ the Fourier transform becomes

$$\mathcal{F}U(p) = \int_G U(g)\, \sigma^p_{g^{-1}}\, \mathrm{d}g, \text{ with } p \in \hat{G}, \tag{5.20}$$

where $\hat{G}$ is the dual of group of G consisting of irreducible representations of G. Now, as $SE(2)$ is type I, we have by Theorem 2.2 that the inverse Fourier transform is given by

$$U(g) = \int_{\hat{G}} \mathrm{trace}\{\sigma^p_g \circ \mathcal{F}U(p)\}\, \mathrm{d}\nu(p) \tag{5.21}$$

with dual measure $\nu(p)$.

All UIRs (up to unitary equivalence) of $SE(2)$ are indexed by $p \in \hat{G} = \mathbb{R}^+$ and given by

$$(\sigma_g^p \phi)(\mathbf{n}) = e^{ip\boldsymbol{x}\cdot\mathbf{n}}\phi(R_\theta^{-1}\mathbf{n}), \qquad g = (\boldsymbol{x}, \theta) \tag{5.22}$$

for all $\phi \in L^2(S^1), g = (\boldsymbol{x}, R_\theta)$. This is a well-known result [Sug90],[CK01, ch: 10.2], but the next remark gives a quick view on where these UIRs originate from and how they relate 1-to-1 to the dual orbits of $SO(2)$ acting on dual spatial group $\hat{\mathbb{R}}^2$.

Remark 5.7. *These representations σ^p are unitary equivalent to the representations $\tilde{\sigma}^{\rho=p}$ that arise by direct integral decompositions [FF05] of the left-regular unitary representation $\mathcal{U} : G \to L^2(\mathbb{R}^2)$ in the spatial Fourier domain. Here*

$$(\mathcal{U}_g\psi)(\mathbf{v}) = \psi(R_\theta^{-1}(\mathbf{v} - \boldsymbol{x})),$$

and $g = (\boldsymbol{x}, R_\theta) \in G = SE(2)$, $\psi \in L^2(\mathbb{R}^2)$. Indeed

$$\mathcal{F} \circ \mathcal{U}_g \circ \mathcal{F}^{-1} = \int_{\mathbb{R}^+}^{\oplus} \tilde{\sigma}^p p\mathrm{d}p$$

where $\mathcal{F} = \mathcal{F}_L^{\mathbb{R}^2}$ where $\tilde{\sigma}^\rho : G \to L^2(S_\rho^1)$ is given by

$$\tilde{\sigma}_g^\rho F(\rho\mathbf{n}(\varphi)) = e^{i(\rho\mathbf{n}(\varphi))\cdot\boldsymbol{x}}\, F(\rho\mathbf{n}(\varphi - \theta))),$$

for all $g = (\boldsymbol{x}, \theta) \in G$, $\omega = \rho\mathbf{n}(\varphi) \in \hat{\mathbb{R}}^2$, $F \in L^2(S_\rho^1)$.

Remark 5.8. *The unitary equivalence mentioned in the previous remark follows by scaling. Indeed $\tilde{\sigma}_g^\rho = \mathcal{D}_\rho \circ \sigma_g^\rho \circ \mathcal{D}_\rho^{-1}$ with $\mathcal{D}_\rho : L^2(S_1^1) \to L^2(S_\rho^1)$ given by $\mathcal{D}_\rho\phi(\mathbf{v}) = \rho^{-\frac{1}{2}}\,\phi(\rho^{-1}\mathbf{v})$.*

So we have the linear Fourier transform on $SE(2)$:

$$\mathcal{F}_L^G U(p) = \int_G U(g)\,\sigma_{g^{-1}}^p\,\mathrm{d}g, \text{ with } p \in \mathbb{R}^+, \tag{5.23}$$

and its inverse Fourier transform is given by:

$$U(g) = \int_{\mathbb{R}^+} \mathrm{trace}\{\sigma_g^p \circ \mathcal{F}_L^G U(p)\}\,p\mathrm{d}p. \tag{5.24}$$

The linear Fourier transform on $G = SE(2)$ has a fundamental property that allows us to solve convection-diffusions and fractional diffusions on $SE(2)$. It intertwines the right-regular representations (and their generators: the left-invariant vector fields) with irreducible representations (and their generators):

Lemma 5.5. *Set $G = SE(2)$. For all $f \in L^2(G)$, $g \in G$, $p \in \mathbb{R}^+ \equiv \hat{G}$ and all $A \in T_e(G)$ we have:*

$$\begin{aligned}(\mathcal{F} \circ \mathcal{R}_g f)(p) &= \sigma_g^p \circ (\mathcal{F}f)(p),\\ \mathcal{F} \circ \mathrm{d}\mathcal{R}(A) &= \mathrm{d}\sigma^p(A) \circ \mathcal{F}\end{aligned}$$

where in the 2nd row the unbounded operators have a domain (dense in $L^2(G)$) consisting of all $f \in L^2(G)$ s.t. the limit

$$\mathrm{d}\mathcal{R}(A)f = \lim_{t \downarrow 0} \frac{\mathcal{R}_{e^{tA}} f - f}{t} \text{ exists.}$$

Proof. The second identity follows by the first identity and the continuity of unitary operator $\mathcal{F}$. The first identity is simple is due to the fact G is unimodular: right Haar measure is left Haar measure:

$$\begin{aligned}(\mathcal{F}_L^G \mathcal{R}_g U)(p) &= \int_G U(gh)\, \sigma^p_{g^{-1}}\, \mathrm{d}g = \int_G U(q)\, \sigma^p_{hq^{-1}}\, \mathrm{d}q \\ &= \int_G U(q)\, \sigma^p_h \circ \sigma^p_{q^{-1}}\, \mathrm{d}q = \sigma^p_h \circ \mathcal{F}_L^G U(p)\end{aligned}$$

and the result follows. □

Corollary 5.2. *The previous lemma applies in particular to the basis $\{\mathcal{A}_i = \mathrm{d}\mathcal{R}(A_i)\}_{i=1}^3$ given in (2.2) and the corresponding operators $\{\mathrm{d}\sigma^p(A_i)\}_{i=1}^3$ in the Fourier domain are*

$$\begin{aligned}\mathrm{d}\sigma^p(A_1) &= ip\cos(\psi), \\ \mathrm{d}\sigma^p(A_2) &= ip\sin(\psi), \\ \mathrm{d}\sigma^p(A_3) &= \partial_\psi\end{aligned} \tag{5.25}$$

and they obey the same commutator relations as both $\mathrm{d}\mathcal{R}$ and $\mathrm{d}\sigma^p$ are Lie algebra isomorphisms.

Proof. We will prove $A_i = A_1$ as the rest follow similarly. As we have that $A_1 = \partial_x|_e$ and $e = ((0,0),0)$ we have that $e^{tA_1} = ((t,0),0)$. Now using (5.22), we get that

$$(\sigma^p_{e^{tA_1}}\phi)(\mathbf{n}(\psi)) = e^{ipt(1,0)\cdot(\cos(\psi),\sin(\psi))}\phi(\mathbf{n})$$

Thus using the definition, we have

$$\begin{aligned}\mathrm{d}\sigma^p(A_1)(\phi)(\mathbf{n}) &= \lim_{t\to 0}\left(\frac{e^{ipt\cos(\psi)}\phi(\mathbf{n}) - \phi(\mathbf{n})}{t}\right) \\ &= ip\cos(\psi)\phi(\mathbf{n})\end{aligned}$$

□

Corollary 5.3. *The previous corollary shows that both convection-diffusions and fractional diffusions allow for exact solutions via spectral decompositions of their kernels in the Fourier domain. Indeed, their linear PDE generators (5.5) satisfy:*

$$\Psi := \Delta_{\mathcal{G}} = \sum_{i,j=1}^3 g^{ij}\mathcal{A}_i\mathcal{A}_j \iff$$

$$\hat{\Psi}_p := (\mathcal{F}_L^G \circ \Psi \circ (\mathcal{F}_L^G)^{-1})(\sigma_p) = \sum_{i,j=1}^3 g^{ij}\mathrm{d}\sigma^p(A_i)\mathrm{d}\sigma^p(A_j)$$

and the right-hand side is a self-adjoint Sturm-Liouville operator of Mathieu type with periodic boundary conditions and allows for expansion in a complete ONB $\{\phi^p_k\}_{k\in\mathbb{N}}$. For explicit formulas see [DF10b; Agr+08] *summarized in [Zha+16]. One has (with normalised ϕ_k^p's):*

$$\Phi_t = e^{-t|\Psi|^{\frac{\alpha}{2}}} \iff \tag{5.26}$$

$$\hat{\Phi}_t^p = e^{-t|\hat{\Psi}_p|^{\alpha}} = \sum_{k\in\mathbb{Z}} e^{-(-\lambda_k^{\mathcal{G}}(p))^{\frac{\alpha}{2}} t} \phi_k^p \otimes \phi_k^p \tag{5.27}$$

Proof. Apply Lemma 5.5 to generator Ψ from Lemma 5.4, and realize that the Laplacian Δ_G is as a Laplace-Beltrami operator expressed in the left-invariant frame by (5.9). For the expansions into the Mathieu functions ϕ_n, with negative eigenvalues related to the Mathieu characteristics, see [DF10a; Agr+08]. □

Corollary 5.4. *The kernels of the left-invariant diffusions on $G = SE(2)$ are given by*

$$\hat{k}_t(\sigma_p) = e^{-t|\hat{\Psi}_p^{\mathcal{G}}|^{\frac{\alpha}{2}}} \tag{5.28}$$

with isotropic generator

$$|\hat{\Psi}_p^{\mathcal{G}}|^{\alpha} = -\left(-\sum_{i,j=1}^{3} g^{ij} \mathrm{d}\sigma^p(A_i) \otimes \mathrm{d}\sigma^p(A_j)\right)^{\frac{\alpha}{2}}$$

with $\alpha > 0$. This gives the following expressions for the diffusion kernels on $G = SE(2)$:

$$k_t(g) = \int_0^\infty \sum_{k\in\mathbb{Z}} e^{-t|\lambda_k^{\mathcal{G}}(p)|^{\frac{\alpha}{2}}} (\sigma_g^p)_{k,k}\, p\mathrm{d}p, \tag{5.29}$$

where $(\sigma_g^p)_{k,k} = \overline{(\sigma_g^p)_{-k,-k}} \in \mathbb{C}$ are pre-computable complex numbers and given by $(\sigma_g^p)_{kk} = \left(\sigma_g^p \phi_k^p\ ,\ \phi_k^p\right)_{L^2(S^1)}$.

Proof. We have that

$$k_t(g) = (\mathcal{F}_G^{-1}\hat{k}_t)(g) = \int_{\mathbb{R}^+} \mathrm{tr}(\sigma_g^p \circ \hat{k}_t(\sigma^p)) p dp$$

Now, we have

$$\begin{aligned} \mathrm{tr}(\sigma_g^p \circ \hat{k}_t(\sigma^p)) p &= \sum_{m\in\mathbb{Z}} (\sigma_g^p \circ \hat{k}_t(\sigma^p))_{mm} \\ &= \sum_{m\in\mathbb{Z}} e^{-t|\lambda_m(p)|^{\alpha/2}} (\sigma_g^p)_{mm} \end{aligned}$$

and we get the result. □

Remark 5.9. *The series (5.29) is not converging very fast for very small $t > 0$. It has been used though for numerics [Zha+16], and there is a technique of unfolding the circle and periodise afterwards that provides a rapidly decaying series that can even be computed by the Floquet theorem [DV08; DF10c]. However, in PDE-G-CNNs we need fast computations on the GPU, and rather rely on quick Gaussian approximations [RDu+21].*

Lemma 5.6. *Suppose we have a spatially isotropic metric tensor field*

$$[\mathcal{G}] = diag(g_{11}, g_{11}, g_{33}) \iff \mathcal{G} = \sum_{i=1}^{3} g_{ii}\, \omega^i \otimes \omega^i,$$

with corresponding linear generator (a fractional Laplacian)

$$-|\Delta_{\mathcal{G}}|^{\frac{\alpha}{2}} = -|g^{11}(\partial_x^2+\partial_y^2)+g^{33}\partial_\theta^2|^{\frac{\alpha}{2}}, \qquad g^{ii}=g_{ii}^{-1},$$

then the corresponding kernel left-invariant evolution kernel $k_t^\alpha = e^{-t|\Delta_{\mathcal{G}}|^{\frac{\alpha}{2}}}\delta_e$ *is no longer positive if* $\alpha>2$.

Proof. The fact that $\Delta_{\mathcal{G}} = g^{11}(\partial_x^2+\partial_y^2)+g^{33}\partial_\theta^2$ implies that in the Fourier domain, using (5.25),

$$\begin{aligned}\hat{\Delta}_{\mathcal{G}} &= g^{11}(\mathrm{d}\sigma^p(A_1)^2+\mathrm{d}\sigma^p(A_2)^2)+g^{33}\mathrm{d}\sigma^p(A_3)^2\\ &= -g^{11}p^2(\cos^2(\psi)+\sin^2(\psi))+g^{33}\partial_\psi^2\\ &= -g^{11}p^2+g^{33}\partial_\psi^2\ .\end{aligned}$$

Now let $\phi_k\in L^2(S^1)$ be orthonormal eigenfunctions of $-|\hat{\Delta}_{\mathcal{G}}|^{\frac{\alpha}{2}}$ with corresponding eigenvalues $-|\lambda_k(p)|^{\frac{\alpha}{2}}$. Then by [Agr+08, Corollary.28] we have that the corresponding linear kernel is given by

$$k_t^\alpha(g) = \int_0^\infty \sum_{k\in\mathbb{Z}} e^{-t|\lambda_k^{\mathcal{G}}(p)|^{\frac{\alpha}{2}}}\,(\sigma_g^p\phi_k,\phi_k)\ p\mathrm{d}p$$

Now we note that the 2π-periodic orthonormal eigenfunctions of operator $\hat{\Delta}_{\mathcal{G}}$ take (due to the isotropy constraint $g_{11}=g_{22}$) a rather simple form:

$$\phi_k(\psi)=\frac{1}{\sqrt{2\pi}}e^{ik\psi}, \qquad k\in\mathbb{Z},$$

where the corresponding eigenvalues are

$$\lambda_k^{\mathcal{G}}(p) = g^{33}\lambda_k - g^{11}p^2 = -g^{33}k^2-g^{11}p^2$$

where $\lambda_k=-k^2$ gives the discrete spectrum of differential operator $\frac{d^2}{d\psi^2}$ with 2π periodic boundary conditions. We note that the functions ϕ_k actually do not depend on p while the eigenvalues $\lambda_k(p)$ do. But then we have for $g=(x,y,\theta)$

$$(\sigma_g^p\phi)(\psi) = e^{ip(x\sin(\psi)+y\cos(\psi))}\phi(\psi-\theta)$$

But then we get (only) for $\alpha>2$ that

$$0 = \partial_p^2 e^{-t(g^{11})^{\alpha/2}|p|^\alpha}\Big|_{p=0} = \partial_p^2 e^{-t|\lambda_0^{\mathcal{G}}(p)|^{\frac{\alpha}{2}}}\Big|_{p=0} \tag{5.30}$$

$$= \partial_p^2(\phi_0,\hat{k}_t^\alpha(\sigma_p)\phi_0)\Big|_{p=0} \tag{5.31}$$

$$\Longleftrightarrow \tag{5.32}$$

$$0 = \int_G k_t^\alpha(g)\ \ \partial_p^2(\sigma_{g^{-1}}^p(\phi_0),\phi_0)\Big|_{p=0}\ \mathrm{d}g \tag{5.33}$$

Now, we have $\phi_0=\frac{1}{\sqrt{2\pi}}$, $\lambda_0=0$, $\lambda_0^{\mathcal{G}}(p)=-g^{11}p^2$ and

$$\begin{aligned}\partial_p^2(\sigma_{g^{-1}}^p(\phi_0),\phi_0)\Big|_{p=0} &= (\partial_p^2\sigma_{g^{-1}}^p(\phi_0),\phi_0)\Big|_{p=0}\\ &= -((x\sin(\cdot)+y\cos(\cdot))^2\phi_0(\cdot+\phi),\phi_0(\cdot))\\ &= -\frac{1}{2\pi}\int_0^{2\pi}|x\cos\psi+y\sin\psi|^2\mathrm{d}\psi\\ &= -\frac{1}{2}(x^2+y^2)\le 0.\end{aligned}$$

With respect to the second identity we note that $g=(\mathbf{x},\theta) \iff g^{-1}=(R_\theta^{-1}\mathbf{x},-\theta)$ and thereby the square spatial norm $\|\mathbf{x}\|^2=x^2+y^2$ is invariant under group inversion.

We finally conclude that by (5.30) the kernel must have zero crossings for $\alpha>2$ as the total integrand cannot have the same sign to produce a zero value. □

Next we generalize the previous lemma to the general left-invariant (diagonal) Riemannian case. We also include a sub-Riemannian extension (where either g^{11} or g^{22} is zero).

Theorem 5.2. *Let $\infty \geq g_{11} \geq 0$, $\infty \geq g_{22} \geq 0$, $g_{33}>0$, $g^{11}+g^{22}>0$ to ensures that the Hörmander condition is satisfied.*

Suppose we have a diagonal metric tensor field

$$[\mathcal{G}] = diag(g_{11}, g_{22}, g_{33}) \iff \mathcal{G} = \sum_{i=1}^{3} g_{ii}\, \omega^i \otimes \omega^i,$$

with corresponding linear generator (again $g^{ii}=g_{ii}^{-1}$):

$$-|\Delta_{\mathcal{G}}|^{\frac{\alpha}{2}} = -|g^{11}\mathcal{A}_1^2 + g^{22}\mathcal{A}_2^2 + g^{33}\mathcal{A}_3^2|^{\frac{\alpha}{2}},$$

with left-invariant vector fields $\{\mathcal{A}_i\}_{i=1}^3$ given by (2.2).

Then the corresponding left-invariant evolution kernel $k_t^\alpha = e^{-t|\Delta_{\mathcal{G}}|^{\frac{\alpha}{2}}}\delta_e$ is no longer positive if $\alpha>2$.

Proof. We follow the same approach as the previous lemma. To arrive at a contradiction, we assume positivity of the kernel and we will first show that if $\alpha>2$,

$$\partial_p^2 \left(\phi_0^p \,,\, \hat{k}_t^\alpha(\phi_0^p)\right)_{L^2}\Big|_{p=0} = 0 \tag{5.34}$$

where ϕ_0^p is the zeroth eigenfunction of the operator $-|\hat{\Delta}_{\mathcal{G}}|^{\frac{\alpha}{2}}$. We have from (5.26) that

$$\hat{k}_t^\alpha = e^{-t|\hat{\Psi}_p|^\alpha} = \sum_{k\in\mathbb{Z}} e^{-(-\lambda_k^{\mathcal{G}}(p))^{\frac{\alpha}{2}} t} \phi_k^p \otimes \phi_k^p$$

Now from orthonormality of the eigenfunctions we immediately see that

$$\left(\phi_0^p \,,\, \hat{k}_t^\alpha(\phi_0^p)\right)_{L^2} = e^{-(-\lambda_0^{\mathcal{G}}(p))^{\frac{\alpha}{2}} t}.$$

Now in order to show that

$$\partial_p^2 \left(e^{-(-\lambda_0^{\mathcal{G}}(p))^{\frac{\alpha}{2}} t}\right)\Big|_{p=0} = 0,$$

Since for an arbitrary twice continuously differentiable f, we have

$$\partial_p^2 e^{f(p)} = f''(p)e^{f(p)} + (f'(p))^2 e^{f(p)}.$$

it suffices to show that $\partial_p(-(-\lambda_0^{\mathcal{G}}(p))^{\frac{\alpha}{2}}t)|_{p=0}=0$ and $\partial_p^2(-(-\lambda_0^{\mathcal{G}}(p))^{\frac{\alpha}{2}}t)|_{p=0}=0$.

Now we have using (5.25),

$$\begin{aligned}
\hat{\Delta}_{\mathcal{G}} &= \partial_\psi^2 - p^2 \tfrac{1}{g^{33}}(g^{11}\cos^2(\psi) + g^{22}\sin^2(\psi)) \\
&= \partial_\psi^2 - p^2 \tfrac{1}{g^{33}}(g^{11}(\tfrac{1}{2}+\tfrac{1}{2}\cos^2(2\psi)) + g^{22}(\tfrac{1}{2}-\tfrac{1}{2}\cos(2\psi))) \\
&= \partial_\psi^2 - \tfrac{1}{2}\left(\tfrac{g^{11}+g^{22}}{g^{33}}\right)p^2 - \tfrac{1}{2}\left(\tfrac{g^{11}-g^{22}}{g^{33}}\right)p^2\cos(2\psi).
\end{aligned}$$

Thereby the corresponding eigenvalues of the orthonormal eigenfunctions ϕ_n^p of operator $\hat{\Delta}_{\mathcal{G}}$ are given by

$$\lambda_n^{\mathcal{G}}(p) = \pm a_n(q) + \tfrac{1}{2}\left(\tfrac{g^{22}+g^{11}}{g^{33}}\right)p^2$$

where $q = \frac{1}{4}\left(\frac{g^{22}-g^{11}}{g^{33}}\right)p^2$ and $a_n(q)$'s are the eigenvalues corresponding to the operator $\partial_\psi^2 - 2q\cos(2\psi)$ which are precisely the characteristic values of the Mathieu equation. Now the series expansion of $a_0(q)$ is given by [AS64]

$$a_0(q) = -\frac{q^2}{2} + O(q^4).$$

From this, it is evident that $\partial_p a_0(q)|_{p=0} = \partial_p^2 a_0(q)|_{p=0} = 0$ and hence $\partial_p \lambda_n^{\mathcal{G}}(p)|_{p=0} = \partial_p^2 \lambda_n^{\mathcal{G}}(p)|_{p=0} = 0$. But then this implies that $\partial_p(-(-\lambda_0^{\mathcal{G}}(p))^{\frac{\alpha}{2}}t)|_{p=0}$ and $\partial_p^2(-(-\lambda_0^{\mathcal{G}}(p))^{\frac{\alpha}{2}}t)|_{p=0}$ both equal 0 if $\alpha > 2$ which implies (5.34) was to be shown.

Now, on the other hand, we have by definition,

$$\hat{k}_t^\alpha(p)\phi_0^p = \int_G k_t^\alpha(g)\sigma_{g^{-1}}^p(\phi_0^p)\,\mathrm{d}g$$

and as a vector-valued integral, we get

$$\left(\phi_0^p\ ,\ \hat{k}_t^\alpha(\phi_0^p)\right)_{L^2} = \int_G k_t^\alpha(g)\left(\phi_0^p\ ,\ \sigma_{g^{-1}}^p(\phi_0^p)\right)_{L^2}\ \mathrm{d}g.$$

Now, we claim that

$$\begin{aligned}
&\partial_p^2\left(\phi_0^p\ ,\ \sigma_{g^{-1}}^p(\phi_0^p)\right)_{L^2}\Big|_{p=0} = \partial_p^2\left(\phi_0^p\ ,\ \sigma_g^p(\phi_0^p)\right)_{L^2}\Big|_{p=0}\\
&= \int_{S^1}\partial_p^2\left(e^{ip(x\sin(\psi)+y\cos(\psi))}\phi_0^p(\psi)\overline{\phi_0^p}(\psi-\theta)\right)\Big|_{p=0}\mathrm{d}\psi\\
&\neq 0.
\end{aligned}$$

We have that the power series expansion of ϕ_0^p in p is given by [AS64]

$$\phi_0^p(\psi) = \overline{\phi_0^p}(\psi) = \frac{1}{\sqrt{2\pi}}(1 - \frac{q}{2}\cos(2\psi) + O(q^2))$$

Now using this, we compute the integrand:

$$\begin{aligned}
&\partial_p^2\left(e^{ip(x\sin(\psi)+y\cos(\psi))}\overline{\phi_0^p}(\psi)\phi_0^p(\psi-\theta)\right)\Big|_{p=0}\\
&= \frac{1}{8\pi}\left(-4C\cos(2(\theta-\psi)) - 4C\cos(2\psi) - 4\sin^2(\psi)(x+y)^2\right)
\end{aligned}$$

where $C = \frac{1}{4}\left(\frac{g^{22}-g^{11}}{g^{33}}\right)$. Now, integrating this over S^1 yields

$$\begin{aligned}
&\int_{S^1}\partial_p^2\left(e^{ip(x\sin(\psi)+y\cos(\psi))}\overline{\phi_0^p}(\psi)\phi_0^p(\psi-\theta)\right)\Big|_{p=0}\mathrm{d}\psi\\
&= -\frac{1}{2}(x^2+y^2)
\end{aligned}$$

which is strictly non-positive. Now, by

$$0 = \partial_p^2 \int_G k_t^\alpha(g) \left(\phi_0^p \, , \, \sigma_{g^{-1}}^p(\phi_0^p) \right)_{L^2} \mathrm{d}g \Big|_{p=0}$$
$$= - \int_G \frac{1}{2} k_t^\alpha(g)(x+y)^2 \, \mathrm{d}g$$

we see that the kernel cannot be strictly positive.

□

Theorem 5.3. *Set $R = L$, $G = SE(2)$.*
Then the PDE-CNN Axioms (1,2,3,4,5) force solutions

$$\Phi_t f = k_t^{L,\alpha} *_G^L f, \text{ with } \alpha > 0$$

where the possibly complex-valued kernels satisfy

$$k_t^{L,\alpha} \circledast k_s^{L,\alpha} = k_{s+t}^{L,\alpha} \text{ for all } s,t > 0$$

and where their Fourier transforms $\hat{k}_t^{L,\alpha}(p)$ are given by (5.28). Moreover, if the metric tensor field is diagonal w.r.t. left-invariant frame, then the kernels are positive for $0 < \alpha \leq 2$.

Proof. By (5.13), axioms 1,2,3,4,5 implies forces the generator $\Psi = \nabla_{\mathcal{G}}^{\alpha/2}$ and Corollary 5.3 and Corollary 5.4 implies that

$$\Phi_t f = k_t^\alpha * f.$$

But then this combined with semi-group property, implies that

$$k_t^\alpha * k_s^\alpha = k_{t+s}^\alpha$$

which completes the result. The positivity follows from the previous theorem. □

Chapter 6

Experiment PDE-CNN on $\mathbb{R}^2$: Vessel Segmentation

For this experiment project, we will investigate if the PDE-CNNs have similar benefits as PDE G-CNNs when it comes to training data reduction. It was already reported by [Cas21] that PDE-CNNs yield a reduction in network complexity compared to CNNs with similar performance.

6.1 Experiment: Vessel Segmentation

In this section, we compare the performance of PDE-CNNs with CNNs on vessel segmentation tasks. In this experiment, all the models used exclude diffusion as it was found by Castella [Cas21] that diffusion usually leads to weaker performance. As diffusion is excluded, each layer in the network is comprised of convection, erosion and dilation. For the PDE-CNNS, the models are divided into two types. First is, the GEN model where each layer consists of the same modules which are any combination of erosion, dilation and convection. The second is the DE(diffusion erosion) model where the first and the last layer is the convection dilation module and the layers between them can be any combination that always includes convection. Unlike the GEN model, these middle layers need not be the same. To describe the specific structures of the network, a binary notation is employed to define the three layers that vary. For a single layer, a binary number of length 2 is utilized to indicate whether dilation and/or erosion are utilized. The first digit represents dilation, while the second digit represents erosion. Therefore, a binary number of 10 implies that dilation is used, but erosion is not. The structure of the three layers is defined by a binary number of length 6 in this manner. With a binary value of 101101, the second layer employs dilation, the third layer employs dilation and erosion, and finally, the fourth layer uses erosion. In the paper [Cas21], it was found that combination 111011 on average was the best-performing model and as a result, we will use this model for our comparison with CNNs along with one basic GEN model without diffusion. It was also found that for models with comparable architectures namely, a comparable number of layers and channels, PDE-CNN has significantly lower network complexity than the CNN counterpart. In the first part of this section, we confirm this finding and assess the extent to which the decrease in network complexity decreases the performance. We then explore the consequences of a reduction in training data. We then conclude the section with some quantitative comparisons by plotting images which indicate the true positives (green), false positives (red), true negatives (white), and false negatives (blue). For the testing of the network performance, we use the DICE score.

6.2 Network Complexity Reduction

Model	Layers	Channels	Parameters	Dice score
CNN	6	14	14946	0.80160632
PDE-CNN	6	30	4998	0.80147527
CNN	9	10	16818	0.80477967
PDE-CNN	9	24	5922	0.80562725
CNN	12	18	76242	0.811019685
PDE-CNN	12	34	15318	0.812743681

Table 6.1: performance comparison for 6, 9 and 12 layered PDE-CNNs and CNNs.

In this section, we investigate the reduction in the network complexity for PDE-CNNs compared to CNNs for respective architectures with 6 and 9 layers. Namely, Given 6 layered PDE-CNN with 24 channels and 9 layered PDE-CNN with 20 channels both with the DE configuration 111011, we investigate how many channels are required for CNNs with the same number of layers to perform equally. The DICE score of CNN is plotted against the number of parameters for both 6-layered CNN and 9-layered CNN respectively. From the plots from Figure 6.1, we see that PDE-CNNs have significantly lower network complexity than CNNs with comparable performance. There are 3 fold, 2.8 fold and 5 fold decrease in the respective network complexity for 6,9 and 12 layered PDE-CNNs with each having a similar performance to the CNN counterpart.

Model	6 layers	12 layers
PDE-CNN	3 fold	5 fold
PDE-G-CNN	12 fold	32 fold

Table 6.2: Comparison in the decrease in network complexity for PDE-CNN and PDE-G-CNN. While PDE-CNN has good decrease in network complexity, it was shown by Smets [Sme+21] that PDE-G-CNN has significantly more.

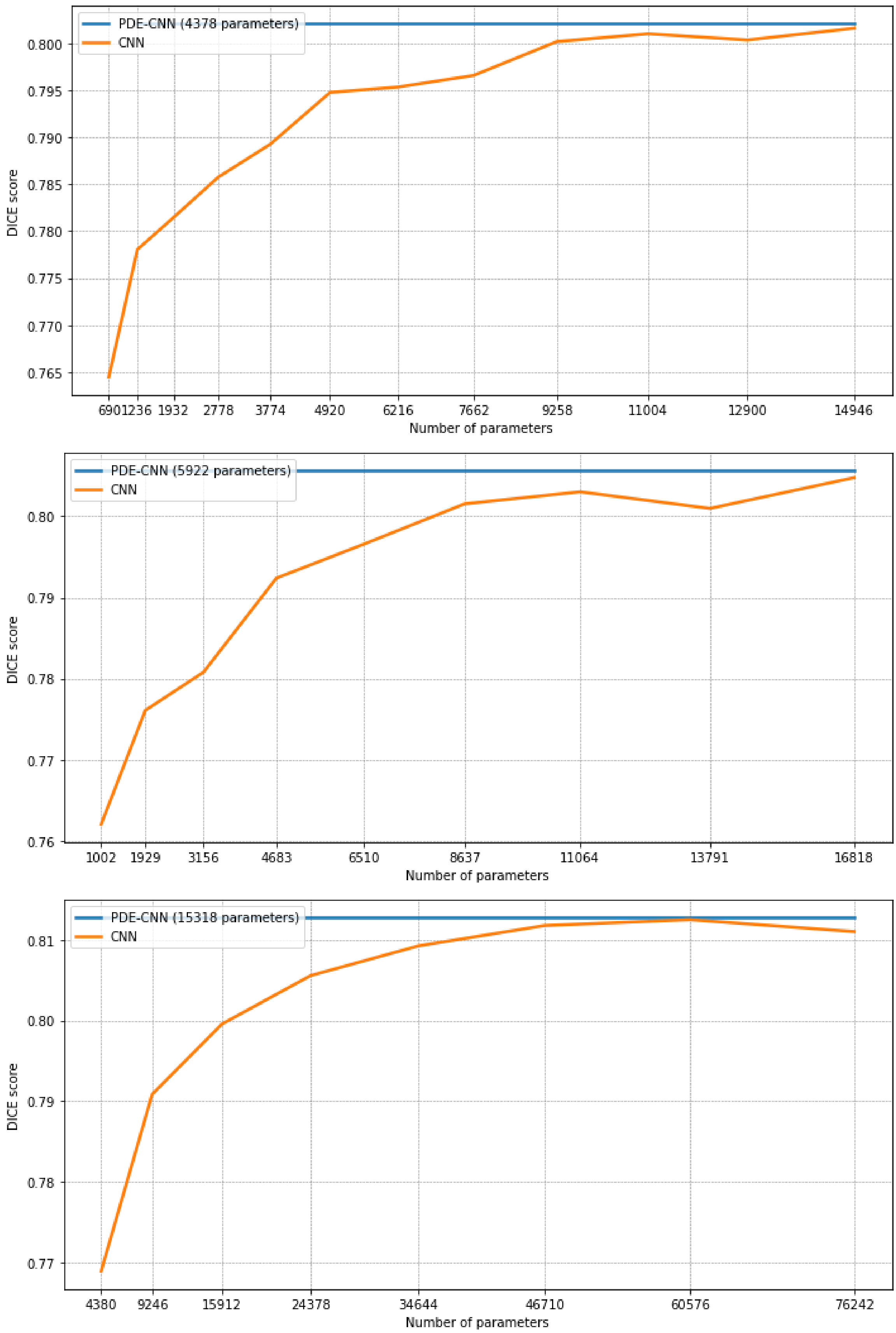

Figure 6.1: We compare the performance of a 6-layered PDE-CNN with 30 channels(top), a 9-layered PDE-CNN(bottom) with 30 channels and a 12-layered PDE-CNN(bottom) with 34 channels with CNNs of the same layers with varying channel sizes. All networks are trained on DRIVE and training data testing data consist of 20(separate) images respectively. The score is taken to be the mean of 3 experiments.

6.3 Training Data Reduction

As we have seen in the previous section, there is a significant reduction in network complexity for PDE-CNNs with CNNs with comparable performance. In this section, we will investigate the performance of PDE-CNNS compared to CNNs in terms of training data reduction.

6.3.1 Comparable Network Architecture

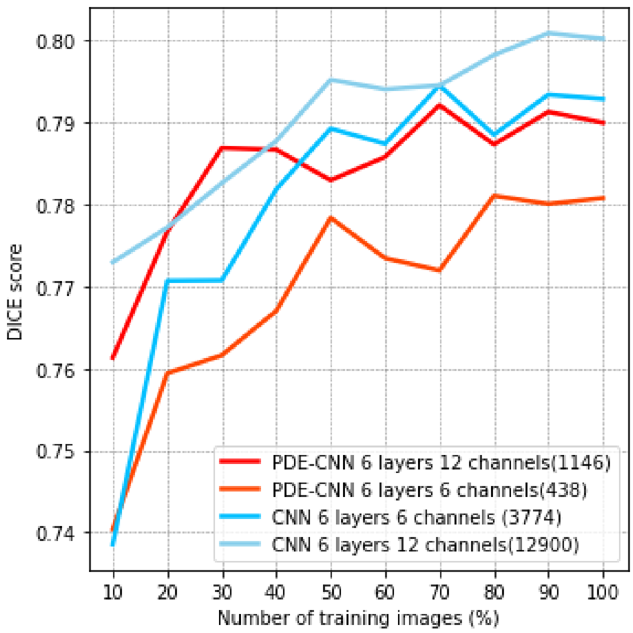

Figure 6.2: The mean dice score of PDE-CNNs compared to CNNs with the same network width and depth and the total number of parameters for each network.

First, we compare PDE-CNNs and CNNs with a comparable number of layers and channels. We namely compare the performance of a 6-layer CNN and a 6-layer PDE-CNN on two different occasions, with 6 channels and 12 channels by reducing the amount of training data. We conduct experiments on the well-known DRIVE [Sta+04] datasets for the vessel segmentation task. For default full training data set, we use a 50-50 split of 40 images for training and testing and employ overlapping patches of dimension 64 x 64 with a patch overlap of 16 for training the networks. To evaluate the impact of reduced input data on the networks' performance, we randomly shuffle the training patches in the DRIVE dataset and gradually compile 10% to 100% of the total patches, and then use this reduced data for training all the networks. The testing is done on the full testing data set of 20 images for each reduction. As expected, CNNs with higher network complexity outperforms PDE-CNN with significantly lower network complexity.

It is worth mentioning that for these network configurations, the difference between the dice score at 10% compared to 100% is smaller for PDE-CNNs.

6.3.2 Comparable Network Complexity

Now, we compare the performances of PDE-CNNs and CNNs with comparable network complexity. We compare them in two scenarios with high network complexity and low network complexity for both.

We see that for comparable number of parameters, PDE-CNNs are superior to CNNs especially for low training data. Especially for low network complexity, when PDE-CNN and CNN have the same complexity, CNN cannot even compete with PDE-CNN as the highest score it reaches at 100% of the training data is barely higher than the score of PDE-CNN with 10% training data. The plots in Figure 6.3 also indicate that the performances of PDE-CNNs with low training data approaches the performance with full training data quicker than the CNNs. We again note that the difference between the dice score at 10% compared to 100% is smaller for PDE-CNNs. The differences are summarised in Table 6.4. In Figure 6.4, Pai compared [Pai+23] the training data reduction of a CNN, a G-CNN and a PDE-G-CNN in the same fashion as our comparison of data reduction. We see that although PDE-CNN does not have a similar level of improvement as PDE-G-CNN, at least for networks with similar level of complexity, there is still a large improvement in performance with low training data. It seems according to Figure 6.4, the benefit of PDE-CNN lies especially in low training data as for PDE-G-CNN, the difference between the dice score of 10% and 100% training data is not as significant as PDE-CNNs compared to CNNs. Although PDE-G-CNN is still clearly superior to PDE-CNN as it achieves higher dice score with lower network complexity. This suggests that the improvement in PDE-CNN with more training data tapers off for PDE-CNNs as we increase the training data size.

Now to more closely investigate the effects of low training data, we look at the performance of PDE-CNN compared with CNN through 1 to 5 training images. The results are shown in Figure 6.5. One notable finding from this experiment is the variance of dice score. On average, there seems to be a less variation in performance for PDE-CNNs compared to CNNs with similar network complexity. In Figure the camprison for PDE-CNN with 6 layers and 6 channels and CNN with 6 layers and 2 channels is made.

	data = 1	data = 2	data = 3	data = 4	data = 5
PDE-CNN	0.0360	0.0210	0.0102	0.0141	0.0065
CNN	0.0880	0.03532	0.0382	0.0212	0.0275

Table 6.3: Standard deviation of dice score over 5 experiments.

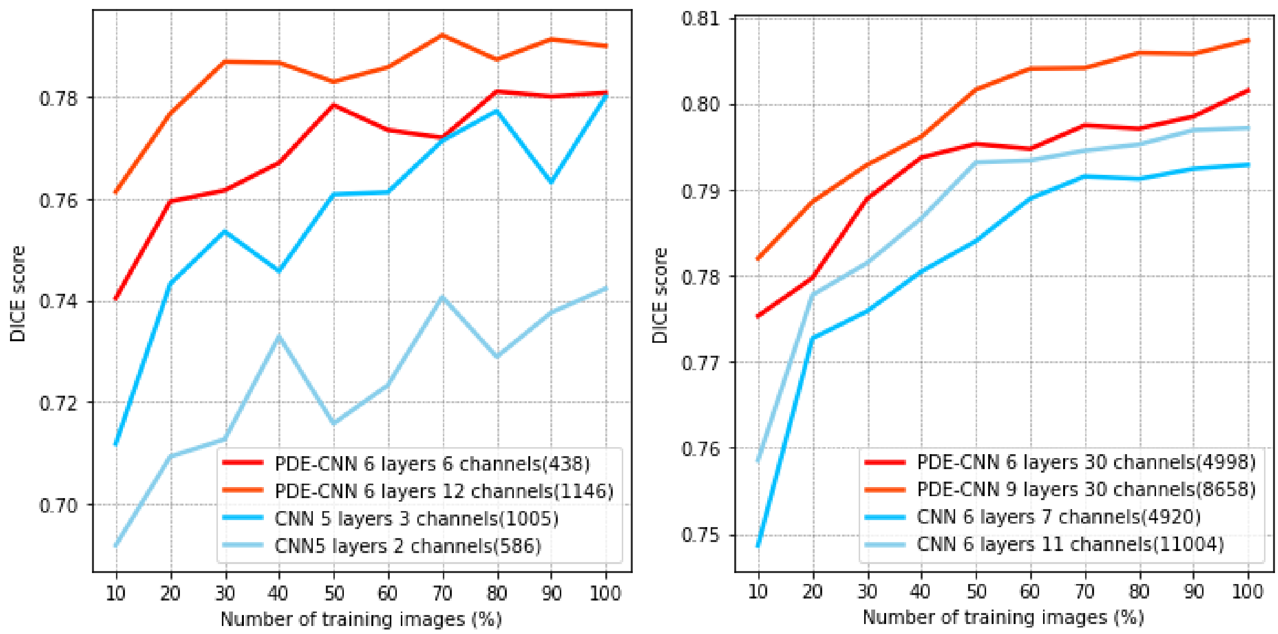

Figure 6.3: The mean dice score of PDE-CNNs compared to CNNs with comparable network complexity. The number of parameters for each network is indicated. On the left we have comparable networks with low complexity, and on the right, we have comparable networks with high network complexity

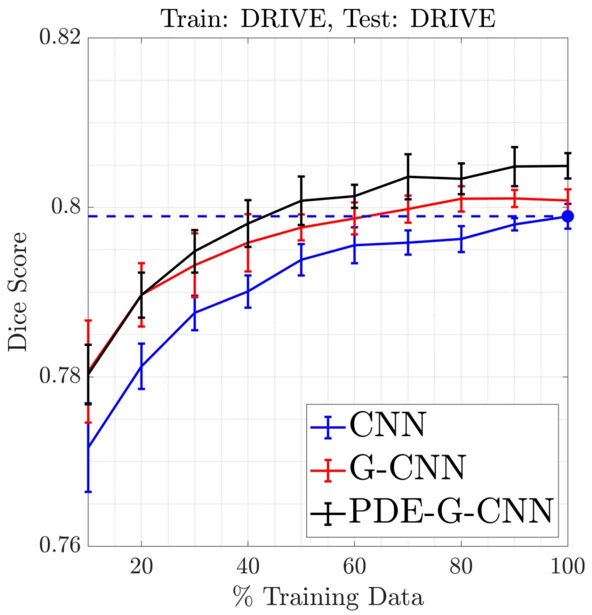

Figure 6.4: The comparison of a 6-layer: CNN (25662 parameters), G-CNN (24632 parameters), and a PDE-G-CNN (2560 parameters) architectures with varying amounts of training data. All networks are trained only on DRIVE.

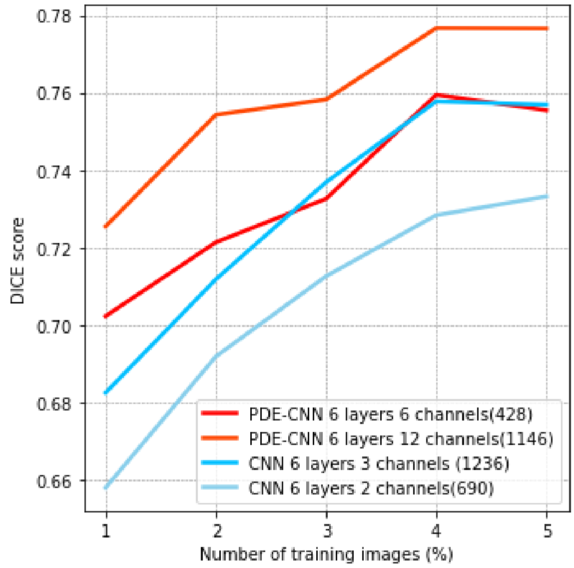

Figure 6.5: Mean die score over 5 experiments of PDE-CNNs compared to CNNs. For each experiment, specified number of random images are selected from the full training set for training.

Model	Layers	Channels	Parameters	score difference
PDE-CNN	6	6	438	0.040379
CNN	3	2	690	0.068294
PDE-CNN	5	3	1146	0.028646
CNN	5	3	1236	0.050465
PDE-CNN	6	30	4998	0.026177
CNN	6	7	4920	0.044212
PDE-CNN	9	30	8658	0.025325
CNN	6	11	11004	0.038585

Table 6.4: Side by side comparison of PDE-CNNs with CNNs with comparable network complexity. The difference in the dice score with 10 % training data and 100 % training data is shown on the right most column.

Chapter 7
Conclusion

In this thesis, we aimed to underpin the PDEs used in PDE-G-CNNs. In particular, we aimed to derive them from a list of axioms, when fixing the Lie group G and semiring R. As this is an ambitious goal, we decided to constrain ourselves to the cases listed in Section 1.2.

In Chapter 3, we laid a foundation for this approach by introducing the notion of semirings, and the corresponding semiring transform on a Lie group G. There, we have observed that two different types of scale space operations namely, the diffusion operations and the morphological operations both arise from semiring kernel operators. We have seen that the tropical semirings gives rise to morphological scale-spaces and the linear semiring to the diffusion scale-spaces.

We then established the axioms for the case $G = \mathbb{R}^2$ in Chapter 4. Upon their introduction, it was shown sequentially, how each axiom narrows down the possibility for the kernels which define the scale space operations. The linear semiring leads to the kernel given by (4.23) in the Fourier domain which depends on a parameter α, and trainable parameters $\mathbf{c}$ which corresponds to convection and $\mathcal{G}$ which is the metric tensor field which controls the anisotropy. For the morphological semiring, the kernel in the Fourier domain was found to be given by (4.24) where α, $\mathbf{c}$ and $\mathcal{G}$ respectively play the same role as in the linear case. It was then shown that these kernels give rise to evolution PDEs: for the linear case, quantum mechanical wave propagation and for the morphological case, HJB-PDE.

Next, we tackled the case for $G = SE(2)$. We have seen that the axioms needed to be adapted as the group structure of $SE(2)$ was quite different from that of $\mathbb{R}^n$. For example, in order to make sense of the notion of localised isotropy, we introduced the concept of local radial isometry. Furthermore, we saw that this isotropy was established in terms of a generator. Upon their introduction, for the case $R = R_L$, we showed that the generator must be of the form given by (5.12) given the axioms. We then proved main Theorem 5.1 to show that in the linear case, the axioms force a PDE given by (5.13). Here, we have noted that if we assume that if the PDE for the morphological case is given by a convex superlinear Hamiltonian $\mathcal{H}_{\mathcal{G}}$, then $\mathcal{H}_{\mathcal{G}}(p) = \frac{1}{\alpha}\|p\|^{\alpha}_{\mathcal{G}^{-1}}$. Furthermore, we have seen in Theorem 5.2 that the optional positivity axiom forces the condition $\alpha \leq 2$. The overall axiomatic derivation is summarised in Theorem 5.3.

Finally, in Chapter 6, we examine the effects of PDE-CNN on the network complexity and training data and we compare them to those of CNNs and PDE-G-CNNs. We see that while compared to PDE-G-CNNs, PDE-CNNs do not have the same level of decrease in network complexity and training data compared to similarly performing CNNs, they still outperform CNNs and G-CNN by a significant margin and their benefits especially come into play when testing on small training data.

Future Research

While we have made a decent progress in underpinning the PDEs used in PDE-G-CNNs axiomatically, there is still quite some work to be done.

Firstly, while we have managed a fully axiomatic derivation for the case $G = \mathbb{R}^n$, some of the results we used rely on the fact that we assume the underlying ring to be either the linear or the tropical semiring. While the scale-space operations arising from these semirings are of main interests in image processing, we would like to extend the approach towards a broader class of semirings, including the logarithmic semiring (associated to Hopf-Cole transformations) or the p-root semirings (that originate from the linear semiring by conjugation with taking a p-th power). This is ongoing research by my supervisors.

On top of this, for $SE(2)$, while we did manage an axiomatic derivation for the linear case and prove a result (Theoem 5.1) about the tropical case, the assumption we relied on (convex superlinear Hamiltonian) for the tropical case is a very strong one and there is still a big gap to fill. Furthermore, it is unclear if the axioms that we used to derive the linear case are the right axioms for the tropical case. For example, in the linear case, we relied on the linearity of the generator to use the Peetre theorem which was essential in our argument.

In the tropical case, the quasi- linearity (i.e. linearity w.r.t. semiring R) of Φ does not imply the quasi-linearity of the generator and on top of that, there seems to be no tropical equivalent of the Peetre theorem. These facts suggest that the approach requires serious adaptation for the tropical case and they still pose a highly significant challenge, even though one would expect it involves a decomposition into tropical (quasi-linear) version of the linear irreducible representations (5.22).

Bibliography

[AS64] M. Abramowitz and A. Stegun. *Handbook of Mathematical Functions*. New York: Dover, 1964.

[Yos68] Kosaku Yosida. *Functional analysis*. springer, 1968.

[Fol73] Gerald B Folland. "A fundamental solution for a subelliptic operator". In: *Bulletin of the American Mathematical Society* 79.2 (1973), pp. 373–376.

[Mac76] G.W. Mackey. *The Theory of Unitary Group Representations*. UCP, 1976.

[Rud87] Walter Rudin. *Real and Complex Analysis, 3rd Ed.* USA: McGraw-Hill, Inc., 1987. ISBN: 0070542341.

[Sug90] M. Sugiura. *Unitary representations and harmonic analysis*. second. Amsterdam, Kodansha, Tokyo: North-Holland Mathematical Library, 44., 1990.

[AQV94] Marianne Akian, Jean-Pierre Quadrat, and Michel Viot. "Bellman processes". In: *11th International Conference on Analysis and Optimization of Systems Discrete Event Systems*. Springer. 1994, pp. 302–311.

[Fol95] G.B. Folland. *A Course in Abstract Harmonic Analysis*. CRC Press, 1995.

[Pau+95] E.J. Pauwels et al. "An extended class of scale-invariant and recursive scale space filters". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17.7 (1995), pp. 691–701.

[KM97] Vassili N. Kolokoltsov and Victor P. Maslov. *Idempotent Analysis and Its Applications*. Springer Dordrecht, 1997. ISBN: 978-0-7923-4509-1. DOI: 10.1007/978-94-015-8901-7.

[Eva98] Lawrence C Evans. "Partial differential equations". In: *Providence, RI* (1998).

[CK01] G. S. Chirikjian and A. B. Kyatkin. *Engineering applications of noncommutitative harmonic analysis: with emphasis on rotation and motion groups*. Boca Raton CRC Press, 2001. ISBN: 0-8493-0748-1.

[Dud02] R. Dudley. *Real Analysis and Probability, 2nd Ed.* Cambridge, UK: CUP, 2002.

[Dui+04] R. Duits et al. "On the axioms of scale space theory". In: *Journal of Mathematical Imaging and Vision* 20 (2004), pp. 267–298.

[Sta+04] Joes Staal et al. "Ridge-based vessel segmentation in color images of the retina". In: *IEEE transactions on medical imaging* 23.4 (2004), pp. 501–509. URL: https://www.isi.uu.nl/Research/Databases/DRIVE/.

[FF05] Hartmut Führ and Hartmut Fuehr. *Abstract harmonic analysis of continuous wavelet transforms*. 1863. Springer Science & Business Media, 2005.

[FM07] Albert Fathi and Ezequiel Maderna. "Weak kam theorem on non compact manifolds". In: *Nonlinear Differential Equations and Applications NoDEA* 14.1 (Oct. 2007), pp. 1–27. ISSN: 1420-9004. DOI: 10.1007/s00030-007-2047-6. URL: https://doi.org/10.1007/s00030-007-2047-6.

[Agr+08] Andrei Agrachev et al. *The intrinsic hypoelliptic Laplacian and its heat kernel on unimodular Lie groups*. 2008. arXiv: 0806.0734 `[math.AP]`.

[DV08] Remco Duits and Markus Van Almsick. "The explicit solutions of linear left-invariant second order stochastic evolution equations on the 2D Euclidean motion group". In: *Quarterly of Applied Mathematics* 66.1 (2008), pp. 27–67.

[Ban10] Erik van den Ban. *Lie Groups*. University Course Notes UU, 2010. URL: https://webspace.science.uu.nl/~ban00101/lecnotes/lie2010.pdf.

[DF10a] R. Duits and E. Franken. "Left-Invariant Parabolic Evolution Equations on $SE(2)$ and Contour Enhancement via Orientation Scores". In: *QAM-AMS* 68 (2010), pp. 255–331.

[DF10b] R. Duits and E. M. Franken. “Left Invariant Parabolic Evolution Equations on $SE(2)$ and Contour Enhancement via Invertible Orientation Scores, Part I: Linear Left-Invariant Diffusion Equations on $SE(2)$”. In: *Quarterly of Applied mathematics, AMS* 68 (June 2010), pp. 255–292.

[DF10c] Remco Duits and Erik Franken. “Left-invariant parabolic evolutions on SE(2) and contour enhancement via invertible orientation scores Part II: Nonlinear left-invariant diffusions on invertible orientation scores”. In: *Quarterly of applied mathematics* 68.2 (2010), pp. 293–331.

[Bal+12] Zoltán M Balogh et al. “Functional inequalities and Hamilton–Jacobi equations in geodesic spaces”. In: *Potential analysis* 36.2 (2012), pp. 317–337.

[CW16] Taco S Cohen and Max Welling. “Group equivariant convolutional networks”. In: *Int. Conf. on Machine Learning*. 2016, pp. 2990–2999.

[SW16] Martin Schmidt and Joachim Weickert. “Morphological counterparts of linear shift-invariant scale-spaces”. In: *Journal of Mathematical Imaging and Vision* 56.2 (2016), pp. 352–366.

[Zha+16] Jiong Zhang et al. “Robust Retinal Vessel Segmentation via Locally Adaptive Derivative Frames in Orientation Scores”. In: *IEEE Transactions on medical imaging* 35.12 (2016), pp. 2631–2644. DOI: https://doi.org/10.1109/TMI.2016.2587062.

[PD17] J. M. Portegies and R. Duits. “New exact and numerical solutions of the (convection-) diffusion kernels on SE(3)”. In: *Differential Geometry and its Applications* 53 (2017), pp. 182–219.

[DBM19] Remco Duits, Erik Bekkers, and Alexey Mashtakov. “Fourier transform on the homogeneous space of 3D positions and orientations for exact solutions to linear PDEs”. In: *Entropy* 21.1 (2019), p. 38.

[Sme+20] Bart Smets et al. “PDE-based group equivariant convolutional neural networks”. In: *arXiv preprint arXiv:2001.09046* (2020).

[Cas21] Adrien Castella. “Introduction to Shift-Invariant PDE-based Convolutional Neural Networks and experimentation with Vessel Segmentation”. Bachelor Final Project. Eindhoven University of Technology, 2021.

[RDu+21] R.Duits et al. “Equivariant Deep Learning via Morphological and Linear Scale Space PDEs on the Space of Positions and Orientations”. In: *LNCS* 12679 (2021), pp. 27–39.

[Sme+21] Bart MN Smets et al. “Total Variation and Mean Curvature PDEs on the Homogeneous Space of Positions and Orientations”. In: *Journal of Mathematical Imaging and Vision* 63.2 (2021), pp. 237–262.

[XV21] Yaoda Xu and Maryam Vaziri-Pashkam. “Limits to visual representational correspondence between convolutional neural networks and the human brain”. In: *Nature Communications* 12 (Apr. 2021). ISSN: 2041-1723. DOI: 10.1038/s41467-021-22244-7. URL: https://doi.org/10.1038/s41467-021-22244-7.

[Sme+22] Bart Smets et al. “PDE-based group equivariant convolutional neural networks”. In: *JMIV* (2022), pp. 1–31.

[BPD23] Gijs Bellaard, Gautam Pai, and Remco Duits. “Geometric Adaptations of PDE-G-CNNs”. In: *9th International Conference on Scale Space and Variational Methods in Computer Vision*. 2023.

[Bel+23] Gijs Bellaard et al. “Analysis of (sub-)Riemannian PDE-G-CNNs”. In: *Journal of Mathematical Imaging and Vision* 65.6 (Dec. 2023), pp. 819–843. ISSN: 1573-7683. DOI: 10.1007/s10851-023-01147-w. URL: https://doi.org/10.1007/s10851-023-01147-w.

[Pai+23] Gautam Pai et al. “Functional Properties of PDE based Group Equivariant Convolutional Neural Networks”. In: *Proceedings of the 6th International Conference on Geometric Science of Information*. 2023.

Appendix A

A.1 Measurable Semiring Valued Functions on a Lie group

Let G be a Lie group, and R a semiring. Let $e_{\otimes} \in R$ denote the unity element of $\otimes$ in the semiring. Let $e_{\oplus} \in R$ denote the unity element of $\oplus$ in the semiring R.

Definition A.1 (Indicator). *Let $\omega \subseteq G$ be **any** set and R a semiring. The indicator function $I_\omega : G \to R$ of ω is given by*

$$1_{\omega}^{R,G}(x) = \begin{cases} e_{\otimes} & \text{for } x \in \omega \\ e_{\oplus} & \text{for } x \notin \omega \end{cases} \tag{A.1}$$

Definition A.2 (Simple Function). *Finite quasi-linear combinations of indicator functions of **measurable** sets A_i will be called simple functions.*

$$\bigoplus_{i=1}^{n} a_i \otimes 1_{A_i} \tag{A.2}$$

Definition A.3 (Sum Approachable). *A function $f : G \to R$ is called sum approachable if there exists $a_i \in R$ and $A_i \subseteq G$ **open**, such that*

$$f(g) = \lim_{n\to\infty} \bigoplus_{i=1}^{n} a_i \otimes 1_{A_i}(g) \tag{A.3}$$

The set of all sum-approachable functions from G to R is denoted by S_R.

Remark A.1. *In the linear semiring case every pointwise limit of simple functions is sum-approachable. However, in general this is not the case. For instance in the tropical. Now let us consider $R = T_+$. Then the morphological delta in the tropical max semiring is*

$$\delta^{T_+}(g) = \begin{cases} 0 & \text{if } g = e \\ -\infty & \text{otherwise} \end{cases} \tag{A.4}$$

*This function is a pointwise of simple functions (take the sequence $f_n = 1^{T_+}_{(-1/n,1/n)}$ for $n \in \mathbb{N}$). However, the morphological delta is **not** sum-approachable as we will see below in Cor. A.1.*

Definition A.4 (Measurable). *A function $f : G \to R$ is measurable if the pre-image of measurable sets is measurable. Here we use the Borel sigma algebra generated from the standard topology given by the metric.*

Lemma A.1. *A sum-approachable function $f : G \to R$ is measurable.*

Proof. Let X be a measure space and Y a metric space. The pointwise limit given by $f(x) = \lim_{n\to\infty} f_n(x)$ of measurable functions $f_n : X \to Y$ is again measurable, and the sums in the right hand side of (A.3) are measurable. $\square$

Next we investigate the tropical semiring cases.

Lemma A.2. *The pointwise supremum of lower-semi continuous (l.s.c.) functions from G to $\mathbb{R}$ that are bounded from above is again a l.s.c. function from G to $\mathbb{R}$ that is bounded from above.*

Proof. A function $f : G \to \mathbb{R}$ is l.s.c. iff given $g \in G$ and $r < f(g)$ there is an open neighborhood U of g such that $r < f(g)$ for all $g \in U$.
Taking $g \in G$ and any $r < f(g) := \sup_{i\in I} f_i(g)$ it must be that there exists and $i \in I$ (the index set) s.t. $r < f_i(g)$. Since f_i is lower semicontinuous there is an open neighborhood U of g such that $r < f_i(g)$ for all $g \in U$. As $f_i(g) \le f(g)$ for all g, it follows that $r < f(g)$ for each $g \in U$. $\square$

Lemma A.3. *A sum approachable function in the tropical semiring T^+ that is bounded from above is l.s.c. and bounded from above A sum approachable function in the tropical semiring T^- that is bounded from below is u.s.c. and bounded from below.*

Proof. In the tropical semiring T^+ we have that (A.3) boils down to

$$f(g) = \lim_{n\to\infty} \max_{i\in\{1,\dots,n\}} a_i + 1_{A_i}(g) = \sup_{i\in\mathbb{N}} a_i + 1_{A_i}(g) =: \sup_{i\in\mathbb{N}} f_i(g)$$

Now the functions f_i behind the supremum are lower semi-continuous as all 1_{A_i} given by (A.1) with $e_{\otimes}^{T_+} = 0$ and $e_{\oplus}^{T_+} = -\infty$ are lower-semi continuous, so that the result for T_+ follows by Lemma A.2.

Now the result for T_- follows from the result for T_+ and:

- $\min(x) = -\max(-x)$,
- f is l.s.c. if and only of $-f$ is u.s.c,
- $e_{\oplus}^{T_-} = -e_{\oplus}^{T_+} = \infty$.

so that the result follows. $\square$

Corollary A.1. *The morphological delta δ^{T_+} is not sum-approachable in the the tropical semiring T_+*

Proof. Suppose it was sum-approachable then by Lemma A.2 it would be lower-semicontinuous which is not the case. $\square$

Lemma A.4. *In the linear semiring case a measurable function is sum-approachable. In the tropical semiring case a measurable function need not be sum-approachable.*

Proof. The first statement for $R = R_L$ is a well-known property of measurable functions from metric spaces to metric spaces (endowed with corresponding Borel sigma algebra's), [Dud02].

Now let us consider $R = T_+$. Then the *morphological delta* in the tropical max semiring is

$$\delta^{T_+}(g) = \begin{cases} 0 & \text{if } g = e \\ -\infty & \text{otherwise} \end{cases} \tag{A.5}$$

This function is clearly measurable. However, it not sum-approachable as it is not l.s.c, recall the previous lemma. □

A.2 Semiring Integration on a unimodular Lie group

For convenience we restrict ourselves to unimodular Lie groups G. This allows us to talk about *the Haar measure* on G (up to scalar multiplication) without a need to distinguish between left and right Haar measures.

We define the integral $I_R^G : S_R \to R$ by *requiring* that

1. I_R^G is quasi-linear,
2. I_R^G is continuous.

The above definition is compatible with measurable functions in the linear semiring where $I_L^G(1_\omega^{L,G}) = \mu_G(\omega)$ and $I_R^G(f) = \int_\omega f(x)\, \mathrm{d}\mu_G(x)$ with μ_G the up to scalar multiplication unique left-Haar measure on Lie group G. We write $I_R^G(f)$ symbolically as $I_L^G(f) =: \bigoplus\limits_{x\in G} f(x)$.

In the tropical setting we define $I_R^G(1_A^{T_+,G}) = 0$ for all open sets A in G.

Lemma A.5. *For $f : G \to \mathbb{R}$ with $f \in S_{T_-}$ one has*

$$I_{T_-}^G(f) = \inf_{x\in G} f(x)$$

For $f : G \to \mathbb{R}$ with $f \in S_{T_+}$

$$I_{T_-}^G(f) = \sup_{x\in G} f(x)$$

Proof. Set $R = T_-$. Let $f \in S_R$. Then by Def. A.3 and Lem. A.3 we approximate u.s.c. f by u.s.c. simple functions

$$f = \lim_{n\to\infty} \bigoplus_{i=1}^{n} c_n^i \otimes 1_{\omega_i}^R.$$

By quasi-linearity $I_R^G(1_{\omega_i}^R) = I_R^G(1_{\omega_i}^R \otimes 1_{\omega_i}^R) = I_R^G(1_{\omega_i}^R) \otimes I_R^G(1_{\omega_i}^R) \Rightarrow I_R^G(1_{\omega_i}^R) = 0$. Then by continuity of I_R^G, resp. quasi-linearity of I_R^G we have

$$\begin{aligned} I_R^G\left(\lim_{n\to\infty} \bigoplus_{i=1}^{n} c_n^i \otimes 1_{\omega_i}^R\right) &= \lim_{n\to\infty} I_R^G\left(\bigoplus_{i=1}^{n} c_n^i \otimes 1_{\omega_i}^R\right) \\ &= \lim_{n\to\infty} \bigoplus_{i=1}^{n} c_n^i \otimes I_R^G(1_{\omega_i}^R) = \lim_{n\to\infty} \inf_{i\in\{1,\dots,n\}} c_n^i \\ &= \inf_{x\in G} f(x). \end{aligned}$$

from which the first claim follows. The second claim is tangential. □

A.3 Relating morphological convolutions on $(G, \mathcal{G})$ to morphological convolutions on $\mathbb{R}$ via the distance map

Lemma A.6. *A kernel corresponding to a convex, superlinear Hamiltonian $H_{\mathcal{G}}(p) = H^{1D}(||p||_{\mathcal{G}^{-1}})$ depends only on the norm of the argument, i.e. $k_t(g) = \kappa_t^{1D}(d_{\mathcal{G}}(g,e))$.*

Furthermore both kernels satisfy the semigroup property on resp. G and $\mathbb{R}^+$ and relate via (5.16):

$$k_{t+s}(g) = (k_t \circledast_R^G k_s)(g) = \kappa_{t+s}^{1D}(d_{\mathcal{G}}(g,e)) = (\kappa_t^{1D} *_R^{\mathbb{R}} \kappa_s^{1D})(d_{\mathcal{G}}(g,e)).$$

Proof. We note that from [Bel+23, Prop.1], that we have

$$\kappa_t^{1D}(d_{\mathcal{G}}(g,e)) = t\mathcal{L}^{1D}(d(g,e)/t)$$

where $\mathcal{L}^{1D}$ denotes the corresponding Lagrangian of H^{1D} Now, our assumption of convex superlinearity of H implies that this condition also holds for the corresponding Lagrangian $\mathcal{L}$. Now, note that if $H_{\mathcal{G}}$ is convex, then it is easy to see that H^{1D} is convex as well. Indeed, for $x, y \in \mathbb{R}_+$, let p such that $||p|| = x$, then

$$\begin{aligned}
H^{1D}(tx + (1-t)y) &= H^{1D}(t||p|| + (1-t)\frac{y}{x}||p||) \\
&= H^{1D}(||tp + (1-t)\frac{y}{x}p||) \\
&= H_{\mathcal{G}}(tp + (1-t)\frac{y}{x}p) \\
&\leq tH^{1D}(x) + (1-t)H^{1D}(y)).
\end{aligned}$$

With a similar argument, we can show that H^{1D} is superlinear as well. Therefore, we have that $\mathcal{L}^{1D}$ convex super linear and hence κ_t^{1D} as well. But then we have that κ_t^{1D} is monotonically increasing on $\mathbb{R}_+$. Now, we have (omitting labels 1D in κ and $\mathcal{G}$ in d)

$$\begin{aligned}
\kappa_{t+s}(d(g,e)) &= k_{t+s}(g) = k_t \circledast_R^G k_s(g) \\
&= \inf_{h \in G} \kappa_t(d(g,h)) + \kappa_s(d(h,e)) \\
&= \inf_{h \in \gamma_{e,g}^{\min}} \kappa_t(d(g,h)) + \kappa_s(d(h,e)) \\
&= \inf_{h \in \gamma_{e,g}^{\min}} \kappa_t(d(g,e) - d(h,e)) + \kappa_s(d(h,e)) \\
&= \inf_{0 \leq r \leq d(g,e)} \kappa_t(d(g,e) - r) + \kappa_s(r) \\
&= (\kappa_t^{1D} *^{\mathbb{R}} \kappa_s^{1D})(d(g,e))
\end{aligned}$$

The third equality follows from the fact that κ_t and κ_s are monotonically increasing and the fact that for any $h \in G$ there is an element $\tilde{h} \in \gamma_{e,g}^{\min}$ such that $d(g,\tilde{h}) \leq d(g,h)$ and $d(e,\tilde{h}) \leq d(e,h)$. The fourth equality holds since $h \in \gamma_{e,g}^{\min}$ implies $d(g,e) = d(g,h) + d(h,e)$.

Now for the final equality we must motivate that the constraint $0 \leq r \leq d_{\mathcal{G}}(g,e)$ is a redundant one. To this end we define

$$G(r) := \kappa_t(d-r) + \kappa_s(r) = t\mathcal{L}^{1D}\left((d-r)/t\right) + s\mathcal{L}^{1D}(r/s)$$

implies

$$G'(r) = 0 \iff \mathcal{L}^{1D'}((d-r)/t)) = \mathcal{L}^{1D'}(r/s)$$

Now we have that $\mathcal{L}^{1D}$ is convex on $\mathbb{R}$. But then $\mathcal{L}^{1D}$ monotonically increasing. But then it is injective and hence we have

$$G'(r) = 0 \iff \frac{d-r}{t} = \frac{r}{s} \iff r = \frac{d}{\frac{t}{s}+1}.$$

Therefore, as $t, s \in \mathbb{R}_+$, we get that $G(r)$ attains the global minimum for $0 \le r \le d$. $\square$