



# PDE-CNNs: Axiomatic Derivations and Applications

Gijs Bellaard<sup>1</sup> · Sei Sakata<sup>1</sup> · Bart M. N. Smets<sup>1</sup> · Remco Duits<sup>1</sup>

Received: 15 April 2024 / Accepted: 15 January 2025  
© The Author(s) 2025

## Abstract

PDE-based group convolutional neural networks (PDE-G-CNNs) use solvers of evolution PDEs as substitutes for the conventional components in G-CNNs. PDE-G-CNNs can offer several benefits simultaneously: fewer parameters, inherent equivariance, better accuracy, and data efficiency. In this article, we focus on Euclidean equivariant PDE-G-CNNs where the feature maps are two-dimensional throughout. We call this variant of the framework a PDE-CNN. From a machine learning perspective, we list several practically desirable axioms and derive from these which PDEs should be used in a PDE-CNN, this being our main contribution. Our approach to geometric learning via PDEs is inspired by the axioms of scale-space theory, which we generalize by introducing semifield-valued signals. Our theory reveals new PDEs that can be used in PDE-CNNs and we experimentally examine what impact these have on the accuracy of PDE-CNNs. We also confirm for small networks that PDE-CNNs offer fewer parameters, increased accuracy, and better data efficiency when compared to CNNs.

**Keywords** PDE · Scale-space · Semifield · Equivariance · Neural network · Machine learning · Computer vision · Convolution · Tropical semiring · Morphology

## 1 Introduction

Recently, PDE-based group equivariant convolution neural networks (PDE-G-CNNs) [1] were introduced. PDE-G-CNNs belong to the broad family of group equivariant convolution neural networks (G-CNNs) [2]. Unlike traditional CNNs, PDE-based networks replace the usual components that make up a CNN layer that being convolutions, max pooling, and nonlinear activation functions, by solvers of evolution PDEs. The coefficients that govern the effect of the PDEs serve as the trainable parameters. Figure 1 contains a diagram of an example CNN layer and PDE layer, intended to illustrate the similarities and differences between them.

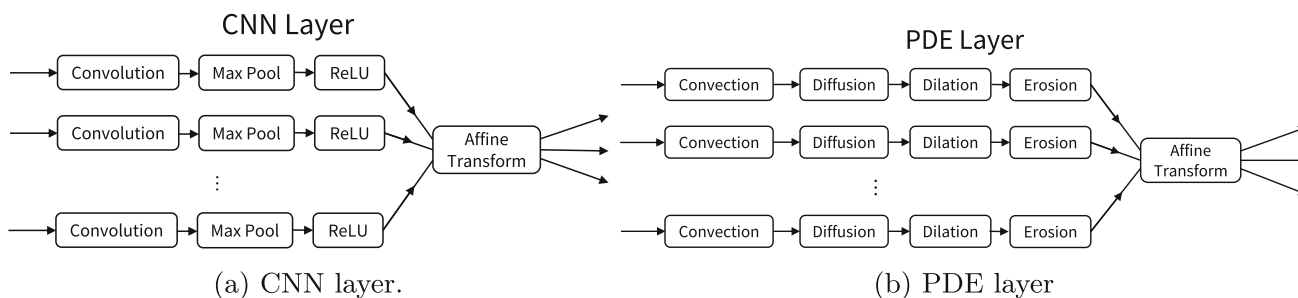
It is shown in [1, 3–5] that, for vessel segmentation in medical images and digit classification problems, PDE-G-CNNs—in addition to being inherently equivariant—require fewer parameters, achieve higher accuracy, and are more data efficient, in comparison with CNNs and G-CNNs. From this perspective, PDE-G-CNNs can be preferable over other architectures in the aforementioned image processing tasks.

The PDE-G-CNN architecture is general in the sense that the feature maps  $f : M \rightarrow \mathbb{R}$  are defined on an arbitrary *homogeneous space*  $M$  on which a *Lie group*  $G$  acts. However, the existing literature [1, 3–5] mainly concerns itself with  $M = \mathbb{M}_2 = \mathbb{R}^2 \times S^1$ , the space of two-dimensional positions and orientations, together with  $G = \text{SE}(2) = \mathbb{R}^2 \rtimes \text{SO}(2)$ , the group of two-dimensional rotations and translations. In this article, we will *not* consider the general setting, or  $M = \mathbb{M}_2$  for that matter, and restrict ourselves to  $M = \mathbb{R}^2$  and  $G = \text{SE}(2)$ , i.e., standard two-dimensional Euclidean space with its roto-translation symmetries, for simplicity. We call this specific instance a PDE-CNN.

---

✉ Gijs Bellaard  
g.bellaard@tue.nl  
Sei Sakata  
seisakata@gmail.com  
Bart M. N. Smets  
b.m.n.smets@tue.nl  
Remco Duits  
r.duits@tue.nl

<sup>1</sup> Department of Mathematics and Computer Science, CASA,  
Eindhoven University of Technology, Eindhoven, The  
Netherlands



**Fig. 1** Diagram of an example CNN layer and PDE layer. The vertical direction represents the channels. The arrows represent the “flow” of the feature maps through the parts that make up a layer. In machine learning terms, the affine transformation block is equivalent to a 2D convolution module with bias and 1x1 kernels. PDE-based networks replace the usual components that make up a CNN layer that being con-

volution, max pooling, and nonlinear activation functions, by solvers of evolution PDEs. The PDEs here are convection, diffusion, dilation, and erosion (1). With “solvers,” we mean the mapping from the initial condition  $f|_{t=0}$  to  $f|_{t=T}$ . We can take  $T = 1$  without loss of generality due to the scale equivariance property of the PDEs (Axiom 4)

In [1], the evolution PDEs that are used in the PDE-G-CNN architecture are

$$\text{convection} \quad \frac{\partial f}{\partial t} = v \cdot \nabla f \tag{1a}$$

$$\alpha - \text{diffusion} \quad \frac{\partial f}{\partial t} = -\frac{1}{\alpha}(-\Delta)^{\alpha/2} f, \alpha > 0 \tag{1b}$$

$$\alpha - \text{dilation} \quad \frac{\partial f}{\partial t} = +\frac{1}{\alpha}\|\nabla f\|^{\alpha}, \alpha > 1 \tag{1c}$$

$$\alpha - \text{erosion} \quad \frac{\partial f}{\partial t} = -\frac{1}{\alpha}\|\nabla f\|^{\alpha}, \alpha > 1. \tag{1d}$$

Here  $f : M \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$  is some scalar field on  $M$  evolving over time  $t \geq 0$ , with  $f(\cdot, 0)$  set to an initial condition. In the convection,  $v : M \rightarrow TM$  denotes a vector field, and in the diffusion  $-(-\Delta)^{\alpha/2}$  denotes a (fractional) power of the Laplacian. Intuitively, the PDEs, respectively, correspond to shifting, blurring, max pooling, and min pooling.

Importantly, the PDEs (1) are (implicitly) dependent on the Riemannian metric tensor field  $\mathcal{G}$  that is chosen on the homogeneous space  $M$ . When using different Riemannian metrics, concepts such as Laplacian  $\Delta$ , gradient  $\nabla$ , and norm  $\|\cdot\|$  change accordingly, consequently altering the effect of the PDEs. The parameters that determine the Riemannian metrics  $\mathcal{G}$  are learned during the training of a PDE-based neural network. The metric tensor field  $\mathcal{G}$  is designed to be invariant to the Lie group  $G$ , resulting in  $G$ -equivariant processing of the signals  $f : M \rightarrow \mathbb{R}$  [1].

The diffusion, dilation, and erosion PDEs (1) used in [1] were not chosen arbitrarily; they satisfy properties considered desirable from a machine learning perspective. For example, the PDEs are *quasilinear* and *equivariant* meaning that they 1) can be solved using convolutional-like operations allowing for fast parallel computation, and 2) allow for the design of inherently equivariant networks, resulting in an architecture that is robust and data efficient [6, 7]. In fact, the desirable properties that we want the PDEs in PDE-

based neural networks to have are essentially the axioms of *scale-space theory* [8–14]. Later, in Sect. 4.1 we explore and motivate these properties in more detail.

In this article, we will be deriving in an axiomatic way which PDEs should be used in PDE-based neural networks. This framework includes the PDEs that are currently already used (1), but also reveals previously unused PDEs, meaning that the accuracy of PDE-G-CNNs could possibly be improved by adding them. Our approach is inspired by the axioms of scale-space theory, which we generalize by introducing *semifield-valued* signals, and motivated from a machine learning perspective.

### 1.1 Contributions

We list six axioms (Definition 27) that a PDE used in a PDE-based neural network should satisfy. The axioms are closely related to classical scale-space theory, but are more general in the sense that we permit semifield-valued signals. The goal of this generalization is to allow for the discovery (or invention) of new PDEs that can be used in the design of PDE-based neural networks.

We will only consider semifields that are commutative and one-dimensional, and the domain of the PDEs will be the two-dimensional Euclidean space  $\mathbb{R}^2$ . To maintain a practical perspective, we will consistently connect the overarching theory using five example semifields: the linear, root, logarithmic, tropical min, and tropical max semifields.

From the axioms, we demonstrate in Theorem 1 that every semifield corresponds to a unique family of scale-spaces, this being the main theoretical contribution of the article. This shows that PDE-based neural networks, in their current form, can be extended greatly by adding new PDEs that generate currently unused scale-spaces.

We experimentally assess how effective the incorporation of new semifields and their corresponding PDEs is in Sect. 7.1.

In Sect. 7.2, we verify that PDE-CNNs exhibit superior data efficiency, reduced parameter count, and competitive accuracy compared to traditional CNNs.

### 1.2 Short Outline

In Sect. 2, we provide background on scale-spaces, semifields, and a non-exhaustive list of related literature. In Sect. 3, we define semifields and all related structures and operations. In Sect. 4, we state the semifield scale-space axioms. In Sect. 5, we show that once a semifield is chosen, a unique (one-parameter) family of scale-spaces arise (Theorem 1). In Sect. 6, we briefly note on the architectural design of PDE-CNNs. In Sect. 7, we lay out two experiments and discuss their results. In Sect. 8, we conclude the article.

## 2 Background

### 2.1 Scale-Spaces

The desired properties of PDEs in PDE-based neural networks are closely related to those of scale-space representations. In fact, there is a one-to-one correspondence between scale-spaces and PDEs used in PDE-based neural networks. In this section, we will introduce and motivate the concept of scale-spaces, providing a few examples, and show to which PDE they correspond.

Real-world scenes contain many different objects at different scales. When a computer is tasked with analyzing an image of a scene, there is no way for it to know beforehand at which scale(s) the interesting structures live. One way to tackle this problem is to analyze the image of interest at *all* scales.

In broad terms, a *scale-space representation* of an image  $f_0$  is an ordered collection of images  $f_t$  where each successive image contains less and less detail, that is, the smaller scales have been processed away. The collection of images is usually indexed by the *scale parameter*  $t \geq 0$  with  $t = 0$  being the original image.

Scale-spaces are a natural choice for computer vision solutions (either neural networks or classical methods) as they respect the inherent symmetries of images, that being translation, rotational, and scaling symmetries. What we mean by this mathematically is that, for example, the scale-space  $g_t$  of a translated image  $g_0 = T_v f_0$ , is equal to the translated scale-space of the original image:  $g_t = T_v f_t$ . Here  $T_v$  is the translation operator defined by  $(T_v f)(x) = f(x - v)$ . Analogous statements hold for the rotation and scaling symmetries. We say that creating the scale-space representation

of an image is *equivariant* with respect to translation, rotations, and scalings.

The prototypical, and most likely first [9, 15, 16], example of a scale-space is the *Gaussian scale-space* made by successive *diffusing* (i.e., *blurring* or *smoothing*) of the original image. The Gaussian scale-space  $f_t$  of a two-dimensional image  $f_0 : \mathbb{R}^2 \rightarrow \mathbb{R}$  can be written as a *linear convolution*  $*$  with a Gaussian *kernel*  $k_t$ :

$$\begin{aligned} \frac{\partial f}{\partial t} &= \frac{1}{2} \Delta f, \\ f_t &= k_t * f_0, \quad k_t(x) = \frac{1}{2\pi t} \exp\left(-\frac{\|x\|^2}{2t}\right), \\ (k_t * f_0)(x) &= \int_{\mathbb{R}^2} k_t(x - y) f_0(y) dy, \end{aligned} \tag{2}$$

where we used the notation  $f(x, t) = f_t(x)$ .

Two other examples are the *morphological scale-space* representations [17] made by successively *dilating* or *eroding* the original image. The  $\alpha$ -dilation scale-space can be written as a nonlinear *dilating convolution*  $\boxplus$  with a kernel:

$$\begin{aligned} \frac{\partial f}{\partial t} &= +\frac{1}{\alpha} \|\nabla f\|^\alpha, \\ f_t &= k_t \boxplus f_0, \quad k_t(x) = -\frac{t}{\beta} \left(\frac{\|x\|}{t}\right)^\beta, \\ (k_t \boxplus f_0)(x) &= \sup_{y \in \mathbb{R}^2} k_t(x - y) + f_0(y), \end{aligned} \tag{3}$$

where  $\beta$  is such that  $1/\alpha + 1/\beta = 1$ . The  $\alpha$ -erosion scale-space is created using an *eroding convolution*  $\boxminus$ :

$$\begin{aligned} \frac{\partial f}{\partial t} &= -\frac{1}{\alpha} \|\nabla f\|^\alpha, \\ f_t &= k_t \boxminus f_0, \quad k_t(x) = \frac{t}{\beta} \left(\frac{\|x\|}{t}\right)^\beta, \\ (k_t \boxminus f_0)(x) &= \inf_{y \in \mathbb{R}^2} k_t(x - y) + f_0(y). \end{aligned} \tag{4}$$

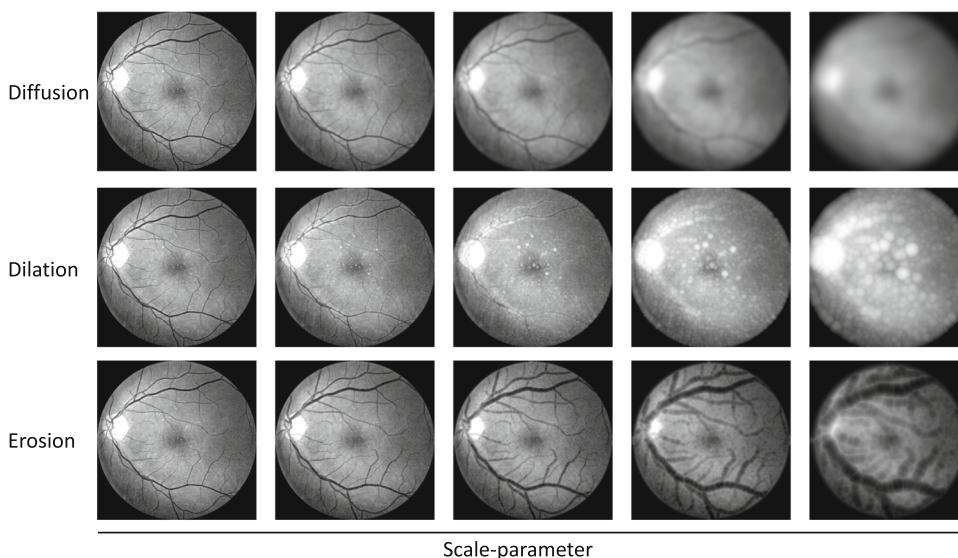
The dilating and eroding convolutions are collected under the umbrella term *morphological convolution*. This is because they are related by the identity  $-(-f \boxplus -g) = f \boxminus g$ .

In Fig. 2, the Gaussian, quadratic ( $\alpha = 2$ ) dilation, and quadratic erosion scale-space representations are visualized of a grayscale image of the fundus of the eye [18].

### 2.2 Semifields & Quasilinearity

Every scale-space representation has a natural corresponding algebra called a *semifield*. In this section, we will show which semifields correspond to the diffusion, dilation, and erosion scale-spaces.

**Fig. 2** Gaussian (2), quadratic ( $\alpha = 2$ ) dilation (3), and quadratic erosion (4) scale-space representations of a grayscale image of the fundus of the eye at various scale parameters. In the Gaussian scale-space, both white and black features fade away toward a uniform image. In the dilation scale-space, the black details (low values), such as the vessels, vanish at bigger scales. In the erosion scale-space, the white details (high values), such as the space between vessels, are removed at higher scales



The Gaussian scale-space representation (2) is linear in the sense that if one takes two images  $f_0, g_0 : \mathbb{R}^2 \rightarrow \mathbb{R}$  and two scalars  $a, b \in \mathbb{R}$ , then the scale-space  $h_t$  of the image  $h_0 = af_0 + bg_0$  is equal to  $h_t = af_t + bg_t$ . But, in an analogous manner, the dilation scale-space (3) is *quasilinear* in the sense that the scale-space of the image  $h_0 = \max\{a + f_0, b + g_0\}$ , where we interpret the maximum pointwise, is equal to  $h_t = \max\{a + f_t, b + g_t\}$ . In the same way, the erosion scale-space (4) is quasilinear in the min sense. To define what we mean with quasilinear more precisely, we need to introduce semifields.

A *semifield*  $(R, 0, 1, \oplus, \otimes)$  is an algebraic structure like a field but where we relax the requirement that the addition  $\oplus$  has inverses. The prototypical example of a semifield is the nonnegative real numbers  $L_{\geq 0} = (\mathbb{R}_{\geq 0}, 0, 1, +, \times)$  with standard addition and multiplication. We have already seen two other examples of semifields in the dilation and erosion scale-spaces. Namely, the so-called *tropical max* semifield defined as  $T_+ = (\mathbb{R} \cup \{-\infty\}, -\infty, 0, \max, +)$  and the *tropical min* semifield  $T_- = (\mathbb{R} \cup \{\infty\}, \infty, 0, \min, +)$ . In the tropical semifields, the minimum (or maximum) of two numbers becomes semifield addition, and normal addition becomes semifield multiplication.

With the definition of a semifield, we can state the quasilinearity of a scale-space formally as *semifield R-linearity*. So, like before, consider a semifield  $R$  and two semifield-valued images  $f_0, g_0 : \mathbb{R}^2 \rightarrow R$  and two elements  $a, b \in R$ . Then by a scale-space being  $R$ -linear, we mean that the scale-space of the  $R$ -linear combination of images  $h_0 = (a \otimes f_0) \oplus (b \otimes g_0)$  is equal to the  $R$ -linear combination of scale-spaces  $h_t = (a \otimes f_t) \oplus (b \otimes g_t)$ . In other words, the operation that takes an image and returns its scale-space representation is a semifield linear operator. For example, the Gaussian scale-space

is  $L_{\geq 0}$ -linear, the quadratic dilation scale-space is  $T_+$ -linear, and the quadratic erosion scale-space is  $T_-$ -linear.

In [8, 13], it is argued in an axiomatic way that the only linear scale-space representations correspond to solutions of the fractional diffusion (pseudo-)PDE system (1b). In a completely analogous manner, one can show [19, 20] that the only morphological scale-spaces, that being scale-spaces that are  $T_+$  or  $T_-$  linear, correspond to (viscosity) solutions of the  $\alpha$ -dilation (1c) and  $\alpha$ -erosion PDE (1d).

These facts reveal something important: to discover new PDEs that can be used in PDE-based neural network we *need* to generalize scale-space theory to semifields other than just  $L_{\geq 0}, T_+$ , and  $T_-$ .

### 2.3 Related Work

In this section, we provide a non-exhaustive list of related scale-space literature.

*Linear Scale-Spaces.* In [9], the first [15] axiomatic treatment of linear scale-space theory is presented. Axioms such as linearity, roto-translation equivariance, one-parameter semigroup property, and most notably, the scale equivariance, can all be found in Iijima’s article, axioms we will also be using. Iijima shows that the Gaussian scale-space arises from his axioms, and that the Laplacian generates it. In [8, 13], an extended class of linear scale-spaces is explored. They derive that (fractional) powers of the Laplacian (1b) are valid linear scale-space generators. Analysis of the scale-space axioms in the Fourier domain is extensively used, an approach we apply in the broader semifield setting.

*Morphological Scale-Spaces.* In [17], it was shown for the first time [12] that morphological operators like dilations and erosions in image processing can be described in terms of PDEs. In [21], the slope transform is shown to be

the morphological counterpart of the Fourier transform and related to the Legendre–Fenchel transform. The semifield Fourier transform we introduce reduces to the Legendre–Fenchel transform in the tropical semifield cases. In [20, 22], a connection between linear and morphological scale-spaces is described using the Cramér transform. The Cramér transform gives us a way to translate between the kernels of the linear and morphological scale-spaces. We will show that the kernels of all semifield scale-spaces have the same form in the Fourier domain (this being our main theorem), illuminating further the connection between the linear and morphological world.

*Other Scale-Space Theory.* In [10], an axiomatic approach to PDE-based scale-spaces is described. The strength of this approach is that it also includes mean curvatures flows [23] as highly powerful nonlinear PDEs (also on Lie groups [24]). Solutions of such non-linear PDEs may be solved with median filtering [25]; however, they lack a semifield structure (taking mean/median are not associative binary operations), thus falling outside the scope of the theory presented here. In [11], nonlinear scale-spaces are obtained by performing a monotonic transformation (known as a “Cole–Hopf” transform [26, Ch.4.4]) on the gray values of a standard linear scale-space and deducing what nonlinear PDE corresponds to the obtained evolution. This transformation neatly bridges linear, logarithmic, and in the extreme cases, morphological scale-spaces, and we will also use this link. In [12], an algebraic framework for scale-spaces is given. Importantly, their perspective is (initially) totally divorced from PDEs, convolutions, and kernels, and focuses solely on the evolution operator. We will define our semifield scale-spaces in the same manner.

*Scale-Spaces in Machine Learning.* In [27–30], the Gaussian scale-space and its spatial (fractional) derivatives are employed to design architectures that can learn filters at the appropriate scale by optimizing the scale parameter(s) during training. Architectures such as [31–33] also learn scale parameters. PDE-G-CNNs learn Riemannian metric tensor fields, which, due to the scale equivariance of scale-space representations, is equivalent to learning scale parameters. In this sense, PDE-based neural networks are closely related to these “scale learning” architectures. In [34–37], (discrete) Gaussian and morphological scale-space representations are used to create architectures that are scale equivariant.

### 3 Semifield Theory

In this section, we define semifields (Definition 1) and all mathematical structures and operations made from them. This includes important concepts such as semimodules (Definition 7), linearity (Definition 8), measures (Definition 16),

integration (Definition 18), convolution (Definition 20), and Fourier transforms (Definition 25).

#### 3.1 Semifield, Semimodules & Linearity

**Definition 1** (Semifield) A (commutative) semifield  $R$  is a tuple  $R = (R, 0, 1, \oplus, \otimes)$  where  $\oplus, \otimes : R \times R \rightarrow R$  are two commutative and associative binary operations on  $R$  called semifield addition and multiplication, such that for all  $a, b, c \in R$ :

$$\begin{aligned} a \oplus 0 &= a, \\ a \otimes 1 &= a, \\ a \neq 0 &: \exists a^{-1} : a \otimes a^{-1} = 1, \\ a \otimes 0 &= 0, \\ a \otimes (b \oplus c) &= (a \otimes b) \oplus (a \otimes c). \end{aligned}$$

In other words, a semifield is a field where we do not require to have “negative elements,” that being additive inverses.

Throughout the article, we will denote an arbitrary semifield-related operation with a circled version of the most closely related linear counterpart. Some example symbols are  $\oplus, \otimes, \oint,$  and  $\circledast$ , which, respectively, correspond to semifield addition, multiplication, integration, and convolution.

In this article, we mainly consider the following semifields:

**Definition 2** (Semifields of Interest)

- a) The linear semifield  $L = (\mathbb{R}, 0, 1, +, \times)$  with the usual addition  $+$  and multiplication  $\times$ . We can restrict the set to  $\mathbb{R}_{\geq 0}$  and we write  $L_{\geq 0}$  in that case.
- b) The root semifields  $R_p = (\mathbb{R}_{\geq 0}, 0, 1, \oplus_p, \times)$  with  $p \neq 0$  where semifield addition is  $a \oplus_p b := \sqrt[p]{a^p + b^p}$ , and where semifield multiplication is normal multiplication.
- c) The logarithmic semifields  $L_\mu = (\mathbb{R} \cup \{\pm\infty\}, \pm\infty, 0, \oplus_\mu, +)$  with  $\mu \neq 0$  where semifield addition is  $a \oplus_\mu b := \frac{1}{\mu} \ln(e^{\mu a} + e^{\mu b})$ , and where semifield multiplication is normal addition. If  $\mu > 0$  we add  $-\infty$  to the ring to act as the additive identity, and if  $\mu < 0$  we add  $+\infty$ .
- d) The tropical max semifield  $T_+ = (\mathbb{R} \cup \{-\infty\}, -\infty, 0, \max, +)$ , where  $\max$  is semifield addition, and usual addition is semifield multiplication.
- e) The tropical min semifield  $T_- = (\mathbb{R} \cup \{\infty\}, \infty, 0, \min, +)$ , where  $\min$  is semifield addition, and usual addition is semifield multiplication.

The family of logarithmic semifields is interesting as in the limits one has:

$$\begin{aligned} \lim_{\mu \rightarrow +\infty} a \oplus_\mu b &= \max(a, b), \\ \lim_{\mu \rightarrow -\infty} a \oplus_\mu b &= \min(a, b). \end{aligned} \tag{5}$$

Thereby, the family of logarithmic semifields  $L_\mu$  relate to the tropical semifields  $T_\pm$  in the extreme cases of  $\mu$ .

**Definition 3 (Semifield Isomorphism)** Let  $R = (R, 0, 1, \oplus, \otimes)$  and  $\tilde{R} = (\tilde{R}, \tilde{0}, \tilde{1}, \tilde{\oplus}, \tilde{\otimes})$  be two semifields. A semifield isomorphism  $\varphi : R \rightarrow \tilde{R}$  is a bijective mapping that satisfies for all  $a, b \in R$ :

- $\varphi(0) = \tilde{0}$ ,
- $\varphi(1) = \tilde{1}$ ,
- $\varphi(a \oplus b) = \varphi(a) \tilde{\oplus} \varphi(b)$ ,
- $\varphi(a \otimes b) = \varphi(a) \tilde{\otimes} \varphi(b)$ .

If there exists a semifield isomorphism between two semifields, they are called isomorphic.

**Proposition 1 (Some Semifields Isomorphism)**

- The root semifields  $R_p$  are isomorphic to the nonnegative linear semifield  $L_{\geq 0}$ , with the isomorphism  $\varphi_p : R_p \rightarrow L_{\geq 0}$  being  $\varphi_p(x) = x^p$ .
- The logarithmic semifields  $L_\mu$  are isomorphic to the nonnegative linear semifield  $L_{\geq 0}$ , with the isomorphism  $\varphi_\mu : L_\mu \rightarrow L_{\geq 0}$  being  $\varphi_\mu(x) = e^{\mu x}$ .
- The tropical max semifield  $T_+$  is isomorphic to the tropical min semifield  $T_-$ , with the isomorphism  $\varphi : T_+ \rightarrow T_-$  being  $\varphi(x) = -x$ .
- Informally, in the limit  $\mu \rightarrow \pm\infty$  the logarithmic semifields  $L_\mu$  “converge” to the tropical semifields  $T_\pm$ .

The above proposition shows that although we defined five semifields of interest, as listed in Definition 2, we are, in fact, only working with 2 non-isomorphic ones.

For the purpose of analysis, we endow the semifields with a metric. Normally, a linear structure  $X$  is endowed with a norm  $\|\cdot\| : X \rightarrow \mathbb{R}_{\geq 0}$ , and afterward, a metric  $d$  is defined through  $d(a, b) = \|a - b\|$ . This is not possible in our semifield setting as we do not necessarily have additive inverses (consider for example the tropical semifields).

**Definition 4 (Semifield Metric)** Let  $R$  be a semifield. A semifield metric  $\rho : R \times R \rightarrow \mathbb{R}_{\geq 0}$  is a metric such that for all  $a, b, c \in R$  we have:

$$\begin{aligned} \rho(c \oplus a, c \oplus b) &\leq \rho(a, b), \\ \rho(c \otimes a, c \otimes b) &= \rho(c, 0)\rho(a, b). \end{aligned}$$

These properties are direct generalizations of the common notions of translation invariance and absolute homogeneity. More importantly, they ensure that semifield addition  $\oplus$  and multiplication  $\otimes$  are continuous (w.r.t the metric).

**Definition 5 (Employed Semifield Metrics)**

- a) In the linear semifield  $L$  case, we use the metric  $\rho_L(a, b) = |a - b|$ .
- b) In the root semifields  $R_p$  case, we use the metric  $\rho_{R_p}(a, b) = |a^p - b^p|$ .
- c) In the logarithmic semifields  $L_\mu$  case, we use the metric  $\rho_{L_\mu}(a, b) = |e^{\mu a} - e^{\mu b}|$ .
- d) In the tropical max semifield  $T_+$  case, we use the metric  $\rho_{T_+}(a, b) = |e^a - e^b|$ .
- e) In the tropical min semifield  $T_-$  case, we use the metric  $\rho_{T_-}(a, b) = |e^{-a} - e^{-b}|$ .

The root and logarithmic semifield metrics are natural as they borrow the metric on the linear semifield  $L$  through the isomorphisms  $\varphi_p(x) = x^p$  and  $\varphi_\mu(x) = e^{\mu x}$ , see Proposition 1. Similarly, the tropical min and max semifield metrics relate by their isomorphism  $\varphi(x) = -x$ .

**Definition 6 (One-Dimensional Semifield)** Let  $R$  be a metric semifield. If  $R$  as a topological space (with the topology induced by the metric) is locally homeomorphic to one-dimensional Euclidean space, we say it is one-dimensional.

Just as mathematical rings and fields can be used to create modules and vector spaces, we define an analogous structure called a semimodule using semifields.

**Definition 7 (Semimodule)** Let  $R = (R, 0, 1, \oplus, \otimes)$  be a semifield. An  $R$ -semimodule  $V = (V, \oplus_V, \otimes_V, 0_V)$  over  $R$  is a set with a commutative and associative binary operation  $\oplus_V : V \times V \rightarrow V$  called addition, and another binary operation  $\otimes_V : R \times V \rightarrow V$  called (left) scalar multiplication, such that for all  $a, b \in R$  and  $u, v \in V$ :

$$\begin{aligned} v \oplus_V 0_V &= v, \\ 1 \otimes_V v &= v, \\ 0 \otimes_V v &= 0_V, \\ (a \oplus b) \otimes_V v &= (a \otimes_V v) \oplus_V (b \otimes_V v), \\ a \otimes_V (u \oplus_V v) &= (a \otimes_V u) \oplus_V (a \otimes_V v). \end{aligned}$$

We do not write the subscript  $V$  on the operations of a semimodule  $V$  from here on out, as is usual.

Now that we have semimodules we can speak of semifield linearity in its full generality. The notion of semifield linearity is totally analogous to the normal notion of linearity, therefore the name.

**Definition 8 (Semifield Linear)** Let  $V_1, V_2$  be two semimodules over the same semifield  $R$ . A mapping  $\varphi : V_1 \rightarrow V_2$  is called  $R$ -linear if for all  $a, b \in R$  and  $u, v \in V_1$  we have:

$$\varphi(a \otimes u \oplus b \otimes v) = a \otimes \varphi(u) \oplus b \otimes \varphi(v). \tag{6}$$

### 3.2 Functions, Measurability & Integration

The prototypical semimodule over a semifield is the space of all semifield-valued functions on a set.

**Definition 9** (Function Semimodule) Let  $R$  be a semifield. Consider the set  $F(\mathbb{R}^2, R)$  of all  $R$ -valued functions  $f : \mathbb{R}^2 \rightarrow R$ . The set  $F(\mathbb{R}^2, R)$  forms an  $R$ -semimodule under pointwise semifield addition and multiplication. The semimodule  $F(\mathbb{R}^2, R)$  is called the function semimodule over  $\mathbb{R}^2$ . More generally, any subsemimodule of  $F(\mathbb{R}^2, R)$  is also called a function semimodule over  $\mathbb{R}^2$ .

On the function semimodule  $F(\mathbb{R}^2, R)$ , we define the following natural  $R$ -linear domain transformation operators:

**Definition 10** (Operators on Function Semimodule)

- **Translation Operator:** For all translation vectors  $v \in \mathbb{R}^2$ , we define the translation operator  $\mathcal{T}_v$

$$(\mathcal{T}_v f)(x) := f(-v + x). \tag{7}$$

- **Rotoreflexion Operator:** For all orthonormal matrices  $Q \in \mathbb{R}^{2 \times 2}$ , we define the rotoreflexion operator  $\mathcal{R}_Q$

$$(\mathcal{R}_Q f)(x) := f(Q^{-1}x). \tag{8}$$

- **Scaling Operator:** For all scalings  $s \in \mathbb{R}_{>0}$ , we define the scaling operator  $\mathcal{S}_s$

$$(\mathcal{S}_s f)(x) := f\left(\frac{x}{s}\right). \tag{9}$$

- **Pointwise Operator:** For all  $\varphi$ , we define the pointwise operator  $\mathcal{P}_\varphi$

$$(\mathcal{P}_\varphi(f))(x) := \varphi(f(x)). \tag{10}$$

To avoid pathological cases, we introduce standard measure theoretical concepts.

**Definition 11** (Measurable Space & Set) Let  $(X, d)$  be a complete metric space. We equip the space  $X$  with the natural Borel sigma algebra  $B$  induced by the metric  $d$ . This turns  $X$  into a measurable space. A measurable set is any element of the Borel sigma algebra  $B$ .

With the above definition, we can turn both  $\mathbb{R}^2$  and any metric semifield  $R$  into a measurable space.

**Definition 12** (Measurable Function) Let  $R$  be a metric semifield and  $f : \mathbb{R}^2 \rightarrow R$  a function. The function  $f$  is called a measurable function if the pre-image of any measurable set is a measurable set.

The set of measurable functions is broad enough to be well behaved under pointwise limits, as the following lemma describes.

**Lemma 1** Let  $R$  be a metric semifield, and let  $f(x) = \lim_{n \rightarrow \infty} f_n(x)$  be the pointwise limit of measurable functions  $f_n : \mathbb{R}^2 \rightarrow R$ . Then  $f$  is also measurable.

A proof of a generalization of this lemma can be found at [38]. But there exists an even stronger statement that describes measurable functions as pointwise limits of indicator functions simple functions, which are made from indicator functions.

**Definition 13** (Indicator Function) Let  $R$  be a semifield and  $A \subseteq \mathbb{R}^2$  any set. We define the indicator function  $\mathbb{1}_A$  of  $A$  as:

$$\mathbb{1}_A(x) = \begin{cases} \mathbb{1} & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

**Definition 14** (Simple Function) Let  $R$  be a semifield. A simple function  $s : \mathbb{R}^2 \rightarrow R$  is a finite  $R$ -linear combination of indicator functions of measurable sets  $A_i$ .

$$s = \bigoplus_{i=1}^n a_i \otimes \mathbb{1}_{A_i}, \tag{12}$$

where each  $a_i \in R$ .

The link between measurable and simple functions is a follows.

**Lemma 2** Let  $R$  be a metric semifield and consider semifield-valued functions on  $\mathbb{R}^2$ . The pointwise limit of a sequence of simple functions is measurable. Every measurable function is the pointwise limit of a sequence of simple functions.

**Proof** The pointwise limit of a sequence of simple functions being measurable follows immediately from Lemma 1, as simple functions are measurable. Showing that every measurable function is the pointwise limit of simple functions goes via a straightforward construction.  $\square$

For every semifield, there is a natural associated class of functions. We would like to specify this class in an axiomatic sense. This is where the sum-approachable definition comes into play. It is a restriction of the well-known statement that “every measurable function is the limit of simple functions.”

**Definition 15** (Sum-Approachable) Let  $R$  be a metric semifield. A function  $f : \mathbb{R}^2 \rightarrow R$  is sum-approachable if there exist  $a_i \in R$  and  $A_i \subseteq \mathbb{R}^2$  open such that we have

$$f(x) = \lim_{n \rightarrow \infty} \bigoplus_{i=1}^n a_i \otimes \mathbb{1}_{A_i}(x). \tag{13}$$

The semimodule of all sum-approachable functions  $f : \mathbb{R}^2 \rightarrow R$  is denoted by  $S(\mathbb{R}^2, R)$ .

There are two differences between sum-approachable and measurable: we only consider open sets, not measurable sets, and we have a limit of a semifield sum of indicator functions, not just a limit.

A function being sum-approachable is more restrictive than one might think at first sight. The following lemma illustrates this by showing that in the tropical cases the sum-approachable functions enjoy the property of being semicontinuous, something that does *not* happen in the linear case.

**Lemma 3** *A sum-approachable function  $f : \mathbb{R}^2 \rightarrow T_+$  is lower semicontinuous. A sum-approachable function  $f : \mathbb{R}^2 \rightarrow T_-$  is upper semicontinuous.*

**Proof** Consider the tropical max semifield case for the moment. Every indicator function  $\mathbb{1}_A(x)$  with  $A \subseteq \mathbb{R}^2$  open is lower semicontinuous in this case. The limit semifield sum in the definition of sum-approachable turns into a pointwise supremum in this case. The pointwise supremum of lower semicontinuous functions is again lower semicontinuous.<sup>1</sup> Thus, every sum-approachable function  $f : \mathbb{R}^2 \rightarrow T_+$  is lower semicontinuous. *Mutatis mutandis*, the exact same argument holds in the tropical min semifield case.  $\square$

**Definition 16** (Semifield Measure) Let  $\Sigma$  be the Borel sigma algebra on  $\mathbb{R}^2$  and  $R = (R, 0, \mathbb{1}, \oplus, \otimes)$  a semifield. A semifield measure  $\mu : \Sigma \rightarrow R$  is a mapping that satisfies the following properties.

• **Nullity of Empty Set:**

$$\mu(\emptyset) = 0. \tag{14}$$

• **Disjoint Additivity:** For all disjoint sets  $A, B \in \Sigma$ :

$$\mu(A \cup B) = \mu(A) \oplus \mu(B), \tag{15}$$

which we extend to countable collections of pairwise disjoint sets.

• **Unity of Unit Square:**

$$\mu([0, 1]^2) = \mathbb{1}. \tag{16}$$

• **Translation Invariance:** For all  $A \in \Sigma$  and  $v \in \mathbb{R}^2$ :

$$\mu(A + v) = \mu(A). \tag{17}$$

<sup>1</sup> Let  $f(x) = \sup_n f_n(x)$ . Let  $\varepsilon > 0$  and  $x_0 \in \mathbb{R}^2$ . Choose  $N$  such that  $f_N(x_0) > f(x_0) - \varepsilon/2$ . Choose  $\delta > 0$  such that  $f_N(x) > f_N(x_0) - \varepsilon/2$  when  $|x - x_0| < \delta$ . Then  $f(x) > f_N(x) > f_N(x_0) - \varepsilon/2 > f(x_0) - \varepsilon$  [39].

• **Rotoreflexion Invariance:** For all  $A \in \Sigma$  and all orthonormal matrices  $Q \in \mathbb{R}^{2 \times 2}$ :

$$\mu(QA) = \mu(A). \tag{18}$$

• **Scaling Equivariance:** There exist a group homomorphism  $\chi : (\mathbb{R}_{>0}, \times) \rightarrow (R \setminus \{0\}, \otimes)$  such that for all scalings  $s \in \mathbb{R}_{>0}$  and all  $A \in \Sigma$ :

$$\mu(sA) = \chi(s) \otimes \mu(A). \tag{19}$$

**Definition 17** (Employed Semifield Measure)

- a) In the linear semifield  $L$  case, we use standard Lebesgue measure  $\lambda$ .  $\mu_L(A) = \lambda(A)$ . The scaling factor is  $\chi(s) = s^2$ .
- b) In the root semifields  $R_p$  cases, we use  $\mu_{R_p}(A) = \sqrt[p]{\lambda(A)}$ . The scaling factor is  $\chi(s) = \sqrt[p]{s^2}$ .
- c) In the logarithmic semifields  $L_\mu$  cases, we use  $\mu_{L_\mu}(A) = \frac{1}{\mu} \ln \lambda(A)$ . The scaling factor is  $\chi(s) = \frac{1}{\mu} \ln s^2$ .
- d) In the tropical max semifield  $T_+$  case, we use  $\mu_{T_+}(A) = 0$ .<sup>2</sup> The scaling factor is  $\chi(s) = 0$ .
- e) In the tropical min semifield  $T_-$  case, we use  $\mu_{T_-}(A) = 0$ . The scaling factor is  $\chi(s) = 0$ .

**Definition 18** (Semifield Integration) Let  $R$  be a metric semifield,  $S = S(\mathbb{R}^2, R)$  the space of sum-approachable functions, and  $\mu$  a semifield measure. Let  $\oint : (\text{dom}(\oint) \subseteq S) \rightarrow R$  be a functional with the following properties.

• **Semifield Linearity:** For all  $a, b \in R$  and  $f, g \in \text{dom}(\oint)$

$$\oint a \otimes f \oplus b \otimes g = a \otimes \left(\oint f\right) \oplus b \otimes \left(\oint g\right). \tag{20}$$

• **Indicator Function:** For all measurable sets  $A \subseteq \mathbb{R}^2$ , we have

$$\oint \mathbb{1}_A = \mu(A). \tag{21}$$

• **Translation Invariance:** For all  $v \in \mathbb{R}^2$  and  $f \in \text{dom}(\oint)$

$$\oint \mathcal{T}_v f = \oint f. \tag{22}$$

• **Rotoreflexion Invariance:** For all orthonormal matrices  $Q \in \mathbb{R}^{2 \times 2}$  and  $f \in \text{dom}(\oint)$ :

$$\oint \mathcal{R}_Q f = \oint f. \tag{23}$$

<sup>2</sup> Remember that the semifield one  $\mathbb{1}$  in the tropical max semifield  $T_+$  case is 0 (Definition 2).

- **Scaling Equivariance:** For all scalings  $s \in \mathbb{R}_{>0}$  and  $f \in \text{dom}(\oint)$ :

$$\oint \mathcal{S}_s f = \chi(s) \otimes \oint f, \tag{24}$$

where  $\chi(s)$  is the scaling of the semifield measure  $\oint$  (Definition 16).

- **Fubini:** For all  $f : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow R$  with both  $f(\cdot, y), f(x, \cdot) \in \text{dom}(\oint)$ , if one of the following integrals exists, then they are equal:

$$\oint_y \oint_x f(x, y) = \oint_x \oint_y f(x, y). \tag{25}$$

We say that such a functional is a semifield integration. A function  $f$  that is in the domain of the semifield integration is called integrable.

To emphasize over what slot we are integrating, we may also write  $\oint_{x \in \mathbb{R}^2} f(x) = \oint f$ . To emphasize over what semifield  $R$  the integration is taking place, we may also write  $\oint = \oint^R$ .

The first two properties of the semifield integration essentially nail down what the integration has to be. That is, for every simple function  $s(x) = \bigoplus_{i=1}^n a_i \otimes \mathbb{1}_{A_i}(x)$  we have

$$\oint s = \bigoplus_{i=1}^n a_i \otimes \mu(A_i) \tag{26}$$

by semifield linearity. This then extends naturally to sum-approachable functions  $f(x) = \lim_{n \rightarrow \infty} \bigoplus_{i=1}^n a_i \otimes \mathbb{1}_{A_i}(x)$  by defining (with some caveats)

$$\oint f = \lim_{n \rightarrow \infty} \bigoplus_{i=1}^n a_i \otimes \mu(A_i). \tag{27}$$

The caveats here being that we need requirements on the exact nature of the sequence of simple functions for the above to be well defined. To not get bogged down into the details, we will just state what integration we will use for our relevant semifields, together with their domain of definition. In the case of the tropical semifields, we show in Appendix B that the upcoming semifield integration is indeed the correct one.

**Definition 19** (Employed Semifield Integration)

- a) In the linear semifield  $L$  case, we use standard Lebesgue integration

$$\oint^L f = \int_{x \in \mathbb{R}^2} f(x) dx. \tag{28}$$

The domain  $\text{dom}(\oint^L)$  is the space of Lebesgue integrable functions.

- b) In the root semifields  $R_p$  cases, we use

$$\oint^{R_p} f = \sqrt[p]{\int_{x \in \mathbb{R}^2} f(x)^p dx}. \tag{29}$$

The domain  $\text{dom}(\oint^{R_p})$  consists of all functions  $f$  such that  $f^p$  is Lebesgue integrable.

- c) In the logarithmic semifields  $L_\mu$  cases, we use

$$\oint^{L_\mu} f = \frac{1}{\mu} \ln \int_{x \in \mathbb{R}^2} e^{\mu f(x)} dx. \tag{30}$$

The domain  $\text{dom}(\oint^{L_\mu})$  consists of all functions  $f$  such that  $e^{\mu f}$  is Lebesgue integrable.

- d) In the tropical max semifield  $T_+$  case, we use the supremum sup.

$$\oint^{T_+} f = \sup_{x \in \mathbb{R}^2} f(x). \tag{31}$$

The domain  $\text{dom}(\oint^{T_+})$  consists of all functions  $f$  that are bounded from above.

- e) In the tropical min semifield  $T_-$  case, we use the infimum inf.

$$\oint^{T_-} f = \inf_{x \in \mathbb{R}^2} f(x). \tag{32}$$

The domain  $\text{dom}(\oint^{T_-})$  consists of all functions  $f$  that are bounded from below.

The logarithmic and root semifield integration is natural as these semifields are isomorphic to the linear semifield, see Proposition 1. Additionally, the tropical max and min semifield integration is related through their isomorphism  $\varphi(x) = -x$ . Indeed, one has  $\sup_{s \in S} s = -\inf_{s \in S} \{-s\}$ .

**Definition 20** (Semifield Convolution) Let  $R$  be a metric semifield. We define the semifield convolution  $\circledast$  of two integrable functions  $f, g \in \text{dom}(\oint)$  as the new function  $f \circledast g \in \text{dom}(\oint)$ :

$$(f \circledast g)(x) := \oint_{y \in \mathbb{R}^2} f(x - y) \otimes g(y). \tag{33}$$

Showing that  $f \circledast g$  is indeed in  $\text{dom}(\oint)$  is an immediate consequence of the Fubini property of semifield integration (25). Moreover, the Fubini property gives us that the semifield convolution is associative:

$$f \circledast (g \circledast h) = (f \circledast g) \circledast h, \tag{34}$$

and the translation invariance of semifield integration together with the commutativity of the semifield multiplication gives us that the semifield convolution is commutative.

We want to perform some analysis in our function spaces, so we need a (pseudo)metric  $\delta : S(\mathbb{R}^2, R) \times S(\mathbb{R}^2, R) \rightarrow \mathbb{R}_{\geq 0}$  (possibly with a restricted domain). Similarly as before, when we introduced a metric on the semifields, we cannot make due with a norm on the function space as we have no additive inverses to turn the norm into a metric.

**Definition 21 (Function Pseudometric)** Let  $R$  be a semifield with metric  $\rho$  and  $S = S(\mathbb{R}^2, R)$  the space of sum-approachable functions. A function (pseudo)metric  $\delta : S \times S \rightarrow \mathbb{R}_{\geq 0} \cup \{\infty\}$  is a (pseudo)metric such that for all  $f, g, h \in S$  and  $a \in R$  we have:

$$\begin{aligned} \delta(h \oplus f, h \oplus g) &\leq \delta(f, g), \\ \delta(a \otimes f, a \otimes g) &= \rho(a, 0)\delta(f, g). \end{aligned}$$

We allow for the (pseudo)metric to return  $\infty$ .

Again, just as in Definition 4, these properties are generalizations of the common notions of translation invariance and absolute homogeneity, and they ensure that both function addition  $\oplus : S \times S \rightarrow S$  and function scalar multiplication  $\otimes : R \times S \rightarrow S$  are continuous (in both slots).

**Definition 22 (Employed Function Pseudometric)**

a) In the linear semifield  $L$  case, we use

$$\delta_L(f, g) = \sqrt{\int_{\mathbb{R}^2} |f(x) - g(x)|^2 dx}. \tag{35}$$

b) In the root semifield  $R_p$  case, we use

$$\delta_{R_p}(f, g) = \sqrt{\int_{\mathbb{R}^2} |f(x)^p - g(x)^p|^2 dx}. \tag{36}$$

c) In the logarithmic semifields  $L_\mu$  case, we use

$$\delta_{L_\mu}(f, g) = \sqrt{\int_{\mathbb{R}^2} |e^{\mu f(x)} - e^{\mu g(x)}|^2 dx}. \tag{37}$$

d) In the tropical max semifield  $T_+$  case, we use

$$\delta_{T_+}(f, g) = \sup_{x \in \mathbb{R}^2} |e^{f(x)} - e^{g(x)}|. \tag{38}$$

e) In the tropical min semifield  $T_-$  case, we use

$$\delta_{T_-}(f, g) = \sup_{x \in \mathbb{R}^2} |e^{-f(x)} - e^{-g(x)}|. \tag{39}$$

Using the function (pseudo)metric, we can make an appropriate function space:

**Definition 23 (Metric Function Space)** Let  $R$  be a metric semifield,  $S = S(\mathbb{R}^2, R)$  the space of sum-approachable functions, and  $\delta : S \times S \rightarrow \mathbb{R}_{\geq 0} \cup \{\infty\}$  a function (pseudo)metric. The function (pseudo)metric space  $H = H(\mathbb{R}^2, R, \delta)$  is defined as

$$H := \{f \in S \mid \delta(0, f) < \infty\}. \tag{40}$$

To turn it into an actual metric space, we need to identify elements using the following natural equivalence relation  $\sim$ .

$$f \sim g \iff \delta(f, g) = 0. \tag{41}$$

This function space will be denoted with  $\mathcal{H} = H / \sim$ .

**Definition 24 (Employed Function Spaces)**

a) In the linear semifield  $L$  case, we have

$$\mathcal{H}_L = \mathbb{L}^2(\mathbb{R}^2). \tag{42}$$

b) In the root semifield  $R_p$  case, we have

$$\mathcal{H}_{R_p} = \{f : \mathbb{R}^2 \rightarrow R_p \mid e^{\mu f} \in \mathbb{L}^2(\mathbb{R}^2)\}. \tag{43}$$

c) In the logarithmic semifield  $L_\mu$  case, we have

$$\mathcal{H}_{L_\mu} = \{f : \mathbb{R}^2 \rightarrow L_\mu \mid f^p \in \mathbb{L}^2(\mathbb{R}^2)\}. \tag{44}$$

d) In the tropical max semifield  $T_+$  case, we have

$$\mathcal{H}_{T_+} = \{f : \mathbb{R}^2 \rightarrow T_+ \text{ l.s.c and b.f.a } \}, \tag{45}$$

where l.s.c means lower semicontinuous and b.f.a means bounded from above.

e) In the tropical min semifield  $T_-$  case we have

$$\mathcal{H}_{T_-} = \{f : \mathbb{R}^2 \rightarrow T_- \text{ u.s.c and b.f.b } \}, \tag{46}$$

where u.s.c means upper semicontinuous and b.f.b means bounded from below.

### 3.3 Fourier Transform

We assume the existence of an injective Fourier transform that need only work on a very restricted class of semifield integrable functions.

**Definition 25 (Semifield Fourier Transform)** Let  $R$  be a metric semifield. A semifield Fourier transform  $\mathcal{F}_R : (\text{dom}(\mathcal{F}_R) \subseteq \text{dom}(\mathcal{f})) \rightarrow \text{dom}(\mathcal{f})$  is an operator satisfying (where we the drop the subscript  $R$  for conciseness):

- **Semifield Linearity:** For all  $a, b \in R$  and  $f, g \in \text{dom}(\mathcal{F})$

$$\mathcal{F}(a \otimes f \oplus b \otimes g) = a \otimes (\mathcal{F}f) \oplus b \otimes (\mathcal{F}g). \tag{47}$$

- **Convolution Property:** For all  $f, g \in \text{dom}(\mathcal{F})$  with  $f \otimes g \in \text{dom}(\mathcal{F})$

$$\mathcal{F}(f \otimes g) = (\mathcal{F}f) \otimes (\mathcal{F}g). \tag{48}$$

- **Rotoreflexion Equivariance:** For all orthonormal matrices  $Q \in \mathbb{R}^{2 \times 2}$

$$\mathcal{F} \circ \mathcal{R}_Q = \mathcal{R}_{Q^{-T}} \circ \mathcal{F}. \tag{49}$$

- **Scaling Equivariance:** For all scalings  $s \in \mathbb{R}_{>0}$ :

$$\mathcal{F} \circ \mathcal{S}_s = \chi(s) \otimes \mathcal{S}_{1/s} \circ \mathcal{F}, \tag{50}$$

where  $\chi(s)$  is the scaling of the semifield measure (Definition 16).

- **Invertibility:** The domain  $\text{dom}(\mathcal{F})$  is chosen such that the transform is injective and thus invertible on its image.

In the next definition, we will specify the choice of semifield Fourier transform together with its appropriate choice of domain for all the semifields we consider (Definition 2). The choices we make here are sometimes more restrictive than strictly needed, but, as we will see in Sect. 5.2, we only need to be able to take the semifield Fourier transform of a very “small” set of functions.

**Definition 26** (Employed Semifield Fourier Transform)

- a) In the linear semifield  $L$  case, we use

$$(\mathcal{F}_L f)(\omega) = \int_{\mathbb{R}^2} f(x) e^{-i\omega \cdot x} dx. \tag{51}$$

The domain  $\text{dom}(\mathcal{F}_L)$  is chosen to be the space of even, continuous, and absolutely integrable functions, with absolutely integrable Fourier transforms. The inverse on its image is

$$(\mathcal{F}_L^{-1} \hat{f})(x) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} \hat{f}(\omega) e^{i\omega \cdot x} d\omega. \tag{52}$$

- b) In the root semifield  $R_p$  case, we use

$$(\mathcal{F}_{R_p} f)(\omega) = \sqrt[p]{\int_{\mathbb{R}^2} f(x)^p e^{-i\omega \cdot x} dx}. \tag{53}$$

The domain  $\text{dom}(\mathcal{F}_{R_p})$  is chosen such that  $f^p$  is in the domain of the linear Fourier transform  $\text{dom}(\mathcal{F}_L)$ , together

with the restriction that the input of the  $p$ 'th root is non-negative. The inverse on its image is

$$(\mathcal{F}_{R_p}^{-1} \hat{f})(x) = \sqrt[p]{\frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} \hat{f}(\omega)^p e^{i\omega \cdot x} d\omega}. \tag{54}$$

- c) In the logarithmic semifield  $L_\mu$  case, we use

$$(\mathcal{F}_{L_\mu} f)(\omega) = \frac{1}{\mu} \ln \int_{x \in \mathbb{R}^2} e^{\mu f(x)} e^{-i\omega \cdot x} dx. \tag{55}$$

The domain  $\text{dom}(\mathcal{F}_{L_\mu})$  is chosen such that  $e^{\mu f}$  is in the domain of the linear Fourier transform  $\text{dom}(\mathcal{F}_L)$ , together with the restriction that the input of the natural logarithm is positive. The inverse on its image is

$$(\mathcal{F}_{L_\mu}^{-1} \hat{f})(x) = \frac{1}{\mu} \ln \left( \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} e^{\mu \hat{f}(\omega)} e^{i\omega \cdot x} d\omega \right). \tag{56}$$

- d) In the tropical max semifield  $T_+$  case, we use

$$(\mathcal{F}_{T_+} f)(\omega) = \sup_{x \in \mathbb{R}^2} f(x) - \omega \cdot x. \tag{57}$$

The domain  $\text{dom}(\mathcal{F}_{T_+})$  is chosen to be the space of even, continuous, concave, superlinear functions. The inverse on its image is

$$(\mathcal{F}_{T_+}^{-1} \hat{f})(x) = \inf_{\omega \in \mathbb{R}^2} \hat{f}(\omega) + \omega \cdot x. \tag{58}$$

- e) In the tropical min semifield  $T_-$  case, we use

$$(\mathcal{F}_{T_-} f)(\omega) = \inf_{x \in \mathbb{R}^2} f(x) - \omega \cdot x. \tag{59}$$

The domain  $\text{dom}(\mathcal{F}_{T_-})$  is chosen to be the space of even, continuous, convex, superlinear functions. The inverse on its image is

$$(\mathcal{F}_{T_-}^{-1} \hat{f})(x) = \sup_{\omega \in \mathbb{R}^2} \hat{f}(\omega) + \omega \cdot x. \tag{60}$$

A proof that these transforms satisfy the definition is found in Appendix A.

**Remark 1** The Laplace-like transform

$$(\mathcal{L}f)(\omega) = \int_{\mathbb{R}^2} f(x) e^{-\omega \cdot x} dx \tag{61}$$

also satisfies Definition 25. However, and this is also mentioned in [20], this transform is limited in its applicability

because it is only finitely valued for functions with super-exponential decay.<sup>3</sup> Given this limitation of this transform, we instead use the normal Fourier transform.

**Remark 2** The above semifield Fourier transforms typically relate to transforms of the type

$$(\mathcal{F}f)(\omega) = \oint f(x) \otimes \chi_\omega(x),$$

where  $\chi_\omega$  is an irreducible semifield linear representation of  $\mathbb{R}^2$ , but we choose to express them in common Fourier/Fenchel transforms to keep a clear track of function space restrictions.

Even though we have used complex numbers in the Fourier transforms, the resulting transformed functions are always of the proper form  $\mathbb{R}^2 \rightarrow R$  due the domain consisting of even functions. This means that we could have freely replaced the  $e^{-i\omega \cdot x}$  with  $\cos(\omega \cdot x)$ . In other words, we could have instead used the *Fourier cosine transform*.

### 4 Semifield Scale-space

In this section, we will state and motivate the semifield scale-space axioms, consider some examples semifield scale-spaces, and define what we mean with isomorphic scale-spaces.

#### 4.1 Axioms

In [8], it is stated that “The only really nontrivial (and possibly too restrictive) assumption imposed on the scale-space operators is that of linearity.” By generalizing to semifield linearity, we sidestep this restrictive assumption, without making the theory too abstract to be practically useful.

Let us shortly motivate the semifield scale-space axioms from a machine learning perspective (building upon similar findings in mathematical deep learning [34–37]). The semifield linearity (Axiom 1) together with the translation equivariance (Axiom 5) will induce semifield convolutions that allow for fast parallel computations. The one-parameter semigroup property (Axiom 2) together with the strong continuity (Axiom 3) provides consistency and stability over evolution time. The one-parameter semigroup property (Axiom 2) together with the scaling equivariance (Axiom 4) allows us to constrain ourselves to a fixed end-time in a PDE sublayer, say  $t = 1$ , without loss of generality, further reducing the total parameter count. The scaling, translation,

and rotoreflection equivariance (Axioms 4, 5 and 6) of the scale-space allows for the design of inherently equivariant networks, resulting in an architecture that is robust and data efficient [6, 7].

**Definition 27** (Semifield Scale-space) Let  $R$  be a one-dimensional metric semifield (Definition 6),  $\mathcal{H} = \mathcal{H}(\mathbb{R}^2, R, \delta)$  a corresponding metric function space (Definition 23), and  $\Phi_t : \mathcal{H} \rightarrow \mathcal{H}$  be a family of operators, indexed by  $t \geq 0$ . We call  $\Phi_t$  a semifield scale-space if it satisfies the following axioms:

1. **Semifield Linearity and Integral Operator:** We require that  $\Phi$  is  $R$ -linear, that is, for all  $f, g \in \mathcal{H}$  and  $a, b \in R$ :

$$\Phi_t(a \otimes f \oplus b \otimes g) = a \otimes (\Phi_t f) \oplus b \otimes (\Phi_t g). \tag{62}$$

More specifically, for positive time  $t > 0$  we will assume that  $\Phi_t$  can be written as an integral operator:

$$(\Phi_t f)(x) = \oint_{y \in \mathbb{R}^2} \kappa_t(x, y) \otimes f(y), \tag{63}$$

for some continuous kernel  $\kappa_t : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow R$  with  $\kappa_t(x, \cdot)$  within the domain of the semifield Fourier transform  $\mathcal{F}_R$  (Definition 25).

2. **One-Parameter Semigroup:** We require that  $\Phi_t$  forms a one-parameter semigroup, that is, for all  $t, s \geq 0$ :

$$\Phi_t \circ \Phi_s = \Phi_{t+s} \text{ and } \Phi_0 = \text{id}, \tag{64}$$

where  $\text{id}$  is the identity map on  $\mathcal{H}$ .

3. **Strong Continuity:** We require that  $\Phi_t f$  is continuous w.r.t. time  $t$  at any  $t_0 > 0$  for all  $f \in \mathcal{H}$ :

$$\lim_{t \rightarrow t_0} (\Phi_t f) = \Phi_{t_0} f, \tag{65}$$

where the limit is taken in the metric function space  $\mathcal{H}$  (Definition 23).

4. **Scaling Equivariance:** There exists a *scaling power*  $\alpha > 0$  such that for all scalings  $s > 0$  and all times  $t \geq 0$ :

$$\Phi_t \circ \mathcal{S}_s = \mathcal{S}_s \circ \Phi_{t/s^\alpha}, \tag{66}$$

where  $\mathcal{S}_s$  is the scaling operator (9).

5. **Translation Equivariance:** We require that  $\Phi$  commutes with all translations  $v \in \mathbb{R}^2$ :

$$\Phi_t \circ \mathcal{T}_v = \mathcal{T}_v \circ \Phi_t, \tag{67}$$

where  $\mathcal{T}_v$  is the translation operator (7).

<sup>3</sup> This Laplace transform is two-sided thus resulting in this extreme condition. Also, we do not regard the transform as a conditionally convergent improper integral.

6. **Rotoreflexion Equivariance:** We require that  $\Phi$  commutes with all orthonormal matrices  $Q \in \mathbb{R}^{2 \times 2}$ :

$$\Phi_t \circ \mathcal{R}_Q = \mathcal{R}_Q \circ \Phi_t, \tag{68}$$

where  $\mathcal{R}_Q$  is the rotoreflexion operator (8).

Note that in our axioms we do not impose any restriction on the creation of new structures when transitioning from finer to coarser scales. This is in contrast with the requirement of *causality* or *non-enhancement of local extrema* in [16, 40].

In the linear semifield  $L$  case, the linearity axiom in some sense already implies the integral operator axiom. The precise statement is known as the Schwartz kernel theorem, a main result in the theory of generalized functions/distributions. In the tropical semifield case, a similar statement can be made, as demonstrated in [41, Thm.2.1]. But for other semifields, such a statement cannot be made just yet. For simplicity, and to be on the safe side, we therefore assume the integral operator axiom.

The one-parameter semigroup property is a natural axiom in the sense that it implies that the (infinitesimal) evolution “looks the same” at all times  $t$ . More precisely, the (strongly continuous) one-parameter semigroup property relates to the existence of a single generator that encapsulates the whole operator family. To understand, consider the linear semifield case, some initial  $f_0 \in \mathcal{H}$ , and its evolution  $f_t := \Phi_t(f_0)$ . From the one-parameter semigroup axiom, we have:

$$\begin{aligned} f_{t+h} - f_t &= \Phi_{h+t} f_0 - \Phi_t f_0 \\ &= \Phi_h \Phi_t f_0 - \Phi_t f_0 \\ &= (\Phi_h - \Phi_0) f_t. \end{aligned} \tag{69}$$

dividing by  $h$  and taking the limit  $h \downarrow 0$  in conjunction with strong continuity, we get the time-invariant evolution equation:

$$\frac{df_t}{dt} = \Psi f_t, \text{ where } \Psi := \lim_{h \downarrow 0} \frac{\Phi_h - \Phi_0}{h}. \tag{70}$$

The operator  $\Psi : D(\Psi) \rightarrow \mathcal{H}$  is called the generator of the operator family  $\Phi_t$  and its natural domain  $D(\Psi)$  consists of all functions  $f \in \mathcal{H}$  for which the above limit makes sense. Typically, this domain will be dense in  $\mathcal{H}$ .

Thus, we can interpret  $\Phi_t$  as the solution operator of an evolution equation. Given that the generator exists, it is possible through various means, for example, the spectral theorem [42], to give meaning to the expression:

$$\Phi_t = e^{t\Psi}, \tag{71}$$

which can be used to quickly confirm (at least formally) that:

$$\frac{d\Phi_t}{dt} = \frac{d}{dt}(e^{t\Psi}) = \Psi e^{t\Psi} = \Psi \Phi_t, \tag{72}$$

which corresponds what we already saw in (70).

**Remark 3** Given the existence of a generator  $\Psi$ , the scaling equivariance axiom can be equivalently written as  $\Psi \circ \mathcal{S}_s = \frac{1}{s^\alpha} \mathcal{S}_s \circ \Psi$ , revealing that the scaling equivariance can also be understood as a sort of  $\alpha$ -homogeneity of the generator.

An easy and illustrative example of a generator together with its operator family is the derivative operator and the family of translation operators in one-dimensional space:

$$(\Phi_t f)(x) = f(x + t), \quad \Psi = \frac{d}{dx}. \tag{73}$$

A well-known related theorem in functional analysis is Stone’s Theorem. This theorem shows that there is a one-to-one correspondence between strongly continuous unitary one-parameter semigroups and (possibly unbounded) densely defined self-adjoint operators on a Hilbert space.

In our case, getting everything precise is made difficult by the fact that we want to generalize to semifields other than the linear semifield. For example, we cannot even directly make sense of (70) for general semifields as there is not necessarily a operation: we only have  $\oplus$ . Given these obstacles, we will not attempt to rigorously prove that every semifield scale-spaces corresponds to a PDE, but will state the related PDEs in our primary cases of interest (Definition 28).

The scaling equivariance says that the scale-space representation of a scaled image should be a scaled version of the scale-space representation of the original image. In a sense we want a scale-space that does not “care” about absolute scale: it should qualitatively looks the same no matter the starting scale of the input. The translation and rotoreflexion equivariance requirements are also not surprising: the Euclidean plane has its natural translation and rotoreflexion symmetries, and demanding the scale-space to respect these is commonplace.

In PDE-based neural networks, the trainable parameters take the form of Riemannian metrics  $\mathcal{G}$  on a homogeneous space  $M$ . In the case of PDE-CNNs that being  $M = \mathbb{R}^2$ , this reduces to an inner product  $\mathcal{G} : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$  which we can always write as  $\mathcal{G}(x, y) = x^\top G y$ , where  $G$  is the corresponding *Gram matrix*. In the axioms, we implicitly make, without loss of generality, the assumption that we use the “standard” inner product on  $\mathbb{R}^2$ , namely  $\mathcal{G}(x, y) = x^\top y$ . Later in Sect. 6 we explain how we bring back general inner products in the PDE-CNN architecture.

### 4.2 Examples

**Definition 28** (Scale-spaces of Interest)

a) The Gaussian scale-space over the linear semifield  $L$ :

$$(\Phi_t f)(x) = \int_{y \in \mathbb{R}^2} \kappa_t(x, y) \times f(y) \, dy,$$

$$\kappa_t(x, y) = \frac{1}{2\pi t} \exp\left(-\frac{1}{2} \frac{\|x - y\|^2}{t}\right),$$

which correspond to solutions of

$$\frac{\partial f}{\partial t} = \frac{1}{2} \Delta f. \tag{74}$$

The scaling power is  $\alpha = 2$ .

b) The (quadratic) root scale-spaces over the root semifields  $R_p$ :

$$(\Phi_t f)(x) = \oint_{y \in \mathbb{R}^2}^{R_p} \kappa_t(x, y) \times f(y),$$

$$\kappa_t(x, y) = \frac{1}{\sqrt[p]{2\pi t}} \exp\left(-\frac{1}{2p} \frac{\|x - y\|^2}{t}\right),$$

which correspond to solutions of

$$\frac{\partial f}{\partial t} = \frac{p-1}{f} \frac{1}{2} \|\nabla f\|^2 + \frac{1}{2} \Delta f. \tag{75}$$

The scaling power is  $\alpha = 2$ .

c) The (quadratic) logarithmic scale-spaces over the logarithmic semifields  $L_\mu$ :

$$(\Phi_t f)(x) = \oint_{y \in \mathbb{R}^2}^{L_\mu} \kappa_t(x, y) + f(y),$$

$$\kappa_t(x, y) = -\frac{1}{\mu} \ln(2\pi t) - \frac{1}{2\mu} \frac{\|x - y\|^2}{t},$$

which correspond to solutions of

$$\frac{\partial f}{\partial t} = \mu \frac{1}{2} \|\nabla f\|^2 + \frac{1}{2} \Delta f. \tag{76}$$

The scaling power is  $\alpha = 2$ .

d) The  $\alpha$ -dilation scale-space over the tropical max semifield  $T_+$ :

$$(\Phi_t f)(x) = \sup_{y \in \mathbb{R}^2} \kappa_t(x, y) + f(y),$$

$$\kappa_t(x, y) = -\frac{t}{\beta} \left(\frac{\|x - y\|}{t}\right)^\beta,$$

with  $1/\alpha + 1/\beta = 1$ , which correspond to (viscosity) solutions of

$$\frac{\partial f}{\partial t} = \frac{1}{\alpha} \|\nabla f\|^\alpha. \tag{77}$$

The scaling power is  $\alpha$ .

e) The  $\alpha$ -erosion scale-space over the tropical min semifield  $T_-$ :

$$(\Phi_t f)(x) = \inf_{y \in \mathbb{R}^2} \kappa_t(x, y) + f(y),$$

$$\kappa_t(x, y) = \frac{t}{\beta} \left(\frac{\|x - y\|}{t}\right)^\beta,$$

with  $1/\alpha + 1/\beta = 1$ , which correspond to (viscosity) solutions of

$$\frac{\partial f}{\partial t} = -\frac{1}{\alpha} \|\nabla f\|^\alpha. \tag{78}$$

The scaling power is  $\alpha$ .

The operators  $\Phi_t$  above solve the corresponding PDEs and one readily checks that the kernels  $\kappa_t(\cdot, y)$  satisfy the PDE for all  $y \in \mathbb{R}^2$ . For example, consider the quadratic ( $\alpha = 2$ ) dilation scale-space and  $k_t(x) := \kappa_t(x, 0) = -\frac{1}{2} \frac{\|x\|^2}{t}$ :

$$\frac{\partial k_t}{\partial t} = \frac{1}{2} \frac{\|x\|}{t^2}, \quad \|\nabla k_t\|^2 = \frac{\|x\|^2}{t^2}. \tag{79}$$

So, indeed,  $k_t$  satisfies the dilation PDE. The same check can be done for the other scale-spaces.

**Remark 4** The Schrödinger equation also generates a scale-space representation in the space  $\mathbb{L}^2(\mathbb{R}^2; \mathbb{C})$ , in the sense that it satisfies the linearity axiom and axioms 2-6. However, it does not fit in the theory here as the complex numbers do not form a one-dimensional semifield, and the corresponding kernel  $\kappa_t(x, \cdot)$  is not square integrable. In T. Kraakman's master's thesis [43] PDE-based neural networks using the Schrödinger equation are investigated and implemented. They require more memory than our classical PDE-Based CNNs for only a small accuracy gain in practice so far.

### 4.3 Isomorphic Scale-spaces

In Proposition 1, we saw that the nonnegative linear, root, and logarithmic semifields are isomorphic, with the same being true for the tropical ones. It seems natural then that there also exist isomorphisms between the corresponding semifield scale-spaces. Let us start by clarifying what we mean by two semifield scale-spaces being isomorphic.

**Definition 29** (Semifield Scale-space Isomorphism) Let  $R$  and  $\tilde{R}$  be two semifields, and let  $\Phi_t$  and  $\tilde{\Phi}_t$  be two semifield scale-spaces over  $R$  and  $\tilde{R}$ , respectively. We say the two scale-spaces are isomorphic if there exists a semifield isomorphism  $\varphi : \tilde{R} \rightarrow R$  such that

$$\Phi_t \circ \mathcal{P}_\varphi = \mathcal{P}_\varphi \circ \tilde{\Phi}_t, \tag{80}$$

where  $\mathcal{P}_\varphi : F(\mathbb{R}^2, \tilde{R}) \rightarrow F(\mathbb{R}^2, R)$  is the pointwise operator (10).

Indeed, one can check that in this sense the scale-spaces of interest are isomorphic in the following way, akin to Proposition 1.

**Proposition 2 (Scale-Space Isomorphisms)**

- The quadratic root scale-spaces over the root semifields  $R_p$  are isomorphic to the Gaussian scale-space over the nonnegative linear semifield  $L_{\geq 0}$ .
- The quadratic logarithmic scale-spaces over the logarithmic semifields  $L_\mu$  are isomorphic to the Gaussian scale-space over the nonnegative linear semifield  $L_{\geq 0}$ .
- The  $\alpha$ -dilation scale-space over the tropical max semifield  $T_+$  is isomorphic to the  $\alpha$ -erosion scale-space over the tropical min semifield  $T_-$ .
- Informally, in the limit  $\mu \rightarrow \pm\infty$  the quadratic logarithmic scale-spaces over the logarithmic semifields  $L_\mu$  “converge” to the quadratic ( $\alpha = 2$ ) dilation and erosion scale-spaces of the tropical semifields  $T_\pm$ .

Definition 29 also gives us a way to create new (isomorphic) semifield scale-spaces from existing ones in the following way. Take any existing semifield scale-space  $\Phi_t$  over a semifield  $R$ , and let  $\varphi : \tilde{R} \rightarrow R$  be any semifield isomorphism. We then simply define the new scale-space  $\tilde{\Phi}_t = \mathcal{P}_{\varphi^{-1}} \circ \Phi_t \circ \mathcal{P}_\varphi$  over the semifield  $\tilde{R}$ .

In [11], Florack creates nonlinear scale-spaces in exactly this way by performing a pointwise transformation on the Gaussian scale-space and deducing what nonlinear PDE corresponds to the obtained evolution.

More specifically, let  $\varphi : R \rightarrow L$  be a monotonic twice continuously differentiable transformation, where  $R \subseteq \mathbb{R}$  is some subset of the reals. We start with the isotropic diffusion PDE on  $\mathbb{R}^2$ :

$$\frac{\partial f}{\partial t} = \frac{1}{2} \Delta f. \tag{81}$$

We then define a new evolution  $g(x, t) := \varphi^{-1}(f(x, t))$ . Let us derive what PDE  $g$  obeys. Using the shorthand  $\varphi(g) := \mathcal{P}_\varphi(g)$  for a moment, it follows from the equalities:

$$\begin{aligned} \frac{\partial g}{\partial t} &= \frac{\partial}{\partial t} \varphi^{-1}(f) = \frac{1}{\varphi'(g)} \frac{\partial f}{\partial t}, \\ \Delta f &= \Delta(\varphi(g)) = \varphi''(g) \|\nabla g\|^2 + \varphi'(g) \Delta g, \end{aligned}$$

that the PDE that describes the evolution of  $g$  is:

$$\frac{\partial g}{\partial t} = \frac{\varphi''(g)}{\varphi'(g)} \frac{1}{2} \|\nabla g\|^2 + \frac{1}{2} \Delta g. \tag{82}$$

Florack suggests setting  $\mu := \varphi''/\varphi' = (\ln \varphi)'$  as a constant, as the class of nontrivial (that being non-affine)  $\varphi$ 's is then:

$$\varphi(x) = e^{\mu x} \text{ with } \mu \neq 0, \mu \in \mathbb{R}, \tag{83}$$

up to affine transformations. This transformation of the PDE is known as the *Cole-Hopf transformation* [26, p.195]. This is exactly the isomorphism between the logarithmic and linear semifield as seen in Proposition 1, and indeed, with this  $\varphi$  we get the quadratic logarithmic scales spaces:

$$\frac{\partial g}{\partial t} = \mu \frac{1}{2} \|\nabla g\|^2 + \frac{1}{2} \Delta g. \tag{84}$$

In the extreme cases of the transformation that being  $\mu = \pm\infty$ , the diffusion part becomes negligible in comparison with the erosion/dilation part, and one can say that the morphological scale-spaces arise.

If we instead choose  $\varphi(x) = x^p$ , the quadratic root scale-spaces arise:

$$\frac{\partial g}{\partial t} = \frac{p-1}{g} \frac{1}{2} \|\nabla g\|^2 + \frac{1}{2} \Delta g. \tag{85}$$

## 5 Consequences

In this section, we explore the consequences of the semifield scale-spaces axioms (Definition 27).

We start by showing that the equivariance axioms of the scale-space representation lead to invariance properties of the kernel  $\kappa_t : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow R$ . For example, the translation equivariance of  $\Phi_t$  (Axiom 5) implies that the kernel  $\kappa_t$  (Axiom 1) is translation invariant in the sense that  $\kappa_t(v+x, v+y) = \kappa_t(x, y)$  for all  $x, y, v \in \mathbb{R}^2$ .

We then show that, due to the translation equivariance (Axiom 5), the semifield scale-space can be written as a semifield convolution with a reduced kernel  $k_t : \mathbb{R}^2 \rightarrow R$ . From there on out, we show Theorem 1 which gives an explicit form of the reduced kernel  $k_t$  in the semifield Fourier domain, this being the main theorem of the article.

### 5.1 Equivariance of Operator becomes Invariance of Kernel

In this subsection, we show how the translation, roto-reflection, and scaling equivariance axioms on the operator family  $\Phi_t$  translate to corresponding invariances on the kernel  $\kappa_t$ . The upcoming three lemmas are straightforward, generally known, and basically identical in proof.

**Lemma 4 (Translation Invariance)** *From the integral operator (Axiom 1) and translation equivariance (Axiom 5), it*

follows that the kernel is translation invariant, that is:

$$\kappa_t(v + x, v + y) = \kappa_t(x, y), \tag{86}$$

for all for all  $x, y, v \in \mathbb{R}^2$  and  $t > 0$ .

**Proof** We rewrite the translation equivariance (Axiom 5) as

$$\mathcal{T}_{-v} \circ \Phi_t \circ \mathcal{T}_v = \Phi_t. \tag{87}$$

We apply some dummy function  $f \in \mathcal{H}$  and evaluate it at some dummy position  $x \in \mathbb{R}^2$ :

$$((\mathcal{T}_{-v} \circ \Phi_t \circ \mathcal{T}_v)(f))(x) = (\Phi_t(f))(x). \tag{88}$$

Using the definition of the translation operator  $\mathcal{T}_v$  (7) and the integral operator axiom, we expand this into:

$$\oint_{y \in \mathbb{R}^2} \kappa_t(v+x, y) \otimes f(-v+y) = \oint_{y \in \mathbb{R}^2} \kappa_t(x, y) \otimes f(y). \tag{89}$$

Using the translation invariance property (22) of the semifield integration gives:

$$\oint_{y \in \mathbb{R}^2} \kappa_t(v+x, v+y) \otimes f(y) = \oint_{y \in \mathbb{R}^2} \kappa_t(x, y) \otimes f(y). \tag{90}$$

Given that this should hold for all  $f \in \mathcal{H}$ , we can conclude:

$$\kappa_t(v + x, v + y) = \kappa_t(x, y). \tag{91}$$

□

The last step of the proof can be understood as a semifield version of *Fundamental lemma of the calculus of variations*. There are several versions of this lemma, but generally they are of the form

$$\forall g \int fg = 0 \Rightarrow f = 0, \tag{92}$$

with some assumptions on the nature of  $f$  and  $g$ . Using the substitution  $f = f_1 - f_2$ , we can equivalently write

$$\forall g \int f_1g = \int f_2g \Rightarrow f_1 = f_2. \tag{93}$$

The lemma in its second form (93) is used in the last step of the proof of Lemma 4, specifically when transitioning from (90) to (91). We will also apply it in the upcoming proofs of Lemma 5 and Lemma 6. The proof of the lemma is straightforward in our setting; the kernel  $\kappa_t$  is assumed to be continuous (Axiom 1), and the space  $\mathcal{H}$  contains all indicator

functions, meaning that a standard “concentration” argument works.

**Lemma 5 (Rotoreflexion Invariance)** *From the integral operator (Axiom 1) and the rotoreflexion equivariance (Axiom 6), it follows that the kernel is rotoreflexion invariant, that is:*

$$\kappa_t(Qx, Qy) = \kappa_t(x, y), \tag{94}$$

for all orthonormal  $Q \in \mathbb{R}^{2 \times 2}$ ,  $x, y \in \mathbb{R}^2$ , and  $t > 0$ .

**Proof** We rewrite the rotoreflexion equivariance (Axiom 6) as

$$\mathcal{R}_{Q^{-1}} \circ \Phi_t \circ \mathcal{R}_Q = \Phi_t. \tag{95}$$

We apply some dummy function  $f \in \mathcal{H}$  and evaluate it at some dummy position  $x \in \mathbb{R}^2$ :

$$((\mathcal{R}_{Q^{-1}} \circ \Phi_t \circ \mathcal{R}_Q)(f))(x) = (\Phi_t(f))(x). \tag{96}$$

Using the definition of the rotoreflexion operator  $\mathcal{R}_Q$  (8) and integral operator axiom, we expand this to:

$$\oint_{y \in \mathbb{R}^2} \kappa_t(Qx, -y) \otimes f(Q^{-1}y) = \oint_{y \in \mathbb{R}^2} \kappa_t(x, y) \otimes f(y). \tag{97}$$

Using the rotoreflexion invariance property (23) of the semifield integration gives:

$$\oint_{y \in \mathbb{R}^2} \kappa_t(Qx, Qy) \otimes f(y) = \oint_{y \in \mathbb{R}^2} \kappa_t(x, y) \otimes f(y). \tag{98}$$

Given that this should hold for all  $f \in \mathcal{H}$ , we can conclude:

$$\kappa_t(Qx, Qy) = \kappa_t(x, y). \tag{99}$$

□

**Lemma 6 (Scale Invariance)** *From the integral operator (Axiom 1) and scaling equivariance (Axiom 4), it follows that the kernel is scale invariant, that is:*

$$\chi(s) \otimes \kappa_{s^{\alpha_t}}(sx, sy) = \kappa_t(x, y). \tag{100}$$

for all  $x, y \in \mathbb{R}^2$  and  $s, t > 0$ .

**Proof** We rewrite the scaling equivariance (Axiom 4) as

$$\mathcal{S}_{1/s} \circ \Phi_{s^{\alpha_t}} \circ \mathcal{S}_s = \Phi_t. \tag{101}$$

We apply some dummy function  $f \in \mathcal{H}$  and evaluate it at some dummy position  $x \in \mathbb{R}^2$ :

$$((\mathcal{S}_{1/s} \circ \Phi_{s^{\alpha_t}} \circ \mathcal{S}_s)(f))(x) = (\Phi_t(f))(x). \tag{102}$$

Using the definition of the scaling operator  $\mathcal{S}_s$  (9) and integral operator axiom, we expand this to:

$$\oint_{y \in \mathbb{R}^2} \kappa_{s^{\alpha_t}}(sx, y) \otimes f(y/s) = \oint_{y \in \mathbb{R}^2} \kappa_t(x, y) \otimes f(y). \tag{103}$$

Using the scaling property (24) of the semifield integration gives:

$$\chi(s) \otimes \oint_y \kappa_{s^{\alpha_t}}(sx, sy) \otimes f(y) = \oint_y \kappa_t(x, y) \otimes f(y). \tag{104}$$

Given that this should hold for all  $f \in \mathcal{H}$ , we conclude:

$$\chi(s) \otimes \kappa_{s^{\alpha_t}}(sx, sy) = \kappa_t(x, y). \tag{105}$$

□

Consider now Lemma 4. Because we can freely choose the translation  $v$ , we can also choose  $v = -y$ :

$$\kappa_t(x, y) = \kappa_t(x - y, 0) =: k_t(x - y). \tag{106}$$

We thus see that  $\kappa_t$  is completely characterized by its behavior on  $\kappa_t(\cdot, 0)$ , which we define as the reduced kernel  $k_t : \mathbb{R}^2 \rightarrow R$ . Plugging this newfound knowledge back into the integral operator axiom, we get the following result.

**Lemma 7 (Translation Equivariance implies Semifield Convolution)** Consider the integral operator (Axiom 1) and translation equivariance (Axiom 5). Define the reduced kernel  $k_t(x) = \kappa_t(x, 0)$ . We can write the scale-space operator  $\Phi_t$  as a semifield convolution:

$$\Phi_t f = k_t \circledast f. \tag{107}$$

**Proof** This follows immediately from Lemma 4.

$$\begin{aligned} (\Phi_t f)(x) &= \oint_{y \in \mathbb{R}^2} \kappa_t(x, y) \otimes f(y) \\ &= \oint_{y \in \mathbb{R}^2} \kappa_t(x - y, 0) \otimes f(y) \\ &= \oint_{y \in \mathbb{R}^2} k_t(-y + x) \otimes f(y) \\ &= (k_t \circledast f)(x). \end{aligned} \tag{108}$$

□

**Lemma 8 (Convolution Property of Reduced Kernel)** From the integral operator (Axiom 1), the one-parameter semigroup (Axiom 2), the strong continuity (Axiom 3), and the translation equivariance (Axiom 5), it follows that the reduced kernel  $k_t(x) = \kappa_t(x, 0)$  satisfies:

$$k_s \circledast k_t = k_{s+t} \text{ for all } t, s > 0. \tag{109}$$

**Proof** We have already seen in Lemma 7 that the integral operator axiom and the translation equivariance axiom imply that

$$\Phi_t f = k_t \circledast f. \tag{110}$$

If we use this formula together with the one-parameter semigroup property axiom we get

$$k_s \circledast (k_t \circledast f) = k_{s+t} \circledast f. \tag{111}$$

Using associativity of semifield convolution (34) on the l.h.s.:

$$(k_s \circledast k_t) \circledast f = k_{s+t} \circledast f. \tag{112}$$

We are free to choose  $f = k_\varepsilon$  for  $\varepsilon > 0$ :

$$(k_s \circledast k_t) \circledast k_\varepsilon = k_{s+t} \circledast k_\varepsilon. \tag{113}$$

From the strong continuity axiom, we know that  $\lim_{\varepsilon \rightarrow 0} k_\varepsilon \circledast k_t = k_t$ , thus, after taking this limit, we can conclude:

$$k_s \circledast k_t = k_{s+t}. \tag{114}$$

□

### 5.2 Toward the Semifield Fourier Domain

We see that to move forward we need a way to efficiently work with semifield convolutions. The semifield Fourier transform (Definition 25) has the important property of turning convolutions into much more wieldy pointwise multiplication. This is why, we translate all previous lemmas to the semifield Fourier domain using the semifield Fourier transform.

**Lemma 9** Consider all axioms of a semifield scale-space (Definition 27). The reduced kernel  $k_t(x) := \kappa_t(\cdot, 0)$  in the semifield Fourier domain  $\hat{k}_t = \mathcal{F}_R(k_t)$  satisfies

$$\hat{k}_t(Q\omega) = \hat{k}_t(\omega), \tag{115}$$

$$\hat{k}_{s^{\alpha_t}}(\omega/s) = \hat{k}_t(\omega), \tag{116}$$

$$\hat{k}_s(\omega) \otimes \hat{k}_t(\omega) = \hat{k}_{s+t}(\omega), \tag{117}$$

□ for all orthonormal  $Q \in \mathbb{R}^{2 \times 2}$ ,  $\omega \in \mathbb{R}^2$ , and  $s, t > 0$ .

**Proof** Lemma 5 tells us that

$$\kappa_t(Qx, Qy) = \kappa_t(x, y). \tag{118}$$

Translating this to the reduced kernel  $k_t$  gives

$$k_t(Qx) = k_t(x). \tag{119}$$

Taking the semifield Fourier transform on both sides, and using the roto-reflection equivariance property of the Fourier transform (49), we get:

$$\hat{k}_t(Q\omega) = \hat{k}_t(\omega). \tag{120}$$

Lemma 6 also tells us that

$$\chi(s) \otimes \kappa_{s\alpha_t}(sx, sy) = \kappa_t(x, y). \tag{121}$$

Translating this to the reduced kernel  $k_t$  gives

$$\chi(s) \otimes k_{s\alpha_t}(sx) = k_t(x). \tag{122}$$

Taking the semifield Fourier transform on both sides, and using the scaling equivariance property of the Fourier transform (50), we get:

$$\chi(s) \otimes \chi(1/s) \otimes \hat{k}_{s\alpha_t}(\omega/s) = \hat{k}_t(\omega). \tag{123}$$

Using that  $\chi$  is a homomorphism we know that  $\chi(s) \otimes \chi(1/s) = \chi(s/s) = \chi(1) = \mathbb{1}$ , so we can simplify this to

$$\hat{k}_{s\alpha_t}(\omega/s) = \hat{k}_t(\omega). \tag{124}$$

Lemma 8 tells us that

$$k_s \otimes k_t = k_{s+t}. \tag{125}$$

Taking the semifield Fourier transform on both sides, and using the convolution property of the Fourier transform (48), we get:

$$\hat{k}_s \otimes \hat{k}_t = \hat{k}_{s+t}. \tag{126}$$

□

Let us check if the reduced kernels of the scale-spaces of interest indeed satisfy the properties listed in the previous lemma by inspecting their Fourier transforms.

**Proposition 3** (*Semifield Fourier Transform of Kernels of Interest*) Consider the employed semifield Fourier transforms (Definition 26) and the semifield scale-spaces of interest (Definition 28). Let  $k_t(x) := \kappa_t(\cdot, 0)$  be the reduced kernel and  $\hat{k}_t = \mathcal{F}_R(k_t)$  its semifield Fourier transform.

a) For the Gaussian scale-space over the linear semifield  $L$ , we have:

$$\hat{k}_t^L(\omega) = \exp\left(-\frac{1}{2}t\|\omega\|^2\right). \tag{127}$$

b) For the quadratic root scale-spaces over the root semifields  $R_p$ , we have:

$$\hat{k}_t^{R_p}(\omega) = \exp\left(-\frac{1}{2p}t\|\omega\|^2\right). \tag{128}$$

c) For the quadratic logarithmic scale-spaces over the logarithmic semifields  $L_\mu$ , we have:

$$\hat{k}_t^{L_\mu}(\omega) = -\frac{1}{2\mu}t\|\omega\|^2. \tag{129}$$

d) For the  $\alpha$ -dilation scale-space over the tropical max semifield  $T_+$ , we have:

$$\hat{k}_t^{T_+}(\omega) = \frac{1}{\alpha}t\|\omega\|^\alpha. \tag{130}$$

e) For the  $\alpha$ -erosion scale-space over the tropical min semifield  $T_-$ , we have:

$$\hat{k}_t^{T_-}(\omega) = -\frac{1}{\alpha}t\|\omega\|^\alpha. \tag{131}$$

### 5.3 Explicit Form of the Reduced Kernel

With the results derived above, we are now ready to derive the explicit form of the reduced kernel in the Fourier domain  $\hat{k}_t$ , and, in turn, an expression for the reduced kernel  $k_t$ . But to succinctly state this explicit form we need one extra ingredient: semifield exponentiation, i.e., a generalization of repeated semifield multiplication.

**Definition 30** (Semifield Exponentiation) Let  $R = (R, \mathbb{0}, \mathbb{1}, \oplus, \otimes)$  be a one-dimensional metric semifield. The semifield exponentiation  $\exp_R : \mathbb{R} \rightarrow R$  is defined as the (up to time scaling unique<sup>4</sup>) mapping that satisfies:

$$\begin{aligned} \exp_R(s) \otimes \exp_R(t) &= \exp_R(s + t) \text{ for all } s, t \in \mathbb{R}, \\ \exp_R(0) &= \mathbb{1}, \\ \lim_{t \rightarrow \infty} \exp_R(t) &= \mathbb{0}. \end{aligned}$$

To distinguish the semifield exponentiation from regular exponentiation, we always indicate the former with the semifield in the subscript and the latter without any subscript.

<sup>4</sup> The connected part of  $(R, \otimes)$  that contains  $\mathbb{1}$  is a one-dimensional Lie group. Semifield exponentiation is the Lie group exponential and is determined by a one-dimensional tangent vector at the identity, thus giving us the up to scaling uniqueness.

**Definition 31** (Employed Semifield Exponentiation)

- a) In the (nonnegative) linear semifield  $L$  case, we have  $\exp_L(t) = \exp(-t)$ .
- b) In the root semifields  $R_p$  case, we have  $\exp_{R_p}(t) = \exp(-t)$ .
- c) In the logarithmic semifields  $L_\mu$  case, we have  $\exp_{L_\mu}(t) = -\text{sign}(\mu)t$ .
- d) In the tropical max semifield  $T_+$  case, we have  $\exp_{T_+}(t) = -t$ .
- e) In the tropical min semifield  $T_-$  case, we have  $\exp_{T_-}(t) = t$ .

**Theorem 1** (Explicit form Reduced Kernel) *Let  $R$  be a one-dimensional metric semifield,  $\mathcal{F}_R$  the semifield Fourier transform (Definition 25), and  $\exp_R$  the semifield exponentiation (Definition 30). Consider all axioms of a semifield scale-space (Definition 27). We have that the reduced kernel  $k_t$  is (up to a time scaling) equal to:*

$$\begin{aligned} \hat{k}_t(\omega) &= \exp_R(\|\omega\|^\alpha t), \\ k_t(x) &= (\mathcal{F}_R^{-1} \hat{k}_t)(x), \end{aligned} \tag{132}$$

where  $\alpha$  is the scaling power (Axiom 4).

Before we continue with the proof, we can check that, indeed, all semifield Fourier transforms of the reduced kernels, as listed in Proposition 3, have this stated form.

**Proof** Consider the reduced kernel  $\hat{k}_t$  in the Fourier domain and all its properties as listed in Lemma 9. Due to the rotoreflectional symmetry (115), we will abuse notation slightly and write

$$\hat{k}_t(\omega) = \hat{k}_t(\|\omega\|) = \hat{k}_t(r), \tag{133}$$

where  $r = \|\omega\|$ .

Taking  $s = r > 0$  in the scaling invariance (116) we get

$$\hat{k}_t(r) = \hat{k}_{r^\alpha t}(1) \text{ for } r > 0. \tag{134}$$

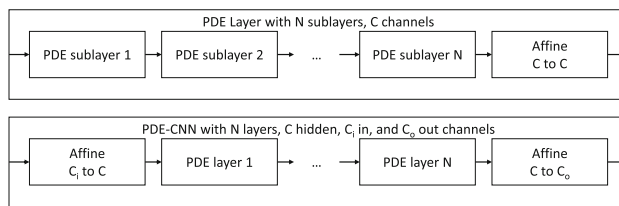
Due to the one-parameter semigroup property (117) and  $t \mapsto \hat{k}_t(\omega)$  being continuous, we have (up to a time scaling)

$$\hat{k}_t(1) = \exp_R(t) \text{ or } \hat{k}_t(1) = 0. \tag{135}$$

We are not interested in the  $\hat{k}_t(1) = 0$  case as it would imply, together with the previous equation, that  $\hat{k}_t$  is identically zero, corresponding to a non-relevant scale-space. Combining the equations found so far, we get

$$\hat{k}_t(r) = \exp_R(r^\alpha t) \text{ for } r > 0. \tag{136}$$

which we can extend to  $r = 0$  by continuity. Taking the inverse semifield Fourier transform concludes the proof.  $\square$



**Fig. 3** Architecture of a  $N$ -layer PDE-CNN with  $C$  channels. A PDE sublayer is either of the form (139) or (140)

The up-to-a-time-scaling non-uniqueness is something we can not avoid as every semifield scale-space  $\Phi_t$  corresponds to an infinite family of scale-spaces  $\tilde{\Phi}_t = \Phi_{st}$  for every  $s > 0$ .

The above theorem shows that every (one-dimensional metric) semifield  $R$  corresponds to a unique one-parameter family of semifield scale-spaces, where the scaling power  $\alpha$  acts as the parameter.

There is one caveat here though, and that is that not every  $\alpha$  necessarily results in a  $\tilde{k}_t$  which is in the domain of the used inverse semifield Fourier transform. For example, in the tropical semifields the case  $\alpha = 1$  results in a  $\tilde{k}_t$  which we cannot insert into the inverse transforms listed in Definition 26.

### 6 Architecture

In this section, we will briefly discuss the PDE-CNN architecture by defining the PDE sublayers corresponding to semifield scale-spaces (139), the PDE sublayer corresponding to convection (140), and the affine sublayer (141). We also illustrate how these sublayers combine to form a PDE layer, and how multiple PDE layers come together to construct a PDE-CNN in Fig. 3.

Consider any one-dimensional metric semifield  $R$  and a corresponding semifield scale-space  $\Phi_t$ . As Lemma 7 shows, we can write the scale-space  $\Phi_t f$  of a (appropriate) function  $f : \mathbb{R}^2 \rightarrow R$  as  $\Phi_t f = k_t \otimes f$  where  $k_t : \mathbb{R}^2 \rightarrow R$  is the reduced kernel, and  $\otimes$  is the semifield convolution. We want to implement these semifield scale-spaces to use them within the design of our PDE-CNNs. However, in practice, we cannot work with general signals defined on the continuum of  $\mathbb{R}^2$ , and we need to discretize our setting.

As is usual in machine learning we imagine images  $f : \mathbb{R}^2 \rightarrow R$  as sampled images  $\tilde{f} : Z \rightarrow R$  on a grid  $Z \subset \mathbb{R}^2$  such that  $\tilde{f}(z) = f(z)$  for all grid points  $z \in Z$ . We idealize the grid as the infinite integer grid  $\mathbb{Z}^2$  here for the sake of simplicity (we have no boundary concerns). Let  $\tilde{f}, \tilde{k} : \mathbb{Z}^2 \rightarrow R$  be the discretized versions of an image  $f$  and any kernel  $k$ . The discrete semifield convolution is defined as

$$(\tilde{k} \otimes \tilde{f})[i, j] = \bigoplus_{m, n \in \mathbb{Z}^2} \tilde{k}[-m + i, -n + j] \otimes \tilde{f}[m, n], \tag{137}$$

where we have used the notation  $[\cdot, \cdot]$  to emphasize the discrete nature.

With the discrete semifield convolution we can write down the formula for a PDE sublayer in the PDE-CNN. The input of a PDE sublayer consists of signals  $\tilde{f}_i : \mathbb{Z}^2 \rightarrow R$  and matrices  $H_i \in \mathbb{R}^{2 \times 2}$  with  $i = 1, \dots, C$ , and  $C$  being the amount of channels. The matrices  $H_i$  act as the learnable parameters of the layer. We consider the continuous scale-space kernel  $k_t : \mathbb{R}^2 \rightarrow R$  and create the discretized kernels  $\tilde{k}_i : \mathbb{Z}^2 \rightarrow R$  by defining

$$\tilde{k}_i(x) = k_1(H_i x). \tag{138}$$

Without loss of generality, we may take  $t = 1$  in our scale-space kernel  $k_t$  as a scaling in  $t$  can be captured in  $H_i$ . We then perform the discrete semifield convolutions to acquire our outputs  $\tilde{g}_i : \mathbb{Z}^2 \rightarrow R$ :

PDE sublayer: (139)  

$$\tilde{g}_i = \tilde{k}_i \circledast \tilde{f}_i.$$

Note that every input channel is only convolved with a single kernel, this is also known as a *depthwise convolution*.

The matrices  $H_i$  require some explanation. As already touched upon in Sect. 4.1, in the semifield scale-space axioms we implicitly assumed the standard inner product  $\mathcal{G}(x, y) = x^\top y$  on  $\mathbb{R}^2$ , but this is not the only one we can choose. By choosing the inner product to be

$$\mathcal{G}(x, y) = x^\top G y, \quad G = H^\top H$$

for any matrix  $H$  we get a scale-space representation that is complete identical, albeit “stretched” with respect to the coordinates. The Gram matrix  $G = H^\top H$  is by construction symmetric positive definite (SPD) as required, and relieves us from coding a SPD constraint. By considering kernels of the form  $k_t(Hx)$  we effectively include the possibility of processing the image with a different inner product. The “stretching” induced by the matrix  $H$  in the Gaussian scale-space case corresponds to the *affine Gaussian scale-space* studied in [44].

The PDE sublayer, as described in formula (139), seems to have nothing to do with a PDE at first glance. However, remember that every semifield scale-space  $\Phi_t$  over a semifield  $R$  can be associated with a PDE, and the PDE sublayer is effectively a solver for the corresponding initial value problem. The examples in Definition 28 clarify this.

Alongside the PDE sublayers that correspond to semifield scale-spaces, we also have the convection PDE sublayer. The convection sublayer effectively solves the convection PDE by translating images. The input of a convection PDE sublayer consists of images  $\tilde{f}_i : \mathbb{Z}^2 \rightarrow R$  and vectors  $v_i \in \mathbb{R}^2$  with  $i = 1, \dots, C$ , and  $C$  being the amount of channels. The

vectors  $v_i$  act as the learnable parameters of the layer. The output signals  $\tilde{g}_i : \mathbb{Z}^2 \rightarrow R$  are obtained through bilinear interpolating (Interp) the inputs at the appropriate positions:

Convection sublayer: (140)  

$$\tilde{g}_i[m, n] = \text{Interp} \left( \tilde{f}_i, m - (v_i)_1, n - (v_i)_2 \right).$$

The final ingredient we need is the affine layer which is defined as follows. The input consists of channels  $\tilde{f}_i : \mathbb{Z}^2 \rightarrow \mathbb{R}$ , weights  $w_{ij} \in \mathbb{R}$ , and biases  $b_j \in \mathbb{R}$ , with  $i = 1, \dots, C_i$ ,  $j = 1, \dots, C_o$ , and  $C_i, C_o$  being the amount of input and output channels, respectively. The weights  $w_{ij}$  and biases  $b_j$  act as the learnable parameters of the layer. We then perform the following pointwise computation to get the  $C_o$  outputs  $\tilde{g}_j : \mathbb{Z}^2 \rightarrow R$

Affine layer: (141)  

$$\tilde{g}_j = b_j + \sum_{i=1}^{C_i} w_{ij} \tilde{f}_i.$$

By concatenating various PDE sublayers that being either a sublayer that corresponds to a scale-space or a convection sublayer, with an affine combination layer at the end we form a PDE layer. Multiple PDE layers after each other with an affine layer at the start and end creates a PDE-CNN. The architecture is illustrated in Fig. 3.

## 7 Experiments

### 7.1 Including New PDEs in PDE-CNNs

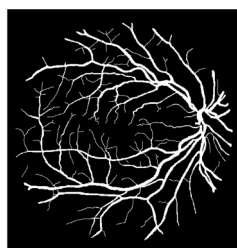
Current PDE-based neural networks employ three PDEs that generate scale-spaces: diffusion, dilation, and erosion (1), which correspond, respectively, to the linear semifield  $L$ , the tropical min semifield  $T_-$ , and the tropical max semifield  $T_+$ . Our theory reveals at least two PDEs that have not yet been used in PDE-based neural networks: the PDE that generates the root scale-space (75) and the PDE that generates the logarithmic scale-space (76), which arise naturally from the root semifield  $R_\rho$  and logarithm semifield  $L_\mu$ . Our first experiment will examine how the inclusion of these PDEs affects the accuracy of the PDE-CNN architecture, see Fig. 3.

The networks that we consider always include the convection PDE sublayer at the start of the PDE layer. The PDE-CNNs will consist of 6 PDE layers, 32 channels, and have an average parameter count of approximately 9 500, which changes with the amount of PDEs we add to its PDE layers.

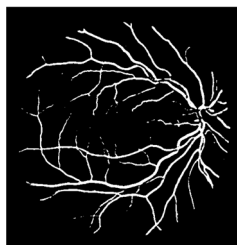
We will test the networks on the DRIVE dataset [18]. The dataset consists of fundus images, with the goal being



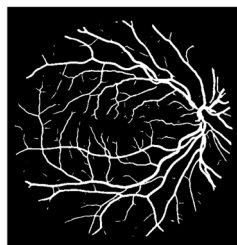
(a) Input image of the fundus of the eye.



(b) Target vessel segmentation.



(c) Example output of a network with a dice coefficient of  $\approx 0.78$ . Notice how small vessels are missed.



(d) Example output of a network with a dice coefficient of  $\approx 0.81$ . Notice how more small vessels are captured.

**Fig. 4** One instance of an input and its corresponding target segmentation from the DRIVE dataset, together with two example outputs of networks with different dice coefficients

vessel segmentation. In Fig. 4, one can see an example of such an image and its segmentation. We selected the DRIVE dataset because its images contain vessels at varying scales, making it well-suited for applying scale-space techniques, including PDE-based neural networks. It also allows for comparison with the results in [1, 3–5], which also use the DRIVE dataset.

The dataset consists of a training set of 20 images and a test set of 20 images. All images are  $584 \times 565$  pixels in 8-bit RGB color, which we rescale to the  $[0, 1]$  range by dividing by 255. We divided the training set into 2 880 overlapping patches of  $64 \times 64$ . Patches that contain no annotation, i.e., patches that are essentially completely within the black mask (Fig. 4), are removed, leaving us with 2 409 patches.

We train on batches consisting of 8 patches. We empirically found that a higher batch size results in a worse test set accuracy. We use the AdamW optimizer with an initial learning rate of 0.01 that decays linearly to 0.001 over the first 1 000 batches. The beta, epsilon, and weight decay parameters of AdamW are kept at their (PyTorch) default values of (0.9, 0.999),  $10^{-8}$ , and 0.01. During training, we keep track of the Dice coefficient on the test set and the best one is stored. We train until the Dice coefficient on the test set no longer increases, which happens within 20,000 batches. We then repeat this 5 times for every possible situation.

We empirically found that adding the linear semifield  $L$ , that is, we add a PDE sublayer corresponding to the Gaussian scale-space, does *not* affect the accuracy of any the networks. This can be explained by noticing that such a PDE sublayer can be emulated completely and effectively by the convection sublayer (140) together with the affine sublayer (141), which are always components of the PDE-CNNs we consider here. For this reason, we have omitted the linear semifield PDE sublayers altogether. This is in agreement with the results found in [45, p.28].

The result is found in Fig. 5. We observe multiple things:

- Adding semifields to the existing PDE-CNN architecture, which only employ the tropical semifields and convection, may enhance accuracy, albeit not significantly, as observed in the case of going from  $\{T_+, T_-\}$  to  $\{T_+, T_-, L_\mu\}$ .
- The inclusion of the tropical min semifield  $T_-$  *always* increases accuracy, most starkly seen when going from  $\{L_\mu\}$  to  $\{T_-, L_\mu\}$ .
- Adding semifields does *not* necessarily improve accuracy, as is evident from the last row when compared to the two semifield models in the middle rows.
- The inclusion of the root semifield  $R_p$  seems to make the training less stable, as indicated by the increase in spread within the scatter plot at the respective rows.

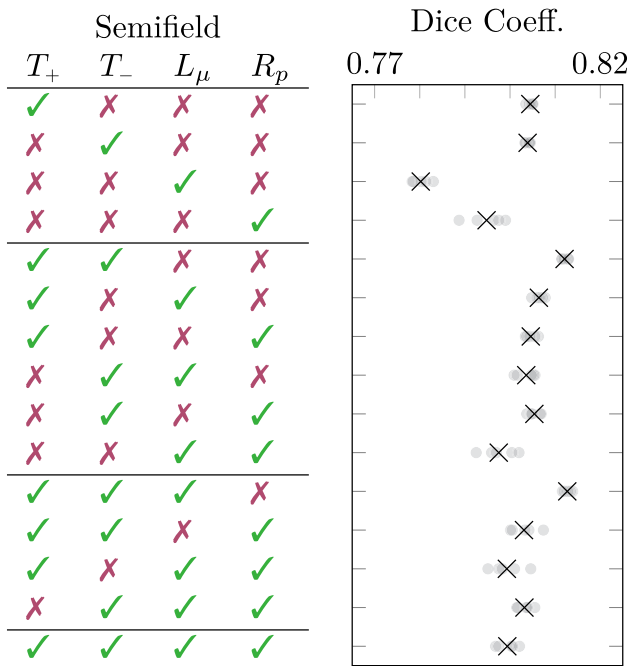
It is worth mentioning, however, that these results might be specific to the DRIVE dataset.

## 7.2 Data Efficiency of PDE-CNNs

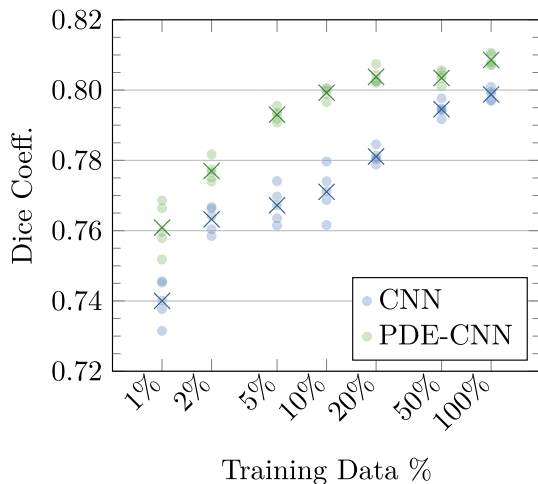
The data efficiency of PDE-G-CNNs on  $\mathbb{M}_2$  is already verified in [3], but whether this desirable property holds in the PDE-CNN case is still left untested. Our second experiment is therefore testing the data efficiency of a PDE-CNN on the DRIVE dataset. The PDE-CNN we consider here employs three PDEs within its PDE layers: convection, dilation, and erosion, just as in the papers [1, 3–5].

As a baseline, we consider a CNN with 31 488 parameters and compare this against a PDE-CNN with 5 280 parameters. To make the comparison fair, both networks have 6 layers and 24 channels. The only difference between the CNN and PDE-CNN is the kind of layer that is used. In the CNN, we use a standard 2D convolutional module with  $3 \times 3$  kernels together with a nonlinear activation function. In the PDE-CNN, we use a PDE layer as described in Fig. 3. The size of the networks has been chosen this way such that both give a satisfactory Dice coefficient of  $\gtrsim 0.80$  on the test set when trained on the complete training set.

Following the method in [3], we randomly take 1% to 100% of the training data. Other than that our methodology is identical to Sect. 7.1.



**Fig. 5** A scatterplot of the accuracy of a 6-layer PDE-CNN on the DRIVE dataset, with various designs of the PDE layer as indicated in the table on the left. The crosses indicate the mean. The rows are organized according to the amount of semifields included in the model



**Fig. 6** A scatterplot of the accuracy of a 6-layer 24-channel CNN (31 488 parameters) and PDE-CNN (5 280 parameters) on the DRIVE dataset, when trained multiple times on varying amounts of training data. The crosses indicate the mean

The result is found in Fig. 6. We see that on the DRIVE dataset, in comparison with a standard CNN, the PDE-CNN not only features fewer parameters but also showcases competitive accuracy and increased data efficiency. This mirrors the results found in [3], but this time for a PDE-CNN instead of the  $M = \mathbb{M}_2$  PDE-G-CNN considered there.

### 8 Conclusion

PDE-CNNs are an interesting alternative to CNNs in the sense that their constituents, this being solvers of PDEs that generate scale-spaces, are geometrically meaningful and interpretable.

The existing PDE-CNN framework uses four PDEs: convection, diffusion, dilation, and erosion. Through the introduction of semifield scale-spaces (Definition 27), we demonstrate the presence of a broad class of PDEs that remain unused within the PDE-CNN paradigm.

The theory of semifields scale-spaces is expressive and encapsulates a large class of known scale-spaces. Theorem 1 shows that every semifield gives rise to a one-parameter family of semifield scale-spaces. This indicates that the generalization to semifields is one that is not too general and definitely fruitful.

In Sect. 7.2, we empirically verified that on the DRIVE dataset that PDE-CNNs, just like PDE-G-CNNs, when compared to traditional CNNs, require less training data, have fewer parameters, and increased accuracy.

In Sect. 7.1, we experimented on the inclusion of various semifields and their corresponding scale-spaces within PDE layers of a PDE-CNN. We see that the thought “more semifields means better accuracy” is incorrect and that it is not clear if the addition of more semifields into the already existing PDE-CNN framework is worth the effort. However, in all cases inclusion of the tropical semifield improved the result, advocating for tropical algebras in PDE-based neural networks.

### Further Research

When comparing the results of PDE-CNNs on the DRIVE dataset here to the  $G = SE(2)$  PDE-G-CNNs results in [3], the accuracy is essentially the same (Dice  $\approx 0.81$ ), but there is a trade-off between memory usage and parameter reduction.

The  $SE(2)$  variant has less parameters (2 560) but, due to the feature maps being scalar fields on  $SE(2)$ , uses more memory, that being  $O \times H \times W$  scalars per feature map. Here  $O$  refers to the amount of orientations (typically 8),  $H$  to the height of the images, and  $W$  to the width.

Conversely, the  $\mathbb{R}^2$  variant has more parameters (5 280) but uses much less memory;  $H \times W$  scalars per feature map. This means that in some applications the PDE-CNN architecture might be preferable. However, the goal of the work here was not to compare PDE-CNNs to PDE-G-CNNs and the observations here only apply to the DRIVE dataset. Further research is needed to properly compare both architectures.

### Appendix A Semifield Fourier Transforms

**Lemma 10** *The employed semifield Fourier transforms satisfy Definition 25.*

**Proof** In the linear semifield  $L$  case, we know that the familiar Fourier transform satisfies the definition.

As for the root and logarithmic semifields, being isomorphic to the linear semifield, we can quickly deduce that they also satisfy the definitions through the equalities

$$\begin{aligned} \mathcal{F}_{L_\mu} &= \mathcal{P}_{\varphi_\mu^{-1}} \circ \mathcal{F}_L \circ \mathcal{P}_{\varphi_\mu}, \\ \mathcal{F}_{R_p} &= \mathcal{P}_{\varphi_p^{-1}} \circ \mathcal{F}_L \circ \mathcal{P}_{\varphi_p}, \end{aligned} \tag{A1}$$

where  $\mathcal{P}$  is the pointwise operator (10),  $\varphi_\mu(x) = e^{\mu x}$  the semifield isomorphism  $\varphi_\mu : L_\mu \rightarrow L_{\geq 0}$ , and  $\varphi_p(x) = x^p$  the semifield isomorphism  $\varphi_p : R_p \rightarrow L_{\geq 0}$ . For example, to show that  $\mathcal{F}_{L_\mu}$  satisfies the convolution property:

$$\begin{aligned} \mathcal{F}_{L_\mu}(f \otimes g) &= \mathcal{P}_{\varphi_\mu^{-1}} \mathcal{F}_L \mathcal{P}_{\varphi_\mu}(f \otimes g) \\ &= \mathcal{P}_{\varphi_\mu^{-1}} \mathcal{F}_L((\mathcal{P}_{\varphi_\mu} f) * (\mathcal{P}_{\varphi_\mu} g)) \\ &= \mathcal{P}_{\varphi_\mu^{-1}}((\mathcal{F}_L \mathcal{P}_{\varphi_\mu} f) \times (\mathcal{F}_L \mathcal{P}_{\varphi_\mu} g)) \\ &= (\mathcal{P}_{\varphi_\mu^{-1}} \mathcal{F}_L \mathcal{P}_{\varphi_\mu} f) \otimes (\mathcal{P}_{\varphi_\mu^{-1}} \mathcal{F}_L \mathcal{P}_{\varphi_\mu} g) \\ &= (\mathcal{F}_{L_\mu} f) \otimes (\mathcal{F}_{L_\mu} g), \end{aligned} \tag{A2}$$

where  $\otimes$  and  $\times$  are the semifield convolution and multiplication of  $L_\mu$  and where  $\times$  denotes the standard pointwise product of functions. In the above derivation, we have used that

$$\begin{aligned} \mathcal{P}_{\varphi_\mu}(f \otimes g) &= (\mathcal{P}_{\varphi_\mu} f) * (\mathcal{P}_{\varphi_\mu} g), \\ \mathcal{P}_{\varphi_\mu^{-1}}(f \times g) &= (\mathcal{P}_{\varphi_\mu^{-1}} f) \otimes (\mathcal{P}_{\varphi_\mu^{-1}} g), \end{aligned} \tag{A3}$$

and that  $\mathcal{F}_L$  has the convolution property.

Consider now the tropical max semifield  $T_+$ . That  $\mathcal{F}_{T_+}$  satisfies the linearity, equivariences, and the zero-frequency properties is immediate. As for the convolution property, we have

$$\begin{aligned} (\mathcal{F}_{T_+}(f \otimes g))(\omega) &= \sup_x (f \otimes g)(x) - \omega \cdot x \\ &= \sup_x (\sup_y f(x - y) + g(y)) - \omega \cdot x \\ &= \sup_x (\sup_{x_1+x_2=x} f(x_1) + g(x_2)) - \omega \cdot x \\ &= \sup_{x_1, x_2} f(x_1) + g(x_2) - \omega \cdot (x_1 + x_2) \\ &= (\sup_{x_1} f(x_1) - \omega \cdot x_1) + (\sup_{x_2} g(x_2) - \omega \cdot x_2) \\ &= (\mathcal{F}_{T_+} f)(\omega) \otimes (\mathcal{F}_{T_+} g)(\omega), \end{aligned} \tag{A4}$$

where  $\otimes$  and  $*$  are the tropical max  $T_+$  multiplication and convolution. For the invertibility, we refer to the Fenchel biconjugation theorem [46, Thm.4.2.1]. That the tropical min semifield Fourier transform  $\mathcal{F}_{T_-}$  satisfies all properties follows from the fact that  $T_-$  is semifield isomorphic to  $T_+$  with the isomorphism being  $\phi(x) = -x$ .  $\square$

### Appendix B Tropical Integration

**Proposition 4** *The natural integration of sum-approachable and bounded from above functions  $f : \mathbb{R}^2 \rightarrow T_+$  is*

$$\oint^{T_+} f = \sup_{x \in \mathbb{R}^2} f(x). \tag{B5}$$

*The natural integration of sum-approachable and bounded from below functions  $f : \mathbb{R}^2 \rightarrow T_-$  is*

$$\oint^{T_-} f = \inf_{x \in \mathbb{R}^2} f(x). \tag{B6}$$

**Proof** We will only prove this for the tropical max semifield  $T_+$ , the tropical min semifield case goes completely analogously. As we are working with  $T_+$  we remind ourselves that we have  $\oplus = \max$ ,  $\otimes = +$ ,  $\mu_{T_+}(A) = 0$ , and

$$1_A(x) = \begin{cases} 0 & \text{if } x \in A \\ -\infty & \text{otherwise} \end{cases}.$$

The function  $f$ , being sum-approachable and bounded from above, is pointwise defined by the limit

$$f(x) = \lim_{n \rightarrow \infty} \bigoplus_{i=1}^n a_i \otimes 1_{A_i}(x), \tag{B7}$$

with  $A_i$  non-empty and  $a_i$  bounded from above. We define its natural integral by

$$\begin{aligned} \oint f &:= \lim_{n \rightarrow \infty} \oint \left( \bigoplus_{i=1}^n a_i \otimes 1_{A_i} \right) \\ &= \lim_{n \rightarrow \infty} \bigoplus_{i=1}^n a_i \otimes \mu(A) \\ &= \lim_{n \rightarrow \infty} \max_{i=1, \dots, n} a_i + 0 \\ &= \sup_{i \in \mathbb{N}} a_i, \end{aligned} \tag{B8}$$

where the second equality is by the linearity of the integration (20) and the indicator function property (21).

Similarly, we have

$$\begin{aligned}
\sup_{x \in \mathbb{R}^2} f(x) &= \sup_{x \in \mathbb{R}^2} \lim_{n \rightarrow \infty} \bigoplus_{i=1}^n a_i \otimes \mathbb{1}_{A_i}(x) \\
&= \sup_{x \in \mathbb{R}^2} \lim_{n \rightarrow \infty} \max_{i=1, \dots, n} a_i + \mathbb{1}_{A_i}(x) \\
&= \sup_{x \in \mathbb{R}^2} \sup_{i \in \mathbb{N}} a_i + \mathbb{1}_{A_i}(x) \\
&= \sup_{i \in \mathbb{N}} \sup_{x \in \mathbb{R}^2} a_i + \mathbb{1}_{A_i}(x) \\
&= \sup_{i \in \mathbb{N}} a_i,
\end{aligned} \tag{B9}$$

where in the fourth equality we interchanged the order of suprema, and in the fifth equality we used the definition of  $\mathbb{1}_{A_i}$ .

Combining these two results, it follows that

$$\oint f = \sup_{x \in \mathbb{R}^2} f(x) \tag{B10}$$

is the natural tropical max semifield integration.  $\square$

**Acknowledgements** We thank Adrien Castella [45] for the initial Python implementation of the  $\mathbb{R}^2$  convection, diffusion, dilation and erosion PDE sublayers <https://github.com/adrien-castella/PDE-based-CNNs>.

**Author Contributions** G. Bellaard is the first author and main writer of the manuscript. G. Bellaard, S. Sakata, B. Smets, and R. Duits contributed to the semifield theory. G. Bellaard, S. Sakata, and R. Duits contributed to the semifield scale-space theory. S. Sakata performed the initial exploratory experiments. G. Bellaard prepared, conducted, and analyzed the final experiments. B. Smets provided important practical advice on the experimental aspects. All authors collaborated closely and reviewed the manuscript carefully. Author contributions per section. Section 1: G. Bellaard, R. Duits; Sect. 2: G. Bellaard, R. Duits; Sect. 3: G. Bellaard, S. Sakata, B. Smets, R. Duits; Sect. 4: G. Bellaard, S. Sakata, R. Duits; Sect. 5: G. Bellaard, S. Sakata, R. Duits; Sect. 6: G. Bellaard, B. Smets, R. Duits; Sect. 7: G. Bellaard, S. Sakata, B. Smets; Sect. 8: G. Bellaard, R. Duits.

**Funding** The EU is gratefully acknowledged for financial support through the REMODEL project (MSCA-SE 101131557).

We gratefully acknowledge the Dutch Foundation of Science NWO for funding of VICI 2020 Exact Sciences (Duits, Geometric learning for Image Analysis, VIC.202-031 <https://www.nwo.nl/en/projects/vic202031>).

**Availability of Data and Code** All code be found at <https://gitlab.com/gijsbel/semifield-pde-cnns>. The DRIVE dataset [18] can (The old address was <https://web.archive.org/web/20191003101812/http://www.isi.uu.nl/Research/Databases/DRIVE/>) be found at <https://drive.grand-challenge.org/>. The LieTorch package is public and can be found at <https://gitlab.com/bismetsjr/lietorch>. The original PDE-CNN implementation can be found at <https://github.com/adrien-castella/PDE-based-CNNs>.

## Declarations

**Conflict of interest** R. Duits is a member of the editorial board of Journal of Mathematical Imaging and Vision (JMIV).

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Smets, B., Portegies, J., Bekkers, E., Duits, R.: PDE-based group equivariant convolutional neural networks. *J. Math. Imaging Vis.* **65**, 209–239 (2023). <https://doi.org/10.1007/s10851-022-01114-x>
- Cohen, T., Welling, M.: *Group equivariant convolutional networks*. (eds Balcan, M.F., Weinberger, K.Q.) Proceedings of The 33rd International Conference on Machine Learning, Vol. 48 of Proceedings of Machine Learning Research, 2990–2999 (PMLR, New York, New York, USA, 2016)
- Pai, G., Bellaard, G., Smets, B., Duits, R.: *Functional properties of PDE-based group equivariant convolutional neural networks*. (eds Nielsen, F. & Barbaresco, F.) Geometric Science of Information, 63–72 (Springer Nature Switzerland, Cham, 2023). [https://doi.org/10.1007/978-3-031-38271-0\\_7](https://doi.org/10.1007/978-3-031-38271-0_7)
- Bellaard, G., Bon, D., Pai, G., Smets, B., Duits, R.: Analysis of (sub-)Riemannian PDE-G-CNNs. *J. Math. Imaging Vis.* **65**, 819–843 (2023). <https://doi.org/10.1007/s10851-023-01147-w>
- Bellaard, G., Pai, G., Bescos, J., Duits, R.: *Geometric adaptations of PDE-G-CNNs*. (eds Calatroni, L., Donatelli, M., Morigi, S., Prato, M. & Santacesaria, M.) Scale Space and Variational Methods in Computer Vision, 538–550 (Springer International Publishing, Cham, 2023). [https://doi.org/10.1007/978-3-031-31975-4\\_41](https://doi.org/10.1007/978-3-031-31975-4_41)
- Mohamed, M., Cesa, G., Cohen, T., Welling, M.: A data and compute efficient design for limited-resources deep learning (2020). <https://doi.org/10.48550/arXiv.2004.09691>
- Cohen, T., Weiler, M., Kicanaoglu, B., Welling, M.: *Gauge equivariant convolutional networks and the icosahedral CNN*. (eds Chaudhuri, K. & Salakhutdinov, R.) In Proceedings of the 36th International Conference on Machine Learning, Vol. 97, 1321–1330 (2019). <https://proceedings.mlr.press/v97/cohen19d.html>
- Pauwels, E., van Gool, L., Fiddelaers, P., Moons, T.: An extended class of scale-invariant and recursive scale space filters. *IEEE Trans. Pattern Anal. Mach. Intell.* **17**, 691–701 (1995). <https://doi.org/10.1109/34.391411>
- Iijima, T.: Basic theory of pattern observation. *Papers of Technical Group on Automata and Automatic Control* (1959)
- Alvarez, L., Guichard, F., Lions, P., Morel, J.: Axioms and fundamental equations of image processing. *Arch. Ration. Mech. Anal.* **123**, 199–257 (1993). <https://doi.org/10.1007/BF00375127>
- Florack, L.: Non-linear scale-spaces isomorphic to the linear case with applications to scalar, vector and multispectral images. *Int. J. Comput. Vision* **42**, 39–53 (2001). <https://doi.org/10.1023/A:1011185417206>
- Heijmans, H., Boomgaard, R.: Algebraic framework for linear and morphological scale-spaces. *J. Vis. Commun. Image Represent.* **13**, 269–301 (2002). <https://doi.org/10.1006/jvci.2001.0480>

13. Duits, R., Florack, L., de Graaf, J., Haar Romenij, B.: On the axioms of scale space theory. *J. Math. Imaging Vis.* **20**, 267–298 (2004). <https://doi.org/10.1023/B:JMIV.0000024043.96722.aa>
14. Felsberg, M., Sommer, G.: The monogenic scale-space: a unifying approach to phase-based image processing in scale-space. *J. Math. Imaging Vis.* **21**, 5–26 (2004). <https://doi.org/10.1023/B:JMIV.0000026554.79537.35>
15. Weickert, J., Ishikawa, S., Imiya, A.: Linear scale-space has first been proposed in japan. *J. Math. Imaging Vis.* **10**, 237–252 (1999). <https://doi.org/10.1023/A:1008344623873>
16. Koenderink, J.: The structure of images. *Biol. Cybern.* **50**, 363–370 (1984). <https://doi.org/10.1007/BF00336961>
17. Brockett, R., Maragos, P.: *Evolution equations for continuous-scale morphology*. (ed. IEEE) In 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 3, 125–128 vol. 3 (1992). <https://doi.org/10.1109/ICASSP.1992.226260>
18. Staal, J., Abramoff, M., Niemeijer, M., Viergever, M., van Ginneken, B.: Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging* **23**, 501–509 (2004). <https://doi.org/10.1109/TMI.2004.825627>
19. Welk, M.: *Families of generalised morphological scale spaces*. (eds Griffin, L. & Lillholm, M.) *Scale Space Methods in Computer Vision*, 770–784 (Springer, Berlin, Heidelberg, 2003). [https://doi.org/10.1007/3-540-44935-3\\_54](https://doi.org/10.1007/3-540-44935-3_54)
20. Schmidt, M., Weickert, J.: Morphological counterparts of linear shift-invariant scale-spaces. *J. Mat. Imaging Vis.* **56**, 352–366 (2016). <https://doi.org/10.1007/s10851-016-0646-8>
21. Dorst, L., Boomgaard, R.: Morphological signal processing and the slope transform. *Signal Process.* **38**, 79–98 (1994). [https://doi.org/10.1016/0165-1684\(94\)90058-2](https://doi.org/10.1016/0165-1684(94)90058-2)
22. Burgeth, B., Weickert, J.: An explanation for the logarithmic connection between linear and morphological system theory. *Int. J. Comput. Vision* **64**, 157–169 (2005). <https://doi.org/10.1007/s11263-005-1841-z>
23. Sapiro, G.: *Geometric Partial Differential Equations and Image Analysis* (Cambridge University Press, 2001). <https://doi.org/10.1017/CBO9780511626319>
24. Citti, G., Franceschiello, B., Sanguinetti, G., Sarti, A.: Sub-riemannian mean curvature flow for image processing. *SIAM J. Imag. Sci.* **9**, 212–237 (2016). <https://doi.org/10.1137/15M1013572>
25. Guichard, F., Morel, J.: *Partial differential equations and image iterative filtering*. (eds Duff, I. & Watson, G.) *The State of the Art in Numerical Analysis*, Vol. 63, 252–562 (Clarendon Press, 1997). <https://doi.org/10.1093/oso/9780198500148.003.0020>
26. Evans, L.: *Partial Differential Equations* Graduate studies in mathematics (American Mathematical Society, 2010). <https://doi.org/10.1090/gsm/019>
27. Jacobsen, J.-H., Van Gemert, J., Lou, Z., Smeulders, A.W.M.: *Structured receptive fields in cnns*. (ed. IEEE) In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Conference on Computer Vision and Pattern Recognition, 2610–2619 (IEEE, 2016). <https://doi.org/10.1109/CVPR.2016.286>
28. Pinteá, S.L., Tomen, N., Goes, S.F., Loog, M., van Gemert, J.C.: Resolution learning in deep convolutional networks using Scale-Space theory. *IEEE Trans. Image Process.* **30**, 8342–8353 (2021). <https://doi.org/10.1109/TIP.2021.3115001>
29. Tomen, N., Pinteá, S.-L., Van Gemert, J.: *Deep continuous networks*. (eds Meila, M. & Zhang, T.) In Proceedings of the 38th International Conference on Machine Learning, Vol. 139 of Proceedings of Machine Learning Research, 10324–10335 (2021). <https://proceedings.mlr.press/v139/tomen21a.html>
30. Saldanha, N., Pinteá, S.L., van Gemert, J.C., Tomen, N.: Frequency learning for structured cnn filters with gaussian fractional derivatives (2021). <https://arxiv.org/abs/2111.06660>. arXiv:2111.06660
31. Basting, M., et al.: *Scale Learning in Scale-Equivariant Convolutional Networks*. Master’s thesis, Delft University of Technology (2023)
32. Luan, S., Chen, C., Zhang, B., Han, J., Liu, J.: Gabor convolutional networks. *IEEE Trans. Image Process.* **27**, 4357–4366 (2018). <https://doi.org/10.1109/TIP.2018.2835143>
33. Romero, D.W., et al.: Flexconv: Continuous kernel convolutions with differentiable kernel sizes. (ed. ICLR) In 10th International Conference on Learning Representations (ICLR) (2022). <https://openreview.net/forum?id=3jooF27-0Wy>
34. Worrall, D., Welling, M.: *Deep scale-spaces: Equivariance over scale*. (eds Wallach, H. et al.) *Advances in Neural Information Processing Systems*, Vol. 32 (Curran Associates, Inc.). (2019) [https://proceedings.neurips.cc/paper\\_files/paper/2019/file/f04cd7399b2b0128970efb6d20b5c551-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2019/file/f04cd7399b2b0128970efb6d20b5c551-Paper.pdf)
35. Sosnovik, I., Szmaja, M., Smeulders, A.: *Scale-equivariant steerable networks*. (ed. ICLR) In International Conference on Learning Representations (2020). <https://openreview.net/forum?id=HJgpgurKPS>
36. Lindeberg, T.: Scale-covariant and scale-invariant gaussian derivative networks. *J. Math. Imaging Vis.* **64**, 223–242 (2022). <https://doi.org/10.1007/s10851-021-01057-9>
37. Sangalli, M., Blusseau, S., Velasco-Forero, S., Angulo, J.: *Scale equivariant neural networks with morphological scale-spaces*. (eds Lindblad, J., Malmberg, F. & Sladoje, N.) *Discrete Geometry and Mathematical Morphology*, 483–495 (Springer International Publishing, Cham, 2021). [https://doi.org/10.1007/978-3-030-76657-3\\_35](https://doi.org/10.1007/978-3-030-76657-3_35)
38. saz. Measurability of a pointwise limit of measurable functions. *Mathematics Stack Exchange*. <https://math.stackexchange.com/q/921576>
39. Matematleta. a proof that the pointwise limits of lower semicontinuous (lsc) functions is lsc. *Mathematics Stack Exchange*. <https://math.stackexchange.com/q/2968282>
40. Lindeberg, T.: *On the Axiomatic Foundations of Linear Scale-Space*, 75–97 (Springer Netherlands, Dordrecht, 1997). [https://doi.org/10.1007/978-94-015-8802-7\\_6](https://doi.org/10.1007/978-94-015-8802-7_6)
41. Kolokoltsov, V., Maslov, V.: *Idempotent Analysis and Its Applications* (Springer, Dordrecht, 1997). <https://doi.org/10.1007/978-94-015-8901-7>
42. Rudin, W.: *Functional Analysis* (McGraw-Hill, 1973, 1991)
43. Kraakman, T.B.: *On the Construction of Classical and Quantum PDE-Based Neural Networks*. Master’s thesis, Eindhoven University of Technology (2023)
44. Lindeberg, T., Gårding, J.: Shape-adapted smoothing in estimation of 3-d shape cues from affine deformations of local 2-d brightness structure. *Image and Vision Computing* **15**, 415–434 (1997). <https://www.sciencedirect.com/science/article/pii/S026288569701144X>. [https://doi.org/10.1016/S0262-8856\(97\)01144-X](https://doi.org/10.1016/S0262-8856(97)01144-X)
45. Castella, A.: *Introduction to Shift-Invariant PDE-based Convolutional Neural Networks and experimentation with Vessel Segmentation*. Bachelor’s thesis, Eindhoven University of Technology (2021)
46. Borwein, J., Lewis, A.: *Convex Analysis and Nonlinear Optimization* (Springer New York, NY, 2016). <https://doi.org/10.1007/978-0-387-31256-9>