

**Dissertation**

**Nonlocal Evolutions in Image  
Processing**

Giovanno Marcelo Renato Cárdenas

Saarbrücken

2018

Dissertation  
zur Erlangung des Grades des  
Doktors der Naturwissenschaften  
der Fakultät für Mathematik und Informatik  
der Universität des Saarlandes

*Day of Colloquium*

12 July 2018

*Dean of Faculty*

Univ.-Prof. Dr. Sebastian Hack

*Chair of the Committee*

Prof. Dr. Sergej Rjasanow

*Reviewers*

Prof. Dr. Joachim Weickert

Universität des Saarlandes

Prof. Dr. Guy Gilboa

Technion–Israel Institute of Technology

*Academic Assistant*

Dr. Pascal Peter

# Acknowledgments

I would like to express my sincere gratitude to Prof. Dr. Joachim Weickert for his guidance and constant support during the development of this thesis. I would also like to take opportunity to thank all the members of the Mathematical Image Analysis Group with whom I had many interesting scientific discussions during the past years and who always created such a nice atmosphere at the work place. Last but not least I would like to thank my family. Ale, this would have not been possible without your support.



# Short Abstract

The main topic of this thesis is to study a general framework which encompasses a wide class of nonlocal filters. For that matter, we introduce a general initial value problem, defined in terms of integro-differential equations and following the work of Weickert [1], we impose a set of basic assumptions that turn it into a well-posed model and develop nonlocal scale-space theory. Moreover, we go one step further and consider the consequences of relaxing some of this initial set of requirements. With each particular modification of the initial requirements, we obtain a particular framework which encompasses a more specific, yet wide, family of nonlocal processes.



# Kurzzusammenfassung

Das Hauptthema dieser Arbeit ist die Untersuchung eines allgemeinen Rahmens für eine breite Klasse nichtlokaler Filter. Zuerst führen wir ein Modell ein, das auf Integro-Differentialgleichungen basiert. Wir ergänzen es mit einer Reihe von Grundannahmen, die es uns ermöglichen, eine nichtlokale Skalenraumtheorie zu entwickeln, wie in [1]. Außerdem betrachten wir die Konsequenzen der Abschwächung einiger dieser Annahmen. Wir stellen verschiedene Beispiele für die mit jeder Relaxation erhaltenen Prozesse vor.



# Abstract

Many filters encountered in image processing, usually exhibit a local behaviour which can be described in terms of partial differential equations (PDEs). Since more than a decade, however, nonlocal filters have been introduced and represent an important alternative to local filters like nonlinear diffusion, specially for task in which the restoration of texture information is important. Even though they have been widely studied in the literature, less is known about their scale-space properties. The main topic of this thesis is to study a general framework which encompasses a wide class of nonlocal filters. For that purpose, we introduce a general initial value problem, defined in terms of integro-differential equations and following [1], we impose a set of basic assumptions that turn it into a well-posed model. This allows us to develop a complete nonlocal scale-space theory. Moreover, we go one step further and consider the consequences of relaxing some of the initial set of requirements. With each particular modification, we obtain a more specific framework which encompasses a more specific, yet wide, family of processes satisfying a useful set of properties. Specifically, the basic general framework that we study in this work, is given by the following initial value problem:

$$u_t(\mathbf{x}, t) = \int_{\Omega} K(\mathbf{x}, \mathbf{y}, u(\cdot, t))(u(\mathbf{y}, t) - u(\mathbf{x}, t))d\mu(\mathbf{y}) \quad \text{in } \bar{\Omega} \times [0, T], \quad (1)$$

$$u(\mathbf{x}, 0) = f(\mathbf{x}) \quad \text{in } \bar{\Omega}, \quad (2)$$

where  $\Omega$  represents the domain which contains the input image and  $f$  represents the input image itself. The variables  $\mathbf{x}, \mathbf{y}$  represent the locations in the domain  $\Omega$  and the variable  $t$  represents time. The measure  $\mu$  is a general measure which allows us to analyse a spatially continuous formulation and a corresponding space discrete formulation contemporaneously. The function

$K$ , represents a weight between points inside the image domain. Its value controls the strength with which two different points of the image should influence each other through the integro-differential evolution. All the assumptions used to define the specific families involve constraints about the weight function  $K$ .

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Scale-spaces for symmetric nonlocal evolutions</b>	<b>7</b>
2.1	Well-posedness of the general framework . . . . .	8
2.2	Nonlocal scale spaces: symmetric evolutions . . . . .	14
2.3	Examples . . . . .	19
2.3.1	Linear continuous setting with shift-invariant weight functions . . . . .	19
2.3.2	Space discrete setting . . . . .	20
2.3.3	A linear scale-space variant of bilateral filtering . . . . .	21
2.3.4	Robustified bilateral scale-space . . . . .	23
2.3.5	NL-means scale-space . . . . .	24
2.4	Conclusions . . . . .	25
<b>3</b>	<b>Nonsymmetric nonlocal evolutions</b>	<b>27</b>
3.1	Preliminaries . . . . .	28
3.2	Particle swarm systems: consensus problem . . . . .	31

---

3.3	Nonlocal nonsymmetric processes: convergence results . . . . .	34
3.3.1	Vector valued processes . . . . .	38
3.4	Application: sector diffusion . . . . .	39
3.4.1	Semidiscrete modelling of sector diffusion . . . . .	41
3.4.2	Time discretisation of the sector diffusion scheme . . . . .	45
3.4.3	Influence of the number of sectors on denoising . . . . .	49
3.4.4	Parameter selection . . . . .	49
3.4.5	Denoising evaluation . . . . .	50
3.4.6	Datasets and implementations . . . . .	51
3.5	Application: coherence enhancing image filter . . . . .	51
3.5.1	The two-step approach in detail . . . . .	53
3.5.2	Parameter selection and influence . . . . .	58
3.6	Applications: bilateral filtering and NL-means evolutions . . . . .	63
3.7	Conclusions . . . . .	65
<b>4</b>	<b>Nonlocal evolutions: beyond nonnegativity</b>	<b>67</b>
4.1	Forwards and backwards nonlocal processes . . . . .	68
4.1.1	Stabilization . . . . .	70
4.1.2	Explicit schemes and experiments . . . . .	74
4.2	Backward nonlocal processes . . . . .	76
4.2.1	Basic properties . . . . .	80
4.2.2	Generalisation with weights . . . . .	84
4.2.3	Connections to the nonlocal framework . . . . .	87

---

4.2.4	Explicit scheme and experiments . . . . .	88
4.3	Higher order nonlocal equations . . . . .	89
4.4	Conclusions . . . . .	90
<b>5</b>	<b>Summary and outlook</b>	<b>93</b>



# List of Figures

2.1	Bilateral scale-space evolution . . . . .	22
2.2	Robustified bilateral process . . . . .	23
2.3	Nonlinear NL-means evolution . . . . .	25
3.1	Mapping of pixels to particle system in the interval $[0,255]$ . .	35
3.2	Mapping of pixels to sectors . . . . .	42
3.3	Plot of function $h$ . . . . .	43
3.4	Peppers image denoising . . . . .	46
3.5	Denoising of peppers image with sector diffusion . . . . .	47
3.6	Corner preservation and denoising . . . . .	47
3.7	Denoising peppers (std. dev. = 40) . . . . .	48
3.8	Idea of coherence enhancing image filtering . . . . .	51
3.9	Kernel function $k_1$ as given in (3.41) . . . . .	55
3.10	Greyscale image of a fingerprint . . . . .	57
3.11	Colour coding . . . . .	58
3.12	Local dominant gradient orientation . . . . .	59

3.13	Grey value evolution of coherence enhancement . . . . .	62
3.14	Comparison to coherence enhancing diffusion . . . . .	63
4.1	Mapping of pixels with reflections . . . . .	71
4.2	Mapping of pixels to particle system in the interval $[0,255]$ . .	72
4.3	Corresponding diffusivity and flux . . . . .	78
4.4	Exemplar penaliser . . . . .	80
4.5	Processing a photography of “flatowturm (potsdam)” . . . . .	88

# List of Tables

1.1	Structure of the work . . . . .	5
3.1	Denoising evaluation of sector diffusion with NL-means . . . . .	46
5.1	Summary of the main assumptions . . . . .	94
5.2	Summary of the chapters . . . . .	95



# Chapter 1

## Introduction

Diffusion models have a long tradition in different fields like Physics, Biology and Engineering as several different types of natural phenomena find a good description in terms of diffusion processes. One important example is that of the heat equation,

$$\partial_t u = \Delta u,$$

which models the change of temperature in time  $t$  of each location inside a given region  $\Omega$  and is the prototypical parabolic partial differential equation. If  $\Omega$  is a subset of  $\mathbb{R}^N$ , the operator in the r.h.s. is given by the Laplacian  $\Delta u := \sum_{i=1}^N \frac{\partial^2 u}{\partial x_i^2}$ . The complete model for the rate of change of temperature inside some region  $\Omega$  is obtained by supplementing this linear differential equation with an initial temperature distribution  $f$ , and some conditions at the boundary  $\partial\Omega$  of the region of interest. One example of such an initial value problem is obtained when there is no heat flux between the region  $\Omega$  and its surroundings. This is modelled by supplementing the equation with the boundary condition given by  $\frac{du}{d\nu} = 0$ , where  $\nu$  denotes the outer normal vector to  $\partial\Omega$ . In this case, the problem is usually referred to as the Neumann initial value boundary problem. With all its ingredients, it is given by

$$\begin{aligned} \partial_t u &= \Delta u & \text{in } \Omega \times ]0, T], \\ \frac{du}{d\nu} &= 0 & \text{on } \partial\Omega, \\ u(\mathbf{x}, 0) &= f(\mathbf{x}) & \text{in } \Omega. \end{aligned}$$

Starting from the work of Iijima's pioneering work in 1962 [2], this type of problem has found its way to image processing applications as a filter for denoising and simplifying images. In that context, the function  $f$  represents an input image rather than a temperature distribution and the diffusion equation acts diffusing, or locally averaging, its grey values. In the case of denoising, the given input  $f$  represents some image corrupted with noise and the goal is that of removing the noise as much as possible, while at the same time preserving, as best as possible, the information of the original uncorrupted image.

A major drawback about the above filter, however, is the fact that the Laplacian operator  $\Delta$  is homogeneous, in the sense that it acts uniformly over the whole image. Hence, it has no chance of steering the way in which diffusion is done and is thus unable to take into account semantically important information about the image. For example, it is unable to recognize edges. The filter does indeed remove noise thanks to the local averaging characteristic, but at the same time, it also cancels important information.

An alternative family of filter which seeks to overcome this problem, is given by nonlinear diffusion filters. Starting from Perona and Malik [3], different types of nonlinear diffusion filters have been proposed in order to control the amount of diffusion which should be applied to each part of the image, by taking into account semantically important information. The prototypical form of nonlinear diffusion is given by

$$u_t = \operatorname{div}(g(|\nabla u|^2)\nabla u) \quad \text{in } \Omega \times [0, T], \quad (1.1)$$

$$u(\mathbf{x}, 0) = f(\mathbf{x}) \quad \text{in } \Omega. \quad (1.2)$$

The main mechanisms of this filter is given by the diffusivity  $g$  which controls the strength of the diffusion at a given location, depending on the likelihood, measured with  $|\nabla u|^2$ , that it is part of an edge. This way, the filter is able to smooth homogeneous regions of the image, while preserving edge information.

For some applications, it may be desirable to steer the diffusion in such a way that the orientation of some interesting features is taken into account. Nonlinear diffusion filters like the one of Perona Malik, however, are unable to achieve such behaviour because of the scalar-valued nature of  $g$ . Anisotropic filters are designed to model exactly this type of situation. A general theoretical framework for anisotropic diffusion filters was studied in [1]. The

basic model for the class of processes studied in that work is given by

$$\begin{aligned} \partial_t u &= \operatorname{div}(D(J_\rho(\nabla u_\sigma))\nabla u) && \text{in } ]0, T] \times \Omega, \\ u(\mathbf{x}, 0) &= f(\mathbf{x}) && \text{in } \Omega, \\ \langle D(J_\rho(\nabla u_\sigma))\nabla u, \boldsymbol{\nu} \rangle &= 0 && \text{in } \partial\Omega, \end{aligned}$$

where  $\boldsymbol{\nu}$  represents the outer normal vector at each point of  $\partial\Omega$ , and  $\sigma, \rho$ , represent a small spatial smoothing. In contrast to the previous diffusion filter, the diffusivity  $g$  is replaced by the compositions of the structure tensor  $J$  which is able to identify features such as corners, by taking into account how the orientation of the gradient changes in the vicinity of any investigated point, and the diffusion tensor  $D$ . The author shows that it is possible, under rather weak assumptions, to establish well-posedness of the model and to ensure that its solutions satisfy a set of scale-space properties. The author also goes one step further and studies a corresponding semidiscrete framework, where the image is sampled on a finite grid which represents its pixels positions. This leads to a systems of ordinary differential equations (ODEs) in which the image is now represented as a vector  $\mathbf{u} \in C([0, \infty[, \mathbb{R}^N)$ . The general model reads,

$$\begin{aligned} \dot{\mathbf{u}} &= \mathbf{A}(\mathbf{u})\mathbf{u}, \\ \mathbf{u}(0) &= \mathbf{f}, \end{aligned}$$

where  $\mathbf{f} \in \mathbb{R}^N$  and  $\mathbf{A}$  is a matrix and  $\dot{\mathbf{u}}$  denotes the derivative w.r.t time. The author states a basic set of conditions about the matrix  $\mathbf{A}$ , and shows that under those conditions the model is well-posed, and its unique solution satisfies a maximum-minimum principle, preserves the average grey value, admits a large set of Lyapunov functionals and converges to a constant steady state. An important point to notice about the semidiscrete framework, is that of its generality. In fact, even though in that work it was used as a basis for framing the spatial discretisation obtained from the continuous anisotropic diffusion filters. The actual assumptions about the matrix  $\mathbf{A}$  are rather general and it is apparent that a larger class of processes would give rise to space discretisations that satisfy them. Having this considerations in mind, let us introduce the idea of nonlocal processes. Many filters encountered in image processing, including the nonlinear anisotropic filters discussed above, usually exhibit a local behaviour which can be described in terms of differential equations (PDEs). Recently, however, nonlocal filters have been introduced [4–7].

Their nonlocal nature derives from the fact that pixels are averaged based on information which goes from large neighbourhood, consisting of more than its closest neighbouring pixels, all the way to information about the whole image. Some types of nonlocal filter, like the bilateral filter [5], are based on grey value differences between pixels. Others are based on the similarities between patches. Namely, small sub-images surrounding a given pixel. In the last case, filters like NL-means [6], are able to take advantage of the self-similarity which is prevalent in many type of natural images. Indeed, they have a superior performance in terms of preserving and denoising texture information compared to local nonlinear diffusion filters.

Even though nonlocal filters have been widely studied in the literature, less is known about their scale-space properties. Hence, the observation about the generality of the semidiscrete theory of [1] becomes quite relevant in this context.

The main topic of this thesis is to study a general framework which encompasses a wide class of nonlocal filters. We will introduce a general initial value problem, defined in terms of integro-differential equations and following [1], we will impose a set of basic assumptions that will turn it into a well-posed model. We will initially state a set of assumptions that will allow us to obtain a nonlocal scale-space theory. However, we will go one step further and consider the consequences of relaxing some of this initial set of requirements. Thus obtaining, with each particular modification, an alternative framework which encompasses a different family of processes that satisfy a specific set of properties. Specifically, the basic general framework that we study in this work, is given by the following initial value problem:

$$u_t(\mathbf{x}, t) = \int_{\Omega} K(\mathbf{x}, \mathbf{y}, u(\cdot, t))(u(\mathbf{y}, t) - u(\mathbf{x}, t))d\mu(\mathbf{y}) \quad \text{in } \bar{\Omega} \times [0, T], \quad (1.3)$$

$$u(\mathbf{x}, 0) = f(\mathbf{x}) \quad \text{in } \bar{\Omega}, \quad (1.4)$$

where  $\Omega$  represents the domain which contains the input image and  $f$  represents the input image itself. The variables  $\mathbf{x}, \mathbf{y}$  represent the locations in the domain  $\Omega$  and the variable  $t$  represents time. The introduction of a general measure  $\mu$ , instead of the usual Lebesgue integral, will allow us to study contemporaneously, both a spatially continuous formulation of the framework and a corresponding space discrete formulation, which can be easily obtained by restricting the measure  $\mu$  to a finite grid given by the positions

of the image pixels. The function  $K$  on the other hand, plays the role of the diffusivity. In this setting it represents a weight between points inside the image domain. Its value controls the strength with which two different points of the image should be averaged. All the assumptions used to defined the specific families involve constrains about the weight function  $K$

### Structure of the work

The guideline of the work is the general framework (1.3)–(1.4). In each of the chapters we will focus on a specific set of assumptions about its weight  $K$ . In the second chapter we study a subclass of (1.3)–(1.4) which gives rise to scale-spaces. These will be mainly characterized by the fact that they involve a symmetric  $K$ . In the third chapter we relax the symmetry assumption and replace it with less restrictive alternatives. Finally, in Chapter 4 we consider processes which involve a weight function with no restrictions about its sign. A summary of this organization is given in Table 1.1. A summary of the different properties is given in Table 5.1 of Chapter 5.

	Chapter 2	Chapter 3	Chapter 4
Regularity (NL1):	yes	yes	yes/no
Non-negativity (NL2):	yes	yes	no
Symmetry (NL3):	yes	no	no
Irreducibility (NL4):	yes	no	no
Type-symmetry (NL5):	no	yes	no
Local boundedness (NL6):	no	yes/no	no

Table 1.1: Summary of the work. The option yes/no means that both possibilities are considered in that chapter.



## Chapter 2

# Scale-spaces for symmetric nonlocal evolutions

Starting with Iijima's pioneering work in 1962 [2] and its western counterparts by Witkin [8] and Koenderink [9] two decades later, the scale-space concept has become an integral part of many image processing and computer vision methods. For example, it is the backbone of the widely used SIFT detector for feature matching [10].

Scale-spaces embed an original image  $f$  into a family  $\{T_t f \mid t \geq 0\}$  such that  $T_0 f = f$  and larger values of  $t$  correspond to simpler representations of  $f$ . Numerous attempts have been made to formalise this idea and supplement it with additional assumptions in order to restrict the scale-space evolution to a specific class of processes, or even single out a unique scale-space in an axiomatic way. Such evolutions include linear processes such as Gaussian scale-space [2, 8, 9, 11–13], the Poisson scale-space [14] and its generalisation to  $\alpha$ -scale-spaces [15]. Typical representatives of nonlinear scale-spaces are given by nonlinear diffusion scale-spaces [1], the morphological equivalent of Gaussian scale-space [16], and curvature-driven evolutions such as the affine morphological scale-space [17]. Moreover, also spatio-temporal scale-spaces have been considered [18, 19], regularisation methods have been identified as scale-spaces [20], and spatially varying dominant scales have been proposed in [21].

Many of these processes exhibit a local behaviour and can be described in terms of partial differential equations (PDEs) or pseudodifferential equations.

More recently, however, nonlocal processes have become very popular in research. For instance, bilateral filters [4, 5] and patch-based methods [6, 7] are widely-used in image processing applications, and classical PDEs and variational methods have been generalised to nonlocal evolutions [22, 23]. However, less is known about scale-space theory for nonlocal processes. Related work can be found in [24], where the authors develop nonlocal morphological scale-spaces as an extension of [17].

The goal of this chapter is to address this issue from the point of view of diffusion processes. To achieve this, we impose a set of constraints to the general framework (1.3)–(1.4), in order to define a general class of well-posed nonlocal evolutions which satisfy scale-space properties that are in analogy to the diffusion framework by Weickert [1]. These includes existence and uniqueness, preservation of the average grey value, an extremum principle, a large class of Lyapunov functionals, and convergence to a flat steady state. We show that our framework covers nonlocal generalisations of Gaussian scale-space as well as space-discrete diffusion scale-spaces. Moreover, we introduce modifications of bilateral filtering and NL-means that are in accordance with our theory and can be robust under noise.

The results of this chapter were initially published in [25]. In that work only linear nonlocal scale-spaces were considered. Here the more general situation of non-linear processes is addressed.

## 2.1 Well-posedness of the general framework

We now define in detail the general class of nonlocal processes that will be discussed in the rest of this chapter. It is obtained by making a set of basic assumptions to (1.3)–(1.4) which turn it into the well-posed initial value problem which we now describe in detail. For that matter, let  $\Omega$  be an open bounded subset of  $\mathbb{R}^N$ , let  $T > 0$  be a constant, and let  $\mu$  be a locally finite Borel measure. Let  $f \in C(\bar{\Omega})$ , and consider the following problem.

$$\partial_t u(\mathbf{x}, t) = \int_{\Omega} K(\mathbf{x}, \mathbf{y}, u(\cdot, t))(u(\mathbf{y}, t) - u(\mathbf{x}, t)) d\mu(\mathbf{y}) \quad \text{in } \bar{\Omega} \times [0, T], \quad (2.1)$$

$$u(0, \mathbf{x}) = f(\mathbf{x}) \quad \text{in } \bar{\Omega}. \quad (2.2)$$

In order to ensure the well-posedness of the above initial value problem, we make the following set of basic assumptions.

**(NL1) Regularity:**  $K : \bar{\Omega} \times \bar{\Omega} \times C(\bar{\Omega}) \rightarrow \mathbb{R}$ , is Lipschitz continuous and bounded.

**(NL2) Non-Negativity:** The function  $K$  is non-negative.

With the above assumption, we can ensure that problem (2.1)–(2.2) is well-posed and has a unique solution in the sense of the following definition.

**Definition 1** (Solution). *Let  $T > 0$ . We say that  $u \in C(\bar{\Omega} \times [0, T])$  is a solution of problem (2.1)–(2.2) if for  $0 \leq t \leq T$ , the following holds.*

$$u(\mathbf{x}, t) = f(\mathbf{x}) + \int_0^t \int_{\Omega} K(\mathbf{x}, \mathbf{y}, u(\cdot, s))(u(\mathbf{y}, s) - u(\mathbf{x}, s)) d\mu(\mathbf{y}) ds. \quad (2.3)$$

More precisely, we obtain the following result.

**Theorem 1** (Well-posedness). *For any given  $T > 0$ , the initial value problem (2.1)–(2.2) supplemented with the assumptions (NL1)–(NL2) has a unique solution in the sense of Definition 1. Moreover, this unique solution depends continuously on the initial value  $f$ .*

We will now prove Theorem 1. Let us first introduce some definitions which will be used for that purpose. Let  $\Omega$  be as above and  $M$  be a positive constant. Consider the metric space defined as

$$C_M(\bar{\Omega} \times [0, T]) := \{u \in C(\bar{\Omega}) : \|u\|_{\infty} \leq M\},$$

with the metric induced by the  $\|\cdot\|_{\infty}$  norm.

The elements of this space correspond to the closed ball with radius  $M$  of the complete metric space  $C(\bar{\Omega} \times [0, T])$ . Therefore, the space  $C_M$  is also a complete metric space.

Moreover, given fixed values  $T, M > 0$ , and  $f \in C(\bar{\Omega})$ . Let  $H_f$  be the operator on  $C(\bar{\Omega} \times [0, T])$  defined as,

$$H_f(u)(\mathbf{x}, t) := f(\mathbf{x}) + \int_0^t \int_{\Omega} K(\mathbf{x}, \mathbf{y}, u(\cdot, s))(u(\mathbf{y}, s) - u(\mathbf{x}, s)) d\mu(\mathbf{y}) ds. \quad (2.4)$$

We may now state the following intermediate lemmas which will allow us to establish Theorem 1.

**Lemma 1.** *For every  $f \in C(\bar{\Omega})$ , and every  $u \in C(\bar{\Omega} \times [0, T])$ , the function  $H_f(u)$  belongs to  $C(\bar{\Omega} \times [0, T])$ .*

*Proof.* Let  $(\mathbf{x}, t), (\mathbf{z}, r) \in \bar{\Omega} \times [0, T]$ . From (2.4) we know that

$$H_f(\mathbf{z}, r) = f(\mathbf{z}) + \int_0^r \int_{\Omega} K(\mathbf{z}, \mathbf{y}, u(\cdot, s))(u(\mathbf{y}, s) - u(\mathbf{z}, s)) d\mu(\mathbf{y}) ds. \quad (2.5)$$

We then split the outer integral in the r.h.s. of the previous equality and rewrite it as

$$\begin{aligned} & \int_0^t \int_{\Omega} K(\mathbf{z}, \mathbf{y}, u(\cdot, s))(u(\mathbf{y}, s) + u(\mathbf{x}, s) - u(\mathbf{x}, s) - u(\mathbf{z}, s)) d\mu(\mathbf{y}) ds + \\ & \int_t^r \int_{\Omega} K(\mathbf{z}, \mathbf{y}, u(\cdot, s))(u(\mathbf{y}, s) - u(\mathbf{z}, s)) d\mu(\mathbf{y}) ds, \end{aligned}$$

where we have added and subtracted  $u(\mathbf{x}, s)$  inside the first integral. Using this last expression, and the definition (2.4) of  $H_f(\mathbf{x}, t)$ , we obtain that

$$\begin{aligned} |H_f(\mathbf{z}, r) - H_f(\mathbf{x}, t)| & \leq \int_t^r \int_{\Omega} |K(\mathbf{z}, \mathbf{y}, u(\cdot, s))| |u(\mathbf{y}, s) - u(\mathbf{z}, s)| d\mu(\mathbf{y}) ds + \\ & \int_0^t \int_{\Omega} |K(\mathbf{z}, \mathbf{y}, u(\cdot, s)) - K(\mathbf{x}, \mathbf{y}, u(\cdot, s))| |u(\mathbf{y}, s) - u(\mathbf{x}, s)| d\mu(\mathbf{y}) ds + \\ & \int_0^t \int_{\Omega} |K(\mathbf{z}, \mathbf{y}, u(\cdot, s))| |u(\mathbf{x}, s) - u(\mathbf{z}, s)| d\mu(\mathbf{y}) ds. \end{aligned}$$

Finally, if we assume that  $L > 0$  is the Lipschitz constant of  $K$ , and that  $B > 0$  is an upper bound for  $K$ , we obtain that

$$|H_f(\mathbf{z}, r) - H_f(\mathbf{x}, t)| \leq |t - r| \mu(\Omega) B^2 \|u\|_{\infty} +$$

$$|z - x|L2\|u\|_\infty|T|\mu(\Omega)+ \\ |T|\mu(\Omega) \int_0^t |u(\mathbf{x}, s) - u(\mathbf{z}, s)| ds.$$

The continuity of  $H_f$  thus follows from the uniform continuity of  $u$ .  $\square$

**Lemma 2.** *Given fixed constants  $T, M > 0$ . Let  $f, g \in C(\bar{\Omega})$ , and  $u, v \in C_M(\bar{\Omega} \times [0, T])$ . There exists a constant  $c$ , depending only on  $\Omega$  and  $K$  such that,*

$$\|H_f(u) - H_g(v)\|_\infty \leq \|f - g\|_\infty + cT(1 + M)\|u - v\|_\infty. \quad (2.6)$$

*Proof.* From (2.4), we obtain by adding and subtracting the expression  $\int_0^T \int_\Omega K(x, y, v)(u(y, s) - u(x, s)) d\mu(y) ds$ , that

$$H_f(u)(x, T) - H_g(v)(x, T) = f(x) - g(x) + \\ \int_0^T \int_\Omega (K(x, y, u(\cdot, s)) - K(x, y, v(\cdot, s))) (u(y, s) - u(x, s)) d\mu(y) ds + \\ \int_0^T \int_\Omega K(x, y, v(\cdot, s))(u(y, s) - u(x, s) - v(y, s) + v(x, s)) d\mu(y) ds.$$

Therefore,

$$|H_f(u)(x, T) - H_g(v)(x, T)| \leq |f(x) - g(x)| + \\ \int_0^T \int_\Omega |K(x, y, u(\cdot, s)) - K(x, y, v(\cdot, s))| 2\|u\|_\infty d\mu(y) ds + \\ \int_0^T \int_\Omega |K(x, y, v(\cdot, s))| (|u(y) - v(y)| + |u(x) - v(x)|) d\mu(y) ds.$$

Hence, if  $B$  and  $L$  are an upper bound and the Lipschitz constant of  $K$  respectively, we obtain that

$$|H_f(u) - H_g(v)| \leq \|f - g\|_\infty + \\ T|\Omega|L2\|u\|_\infty\|u - v\|_\infty + \\ T|\Omega|B2\|u - v\|_\infty.$$

Finally, since  $\|u\|_\infty < M$ , we obtain the result.  $\square$

**Lemma 3.** *Let  $f \in C_{\frac{M}{2}}(\bar{\Omega})$  and  $u \in C_M(\bar{\Omega} \times [0, T])$ . Then,*

$$\|H_f(u)(\cdot, t)\|_\infty \leq \left(\frac{1}{2} + Tc(M+1)\right)M, \quad (2.7)$$

where  $c$  is a constant which depends only on  $\Omega$  and  $K$ . Moreover,

$$\|H_f(u) - H_f(v)\|_\infty \leq cT(1+M)\|u - v\|_\infty. \quad (2.8)$$

Therefore,  $H_f$  is a continuous operator from  $C_M(\bar{\Omega} \times [0, T])$  to  $C(\bar{\Omega} \times [0, T])$ .

*Proof.* Setting  $g = 0$  in (2.6) of Lemma 2, we obtain that

$$\|H_f(u)\|_\infty \leq \|f\|_\infty + Tc(M+1)\|u\|_\infty \leq \left(\frac{1}{2} + Tc(M+1)\right)M.$$

Furthermore, we already know from Lemma 1 that  $H_f(u)$  belongs to  $C(\bar{\Omega} \times [0, T])$ . Setting  $g = f$  in (2.6) we obtain (2.8).  $\square$

**Lemma 4.** *Given  $f \in C(\bar{\Omega})$ , there exists a  $t_0 > 0$  such that the initial value problem (2.1)–(2.2) has a unique solution  $u \in C(\bar{\Omega} \times [0, t_0])$ .*

*Proof.* Setting  $M = 2\|f\|_\infty$  and choosing  $t_0$  such that  $ct_0M(1+M) < \frac{1}{2}$  and  $ct_0(M+1) < 1$  in (2.7) and (2.8) respectively. It follows that the operator  $H_f$  is a contraction from  $C_M(\bar{\Omega} \times [0, t_0])$  to itself. The result thus follows from Banach's fixed point theorem.  $\square$

The previous lemma shows that the initial value problem has a unique solution locally. In order to extend it to a global well-posedness result we need the following stability property.

**Lemma 5.** *If  $f \in C(\bar{\Omega})$ , the local solution  $u$  of (2.1)–(2.2) given by Lemma 4 is such that*

$$\min_{z \in \bar{\Omega}} f(z) \leq u(\mathbf{x}, t) \leq \max_{z \in \bar{\Omega}} f(z) \quad \forall (\mathbf{x}, t) \in \bar{\Omega} \times [0, t_0]. \quad (2.9)$$

*Proof.* First we show that if  $f \geq 0$ , then  $u \geq 0$ . In fact, assume this is not the case and that  $u$  is negative somewhere. Define  $v(\mathbf{x}, t) = u(\mathbf{x}, t) + t\epsilon$ , with

a small  $\epsilon$  such that  $v$  is negative somewhere. Let  $(\mathbf{x}_0, t_0)$  be the point where  $v$  attains its minimum negative value, then  $t_0 > 0$  and

$$\begin{aligned} v_t(\mathbf{x}_0, t_0) &= u_t(\mathbf{x}_0, t_0) + \epsilon > \int_{\Omega} K(\mathbf{x}, \mathbf{y}, u)(u(\mathbf{y}, t_0) - u(\mathbf{x}_0, t_0))d\mu\mathbf{y} \\ &= \int_{\Omega} K(\mathbf{x}, \mathbf{y}, u)(v(\mathbf{y}, t_0) - v(\mathbf{x}_0, t_0))d\mu(\mathbf{y}), \end{aligned}$$

which is a contradiction.  $\square$

We are now ready to proof the main result of this section.

*Proof of Theorem 1.* The result is a consequence of the local existence result given by Lemma 4, and the fact that the  $\|u\|_{\infty}$  norm of the local solution is bounded by  $\|f\|_{\infty}$  as a consequence of Lemma 5. The continuous dependence of the solution on the initial value is given by Lemma 2.  $\square$

As a direct consequence of Lemma 5, we also obtain that the unique solution satisfies the following two properties.

**Proposition 1** (Preservation of Nonnegativity). *If  $u$  is a solution of (2.1)–(2.2) with  $f \geq 0$ , then*

$$\min_{(\mathbf{x}, t) \in \bar{\Omega} \times [0, T]} u(\mathbf{x}, t) \geq 0. \quad (2.10)$$

**Proposition 2** (Maximum–Minimum Principle). *If  $u$  is a solution of (2.1)–(2.2), then*

$$\min_{\mathbf{z} \in \bar{\Omega}} f(\mathbf{z}) \leq u(\mathbf{x}, t) \leq \max_{\mathbf{z} \in \bar{\Omega}} f(\mathbf{z}) \quad \forall (\mathbf{x}, t) \in \bar{\Omega} \times [0, T]. \quad (2.11)$$

We have thus obtained a general class of well-posed nonlocal processes given by the solutions of the initial value problem (2.3)–(2.3), supplemented with the assumptions (NL1) and (NL2). In the following section, we will proceed with the analysis of more specific class of processes, namely nonlocal scale-spaces. These will be obtained by imposing additional symmetry and nonnegativity constraint to the weight function  $K$  of (2.3)–(2.3).

## 2.2 Nonlocal scale spaces: symmetric evolutions

The goal of this section is to define a general class of nonlocal processes which satisfy scale-space properties, analogous to the ones described for anisotropic diffusion processes in [1]. We start by considering (2.1)–(2.2), together with (NL1) and (NL2), thus ensuring that we are dealing with well-posed processes. These two assumptions alone, however, are not sufficient to ensure that the corresponding unique solutions of the initial value problem gives rise to evolutions which satisfy the desired scale-space properties. For that matter, we make the following further assumptions.

- (NL3) Symmetry:**  $K(\mathbf{x}, \mathbf{y}, u) = K(\mathbf{y}, \mathbf{x}, u)$  in  $\bar{\Omega} \times \bar{\Omega}$ , for every  $u \in C(\bar{\Omega})$ .
- (NL4) Irreducibility:** There exists a finite family of  $\mu$ -measurable sets  $\mathcal{F} := \{B_i \subset \Omega : 1 \leq i \leq p\}$ , such that:
- (i) There exists a constant  $c > 0$ , independent of  $T$ , such that  $K(\mathbf{x}, \mathbf{y}, u(\cdot, s)) \geq c$  for  $0 \leq s \leq T$ , whenever  $u \in C(\bar{\Omega} \times [0, T])$ ,  $B \in \mathcal{F}$  and  $\mathbf{x}, \mathbf{y} \in B$ .
  - (ii)  $\Omega = \bigcup_{i=1}^p B_i$  and  $\mu(B_i \cap B_{i+1}) > 0$  for  $1 \leq i \leq p - 1$ .

Let us briefly describe the reason for supplementing (2.1)–(2.2) with the different assumptions introduced so far, by summarising their connections to the related scale-space properties discussed below. Condition (NL2) will be used to obtain a maximum-minimum principle. Condition (NL3) represents a relation of symmetry of the interaction between different parts of the domain where the evolution takes place. It is used in the proof of average grey value preservation. Condition (NL4) encodes a connectivity condition between different points of the image domain, and together with (NL3) is used to determine the long time behaviour of the solution. Under these assumptions it is shown that the evolution converges to the average grey value.

Let us now begin to describe in detail the different scale-space properties.

**Proposition 3** (Preservation of the Average Grey Value). *The solution of (2.1)–(2.2) preserves the average grey value:*

$$\frac{1}{\mu(\Omega)} \int_{\Omega} u(\mathbf{x}, t) d\mu(\mathbf{x}) = \frac{1}{\mu(\Omega)} \int_{\Omega} f(\mathbf{x}) d\mu(\mathbf{x}) \quad \text{for } 0 \leq t \leq T. \quad (2.12)$$

*Proof.* Integrating (2.3) over  $\Omega$  with respect to  $\mu$  and applying the Tonelli-Fubini theorem together with (NL3), we obtain

$$\begin{aligned} & \int_{\Omega} (u(\mathbf{x}, t) - f(\mathbf{x})) d\mu(\mathbf{x}) \\ &= \int_0^t \int_{\Omega} \int_{\Omega} K(\mathbf{x}, \mathbf{y}, u(\cdot, s))(u(\mathbf{y}, s) - u(\mathbf{x}, s)) d\mu(\mathbf{y}) d\mu(\mathbf{x}) ds. \\ &= - \int_0^t \int_{\Omega} \int_{\Omega} K(\mathbf{x}, \mathbf{y}, u(\cdot, s))(u(\mathbf{y}, s) - u(\mathbf{x}, s)) d\mu(\mathbf{y}) d\mu(\mathbf{x}) ds. \end{aligned} \quad (2.13)$$

This implies the result.  $\square$

Our next goal is to analyse the behaviour of the solution of (2.1)–(2.2) as  $T \rightarrow \infty$ . We will need the following lemma.

**Lemma 6.** *Let  $r : \mathbb{R} \rightarrow \mathbb{R}$  be a convex  $C^2$  function. If  $u$  is a solution of (2.1)–(2.2), then  $\frac{d}{dt} \int_{\Omega} r(u(\mathbf{x}, t)) d\mu(\mathbf{x})$  exists for  $t \in [0, T]$  (here we mean the right and left derivative for  $t = 0$  and  $t = T$ , respectively). Moreover, this expression is equal to  $\int_{\Omega} r'(u(\mathbf{x}, t)) \partial_t u(\mathbf{x}, t) d\mu(\mathbf{x})$ .*

*Proof.* Let  $t \in [0, T]$  and define  $F_h(\mathbf{x}) := \frac{1}{h} (r(u(\mathbf{x}, t+h)) - r(u(\mathbf{x}, t)))$ , for  $t, t+h \in [0, T]$ . Then, since  $r \in C^2$ , we obtain that  $\partial_t r(u(\mathbf{x}, t))$  exists for every  $(\mathbf{x}, t) \in \bar{\Omega} \times [0, T]$  and is equal to  $r'(u(\mathbf{x}, t)) \partial_t u(\mathbf{x}, t) = \lim_{h \rightarrow 0} F_h(\mathbf{x})$ . On the other hand, since  $u \in C(\bar{\Omega} \times [0, T])$ , we also know that

$$|F_h(\mathbf{x})| = |r'(u(\mathbf{x}, t_x))| \quad (2.14)$$

for some  $t_x \in [0, T]$  such that  $|t_x - t| < h$ . Thus, we may bound  $F_h(\mathbf{x})$  with a constant  $M > 0$  which is independent of  $x$  and  $h$ . This allows us to apply Lebesgue's convergence theorem to obtain that

$$\lim_{h \rightarrow 0} \int_{\Omega} F_h(\mathbf{x}) d\mu(\mathbf{x}) = \int_{\Omega} \lim_{h \rightarrow 0} F_h(\mathbf{x}) d\mu(\mathbf{x}), \quad (2.15)$$

as wanted.  $\square$

In what follows we will denote the constant function that is equal to the average grey value of  $f$  by

$$\tilde{u}(\mathbf{x}) := \frac{1}{\mu(\Omega)} \int_{\Omega} f(\mathbf{z}) d\mu(\mathbf{z}) \quad \forall \mathbf{x} \in \bar{\Omega}. \quad (2.16)$$

With this notation we can state the following result.

**Proposition 4** (Lyapunov Functionals). *Let  $u$  be the solution of (2.1)–(2.2). For any convex  $C^2$  function  $r : \mathbb{R} \rightarrow \mathbb{R}$ , the expression*

$$V(t) = \Phi(u(\cdot, t)) := \int_{\Omega} r(u(\mathbf{x}, t)) d\mu(\mathbf{x}) \quad (2.17)$$

*is a Lyapunov functional, in the sense that it satisfies the following properties.*

- (i)  $\Phi(u(\cdot, t)) \geq \Phi(\tilde{u})$  for all  $t \geq 0$ .
- (ii)  $V \in C^1[0, \infty[$  and  $V'(t) \leq 0$  for all  $t \geq 0$ .

*Moreover, if  $r'' > 0$ , then  $V(t)$  is even a strict Lyapunov functional, i.e.*

- (iii) *For all  $t \geq 0$  we have that  $\Phi(u(\cdot, t)) = \Phi(\tilde{u})$ , if and only if  $u(\cdot, t) = \tilde{u}$   $\mu$ -a.e. in  $\Omega$ .*
- (iv) *If  $t \geq 0$ , then  $V'(t) = 0$ , if and only if  $u(t) = \tilde{u}$   $\mu$ -a.e. in  $\Omega$ .*
- (v)  *$V(0) = V(T)$  for  $T > 0$ , if and only if  $\forall t \in [0, T] : u(\mathbf{x}, t) = \tilde{u}$   $\mu$ -a.e. in  $\Omega$ .*

*Proof.* (i) From Jensen's inequality and the preservation of the average grey value we obtain that

$$\Phi(u(\cdot, t)) = \int_{\Omega} r(u(\mathbf{z}, t)) d\mu(\mathbf{z}) \geq \int_{\Omega} r \left( \int_{\Omega} \frac{u(\mathbf{z}, t)}{\mu(\Omega)} d\mu(\mathbf{z}) \right) d\mu(\mathbf{y}) = \Phi(\tilde{u}). \quad (2.18)$$

(ii) From Lemma 6 we know that

$$V'(t) = \int_{\Omega} r'(u(\mathbf{x}, t)) \frac{d}{dt} u(\mathbf{x}, t) d\mu(\mathbf{x}). \quad (2.19)$$

Then, from (2.1)–(2.2) we obtain that

$$\begin{aligned} 2V'(t) &= 2 \int_{\Omega} \int_{\Omega} K(\mathbf{y}, \mathbf{x}, u(\cdot, t)) r'(u(\mathbf{x}, t))(u(\mathbf{y}, t) - u(\mathbf{x}, t)) d\mu(\mathbf{y}) d\mu(\mathbf{x}) \\ &= \int_{\Omega} \int_{\Omega} K(\mathbf{y}, \mathbf{x}, u(\cdot, t))(r'(u(\mathbf{x}, t)) - r'(u(\mathbf{y}, t))) \cdot \\ &\quad \cdot (u(\mathbf{y}, t) - u(\mathbf{x}, t)) d\mu(\mathbf{y}) d\mu(\mathbf{x}) \end{aligned} \quad (2.20)$$

where we used (NL3) for the second equality. Since  $r$  is convex we know that  $r'$  is nondecreasing. Therefore, the quantity  $(r'(u(\mathbf{x}, t)) - r'(u(\mathbf{y}, t))) \cdot (u(\mathbf{y}, t) - u(\mathbf{x}, t))$  is always nonpositive and it follows from the nonnegativity of  $K$  (NL2) that  $V'(t) \leq 0$ . Continuity of  $V(t)$  and  $V'(t)$  follows from the uniform continuity of  $u$  in  $\bar{\Omega} \times [0, T]$  and (2.20).

- (iii) If we assume that  $r$  is strictly convex, then we obtain from the strict Jensen's inequality that  $\Phi(u(\cdot, t)) = \Phi(\tilde{u})$  if and only if  $u(\mathbf{x}, t) = C$   $\mu$ -a.e. for some constant  $C$ . However, from the preservation of the average grey value (Proposition 3) we conclude that the only possibility is  $C = \tilde{u}$ , as wanted.
- (iv) From (2.20) and the irreducibility condition (NL4) we obtain that  $u$  is  $\mu$ -a.e. equal to a constant. However, this constant can only be  $\tilde{u}$  because of the preservation of the average grey value. This proves the result.
- (v) We use the fact that  $V$  is nonincreasing together with (iv).

□

As explained in [1], Lyapunov functionals guarantee that a scale-space acts image simplifying in many ways. By choosing specific strictly convex functions for  $r$ , it follows that the scale-space evolution reduces all  $L^p$  norms for  $p \geq 2$ , all even central moments, and it increases the entropy of the image. Last but not least, Lyapunov functionals are also useful for proving the following convergence result.

**Proposition 5** (Convergence). *Let  $u$  be a solution of (2.1)–(2.2). Then*

$$\lim_{t \rightarrow \infty} \|u(t) - \tilde{u}\|_{L^2(\Omega, \mu)} = 0. \quad (2.21)$$

*Proof.* Let  $v = u - \tilde{u}$  be the solution of (2.1)–(2.2) when  $f$  is replaced by  $f - \tilde{u}$ . If we consider the Lyapunov functional of Proposition 4 for the solution  $v$ , with the particular choice  $r(x) = x^2$  in the definition (2.17) of  $V$ , we get that

$$\lim_{t \rightarrow \infty} \|u(\cdot, t) - \tilde{u}\|_{L^2(\Omega, \mu)} = \ell, \quad (2.22)$$

for some finite value  $\ell \geq 0$ , as a consequence of (i) and (ii) of Proposition 4. Moreover, we know that

$$\int_0^\infty |V'(t)| dt \leq V(0) - \lim_{t \rightarrow \infty} V(t) < \infty. \quad (2.23)$$

This implies that there exists a sequence  $t_i$  such that  $\lim_{i \rightarrow \infty} t_i = \infty$  and  $\lim_{i \rightarrow \infty} V'(t_i) = 0$ , or equivalently,

$$\lim_{i \rightarrow \infty} \int_\Omega \int_\Omega K(\mathbf{y}, \mathbf{x}, u(\cdot, t_i)) (u(\mathbf{y}, t_i) - u(\mathbf{x}, t_i))^2 d\mu(\mathbf{y}) d\mu(\mathbf{x}) = 0. \quad (2.24)$$

Now, for every  $B$  in the family  $\mathcal{F}$  of condition (NL4) we may apply the Cauchy-Schwartz inequality to obtain

$$\begin{aligned} & \int_B \left| u(\mathbf{x}, t) - \frac{1}{\mu(B)} \int_B u(\mathbf{y}, t) d\mu(\mathbf{y}) \right|^2 d\mu(\mathbf{x}) \\ & \leq \frac{1}{\mu(B)} \int_B \int_B |u(\mathbf{y}, t) - u(\mathbf{x}, t)|^2 d\mu(\mathbf{x}) d\mu(\mathbf{y}) \\ & \leq \frac{1}{\mu(B)} \int_B \int_B \frac{K(\mathbf{x}, \mathbf{y}, u(\cdot, t))}{c} |u(\mathbf{y}, t) - u(\mathbf{x}, t)|^2 d\mu(\mathbf{x}) d\mu(\mathbf{y}), \end{aligned} \quad (2.25)$$

where  $c > 0$  is the lower bound for  $K$  in condition (NL4). Let us denote by  $h_k(t)$  the constant function defined on each  $B_k \in \mathcal{F}$  of condition (NL4) that is equal to  $\frac{1}{\mu(B_k)} \int_{B_k} u(\mathbf{x}, t) d\mu(\mathbf{x})$  for  $1 \leq k \leq p$ . The last inequality and (2.24) imply that

$$\|u(\cdot, t_i) - h_k(t_i)\|_{L^2(B_k, \mu)} \rightarrow 0. \quad (2.26)$$

Moreover, from (2.22) we know that  $u(\cdot, t_i)$  is bounded in  $L^2(\Omega, \mu)$ . Therefore, also  $h_k(t_i)$  is bounded. We may choose a subsequence of  $t_i$  which we continue to denote the same way, such that  $\lim_{i \rightarrow \infty} h_k(t_i)$  exists and is finite for  $1 \leq k \leq p$ .

Furthermore, the quantity  $\gamma := \min \{\mu(B_k \cap B_{k+1}) : 1 \leq k \leq p-1\}$  is positive because of (NL4). Therefore, we obtain that

$$\begin{aligned} \gamma |h_k(t) - h_{k+1}(t)|^2 &\leq |h_k(t) - h_{k+1}(t)|^2 \int_{B_k \cap B_{k+1}} d\mu(\mathbf{x}) \\ &= \|h_k(t) - h_{k+1}(t)\|_{L^2(B_k \cap B_{k+1}, \mu)}^2 \\ &\leq \|u(\cdot, t) - h_k(t)\|_{L^2(B_k, \mu)} + \|u(\cdot, t) - h_{k+1}(t)\|_{L^2(B_{k+1}, \mu)} \end{aligned} \quad (2.27)$$

for  $0 \leq k \leq p-1$ . These inequalities, together with (NL4) and (2.26), allow us to conclude that all  $h_k(t_i)$  converge to the same constant. Hence, it follows that  $u$  converges in  $L^2(\Omega, \mu)$  to a constant. This implies that the value  $\ell$  in (2.22) has to be 0 as wanted.  $\square$

## 2.3 Examples

The scale-space framework introduced above is quite general. It is possible to construct a variety processes by defining specific weight functions  $K$  which satisfy the required conditions (NL1)–(NL4). In what follows we will discuss some examples of such specific processes.

### 2.3.1 Linear continuous setting with shift-invariant weight functions

Let us consider the specific problem

$$\partial_t u(\mathbf{x}, t) = \int_{\Omega} J(\mathbf{x} - \mathbf{y}) (u(\mathbf{y}, t) - u(\mathbf{x}, t)) d\mathbf{y}, \quad \text{in } [0, T] \times \bar{\Omega}, \quad (2.28)$$

$$u(\mathbf{x}, 0) = f(\mathbf{x}) \quad \text{in } \bar{\Omega}, \quad (2.29)$$

with  $f \in L^1(\Omega)$  and some nonnegative radial function  $J \in C(\mathbb{R}^N)$  such that  $J(0) > 0$  and  $\int_{\mathbb{R}^N} J(\mathbf{x}) d\mathbf{x} = 1$ . Notice that since we restrict ourselves to continuous initial data, i.e.  $f \in C(\bar{\Omega})$ , it is not difficult to check that (2.28) satisfies all conditions (NL1)–(NL4). Thus, we may apply the results of the previous section, for  $\mu$  equal to the Lebesgue measure of  $\mathbb{R}^N$ .

Interestingly, this process was studied also in [26]. The authors proved that the family of solutions  $u_\epsilon$  of (2.28) with  $J$  replaced by an appropriate rescaled version  $J_\epsilon$ , approximates the solution of the usual Neumann problem for homogeneous diffusion. More precisely, if  $v$  is a solution of

$$\partial_t u = \Delta u \quad \text{in } ]0, T] \times \Omega, \quad (2.30)$$

$$\frac{du}{d\nu} = 0 \quad \text{on } \partial\Omega, \quad (2.31)$$

$$u(\mathbf{x}, 0) = f(\mathbf{x}) \quad \text{in } \Omega, \quad (2.32)$$

where  $\nu$  denotes the outer normal vector to  $\partial\Omega$ , then

$$\lim_{\epsilon \rightarrow 0} \|u_\epsilon - v\|_\infty = 0. \quad (2.33)$$

In other words, the nonlocal problem (2.28) solves a diffusion problem. Observe that using any kernel  $J$  as specified above will always lead us to Gaussian scale-space. This statement has its stochastic counterpart in the central limit theorem, which tells us that an iterated application of a smoothing kernel converges to a Gaussian. This motivates us to consider more general filters below, where the kernel can be a space-variant function of the initial image  $f$ .

### 2.3.2 Space discrete setting

Now we discuss the case when the measure  $\mu$  is a discrete measure concentrated on a finite subset of  $\Omega$ . We will focus on the one-dimensional case since the extension to higher dimension is straightforward.

First of all, let  $\Omega = ]0, 1[$  and let  $h = \frac{1}{M}$  for some fixed integer  $M > 1$ . Moreover, we define  $\mu$  as the restriction to  $\Omega$  of the discrete measure that is concentrated on the set  $\mathcal{Z}_h := \{h(z - \frac{1}{2}); z \in \mathbb{Z}\}$ . In what follows we set  $k_{i,j} := K((i - \frac{1}{2})h, (j - \frac{1}{2})h)$  and  $u_i(t) := u((i - \frac{1}{2})h, t)$  for  $1 \leq i, j \leq M$ . With these choices, the problem (2.1)–(2.2) becomes

$$\dot{u}_i = \sum_{j=1}^M K(i, j, \mathbf{u})(u_j - u_i) \quad (1 \leq i \leq M), \quad (2.34)$$

$$u_i(0) = f_i \quad (1 \leq i \leq M). \quad (2.35)$$

This is a semidiscrete evolution process for the vector  $\mathbf{u} := (u_1, u_2, \dots, u_M)^\top$ . Conditions (NL1)-(NL4) imply that the matrix  $\mathbf{K} = (k_{i,j})_{i,j=1}^M$  is symmetric, nonnegative, and irreducible. Notice that (2.34) can be written as

$$\dot{\mathbf{u}}(t) = \mathbf{A}\mathbf{u}(t), \quad (2.36)$$

$$\mathbf{u}(0) = \mathbf{f}, \quad (2.37)$$

where  $\mathbf{f} = (f_1, f_2, \dots, f_M)^\top$  and  $\mathbf{A} = (a_{i,j})_{i,j=1}^M$  is the matrix with entries

$$a_{i,j} = \begin{cases} k_{i,j} & (i \neq j), \\ -\sum_{n \neq i} k_{i,n} & (i = j). \end{cases} \quad (2.38)$$

This process satisfies all the properties of the semidiscrete framework for anisotropic diffusion considered in [1]. In fact, since  $\mathbf{K}$  is a matrix, the corresponding linear operator is Lipschitz-continuous. Moreover,  $\mathbf{K}$  is symmetric, has nonnegative entries, and is irreducible. Thus, it follows that  $\mathbf{A}$  is Lipschitz-continuous, symmetric, has nonnegative off-diagonal entries, and is irreducible. Moreover, (2.38) implies that  $\mathbf{A}$  has zero row sums. These are the conditions required in [1].

**Remark 1.** (a) *Note that the fact that in the linear case, the semidiscrete diffusion framework is covered by our nonlocal continuous framework is a benefit of our formulation in terms of measures.*

(b) *This also shows that Weickert's semidiscrete diffusion theory is more general than his continuous one, which requires local processes in terms of PDEs.*

(c) *Extensions to higher dimensions can be obtained by choosing  $\Omega \subset \mathbb{R}^N$  and the measure  $\mu_n := \mu \times \mu \times \dots \times \mu$ , where the product is taken  $n$  times, and  $\mu$  is a discrete measure concentrated on  $\mathcal{Z}_h := \{h(z - \frac{1}{2}); z \in \mathbb{Z}\}$  as above.*

### 2.3.3 A linear scale-space variant of bilateral filtering

Bilateral filtering goes back to Aurich and Weule [4] and became popular by a paper of Tomasi and Manduchi [5]. In a continuous notation, it filters a



Figure 2.1: Bilateral scale-space evolution (2.40) of a test image ( $256 \times 256$  pixels,  $\rho = 5$ ,  $\lambda = 10$ ). **From left to right:**  $t = 0$ , 500, 10000, and 200000.

greyscale image  $f : \Omega \rightarrow \mathbb{R}$  by means of spatial and tonal averaging with Gaussian weights:

$$u(\mathbf{x}) = \frac{\int_{\Omega} g_{\lambda}(|f(\mathbf{y}) - f(\mathbf{x})|) g_{\rho}(|\mathbf{y} - \mathbf{x}|) f(\mathbf{y}) d\mathbf{y}}{\int_{\Omega} g_{\lambda}(|f(\mathbf{y}) - f(\mathbf{x})|) g_{\rho}(|\mathbf{y} - \mathbf{x}|) d\mathbf{y}}, \quad (2.39)$$

where  $g_{\rho}(s) := \exp(-s^2/(2\rho^2))$  and  $g_{\lambda}(s) := \exp(-s^2/(2\lambda^2))$ . The positive constants  $\rho$  and  $\lambda$  control the range of values determined by the spatial and tonal weights respectively. While bilateral filtering is a nonlocal process, it does not preserve the average grey value. Moreover, it is typically applied in a noniterative way.

We propose the following modification that leads to an evolution equation:

$$\partial_t u(\mathbf{x}, t) = \frac{1}{c} \int_{\Omega} g_{\lambda}(|f(\mathbf{y}) - f(\mathbf{x})|) g_{\rho}(|\mathbf{y} - \mathbf{x}|) (u(\mathbf{y}, t) - u(\mathbf{x}, t)) d\mathbf{y}, \quad (2.40)$$

where  $c := \int_{\Omega} g_{\rho}(\mathbf{y}) d\mathbf{y}$  performs a normalisation of the spatial weighting. In our terminology, this is a nonlocal linear scale-space with the specific kernel  $K(\mathbf{x}, \mathbf{y}) = \frac{1}{c} g_{\lambda}(|f(\mathbf{y}) - f(\mathbf{x})|) g_{\rho}(|\mathbf{y} - \mathbf{x}|)$  and the Lebesgue measure  $\mu$ . It is straightforward to check that it satisfies the requirements (NL1)–(NL4) of our theory. This implies e.g. that it preserves the average grey value.

Figure 2.1 illustrates such a scale-space evolution. It has been obtained with an explicit finite difference scheme. As predicted by the theory, we observe that the image is gradually simplified. For  $t \rightarrow \infty$ , it converges to a flat steady state with the same average grey value as the initial image. It is noticeable how well the localisation of edges is preserved.



Figure 2.2: **(a) Left:** Noisy test image,  $128 \times 128$  pixels. **(b) Middle:** After processing with the robustified bilateral process (2.41) with  $\sigma = 2$ ,  $\rho = 5$ ,  $\lambda = 1.4$ , and  $t = 500$ . **(c) Right:** After rescaling the filtered result from (b) to the greyscale interval  $[0, 255]$ .

### 2.3.4 Robustified bilateral scale-space

While our bilateral integro-differential equation (2.40) gives an interesting scale-space evolution, its performance under noise is less favourable. The reason is easy to understand: Noise creates large values for  $|f(\mathbf{y}) - f(\mathbf{x})|$ , such that the corresponding tonal weight  $g_\lambda(|f(\mathbf{y}) - f(\mathbf{x})|)$  becomes very small. As a result, noisy structures are rewarded by a longer lifetime in scale-space. Similar problems are also well-known for the Perona–Malik diffusion filter [3]. Therefore, we can also use a similar strategy to overcome this problem: Following Catté et al. [27], we replace the image  $f$  in the argument of the tonal weight  $g_\lambda$  by a Gaussian-smoothed variant  $f_\sigma$ , where  $\sigma$  denotes the standard deviation of the Gaussian. Hence, our robustified bilateral evolution is given by

$$\partial_t u(\mathbf{x}, t) = \frac{1}{c} \int_{\Omega} g_\lambda(|f_\sigma(\mathbf{y}) - f_\sigma(\mathbf{x})|) g_\rho(|\mathbf{y} - \mathbf{x}|) (u(\mathbf{y}, t) - u(\mathbf{x}, t)) d\mathbf{y}. \quad (2.41)$$

Its behaviour is illustrated in Fig. 2.2. We observe that this process is well-suited for removing even a large amount of noise, while keeping the semantically important edge structures. Its performance is comparable to the edge-enhancing anisotropic nonlinear diffusion filter from [28]. However, this is achieved with a linear process, that does not require to struggle with the numerical challenges of implementing anisotropic filters with a diffusion tensor.

### 2.3.5 NL-means scale-space

The NL-means denoising algorithm [6] is given by the closed form,

$$NL(u)(\mathbf{x}) = \frac{1}{C(\mathbf{x})} \int_{\Omega} \exp\left(-\frac{(G_a(\mathbf{z}) * |u(\mathbf{x} + \mathbf{z}) - u(\mathbf{y} + \mathbf{z})|^2)(0)}{\lambda}\right) u(\mathbf{y}) d\mathbf{y},$$

where  $G_a$  is a Gaussian kernel with standard deviation  $a$ ,  $\lambda$  is a positive parameter and  $C(\mathbf{x}) = \int_{\Omega} \exp\left(-\frac{(G_a * |u(\mathbf{x} - \mathbf{z}) - u(\mathbf{y} - \mathbf{z})|^2)(0)}{\lambda}\right) d\mathbf{y}$ . The symbol  $*$  indicates a convolution w.r.t. the variable  $\mathbf{z}$ .

We propose a modification of NL-means denoising given by the evolution equation:

$$\partial_t u(\mathbf{x}) = \frac{1}{c} \int_{\Omega} K(\mathbf{x}, \mathbf{y}, u)(u(\mathbf{y}) - u(\mathbf{x})) d\mathbf{y}, \quad (2.42)$$

where  $K = \exp\left(-\frac{(G_a * |u(\mathbf{x} + \cdot) - u(\mathbf{y} + \cdot)|^2)(0)}{\lambda}\right) g_{\rho}(|\mathbf{y} - \mathbf{x}|)$ . Here, the first and second terms represent the exponentially weighted distance between patches used in NL-means and a spatial weigh respectively. Moreover,  $c := \int_{\Omega} g_{\rho}(\mathbf{y}) d\mathbf{y}$  performs a normalisation of the spatial weighting.

Unfortunately, denoising results are not comparable to NL-means results in terms of MSE, especially for higher levels of noise. We have partially found an explanation for this behaviour in the normalization constant  $c$ . In order to keep our process symmetric, and thus consistent with our framework, we use the same normalising constant on each position of the domain. This is of course in contrast to the position dependent normalising constant of the original NL-means algorithm, which creates a non-symmetric weighting function. This property translates in the fact that the our process acts extremely fast at homogeneous regions, with respect to regions which have jumps or texture. In fact, textures which do not have lot of similar patches locally, will only be denoised properly after the homogeneous regions have been flattened. This undesired behaviour is exemplified in Figure 2.3. The non-linearity seems to amplify this behaviour. This last statements was also observed when comparing the results of the linear bilateral scale-space above, to the denoising results of its nonlinear version given by

$$\partial_t u(\mathbf{x}, t) = \frac{1}{c} \int_{\Omega} g_{\lambda}(|u(\mathbf{y}) - u(\mathbf{x})|) g_{\rho}(|\mathbf{y} - \mathbf{x}|) (u(\mathbf{y}, t) - u(\mathbf{x}, t)) d\mathbf{y}. \quad (2.43)$$

One possible alternative evolution which considers a positions dependent normalisation which, however, preserves the symmetry of the weight, is given

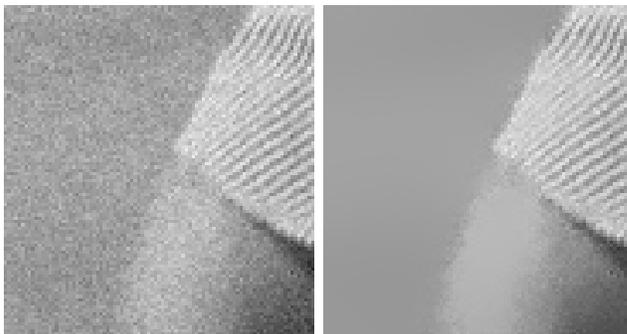


Figure 2.3: Nonlinear NL-means evolution.

by

$$\partial_t u(\mathbf{x}) = \int_{\Omega} \tilde{K}(\mathbf{x}, \mathbf{y}, u)(u(\mathbf{y}) - u(\mathbf{x})) d\mathbf{y},$$

with the weight  $\tilde{K}$  is given by the expression

$$\tilde{K}(\mathbf{x}, \mathbf{y}, u) = \frac{K(\mathbf{x}, \mathbf{y})}{\max\{\int_{\Omega} K(\mathbf{x}, \mathbf{z}) d\mathbf{z}, \int_{\Omega} K(\mathbf{z}, \mathbf{y}) d\mathbf{z}\}},$$

and the weight  $K$  defined as above.

Even though we obtained comparable results to NL-means for low levels of noise, this modification is not sufficient by itself to improve the performance of the process for higher levels of noise. Instead this seems to be an inherent behaviour of the symmetric nature of the weighting function. By enforcing symmetry, we obtain a process which has a smoothing effect with a strength which is proportional to the number of similar patches found in a particular region. Hence, homogeneous regions will be flattened at much higher speed, with respect to other parts of the image with contained less similar patches, like edges or textured regions. This behaviour can be observed in Figure 2.3.

## 2.4 Conclusions

In this chapter we have introduced the initial value problem (2.1)–(2.2) as a general framework for nonlocal scale spaces. We defined this framework by supplementing the abstract process (1.3)–(1.4) with a series of conditions about its weighting function  $K$ . First of all, we required the regularity and

nonnegativity conditions (NL1)–(NL2) in order to obtain the well-posedness of (2.1)–(2.2). These two conditions alone enabled us to prove Theorem 1, which guarantees the existence and uniqueness of a global solution for the initial value problem. Moreover, we also assumed the symmetry and irreducibility conditions (NL3)–(NL4) in order to obtain the series of scale-space properties studied in Section 2.2 which are in accordance with the theory of [29]. Finally, we showed different specific processes fulfilling the required assumptions of the general nonlocal scale-space framework. In particular, we introduced a linear scale-space variant of the bilateral filter and showed that its robustified version (2.41) has denoising capabilities which are comparable to non-linear diffusion. Furthermore, we also introduced a variant of nonlocal means denoising (2.42). We observed that although it has some denoising capabilities, the quality of the results is less satisfactory when compared to the original nonlocal means algorithm. We also identified the symmetry constraint (NL3) as a plausible explanation for this behaviour.

## Chapter 3

# Nonsymmetric nonlocal evolutions

The two main assumptions which allowed us to develop the scale-space theory of the previous chapter, were given by the symmetry conditions (NL3) and the irreducibility (NL4). We saw that both assumptions were fundamental, in order to show the average grey value preservation, the existence of different Lyapunov functionals, and the convergence of the solutions towards the average grey value of the input image as  $t \rightarrow \infty$ . In this chapter we will study the consequences of relaxing these requirements. We will continue working with the general framework (1.3)–(1.4) and keep the assumptions (NL1) and (NL2) in order to ensuring the well-posedness given by Theorem 1. However, the other two conditions will be replaced with less restrictive one, which will allow us to consider more general processes.

Regarding the relaxation of condition (NL3), previous work has been done for processes of a non-symmetric nature in the context of osmosis [30, 31]. In those papers the authors consider a type of linear drift diffusion PDE whose spatial discretisation leads to a semidiscrete process which is indeed non-symmetric. A fundamental difference with the processes studied in this chapter, lays in the fact that osmosis, on the one hand, is a linear process, and secondly, comes from a PDE whose differential operator is in divergence form. In particular, that means that it always preserves the average grey value of the initial image. Instead, the type of processes considered in this work, other than being non-linear, are also not necessarily in divergence form.

Among other things this means that while the non-symmetric nature of the process may lead to non-constant steady states, these could have a different average grey value than that of the input image. In terms of explicit finite difference schemes, osmosis leads to matrices with column sums 1, while the processes considered here lead to matrices with row sums 1. In that sense both theories are complementary.

The relaxation of condition (NL4) implies that we allow the possibility that at certain times of the evolution, parts of the domain may not be connected through the weight function  $K$ . These new type of more general processes will not necessarily fulfill the scale-space properties discussed in the previous chapter. Nevertheless, we will be able to prove other equally interesting results about their behaviour, in particular about their long time behaviour. In order to do that, we will first present some basic concepts about particle swarm systems and describe how they closely relate to our framework. We will then define substitute conditions for (NL3)–(NL4), which in turn will allow us to apply the general convergence results [32] which were developed in the context of swarm systems. The connections between particles swarms systems and nonlocal processes will be meaningful for the semidiscrete case, obtained when choosing a measure concentrated on a square grid which represents the pixel positions of an image, as described in Section 2.3.2. The convergence results will hence be valid in that context. We will also present two example applications of our framework for image processing tasks. First, in Section 3.4, we discuss the so called sector diffusion model, an alternative nonlinear diffusion filter which is specifically designed to preserve corners. Then in Section 3.5, we show coherence enhancement applications with comparable results to [33], but based on the non-symmetric filters considered in this chapter.

### 3.1 Preliminaries

Our basic framework will continue to be given by the initial value problem (2.1)–(2.2), which we now briefly recall. Given an image domain  $\Omega$  which is an open bounded subset of  $\mathbb{R}^N$ , a constant  $T > 0$ , and a locally finite Borel measure  $\mu$ . Let  $f \in C(\bar{\Omega})$  represent the input image, and consider the

following evolution process.

$$\partial_t u(\mathbf{x}, t) = \int_{\Omega} K(\mathbf{x}, \mathbf{y}, u(\cdot, t))(u(\mathbf{y}, t) - u(\mathbf{x}, t)) d\mu(\mathbf{y}) \quad \text{in } \bar{\Omega} \times [0, T], \quad (3.1)$$

$$u(\mathbf{x}, 0) = f(\mathbf{x}) \quad \text{in } \bar{\Omega} \quad (3.2)$$

Furthermore, recall the following regularity and non-negativity assumptions.

**(NL1) Regularity:**  $K : \bar{\Omega} \times \bar{\Omega} \times C(\bar{\Omega}) \rightarrow \mathbb{R}$ , is Lipschitz continuous and bounded.

**(NL2) Non-Negativity:** The function  $K$  is non-negative.

We will impose both conditions (NL1)–(NL2) in order to ensure that (3.1)–(3.2) is well-posed. In fact, we know from the theory developed in the previous chapter, that under these assumptions, the initial value problem has a unique global solution, as stated in Theorem 1. Moreover, these two assumptions are also sufficient to ensure a maximum–minimum principle, as given by Lemma 5. In other words we have the following results.

**Theorem 2** (Well-posedness). *For any given  $T > 0$ , the initial value problem (3.1)–(3.2) supplemented with the assumptions (NL1)–(NL2) has a unique solution. Moreover, this unique solution depends continuously on the initial value  $f$ .*

**Proposition 6** (Preservation of Nonnegativity). *If  $u$  is the solution of (3.1)–(3.2) with  $f \geq 0$ , then*

$$\min_{(\mathbf{x}, t) \in \bar{\Omega} \times [0, T]} u(\mathbf{x}, t) \geq 0. \quad (3.3)$$

**Proposition 7** (Maximum–Minimum Principle). *If  $u$  is the solution of (3.1)–(3.2), then*

$$\min_{z \in \bar{\Omega}} f(z) \leq u(\mathbf{x}, t) \leq \max_{z \in \bar{\Omega}} f(z) \quad \forall (\mathbf{x}, t) \in \bar{\Omega} \times [0, T]. \quad (3.4)$$

On the other hand, if the symmetry and irreducibility assumptions (NL3)–(NL4) are removed, it is clear that the rest of the scale-space properties obtained for the processes of the previous chapter cannot longer be guaranteed, since both assumptions play a main role in their proofs. In what follows, we will choose appropriate substitute assumptions that will allow us to develop a meaningful theory for more general processes than the ones treated so far. Our focus here is mainly on semidiscrete nonlocal processes. Some of the analysis carried out in the rest of this chapter may be adapted to a the general case of a measure  $\mu$ . However, the main convergence results are derived from concepts which are specific for the semidiscrete setting. A complete convergence result for a general  $\mu$  is therefore out of the scope of this work.

Let us thus begin by describing in detail the basic framework of the semidiscrete nonlocal process which will be discussed in the rest of this chapter. Let us assume that  $\Omega$  is a rectangle representing the domain of an image  $f$ . Moreover, let  $\mu$  be given by the discrete measure that is concentrated on an  $M \times N$  regular grid which represents the set of coordinates of its pixels values. This grid can be constructed by adapting the procedure described in Section 2.3.2, to the two dimensional setting of a rectangular image. Moreover, let  $\pi : C(\bar{\Omega}) \rightarrow \mathbb{R}^{MN}$  be the projection that maps a continuous function to its set of values at the  $M \times N$  grid coordinates given by  $\mu$ . We represent the projected values as a single vector in  $\mathbb{R}^{NM}$  using a lexicographical ordering of their grid coordinates. We do the same for the grid point values of  $f$ , in such a way that a single index can be use to run over all the set of pixels. With these choices, process (3.1)–(3.2) becomes

$$\dot{u}_i = \sum_{j=1}^{MN} K(i, j, \pi(u))(u_j - u_i) \quad (1 \leq i \leq MN), \quad (3.5)$$

$$u_i(0) = f_i \quad (1 \leq i \leq MN). \quad (3.6)$$

Since this process is a special case of the general framework, and we are still assuming (NL1)–(NL2). All the properties valid for the general process are also valid in this case. In particular, a continuous solution  $\mathbf{u}(t) \in \mathbb{R}^{NM}$  of the process exists for all times  $t$ , and it fulfills a maximum-minimum principle. The upcoming analysis will be derived from theories first studied in the context of systems control and in particular on the analysis of swarm systems. For that reason, in the following section we proceed with the introduction

of some basic concepts and results about that topic. In particular, we will introduce the consensus problem [34], a widely studied topic in the systems control literature. This will give us enough ground to connect our nonlocal processes to such theories, and will allow us to apply the general results developed in [32].

## 3.2 Particle swarm systems: consensus problem

Literature on particle swarm systems goes back a long time, and several researchers have studied the dynamics of such models in contexts as diverse as Biology, Physics, Mathematics, Economy or Engineering [35–37]. Even though its wide spectrum of applications makes swarm model quite diverse, they can generally be described as a set of particles, usually modelled as points in some Euclidean space, together with a set of rules defining the interaction between them. The goal is then to study the dynamics of these points based on the given rules, with the aim to characterize as much as possible their behaviour.

In this work we will mainly be concerned with the so called consensus problem which has been actively studied in systems control during the last years [34]. The consensus problem refers to the study of the long time behaviour of interacting swarm particle system (also denoted as multiagent systems in the control literature), with the aim of characterizing sets of interaction rules, which would lead to an agreement of the asymptotic position of all particles. More precisely, we will be concerned with the consensus problem for particle systems whose positions at time  $t$  are represented as a set of  $M$  points  $\mathbf{x}_1(t), \mathbf{x}_2(t), \dots, \mathbf{x}_M(t) \in \mathbb{R}^N$ . Each of the particles updates its position based on its current position and time varying attractive strengths towards the other particles, encoded in a system of ordinary differential equations as follows.

$$\dot{\mathbf{x}}_i(t) = \sum_{j=1}^M a_{ij}(t)(\mathbf{x}_j(t) - \mathbf{x}_i(t)), \quad (1 \leq i \leq M), \quad (3.7)$$

$$\mathbf{x}_i(0) = f_i \quad (1 \leq i \leq M). \quad (3.8)$$

The values  $a_{ij}$  represent a strength of attraction that the particle with index  $j$  has over the particle with index  $i$ , at any given time  $t$ .

The attractive nature of the interaction comes from the fact that the values  $a_{ij}(t)$  are assumed to be non negative, thus encouraging the attraction between particles which have similar positions. In general these values may change over time. In that case, we assume that they change continuously in  $t$ .

**Connectivity Graph:** A very useful concept related to the attractive weights in (3.7)–(3.8), is that of a connectivity graph. The main idea is that for every fixed time, we consider the particles as nodes of a graph. The values  $a_{ij}(t)$  thus represent directed arcs with weights going from the node corresponding to  $\mathbf{x}_j$  to the one corresponding to  $\mathbf{x}_i$ . A weight value of zero would mean that there is no attractive strength, or equivalently, no directed arc from  $\mathbf{x}_j$  to  $\mathbf{x}_i$ . We thus obtain at any given time, a corresponding connectivity graph which represents the instantaneous interactions between the particles.

**Graph Laplacian:** The connectivity graph also allow us to introduce the concept of the graph Laplacian. It is a useful representation that encodes the structure of a weighted graph, in terms of a matrix. It is widely used in topics like spectral graph theory [38], [39], [40].

In general, the construction of the graph Laplacian can be done in the following way. If we have a graph consisting of  $M$  nodes, the graph Laplacian is defined as an  $M \times M$  matrix  $\mathbf{A}$ , whose non-diagonal values in the row  $i$  and column  $j$ , are give by the weight of the directed arc going from node  $j$  to node  $i$  (zero if there is no directed arc between the corresponding nodes).

Moreover, its diagonal values are such that its row sums are zero.

In the case of our system (3.7)–(3.8) , we will have a different graph Laplacian  $\mathbf{A}(t)$  for each given time. This set of matrices allows us to rewrite the system it the following way when  $N = 1$  and the particles live in  $\mathbb{R}$ .

$$\dot{\mathbf{x}} = \mathbf{A}(t)\mathbf{x}, \tag{3.9}$$

$$\mathbf{x}(0) = f. \tag{3.10}$$

Here  $\mathbf{x} \in \mathbb{R}^M$  is a vector having as components the positions of the single particles. The previous discussion gives us enough ground for introducing the

main result of Hendrickx and Tsitsiklis [32], which we will apply in the next section to our nonlocal framework (3.5)–(3.6). Let us begin by introducing the following definition which is valid for any graph and its corresponding graph Laplacian  $\mathbf{A}$ .

**Definition 2.** (*Directed path*) We say there is a directed path from node  $i$  to node  $j$  if there exists a sequence of distinct indexes  $i = i_1, i_2, \dots, i_m = j$ , such that  $a_{i_k, i_{k+1}} > 0$  for  $1 \leq k < m$ . In other words, if it is possible to go from node  $i$  to node  $j$  following arcs which have a positive weight.

The following two definitions concern the set of graphs obtained from the system (3.9)–(3.10) and their corresponding graph Laplacians.

**Definition 3.** (*Type-symmetry*) We say that the swarm system with dynamics (3.9)–(3.10) satisfies the type-symmetry condition, if there exists a fixed positive constant  $c$  such that  $a_{i,j}(t) \leq c \cdot a_{j,i}(t)$ , for every pair  $i, j$  and every time  $t$ .

**Definition 4.** (*Unbounded interaction graph*) We define the unbounded interactions graph,  $G = (\{1, \dots, M\}, E)$ , by letting  $(i, j) \in E$  if  $\int_0^\infty a_{ij}(t) dt = \infty$ . Moreover, we say  $i$ , and  $j$  belong to the same connected component of  $G$  if there is a directed path from  $i$  to  $j$  or from  $j$  to  $i$ .

We also need the following lemma.

**Lemma 7.** Assume that the swarm system satisfies the type-symmetry condition. Then, if there is a directed path from node  $i$  to node  $j$  in its unbounded interaction graph, there is also a path from node  $j$  to node  $i$ .

*Proof.* See [32]. □

With the above premises, we are now ready to state the main result.

**Theorem 3.** Assume that the swarm system (3.7)–(3.8), equivalently (3.9)–(3.10), satisfies the type-symmetry condition. Then, the long time behaviour of its solution has the following properties:

(i) The limit  $x_i^* = \lim_{t \rightarrow \infty} x_i(t)$  exists.

(ii) If  $i$  and  $j$  belong to the same connected component of the unbounded interaction graph  $G$ , then  $x_i^* = x_j^*$ .

*Proof.* See [32]. □

Now that we have stated the main technical tools, the next step will be to establish a connection between our process (3.5)–(3.6) and the theory of swarm system developed in this section.

### 3.3 Nonlocal nonsymmetric processes: convergence results

We now return to the main topic of the chapter and continue analysing the semidiscrete framework (3.5)–(3.6). Recall that our goal is that of defining a new set of assumptions which serve as alternatives for the symmetry and irreducibility conditions (NL3)–(NL4). The theory of swarms systems discussed in the previous section has the right ingredients for that purpose. For that reason, let us proceed by establishing a clear relation between both worlds. A strong connection can already be seen by the fact that both processes can be represented as solutions of systems of the ordinary differential equations (3.5)–(3.6) and (3.7)–(3.8). The two systems are essentially the same if we restrict the second to scalar valued components by choosing  $P = 1$ . The main difference is given by the fact that the first one is nonlinear while the second one is linear. We can make this relationship even stronger by reinterpreting the nonlocal semidiscrete system as a particle swarm system. For that matter, consider the evolution of the image pixels, given by process (3.5)–(3.6). Each pixel has some fixed grid coordinates  $i$  (here we use a single index which runs through all pair of coordinates of the two-dimensional image grid), and a corresponding grey value  $u_i(t)$  which varies with time. Hence, among the quantities related to a given pixel, the only one that changes with time is that of its grey value. We may therefore interpret the whole image process as a system of particles with positions in the interval  $I := [0, 255]$ . Their interaction rule is then inherited from the function  $K$  and depends on their current positions in the interval  $I$  and their former coordinates on the image grid. The former coordinate, can now be interpreted as fixed properties and can be represented as tags attached to each of the particles constituting the one-dimensional system. See Figure 3.1.

Summing up, we may conclude that the process (3.5)–(3.6) can be seen as

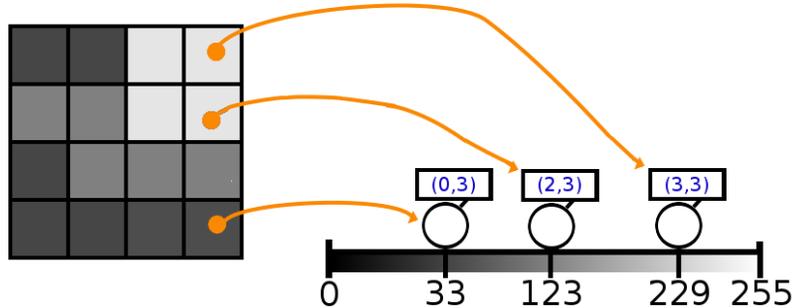


Figure 3.1: Mapping of pixels to particle system in the interval  $[0,255]$

an attractive process between particles contained in  $I$ , where the attractive strength at any given time, is defined by the function  $K$ . Under this interpretation, we can also consider a graph Laplacian  $\mathbf{A}$  for each given  $t$ , like we did in the previous section. However, in contrast to (3.9)–(3.10), here the process is non-linear. Therefore, the matrix  $\mathbf{A}$  depends explicitly on the current particle positions, and the dynamics are thus given by an expression of the following type:

$$\dot{\mathbf{u}} = \mathbf{A}(t, \mathbf{u})\mathbf{u},$$

where  $\mathbf{u}$  is a vector of  $M \times N$  coordinates which contains the positions of all particles.

The nonlinear nature of the nonlocal semidiscrete system is however only apparent. In fact, we may reinterpret it as an equivalent linear process. More precisely, we know from Theorem 1 that the solution of the system exists for all times and that it is bounded. This, together with (NL1), implies that there exists a bounded continuous matrix valued function  $\tilde{\mathbf{A}}(t)$  such that  $\tilde{\mathbf{A}}(t) = \mathbf{A}(t, \mathbf{u}(t))$  for all  $t \geq 0$ . In other words we have the following result.

**Lemma 8.** *The solution of process (3.5)–(3.6) satisfy a non-autonomous linear equation*

$$\dot{\mathbf{u}} = \tilde{\mathbf{A}}(t)\mathbf{u}, \quad (3.11)$$

where  $\tilde{\mathbf{A}} : [0, \infty[ \rightarrow \mathbb{R}^{MN \times MN}$  is a continuous matrix valued function, having non-negative off-diagonal elements and zero row sums at every time  $t$ .

We are now ready to introduce the first substitute for the symmetry and irreducibility assumptions of the first chapter. It will allow us to apply the

results from consensus theory and will be sufficient to characterize the long time behaviour of the solutions of (3.5)–(3.6).

**(NL5) Type-symmetry:** There exists a constant  $c > 0$  such that  $c^{-1}k(i, j, \cdot) < k(i, j, \cdot) < ck(i, j, \cdot)$ .

Condition (NL5) means there are positive upper and lower bounds for the ratio between the attractive strength from  $i$  to  $j$  and the attractive strength from  $j$  to  $i$ . In particular if one of them is positive, the other one must be positive. With this condition, we obtain the following result.

**Theorem 4.** (*Convergence with type-symmetry*) Assume the evolution (3.5)–(3.6) satisfies condition (NL5). Then, the limits  $u_i^* := \lim_{t \rightarrow \infty} u_i(t)$  exist for every  $1 \leq i \leq NM$ . Moreover, if  $i$  and  $j$  belong to the same connected component of the corresponding unbounded interaction graph, it follows that  $u_i^* = u_j^*$ .

*Proof.* The result is mainly a consequence of the above discussion. In fact, we have seen that (3.5)–(3.6) can be represented as an attractive swarm system. Moreover, thanks to Lemma 8 we know that it satisfies the same type of non autonomous linear equation as (3.9)–(3.10). It is therefore enough to notice that the type-symmetry condition (NL5) is actually equivalent to the one required for Theorem 3. In fact, in this case the components  $\tilde{a}_{i,j}(t)$  of the matrix  $\tilde{\mathbf{A}}$  of Lemma 8 are given by  $k(i, j, \pi(u))$ , and the theorem follows.  $\square$

We may say even more about the process if instead of the type symmetry, we replace it with the following assumption.

**(NL6) Local bounded connectivity:** There exists positive values  $P, p > 0$  such that for every pair of pixels  $i, j$ , either  $p \leq K(i, j, \pi(u)) \leq P$  or  $K(i, j, \pi(u)) = 0$ . Moreover, if  $i$  and  $j$  are neighbouring pixels in the image grid, then  $k(i, j, \pi(u)) > p$ .

The above condition is stronger than condition (NL5) since bounding the attractive strengths bounds also their ratios. Moreover, the fact that neigh-

bouring pixels have positive attractive strength implies that the corresponding unbounded interaction graph is connected. With this assumption we obtain the following stronger convergence result.

**Theorem 5.** (*Convergence with local bounded connectivity*) *Assume the evolution (3.5)–(3.6) satisfies condition (NL5). Then, there exists a value  $0 \leq Q \leq 255$ , such that  $u_i(t) \rightarrow Q$ , as  $t \rightarrow \infty$ , for every  $0 \leq i \leq NM$ .*

*Proof.* Like in the previous proof, we start by noticing that (3.5)–(3.6) can be represented as an attractive swarm system, and that Lemma 8 implies that it satisfies the same type of nonautonomous linear equation as (3.9)–(3.10). Moreover, condition (NL5) implies that for any given  $t$ , the off-diagonals of the matrix  $\tilde{\mathbf{A}}(t)$  are either confined to some interval  $[p, P]$  or 0. This last property implies the type-symmetric property (NL5) and Theorem 4 may be applied. Furthermore, the fact that neighbouring pixels always have positive attractive weight greater than  $p > 0$ , implies that the corresponding unbounded connectivity graph of the system has only one connected component. The existence of the value  $Q$  thus follows applying the second part of Theorem 4. The fact that  $Q \in [0, 255]$  follows from the maximum-minimum principle.  $\square$

In the special case when the interaction between pixels is symmetric, namely, when the corresponding matrix  $\mathbf{A}$  is symmetric, we recover the long time behaviour of the framework developed in Chapter 2 and in previous works [29], [25]. In fact, notice that assumption (NL5) is stronger than the irreducibility condition (NL4). Therefore, it is clear that reintroducing symmetry leads as once again to the setting of semidiscrete scale-spaces. Here, however, we obtained the convergence result using methods of particle swarm systems. More precisely we have the following result.

**Collorary 1.** (*Symmetric weights*) *With the same conditions of the last theorem, assume we have the additional property that  $k(i, j, \cdot) = k(j, i, \cdot)$ , for every pair of pixel indexes  $i, j$ . Then, the value  $Q$  is given by the average of the initial input image  $f$  and*

$$\lim_{t \rightarrow \infty} u_i(t) = \frac{1}{NM} \sum_{k=1}^{NM} f_k,$$

for every  $1 \leq i \leq NM$ .

*Proof.* It is enough to notice that with this additional condition, the matrix  $\tilde{\mathbf{A}}$  of Lemma 8 becomes symmetric and has zero column sums. Multiplying (3.11) to the left with a vector  $(1, 1, \dots, 1) \in \mathbb{R}^{NM}$ , we obtain that

$$\sum_{k=1}^{MN} \dot{u}_k = 0.$$

This proves the result.  $\square$

### 3.3.1 Vector valued processes

The main idea behind the convergence results for the nonlocal semidiscrete system (3.5)–(3.6), was that of reinterpreting it as a swarm system. Notice, however, that even though the former system is scalar, since it deals with the evolution of grey valued images, the swarm system of Section 3.2 deals with swarm system in which the particles are represented as points moving in a general Euclidean space of dimension  $P$ . We could thus relate also vector valued image evolutions to swarm system of particles in  $\mathbb{R}^P$ , with  $P > 1$  and obtain analogous statements like Theorem 4 and Theorem 5. For example, if we were dealing with colors images, we could represent colors in the RGB space and relate the image evolution to the system (3.7)–(3.8) choosing  $P = 3$ .

In that case, the corresponding vector valued system would be given by,

$$\dot{\mathbf{u}}_i = \sum_{j=1}^{MN} K(i, j, \pi(\mathbf{U}))(\mathbf{u}_j - \mathbf{u}_i) \quad (1 \leq i \leq MN), \quad (3.12)$$

$$\mathbf{u}_i(0) = \mathbf{f}_i \quad (1 \leq i \leq MN). \quad (3.13)$$

where this time  $\mathbf{u}_i$  represent vector values, instead of grey values, attached to each pixel. A corresponding representation analogous to that of Lemma (8) would be given by

$$\dot{\mathbf{U}} = \tilde{\mathbf{A}}(t)\mathbf{U},$$

where  $\mathbf{U}$  is a matrix with rows given by the vectors  $\mathbf{u}_i$ ,  $0 \leq i \leq NM$ . Let us finish this section by discussing another type of image evolution involving vectors. It will be useful for the example discussed in Section 3.5. It is given

by the solution of

$$\begin{aligned}\dot{\mathbf{u}}_i &= \sum_{\substack{\pm \mathbf{u}_j \\ j \in \mathcal{N}_i}} K(i, j, \mathbf{u}_i, \mathbf{u}_j) (\mathbf{u}_j - \mathbf{u}_i), \\ \mathbf{u}_i(0) &= \nabla f_i,\end{aligned}\tag{3.14}$$

where  $\nabla f_i$  represents the gradients of some input image, and  $\mathcal{N}_i$  represents a neighbourhood in the image grid of the pixel  $i$ . This process cannot be directly obtained from system (3.12), because of the  $\pm \mathbf{u}_j$  appearing under the sum symbol. However, we can still interpret it as a swarm system by defining two different particles for each pixel. One with position  $\mathbf{u}_j$  and one with position  $-\mathbf{u}_j$ .

### 3.4 Application: sector diffusion

The main topic of this section will be to introduce the so called Sector Diffusion, as an alternative diffusion model aimed at having good performance in terms of corner preservation. The content of this section is the result of a joint work with Kireeti Bodduna and Joachim Weickert.

Before we introduce the actual model, let us begin by discussing some preliminary motivations regarding diffusion and corner preservation.

Non-linear diffusion models are able to process the variation of an image in all orientations, in such a way that diffusion is made along edges, but not across them. However, even though anisotropic diffusion filters have a good performance in terms of these requirements, there is still room for improvement in terms of corner preservation. Corner enhancing/preserving filtering is a less explored and difficult research area when compared to edge enhancing/preserving filtering. Corners form the basis for most of the present day computer vision systems which are built for object detection, motion detection, image registration, video tracking and a plethora of other applications. The most common definition of a corner and the definition we adhere to in this work is : "a corner is a point where multiple dominant directions or edges exist in the local neighborhood".

A few works which propose filters that are motivated on the idea to preserve corners are [41], [42], [43] and [44]. The first three propose anisotropic smoothing methods which initially determine whether a particular location

is a corner or not and perform smoothing accordingly. In [44], a novel anisotropic smoothing based energy minimisation method is developed. They find the two dominant directions in a local neighbourhood in which anisotropic smoothing is simulated, thus preserving the corners. However, there can be situations where more than two edges intersect. In such situations one needs to perform smoothing in more than two directions. Also, none of the above mentioned methods claim to have a denoising performance equivalent to that of well known filters like NL Means [45], [6] [46]. In this Section we propose the Sector Diffusion Model which preserves corners, can smooth in multiple directions, does not additionally require to determine whether a particular location belongs to a corner region, and most importantly, it has superior denoising capability than NL Means filter.

Our model is directly inspired by the work of Weickert [47]. In that paper, the author studied the anisotropic diffusion given by the following evolution process

$$\partial_t u(\mathbf{x}, t) = \frac{1}{\pi} \int_0^\pi \partial_\theta (g(\partial_\theta u_\sigma) \partial_\theta u) d\theta, \quad (3.15)$$

where  $\partial_\theta$  stands for the directional derivative in the direction represented by angle  $\theta$ ,  $g$  is the diffusivity function and  $u_\sigma$  denotes a convoluted version of  $u$  with a Gaussian of standard deviation  $\sigma$ .

This model considers each orientation separately and is thus capable of diffusing along edges, but not across them. For our purposes, the main point which could be an advantage in terms of corner preservation is to treat directions, rather than orientation, individually. One possible attempt in that direction is given by the following modification of (3.15),

$$\partial_t u(\mathbf{x}, t) = \frac{1}{2\pi} \int_0^{2\pi} \partial_\theta^+ (g(\partial_\theta^+ u_\sigma^+) \partial_\theta^+ u) d\theta, \quad (3.16)$$

where this time  $u_\sigma^+$  represents a one sided smoothing of  $u$  in the direction given by the angle  $\theta$ , and  $\partial_\theta^+$  denotes a one sided derivative in the same direction. In contrast to the usual Gaussian smoothing applied to (3.15), this one sided smoothing allows the filter to distinguish two different derivatives for a given orientation, one in the direction of  $\theta$ , and the other in the direction of  $\theta + \pi$ . A formal definition of these concepts, could be made by considering the restriction of  $u$  to the corresponding ray starting at  $\mathbf{x}$ , in the direction of each  $\theta$ . Namely, for fixed  $\mathbf{x}, t, \theta$ , we consider  $u(h; \mathbf{x}, t) := u(\mathbf{x} + h(\cos(\theta), \sin(\theta))^T, t)$ , for  $h \in [0, \infty]$ . Then, the one sided

directional derivative  $\partial_\theta^+$  could be formally defined as

$$\partial_\theta^+ u := \lim_{h \rightarrow 0^+} \frac{u(h; \mathbf{x}, t) - u(\mathbf{x}, t)}{h}. \quad (3.17)$$

A complete theoretical analysis of this local model is out of the scope of the present work. Here, we would like to combine this idea with the concept of nonlocality by considering the following continuous evolution,

$$\partial_t u(\mathbf{x}, t) = \int_{B_{x,\rho}} \frac{1}{|\mathbf{y} - \mathbf{x}|} g \left( \frac{v_\sigma(\mathbf{x}, \mathbf{y}) - v_\sigma(\mathbf{x}, \mathbf{x})}{|\mathbf{y} - \mathbf{x}|} \right) \frac{u(\mathbf{y}) - u(\mathbf{x})}{|\mathbf{y} - \mathbf{x}|} d\mathbf{y}. \quad (3.18)$$

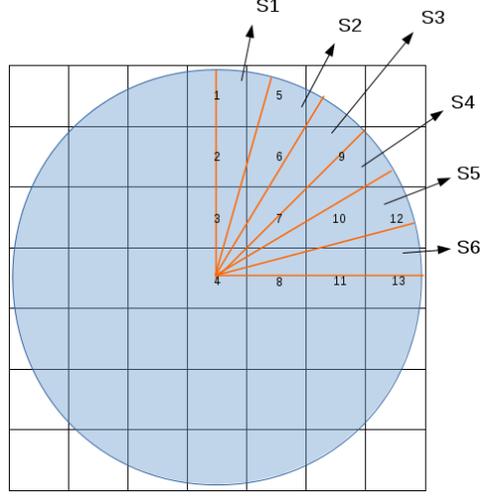
Here,  $B_{x,\rho}$  denotes the ball with center  $x$  of radius  $\rho$ , and the function  $v_\sigma$  is given by

$$v_\sigma(\mathbf{y}, \mathbf{x}) = \frac{1}{c} \int_0^\rho a(\mathbf{x}, \mathbf{y}, \lambda_{\mathbf{x}\mathbf{y}}(s), \sigma) b(\mathbf{z}, \mathbf{y}) u(s) ds.$$

where  $\lambda_{\mathbf{x}\mathbf{y}}(s) = \mathbf{x} + s \frac{(\mathbf{y} - \mathbf{x})}{|\mathbf{y} - \mathbf{x}|}$ ,  $s \in [0, \rho]$ , parametrizes the segment that goes over  $\mathbf{y}$ , starts from the center of the ball, and finishes on its boundary. Moreover,  $c$  is a normalizing constant. The functions  $g$ ,  $a$  and  $b$  are nonlinear weighting functions that will be described precisely in the following subsection, where the corresponding semidiscrete process will be analysed.

### 3.4.1 Semidiscrete modelling of sector diffusion

We first consider a disk shaped neighborhood of radius  $\rho$  around every pixel and divide it into sectors. For such a division, we map a particular pixel to all sectors it has a non-zero area of intersection with. This would give rise to a distribution where each pixel can be mapped to multiple sectors. There are two advantages of such a distribution of pixels over a distribution which maps a pixel to just one sector. First of all, the former distribution always gives rise to non-empty sectors, with enough pixels in each sector. The latter might give a distribution where a sector contains just a few pixels far away from the center pixel. This paves the way for the second advantage: the former distribution also gives a continuous distribution of pixels in each sector which helps in recognizing the edges. This is not possible if we distribute one pixel to just one sector. Figure 3.2 shows the distribution of the pixels in the disk shaped neighbourhood to their respective sectors. Now that we



**Distribution of pixels into sectors:**

S1 – 1, 2, 3, 5, 6, 4    S4 – 8, 7, 10, 9, 4  
 S2 – 3, 2, 7, 6, 5, 4    S5 – 8, 7, 10, 12, 11, 4  
 S3 – 3, 7, 6, 9, 4    S6 – 8, 11, 10, 13, 12, 4

Figure 3.2: Mapping of pixels to sectors

have clarified how the image pixels are distributed into the sectors, we are ready to define the image evolution of Sector Diffusion. It is given by the solution of the following system, where  $f_i$  represents the grey values of an input image and  $u_i(t)$  their evolutions. The index  $i$  goes over all coordinates of the grid defining the pixel positions like in (3.5)–(3.6).

$$\dot{u}_i = \sum_{l=1}^M \sum_{j \in S_l} \frac{1}{|x_j - x_i|} \left( g \left( \frac{u_{\sigma jl} - u_{\sigma il}}{|x_j - x_i|} \right) \frac{u_j - u_i}{|x_j - x_i|} \right), \quad (3.19)$$

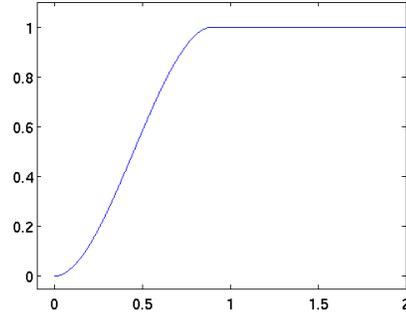
$$u_i(0) = f_i \quad (3.20)$$

Here  $g$  is the diffusivity [29] given by

$$g(s) = 1 - \exp \left( \frac{-3.31488}{\left(\frac{s}{\lambda}\right)^8} \right), \quad (3.21)$$

where  $\lambda$  is a positive parameter,  $M$  is the number of sectors,  $S_l$  the set of pixels within a particular sector  $l$ ,  $x_i$  and  $x_j$  denote the position of the pixels  $i$  and  $j$  in the image grid,  $u_{\sigma jl}$  is defined as

$$u_{\sigma jl} = \frac{1}{c} \sum_{k \in S_l} a(i, j, k, \sigma) b(k, j) u_k, \quad (3.22)$$

Figure 3.3: Plot of function  $h$ 

where  $c$  is a normalization constant and

$$a(i, j, k, \sigma) = h(|k - i|) \exp\left(\frac{-|x_k - x_j|^2}{2\sigma^2}\right), \quad (3.23)$$

$$b(k, j) = (\max - |u_k - u_j|), \quad (3.24)$$

where "max" denotes the highest possible grey value (255.0 for an 8-bit grayscale image), and

$$h(s) = \begin{cases} -2.7435s^3 + 3.7037s^2 & 0 \leq s < 0.9, \\ 1 & \text{otherwise.} \end{cases} \quad (3.25)$$

The value of  $u_{\sigma il}$  in (3.19) is also computed using (3.22).

The following is the intuition behind the modelling of Sector Diffusion: The decisive idea for preserving the corners in the modelling of Sector Diffusion is the distribution of the disk shaped neighborhood into sectors. A corner is the intersection of edges. Hence it is clear that in the local neighborhood where corners exist, regions of similar grey values occur in the shape of sectors. Thus smoothing within these sectors would give better estimates of  $u_{\sigma jl}$  and  $u_{\sigma il}$  using (3.22), instead of smoothing within the entire disk shaped neighbourhood.

Moreover, (3.19) can be seen as is a nonlocal semidiscrete approximation of (3.17). Thus, our process is closer to (3.16), which deals with one sided directional derivatives instead of the directional derivatives of (3.15) which have the characteristic of smoothing over the central pixel and destroying the corners. Also, we use smoothed differences  $\mathbf{u}_{\sigma jl} - \mathbf{u}_{\sigma il}$  instead of  $\mathbf{u}_{jl} - \mathbf{u}_{il}$  to remain robust with respect to noise [27]. The specific smoothing in (3.22) is inspired by bilateral filtering [4, 5] where  $a$  and  $b$  denote the radial and tonal

weights respectively. Such a smoothing was found to be better than a standard Gaussian smoothing because the presence of tonal weights introduces a smoothing that is robust under noise and preserves edges and corners.

The following is the reason for the presence of the function  $h(|k-i|)$  (plotted in Figure 3.3) in the function  $a$ : if one looks closely at equation (3.19),  $\mathbf{u}_{\sigma il}$  is present in every term of the summation. Thus, it is important that the estimate of  $\mathbf{u}_{\sigma il}$  is robust with respect to noise. However, when  $i = k = j$ , the following situation arises in estimating this critical value  $\mathbf{u}_{\sigma il}$  using (3.22): tonal weight  $b$  will have the highest weight (equal to "max"), the radial weight (in the absence of  $h(|k-i|)$ ) will also have the highest weight as the Gaussian has the maximum value at 0 (when  $x_k = x_i$ ). This means that, if the pixel  $\mathbf{u}_i$  is noisy, the entire estimation of  $\mathbf{u}_{\sigma il}$  is incorrect because of the high radial and tonal weights for pixel  $i$ . This was also experimentally verified. Thus the presence of the function  $h(|k-i|)$  in the function  $a$  is critical as it suppresses the contribution of the pixel  $i$ , which is necessary in the above mentioned scenario. This finally means that both the estimates  $u_{\sigma jl}$  and  $u_{\sigma il}$  are calculated without any contribution of pixel  $i$ . For implementation purposes in a discrete setting, not having the function  $h(|k-i|)$  in the (3.23) would be equivalent to excluding the center pixel  $i$  from all the sectors. Moreover, it has to be mentioned that, in applications that do not involve noise, we can ignore  $h(|k-i|)$  in the function  $a$ .

**Connections to the nonlocal semidiscrete theory** The Sector Diffusion evolution (3.19) can be seen as an example of a process (3.5)–(3.6). In this particular case, the function  $K$  is given by

$$K(i, j, \pi(u)) = \sum_{l \in S_j} h(d(i, j)) \cdot g \left( \frac{\mathbf{u}_{\sigma jl} - \mathbf{u}_{\sigma il}}{|\mathbf{x}_j - \mathbf{x}_i|} \right).$$

Here  $h$  is an appropriate continuous cut-off function which depends on  $\rho$  and takes as argument the distance between pixels  $i$  and  $j$  denoted as  $d(i, j)$ . Moreover,  $S_j$  denotes the set of sectors to which pixel  $j$  belongs to.

The Sector Diffusion (3.19) is therefore a well defined process, having a global solution which satisfies a maximum-minimum principle. Furthermore, the argument  $s$  of the diffusivity  $g$  can be at most 255. Since  $g$  is decreasing, the process satisfies (NL6) with values  $p$  and  $P$  given by  $g(256)$  and 1. In

conclusion, applying Theorem 5 we obtain that all pixels converge to some common grey value  $c$  as  $t \rightarrow \infty$ .

### 3.4.2 Time discretisation of the sector diffusion scheme

The explicit scheme of (3.19) can be written as

$$\mathbf{u}_i^{k+1} = \mathbf{u}_i^k + \tau \left( \sum_{l=1}^M \sum_{j \in S_l} g \left( \frac{\mathbf{u}_{\sigma_{jl}^k} - \mathbf{u}_{\sigma_{il}^k}}{|\mathbf{x}_j - \mathbf{x}_i|} \right) \frac{\mathbf{u}_j^k - \mathbf{u}_i^k}{|\mathbf{x}_j - \mathbf{x}_i|^2} \right). \quad (3.26)$$

The above equation can be written in a more compact way using a matrix  $\mathbf{Q}(\mathbf{u}^k) = q_{i,j}(\mathbf{u}^k)$ :

$$\mathbf{u}^{k+1} = \mathbf{Q}(\mathbf{u}^k) \mathbf{u}^k, \quad (3.27)$$

where the entries of the matrix  $\mathbf{Q}(\mathbf{u}^k)$  are given by

$$q_{i,j}(\mathbf{u}^k) = \begin{cases} 1 - \sum_{l=1}^M \sum_{j \in S_l} \frac{\tau \cdot g(\mathbf{u}^k)}{|\mathbf{x}_j - \mathbf{x}_i|^2} & j = i, \\ \frac{-\tau \cdot g(\mathbf{u}^k)}{|\mathbf{x}_j - \mathbf{x}_i|^2} & i \neq j \in B_\rho(i), \\ 0 & \text{otherwise} \end{cases} \quad (3.28)$$

with diffusivity weight  $g(\mathbf{u}^k) = g \left( \frac{\mathbf{u}_{\sigma_j^k} - \mathbf{u}_{\sigma_i^k}}{|\mathbf{x}_j - \mathbf{x}_i|} \right)$ . The values of  $\mathbf{u}_{\sigma_{jl}^k}$  and  $\mathbf{u}_{\sigma_{il}^k}$  would be different in the case where  $u_j$  is the center pixel and  $u_i$  is the center pixel (they would have different set of neighbours in both cases). Thus,  $\mathbf{Q}(\mathbf{u}^k)$  is not symmetric.

Now let us investigate the maximum-minimum principle fulfillment and the average grey value preservation of Sector Diffusion. The row sums of the matrix  $\mathbf{Q}(\mathbf{u}^k)$  are equal to 1:

$$\begin{aligned} & \left( \sum_{l=1}^M \sum_{j \in S_l} \tau \cdot g \left( \frac{\mathbf{u}_{\sigma_j^k} - \mathbf{u}_{\sigma_i^k}}{|\mathbf{x}_j - \mathbf{x}_i|} \right) \frac{1}{|\mathbf{x}_j - \mathbf{x}_i|^2} \right) + 1 - \\ & \left( \sum_{l=1}^M \sum_{j \in S_l} \tau \cdot g \left( \frac{\mathbf{u}_{\sigma_j^k} - \mathbf{u}_{\sigma_i^k}}{|\mathbf{x}_j - \mathbf{x}_i|} \right) \frac{1}{|\mathbf{x}_j - \mathbf{x}_i|^2} \right) = 1. \end{aligned} \quad (3.29)$$

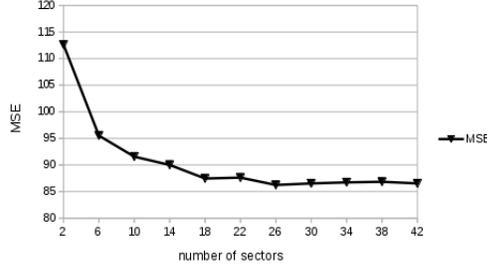


Figure 3.4: Peppers image denoising (std. dev. = 40). Parameters:  $\sigma = 0.6$ ,  $\lambda = 3.4$ .

Noisy Image (std. dev.)	$\sigma$	$\lambda$	$t$	SD (MSE)	NL Means (MSE)
Bridge (40)	0.3	9.5	$3\tau$	<b>261.18</b>	287.62
Bridge (60)	0.6	7.6	$6\tau$	<b>353.89</b>	425.18
Bridge (80)	0.6	7.8	$9\tau$	<b>429.60</b>	523.29
Peppers (40)	0.6	3.4	$11\tau$	<b>86.84</b>	93.75
Peppers (60)	0.6	4.1	$13\tau$	<b>133.08</b>	145.93
Peppers (80)	0.6	4.4	$15\tau$	<b>189.03</b>	204.31

Table 3.1: Denoising evaluation of Sector Diffusion with NL Means

Given that the maximum value of diffusivity is 1 the above upper bound on the time step size simplifies to

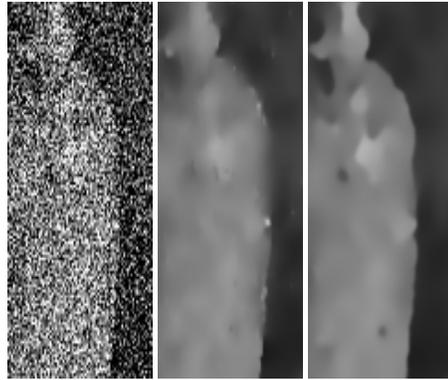
$$\tau \leq \frac{1}{\sum_{l=1}^M \sum_{j \in S_l} \frac{1}{|\mathbf{x}_j - \mathbf{x}_i|^2}}. \quad (3.30)$$

For the maximum-minimum principle to hold, we also additionally need all the entries of the matrix  $\mathbf{Q}(\mathbf{u}^k)$  to be non-negative, for which the following condition must hold:

$$1 - \sum_{l=1}^M \sum_{j \in S_l} \tau \cdot g\left(\frac{\mathbf{u}_{\sigma j}^k - \mathbf{u}_{\sigma i}^k}{|\mathbf{x}_j - \mathbf{x}_i|}\right) \frac{1}{|\mathbf{x}_j - \mathbf{x}_i|^2} \geq 0. \quad (3.31)$$

This gives an upper bound on the time step size

$$\tau \leq \frac{1}{\sum_{l=1}^M \sum_{j \in S_l} g\left(\frac{\mathbf{u}_{\sigma j}^k - \mathbf{u}_{\sigma i}^k}{|\mathbf{x}_j - \mathbf{x}_i|}\right) \frac{1}{|\mathbf{x}_j - \mathbf{x}_i|^2}}. \quad (3.32)$$



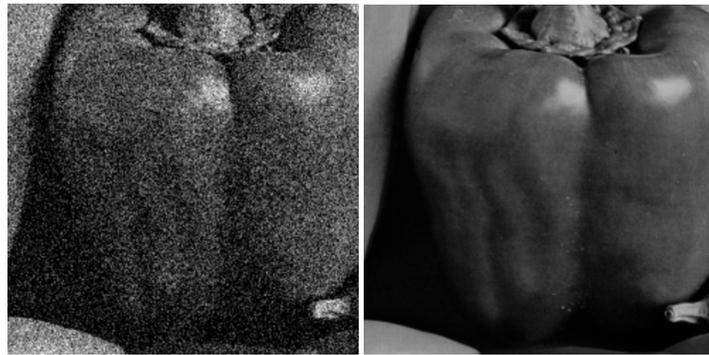
(a) Noisy image (b) 1 sector (c) 36 sectors

Figure 3.5: Denoising of Peppers image with Sector Diffusion. (a) Noisy image (std. dev. = 80), (b) SD with 1 sector ( $\lambda = 17.5$ ,  $\tau = 0.0619$ ,  $\sigma = 0.1$ ,  $t = 7 \cdot \tau$ , **MSE**: 227.47), (c) SD with 36 sectors ( $\lambda = 4.4$ ,  $\tau = 0.0119$ ,  $\sigma = 0.6$ ,  $t = 20 \cdot \tau$ , **MSE**: 189.03).



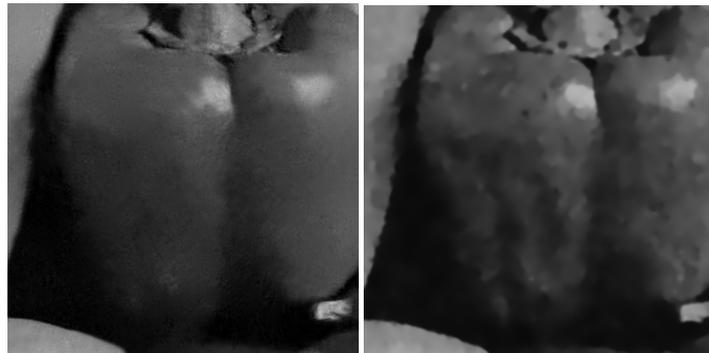
(a) Noisy image (b) EED (c) SD

Figure 3.6: Corner preservation and denoising. (a) Noisy image (std. dev. = 25), (b) Edge enhancing diffusion ( $\lambda = 4.0$ ,  $\tau = 0.2$ ,  $\sigma = 1.3$ ,  $t = 50 \cdot \tau$ ), (c) Sector diffusion ( $\lambda = 1.4$ ,  $\tau = 0.0119$ ,  $\sigma = 0.6$ ,  $t = 26 \cdot \tau$ ).



(a) Noisy Peppers

(b) Original Image



(c) NL means (MSE = 93.75)

(d) SD (MSE = 86.84)

Figure 3.7: Denoising Peppers (std. dev. = 40)

### 3.4.3 Influence of the number of sectors on denoising

Throughout this paper, we have used only white Gaussian noise for corrupting the image. Figure 3.4 shows the Mean Squared Error (MSE) values of the obtained reconstructed Peppers images after denoising, as a function of the number of sectors chosen. It is clear from the figure that a high number of sectors leads to better reconstructed images. One can also see that after a certain number of sectors the quality of the reconstructed image is almost the same. The reason is the following: When many sectors are chosen, it is more likely that all the sectors are divided into regions of similar grey value. Hence, sector smoothing gives correct estimates of  $\mathbf{u}_{\sigma jl}$  and  $\mathbf{u}_{\sigma il}$ , in turn leading to a diffusion process that preserves edges and corners. This is less likely when we do not divide the disk shaped neighbourhood into sectors. Figure 3.5 shows the results of denoising Peppers image with 1 sector and 36 sectors. In the first case the only sector  $S_1$  corresponds to the disk shaped neighborhood of radius  $\rho$  and is thus equivalent to the case of no division in sectors. In that case only the inner summation of (3.19) is relevant and is computed for the whole disk shaped neighborhood. It is clear from the reconstructed images and the MSE values that a high number of sectors leads to superior quality results (the mentioned parameters in Figure 3.5 have been optimized with respect to the MSE value). Both these experiments signify the importance of distribution of the disk shaped neighbourhood into sectors.

### 3.4.4 Parameter selection

We have six parameters: the smoothing parameter  $\sigma$ , the contrast parameter  $\lambda$ , the radius of the disk shaped neighbourhood  $\rho$ , the total diffusion time  $t$ , the number of sectors  $M$  and the time step size  $\tau$ . However, we do not have to optimize all parameters. From Section 3.4.3 we have seen that for a high enough number of sectors we have good reconstruction quality and rotational invariance, thus, in all our further experiments we have used 36 sectors. Also from (3.26), we can see that we have a  $\frac{1}{|\mathbf{x}_j - \mathbf{x}_i|^2}$  dependence. After  $|\mathbf{x}_j - \mathbf{x}_i| > 6$ , these weights drop by 97.5% which decreases the influence of such neighbors on the diffusion process to a negligible amount. Thus, we have chosen the radius of the disk shaped neighbourhood  $\rho = 6$  in all our

experiments. For the time step size  $\tau$ , we have chosen it to be 95% for the upper bound mentioned in (3.30). For 36 sectors and a radius of 6, 95% of the upper bound turns out to be 0.011872. Finally, we have to optimize for only 3 parameters: the smoothing parameter  $\sigma$ , the contrast parameter  $\lambda$  and the diffusion time  $t$ .

### 3.4.5 Denoising evaluation

Figure 3.6 shows the corner preserving and denoising capabilities of Sector Diffusion after denoising. It can be clearly seen that Sector Diffusion is a better corner preserving filter than the edge enhancing diffusion filter [29].

Figure 3.7 shows the results of denoising Peppers image with Sector Diffusion and NL means. Although NL means produces visually pleasing results, it destroys certain image structures which are preserved by Sector Diffusion. Also, from Table 3.1, it is clearly evident from the MSE values that Sector Diffusion outperforms NL means. Especially, the MSE values are significantly better for highly noisy images (with more standard deviation of white Gaussian noise). The parameters  $\lambda$ ,  $\sigma$  and  $t$  have been optimized with respect to the mean squared error. It can be seen from the Table 3.1 that the smoothing parameter  $\sigma$  is almost constant ( $\approx 0.6$ ) irrespective of the standard deviation of noise with which the image has been corrupted. This means we have to optimize for only 2 parameters:  $\lambda$  and  $t$ .

The other surprisingly good result of Sector Diffusion when compared to the denoising with NL means is the following: For a noise of standard deviation 80, the NL means uses a search window of size  $35 \times 35$  pixels and a neighbourhood of size  $11 \times 11$ . This means information from a  $41 \times 41$  neighbourhood is being used for denoising. In Sector Diffusion on the other hand, we have a radius of the disk shaped neighbourhood  $\rho = 6$ . This means we are using just 113 pixels which is 6.7% of the pixels being used by NL means and still getting significantly better results. Thus the obvious future work would be to combine the ideas for patch based methods with Sector Diffusion for best results that are both visually pleasing and preserve image structures.

We attribute the superior corner preservation and denoising capability of Sector Diffusion model to three specific model features: diffusion within a "non-local" disk shaped neighbourhood, distribution of the nonlocal disk shaped



Figure 3.8: **Idea of coherence enhancing image filtering. Left:** input image showing an interrupted curve. **Right:** filtered image with closed gap.

neighbourhood into "sectors" using "area-based" distribution of pixels and finally "robust smoothing within these sectors" with a "bilateral filter" inspired idea.

### 3.4.6 Datasets and implementations

The test images Texmos (texture mosaic) used in Figure 3.6, and the image Peppers used in Figures 3.5 and 3.7 have been taken from the University of Southern California image database (<http://sipi.usc.edu/database/>). We have used the implementation by Buades et al. [46] for simulating NL means algorithm ([http://www.ipol.im/pub/art/2011/bcm\\_nlm/](http://www.ipol.im/pub/art/2011/bcm_nlm/)).

## 3.5 Application: coherence enhancing image filter

In the following, we demonstrate how the nonlocal theory developed in this chapter can be used to construct a coherence enhancing image filter. This is a joint work with Leif Bergerhoff and Joachim Weickert. The aim of coherence enhancing is the filling-in of missing information in a given image while preserving its original structure (cf. Figure 3.8). Challenging conditions during image acquisition often lead to noise or gaps in the image structure. In those cases coherence enhancing filters represent an important and necessary preprocessing step for many computer vision applications. Filters which fulfill the previously mentioned criteria can e.g. be found in [33], [48], and [49]. All of these approaches have in common that they employ the structure

tensor [50] to estimate the so-called coherence orientation in a first step. Afterwards, this information is used in the main filtering phase. We present a novel two-step approach for coherence enhancing image filtering which is based on the nonlocal semidiscrete theory developed earlier in this chapter. We follow the idea that grey values should propagate orthogonally to the gradient direction in order to increase coherence. For this purpose, we first estimate the *coherence orientation* in the image and then define a *grey value evolution* which makes the grey values move along that orientation.

**Connections to the nonlocal semidiscrete theory** Let us begin by describing the general type of process considered in this section. For that matter, let  $\mathbf{u}_i(t)$  denotes a vector value assigned to a pixel and let the index  $i$ ,  $0 \leq i \leq MN$ , run over all pixels of the image grid. Moreover,  $\mathbf{x}_i$ ,  $0 \leq i \leq MN$ , represents the two dimensional coordinates of the pixel  $i$  in the image grid. The general process on which the two steps of the proposed coherence enhancing filter is based, is given by the following system:

$$\dot{\mathbf{u}}_i(t) = \sum_{j \in \mathcal{N}_i} w(\mathbf{u}_i, \mathbf{u}_j) k(|\mathbf{u}_i - \mathbf{u}_j|) (\mathbf{u}_j - \mathbf{u}_i), \quad (3.33)$$

where  $|\cdot|$  denotes the Euclidean norm and we assume that  $\mathbf{u}_i = \mathbf{u}_i(t)$ . An appropriate selection of the neighbourhood  $\mathcal{N}_i \subseteq \{1, \dots, NM\}$  ensures that only nearby pixels influence the behaviour of pixel  $i$ . Notice that this system can be derived from the nonlocal semidiscrete theory discussed in Section (3.3.1). In this setting, the weight function  $K$  is given by  $w(\mathbf{x}_i, \mathbf{x}_j) k(|\mathbf{x}_i - \mathbf{x}_j|) \eta(|\mathbf{x}_i - \mathbf{y}_j|)$ , where  $\mathbf{x}_i, \mathbf{x}_j$  represent the two dimensional positions of the pixels  $i$  and  $j$  in the image grid, and  $\eta$  is a continuous cut-off function which determines the neighbourhoods  $\mathcal{N}_i$ . The system matrix  $\tilde{A}(t) \in \mathbb{R}^{NM \times NM}$ , given by the matrix product representation is the same for all dimensions. In this setting, its entries read

$$\tilde{a}_{ij}(t) = \begin{cases} 0, & \text{if } j \notin \mathcal{N}_i \\ w(\mathbf{u}_i, \mathbf{u}_j) k(|\mathbf{u}_i - \mathbf{u}_j|), & \text{if } j \in \mathcal{N}_i \\ -\sum_{j \in \mathcal{N}_i} w(\mathbf{x}_i, \mathbf{x}_j) k(|\mathbf{u}_i - \mathbf{u}_j|), & \text{if } j = i. \end{cases} \quad (3.34)$$

The specific systems which will be considered in the rest of this section satisfy the type-symmetry condition (NL5). Therefore, a vector valued version of

Theorem 4 holds accordingly. Clearly, neither the particle weights  $w(\mathbf{u}_i, \mathbf{u}_j)$  nor – as a consequence – the matrix  $\tilde{A}(t)$  need to be symmetric.

**Discretisation** In order to solve the initial value problem related to (3.33), we approximate the time derivative by its forward difference and obtain the explicit scheme

$$\mathbf{u}_i^{k+1} = \mathbf{u}_i^k + \tau \cdot \sum_{j \in \mathcal{N}_i} w(\mathbf{u}_i^k, \mathbf{u}_j^k) k(|\mathbf{u}_i^k - \mathbf{u}_j^k|) (\mathbf{u}_j^k - \mathbf{u}_i^k). \quad (3.35)$$

This iterative process can – for every dimension – be written as

$$\mathbf{u}^{k+1} = (\mathbf{I} - \tau \tilde{\mathbf{A}}(\mathbf{u}^k)) \mathbf{u}^k = \mathbf{Q}(\mathbf{u}^k) \mathbf{u}^k, \quad (3.36)$$

with  $\mathbf{I}, \mathbf{Q}(\mathbf{u}^k) \in \mathbb{R}^{NM \times NM}$ , where  $\mathbf{I}$  denotes the identity matrix. Again, the matrix  $\mathbf{Q}(\mathbf{u}^k)$  is the same for all dimensions. Furthermore, this matrix has unit row sums. In order to fulfill the maximum-minimum principle, one also needs to ensure non-negativity of the matrix entries  $q_{ij}^k$ . Thus, we require  $\tau \geq 0$  and for all  $i = 1, \dots, NM$ , that

$$1 - \tau \cdot \sum_{j \in \mathcal{N}_i} w(\mathbf{u}_i^k, \mathbf{u}_j^k) k(|\mathbf{u}_i^k - \mathbf{u}_j^k|) \geq 0. \quad (3.37)$$

Based on this, and the fact that  $0 \leq k(\cdot) \leq 1$ , we derive the following bounds for the time step size  $\tau$ :

$$0 \leq \tau \leq \frac{1}{\max_i \sum_{j \in \mathcal{N}_i} w(\mathbf{u}_i^k, \mathbf{u}_j^k)}. \quad (3.38)$$

### 3.5.1 The two-step approach in detail

Now, that we have drawn the connections to theory, let us take a deeper look at our two-step algorithm for coherence enhancing image filtering.

#### Coherence orientation

As mentioned before, the goal of the first step is the estimation of the coherence orientation. For this purpose, recent filters usually make use of the struc-

ture tensor [50]. We, however, follow a different strategy and employ a nonlocal approach. Subsequently, we design a nonlocal model in such a way that it allows us to estimate the *local dominant gradient orientation*  $\nabla \tilde{f}$ . The latter depends on the gradient of the input image  $\nabla f_\sigma : \{1, \dots, MN\} \rightarrow [-1, 1]^2$  (the subscript  $\sigma$  denotes the standard deviation of a Gaussian kernel which is used for pre-smoothing the input image  $f$ ). Afterwards, the desired coherence orientation can be derived from every vector  $\nabla \tilde{f}_i$  with non-zero magnitude. Namely, it is given by the orthogonal vector denoting the *local dominant tangent orientation*  $\nabla^\perp \tilde{f}_i$ .

Now, given the pre-smoothed gradient field  $\nabla f_\sigma$ , consider the two-dimensional evolution

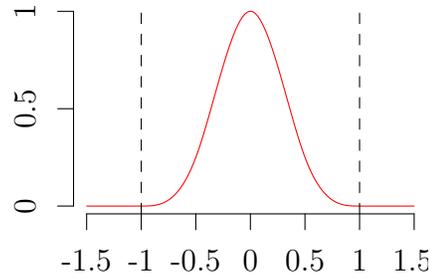
$$\begin{aligned} \dot{\mathbf{u}}_i &= \sum_{\substack{\pm \mathbf{u}_j \\ j \in \mathcal{N}_i}} w(\mathbf{u}_i, \mathbf{u}_j) k_1(|\mathbf{u}_j - \mathbf{u}_i|) (\mathbf{u}_j - \mathbf{u}_i), \\ \mathbf{u}_i(0) &= \nabla f_{\sigma,i}, \end{aligned} \quad (3.39)$$

which describes the local alignment of the gradient vector  $\nabla f_{\sigma,i}$  in each pixel  $i$ , being described by the corresponding particle position  $\mathbf{u}_i$ . The particle velocity  $\dot{\mathbf{x}}_i$  depends on all gradient directions  $\mathbf{u}_j$ , as well as their negative counterparts  $-\mathbf{u}_j$ , of all mates  $j$  that lie within a disk-shaped neighbourhood  $\mathcal{N}_i$  of radius  $d_1$  in the image plane. The weights

$$w(\mathbf{u}_i, \mathbf{u}_j) = \begin{cases} 0, & \text{if } \mathbf{u}_j^T \mathbf{u}_i \leq 0 \\ \frac{|\mathbf{u}_j|^2 + \varepsilon}{\varepsilon} \cdot \mathbf{u}_j^T \mathbf{u}_i, & \text{if } 0 < \mathbf{u}_j^T \mathbf{u}_i \leq \varepsilon \\ |\mathbf{u}_j|^2 + \varepsilon, & \text{else.} \end{cases} \quad (3.40)$$

ensure that  $\mathbf{u}_i$  only aligns with those vectors  $\pm \mathbf{u}_j$ , which deviate from its own orientation by less than  $\pi/2$ . As a consequence, at most one of both vectors  $\mathbf{u}_j$  and  $-\mathbf{u}_j$  can influence the behaviour of  $\mathbf{u}_i$  at a time. Additionally, the weights  $w(\mathbf{u}_i, \mathbf{u}_j)$  approximate the idea that vectors with large magnitude should guide those with small modulus, but not the other way round. This asymmetry is implemented by multiplication with the term  $|\mathbf{u}_j|^2 + \varepsilon$ . By applying a scaled cubic B-spline as kernel function

$$k_1(x) := \begin{cases} 1 - 6x^2 + 6|x|^3, & \text{if } 0 \leq |x| < \frac{1}{2} \\ 2 \cdot (1 - |x|)^3, & \text{if } \frac{1}{2} \leq |x| < 1 \\ 0, & \text{else.} \end{cases} \quad (3.41)$$

Figure 3.9: Kernel function  $k_1$  as given in (3.41).

small differences  $|\mathbf{u}_j - \mathbf{u}_i|$  lead to higher kernel weights, whereas large ones induce small kernel weights (see Figure 3.9). System (3.39) was described in (3.14), where we discussed how it connects to the swarm systems theory and its convergence results. Denote its steady state of by

$$\mathbf{u}^* := \lim_{t \rightarrow \infty} \mathbf{u}(t). \quad (3.42)$$

Since for the second step of our algorithm – the grey value evolution – we are not interested in the magnitude, but only the local dominant gradient orientation, we define  $\nabla \tilde{f}_i$  as

$$\nabla \tilde{f}_i := (\nabla \tilde{f}_{i,1}, \nabla \tilde{f}_{i,2})^T = \begin{cases} (0, 0)^T, & \text{if } |\mathbf{u}^*| = 0 \\ \frac{\mathbf{u}^*}{|\mathbf{u}^*|}, & \text{else.} \end{cases} \quad (3.43)$$

Then, the desired coherence orientation is given by the local dominant tangent orientation

$$\nabla^\perp \tilde{f}_i := (-\nabla \tilde{f}_{i,2}, \nabla \tilde{f}_{i,1})^T. \quad (3.44)$$

### Grey value evolution

In the second step, a swarm model is used to let the grey values of the input image  $f : \{1, \dots, MN\} \rightarrow [0, 1]$  propagate in direction of  $\nabla^\perp \tilde{f}$  (as given in (3.44)). We define a one-dimensional evolution in which every particle  $i$  is connected to a pixel  $i$  in the image plane and moves in the tonal domain:

$$\begin{aligned} \dot{u}_i &= \sum_{j \in \mathcal{N}_i} w_{i,j} \cdot k_2(a \cdot |u_j - u_i|) \cdot (u_j - u_i) \\ u_i(0) &= f_i. \end{aligned} \quad (3.45)$$

In accordance with the first step, the neighbourhood  $\mathcal{N}_i$  contains all neighbours of particle  $i$  that lie within euclidean distance  $d_2$  in the image plane. Within (3.45), the weights  $w_{i,j}$  are non-negative constants given by

$$w_{i,j} = \left| \nabla^\perp \tilde{f}_i^T \mathbf{n}_{i,j} \right|^b \cdot \left| \nabla^\perp \tilde{f}_j^T \mathbf{n}_{j,i} \right|^b, \quad (3.46)$$

where  $\mathbf{n}_{i,j}$  denotes the normal vector pointing from pixel  $i$  to pixel  $j$  in the image plane. Consequently, the weights implement the idea, that – during grey value evolution – directions similar to the coherence orientation should be favoured. The parameter  $b > 0$  offers the possibility to control the allowed directional offset, which determines the anisotropy of the evolution.

In order to enforce the evolution across edges, the value of the kernel function

$$k_2(x) := 1 - k_1(x) \quad (3.47)$$

increases with larger grey value differences  $|u_j - u_i|$ . In the context of (3.45), the scalar value  $a \geq 0$  steers the non-linearity of the model (large values of  $a$  approximate a linear model), and by this the influence of local grey value differences.

The steady state of (3.45),

$$\mathbf{u}^* := \lim_{t \rightarrow \infty} \mathbf{u}(t), \quad (3.48)$$

denotes the coherence enhanced image.

## Experiments

Subsequently, we demonstrate the efficacy of our coherence enhancing image filter, and apply the filter to a greyscale fingerprint image (cf. Figure 3.10).

**Colour coding** However, before discussing the results, let us briefly introduce the colour coding shown in Figure 3.11. In the following, we use it to illustrate the orientation of vectors, where we assume w.l.o.g. that the latter can be understood as an angle between 0 and  $\pi$ . This makes sense because we are not interested in the sign of a vector and thus don't distinguish between polar angles  $\theta \in [0, \pi)$  and  $\theta + n \cdot \pi$  ( $n \in \mathbb{Z}$ ).



Figure 3.10: **Input image.** Greyscale image of a fingerprint with  $300 \times 300$  pixels (high resolution version of the image used in [51]).

**Time step size** For all our experiments we choose – in accordance with (3.38) – constant time step size  $\tau$ , namely

$$\tau = \frac{1}{2 \cdot |\mathcal{N}_i| \cdot (1 + \varepsilon)} \quad (3.49)$$

during the estimation of the local dominant gradient direction and

$$\tau = \frac{1}{2 \cdot \max_i \sum_{j \in \mathcal{N}_i} w_{ij}} \quad (3.50)$$

in the second step for grey value evolution. Note that the weights  $w_{ij}$  are constants (cf. (3.46)) and the maximum term in (3.50) can be precomputed easily.

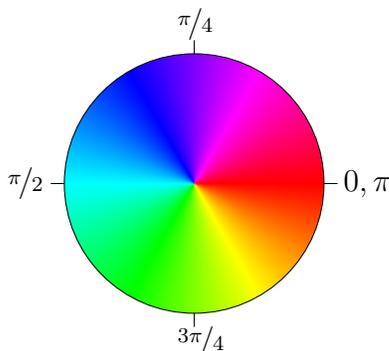


Figure 3.11: **Colour coding.** Colour coding as used for illustrating the local dominant gradient orientation.

**Stopping criterion** Our stopping criterion for the explicit scheme reads in case of the local dominant gradient direction

$$\sum_{i=1}^N |\dot{\mathbf{x}}_i^k|^2 < 10^{-5}. \quad (3.51)$$

For the grey value evolution we end up iterating, if it holds that

$$\sum_{i=1}^N |\dot{u}_i^k|^2 < 10^{-4}. \quad (3.52)$$

**Computational efficiency** Each of both steps of our algorithm has a complexity of  $\mathcal{O}(N \cdot |\mathcal{N}_i| \cdot k)$ , where  $N$  denotes the number of particles – in our case this is equivalent to the number of pixels of  $f$  –, and  $|\mathcal{N}_i|$  represents the size of the neighbourhood of an particle (which is the same for all particles in our setup). The number of needed iterations is given by  $k$ .

### 3.5.2 Parameter selection and influence

During our experiments we make use of a grid size  $h = 1$ . Furthermore, we set  $d_1 = \sqrt{2}$ . Thus, the eight nearest neighbours of every pixel  $i$  in the image plane are considered in the first step. The value of  $\varepsilon$  (cf. (3.40)) is fixed to  $10^{-11}$ .

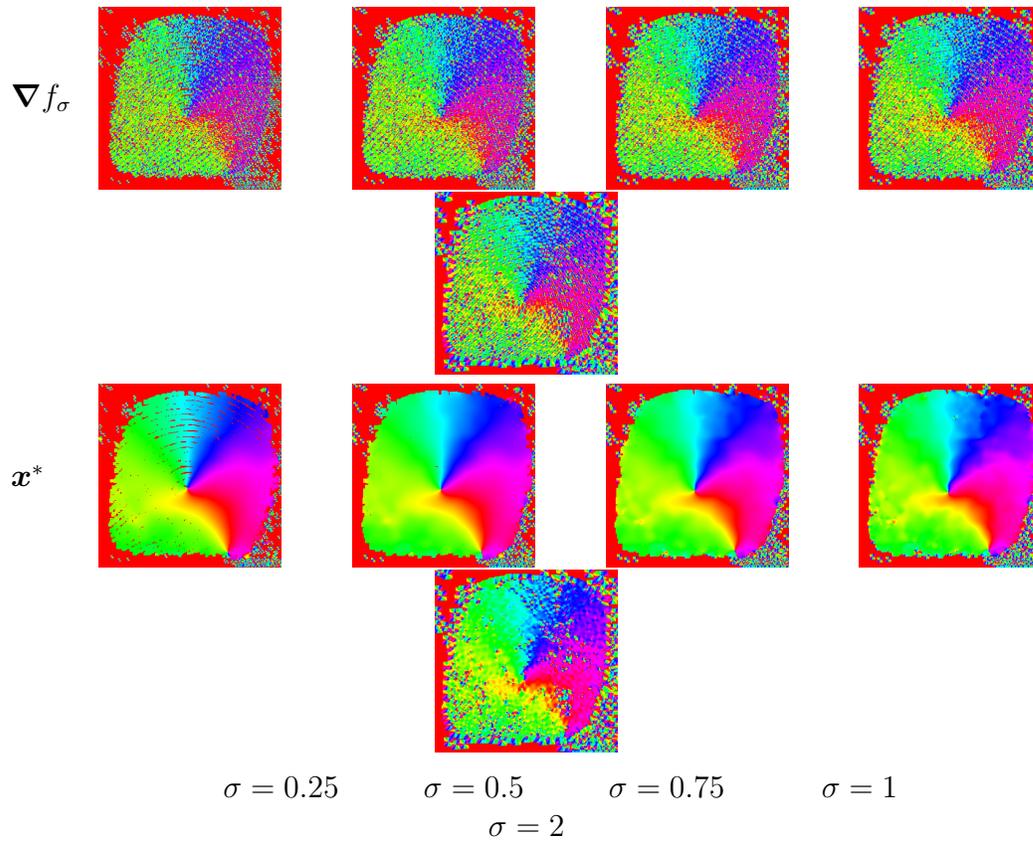


Figure 3.12: **Local Dominant Gradient Orientation.** Influence of parameter  $\sigma$  on the orientation of the pre-smoothed gradient field  $\nabla f_\sigma$  and of  $x^*$  (3.42).

**Local dominant gradient orientation** In dependence on parameter  $\sigma$ , we present different results for the first step in Figure 3.12. The pictures show the orientation of the pre-smoothed gradient vectors, as well as the orientation of the local dominant gradient (steady state of the first evolution (3.39)). The illustrations make use of the previously mentioned colour coding (cf. Figure 3.11). Note, that at all positions with zero gradient magnitude the direction was set to zero (see e.g. the red areas in the surrounding of the fingerprint).

Our results show, that the original gradient field of  $f$  is quite noisy. This justifies the necessity of the first step of our algorithm, since a smooth vector field – describing the local dominant gradient – is essential for the subsequent grey value evolution.

From Figure 3.12 it also becomes clear that an appropriate choice of  $\sigma$  allows to fill in directional information in regions with small or no gradient magnitude (see e.g. the differences between the results for  $\sigma = 0.25$  and  $\sigma = 0.5$ ). This is an important property of our model, since this allows us to smooth not only directly along the edges of an image, but also in between. On the other hand – and in accordance with the findings in [33] – one can also observe, that with increasing value of  $\sigma$  cancellation artifacts for  $\nabla f_\sigma$  appear, which induce an irregular vector field  $\mathbf{x}^*$  (see e.g. the results for  $\sigma = 2$ ). This means that the value of  $\sigma$  should be chosen carefully and be as small as possible.

In the latter case, the steady state of (3.39) represents a smooth vector field (cf. Figure 3.12 for  $\sigma = 0.5$ ). However, note that close to the boundaries of large homogeneous regions irregularities will always appear (e.g. surroundings of the fingerprint). This is related to the fact, that the pre-smoothing might only affect the region boundaries and might not influence its inner parts. By definition of the evolution in (3.39) and (3.40), the particle velocities within those areas are always zero. Consequently, the original gradient orientation – which is undetermined at those positions – can be arbitrary, and will not change (we set the angle to zero as mentioned before). This is highly likely to be in conflict with the orientation of surrounding regions. Fortunately, these artifacts are negligible since the interior of large homogeneous regions plays no role in the subsequent grey value evolution. From (3.45) it is clear that the velocity in those areas is always zero and no transport of grey values takes place.

**Grey value evolution** Based on our previous findings, we choose  $\sigma = 0.5$  to estimate a smooth local dominant gradient field (cf. Figure 3.12). In accordance with [47], we choose a sufficiently large neighbourhood radius  $d_2$  to approximate rotational invariance well. For all further experiments we fix  $d_2 = 5$ . Given this setup, we calculate the results for grey value evolution for varying values of the parameters  $a$  and  $b$  as used in (3.45) and (3.46). The filtered images are presented in Figure 3.13.

Parameter  $a$ , which steers the general amount of attractive forces between the particles, can be seen as the counterpart to the diffusion time in diffusion filters. As one can also see from Figure 3.13, higher values of  $a$  go along with an increased smoothing of the image.

On the other hand, parameter  $b$  allows to control how strict deviations from the local dominant tangent direction should be punished. In the end, this describes how much smoothing should be done in off-tangent direction. Consequently, low values of  $b$  induce a more blurred image, while with increasing value of  $b$  the anisotropy of the filter rises, leading to clearer structures in the image (cf. Figure 3.13).

In conclusion, one can say that – depending on the input image – an adequate weighting of both parameters is important. In our case, we think the usage  $a = 0.2$  and  $b = 6$  offers a good compromise of smoothing and strictness about the smoothing direction. Figure 3.14 illustrates a comparison to coherence enhancing diffusion [33].

**Conclusion** We designed a coherent enhancing image filter whose two steps fully comply with the theory of swarm models. When comparing the input image (cf. Figure 3.10) and the results in Figure 3.13 one can clearly see the efficacy of our suggested filter. Apparent gaps (e.g. at the bottom left or the top of the fingerprint) are closed and overall all lines in the image are smoothed in tangent direction. We still see a lot of potential in the suggested method as one could make the the particles not only move linearly in tangent direction, but also use higher order expressions. We believe, that it is further possible to improve the filter by selecting a different weighting function for the tangent directions in (3.46). This function is of high importance since it does not only determine the amount but also the shape of occurrent anisotropy.

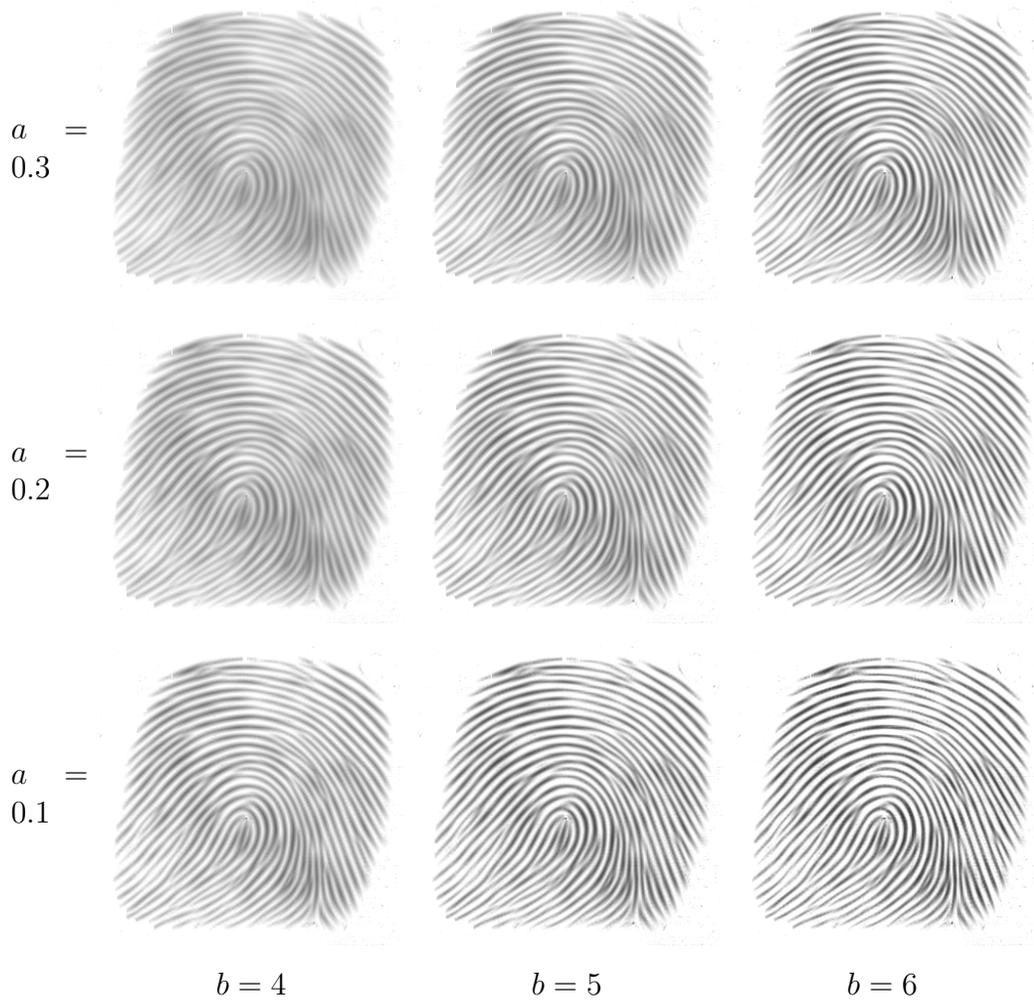


Figure 3.13: **Grey Value Evolution.** Influence of parameters  $a$  and  $b$  on the steady state  $u^*$ .



Figure 3.14: **Comparison to coherence enhancing diffusion.** **Left:** result of our method with parameters  $a = 0.2$  and  $b = 6$ . **Right:** result of coherence enhancing diffusion with parameters  $\sigma = 0.5$ ,  $\rho = 4$  and  $t = 100$ .

### 3.6 Applications: bilateral filtering and NL-means evolutions

In Sections 2.3.4 and 2.3.5 we have presented two processes closely related to the bilateral filter and NL-means respectively, as examples of the general scale-space framework developed in Chapter 2. The nonlocal version of the bilateral filter was given by

$$\partial_t u = \frac{1}{c} \int_{\Omega} g_{\lambda}(|u(\mathbf{y}) - u(\mathbf{x})|) g_{\rho}(|\mathbf{y} - \mathbf{x}|) (u(\mathbf{y}, t) - u(\mathbf{x}, t)) d\mathbf{y},$$

where  $g_{\rho}(s) := \exp(-s^2/(2\rho^2))$ ,  $g_{\lambda}(s) := \exp(-s^2/(2\lambda^2))$  and  $c := \int_{\Omega} g_{\rho}(\mathbf{y}) d\mathbf{y}$  performs a normalisation of the spatial weighting. The nonlocal version of NL-means was given by

$$\partial_t u = \frac{1}{c} \int_{\Omega} K(\mathbf{x}, \mathbf{y}, u) (u(\mathbf{y}) - u(\mathbf{x})) d\mathbf{y},$$

where  $K = \exp\left(-\frac{(G_a * |u(\mathbf{x} + \cdot) - u(\mathbf{y} + \cdot)|^2)(0)}{\lambda}\right) g_{\rho}(|\mathbf{y} - \mathbf{x}|)$  and the first and second terms represent the exponentially weighted distance between patches used in NL-means, and a spatial weight, respectively. Moreover,  $c := \int_{\Omega} g_{\rho}(\mathbf{y}) d\mathbf{y}$  performs a normalisation of the spatial weighting.

These processes were applied to a denoising problem. From the discussion

in Section 2.3.5, we know that in general the results obtained were not good enough, when compared with the corresponding original bilateral and NL-means filters. Moreover, the inherent symmetry of these scale-spaces was one of the reasons for this unsatisfactory behaviour. In this chapter, however, we have developed a theory which allows us to consider a larger set of processes, including non symmetric ones. Therefore, at this point it makes sense to consider also nonsymmetric versions of the above evolutions. Namely

$$\partial_t u = \frac{1}{c(\mathbf{x})} \int_{\Omega} g_{\lambda}(|u(\mathbf{y}) - u(\mathbf{x})|) g_{\rho}(|\mathbf{y} - \mathbf{x}|) (u(\mathbf{y}, t) - u(\mathbf{x}, t)) d\mathbf{y}, \quad (3.53)$$

where this time  $c$  depends on  $\mathbf{x}$  and performs a normalisation of the spatial and tonal weights, and for NL-means,

$$\partial_t u = \frac{1}{c(\mathbf{x})} \int_{\Omega} K(\mathbf{x}, \mathbf{y}, u) (u(\mathbf{y}) - u(\mathbf{x})) d\mathbf{y}, \quad (3.54)$$

with  $K$  as above, and where this time  $c$  depends on  $\mathbf{x}$  and performs a normalisation of the spatial and patch distances weights. Both processes can be seen as cases of the general one given by (3.1)–(3.2). Their corresponding semidiscrete versions are instances of (3.5)–(3.6) and fulfill the type-symmetry condition (NL5). Hence, they satisfy the theory developed in Section 3.3. In particular, their long time behaviour is determined by Theorem 4. Furthermore, by looking at their explicit scheme discretisations, both processes can be seen as damped versions of the corresponding original filters. Let us describe in detail how this comes to be for the case of the bilateral evolution. For NL-means the discussion is essentially the same. If we discretise the time derivative of (3.53) with an explicit scheme, we obtain

$$u^{k+1}(\mathbf{x}) = \tau \frac{1}{c(\mathbf{x})} \int_{\Omega} g_{\lambda}(|u^k(\mathbf{y}) - u^k(\mathbf{x})|) g_{\rho}(|\mathbf{y} - \mathbf{x}|) (u^k(\mathbf{y}, t) - u^k(\mathbf{x}, t)) d\mathbf{y},$$

If we choose a time step  $\tau$  equal to 1, we obtain exactly one iteration of the bilateral filter. Hence, if we instead choose some  $\tau \leq 1$ , it becomes nothing else than a damped bilateral filtering. The same holds for the NL-means nonsymmetric evolution (3.54).

**Accelerated diffusion** The nonsymmetric examples considered here are also related to the work of Didas et. al. [52]. In that work, the authors

interprets the one dimensional bilateral filter as an approximation of:

$$\partial_t u = \frac{1}{g(|\partial_x u|^2)} \partial_x (g(|\partial_x u|^2) \partial_x u).$$

The additional term is shown to act in such a way that it accelerates the diffusion along edges. Moreover the author also showed how the resulting process can be interpreted as a nonsymmetric version of nonlinear diffusion. Two dimensional versions of the bilateral filters were also considered in that work. The authors showed that in that case, the bilateral filter can be interpreted as a crude approximation of the directional anisotropic diffusion of Weickert [47] in which the sector diffusion is inspired.

### 3.7 Conclusions

In this chapter we considered the well-posed initial value problem (3.1)–(3.2), obtained from the general framework (1.3)–(1.4) by supplementing it with assumptions (NL1)–(NL2) but leaving out the symmetry and irreducibility conditions considered in the previous chapter. After noticing that the well-posedness theory, still holds in this case, we focused on the nonlocal semidiscrete version (3.5)–(3.6). First of all, we discussed a different interpretation of the semidiscrete process, in terms of particle swarm systems. This new interpretation paved the road for defining (NL5) or (NL6), as alternatives for the pair of conditions (NL3)–(NL4). We showed that these less restrictive conditions are sufficient to study the long time behaviour of the solutions of (3.5)–(3.6) by means of Theorem 4 and Theorem 5 when condition (NL5) or (NL6) is assumed, respectively. Furthermore, we discussed two different specific applications. Firstly, Sectors Diffusion was introduced as an alternative diffusion filter with good corner preservation properties. Finally a coherence enhancing application was discussed based on two steps, each of which can be interpreted within the studied framework.



## Chapter 4

# Nonlocal evolutions: beyond nonnegativity

In this chapter we will explore the consequences of relaxing the sign condition given by assumption (NL2) of (2.1)–(2.2). We will thus consider the general process allowing the weight function  $K$  to attain negative values.

We already saw in the previous chapters that the symmetry and irreducibility assumptions (NL3)–(NL4) are not necessary in order to ensure well-posedness of (1.3)–(1.4). The non-negativity assumption (NL2), on the other hand, does play an essential role in this respect. In fact, this assumption was essential in order to ensure the stable behaviour of the process, given by the maximum-minimum property of Lemma 5 which constituted a fundamental ingredient in the proof of Theorem 1. Therefore, the first consequence of removing the non-negativity assumption is that of non being able to ensure the global existence of a solution for (1.3)–(1.4).

This absence of well-posedness results is to be expected, however. In fact, as was mentioned in the introductory chapter, it is possible to establish a formal connection between the general framework (1.3)–(1.4) and processes described by PDEs. In that sense, in the context of nonlocal processes, the weight function  $K$  can be formally seen as a diffusivity. This means that the processes considered in the previous chapters, which had only positive weight functions, are related to initial value problems given by forward parabolic equations. It is therefore expected that they would give rise to stable solutions as was indeed verified. On the other hand, the absence of

the non-negativity of  $K$ , leads to processes related to backwards parabolic equations, which are known in to lack in general of a well-posedness theory. It is natural to expect an analogous property for their nonlocal counterpart. Backwards diffusion has been widely studied in previous literature. Its potential applications on problems like the deblurring or sharpening of degraded imagery justify the efforts to overcome its inherent instability. The most widely used strategy in that direction, is to impose constraints at extrema that aim at enforcing a maximum–minimum principle. One example is the inverse diffusion filter of Osher and Rudin [53], which implements backward diffusion everywhere except at extrema, where the evolution is set to zero. Another example is the so-called forward-and-backward (FAB) diffusion of Gilboa et al. [54]. It differs from the closely related Perona–Malik filter [3] by the fact that it uses negative diffusivities for a specific range of gradient magnitudes. However, at extrema where the gradient vanishes, it always avoids explosions by imposing forward diffusion. Another, less widely-used class of stabilisation attempts adds a fidelity term that prevents the backward evolution to move too far away from the original image [55] or from the average grey value of the desired range [56]. In this case the range of the filtered image obviously depends on the weights of the fidelity and the backward diffusion term. Other examples of backwards processes are given in [57], where the authors show that the solution of a complex diffusion equation combines properties of both forward and backward diffusion, and in [58] where the authors consider the flow of a nonconvex triple well potential. The main concern of this chapter will be the stabilisation of our nonlocal process without the non-negativity constraint.

It consists of two parts: In the first half, we will consider a rather general procedure which will be sufficient to stabilise general semidiscrete processes consisting of weight functions with both positive and negative values. In the second half, we will consider a pure backwards process. Namely, one consisting of a non-positive weight function.

## 4.1 Forwards and backwards nonlocal processes

We will now continue the study of the initial value problem (1.3)–(1.4) in the general situation in which no further condition other than the regularity (NL1) is assumed and we allow the weighting function to have negative and

positive values. In what follows we describe a stabilisation procedure which overcomes the lack of a maximum-minimum principle which had previously been obtained with the help of (NL2). Let us start by recalling the general initial valued problem. Let  $\Omega$  be an open bounded subset of  $\mathbb{R}^N$ , let  $T > 0$  be a constant, and let  $\mu$  be a locally finite Borel measure. Let  $f \in C(\bar{\Omega})$ , and consider the following problem:

$$u_t(\mathbf{x}, t) = \int_{\Omega} K(\mathbf{x}, \mathbf{y}, u(\cdot, t))(u(\mathbf{y}, t) - u(\mathbf{x}, t))d\mu(\mathbf{y}) \quad \text{in } \bar{\Omega} \times [0, T], \quad (4.1)$$

$$u(\mathbf{x}, 0) = f(\mathbf{x}) \quad \text{in } \bar{\Omega}. \quad (4.2)$$

The only further assumption here is given by the regularity of  $K$ , namely,

**(NL1) Regularity:**  $K : \bar{\Omega} \times \bar{\Omega} \times C(\bar{\Omega}) \rightarrow \mathbb{R}$ , is Lipschitz continuous and bounded.

The following stabilisation procedure was initially described in [59], where it was used as a stabilization strategy in the case of the specific backwards nonlocal process which will be presented in detail in Section 4.2. Here, we will see that the same construction is sufficient to stabilize quite general nonlocal processes with sign changing values. The theory presented in this chapter is most meaningful for nonlocal semidiscrete processes. Recall that these processes are obtained by choosing the specific measure  $\mu$  concentrated on a grid of points which represent the positions of the image pixels, as was done for (3.5)–(3.6). In detail, we will focus on the following specific type of semidiscrete system.

$$\dot{u}_i = \sum_{j=1}^{MN} \gamma(d(i, j))g(|u_i - u_j|)(u_j - u_i) \quad (1 \leq i \leq MN), \quad (4.3)$$

$$u_i(0) = f_i \quad (1 \leq i \leq MN). \quad (4.4)$$

where,  $d(i, j)$  represents the Euclidean distance between the two dimensional coordinate positions in the image grid of the pixels with indexes  $i$  and  $j$ . Moreover, index  $i$  runs over all the grid positions of the image pixels.

Here we chose the weight function  $K = \gamma(|x - y|)g(|u(x) - u(y)|)$ , with  $\gamma$  representing a spatial weight and  $g$  representing a tonal weight in a similar

way as the bilateral evolutions introduced in (2.40). This weight function is basic enough to make the presentation simple, and at the same time, general enough to cover interesting applications.

### 4.1.1 Stabilization

The idea of the stabilization strategy is mainly based on the reflection of the values of each pixel with respect to the boundaries of the interval containing the range of the co-domain, which in the case of grey valued images is given by the interval  $[0, 255]$ . In order to describe the stabilization construction in detail, let us recall the connection between our nonlocal semidiscrete framework and particle swarm systems described in Section 3.3. This will give us a nice way to introduce and visualize the details of the strategy.

We can apply the same procedure depicted in Figure 3.1, in order to obtain an equivalent particle swarm system for (4.3)–(4.4). The main difference to the swarms considered previously, is given by the fact that the interaction between particles can now be both attractive or repulsive. Moreover, the lack of a maximum-minimum principle translates, in terms of the corresponding particle system, in the fact that we cannot guarantee the permanence of the particles inside the interval  $[0, 255]$ , during the evolution of the system.

From now on we will assume that the grey values of the input images are between 0 and 1, in such a way that the corresponding particle system has an initial configuration of positions within the interval  $[0, 1]$ . This condition will not restrict in any way the analysis and is only introduced as a way to simplify notation.

The proposed strategy is given by the following procedure: Assume we have a set of  $MN$  particles with some given initial positions inside the interval  $[0, 1]$ . We consider  $MN$  new particles obtained by duplicating each of the initial ones and reflecting their positions with respect to the right boundary of the interval  $[0, 1]$ . We then duplicate the initial particles one again, only this time we reflect the new particle with respect to the left boundary of the interval. Thus, if the original swarm system is given by particles  $(x_i, \mathbf{a}_i)$ ,  $0 \leq i \leq MN$ , with each  $x_i$  denoting the position of the particle  $i$  in  $[0, 1]$ , and each  $\mathbf{a}_i$  denoting the fixed two dimensional coordinates of the corresponding pixels positions in the image grid. We obtain a new system of  $3MN$  particles which in addition to the initial ones, contains also  $(-x_i, \mathbf{a}_i)$  and  $(2 - x_i, \mathbf{a}_i)$ ,

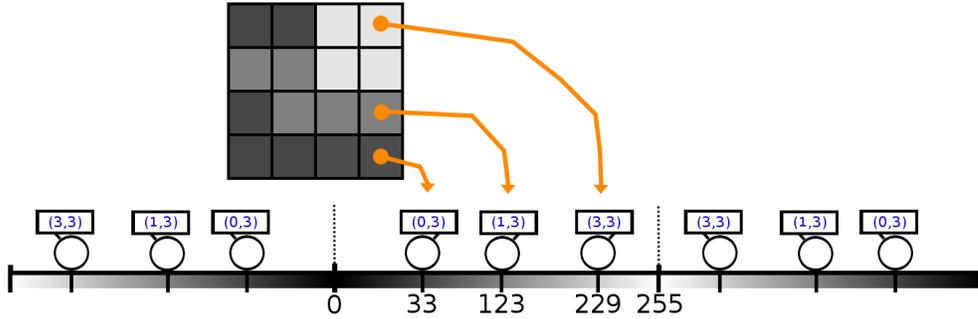


Figure 4.1: Mapping of pixels with reflections

for  $0 \leq i \leq M$ , as showed in Figure 4.1 The evolution of the new system is defined as follows: The particles in the central interval  $[0, 1]$ , move by means of their interaction with all the other particles of the  $3M$  system, according to the weight function given by the original process (4.3)–(4.4). Furthermore, the particles obtained from the reflection procedure, evolve in such a way that the symmetries w.r.t the points 0 and 1, are kept during the whole evolution. More precisely, if the positions of the particles starting inside the central interval are denoted by  $x_i$  and their left and right reflections by  $x_{-i}, x_{2MN+1-i}$ , for  $1 \leq i \leq MN$ , respectively. The evolution of the whole  $3MN$  particles system is given by

$$\dot{x}_i = \sum_{-MN \leq j \leq 2MN} \gamma(d(i, j))g(|x_i - x_j|)(x_j - x_i) \quad (1 \leq i \leq MN), \quad (4.5)$$

with initial conditions

$$\begin{aligned} x_i(0) &= u_i(0), \\ x_{-i}(0) &= -u_i(0), \\ x_{2MN+1-i}(0) &= 2 - u_i(0), \end{aligned}$$

and symmetry conditions,

$$\begin{aligned} \dot{x}_{-i} &= -\dot{x}_i, \\ \dot{x}_{2MN+1-i} &= -\dot{x}_i. \end{aligned}$$

We will only be interested in the case of continuous functions  $\gamma$  which are null for values greater than 1.

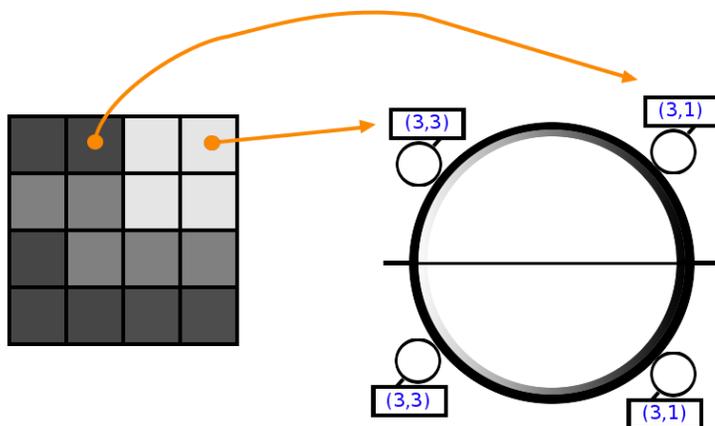


Figure 4.2: Mapping of pixels to particle system in the interval  $[0,255]$

**Remark 2.** *The last condition about the function  $\gamma$ , permits us to give an interesting interpretation of the system, which also allows us to describe it without having to define separately the dynamics of the central particles and that of the reflected ones. In fact, in this case the particle system is equivalent to a system of  $2MN$  particles moving in a circle as shown in Figure 4.2. In this representation, each particle interacts with all the other ones of the circular system and the positions  $x_i$  are converted into angles. The complete dynamics of the  $2MN$  particles are then given as follows.*

$$\dot{x}_i = \sum_{j=1}^{2MN} \gamma(d(i,j))g(|x_i - x_j|)(x_j - x_i) \quad (1 \leq i \leq 2MN), \quad (4.6)$$

$$x_i(0) = f_i \quad (1 \leq i \leq 2MN), \quad (4.7)$$

*The configuration of the particles is symmetric w.r.t to the horizontal axis during the whole evolution, and the dynamics of the ones contained in the upper part of the circle, correspond to those in the central interval  $[0,1]$  of system (4.5).*

We have thus transformed our semidiscrete nonlocal problem into a particle system, in the same spirit of Chapter 3. Here, however, the difference lays in the fact that we directly modified the obtained swarm system, by applying the stabilization procedure. Therefore, the resulting system (4.5) is not completely equivalent to the original nonlocal one (4.3)–(4.4). We can, how-

ever, go the other way around and obtain a corresponding new semidiscrete nonlocal process which is equivalent to (4.5) as follows.

**Lemma 9.** *The particle system (4.5), obtained from (4.3)–(4.4) by applying the stabilization procedure, is equivalent to the nonlocal given by the following system of equations.*

$$\dot{u}_i = \sum_{j=1}^{MN} \gamma(d(i,j))\Phi(u_i - u_j) - \sum_{j=1}^{MN} \gamma(d(i,j))\Phi(u_i + u_j) \quad (1 \leq i \leq MN), \quad (4.8)$$

$$u_i(0) = f_i \quad (1 \leq i \leq MN). \quad (4.9)$$

*Proof.* This is a direct consequence of applying the initial and symmetry conditions of (4.5), thus avoiding the summation over the redundant terms given by the reflected particles. Going from the particle swarm to the nonlocal system, can be obtained by simply renaming the variables from  $x_i$  to  $v_i$ . This inverts the transformation described above and takes into account that the summation is now done over an index which represents the different pixel positions over the image grid.  $\square$

Finally, we can also guarantee that the new system is indeed stable w.r.t. the range constrains.

**Lemma 10** (Stability w.r.t. range constraints). *The solution of system (4.8)–(4.9) is stable w.r.t. the range constrains. Namely, for every  $1 \leq i \leq NM$  we have that*

$$0 \leq u_i(t) \leq 1.$$

*Proof.* Without loss of generality assume that  $0 < \min_{0 \leq k \leq MN} f_k, \max_{0 \leq k \leq MN} f_k < 1$ . From the regularity condition (NL1) of  $K$ , it follows that the evolution of a single particle (grey values)  $x_i$  is continuous in time. This implies that in order to leave the interval  $[0,1]$  there should exist a time  $t_0$  in which the particle stands exactly at the boundary, say at point 0. However, because of the symmetry of the system it follows that for  $t > t_0$  the velocity  $\dot{x}_i(t) = 0$ . This shows that  $x_i$  cannot leave the interval  $[0, 1]$ .  $\square$

We are now able to consider a process which is stable and has no restrictions about the sign of the weight function. At this point, an issue which naturally arises is that of the long time behaviour of the stabilized system. In recent years, different authors have considered the problem of long time behaviour of interacting particles in a circle, of which we know the stabilized system (4.5) to be a particular case, thanks to Remark 2. In the systems control literature, it is possible to find some results about the consensus problems for particles living on a circle or other manifolds [60–62]. However, to the best of our knowledge, the analysis of the long time behaviour of (4.5), or equivalently, of the particle system depicted on Figure 4.2, in the case of a general weight function  $k$  without restrictions on its sign, is an open question. Nevertheless, for some specific shapes of the weight function  $K$  which are important for applications, this issue can be resolved. We will see how in Section 4.2.

## 4.1.2 Explicit schemes and experiments

We would now like to translate system (4.8) into a corresponding stable algorithm, in order to be able to compute experimental results with actual images. The purpose of the following proposition is to show that an explicit scheme is sufficient to obtain an algorithm which is stable. We will see that under adequate constraints about the step size, particle position will always be inside the original interval  $[0, 255]$ . Using forward differences to approximate the time derivative in and using  $\gamma_{ij} := \gamma(d(i, j))$  the explicit scheme of system (4.8) reads

$$u_i^{k+1} = u_i^k + \tau \cdot \sum_{\substack{j=1 \\ j \neq i}}^N \cdot \gamma_{ij} \cdot \Phi(u_j^k - u_i^k) - \tau \cdot \sum_{j=1}^N \cdot \gamma_{ij} \cdot \Phi(u_i^k + u_j^k), \quad i = 1, \dots, N, \quad (4.10)$$

where  $\tau$  denotes the time step size and an upper index  $k$  refers to the time  $k\tau$ .

**Theorem 6** (Stability Guarantees for the Explicit Scheme). *Let  $L_\Phi$  be the Lipschitz constant of  $\Phi$  restricted to the interval  $(0, 2)$ . Moreover, assume that the time step size used in the explicit scheme (4.10) satisfies*

$$\tau < \frac{1}{2 \cdot L_\Phi \cdot \max_{1 \leq j \leq N} \sum_{i=1}^N w_i \gamma_{ij}}.$$

Then the following stability properties hold:

(i) If  $0 < u_i^k < 1$ , then  $0 < u_i^{k+1} < 1$ , for every  $1 \leq i \leq N$ .

(ii) If  $\gamma = 1$  and  $0 < u_i^k < u_j^k < 1$ , then  $u_i^{k+1} < u_j^{k+1}$ .

*Proof.* (i). Let  $0 < u_i^k, u_j^k < 1$ . We have the following cases:  
If  $u_i^k = u_j^k < \frac{1}{2}$ , then  $2v_i^k \in (0, 1]$ . Thus

$$0 \leq -\Phi(2v_i^k) \quad (4.11)$$

If  $\frac{1}{2} < u_i^k = v_j^k$ , then  $2v_i^k \in (1, 2)$ . Thus

$$|\Phi(2v_i^k)| = |\Phi(2v_i^k) - \Phi(1)| \leq |2v_i^k - 1|L_\Phi \leq 2 \cdot v_i^k \cdot L_\Phi \quad (4.12)$$

If  $u_i^k < u_j^k$  then  $u_j^k - u_i^k, u_i^k + u_j^k \in (0, 2)$ . Thus,

$$|\Phi(u_i^k + u_j^k) - \Phi(u_j^k - u_i^k)| < L_\Phi \cdot 2u_i \quad (4.13)$$

If  $u_j^k < u_i^k \leq \frac{1}{2}$ , then  $u_j^k - u_i^k \in (-1, 0)$  and  $u_j^k + u_i^k \in (0, 1)$ . Thus,

$$0 \leq \Phi(u_j^k - u_i^k) - \Phi(u_i^k + u_j^k) \quad \text{and} \quad 0 \leq -\Phi(2u_i) \quad (4.14)$$

Finally, if  $u_j^k < u_i^k$  and  $\frac{1}{2} < u_i^k$ , using  $\Phi(1) = 0$  the periodicity of  $\Phi$  we get

$$\begin{aligned} |\Phi(u_j^k - u_i^k) - \Phi(u_i^k + u_j^k)| &= |\Phi(u_i^k + u_j^k) - \Phi(2 + u_j^k - u_i^k)| < 2 \cdot u_i \cdot L_\Phi \quad \text{and} \\ |\Phi(2u_i)| &= |\Phi(2u_i) - \Phi(1)| \leq |2u_i - 1|L_\Phi \leq 2 \cdot u_i \cdot L_\Phi \quad (4.15) \end{aligned}$$

Let  $J_1 := \{1 \leq j \leq N : u_i \neq u_j\}$  and  $J_2 := \{1 \leq j \leq N : u_i = u_j\}$ . Combining (4.10) with the expressions from (4.11) to (4.15), we obtain that

$$\begin{aligned} u_i^{k+1} - u_i^k &= -\tau \cdot \sum_{J_1} \gamma_{ij} \cdot w_j \cdot (\Phi(u_i^k + u_j^k) - \Phi(u_j^k - u_i^k)) - \tau \cdot \sum_{J_2} \gamma_{ij} \cdot w_i \cdot \Phi(2u_i^k) \\ &\geq \tau \cdot L_\Phi \cdot 2u_i^k \cdot \sum_{\ell=1}^N w_\ell \geq -u_i^k \quad (4.16) \end{aligned}$$

This, shows that  $u_i^{k+1} > 0$ , as claimed.

The proof of  $u_i^{k+1} < 1$  proceeds in a similar way.

(ii). Considering the explicit discretisation of (4.53) for  $\partial_t u_i$  and  $\partial_t u_j$  and  $\gamma = 1$ , we obtain for  $i, j = 1, \dots, N$

$$\begin{aligned} u_j^{k+1} - u_i^{k+1} &= u_j^k - u_i^k + \tau \cdot (w_i + w_j) \cdot \Phi(u_i^k - u_j^k) + \\ &\quad \tau \cdot \sum_{\substack{\ell=1 \\ \ell \neq i, j}}^N w_\ell \cdot (\Phi(u_\ell^k - u_i^k) - \Phi(u_\ell^k - u_j^k)) - \\ &\quad \tau \cdot \sum_{\ell=1}^N w_\ell \cdot (\Phi(u_\ell^k + u_j^k) - \Phi(u_\ell^k + u_i^k)) . \end{aligned} \quad (4.17)$$

Using the fact that  $\Phi$  is Lipschitz in the interval  $(0, 2)$ , we also know that

$$\begin{aligned} &\tau \cdot \sum_{\substack{\ell=1 \\ \ell \neq i, j}}^N w_\ell \cdot |\Phi(u_\ell^k - u_i^k) - \Phi(u_\ell^k - u_j^k)| + \tau \cdot \sum_{\ell=1}^N w_\ell \cdot |\Phi(u_\ell^k + u_j^k) - \Phi(u_\ell^k + u_i^k)| \\ &< \tau \cdot L_\Phi \cdot 2 |u_j^k - u_i^k| \cdot \sum_{\ell=1}^N w_\ell < |u_j^k - u_i^k| . \end{aligned} \quad (4.18)$$

Finally, since  $0 < \Phi(u_i^k - u_j^k)$ , (4.17) and (4.18) imply that  $0 < u_j^{k+1} - u_i^{k+1}$ , as claimed.  $\square$

In the following section, we will consider a special form of the weight function  $K$  for (4.8). The resulting system, other than having a maximum-minimum principle, and a corresponding stable algorithm given by the explicit scheme studied above, will also have a form which will simplify the analysis of its long time behaviour.

## 4.2 Backward nonlocal processes

We will now consider system (4.3)–(4.4) with a more specific choice of the weight function, which will result in a nonlocal process with negative weights. Hence, on backwards diffusion like evolutions. The results presented in this section are based on [59].

Let us recall the basic type of process (4.3) considered in this chapter. It is given by

$$\dot{u}_i = \sum_{j=1}^{MN} \gamma(d(i, j)) \Phi(|u_i - u_j|) \quad (1 \leq i \leq MN), \quad (4.19)$$

$$u_i(0) = f_i \quad (1 \leq i \leq MN). \quad (4.20)$$

where  $d(i, j)$  represents the Euclidean distance between the two dimensional coordinates of the pixel with indexes  $i$  and  $j$ . Here we renamed  $\Phi(s)$  by  $g(s)s$ . In what follows, we only consider functions  $g$  such that the function  $\Phi$  satisfies the following set of properties: It is strictly increasing and continuous in the interval  $]0, 1[$ , non-negative in  $]0, 1[$ , antisymmetric and periodic with period 2. We also assume that  $\Phi(0) = 0$ . In a typical scenario, function  $\Phi$  would look like in Figure 4.3.

Once we apply the stabilization procedure, we end up with a corresponding new process which according to Lemma 9, is equivalent to the following system.

$$\dot{u}_i = \sum_{\substack{j=1 \\ j \neq i}}^{MN} \Phi(|u_i - u_j|) - \sum_{j=1}^{MN} \Phi(|u_i + u_j|) \quad (1 \leq i \leq MN), \quad (4.21)$$

$$u_i(0) = f_i \quad (1 \leq i \leq MN). \quad (4.22)$$

Before we start analysing the properties of the newly proposed process, let us for a moment turn our attention to forward diffusion processes, in order to gain an insight about the weight choices above. This will lead us to the introduction of an equivalent backwards process by means of a minimization problem. We will study some properties of the evolutions obtained from the gradient descent of the minimization problem, including its long time behaviour, and finally discuss its connection to (4.21).

For simplicity we consider a simple 1-D evolution for signal smoothing. It regards the original signal  $f : [a, b] \rightarrow \mathbb{R}$  as initial state of the diffusion equation

$$\partial_t u = \partial_x (g(u_x^2) u_x) \quad (4.23)$$

where  $u = u(x, t)$  is a filtered version of the original signal  $u(x, 0) = f(x)$ ,  $u_x = \partial_x u$ , and reflecting boundary conditions at  $x = a$  and  $x = b$  are imposed. Larger diffusion times  $t$  create simpler representations. The diffusivity

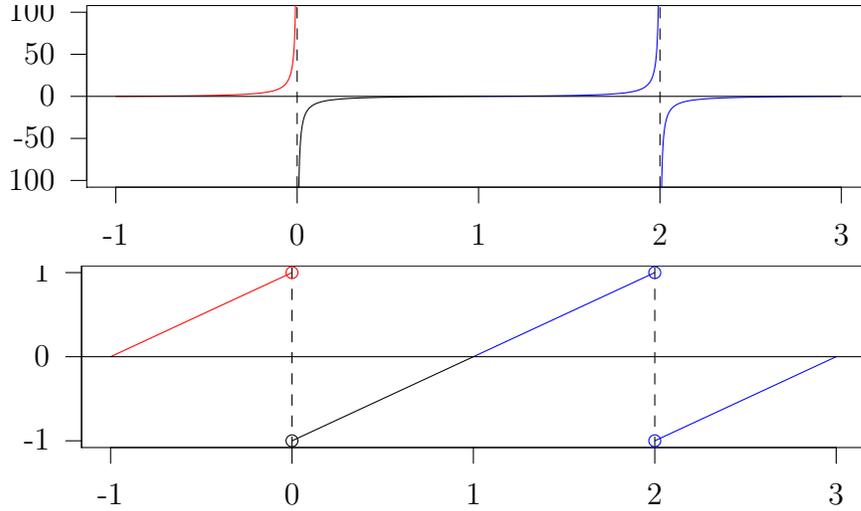


Figure 4.3: **Middle:** Corresponding diffusivity  $\tilde{g}(s) := g(s^2) = \Psi'(s^2)$  where  $\tilde{g}(s) := 1 - 1/s$  for  $s \in (0, 1]$ . **Bottom:** Corresponding flux  $\Phi(s) = \Psi'(s^2)s$  with  $\Phi(s) = s - 1$  for  $s \in (0, 1]$ .

function  $g$  is nonnegative. In order to smooth less at signal edges than in more homogeneous regions, Perona and Malik [3] propose to choose  $g$  as a decreasing function of the contrast  $u_x^2$ . If the flux function  $\Phi(u_x) := g(u_x^2) u_x$  is strictly increasing in  $u_x$  we have a forward diffusion process that cannot sharpen edges. Then the diffusion process can be seen as a gradient descent evolution for minimising the energy

$$E[u] = \int_a^b \Psi(u_x^2) dx \quad (4.24)$$

with a potential function  $\tilde{\Psi}(u_x) = \Psi(u_x^2)$  that is strictly convex in  $u_x$ , increasing in  $u_x^2$ , and satisfies  $\Psi'(u_x^2) = g(u_x^2)$ . Since the energy functional is strictly convex, it has a unique minimiser. This minimiser is given by the (flat) steady state ( $t \rightarrow \infty$ ) of the gradient descent method, and the gradient descent / diffusion evolution is well-posed. Due to this classical appearance of well-posed forward diffusion as a consequence of strictly convex energies, one might be tempted to believe that backward diffusion is necessarily connected to nonconvex energies. Interestingly, this is not correct! Understanding this connection better opens new ways to design stable backward processes. We consider a space-discrete model where we admit *globally negative* diffusivities, corresponding to decreasing penalisers  $\Psi$ . However, we require  $\Psi(u_x^2)$

to be strictly convex in  $u_x$ . We do not rely on stabilisations through zero or forward diffusivities at extrema, and we do not incorporate fidelity terms explicitly. Stabilisation will be achieved in our model on a *global* level by bounding the range of  $u$ . This is achieved by imposing reflecting boundary conditions in the co-domain. We show that this is sufficient to stabilise the inverse diffusion in the space-discrete and time-continuous setting.

We may introduce a dynamical system that is motivated from a spatial discretisation of the energy functional (4.24) with a decreasing penaliser function  $\Psi : \mathbb{R}_0^+ \rightarrow \mathbb{R}$  and a global range constraint on  $u$ . The corresponding flux function  $\Phi$  is given by  $\Phi(s) := \Psi'(s^2)s$ .

We consider vectors  $\mathbf{v} = (v_1, \dots, v_N) \in (0, 1)^N$ , where  $v_i$  for  $i = 1, \dots, N$  are assumed to be distinct. We extend such  $\mathbf{v}$  with the additional coordinates  $v_{N+1}, \dots, v_{2N}$  defined as  $v_{2N+1-i} = 2 - v_i \in (1, 2)$ . For this extended  $\mathbf{v} \in [0, 2]^{2N}$ , we consider the energy function

$$E(\mathbf{v}) = \frac{1}{2} \cdot \sum_{i=1}^{2N} \sum_{\substack{j=1 \\ j \neq i}}^{2N} \Psi((v_j - v_i)^2) . \quad (4.25)$$

A typical scenario for  $\Psi$  is illustrated in Figure 4.3. First, the function  $\Psi(s^2)$  is defined as a continuously differentiable, decreasing, and strictly convex function for  $s \in [0, 1]$  with  $\Psi(0) = 0$  and  $\Phi_-(1) = 0$  (left-sided derivative). It is then extended to  $[-1, 1]$  by symmetry and to  $\mathbb{R}$  by periodicity  $\Psi((2+s)^2) = \Psi(s^2)$ . As a result,  $\Psi(s^2)$  is continuously differentiable everywhere except at even integers, where it is still continuous. Note that  $\Psi(s^2)$  is increasing on  $[-1, 0]$  and  $[1, 2]$ . The flux  $\Phi$  is continuous and increasing in  $(0, 2)$  with jump discontinuities at 0 and 2 (see Figure 4.3). Furthermore, we have that  $\Phi(s) = -\Phi(-s)$  and  $\Phi(2+s) = \Phi(s)$ . A gradient descent for (4.25) is given by

$$\dot{v}_i = -\partial_{v_i} E(\mathbf{v}) = \sum_{\substack{j=1 \\ j \neq i}}^{2N} \Phi(v_j - v_i) , \quad i = 1, \dots, 2N , \quad (4.26)$$

where  $v_i$  now are functions of the time  $t$ . Note that for  $1 \leq i, j \leq N$ , thus  $|v_j - v_i| < 1$ , the flux  $\Phi(v_j - v_i)$  is negative for  $v_j > v_i$  and positive otherwise, thus driving  $v_i$  always away from  $v_j$ . This implies that we have negative diffusivities  $\Psi'$  for all  $|v_j - v_i| < 1$ . Due to the convexity of  $\Psi(s^2)$ , the absolute values of the repulsive forces  $\Phi$  are decreasing with the distance between  $v_i$  and  $v_j$ . We remark that the jumps of  $\Phi$  at 0 and 2 are not problematic here

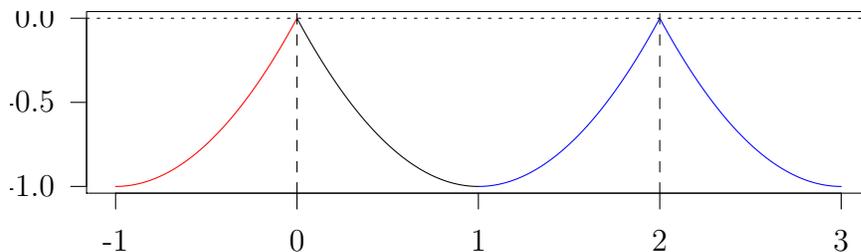


Figure 4.4: **Top:** Exemplary penaliser  $\tilde{\Psi}(s) := \Psi(s^2)$  with  $\tilde{\Psi}(s) = (s-1)^2 - 1$  for  $s \in [0, 1]$ , extended to the interval  $[-1, 3]$  by imposing **symmetry** and  $\Psi((2+s)^2) = \Psi(s^2)$ .

as the  $v_i$  are required to be distinct.

### 4.2.1 Basic properties

Let us discuss shortly how the interval constraint for the  $v_i$ ,  $i = 1, \dots, N$ , is enforced in (4.25) and (4.26). First, notice that  $v_{2N+1-i}$  for  $i = 1, \dots, N$  is the reflection of  $v_i$  on the right interval boundary 1. For  $v_i$  and  $v_{2N+1-j}$  with  $1 \leq i, j \leq N$  and  $v_{2N+1-j} - v_i < 1$  there is a repulsive force due to  $\Phi(v_{2N+1-j} - v_i) < 0$  that drives  $v_i$  and  $v_{2N+1-j}$  away from the *right* interval boundary. The closer  $v_i$  and  $v_{2N+1-j}$  come to this boundary, the stronger is the repulsion. For  $v_{2N+1-j} - v_i > 1$ , we have  $\Phi(v_{2N+1-j} - v_i) > 0$ . By  $\Phi(v_{2N+1-j} - v_i) = \Phi((2-v_j) - v_i) = \Phi((-v_j) - v_i)$ , this can equally be interpreted as a repulsion between  $v_i$  and  $-v_j$  where  $-v_j$  is the reflection of  $v_j$  at the left interval boundary 0. In this case the interaction between  $v_i$  and  $v_{2N+1-j}$  drives  $v_i$  and  $-v_j$  away from the *left* interval boundary. Recapitulating both possible cases, it becomes clear that every  $v_i$  is either repelled from the reflection of  $v_j$  at the left or at the right interval boundary, but never from both at the same time.

As  $\dot{v}_{2N+1-i} = -\dot{v}_i$  holds in (4.26), the symmetry of  $\mathbf{v}$  is preserved. Dropping the redundant entries  $v_{N+1}, \dots, v_{2N}$ , equation (4.26) can be rewritten as

$$\dot{v}_i = \sum_{\substack{j=1 \\ j \neq i}}^N \Phi(v_j - v_i) - \sum_{j=1}^N \Phi(v_i + v_j) , \quad i = 1, \dots, N , \quad (4.27)$$

where the second sum represents the repulsions between original and reflected coordinates in a more symmetric way.

Given an initial vector  $\mathbf{f} \in (0, 1)^N$  with distinct entries  $f_i$ , and initialising  $v_i(0) = f_i$ ,  $v_{2N+1-i}(0) = 2 - f_i$  for  $i = 1, \dots, N$ , the gradient descent (4.26) or (4.27) evolves  $\mathbf{v}$  towards a minimiser of  $E$ .

A more detailed analysis below shows that in the course of the evolution, no  $v_i$  can reach the interval boundaries 0 or 1, and no  $v_i, v_j$  with  $i \neq j$  can ever become equal. Thus the initial rank-order of  $v_i$  is preserved throughout the evolution. Each of the  $N!$  possible rank-orders constitutes a connected component of the configuration space for  $\mathbf{v}$ . There is a unique minimiser of  $E$  in the interior of each connected component due to the strict convexity of  $\Psi(s^2)$ .

**Theorem 7** (Avoidance of Boundaries). *The  $N$  initially distinct positions  $v_i \in (0, 1)$  evolving according to (4.27) never reach the domain boundaries 0 and 1.*

*Proof.* The definition of  $\Phi$  implies that

$$\lim_{h \rightarrow 0^+} \frac{\Psi((0+h)^2) - \Psi(0)}{h} < 0 \quad \text{and} \quad \lim_{h \rightarrow 0^-} \frac{\Psi(2) - \Psi((2+h)^2)}{h} > 0, \quad (4.28)$$

from which it follows that

$$\lim_{v_i \rightarrow 0^+} (-\Phi(2v_i)) > 0 \quad \text{and} \quad \lim_{v_i \rightarrow 1^-} (-\Phi(2v_i)) < 0. \quad (4.29)$$

Equation (4.27) can be written as

$$\dot{v}_i = \sum_{\substack{j=1 \\ j \neq i}}^N (\Phi(v_j - v_i) - \Phi(v_i + v_j)) - \Phi(2v_i). \quad (4.30)$$

Since for  $j = 1, \dots, N$  and  $j \neq i$  one has

$$\lim_{v_i \rightarrow 0^+} \Phi(v_j - v_i) - \Phi(v_i + v_j) = 0 \quad \text{and} \quad \lim_{v_i \rightarrow 1^-} \Phi(v_j - v_i) - \Phi(v_i + v_j) = 0, \quad (4.31)$$

it follows that

$$\lim_{v_i \rightarrow 0^+} \dot{v}_i > 0 \quad \text{and} \quad \lim_{v_i \rightarrow 1^-} \dot{v}_i < 0. \quad (4.32)$$

Consequently,  $v_i$  can never reach the left interval boundary 0 because it will move to the right when getting closer to it. The same holds for the right domain boundary 1 where  $v_i$  will move to the left before reaching it.  $\square$

**Theorem 8** (Nonequality of  $v_i$  and  $v_j$ ). *Among  $N$  initially distinct positions  $v_i \in (0, 1)$  evolving according to (4.27), no two ever become equal.*

*Proof.* Using (4.27) it is possible to derive the difference

$$\begin{aligned} \dot{v}_j - \dot{v}_i &= 2 \cdot \Phi(v_i - v_j) + \sum_{\substack{k=1 \\ k \neq i, j}}^N (\Phi(v_k - v_j) - \Phi(v_k - v_i)) \\ &\quad - \sum_{k=1}^N (\Phi(v_j + v_k) - \Phi(v_i + v_k)) \end{aligned} \quad (4.33)$$

where  $1 \leq i, j \leq N$ . Assume w.l.o.g. that  $v_j > v_i$  and consider (4.33) in the limit  $v_j - v_i \rightarrow 0$ . Then we have

$$\lim_{v_j - v_i \rightarrow 0} \dot{v}_j - \dot{v}_i = \lim_{v_j - v_i \rightarrow 0} 2 \cdot \Phi(v_i - v_j) > 0 . \quad (4.34)$$

The latter inequality follows from the fact that  $\Phi(s) > 0$  for  $s \in (-1, 0)$ . This means that  $v_j$  will always start moving away from  $v_i$  (and vice versa) when the difference between both gets sufficiently small. Since the initial positions are distinct, it follows that  $v_i \neq v_j$  for  $i \neq j$  for all times  $t$ .  $\square$

**Theorem 9** (Explicit Steady-State Solution). *Under the assumption that  $(v_i)$  is in increasing order and that  $\Psi(s^2)$  is twice continuously differentiable in  $(0, 2)$  the unique minimiser of (4.25) is given by  $\mathbf{v}^* = (v_1^*, \dots, v_{2N}^*)T$ ,  $v_i^* = (i - 1/2)/N$ ,  $i = 1, \dots, 2N$ .*

*Proof.* Equation (4.25) can be rewritten without the redundant entries of  $\mathbf{v}$  as

$$E(\mathbf{v}) = 2 \cdot \sum_{i=1}^{N-1} \sum_{j=i+1}^N \Psi((v_j - v_i)^2) + \sum_{i=1}^N \Psi(4v_i^2) + 2 \cdot \sum_{i=1}^{N-1} \sum_{j=i+1}^N \Psi((v_i + v_j)^2) \quad (4.35)$$

from which one can verify by straightforward, albeit lengthy calculations that  $\nabla E(\mathbf{v}^*) = 0$ , and the Hessian of  $E$  at  $\mathbf{v}^*$  is

$$D^2E(\mathbf{v}^*) = \sum_{k=1}^N \mathbf{A}_k \Phi' \left( \frac{k}{N} \right) \quad (4.36)$$

with sparse symmetric  $N \times N$ -matrices

$$\mathbf{A}_k = 4\mathbf{I} - 2\mathbf{T}_k - 2\mathbf{T}_{-k} + 2\mathbf{H}_{k+1} + 2\mathbf{H}_{2N-k+1} \quad , \quad k = 1, \dots, N-1 \quad , \quad (4.37)$$

$$\mathbf{A}_N = 2\mathbf{I} + 2\mathbf{H}_{N+1} \quad (4.38)$$

where the unit matrix  $\mathbf{I}$ , single-diagonal Toeplitz matrices  $\mathbf{T}_k$  and single-anti-diagonal Hankel matrices  $\mathbf{H}_k$  are defined as

$$\mathbf{I} = (\delta_{i,j})_{i,j=1}^N \quad , \quad \mathbf{T}_k = (\delta_{j-i,k})_{i,j=1}^N \quad , \quad \mathbf{H}_k = (\delta_{i+j,k})_{i,j=1}^N \quad . \quad (4.39)$$

Here,  $\delta_{i,j}$  denotes the Kronecker symbol,  $\delta_{i,j} = 1$  if  $i = j$ , and  $\delta_{i,j} = 0$  otherwise. All  $\mathbf{A}_k$ ,  $k = 1, \dots, N$  are weakly diagonally dominant with positive diagonal, thus positive semidefinite by Gershgorin's Theorem. Moreover, the tridiagonal matrix  $\mathbf{A}_1$  is of full rank, thus even positive definite. By strict convexity of  $\Psi(s^2)$ , all  $\Phi'(k/N)$  are positive, thus  $D^2E(\mathbf{v}^*)$  is positive definite.

As a consequence, the steady state of the gradient descent (4.27) for any initial data  $\mathbf{f}$  (with arbitrary rank-order) can be computed directly by sorting the  $f_i$ : Let  $\sigma$  be the permutation of  $\{1, \dots, N\}$  for which  $(f_{\sigma^{-1}(i)})_{i=1, \dots, N}$  is increasing (this is what a sorting algorithm computes), the steady state is given by  $v_i^* = (\sigma(i) - 1/2)/N$  for  $i = 1, \dots, N$ .  $\square$

**Theorem 10** (Convergence). *For  $t \rightarrow \infty$  any initial configuration  $\mathbf{v} \in (0, 1)^N$  with distinct entries converges to a unique steady state  $\mathbf{v}^*$  which is the global minimiser of the energy given in (4.35).*

*Proof.* As a sum of convex functions, (4.35) is convex. Therefore the function  $V(\mathbf{v}) := E(\mathbf{v}) - E(\mathbf{v}^*)$  (where  $\mathbf{v}^*$  is the equilibrium point) is a Lyapunov function with  $V(\mathbf{v}^*) = 0$  and  $V(\mathbf{v}) > 0$  for all  $\mathbf{v} \neq \mathbf{v}^*$ . Furthermore, we have

$$\partial_t V(\mathbf{v}) = - \sum_{i=1}^N (\partial_{v_i} E(\mathbf{v}))^2 \leq 0 \quad . \quad (4.40)$$

Note that due to the positive definiteness of (4.36) we know that  $E(\mathbf{v})$  has a strict (global) minimum which implies that the inequality in (4.40) becomes strict except in case of  $\mathbf{v} = \mathbf{v}^*$ . This guarantees asymptotic Lyapunov stability of  $\mathbf{v}^*$  and thus convergence to  $\mathbf{v}^*$  for  $t \rightarrow \infty$ .  $\square$

## 4.2.2 Generalisation with weights

The model (4.27), although nicely behaved, has a drawback in terms of image modelling. We assumed from the beginning that the  $v_i$  are all distinct. If we were to represent an images grey values with the  $v_i$ , this would mean that we could only consider images which have at most one pixel of a given grey value. In order to overcome this drawback, we now introduce a slightly more general version of the previous model by adding some additional constant weights which will allow us to model the presence of more than one pixel with the same grey value. Let us now consider a generalised version of our model that allows for localisation and different treatment of distinct  $v_i$ . For this purpose we make use of vectors  $\mathbf{w} = (w_1, \dots, w_N)^T \in (0, \infty)^N$  and  $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_N)^T \in (\mathbb{R}^n)^N$  which we extend – similar to  $\mathbf{v}$  – with the coordinates  $w_{N+1}, \dots, w_{2N}$  and  $\mathbf{x}_{N+1}, \dots, \mathbf{x}_{2N}$ . Both are defined as  $w_{2N+1-i} = w_i$  and  $\mathbf{x}_{2N+1-i} = \mathbf{x}_i$ . Each  $w_i$  denotes the weight, or importance, of the corresponding  $v_i$ , whereas the  $\mathbf{x}_i$  provide additional  $n$ -dimensional position information. Neither  $w_i$  nor  $\mathbf{x}_i$  change over time. Additionally, we introduce a weighting function  $\gamma(|x|)$  which is 1 for  $|x| \leq \varrho$  and 0 else for  $\varrho > 0$ . Now regard the adapted variant of (4.25) given by

$$E(\mathbf{p}, \mathbf{x}, \mathbf{w}) = \frac{1}{2} \sum_{i=1}^{2N} \sum_{j=1}^{2N} w_i \cdot w_j \cdot \gamma(|\mathbf{x}_j - \mathbf{x}_i|) \cdot \Psi((v_j - v_i)^2), \quad (4.41)$$

where we make use of the coordinate transform  $p_i := \sqrt{w_i} \cdot v_i$ ,  $i = 1, \dots, 2N$ . Referring to (4.26), a gradient descent can be formulated as

$$\dot{p}_i = \sqrt{w_i} \cdot \sum_{\substack{j=1 \\ j \neq i}}^{2N} w_j \cdot \gamma(|\mathbf{x}_j - \mathbf{x}_i|) \cdot \Phi \left( \left( \frac{p_j}{\sqrt{w_j}} - \frac{p_i}{\sqrt{w_i}} \right)^2 \right) \quad (4.42)$$

for  $i = 1, \dots, 2N$ . Since  $\dot{p}_{2N+1-i} = -\dot{p}_i$ , we can drop the redundant entries and rewrite (4.42) with  $\dot{p}_i = \sqrt{w_i} \cdot \dot{v}_i$  for  $i = 1, \dots, N$  as

$$\dot{v}_i = \sum_{\substack{j=1 \\ j \neq i}}^N w_j \cdot \gamma(|\mathbf{x}_j - \mathbf{x}_i|) \cdot \Phi(v_j - v_i) - \sum_{j=1}^N w_j \cdot \gamma(|\mathbf{x}_j - \mathbf{x}_i|) \cdot \Phi(v_i + v_j) . \quad (4.43)$$

### Properties of the generalised model.

A similar result to Theorem 7 holds also for the generalised model.

**Theorem 11** (Avoidance of Boundaries). *The  $N$  initially distinct positions  $v_i \in (0, 1)$  evolving according to (4.43) never reach the domain boundaries 0 and 1.*

*Proof.* The definition of  $\Phi$  implies that

$$\lim_{h \rightarrow 0^+} \frac{\Psi((0+h)^2) - \Psi(0)}{h} < 0 \quad \text{and} \quad \lim_{h \rightarrow 0^-} \frac{\Psi(2) - \Psi((2+h)^2)}{h} > 0 , \quad (4.44)$$

from which it follows that

$$\lim_{v_i \rightarrow 0^+} (-\Phi(2v_i)) > 0 \quad \text{and} \quad \lim_{v_i \rightarrow 1^-} (-\Phi(2v_i)) < 0 . \quad (4.45)$$

Equation (4.43) can be written as

$$\dot{v}_i = \sum_{\substack{j=1 \\ j \neq i}}^N w_j \cdot \gamma(|\mathbf{x}_j - \mathbf{x}_i|) \cdot (\Phi(v_j - v_i) - \Phi(v_i + v_j)) - w_i \Phi(2v_i) . \quad (4.46)$$

Since for  $j = 1, \dots, N$  and  $j \neq i$  one has

$$\lim_{v_i \rightarrow 0^+} \Phi(v_j - v_i) - \Phi(v_i + v_j) = 0 \quad \text{and} \quad \lim_{v_i \rightarrow 1^-} \Phi(v_j - v_i) - \Phi(v_i + v_j) = 0 , \quad (4.47)$$

it follows that

$$\lim_{v_i \rightarrow 0^+} \dot{v}_i > 0 \quad \text{and} \quad \lim_{v_i \rightarrow 1^-} \dot{v}_i < 0 . \quad (4.48)$$

Consequently,  $v_i$  can never reach the left interval boundary 0 because it will move to the right when getting closer to it. The same holds for the right domain boundary 1 where  $v_i$  will move to the left before reaching it.  $\square$

Moreover, if  $\gamma = 1$  we also have an similar result to Theorem 8.

**Theorem 12** (Nonequality of  $v_i$  and  $v_j$ ). *Among  $N$  initially distinct positions  $v_i \in (0, 1)$  evolving according to (4.43), no two ever become equal.*

*Proof.* Using (4.43) it is possible to derive the difference

$$\begin{aligned} \dot{v}_j - \dot{v}_i &= (w_1 + w_2) \cdot \Phi(v_i - v_j) + \sum_{\substack{k=1 \\ k \neq i, j}}^N w_k (\Phi(v_k - v_j) - \Phi(v_k - v_i)) \\ &\quad - \sum_{k=1}^N w_k (\Phi(v_j + v_k) - \Phi(v_i + v_k)) \end{aligned} \quad (4.49)$$

where  $1 \leq i, j \leq N$ . Assume w.l.o.g. that  $v_j > v_i$  and consider (4.49) in the limit  $v_j - v_i \rightarrow 0$ . Then we have

$$\lim_{v_j - v_i \rightarrow 0} \dot{v}_j - \dot{v}_i = \lim_{v_j - v_i \rightarrow 0} (w_i + w_j) \cdot \Phi(v_i - v_j) > 0 . \quad (4.50)$$

The latter inequality follows from the fact that  $\Phi(s) > 0$  for  $s \in (-1, 0)$ . This means that  $v_j$  will always start moving away from  $v_i$  (and vice versa) when the difference between both gets sufficiently small. Since the initial positions are distinct, it follows that  $v_i \neq v_j$  for  $i \neq j$  for all times  $t$ .  $\square$

A minimiser  $\mathbf{p}^*$  for  $E(\mathbf{p}, \mathbf{x}, \mathbf{w})$  depends in general on the definition of  $\Psi$ . As evident from Theorem 9 this dependency vanishes in the special case  $\gamma = 1$ ,  $w_i = 1$ , for  $i = 1, \dots, N$ . For nontrivial  $w_i$ , if we assume that  $\Phi$  belongs to the class of linear functions, i.e.  $\Phi(s) = a \cdot (s - 1)$ ,  $a > 0$  (cf. Figure 4.3) and that  $\varrho$  is large enough – such that  $\gamma = 1$  for all pairs  $(\mathbf{x}_i, \mathbf{x}_j)$  – we get that the point  $\mathbf{p}^*$ , defined as follows:

$$p_i^* = \sqrt{w_i} \cdot v_i^* = \sqrt{w_i} \cdot \frac{\sum_{j=1}^i w_j - \frac{1}{2}w_i}{\sum_{j=1}^N w_j} , \quad i = 1, \dots, N, \quad (4.51)$$

is a critical points of  $E(\mathbf{p}, \mathbf{x}, \mathbf{w})$ . Moreover, the following theorem implies that the point  $\mathbf{v}^*$  is the unique global minimizer of the energy (4.41).

**Theorem 13** (Convergence). *If we assume that  $\gamma = 1$  and that  $\Phi(s) = a \cdot (s-1)$ ,  $a > 0$ , is a linear function. Then, for  $t \rightarrow \infty$ , any initial configuration  $\mathbf{v} \in (0, 1)^N$  with distinct entries evolving according to (4.43), converges to the unique steady state  $\mathbf{v}^*$  which is the global minimiser of the energy given in (4.41).*

*Proof.* As a sum of convex functions, (4.41) is convex. Therefore the function  $V(\mathbf{v}) := E(\mathbf{v}, \mathbf{x}, \mathbf{w}) - E(\mathbf{v}^*, \mathbf{x}, \mathbf{w})$  (where  $\mathbf{v}^*$  is the equilibrium point) is a Lyapunov function with  $V(\mathbf{v}^*) = 0$  and  $V(\mathbf{v}) > 0$  for all  $\mathbf{v} \neq \mathbf{v}^*$ . Furthermore, we have

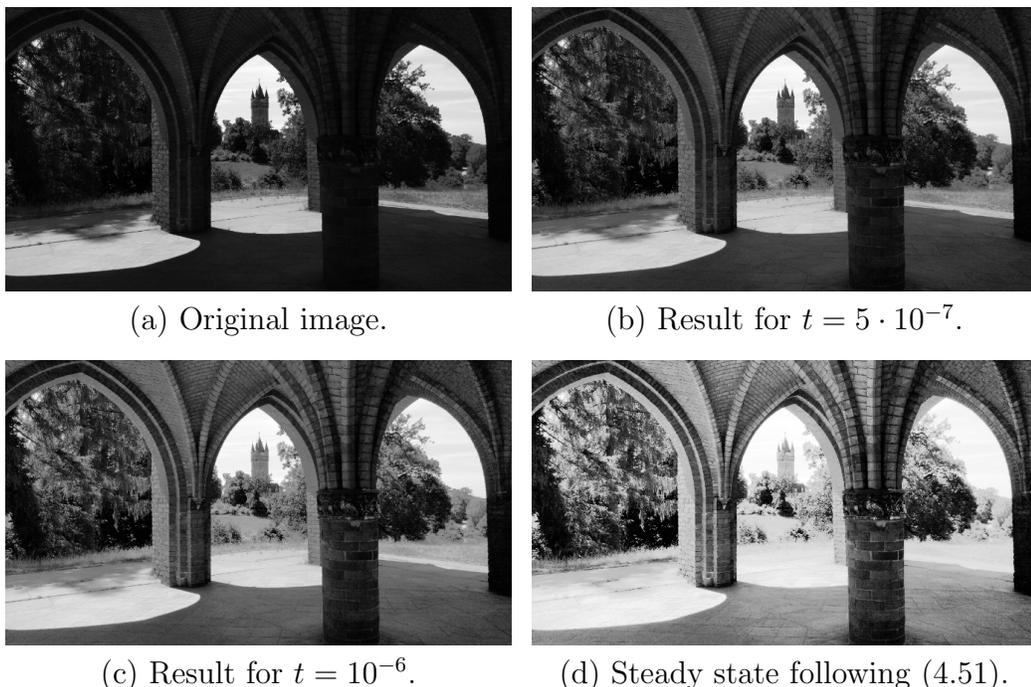
$$\partial_t V(\mathbf{v}) = - \sum_{i=1}^N (\partial_{v_i} E(\mathbf{v}))^2 \leq 0 . \quad (4.52)$$

The positive definiteness of the Hessian of  $E$  at  $\mathbf{v}^*$ ,  $D^2 E(\mathbf{v}^*)$ , follows from the linearity of  $\Phi(s)$ . Therefore, we know that  $E(\mathbf{v}, \mathbf{x}, \mathbf{w})$  has a strict (global) minimum which implies that the inequality in (4.52) becomes strict except in case of  $\mathbf{v} = \mathbf{v}^*$ . This guarantees asymptotic Lyapunov stability of  $\mathbf{v}^*$  and thus convergence to  $\mathbf{v}^*$  for  $t \rightarrow \infty$ .  $\square$

### 4.2.3 Connections to the nonlocal framework

As mentioned at the beginning of Section 4.2, the processes obtained as the gradient descent of the energy function (4.25), respectively (4.41), are directly related to the nonlocal process (4.19). In the first case, this can be easily verified observing that (4.27) corresponds to (4.19) with the special choice of  $\gamma = 1$ , if the variables  $v_i$  are renamed as  $u_i$ . In the case of the generalised model of Section 4.2.2, it can be seen that the expression (4.43) is equivalent to (4.19) if  $\gamma(d(i, j))$  is identified with  $\gamma(|\mathbf{x}_i - \mathbf{x}_j|)$ .

Up to this points we have been considering the nonlocal initial value problem condition (NL1) which assumes the regularity of  $K$ . Here, however, we are dealing with a function  $g$  which is not, as can be seen in Figure 4.3. It is thus worth mentioning that here the existence of the corresponding evolution, comes from the interpretation of the process as a gradient descent of (4.41). Rather than the well-posedness theory developed in Section 2.1.



(a) Original image. (b) Result for  $t = 5 \cdot 10^{-7}$ .  
(c) Result for  $t = 10^{-6}$ . (d) Steady state following (4.51).  
Figure 4.5: Processing a photograph of “Flatowturm (Potsdam)”.

#### 4.2.4 Explicit scheme and experiments

Using forward differences to approximate the time derivative in (4.43) and using  $\gamma_{i\ell} := \gamma(|\mathbf{x}_\ell - \mathbf{x}_i|)$  the explicit scheme of the generalised model reads

$$v_i^{k+1} = v_i^k + \tau \cdot \sum_{\substack{\ell=1 \\ \ell \neq i}}^N w_\ell \cdot \gamma_{i\ell} \cdot \Phi(v_\ell^k - v_i^k) - \tau \cdot \sum_{\ell=1}^N w_\ell \cdot \gamma_{i\ell} \cdot \Phi(v_i^k + v_\ell^k), \quad i = 1, \dots, N, \quad (4.53)$$

where  $\tau$  denotes the time step size and an upper index  $k$  refers to the time  $k\tau$ . Theorem 6 can be easily adapted to this setting to obtain a corresponding bound for the time step which results in a stable explicit scheme.

#### Image enhancement.

Similar to the approach proposed in [63] we apply our model to enhance the *global* contrast of digital grey value images  $f : \{1, \dots, n_x\} \times \{1, \dots, n_y\} \rightarrow$

$[0, 1]$ . As illustrated in Figure 4.5, this can be achieved in two ways. The first option uses the explicit scheme (4.53) to describe the evolution of grey values  $v_i$  up to some time  $t$  (see Figure 4.5 (b) and (c)) where the weights  $w_i$  reflect the multiplicity of each grey value  $v_i$ . Note that all grey values are mapped to the interval  $(0, 1)$  beforehand to ensure the validity of our model and  $\gamma$  is fixed to 1. The amount of contrast enhancement grows with increasing values of  $t$ . In our experiments an image size of  $683 \times 384$  pixels and the application of a flux function  $\Phi$  with  $L_\Phi = 1$  (see Figure 4.3) imply an upper bound of  $1/(2 \cdot 683 \cdot 384)$  for  $\tau$  and allow us to achieve the presented results after just one iteration.

If one is only interested in an enhanced version of the original image with maximum global contrast there is an alternative, namely the derived steady state solution for linear flux functions (4.51). The result is shown in Figure 4.5 (d). Figure 4.5 also confirms that the solution of the explicit scheme (4.53) converges to the steady state solution (4.51) for  $t \rightarrow \infty$ . From (4.51) it is clear that the steady state is equivalent to histogram equalisation. It is therefore interesting to compare our evolution with the histogram modification flow introduced in [56] which can have the same steady state. Indeed, the flow from [56] can also be translated into a combination of repulsion among grey-values and a barrier function. However, as in [56] the repulsive force is constant, and the barrier function quadratic, they cannot be derived from the same kind of interaction between the  $v_i$  and their reflected counterparts.

### 4.3 Higher order nonlocal equations

Starting from the work of Perona and Malik [3], the application of diffusion filtering to image processing has given rise to many different approaches based on partial differential equation. Even though these processes produce high quality results for image simplification and denoising applications, in practical results, Perona-Malik filters have one major drawback. Since their results contain piecewise constant regions, linear gray value transitions are hard to recover. These filters tend to oversegment them, giving rise to the so-called staircasing artifact [64]. The most promising idea to overcome this limitation is that of introducing higher derivative orders in the diffusion filter models [65–67]. Scale space properties and numerical implementations of

these type of higher order filtering, analogous to the ones discussed in Chapter 1, were studied in [68].

Up to this point we have been considering nonlocal processes which are directly related to nonlinear diffusion filtering. However, relaxing the nonnegativity condition (NL2) can lead to processes which are related to higher order diffusion filtering. We will pursue this direction further but, nevertheless, we would like to notice the following result which exemplifies how this connection may come to be. The following process is related to the example about linear shift-invariant operators considered in Section 2.3.1.

$$\partial_t u = \int_{\Omega} J(\mathbf{x} - \mathbf{y}) (u(\mathbf{y}, t) - u(\mathbf{x}, t)) d\mathbf{y}, \quad \text{in } [0, T] \times \Omega, \quad (4.54)$$

$$u(\mathbf{x}, 0) = f(\mathbf{x}) \quad \text{in } \Omega, \quad (4.55)$$

with  $f \in L^1(\Omega)$  and some radial function  $J \in C(\mathbb{R}^N, \mathbb{R})$  such that  $J(0) > 0$  and  $\int_{\mathbb{R}^N} J(\mathbf{x}) d\mathbf{x} = 1$ . Moreover assume that

$$\hat{J}(\eta) = 1 - A|\eta|^\alpha + o(|\eta|^\alpha) \quad \text{for } t \rightarrow \infty.$$

This process was studied in [69]. The authors proved that a family of solutions  $u_\epsilon$  of (4.54) with  $J$  replaced by an appropriate rescaled version  $J_\epsilon$ , approximates the solution of the following higher order initial valued problem.

$$\partial_t u = -A^n (-\Delta)^{\frac{n\alpha}{2}} u(\mathbf{x}, t) \quad \text{in } ]0, T] \times \Omega, \quad (4.56)$$

$$\frac{du}{d\nu} = 0 \quad \text{on } \partial\Omega, \quad (4.57)$$

$$u(\mathbf{x}, 0) = f(\mathbf{x}) \quad \text{in } \Omega, \quad (4.58)$$

then

$$\lim_{\epsilon \rightarrow 0} \|u_\epsilon - v\|_{L^\infty(\Omega \times [0, t_0])} = 0. \quad (4.59)$$

In this sense, the nonlocal problem (4.54) solves a higher order diffusion problem.

## 4.4 Conclusions

In this chapter we have discussed the consequences of relaxing the nonnegativity condition (NL2) of framework (1.3)–(1.4). We identified the loss of

---

the maximum-minimum principle and introduced an alternative stabilisation procedure based on a reflexion on the co-domain of the image evolution. In Section 4.1, we saw that thanks to this stabilization we are able to obtain an explicit scheme discretisation which provides a stable algorithm for a wide range of forward and backwards nonlocal processes provided that the weight function, is regular enough. Moreover, in Section 4.2 we reviewed the results of [59] in which we study the case of a special family of weight functions. In this case, we saw that other than defining a stable process, it is also possible to analyse its long time behaviour by providing an interpretation in terms of convex optimisation.



# Chapter 5

## Summary and outlook

In this work we have studied a wide set of nonlocal processes and their applications as image filters. The guideline of the work was based on a general framework defined as an initial valued problem involving an integro-differential equation,

$$u_t(\mathbf{x}, t) = \int_{\Omega} K(\mathbf{x}, \mathbf{y}, u(\cdot, t))(u(\mathbf{y}, t) - u(\mathbf{x}, t))d\mu(\mathbf{y}) \quad \text{in } \bar{\Omega} \times [0, T],$$
$$u(\mathbf{x}, 0) = f(\mathbf{x}) \quad \text{in } \bar{\Omega}.$$

In each chapter we considered different specific sub families of this general problem, which we obtained by means of imposing a particular set of constraints to the weight function  $K$ . The complete list of assumptions considered along the thesis is summarized in Table 5.1.

The study of the different subclasses is organized in the following way: In Chapter 2 we imposed the assumptions (NL1) to (NL4) and showed that the resulting family give rise to filters which satisfy a complete nonlocal scale-space theory. In Chapter 3 we studied the consequences of removing the symmetry condition (NL3) and the irreducibility condition (NL4). Instead, we introduced the alternative assumptions given by the type-symmetry (NL5) and the local boundedness (NL6). These provided enough ground to study the long time behaviour of semidiscrete processes with a possibly nonsymmetric  $K$ . In the final chapter we consider the case in which the nonnegativity condition (NL2) is removed. A summary of the structure of the work and

the main characteristics of the families studies on each chapter is given in Table 5.1 and Table 5.2. The different processes considered in this work

Assumption	Description
Regularity (NL1):	$K : \bar{\Omega} \times \bar{\Omega} \times C(\bar{\Omega}) \rightarrow \mathbb{R}$ , is Lipschitz continuous and bounded.
Non-negativity (NL2):	The function $K$ is non-negative.
Symmetry (NL3):	$K(\mathbf{x}, \mathbf{y}, u) = K(\mathbf{y}, \mathbf{x}, u)$ in $\bar{\Omega} \times \bar{\Omega}$ , for every $u \in C(\bar{\Omega})$ .
Irreducibility (NL4):	There exists a finite family of $\mu$ -measurable sets $\mathcal{F} := \{B_i \subset \Omega : 1 \leq i \leq p\}$ , such that: There exists a constant $c > 0$ , independent of $T$ , such that $K(\mathbf{x}, \mathbf{y}, u(\cdot, s)) \geq c$ for $0 \leq s \leq T$ , whenever $u \in C(\bar{\Omega} \times [0, T])$ , $B \in \mathcal{F}$ and $\mathbf{x}, \mathbf{y} \in B$ . Moreover, $\Omega = \bigcup_{i=1}^p B_i$ and $\mu(B_i \cap B_{i+1}) > 0$ for $1 \leq i \leq p - 1$ .
Type-symmetry (NL5):	There exists a constant $c > 0$ such that $c^{-1}k(i, j, \cdot) < k(i, j, \cdot) < ck(i, j, \cdot)$ .
Local boundedness (NL6):	There exists positive values $P, p > 0$ such that for every pair of pixels $i, j$ , either $p \leq K(i, j, \cdot) \leq P$ or $K(i, j, \cdot) = 0$ . Moreover, if $i$ and $j$ are neighbouring pixels in the image grid, then $k(i, j, \cdot) > p$ .

Table 5.1: Summary of the main assumptions.

turn up to satisfy a series of interesting properties, as showed in the present work. These, in turn, bring up a series of questions, and imply a series of extensions, which could be the topic of future research:

**About the symmetric processes of Chapter 2:** We notice that even though it was possible to obtain a satisfactory theory for nonlocal nonlinear scale-spaces, the same cannot be said about their applications as image processing filters. In fact, we observed that the proposed scale-spaces based on

	Chapter 2	Chapter 3	Chapter 4
Regularity (NL1):	yes	yes	yes/no
Non-negativity (NL2):	yes	yes	no
Symmetry (NL2):	yes	no	no
Irreducibility (NL3):	yes	no	no
Type-symmetry (NL4):	no	yes	no
Local boundedness (NL5):	no	yes/no	no

Table 5.2: Summary of the chapters. The option yes/no means that both possibilities are considered in that chapter.

bilateral and NL-means filters, did not perform as good as the original filters. Moreover, we identified the symmetry condition might be a plausible explanations for this behaviour. Therefore, it is still an open issue to determine other types of specific filters which may have a better performance than the ones studied in Section 2.3 of Chapter 2, or to identify other applications where the nonlocal scale-space theory would be beneficial.

**About the nonsymmetric processes of Chapter 3:** We noticed that the existence and uniqueness results developed in Chapter 2 are still valid in the case of the nonsymmetric process, given that the regularity and nonnegativity assumptions are still fulfilled. This is true for the general measure  $\mu$ , and thus also in particular for the spatially continuous process obtained by selecting the specific choice of the Lebesgue measure, which lead to the usual integral in  $\mathbb{R}^N$ . The long time results developed on Chapter 3, on the other hand, were obtained only for the semidiscrete process. Therefore, one possible extension of the theory developed for nonsymmetric processes, is that of thoroughly verifying that the long time behaviour of the spatially continuous process corresponds to that of the semidiscrete one. For that matter, it would be necessary to obtained new assumptions that extend the type-symmetry and boundedness conditions (NL5)–(NL6), which originate from the particle swarm systems theory and are inherently semidiscrete. One such spatially continuous process, which would be interesting on its own right, is that of the Sector Diffusion introduced in Section 3.4. We mentioned that it would be possible, at least formally, to consider the spatially continuous version (3.18) of it, and even a localized version (3.16) which would extend the model initially proposed in [47]. This last model is defined by means of one sided

partial differentials operators. These processes could be an alternative to the usual partial derivatives and the resulting equations could be considered as PDE version of the one sided derivatives studied in the ODE literature [70]. To the best of our knowledge, such PDE have not been thoroughly studied before.

**About the nonlocal forward and backward processes of Chapter 4:**

Firstly, similar to the nonnegative nonsymmetric theory, the results developed in this work are most relevant for the semidiscrete case. Hence, the space continuous theory is still open to further analysis. In that case, even the global existence of a solution cannot be taken for granted.

Regarding the long time behaviour of the solutions, we where able to establish a convergence theorem only in the case of a specific family of weight functions. However, we also saw in Section 4.1 that even for quite general weight functions, the proposed stabilization procedure, produced a new semidiscrete process which is stable and produces an explicit scheme which is also stable. In this direction, it would be interesting to establish a more general theory regarding the long time behaviour of the process with more general weight functions, involving positive and negative values.

# List of symbols

The following notation is used throughout the whole thesis. Scalar values and functions are denoted with lower case letters. Vector values with lower case bold letters. Matrices are denoted with upper case letters.

Symbol	Description
$\nabla$	Gradient operator
$C(\Omega)$	Space of continuous function defined in $\Omega$
$\bar{\Omega}$	Closure of a set $\Omega$
$\Delta$	Laplacian operator
$ \cdot $	Norm of a vector. Absolute value in case of a scalar
$\dot{\cdot}$	Derivative with respect to time $t$
$\ \cdot\ _\infty$	Essential supremum. Coincides with $\max_{\mathbf{x} \in \bar{\Omega}} u(\mathbf{x})$ for $u \in C(\bar{\Omega})$
$L^2(\Omega, \mu)$	Space of measurable functions $f$ in $\Omega$ such that $\int_\Omega  f(\mathbf{x}) ^2 d\mu(\mathbf{x}) < \infty$



# Bibliography

- [1] Weickert, J.: Anisotropic Diffusion in Image Processing. Teubner, Stuttgart (1998)
- [2] Iijima, T.: Basic theory on normalization of pattern (in case of typical one-dimensional pattern). Bulletin of the Electrotechnical Laboratory **26** (1962) 368–388 In Japanese.
- [3] Perona, P., Malik, J.: Scale space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence **12** (1990) 629–639
- [4] Aurich, V., Weule, J.: Non-linear Gaussian filters performing edge preserving diffusion. In Sagerer, G., Posch, S., Kummert, F., eds.: Mustererkennung 1995. Springer, Berlin (1995) 538–545
- [5] Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Proc. Sixth International Conference on Computer Vision, Bombay, India, Narosa Publishing House (January 1998) 839–846
- [6] Buades, A., Coll, B., Morel, J.M.: A review of image denoising algorithms, with a new one. Multiscale Modeling and Simulation **4**(2) (2005) 490–530
- [7] Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3D transform-domain collaborative filtering. IEEE Transactions on Image Processing **16**(8) (August 2007) 2080–2095
- [8] Witkin, A.P.: Scale-space filtering. In: Proc. Eighth International Joint Conference on Artificial Intelligence. Volume 2., Karlsruhe, West Germany (August 1983) 945–951

- [9] Koenderink, J.J.: The structure of images. *Biological Cybernetics* **50** (1984) 363–370
- [10] Lowe, D.L.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* **60**(2) (2004) 91–110
- [11] Lindeberg, T.: *Scale-Space Theory in Computer Vision*. Kluwer, Boston (1994)
- [12] Florack, L.: *Image Structure*. Volume 10 of *Computational Imaging and Vision*. Kluwer, Dordrecht (1997)
- [13] Weickert, J., Ishikawa, S., Imiya, A.: Linear scale-space has first been proposed in Japan. *Journal of Mathematical Imaging and Vision* **10**(3) (May 1999) 237–252
- [14] Felsberg, M., Sommer, G.: Scale-adaptive filtering derived from the Laplace equation. In Radig, B., Florczyk, S., eds.: *Pattern Recognition*. Volume 2032 of *Lecture Notes in Computer Science*. Springer, Berlin (2001) 95–106
- [15] Duits, R., Florack, L., de Graaf, J., ter Haar Romeny, B.: On the axioms of scale space theory. *Journal of Mathematical Imaging and Vision* **20** (2004) 267–298
- [16] van den Boomgaard, R.: The morphological equivalent of the Gauss convolution. *Nieuw Archief Voor Wiskunde* **10**(3) (November 1992) 219–236
- [17] Alvarez, L., Guichard, F., Lions, P.L., Morel, J.M.: Axioms and fundamental equations in image processing. *Archive for Rational Mechanics and Analysis* **123** (1993) 199–257
- [18] Guichard, F.: A morphological, affine, and Galilean invariant scale-space for movies. *IEEE Transactions on Image Processing* **7**(3) (March 1998) 444–456
- [19] Lindeberg, T.: Generalized Gaussian scale-space axiomatics comprising linear scale-space, affine scale-space and spatio-temporal scale-space. *Journal of Mathematical Imaging and Vision* **40** (2011) 36–81

- 
- [20] Scherzer, O., Weickert, J.: Relations between regularization and diffusion filtering. *Journal of Mathematical Imaging and Vision* **12**(1) (February 2000) 43–63
- [21] Gilboa, G.: Nonlinear scale space with spatially varying stopping time. *IEEE Transactions on Pattern Analysis and Machine Learning* **30**(12) (2008) 2175–2187
- [22] Gilboa, G., Osher, S.: Nonlocal operators with applications to image processing. *Multiscale Modeling and Simulation* **7**(3) (2008) 1005–1028
- [23] G. Gilboa, S.O.: Nonlocal linear image regularization and supervised segmentation. *Multiscale Modeling and Simulation* **6**(2) (2007) 595–630
- [24] Ballester, C., Calderero, F., Caselles, V., Facciolo, G.: Multiscale analysis of similarities between images on riemannian manifolds. *Multiscale Modeling and Simulation* **12**(2) (2014) 616–649
- [25] Cárdenas, G.M., Weickert, J., Schäffer, S.: A linear scale-space theory for continuous nonlocal evolutions. In J.-F. Aujol, M.N., Papadakis, N., eds.: *Scale Space and Variational Methods in Computer Vision*. Volume 9087. Springer (2015) 103–114
- [26] Andreu-Vaillo, F., Mazón, J.M., Rossi, J.D.: *Nonlocal Diffusion Problems*. American Mathematical Society, Providence (2010)
- [27] Catté, F., Lions, P.L., Morel, J.M., Coll, T.: Image selective smoothing and edge detection by nonlinear diffusion. *SIAM Journal on Numerical Analysis* **32** (1992) 1895–1909
- [28] Weickert, J.: Theoretical foundations of anisotropic diffusion in image processing. *Computing Supplement* **11** (1996) 221–236
- [29] Weickert, J.: *Anisotropic Diffusion in Image Processing*. PhD thesis, Department of Mathematics, University of Kaiserslautern, Germany (January 1996) Revised and extended version published by Teubner, Stuttgart, Germany, 1998.
- [30] Vogel, O., Hagenburg, K., Weickert, J., Setzer, S.: A fully discrete theory for linear osmosis filtering. In Kuijper, A., Bredies, K., Pock, T., Bischof, H., eds.: *Scale Space and Variational Methods in Computer*

- Vision. Volume 7893 of Lecture Notes in Computer Science. Springer, Berlin (2013) 368–379
- [31] Weickert, J., Hagenburg, K., Breuß, M., Vogel, O.: Linear osmosis models for visual computing. In Heyden, A., Kahl, F., Olsson, C., Oskarsson, M., Tai, X.C., eds.: Energy Minimisation Methods in Computer Vision and Pattern Recognition. Volume 8081 of Lecture Notes in Computer Science. Springer, Berlin (2013) 26–39
- [32] Hendrickx, J.M., Tsitsiklis, J.N.: Convergence of type-symmetric and cut-balanced consensus seeking systems. *IEEE Transactions on Automatic Control* **58**(1) (2013) 214–218
- [33] Weickert, J.: Coherence-enhancing diffusion filtering. *International Journal of Computer Vision* **31**(2/3) (April 1999) 111–127
- [34] Ren, W., Beard, R., Atkins, E.: A survey of consensus problems in multi-agent coordination. In: Proc. of the 2005 American Control Conference, Portland (2005) 1859–1864
- [35] Couzin, I., Krause, J., James, R., Ruxton, G., Franks, N.: Collective memory and spatial sorting in animal groups. *Journal of Theoretical Biology* **218**(1) (2002) 1–11
- [36] Cucker, F., Smale, S.: Emergent behavior in flocks. *IEEE Transactions on Automatic Control* **52**(5) (2007) 852–862
- [37] Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I., Shochet, O.: Novel type of phase transition in a system of self-driven particles. *Physical Review Letters* **75** (1995) 1226–1229
- [38] Chung, F.R.K.: Spectral graph theory. In: Regional Conference Series in Mathematics. Number 92. American Mathematical Society (1997)
- [39] Li, J.S., Zhang, X.D.: On the Laplacian eigenvalues of a graph. *Linear Algebra and its Applications* **285** (1998) 305–307
- [40] Belkin, M., Niyogi, P.: Laplacian eigenmaps and spectral techniques for embedding and clustering. *Advances in Neural Information Processing Systems* (2002) 585–591

- 
- [41] Riedel, K.: Corner-preserving anisotropic diffusion and junction detection using the structure tensor. In: *Mustererkennung 1999: DAGM-Symposium Bonn*, Bonn, Springer (1999) 164–171
- [42] Nitzberg, M., Shiota, T.: Nonlinear image filtering with edge and corner enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**(8) (1992) 826–833
- [43] El-Fallah, A.I., Ford, G.E.: The evolution of mean curvature in image filtering. *IEEE International Conference on Image Processing* **1** (1994) 298–302
- [44] Steidl, G., Teuber, T.: Anisotropic smoothing using double orientations. In Tai, X., Morken, K., Lysaker, M., Lie, K., eds.: *Scale Space and Variational Methods in Computer Vision*. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg (2009) 477–489
- [45] Buades, A., Coll, B., Morel, J.: A non-local algorithm for image denoising. In: *Proc. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society Press (2005) 60–65
- [46] Buades, A., Coll, B., Morel, J.M.: Non-local means denoising. *Image Processing On Line* (2011)
- [47] Weickert, J.: Anisotropic diffusion filters for image processing based quality control. In Fasano, A., Primicerio, M., eds.: *Proc. Seventh European Conference on Mathematics in Industry*. Teubner, Stuttgart (1994) 355–362
- [48] Weickert, J.: Coherence-enhancing shock filters. In Michaelis, B., Krell, G., eds.: *Pattern Recognition*. Volume 2781 of *Lecture Notes in Computer Science*. Springer, Berlin (2003) 1–8
- [49] Li, F., Pi, L., Zeng, T.: Explicit coherence enhancing filter with spatial adaptive elliptical kernel. *IEEE Signal Processing Letters* **19**(9) (September 2012) 555–558
- [50] Förstner, W., Gülch, E.: A fast operator for detection and precise location of distinct points, corners and centres of circular features. In:

- Proc. ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland (June 1987) 281–305
- [51] Welk, M., Weickert, J., Steidl, G.: From tensor-driven diffusion to anisotropic wavelet shrinkage. In Bischof, H., Leonardis, A., Pinz, A., eds.: *Computer Vision – ECCV 2006, Part I*. Volume 3951 of *Lecture Notes in Computer Science*. Springer, Berlin (2006) 391–403
- [52] Didas, S., Weickert, J.: From adaptive averaging to accelerated nonlinear diffusion filtering. In Franke, K., Müller, K.R., Nickolay, B., Schäfer, R., eds.: *Pattern Recognition*. Volume 4174 of *Lecture Notes in Computer Science*. Springer, Berlin (2006) 101–110
- [53] Osher, S., Rudin, L.: Shocks and other nonlinear filtering applied to image processing. In Tescher, A.G., ed.: *Applications of Digital Image Processing XIV*. Volume 1567 of *Proceedings of SPIE*. SPIE Press, Bellingham (1991) 414–431
- [54] Gilboa, G., Sochen, N.A., Zeevi, Y.Y.: Forward-and-backward diffusion processes for adaptive image enhancement and denoising. *IEEE Transactions on Image Processing* **11**(7) (2002) 689–703
- [55] Sochen, N.A., Zeevi, Y.Y.: Resolution enhancement of colored images by inverse diffusion processes. In: *Proc. 1998 IEEE International Conference on Acoustics, Speech and Signal Processing*, Seattle, WA (May 1998) 2853–2856
- [56] Sapiro, G., Caselles, V.: Histogram modification via differential equations. *Journal of Differential Equations* **135** (1997) 238–268
- [57] G. Gilboa, J. Darbon, S.O., Chan, T.: Image enhancement and denoising by complex diffusion processes. *IEEE Transactions on Pattern Analysis and Machine Learning* **26**(8) (2004) 1020–1036
- [58] Gilboa, G., Sochen, N., Zeevi, Y.Y.: Image sharpening by flows based on triple well potentials. *Journal of Mathematical Imaging and Vision* **20** (2004) 121–131
- [59] Bergerhoff, L., Cárdenas, M., Weickert, J., Welk, M.: Modelling stable backward diffusion and repulsive swarms with convex energies and

- 
- range constraints. In: To appear in Proc. 11th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition, Venice, Italy (2017)
- [60] Markdahl, J., Goncalves, J.: Global convergence properties of a consensus protocol on the  $n$ -sphere. In: 55th Conference on Decision and Control, Las Vegas, NV, USA, IEEE (2016) 3487–3492
- [61] Hegarty, P., Martinsson, A., Wedin, E.: The Hegselmann-Krause dynamics on the circle converges. *Journal of Difference Equations and Application* **22**(11) (2016) 1720–1731
- [62] Sepulchre, R.: Consensus on nonlinear spaces. *Annual Reviews in Control* **35**(1) (2010) 56–64
- [63] Bergerhoff, L., Weickert, J.: Modelling image processing with discrete first-order swarms. In Pillay, N., Engelbrecht, P.A., Abraham, A., du Plessis, C.M., Snášel, V., Muda, K.A., eds.: *Advances in Nature and Biologically Inspired Computing*. Volume 419. Springer, Cham (2016) 261–270
- [64] Buades, A., Coll, B., Morel, J.M.: The staircasing effect in neighborhood filters and its solution. *IEEE Transactions on Image Processing* **15**(6) (2006) 1499–1505
- [65] Scherzer, O.: Denoising with higher order derivatives of bounded variation and an application to parameter estimation. *Computing* **60** (1998) 1–27
- [66] Didas, S., Weickert, J., Burgeth, B.: Stability and local feature enhancement of higher order nonlinear diffusion filtering. In Kropatsch, W., Sablatnig, R., Hanbury, A., eds.: *Pattern Recognition*. Volume 3663 of *Lecture Notes in Computer Science*. Springer, Berlin (2005) 451–458
- [67] You, Y.L., Kaveh, M.: Fourth-order partial differential equations for noise removal. *IEEE Transactions on Image Processing* **9**(10) (2000) 1723–1730
- [68] Didas, S., Weickert, J., Burgeth, B.: Properties of higher order nonlinear diffusion filtering. *Journal of Mathematical Imaging and Vision* **35** (2009) 208–226

- [69] Rossi, J.D., Schönlieb, C.B.: Nonlocal higher order evolution equations. *Applicable Analysis* **89**(6) (2010) 949–960
- [70] Filippov, A.F.: *Differential Equations with Discontinuous Righthand Sides*. Kluwer, Dordrecht (1988)