PART I:

Geometric Image Processing: Regularization, Representation, and Locally Adaptive Derivatives.

R. Duits · E.J. Bekkers

Abstract We study multi-scale representations of images, so called scale space representations. Such scale representations are typically obtained via PDE-evolutions, where we primarily focus on the case where the PDE-evolution is a diffusion. We will consider linear and nonlinear (data-driven) diffusions and set important connections to popular variational image regularization/denoising techniques. We also address basic geometric image analysis techniques, that comes along with these scale space representations, such as scale selection, well-posed Gaussian derivative operators, local analysis, localized basis expansions, differential invariants and underlying locally adaptive frames of derivatives.

We distinguish between locally adaptive frames of the first kind and of the second kind. The locally adaptive frames of the first kind are computed by eigenvector decomposition of the structure tensor of an image, whereas locally adaptive frames of the second kind are computed by eigenvector decomposition of the Hessian of an image. Such locally adaptive frames provide tools for geometric reasoning (diffusions, enhancement, tracking, detection) that works as long as the local image-structure can be modeled by a single oriented frame per position.

The problem is that, at complex structures in images (e.g. crossings or junctions), multiple (non-orthogonal) orientations are present. Thereby the computation of differential frames of the 1st and 2nd order is not well-defined at these locations. Therefore, in later sections of the lecture notes we generalize the notion of gauge frames on images to gauge frames on data representations defined on the extended space of positions and orientations.

Learning objectives:

- Understand Gaussian scale space representations and master solving linear diffusion systems.
- Understand the notion of Gaussian derivatives.
- Understand basic image regularization techniques and know how they relate.
- Understand practical B-spline expansions of images.
- Master geometric and algebraic treatment of vector fields on manifolds, and know how they relate.
- Understand the optimization problems and Euler-Lagrange formalisms underlying locally adaptive frames, and know how to put them in a more general Lie group formalism.
- Understand possibilities and limitations of locally adaptive frames directly in the image domain.
- Understand the numerical methods for PDE's (nonlinear diffusions and total variation flows).
- Know how to apply variants of nonlinear images diffusions via prepared Mathematica notebooks.
- Understand variational techniques (total variation flow) for image regularization/denoising.

NB:
Technical parts and exercises with extra material are indicated in blue and may be skipped: The black text forms the core of the course and does not rely on the blue parts. Regular exercises are indicated in red.

Introduction

In this first part of the course, we will be concerned with basic techniques on mathematical (PDE-based) image processing. We will not presume pre-knowledge on image processing, but we will presume pre-knowledge on PDE's, analysis and linear algebra. We will explain the following, basic mathematical image analysis techniques:

- **Linear scale space representations**,
  with an emphasis on the Gaussian case in Section 1.
- **Gaussian derivatives**,
  and well-posed differentiation in Gaussian scale spaces in Section 2.
- **Scale selection**
  in Gaussian scale space representations in Section 3.
- **Separable implementation**
  of Gaussian scale space representations/implementations in Section 4.
- **Local analysis and basis expansions**
  of images in Section 5.
- **Differential invariants and nonlinear geometric filtering**
  of images in Section 6.
- **Locally adaptive frames**
  for data-driven derivatives in the image domain, as explained in detail in Section 7. To put this concept gradually in a more general context, we include two *intermezzos* with
  1. *preliminaries on Lie groups*,
  2. *tangent vectors and vector fields, where we explain both the geometric and the algebraic viewpoint*,

  which are both crucial for understanding the geometrical techniques later on in the lecture notes.
- **Diffusion algorithms for image-enhancement**
  in Section 8, with a high emphasis on finite difference schemes, where we distinguish between:
  1. Linear diffusion algorithms in Subsection 8.1.
  2. Nonlinear diffusion with adaptive scalar diffusivity in Subsection 8.2.
  3. Nonlinear diffusion steered by locally adaptive frames in Subsection 8.3.
- **Variational Techniques for image-denoising** in Section 9.

In this first part the geometric processing will only take place in the actual image domain. This image domain is regarded as connected and compact subset of $\mathbb{R}^d$ (typically a rectangle if $d = 2$ or a cube if $d = 3$). In all later parts of the course we shall extend the image domain from $\mathbb{R}^d$ to higher dimensional spaces (Lie groups and Lie group quotients), via lifts (e.g. orientation scores, continuous wavelet transforms or Gabor transforms). As such extensions require more advanced geometrical and PDE-techniques, we will sometimes put the geometric processing in the actual image domain of this section, in a bit more formal context than strictly required. This helps to identify the geometric tools underlying geometric image processing, and provides a basis for better understanding of the more technical generalizations in later parts.

1 Gaussian Scale Space Representations

The Gaussian kernel $G_s : \mathbb{R}^d \to \mathbb{R}^+$ given by

$$G_s(\mathbf{x}) = (4\pi s)^{-d/2} e^{-\frac{\|\mathbf{x}\|^2}{4s}}, \tag{1}$$

can be used to construct a Gaussian scale space representation of a square integrable image $f : \mathbb{R}^d \to \mathbb{R}$:

$$u(\mathbf{x}, s) = (G_s * f)(\mathbf{x}) := \int_{\mathbb{R}^d} G_s(\mathbf{x} - \mathbf{y})\, f(\mathbf{y})\, \mathrm{d}\mathbf{y}. \tag{2}$$

In general a greyscale image f assigns to $(x, y) \in \mathbb{R}^2$ a greyvalue $f(x, y) \in \mathbb{R}$, whereas a scake space representation assigns to each location (x, y) and scale $s > 0$ a greyvalue $u(x, y, s)$. Here $s > 0$ is to be considered at the scale at which the image f is observed. See Figure 1. Typically, one expresses the scale of observation $s = \frac{1}{2}\sigma^2$ in the standard deviation σ of a Gaussian kernel.

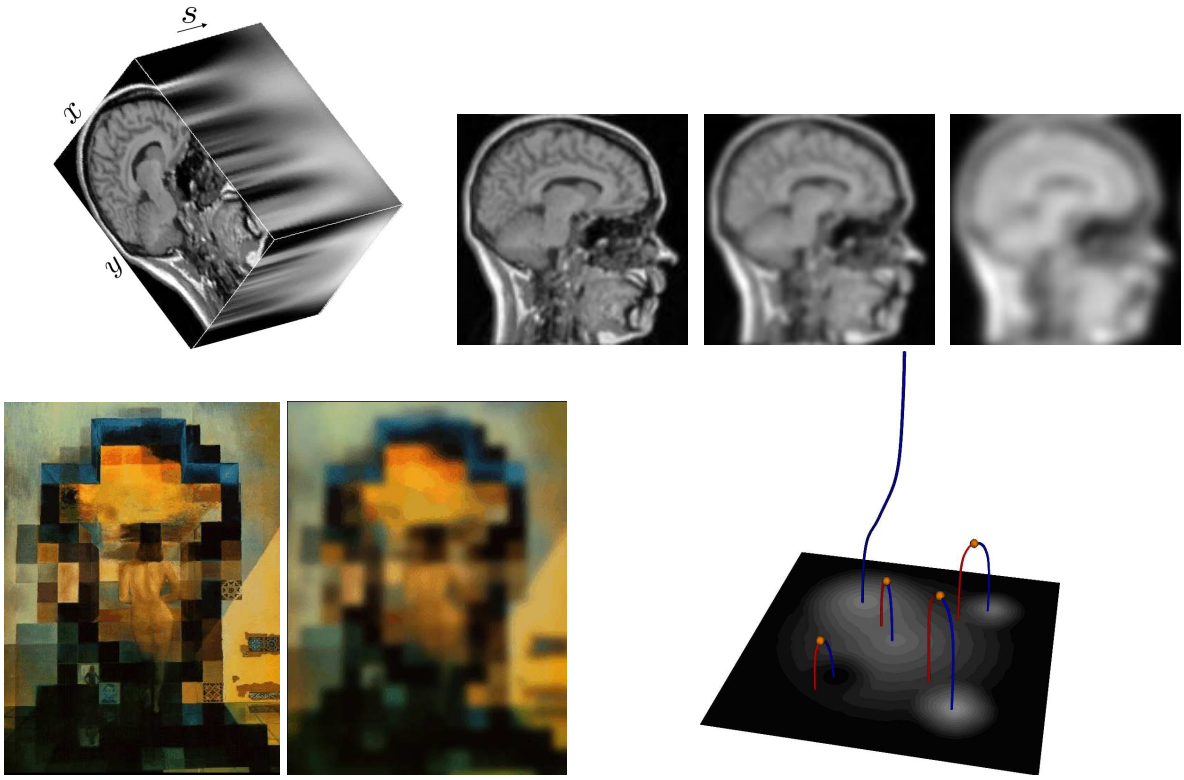

Fig. 1 Top row; left: illustration of a scale space representation $u(x, y, s)$ of a 2D image $f(x, y)$, right: fixed scale layers $u(\cdot, s)$ for increasing $s > 0$; left: two slices (small scale and large scale) in a scale space representation of a painting of Salvador Dali. right: the critical paths of a simple image consisting of only a few blobs, structures vanish as $s > 0$ increases extrema and saddles meet each other at singular points (where up to 2nd order differential structure allows for a reasonable approximative image reconstruction [1–3]).

The above Gaussian scale space representation $u(\mathbf{x}, s)$ satisfies a familiar evolution equation involving the heat-equation

$$\begin{cases} \frac{\partial u}{\partial s}(\mathbf{x}, s) = \Delta u\,(\mathbf{x}, s), & \mathbf{x} \in \mathbb{R}^d, s > 0, \\ u(\mathbf{x}, 0) \quad = f(\mathbf{x}), & \mathbf{x} \in \mathbb{R}^d, f \in \mathbb{L}_2(\mathbb{R}^d), \end{cases} \tag{3}$$

with Laplacian $\Delta = \nabla \cdot \nabla$ which on $\mathbb{R}^d$ can be expressed as $\left(\frac{\partial}{\partial x^1}\right)^2 + \ldots + \left(\frac{\partial}{\partial x^d}\right)^2$ in Euclidean coordinates $\mathbf{x} = (x^1, \ldots x^d) \in \mathbb{R}^d$. Intuitively, function $u(\cdot, s) \in \mathbb{L}_2(\mathbb{R}^d)$ for a fixed $s > 0$ represents the image observed at scale $s = \frac{1}{2}\sigma^2$. A paintings of Dali in Fig. 1 reveals how important this choice of scale of observation can be. At a fine scale we recognize a body, at a larger scale we recognize the portrait of Abraham Lincoln.

Let us recall that Fourier transform $\mathcal{F} : \mathbb{L}_2(\mathbb{R}^d) \to \mathbb{L}_2(\mathbb{R}^d)$ is given by

$$\mathcal{F}f(\boldsymbol{\omega}) = \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^d} f(\mathbf{x}) e^{-i\boldsymbol{\omega}\cdot\mathbf{x}}\, \mathrm{d}\mathbf{x}, \tag{4}$$

for almost every $\boldsymbol{\omega} \in \mathbb{R}^d$. It is a unitary operator $\mathcal{F}^{-1} = \mathcal{F}^*$ and thereby

$$\|\mathcal{F}f\|^2_{\mathbb{L}_2(\mathbb{R}^d)} = \|f\|^2_{\mathbb{L}_2(\mathbb{R}^d)} \text{ for all } f \in \mathbb{L}_2(\mathbb{R}^d). \tag{5}$$

Furthermore, one has the property that a convolution of two elements $f, g \in \mathbb{L}_1(\mathbb{R}^d) \cap \mathbb{L}_2(\mathbb{R}^d)$ in the spatial domain correspond to a direct product of functions in the Fourier domain:

$$\mathcal{F}(f * g) = (2\pi)^{\frac{d}{2}} \mathcal{F}f\, \mathcal{F}g. \tag{6}$$

Furthermore recall that for sufficiently smooth $f \in \mathbb{L}_2(\mathbb{R}^d)$ one has

$$\mathcal{F}\left(\left(\frac{\partial}{\partial x^k}\right)^n f\right)(\boldsymbol{\omega}) = (i\omega^k)^n \ (\mathcal{F}f)(\boldsymbol{\omega}) \text{ for all } k = 1, \dots, d, \tag{7}$$

where the *label* k for contravariant components of the vectors $\mathbf{x} = (x^1, \dots, x^d) \in \mathbb{R}^d$ and covariant components of the vector[1] $\boldsymbol{\omega} = (\omega^1, \dots, \omega^d)$.

Finally, it can be shown (by technical computations and contour-integration in the complex plane) that

$$\mathcal{F}G_s(\boldsymbol{\omega}) = (2\pi)^{-\frac{d}{2}} \ e^{-s\|\boldsymbol{\omega}\|^2}. \tag{8}$$

The identities (5), (7), and (8) can be used without proof throughout this course.

Exercise 1 Derive the solution (2) to the heat evolution (3), using (5), (7), and (8).

Remark 1 **(other linear scale space representations)**
Gaussian scale space representations have been first proposed in image processing by Ijima [4,5]. From a list of scale space axioms (translation-covariance, rotation-covariance, linearity, positivity preservation, semi-group property, average grey-value invariance, scaling property) one can deduce so-called a parameterized class of α-scale space representations. For an overview see [6]. Here $\alpha \in (0, 1]$ bridges the Gaussian scale space representations $\alpha = 1$ to the Poisson scale space representation $\alpha = \frac{1}{2}$.

Poisson scale spaces are harmonic functions with respect to the joined scale space variable $(\mathbf{x}, s)$ and they boil down to harmonic extension of the original image at $s = 0$ to the upper-space $s \geq 0$, as you need to show in Exercise 2a below. They have benefits for phase-based image processing as one can rely on (Clifford)-analytic extensions [7–9]). Such α-scale space representations [10,6,11] are generated by a fractional power $-(-\Delta)^\alpha$ of the minus Laplacian (with $0 < \alpha \leq 1$). They are well-known in probability theory as they are forward Kolmogorov equations of α-stable Lévy processes [12].

If one imposes a strong causality constraint (the non-enhancement of local extrema) [13,14] or a (questionable) finite variance assumption [12], the allowable class of the linear scale spaces is reduced to the Gaussian scale space representation.

Exercise 2 Consider the factorization

$$\Delta + \partial_s^2 = (-\sqrt{-\Delta} + \partial_s)(\sqrt{-\Delta} + \partial_s),$$

where the unbounded pseudo-differential operator $-\sqrt{-\Delta}$ is given by

$$-\sqrt{-\Delta} f = \mathcal{F}^{-1}(\boldsymbol{\omega} \mapsto -\|\boldsymbol{\omega}\| \mathcal{F}f(\boldsymbol{\omega})) \text{ for all } f \in \mathbb{H}^1(\mathbb{R}^d).$$

with $\mathbb{H}_1(\mathbb{R}^d) \subset \mathbb{L}_2(\mathbb{R}^d)$ the first order Sobolev space.

a.) Show that the following systems (where we require $u(\cdot, s) \in \mathbb{L}_2(\mathbb{R}^d)$ for all $s \geq 0$) produce the same solutions $u(\mathbf{x}, s)$:

$$\begin{cases} \partial_s u(\mathbf{x}, s) = -(\sqrt{-\Delta}\, u)(\mathbf{x}, s) & \text{for all } (\mathbf{x}, s) \in \mathbb{R}^d \times \mathbb{R}^+, \\ u(\mathbf{x}, 0) = f(\mathbf{x}) & \text{for allmost every } \mathbf{x} \in \mathbb{R}^d, \end{cases}$$

$$\Leftrightarrow$$

$$\begin{cases} \Delta_{\mathbf{x},s} u(\mathbf{x}, s) := (\partial_s^2 + \Delta) u(\mathbf{x}, s) = 0 & \text{for all } (\mathbf{x}, s) \in \mathbb{R}^d \times \mathbb{R}^+, \\ u(\mathbf{x}, 0) = f(\mathbf{x}) & \text{for allmost every } \mathbf{x} \in \mathbb{R}^d. \end{cases}$$

b.) Their solution u is given in explicit form by a convolution kernel $u(\cdot, s) = H_s * f$ with kernel

$$H_s(\mathbf{x}) = \frac{2}{\sigma_d} \frac{s}{(s^2 + \|\mathbf{x}\|^2)^{\frac{d+1}{2}}}.$$

where $\sigma_{d-1} = 2\, \pi^{\frac{d}{2}} \Gamma\left(\frac{d}{2}\right)$ denotes the surface area of the $d-1$-dimensional sphere embedded in $\mathbb{R}^d$, so $\sigma_1 = 2\pi$ and $\sigma_2 = 4\pi$. You do not need to derive this for general $d \in \mathbb{N}$.

Explicitly derive this convolution kernel yourself from the PDE-systems only for the special case $d = 1$.

[1] We use inner products in our definition of the Fourier transform (4) which requires $\boldsymbol{\omega}$ to be a vector. Sometimes it is worthwhile to consider $\boldsymbol{\omega}$ rather as a co-vector (linear functional) but this is not required here as $(\mathbb{R}^d)^* \equiv \mathbb{R}^d$.

c1.) Show that

$$G_t * G_s = G_{t+s} \text{ and } H_t * H_s = H_{t+s}$$

for all $t, s \geq 0$.

c2.) Show that

$$\int_{\mathbb{R}^d} G_s(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1 \text{ and } \int_{\mathbb{R}^d} H_s(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1.$$

hint: zero frequency in the Fourier domain.
What does this mean for the average grey-value of the Poisson and Gaussian scale space representation as scale increases?

d.) Apply the Gauss divergence theorem to show that the flow-operator in a Gaussian scale space representation is given by the gradient

$$(\mathbf{x}, s) \mapsto \nabla u(\mathbf{x}, s) = (\nabla G_s * f)(\mathbf{x}).$$

hint: $\Delta = \nabla \cdot \nabla = div \circ grad$.

e.) Apply the Gauss divergence theorem to show that the flow-operator in a Poisson scale space representation is given by minus the Riesz-transform

$$(\mathbf{x}, s) \mapsto \mathbf{R}u(\mathbf{x}, s) = (\mathbf{R}H_s * f)(\mathbf{x}),$$

where $\mathbf{R}$ denotes the Riesz transform given by $\mathbf{R}f = \mathcal{F}^{-1}(\boldsymbol{\omega} \mapsto i\frac{\boldsymbol{\omega}}{\|\boldsymbol{\omega}|}\mathcal{F}f(\boldsymbol{\omega}))$.
hint: $-\sqrt{-\Delta} = \nabla \cdot (-\mathbf{R})$.

f.) Show that the flow-kernels in a Poisson scale space representation are given by

$$\begin{aligned} \nabla G_s(\mathbf{x}) &= -\tfrac{\mathbf{x}}{2s} G_s(\mathbf{x}), \\ -\mathbf{R}H_s(\mathbf{x}) &= -\tfrac{\mathbf{x}}{s} H_s(\mathbf{x}), \end{aligned}$$

g.) Show that extrema cannot be enhanced in a Gaussian scale space representation. I.e. show that if $u(\cdot, s_i)$ has an (isolated) maximum in $\mathbf{x} = \mathbf{x}_i$ then $\partial_s u(\mathbf{x}_i, s_i) < 0$ and $u(\cdot, s_i)$ has an (isolated) minimum in $\mathbf{x} = \mathbf{x}_i$ then $\partial_s u(\mathbf{x}_i, s_i) > 0$.

hint: In a Gaussian scale space representation one has $(\partial_s u(\mathbf{x}, s))(\Delta u(\mathbf{x}, s)) \geq 0$ every $(\mathbf{x}, s) \in \mathbb{R}^d \times \mathbb{R}^+$ and $\Delta u(\mathbf{x}, s_i) = trace\{Hu(\mathbf{x}, s_i)\}$, where $Hu(\mathbf{x}, s_i)$ denotes the (usual) Hessian matrix on $\mathbb{R}^d$, see (9).

Exercise 3 As scale increases in a scale space representation we observe that structures can disappear. See for example the critical paths in the lower right-corner of Fig. 1.

Such topological transitions can be mathematically classified by local Morse theory [15] which gives equivalence classes ('germs') of unfoldings $u(x, y, s)$ that satisfy the diffusion equation. Such germs can be represented by local polynomials. In particular, generic topological transitions where bifurcations take place in the critical paths in scale space, are represented[2] by the polynomials u_1 and u_2 defined below:

$$\begin{aligned} u_1(x_1, \ldots, x_d, s) &:= x_1^3 - 6x_1(x_2^2 + s) + Q(x_2, \ldots, x_d, s), \\ u_2(x_1, \ldots, x_d, s) &:= x_1^3 + 6x_1 s + Q(x_2, \ldots, x_d, s). \end{aligned}$$

with quadratic forms $Q(x_2, \ldots, x_d) = \sum_{k=2}^{d} \epsilon_k(x_k^2 + s)$, with $\epsilon_k \in \{-1, 1\}$ for all $k = 2 \ldots, d$.

a.) Verify that u_1 satisfies the diffusion equation.

b.) Study what happens to the critical paths (where $\nabla_{\mathbf{x}} u_1(\cdot, s)$ vanishes) for $s < 0$, $s = 0$ and $s > 0$ and classify their nature (i.e. is it a saddle, an extremum or a singular point).

c.) Verify that u_2 satisfies the diffusion equation.

d.) Study what happens to the critical paths (where $\nabla_{\mathbf{x}} u_2(\cdot, s)$ vanishes) for $s < 0$, $s = 0$ and $s > 0$ and classify their nature (i.e. is it a saddle, an extremum or a singular point).

e.) Which of the two cases above corresponds to what we see in Fig. 1?

[2] Two unfoldings are equivalent if they lie in the same orbit under actions of diffeomorphishms, see [15]. For a definition of the 'germs' and the 'one-sided stable germs' u_1, u_2, and how they follow from the 'cusp-catastrophe' see [16, ch:3.3.1].

f.) Set $d = 2$ and let u be a scale space representation of image f. Motivate that at the singular-points $(\mathbf{x}^*, s^*) \in \mathbb{R}^2 \times \mathbb{R}^+$ one has

$$\nabla u(\mathbf{x}^*, s^*) = \mathbf{0} \text{ and } \det\left(Hu(\mathbf{x}^*, s^*)\right) = 0.$$

with Hu the (spatial) Hessian of u with matrix representation

$$Hu(x, y, s) = \begin{pmatrix} \partial_x^2 u(x, y, s) & \partial_x \partial_y u(x, y, s) \\ \partial_x \partial_y u(x, y, s) & \partial_y^2 u(x, y, s) \end{pmatrix}. \tag{9}$$

1.1 Gaussian Scale Space on the Bounded Domain

We typically model the space of images as $\mathbb{L}_2(\mathbb{R}^d)$. Now in practice such images are always given on a compactly support domain (typically a rectangle or a cube) $\Omega \subset \mathbb{R}^d$. By extending such an image with 0 outside the domain Ω, i.e by setting

$$f(\mathbf{x}) := \begin{cases} \text{the greyvalue at location } \mathbf{x} & \text{if } \mathbf{x} \in \Omega, \\ 0 & \text{else}, \end{cases}$$

one considerably affects the boundaries in a scale space representation.

Alternatively, one can consider the diffusion equation on a bounded domain with boundary conditions (e.g. Neumann or Dirichlet boundary conditions). From the practical point of view this is still somewhat artificial since as the shape of $\partial\Omega$ (which in principle has nothing to do with content in the image) still affects[3] the scale space representation.

Nevertheless, for scale spaces on the bounded domain the choice of Neumann boundary conditions is reasonable, both in view of increase of entropy and in view of maintainance of the average grey-value.

Exercise 4 (scale space representation on a bounded domain)

Consider the following evolution system for a scale space representation on the rectangle $\Omega := [0, d_1] \times [0, d_2]$, with $(d_1, d_2 > 0)$:

$$\begin{cases} \frac{\partial u}{\partial s}(x, y, s) = \Delta u(x, y, s) = \frac{\partial^2 u}{\partial x^2}(x, y, s) + \frac{\partial^2 u}{\partial y^2}(x, y, s), \\ \text{for } \ (x, y) \in [0, d_1] \times [0, d_2], s \geq 0 \ , \\ \frac{\partial u}{\partial x}(0, y, s) = \frac{\partial u}{\partial x}(d_1, y, s) = 0 \text{ for } y \in [0, d_2], \\ \frac{\partial u}{\partial y}(x, 0, s) = \frac{\partial u}{\partial y}(x, d_2, s) = 0 \ \text{ for } x \in [0, d_1], \\ u(x, y, 0) = f(x, y) \qquad \text{for } x \in [0, d_1], y \in [0, d_2], \end{cases} \tag{10}$$

where $f(x, y) \geq 0$ denotes the initial image at $s = 0$.

a.) Use the method of separation to explicitly express the unique solution of (10) in a Fourier series where $f \in \mathbb{L}_2(\Omega) \cap C^2(\Omega)$ satisfies the Neumann boundary conditions.

b.) Explicitly derive the unique solution of (10) in case the initial condition equals

$$f(x, y) = \cos^2\left(\frac{\pi x}{d_1}\right) \cos^2\left(\frac{2\pi y}{d_2}\right) \ ,$$

c.) Show that for any initial image $f \in \mathbb{L}_2(\Omega) \cap C^2(\Omega)$ satisfying the Neumann boundary conditions, the average greyvalue on the rectangle $[0, d_1] \times [0, d_2]$ is preserved over scale, i.e.

$$\frac{1}{d_1 d_2} \int_\Omega u(\mathbf{x}, s) \, \mathrm{d}\mathbf{x} = \int_\Omega f(\mathbf{x}) \cdot 1 \, \mathrm{d}\mathbf{x} \text{ for all } s \geq 0.$$

[3] See for example [16, Fig. 2.6]

d.) Show that for any initial image $f \in \mathbb{L}_2(\Omega) \cap C^2(\Omega)$ with $0 < f \leq 1$ satisfying the Neumann boundary conditions, the entropy increases, i.e.

$$\frac{d}{ds}\mathcal{E}(s) \leq 0$$

for $\mathcal{E}(s) = -\int_\Omega u(\mathbf{x}, s) \log u(\mathbf{x}, s)\, \mathrm{d}\mathbf{x}$.

e.) Re-do the above exercises 4a–4c, when the Gaussian scale space generator Δ in (10) is replaced by the alpha-scale space generator $-(-\Delta)^\alpha$ with $\alpha \in (0, 1]$.

In the remainder of this section we again consider images as compactly supported elements in $\mathbb{L}_2(\mathbb{R}^d)$, unless explicitly stated otherwise.

2 Gaussian Derivatives in Gaussian Scale Space

One of the key advantages of using a scale space representation is that for all $f \in \mathbb{L}_2(\mathbb{R}^d)$ and all $s > 0$ the function $u(\cdot, s)$ which maps $\mathbf{x}$ onto $u(\mathbf{x}, s)$ is infinitely differentiable, and one has put derivatives onto the Gaussian convolution kernel:

$$\frac{\partial^k}{\partial x^k}(G_s * f) = \left(\frac{\partial^k}{\partial x^k} G_s\right) * f$$

We will denote the Gaussian gradient is denoted by

$$\nabla^s f := \nabla G_s * f. \tag{11}$$

Gaussian partial derivatives w.t.r. Euclidean coordinates $(x^1, \ldots, x^d)$ in $\mathbb{R}^d$ at scale $s > 0$ will be denoted briefly by $\partial^s_{x^i}$ where $\partial^s_{x^i} f = \frac{\partial}{\partial x^i} G_s * f$.

One of the key-advantages of Gaussian derivatives over normal derivatives is that they are continuous operators on $\mathbb{L}_2(\mathbb{R}^d)$. Recall a standard result from functional analysis:

Lemma 1 *Let $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$ be a normed spaces. Let $A : X \to Y$ be linear. Then A is continuous if and only if it is bounded (i.e. $\|A\| = \sup_{f \in X, f \neq 0} \frac{\|Af\|_Y}{\|f\|_X} < \infty$).*

Exercise 5 Prove the above lemma.

Exercise 6 Let S be the Schwartz space of infinitely differentiable, rapidly decaying functions on $\mathbb{R}$, i.e.

$$S = \{f \in C^\infty(\mathbb{R}) \mid \forall_{N \in \mathbb{N}} \sup_{|\alpha| \leq N} \sup_{x \in \mathbb{R}} (1 + |x|^2)^N |(D^\alpha f)(x)| < \infty\}$$

with $D^\alpha = \left(\frac{d}{dx}\right)^\alpha$. We equip S with the $\mathbb{L}_2(\mathbb{R})$-norm. You may use that Fourier transform $\mathcal{F}$

- is a unitary operator from $\mathbb{L}_2(\mathbb{R}) \to \mathbb{L}_2(\mathbb{R})$
- maps S into S,
- satisfies $\mathcal{F}_\mathbb{R}(f * g) = \sqrt{2\pi}\, \mathcal{F}_\mathbb{R} f \cdot \mathcal{F}_\mathbb{R} g$, where $*$ represents convolution.

a.) Let $\alpha = 1$. Show that the mapping $S \ni f \mapsto Df \in S$ is *unbounded* on S,
(hint: $x \mapsto e^{-n\,x^2} \in S$ for all $n \in \mathbb{N}$.)

b.) Let $s > 0, \alpha \in \mathbb{N}$. Let $G_s(x)$ denote the 1D Gaussian kernel (1) for $d = 1$ whose Fourier transform is given by (8). Show that the operator $D^{\alpha,s} : \mathbb{L}_2(\mathbb{R}) \to \mathbb{L}_2(\mathbb{R})$ given by

$$D^{\alpha,s}(f) := (D^\alpha G_s) * f, \qquad f \in \mathbb{L}_2(\mathbb{R}), \tag{12}$$

is a *bounded* linear operator.
(hint: First show that $D^{1,s}$ is bounded and then the other cases $\alpha > 1$ follow by $D^{\alpha,s} = (D^{1,\frac{s}{\alpha}})^\alpha$)

c.) Compute the operator norm of $D^{1,s}$, i.e. compute

$$\|D^{1,s}\| = \sup_{f \in S, f \neq 0} \frac{\|D^{1,s} f\|_{\mathbb{L}_2(\mathbb{R})}}{\|f\|_{\mathbb{L}_2(\mathbb{R})}}$$

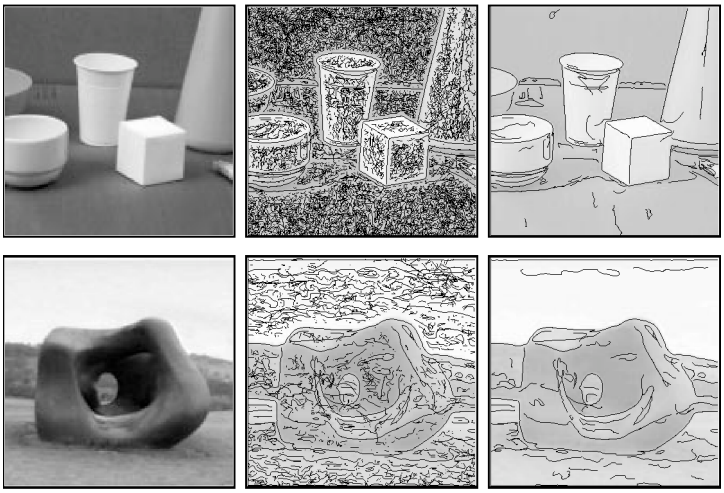

Fig. 2 Figure from an early paper [17] by Tony Lindeberg on the scale selection procedure he proposed. He showed how the scale selection procedure for Gaussian derivatives can be used to extract the relevant edges in images. Left: original image. Middle: all edges in scale space (obtained via a zero-crossings algorithm [17]). Right: the 100 most significant scale space edges according to a significance measure that integrates $\max_{s>0} \sqrt{s}\|\nabla^s f(\cdot)\|$ along each of the edge lines.

3 Scale Selection Principle

A well-established scale selection (i.e. selection of $s > 0$) by Tony Lindeberg [14] is to first express detectors in dimensionless coordinates $X = x/\sqrt{s}$ and $Y = y/\sqrt{s}$ and then optimize over scale s afterwards. For example, the Gaussian Laplacian

$$\Delta^s f = (\partial_x^s)^2 f + (\partial_y^s)^2 f, \tag{13}$$

is a common "blob-detector". Blobs in images f can be detected by a scale derivative in their scale space representation $u(\mathbf{x}, s)$ via the Laplacian since $\Delta u(\mathbf{x}, s) = \frac{\partial}{\partial s} u(\mathbf{x}, s) = (\Delta G_s * f)(\mathbf{x})$. The next exercise shows an analytical example that supports this simple, very general and widely applicable scale selection principle.

Exercise 7 Consider (analytic) 2D image that equals a Gaussian kernel with scale $s_0 > 0$ centered around the origin $(x, y) = (0, 0)$ given by $f(x, y) = G_{s_0}(x, y)$, given by (1) for $d = 2$.

a.) Show that $G_{s_1} * G_{s_2} = G_{s_1+s_2}$ for all $s_1, s_2 > 0$.
b.) Compute

$$\frac{\partial}{\partial s} \left(s\, \Delta_{x,y} G_{s+s_0}(x_0, y_0) \right),$$

with the Laplacian $\Delta_{x,y} = \left(\frac{\partial}{\partial x}\right)^2 + \left(\frac{\partial}{\partial y}\right)^2$ expressed in the regular Euclidean coordinates (x, y) with physical dimension: Length.
c.) Apply the scale selection procedure at position $(x_0, y_0) \in \mathbb{R}^2$ and observe that at $(x_0, y_0) = (0, 0)$ one has that the optimal scale $s > 0$ is indeed equal to $s = s_0$.

In general the choice of scale $s = \frac{1}{2}\sigma^2$ is done in such a way that $\sigma > 0$ is chosen such that width of structures present in the images matches the width of the corresponding Gaussian derivatives.

Once a scale $s = s(x_0, y_0) > 0$ is selected for a position of interest (x_0, y_0) in the image, the corresponding outcome value of the optimization gives a significance for the feature of interest. For example, for edges one applies the scale selection procedure on the Gaussian gradient, and the resulting value $\max_{s>0} \sqrt{s}\|\nabla^s f(x_0, y_0)\|$ gives an edge significance, see Figure 2.

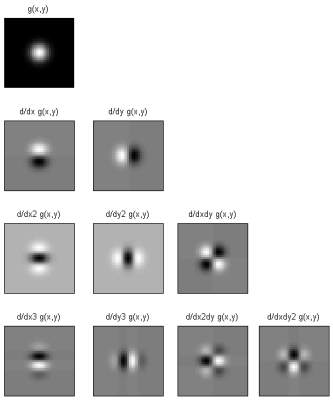

Fig. 3 Equidistantly sampled Gaussian derivatives $\frac{\partial^i}{\partial x^i}\frac{\partial^j}{\partial y^j} g(x,y) = G_s(x,y)$ w.r.t. Cartesian coordinates (x,y) for scale $s = \frac{1}{2}\sigma^2$ with $\sigma \approx 3$ pixels and with differential orders up to $0 \leq i+j \leq 3$.

4 Separability

A practical (though coordinate-dependent) property of a Gaussian kernel is its separability in Euclidean coordinates:

$$G_s^{2D}(x,y) = G_s^{1D}(x)\, G_s^{1D}(y),$$

where we label the Gaussians by 'dD' to stress that they are defined on $\mathbb{R}^d$. It allows one to efficiently compute a Gaussian derivative of order n_1 at scale s_1 in x-direction and a Gaussian derivative of order n_2 by scale s_2 in y-direction by two subsequent 1D convolutions instead of a 2D convolution:

$$\begin{aligned}(D^{(n_1,n_2)}_{(s_1,s_2)} * f)(x,y) &= (D^{(n_1,n_2)} G^{2D}_{(s_1,s_2)} * f)(x,y) = (D^{n_2}_{s_2} \circ D^{n_1}_{s_1} f)(x,y) \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} G^{(n_1)}_{s_1}(x-\tilde{x})\, f(\tilde{x},\tilde{y})\, \mathrm{d}\tilde{x} \right) G^{(n_2)}_{s_2}(y-\tilde{y})\, \mathrm{d}\tilde{y}. \end{aligned} \tag{14}$$

The next exercise shows that this property is rather special.

Exercise 8

a.) Show that a differentiable function $f : \mathbb{R}^2 \to \mathbb{R}$ is isotropic if and only if

$$((x\partial_y - y\partial_x) f)(x,y) = 0 \text{ for all } (x,y) \in \mathbb{R}^2.$$

hint: switch to polar coordinates.

b.) Show that a differentiable function $G : \mathbb{R}^2 \to \mathbb{R}$ which is both isotropic and separable with $G(x,y) = \frac{G(x,0)G(0,y)}{G(0,0)}$ must be a scalar multiple of a Gaussian kernel.

Remark 2 In practice (14) is computed via a Riemann sum approximation with equidistant sampling on the rectangular spatial image domain. When $s_1 = \frac{1}{2}\sigma_1^2, s_2 = \frac{1}{2}\sigma_2^2 > 0$ are large (i.e. σ_i is more than a few pixel sizes and proportional to a quarter of the image dimensions) then one may use (7) and sample in the Fourier domain and apply an inverse Discrete Fourier Transform instead.

5 Local Analysis and Expansions: Gaussian Derivatives, Hermite Functions and B-splines

The Gaussian Derivatives are listed for several orders in Figure 3. There we observe that an n-th order derivative of a Gaussian yields n zero crossings, this is due to the scaled n-th order Hermite polynomial that appears in the derivatives as can be seen in the next exercise. The Hermite polynomial of order $n \in \mathbb{N} \cup \{0\}$ is given by

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2} = H_n(x) = n! \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} \frac{(-1)^k (2x)^{n-2k}}{k!(n-2k)!}.$$

The Hermite polynomials $\{H_n\}_{n=0}^{\infty}$ are orthogonal w.r.t. a Gaussian weightfunction e^{-x^2} and the Hermite basis $\left\{x \mapsto \eta_n(x) := \frac{1}{\sqrt{\sqrt{\pi}2^n n!}} H_n(x) e^{-\frac{x^2}{2}}\right\}_{n=0}^{\infty}$ is a complete orthonormal basis for $\mathbb{L}_2(\mathbb{R})$. However, one must be aware that the Gaussian derivatives given by $\{G_s^{(n)}(x) := \frac{d^n}{dx^n} G_s(x)\}$ are *not* orthogonal w.r.t. the $\mathbb{L}_2(\mathbb{R})$-norm. In fact their correlation equals:

$$r_{nm} := \frac{\int_{\mathbb{R}} G_s^{(n)}(x) G_s^{(m)}(x) \mathrm{d}x}{\int_{\mathbb{R}} |G_s^{(n)}(x)|^2 \mathrm{d}x \int_{\mathbb{R}} |G_s^{(m)}(x)|^2 \mathrm{d}x} = \begin{cases} 0 & \text{if } n-m \text{ odd} \\ \frac{(-1)^{\frac{n-m}{2}} \Gamma\left(\frac{m+n+1}{2}\right)}{\Gamma\left(\frac{2n+1}{2}\right)\Gamma\left(\frac{2m+1}{2}\right)} & \text{if } n-m \text{ even} \end{cases} \tag{15}$$

with $n \geq m$, and with $\Gamma(z)$ the well-known Gamma-function. This makes the Gaussian derivatives (in contrast to the Hermite functions) unsuited for expansions up to high order derivatives. See also Fig. 4.

Exercise 9 Let us consider the 2D Gaussian given by (1) for $d = 2$.

a.) Show that the Gaussian derivatives are given by

$$\frac{d}{dx^n}\frac{d}{dy^m} G_s(x,y) = (-1)^{n+m} (\sqrt{4s})^{n+m} H_n\left(\frac{x}{\sqrt{4s}}\right) H_m\left(\frac{y}{\sqrt{4s}}\right) G_s(x,y).$$

b.) The Hermite functions form an orthonormal basis for $\mathbb{L}_2(\mathbb{R})$ and thereby one has for all $f \in \mathbb{L}_2(\mathbb{R}^2)$:

$$f = \sum_{n=0}^{\infty} (\eta_n, f)_{\mathbb{L}_2(\mathbb{R})}\ \eta_n. \tag{16}$$

What changes to such an expansion when using the basis of Gaussian derivatives?
hint: 'reciprocal basis'.

c.) Motivate both from (15) and Fig. 4 that the computation of the reciprocal basis of the Gaussian derivatives up to high order $n, m \leq N$ (with $N \gg 0$) is relatively ill-posed.

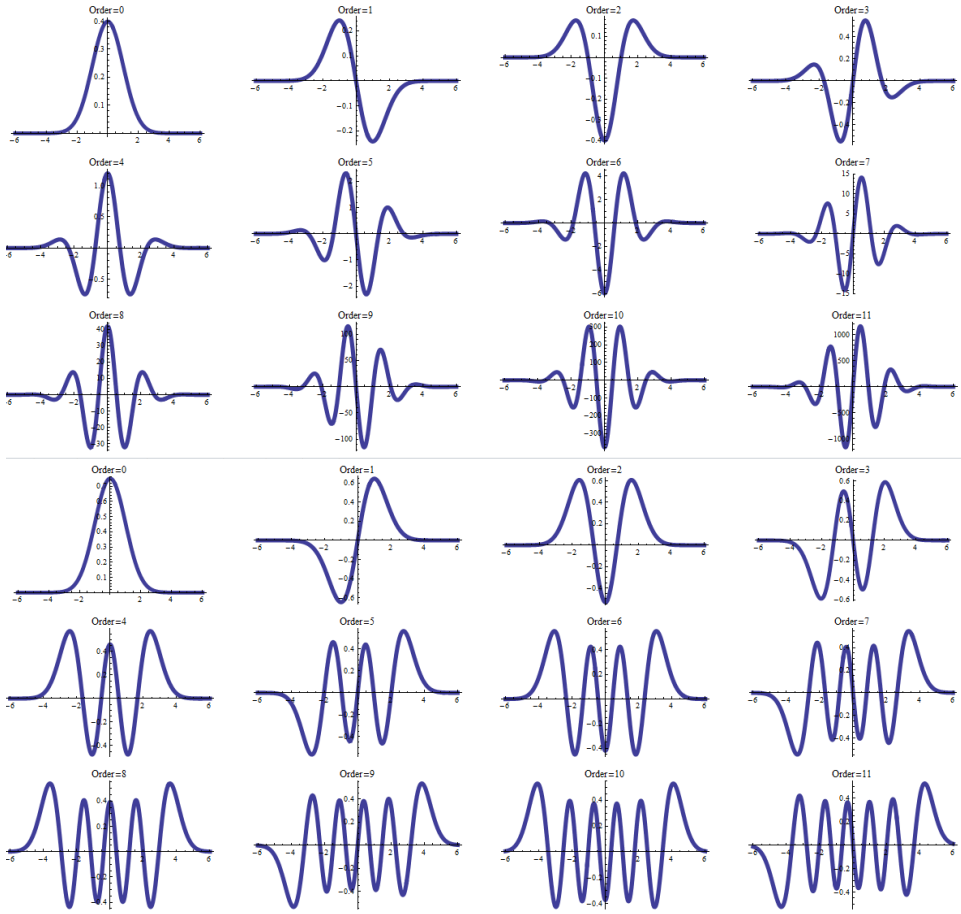

Fig. 4 Top: The Gaussian derivatives up to order 11 for $\sigma = 1$. Bottom: The Hermite functions up to order 11.

Intermezzo: Practical Expansions in a Simple Localized Basis: B-splines

Even though the Hermite functions allow for Fourier series expansions (16), they are oscillatory and not nicely localized which can lead to Gibbs oscilations when series are truncated (in a hard manner).

Furthermore, images are typically given on a discrete grid rather than that they are given as continuous functions in $\mathbb{L}_2(\mathbb{R}^d)$ and numeric approximations of Fourier expansions (e.g. by sampling the series, and/or by approximating the $\mathbb{L}_2$-inner products) typically produces a mismatch (approximation errors) between continuous invertible transforms in theory and corresponding discrete transforms that can be inverted exactly on the grid.

Therefore in image analysis one often relies on a compactly supported B-spline basis (or a wavelet basis) to interpolate, represent and express images. Furthermore, it makes sense in a scale space representation to use less and scaled basis-functions to represent $u(\cdot, s)$ as $s > 0$ increases.

Let B_n denote the nth order B-spline given by the continuous convolution:

$$B_n(x) = (B_{n-1} * B_0)(x), \;\; B_0(x) = \begin{cases} 1 \text{ if } -\frac{1}{2} < x < \frac{1}{2} \\ 0 \text{ otherwise} \end{cases} . \tag{17}$$

This gives rise to a function B_n which is in $C^{n-1}(\mathbb{R})$ and consists of a 'piecewise gluing' of polynomials of degree n on intervals of length 1, and which is supported on $[-(n+1)/2, (n+1)/2]$. The B-spline has the

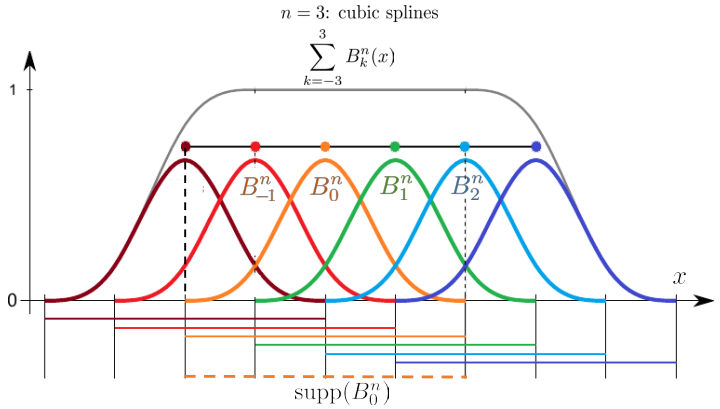

Fig. 5 Visualization of the cubic B-spline basis for $n = 3$.

following important properties:

$$\begin{aligned} &\int_{\mathbb{R}} B_n(x)\mathrm{d}x = 1, \ \sum_{k\in\mathbb{Z}} B_n(x-k) = 1, \\ &\text{with derivative: } B_n'(x) = B_n\left(x+\tfrac{1}{2}\right) - B_n\left(x-\tfrac{1}{2}\right) \text{ for } n \geq 2, \\ &\text{and 2-scale relation: } B_n\left(\tfrac{x}{2^j}\right) = \sum_{k\in\mathbb{Z}} 2^{-n}\binom{n+1}{k} B_n\left(\tfrac{x}{2^{j-1}} - k\right). \end{aligned} \tag{18}$$

Then we define the B-spline basis as:

$$B_{k,l}^n(x,y) = B^n\left(\frac{x}{s_k} - k\right) B^n\left(\frac{y}{s_l} - l\right), \text{ for all } (x,y) \in \mathbb{R}^2,$$

for $k = 1, \ldots N_k$ and $l = 1, \ldots N_l$, where $s_k > 0$ and $s_l > 0$ scale the B-splines and typically depend on the number N_k and N_l of B-spline basis functions. See Figure 5 for an illustration of the shifted B-spline basis in 1 dimension. Now let us enter the real world where images are given on a finite discrete grid. Let the discrete image be given on an equidistant grid where $f(\mathbf{x}_{ij})$ denotes the greyvalue at grid point $\mathbf{x}_{ij} = (i\Delta x, j\Delta y)$ where stepsizes are $\Delta x = \frac{1}{N_x}$, $\Delta y = \frac{1}{N_y}$ and where $i = 1, \ldots, N_x \in \mathbb{N}$, $j = 1, \ldots N_y \in \mathbb{N}$. Then, in the case of strict interpolation one would aim for coefficients α^{kl} such that:

$$f(\mathbf{x}_{ij}) = \sum_{k=1}^{N_k} \sum_{l=1}^{N_l} \alpha^{kl} B_{k,l}^n(\mathbf{x}_{ij}), \tag{19}$$

in which case $\sum_{k=1}^{N_k} \sum_{l=1}^{N_l} \alpha^{kl} B_{k,l}^n(\cdot)$ provides a continuous and $n-1$-times differentiable interpolation function. We store the greyvalues $f(\mathbf{x}_{ij})$ in a single vector $\mathbf{f} \in \mathbb{R}^{N_x N_y}$, and similarly we store the coefficients α^{kl} in a vector $\boldsymbol{\alpha} \in \mathbb{R}^{N_k N_l}$. Then we rewrite the linear equation (19) into vector form

$$\mathbf{f} = B\boldsymbol{\alpha} \ ,$$

with B the $N_x N_y \times N_k N_l$ matrix with entries $B_{k,l}^n(\mathbf{x}_{ij})$ yielding $N_k N_l$ linearly independent columns, and thereby (verify this) matrix $B^T B$ is invertible, Now typically, for a sparse basis representation we have

$$N_k < N_x \text{ and } N_l < N_y, \tag{20}$$

so the columns do not span the entire space $\mathbb{R}^{N_x N_y}$. Now from linear algebra, we know that this means that the corresponding linear map is injective but not surjective, and the inverse of the matrix B does not exist. Therefore, we must relax our too strict demand (19). That is we write

$$f^{int}(x,y) = \sum_{k=1}^{N-k} \sum_{l=1}^{N_l} \alpha^{kl} B_{k,l}^n(x,y), \tag{21}$$

and aim for the optimal coefficients as a solution

$$\boldsymbol{\alpha}_{\text{opt}} = (B^T B)^{-1} B^T \mathbf{f},$$

to a least-square problem, which is the unique solution of

$$\boldsymbol{\alpha}_{\text{opt}} := \underset{\boldsymbol{\alpha}\in\mathbb{R}^{N_k N_l}}{\arg\min} \|B\boldsymbol{\alpha} - \mathbf{f}\|^2 = (B^T B)^{-1} B^T \mathbf{f}. \tag{22}$$

Remark 3 There exist more efficient algorithms for (smoothed) B-spline interpolation via Z-transforms and cardinal splines [18], but this is beyond the scope of this course.

Now let us return to the non-discrete function space setting, likewise the rest of this section. Let V denote the $N_k N_l$-dimensional span of compactly supported B-spline basis functions:

$$V := \text{span}\{B^n_{k,l} \mid k = 1, \ldots, N_k, l = 1, \ldots, N_l\} \subset \mathbb{L}_2(\mathbb{R}^d). \tag{23}$$

We define the expansion operator $\mathcal{E} : \mathbb{R}^{N_k N_l} \to V$ by

$$(\mathcal{E}(\boldsymbol{\alpha}))(x, y) = \sum_{k=1}^{N_k} \sum_{l=1}^{N_l} \alpha^{kl} B^n_{k,l}(x, y).$$

Then $\alpha_{opt} = \mathcal{E}(\boldsymbol{\alpha}_{opt})$ is the minimizer of the corresponding problem in $V \subset \mathbb{L}_2(\mathbb{R}^2)$:

$$\alpha_{opt} = \underset{\alpha\in V}{\arg\min} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} |\alpha(\mathbf{x}_{ij}) - \mathbf{f}(\mathbf{x}_{ij})|^2.$$

Exercise 10 Prove that Eq. (22) indeed holds.

It is also common to include a regularization prior using a matrix $R \geq 0$ (that can be computed analytically [19, App.A]) such that

$$\int_{\mathbb{R}^2} \|\nabla\alpha(\mathbf{x})\|^2 \mathrm{d}\mathbf{x} = \boldsymbol{\alpha}^T R \boldsymbol{\alpha} \quad \text{for all } \alpha = \mathcal{E}(\boldsymbol{\alpha}) \in V.$$

Then we can study the problem of smoothed regression where one enforces regularity and $\mathbb{L}_2$-boundedness on the interpolating function:

$$\alpha_{opt} = \underset{\alpha\in V}{\arg\min} \left\{ \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} |\alpha(\mathbf{x}_{ij}) - \mathbf{f}(\mathbf{x}_{ij})|^2 + \lambda \int_{\mathbb{R}^2} \|\nabla\alpha(\mathbf{x})\|^2 \,\mathrm{d}\mathbf{x} + \mu \int_{\mathbb{R}^2} |\alpha(\mathbf{x})|^2 \,\mathrm{d}\mathbf{x}, \right\},$$

with a given $\lambda > 0$ which can be set manually or which can be adjusted automatically. Then in the finite discrete setting this boils down to

$$\boldsymbol{\alpha}_{opt} = \underset{\boldsymbol{\alpha}\in\mathbb{R}^{N_k N_l}}{\arg\min} \left\{ \|B\boldsymbol{\alpha} - \mathbf{f}\|^2 + \lambda\, \boldsymbol{\alpha}^T R \boldsymbol{\alpha} + \mu \boldsymbol{\alpha}^T \boldsymbol{\alpha} \right\} = (B^T B + \lambda R + \mu I)^{-1} B^T \mathbf{f}. \tag{24}$$

Remark 4 (automatic adjustment of the pair (λ, μ) via GCV).
In many applications, like in template matching and object recognition [20], parameter $\lambda > 0$ is set automatically by *Generalized Cross Validation* [4]. This means that $\lambda > 0$ can be set such that the average error that arises by leaving one of the data-constraints out is minimized. More formally, set influence matrix $A = A(\lambda, \mu) = B(B^T B + \lambda R + \mu I)^{-1} B^T$. which describes the effect of data-vector $\mathbf{f}$ on the data-term of the minimized cost, which is given by $\|\mathbf{f} - \hat{\mathbf{f}}\|^2 = \|(I - A)\mathbf{f}\|^2$ with data-vector estimator $\hat{\mathbf{f}}(\lambda, \mu) := B\boldsymbol{\alpha}_{opt} = A\mathbf{f}$. Let N denote the total amount of data-points (e.g. above we have $N = N_x N_y$). Then minimization of

$$GCV(\lambda, \mu) = \frac{1}{N} \frac{\|(I - A(\lambda, \mu))\mathbf{f}\|^2}{1 - N^{-1}\text{trace}(A(\lambda, \mu))} \approx CV(\lambda, \mu) = \frac{1}{N} \sum_{k=1}^{N} \frac{|f^k - \hat{f}^k(\lambda, \mu)|^2}{|1 - A_{kk}(\lambda, \mu)|^2}, \tag{25}$$

yields the optimal pair $(\lambda, \mu) \in \mathbb{R}^+ \times \mathbb{R}^+$.

[4] The words *"Cross Validation"* in GVC refers to the leave one out principle: how well can the model correct for the missing constraint. *"General"* refers to the rotation invariant estimator $N^{-1}\text{trace}(A)$ used in the approximation (**??**).

Exercise 11 Prove that the matrix $B^T B + \lambda R$ is invertible for all $\lambda \geq 0$ and that Eq. (24) indeed holds.

Exercise-Mathematica 1. *Consider the notebook "*Lecture1_BsplineExpansion.nb*" that can be downloaded from (*http://www.lieanalysis.nl/201-2/*) and apply this B-spline expansion to a photograph of your own*
(NB. Use the "Import" and "Resize" command in Mathematica if needed, press F1 for help index) while increasing $\lambda > 0$ and use it to interpolate a value at a point in between grid-points.

Remark 5 To represent a fixed scale layer (i.e. $u(\cdot, s) = (G_s * f)(\cdot)$ with $s > 0$ fixed) in a scale space representation (Fig. 1) one typically needs much less basis functions N_k and N_l as details are lost (as extrema and saddles 'eat each other' at 'top-points', recall Fig. 1). Furthermore the 2-scaling relation in (18) helps to build an efficient multi-scale representation in a pyramid. It can even be used for approximate image reconstructions from the up to 2nd order differential structure at top-points in scale space. This is beyond the scope of the course, but details can be found in [3, ch:4, ch:2].

6 Differential Invariants and Nonlinear Geometric Filtering

When applying a combination of Gaussian derivatives of an image $f : \mathbb{R}^2 \to \mathbb{R}$ (possibly combined with non-linear co-domain transformations) one can design all kinds of filtering operators $f \mapsto \Phi(f)$. It is important to check whether such an operator commutes with rotations and translations. I.e. applying some rotation and translation to an input image should correspond to applying the same rotation and translation to the output image. Now let us formalize this mathematically.

Let the rotation and translation operator be given by

$$(\mathcal{U}_{(\mathbf{x},\theta)} f)(\mathbf{y}) = f(\mathbf{R}_\theta^{-1}(\mathbf{y} - \mathbf{x})), \qquad \mathbf{y} \in \mathbb{R}^2, \tag{26}$$

with translation vector $\mathbf{x} \in \mathbb{R}^2$ and counter-clockwise rotation[5] $\mathbf{R}_\theta \in SO(2)$ about angle θ. Let $\Phi : \mathbb{L}_2(\mathbb{R}^2) \to \mathbb{L}_2(\mathbb{R}^2)$ be an operator. Then such an operator commutes with roto-translations if and only if

$$\Phi \circ \mathcal{U}_g = \mathcal{U}_g \circ \Phi \ ,$$

for *all* roto-translations $g = (\mathbf{x}, \theta)$ with translation vector $\mathbf{x} \in \mathbb{R}^d$ and rotation angle $\theta \in [0, 2\pi)$.

Exercise-Mathematica 2. *Consider the notebook "Lecture1_Gaugederivatives.nb" that can be downloaded from (*http://www.lieanalysis.nl/201-2/*).*

Load a digital photograph of yourself into Mathematica *and turn it into grayscale-valued function $f : \mathbb{R}^2 \to \mathbb{R}$ with compact support. (NB. Use the "Import" and "Resize" command in Mathematica if needed, press F1 for help index)*
Let $\mathbf{R}_\theta$ denote the counter-clockwise rotation in the plane about angle θ.

a.) Apply the following operators Φf consisting of Gaussian derivatives acting on f:

$$\begin{aligned}
\Phi_1 f &:= \partial_x^s f, \\
\Phi_2 f &:= \partial_y^s f, \\
\Phi_3 f &:= (\partial_x^s f)^2 + (\partial_y^s f)^2 = (\partial_w^s f)^2, \\
\Phi_4 f &:= -(\partial_v^s)^2 f \cdot (\partial_w^s f)^2,
\end{aligned}$$

where v and w are local coordinates along a first order gauge frame with w along the gradient and v orthogonal to the gradient, such that

$$\partial_w^s f = \nabla^s f \cdot \frac{\nabla^s f}{\|\nabla^s f\|} = \|\nabla^s f\| \text{ and } \partial_v^s f = \nabla^s f \cdot \mathbf{R}_{\frac{\pi}{2}} \frac{\nabla^s f}{\|\nabla^s f\|},$$

for various values of $s = \frac{1}{2}\sigma^2$, e.g. $\sigma = 1, 2, 5$ pixel-widths.

b.) Repeat exercise 1a but then using a 90-degree rotated version of f as input.
c.) Which of the operators $\Phi_1, \Phi_2, \Phi_3, \Phi_4$ commute with $\mathcal{U}_{(\mathbf{x},\theta)}$ for all $\mathbf{x} \in \mathbb{R}^2$ and all $\theta \in [0, 2\pi)$?

d.) Motivate your answers at 3c. with a proof or with a counter-example.

Exercise 12 (generators for translation and rotation)

a.) Show that the generator w.r.t. translations $\mathbf{x}_0 = (x_0, y_0) \in \mathbb{R}^2$ equals

$$\lim_{t \downarrow 0} \frac{(\mathcal{U}_{(tx_0, ty_0, 0)} f)(x, y) - f(x, y)}{t} = ((x_0 \partial_x + y_0 \partial_x) f)(x, y) = (\mathbf{x}_0)^T \cdot \nabla f(\mathbf{x})$$

for all differentiable $f : \mathbb{R}^2 \to \mathbb{R}$, with $\mathbf{x} = (x, y) \in \mathbb{R}^2$.

[5] Remember: the Special Orthogonal group $SO(d)$ on $\mathbb{R}^d$ is given by $SO(d) = \{R \in GL(\mathbb{R}^d) \mid R^T = R^{-1} \text{ and } \det(R) = 1\}$ where $GL(\mathbb{R}^d)$ denotes the group of all invertible linear mappings from $\mathbb{R}^d$ onto $\mathbb{R}^d$.

b.) Show that for analytic functions $f : \mathbb{R}^2 \to \mathbb{R}$ one has

$$\begin{aligned}(\mathcal{U}_{(t,0,0)}f)(x,y) &= f(x+t,y)\\ &= (e^{t\,(\partial_x)}f)(x,y) := \sum_{k=0}^{\infty} \tfrac{t^k\,((\partial_x)^k\,f)(x,y)}{k!}\end{aligned}$$

c.) Show that the generator w.r.t. rotations equals

$$\lim_{t\downarrow 0} \frac{(\mathcal{U}_{(0,0,t)}f)(x,y) - f(x,y)}{t} = ((x\partial_y - y\partial_x)f)(x,y)$$

for all differentiable $f : \mathbb{R}^2 \to \mathbb{R}$.

d.) Show that for analytic functions $f : \mathbb{R}^2 \to \mathbb{R}$ one has

$$\begin{aligned}(\mathcal{U}_{(0,0,t)}f)(x,y) &= f(x\cos t - y\sin t, x\sin t + y\cos t)\\ &= (e^{t\,(x\partial_y - y\partial_x)}f)(x,y) := \sum_{k=0}^{\infty} \tfrac{t^k\,((x\partial_y - y\partial_x)^k\,f)(x,y)}{k!} \text{ for all } (x,y)\in\mathbb{R}^2.\end{aligned}$$

7 Locally Adaptive Frames in the Image Domain

7.1 Locally Adaptive Frames from Symmetric Matrix Fields

Recall from linear algebra that a symmetric linear operator $\mathcal{A} : V \to V$ on a dimensional vector space V with finite dimension $n \in \mathbb{N}$ admits a complete orthonormal basis $\alpha := \{\mathbf{v}_i\}_{i=1}^n$ of eigenvectors with corresponding eigenvalues $\{\lambda_i\}_{i=1}^n$, and w.r.t. this basis the corresponding matrix representation takes a diagonal form $\mathbf{A}_\alpha = \mathbf{S} \circ \boldsymbol{\Lambda} \circ \mathbf{S}^T$ with $\mathbf{S}$ the matrix consisting of eigenvector columns $\mathbf{S} = (\mathbf{v}_1 \mid \cdots \mid \mathbf{v}_n)$, and diagonal matrix $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\}$.

Similarly, one can compute for each fixed position $\mathbf{x} \in \mathbb{R}^d$ in the domain of an image $f : \mathbb{R}^d \to \mathbb{R}$ a symmetric matrix field that encodes geometric structure. Based on the local Taylor series expansion of the Gaussian filtered image

$$(G_s * f)(\mathbf{y}) = (G_s * f)(\mathbf{x}) + (\nabla G_s * f)(\mathbf{x}) \cdot (\mathbf{y}-\mathbf{x}) + \frac{1}{2}\,(\mathbf{y}-\mathbf{x})^T\,\mathbf{H}(G_s * f)(\mathbf{x})\,(\mathbf{y}-\mathbf{x}) + O(\,\|\mathbf{x}-\mathbf{y}\|^3\,).$$

Two typical examples for the choice of the symmetric matrix field arise. That is

- the structure matrix $\mathbf{S}^{s,\varrho} f(\mathbf{x})$ (encoding 1st order differential structure and edges).
- the Hessian matrix $\mathbf{H}^s f(\mathbf{x})$ (encoding 2nd order differential structure and lines).

See Figure 6. The structure matrix at $\mathbf{x} \in \mathbb{R}^d$ is defined by

$$\boxed{(\mathbf{S}^{s,\rho}(f))(\mathbf{x}) := (G_\rho * \nabla^s f\,(\nabla^s f)^T)(\mathbf{x}),} \tag{27}$$

with $s = \frac{1}{2}\sigma_s^2$, and $\rho = \frac{1}{2}\sigma_\rho^2$ the scale of regularization typically yielding a non-degenerate and positive definite matrix. Here $s > 0$ is often referred to as an internal regularization parameter, whereas $\rho > 0$ is referred to an external regularization parameter. In the remainder we use short notation $\mathbf{S}^{s,\rho} := \mathbf{S}^{s,\rho}(f)$.

The Hessian matrix includes only one scale parameter and is defined by

$$\boxed{(\mathbf{H}^s(f))(\mathbf{x}) = \left[\partial_{x_j}\partial_{x_i}(G_s * f)(\mathbf{x})\right].} \tag{28}$$

From now on we use short notation $\mathbf{H}^s := \mathbf{H}^s(f)$.

In the subsequent subsections we will show that diagonalization of the structure matrix corresponds to a notion of 1st order exponential curve fits, whereas diagonalization of the Hessian matrix corresponds to a notion of 2nd order exponential curve fits in the Lie group $(\mathbb{R}^2, +)$. At first sight, you may wonder about the formalities and notation, but later it becomes relevant when generalizing the procedure to another Lie group, namely the Lie group of roto-translations. For now let us introduce some preliminaries and definitions on Lie groups in general, that we will need in the remaining sections of this course.

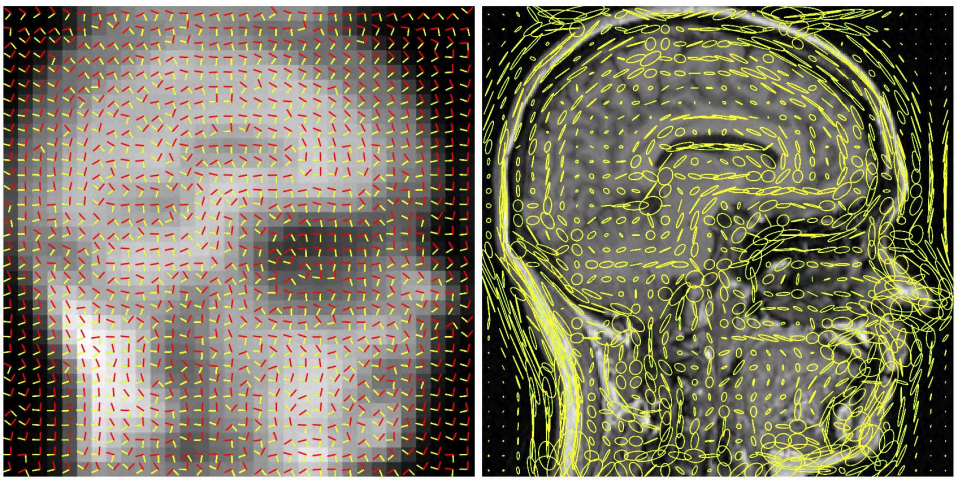

Fig. 6 Left: Locally adaptive frames (gauge frames) in the image domain computed as the eigenvectors of the Hessian of the image at each location. Right: Such gauge frames can be used for adaptive anisotropic diffusion and geometric reasoning. However, at complex structures (blobs/crossings) the gauge frames are ill-defined causing fluctuations.

Intermezzo: Preliminaries on Lie Groups

Definition 1 (Lie Group) A Lie group G is a differentiable manifold equipped with a group product such that the group operations are compatible with the differentiable structure on the manifold.

Example 1 The translation group $(\mathbb{R}^d, +)$ with group product

$$\mathbf{x}_1 \mathbf{x}_2 := \mathbf{x}_1 + \mathbf{x}_2$$

for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^d$.

Example 2 The group of invertible matrices

$$GL(d, \mathbb{R}) = \{M \in \mathbb{R}^{d \times d} \mid M \text{ is invertible } \},$$

equipped with the usual (operator/matrix) product $M_1 \circ M_2$.
The special orthogonal matrix group

$$SO(d) = \{R \in GL(d, \mathbb{R}) \mid R^T = R^{-1} \text{ and } \det(R) = 1\}.$$

again with the usual (operator/matrix) product $R_1 \circ R_2$.

Example 3 Consider the Lie group of rotations and translations of vectors in $\mathbb{R}^d$. The so-called "roto-translation group". Such roto-translations are done as follows

$$\mathbf{v} \mapsto R(\mathbf{v}) + \mathbf{x} \text{ for all } \mathbf{v} \in \mathbb{R}^d, \mathbf{x} \in \mathbb{R}^d, R \in SO(d) \tag{29}$$

Now if we concatenate two of such roto-translations it is again a roto-translation, where we recognize a semi-direct group product: Indeed computation of

$$\mathbf{v} \mapsto R_1(R_2(\mathbf{v}) + \mathbf{x}_2) + \mathbf{x}_1 = (R_1 R_2)\mathbf{v} + (\mathbf{x}_1 + R_1 \mathbf{x}_2),$$

reveals group product

$$(\mathbf{x}_1, R_1)(\mathbf{x}_2, R_2) = (\mathbf{x}_1 + R_1 \mathbf{x}_2, R_1 R_2) \tag{30}$$

on the manifold of roto-translations itself (rather than the space $\mathbb{R}^d$ on which it acts via (29)). It is semi-direct as the rotation part affects the spatial part, and therefore we write $G = SE(d) = \mathbb{R}^d \rtimes SO(d)$ instead[6] of $\mathbb{R}^d \times SO(d)$.

[6] On a direct product structure one would have $(\mathbf{x}_1, R_1)(\mathbf{x}_2, R_2) = (\mathbf{x}_1 + \mathbf{x}_2, R_1 R_2)$ on $\mathbb{R}^d \times SO(d)$.

Definition 2 The unity element of the Lie group is the unique element $e \in G$ such that the product eg equals g for all $g \in G$.

Definition 3 The tangent space at the unity element $T_e(G)$ is the vector space of all tangent vectors of curves passing through the unity element e of the Lie group.

Definition 4 (Exp Curves)
The exponential of $X \in T_e(G)$ is given by

$$\gamma(1) = \exp(X)$$

where $t \mapsto \gamma(t)$ is the unique one-parameter subgroup of G whose tangent vector at the identity is equal to X, i.e. $\gamma'(0) = X$ and $\gamma(t+s) = \gamma(t)\gamma(s)$ for all $s, t \in \mathbb{R}$.

Definition 5 A Lie algebra $\mathfrak{g}$ is a vector space over a field $\mathcal{F}$ equipped with a binary operation (the so-called Lie-bracket) $[\cdot,\cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ such that

$$\begin{array}{ll} \forall_{\alpha,\beta\in\mathcal{F}}\forall_{X,Y\in\mathfrak{g}} & : [\alpha X + \beta Y, Z] = \alpha[X,Z] + \beta[Y,Z] \\ \forall_{\alpha,\beta\in\mathcal{F}}\forall_{X,Y\in\mathfrak{g}} & : [Z, \alpha X + \beta Y] = \alpha[Z,X] + \beta[Z,Y] \\ \forall_{X\in\mathfrak{g}} & : [X,X] = 0, \\ \forall_{X,Y,Z\in\mathfrak{g}} & : [X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] = 0, \end{array}$$

Remark 6 A consequence is a fundamental property of the Lie bracket in general:

$$[X,Y] = -[Y,X]$$

for all $X, Y \in \mathfrak{g}$, since by the first 3 requirements one has: $0 = [X+Y, X+Y] = 0 + 0 + [X,Y] + [Y,X]$.

Within this course we always consider the field $\mathcal{F} = \mathbb{R}$. We often consider a special case where the vector space is the tangent space $T_e(G)$ at the unity element e of the Lie group G.

Lemma 2 *The space $T_e(G)$ is a Lie algebra with Lie-bracket*

$$\begin{aligned} [A,B] &= \tfrac{d}{dt}\big|_{t=0} (\gamma^{-B}(\sqrt{t}) \circ \gamma^{-A}(\sqrt{t}) \circ \gamma^{B}(\sqrt{t}) \circ \gamma^{A}(\sqrt{t}))(e) \\ &= \tfrac{1}{2}\,\tfrac{d^2}{dt^2}\big|_{t=0} (\gamma^{-B}(t) \circ \gamma^{-A}(t) \circ \gamma^{B}(t) \circ \gamma^{A}(t))(e), \end{aligned} \tag{31}$$

where $t \mapsto \gamma^X(t) = e^{tX}$ is a differentiable curve in G with $(\gamma^X)'(0) = X$ for $X = A, B$.
In case G is a matrix group the elements in $T_e(G)$ are matrices and then one simply has

$$[A,B] = AB - BA$$

This result (which is to be remembered) is a bit involved to verify and we refer to standard textbooks [21] on Lie Algebras and Lie groups for the details.

Remark 7 For matrix groups G the exponential map corresponds to the common matrix exponent map

$$\exp(tA) = e^{tA} = \sum_{k=0}^{\infty} \frac{t^k\, A^k}{k!},$$

with satisfies $\exp((t+s)A) = \exp(tA)\exp(sA)$. This holds for example for the the Lie group $GL(3,\mathbb{R})$ and in particular for the matrix subgroup $SO(3) = \{R \in GL(3,\mathbb{R}) \mid R^T = R^{-1} \text{ and } \det(R) = 1\}$, where

$$T_e(SO(3)) = so(3) = \{X \in \mathbb{R}^{3\times 3} \mid X^T = -X\}. \tag{32}$$

Exercise 13 (matrix group exponentials)

a.) Show that

$$(e^{tX})^{-1} = e^{-tX}$$

for all $X \in \mathbb{R}^{3\times 3}$ and all $t \geq 0$.

b.) Show that

$$(e^{tX})^T = e^{tX^T}$$

for all $X \in \mathbb{R}^{3\times 3}$ and all $t \geq 0$.

c.) Show that (32) holds.

d.) Let $X \in \mathbb{R}^{2\times 2}$ be given by

$$X = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Compute e^{tX}. What does the operator $\mathbf{x} \mapsto e^{tX}\mathbf{x}$ do geometrically?

e.) Let $X \in \mathbb{R}^{3\times 3}$ be given by

$$X = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Compute e^{tX}. What does the operator $\mathbf{x} \mapsto e^{tX}\mathbf{x}$ do geometrically?

f.) Let $X \in \mathbb{R}^{3\times 3}$ be given by

$$X = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Compute e^{tX}. What does the operator $\mathbf{x} \mapsto e^{tX}\mathbf{x}$ do geometrically?

Remark 8 The notion of exp curves in the Lie group $\mathbb{R}^d$ is very simple: it leads to straight-lines (verify this!). Indeed we find

$$\exp_{e=(0,0,0)}\left(t\sum_{i=1}^{d} c^i A_i\right) = (0,0,0) + t\,(c^1, c^2, c^3).$$

Due to the commutative structure of the Lie group $(\mathbb{R}^d, +)$ the concept of a straight curve coincides with the concept of a shortest curve. This will no longer be the case for the Lie group $SE(d)$ of roto-translations on $\mathbb{R}^d$ as we will see in later sections of the course-notes.

7.2 Exponential Curve Fits in $\mathbb{R}^d$

In this section we reformulate the classical construction of a locally adaptive frame to image f at location $\mathbf{x} \in \mathbb{R}^d$, in a group-theoretical way.

This reformulation seems technical at first sight, but it will help to understand the formulation of exponential curve fits in the higher dimensional Lie group $SE(d)$, that we will explain later on.

7.3 Exponential Curve Fits in $\mathbb{R}^d$ of the 1st Order

The structure matrix appears in solving the following optimization problem where for all $\mathbf{x} \in \mathbb{R}^d$ we aim to find optimal tangent vector

$$\begin{aligned}\mathbf{c}^*(\mathbf{x}) &= \underset{\substack{\mathbf{c}\in\mathbb{R}^d,\\ \|\mathbf{c}\|=1}}{\arg\min} \int_{\mathbb{R}^d} G_\rho(\mathbf{x}-\mathbf{x}')|\nabla^s f(\mathbf{x}')\cdot\mathbf{c}|^2 \mathrm{d}\mathbf{x}' \\ &= \underset{\substack{\mathbf{c}\in\mathbb{R}^d,\\ \|\mathbf{c}\|=1}}{\arg\min}\ \mathbf{c}^T \mathbf{S}^{s,\rho}(\mathbf{x})\mathbf{c}.\end{aligned} \tag{33}$$

In this optimization problem we find the tangent $\mathbf{c}^*(\mathbf{x})$ which minimizes a (Gaussian) weighted average of the squared directional derivative $|\nabla^s f(\mathbf{x}')\cdot\mathbf{c}|^2$ in the neighborhood of $\mathbf{x}$. The second identity in (33), which directly follows from the definition of the structure matrix, allows us to solve optimization problem (33) via the Euler-Lagrange equation

$$\nabla_{\mathbf{c}}\left(\mathbf{c}^T\mathbf{S}^{s,\rho}(\mathbf{x})\mathbf{c}\right) = \lambda_1\nabla_{\mathbf{c}}(\|\mathbf{c}\|^2-1) \Leftrightarrow \mathbf{S}^{s,\rho}(\mathbf{x})\ \mathbf{c}^*(\mathbf{x}) = \lambda_1\mathbf{c}^*(\mathbf{x}), \tag{34}$$

and the minimizer is found as the eigenvector $\mathbf{c}^*(\mathbf{x})$ with the smallest eigenvalue λ_1.

Now let us put Eq. (33) in group-theoretical form by reformulating it as an exponential curve fitting problem. This is helpful in our subsequent generalizations to $SE(d)$. On $\mathbb{R}^d$ exponential curves are straight lines:

$$\gamma^{\mathbf{c}}_{\mathbf{x}}(t) = \mathbf{x} + \exp_{\mathbb{R}^d}(t\mathbf{c}) = \mathbf{x} + t\mathbf{c}, \tag{35}$$

and on $T(\mathbb{R}^d)$ we impose the standard flat metric tensor $\mathfrak{G}(\mathbf{c},\mathbf{d}) = \sum_{i=1}^d c^i d^i$. In (33) we replace the directional derivative by a time derivative (at $t=0$) when moving over an exponential curve:

$$\boxed{\mathbf{c}^*(\mathbf{x}) = \underset{\mathbf{c}\in\mathbb{R}^d,\|\mathbf{c}\|=1}{\arg\min} \int_{\mathbb{R}^d} G_\rho(\mathbf{x}-\mathbf{x}')\ \Big|\ \tfrac{d}{dt}(G_s * f)(\gamma^{\mathbf{c}}_{\mathbf{x}',\mathbf{x}}(t))\,\Big|_{t=0}\ \Big|^2\ \mathrm{d}\mathbf{x}',} \tag{36}$$

where

$$t \mapsto \gamma^{\mathbf{c}}_{\mathbf{x}',\mathbf{x}}(t) = \gamma^{\mathbf{c}}_{\mathbf{x}}(t) - \mathbf{x} + \mathbf{x}' = \gamma^{\mathbf{c}}_{\mathbf{x}'}(t). \tag{37}$$

Because in (33) we average over directional derivatives in the neighborhood of $\mathbf{x}$ we now average the time derivatives over a *family of neighboring exponential curves* $\gamma^{\mathbf{c}}_{\mathbf{x}',\mathbf{x}}(t)$, which are defined to start at neighboring positions $\mathbf{x}'$ but having the same spatial velocity as $\gamma^{\mathbf{c}}_{\mathbf{x}}(t)$. In $\mathbb{R}^d$ the distinction between $\gamma^{\mathbf{c}}_{\mathbf{x}',\mathbf{x}}(t)$ and $\gamma^{\mathbf{c}}_{\mathbf{x}'}(t)$ is not important but it will be in the $SE(d)$-case.

Definition 6 Let $\mathbf{c}^*(\mathbf{x}) \in T_{\mathbf{x}}(\mathbb{R}^d)$ be the minimizer in (36). We say $\gamma_{\mathbf{x}}(t) = \mathbf{x} + \exp_{\mathbb{R}^d}(t\mathbf{c}^*(\mathbf{x}))$ is the first-order exponential curve fit to image data $f : \mathbb{R}^d \to \mathbb{R}$ at location $\mathbf{x}$.

7.4 Exponential Curve Fits in $\mathbb{R}^d$ of the 2nd Order

For second-order exponential curve fits we need the Hessian matrix defined by

$$(\mathbf{H}^s(f))(\mathbf{x}) = \left[\partial_{x_j}\partial_{x_i}(G_s * f)(\mathbf{x})\right], \tag{38}$$

with G_s the Gaussian kernel given in Eq. (1). From now on we use short notation $\mathbf{H}^s := \mathbf{H}^s(f)$. When using the Hessian matrix for curve fitting we aim to solve

$$\mathbf{c}^*(\mathbf{x}) = \underset{\mathbf{c}\in\mathbb{R}^d,\,\|\mathbf{c}\|=1}{\arg\min} |\mathbf{c}^T\mathbf{H}^s(\mathbf{x})\mathbf{c}|. \tag{39}$$

In this optimization problem we find the tangent $\mathbf{c}^*(\mathbf{x})$ which minimizes the second-order directional derivative of (Gaussian) regularized data $G_s * f$. When all Hessian eigenvalues have the same sign we can solve the optimization problem (39) via the Euler-Lagrange equation

$$\mathbf{H}^s(\mathbf{x})\,\mathbf{c}^*(\mathbf{x}) = \lambda_1 \mathbf{c}^*(\mathbf{x}), \tag{40}$$

and its minimizer is found as the eigenvector $\mathbf{c}^*(\mathbf{x})$ with the smallest eigenvalue λ_1.

Now, we can again put Eq. (39) in group-theoretical form by reformulating it as an exponential curve fitting problem. This is helpful in our subsequent generalizations to $SE(d)$. We again rely on exponential curves as defined in (35). In (39) we replace the second order directional derivative by a second order time derivative (at $t = 0$) when moving over an exponential curve:

$$\boxed{\mathbf{c}^*(\mathbf{x}) = \underset{\mathbf{c}\in\mathbb{R}^d, \|\mathbf{c}\|=1}{\arg\min} \left| \frac{d^2}{dt^2}(G_s * f)(\gamma^{\mathbf{c}}_{\mathbf{x}}(t))\Big|_{t=0} \right|.} \tag{41}$$

Remark 9 In general the eigenvalues of Hessian matrix $\mathbf{H}^s$ do not have the same sign. In this case we still take $\mathbf{c}^*(g)$ as the eigenvector with smallest absolute eigenvalue (representing minimal absolute principal curvature, as will see in Exercise 19) below, though this no longer solves (39).

Definition 7 Let $\mathbf{c}^*(\mathbf{x}) \in T_{\mathbf{x}}(\mathbb{R}^d)$ be the minimizer in (41). We say $\gamma_{\mathbf{x}}(t) = \mathbf{x} + \exp_{\mathbb{R}^d}(t\mathbf{c}^*(\mathbf{x}))$ is the second-order exponential curve fit to image data $f : \mathbb{R}^d \to \mathbb{R}$ at location $\mathbf{x}$.

Remark 10 In order to connect optimization problem (41) with the first order optimization (36) we note that (41) can also be written as an optimization over a family of curves $\gamma^{\mathbf{c}}_{\mathbf{x}',\mathbf{x}}$ defined in (37):

$$\mathbf{c}^*(\mathbf{x}) = \underset{\substack{\mathbf{c}\in\mathbb{R}^d,\\ \|\mathbf{c}\|=1}}{\arg\min} \left| \int_{\mathbb{R}^d} G_s(\mathbf{x}-\mathbf{x}')\, \frac{d^2}{dt^2}(f)(\gamma^{\mathbf{c}}_{\mathbf{x}',\mathbf{x}}(t))\Big|_{t=0} \right| d\mathbf{x}', \tag{42}$$

because of linearity of the second-order time derivative.

Exercise 14 (Exp curve fits of the second order)

a. Verify (40), and show that the solution of (39) is indeed given by the eigenvector of the Hessian with the smallest eigenvalue.
b. Verify equality (42).

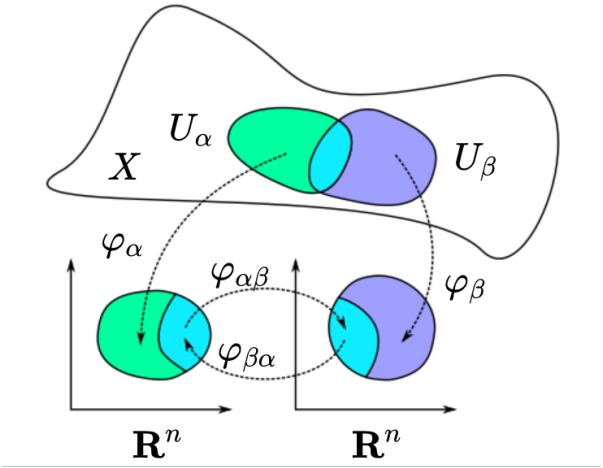

Fig. 7 Transition maps $\varphi_{\alpha\beta} = \varphi_\beta \circ \varphi_\alpha\big|_{\varphi_\alpha(U_\alpha \cap U_\beta)}$ between 2 charts in an atlas

Intermezzo: Tangent Vectors and Vector Fields (The Geometric and the Algebraic Viewpoint)

Let M be an n-dimensional smooth manifold with differentiable charts $\varphi_\alpha : U_\alpha \to \mathbb{R}^n$, with $(U_\alpha)_{\alpha \in I}$ some overlapping open sets covering M for some index set I.

Let $\gamma_1, \gamma_2 \in C^1(\mathbb{R}, M)$ be two differentiable curves on M, with $\gamma_i(0) = p$, $i = 1, 2$, then we define the following equivalence relation

$$\gamma_1 \sim \gamma_2 \Leftrightarrow \frac{d}{dt}(\varphi \circ \gamma_1)\Big|_{t=0} = \frac{d}{dt}(\varphi \circ \gamma_2)\Big|_{t=0}$$

for every chart $\varphi = \varphi_\alpha$ in the differentiable atlas (meaning that all transition maps, see Fig. 7, between charts are differentiable).

The geometric viewpoint:

This gives us a geometric definition of tangent vectors where

$$X_p = \gamma'(0) := \frac{d}{dt}(\varphi \circ \gamma)\Big|_{t=0}$$

for some curve γ from its equivalence class passing through $\gamma(0) = p \in M$, and where $\varphi : U_\alpha \to M$ is a chart s.t. $p \in U_\alpha$. Now as it is usually clear from the context which charts is used, and since one can always use transition maps to switch in a differentiable manner from one chart to the other we very often omit the charts φ in the notation of tangent vectors. The space of all tangent vectors at point $p \in M$ is a tangent space and it is commonly denoted by $T_p(M)$. We denote the tangent bundle by $T(M)$ and write

$$T(M) = \{\ (p, \dot{p}) \mid p \in M,\ \dot{p} \in T_p(M)\ \},$$

where we write $\dot{p} := X_p$ to recall that the tangent vector is obtained by differentiating a(n equivalence class) of curve(s). A vector field X on M is a section[7] in the tangent bundle and it can be seen as a mapping $p \mapsto X_p$ where $X_p \in T_p(M)$ for all p. Typically we consider differentiable vector fields (resp. smooth vector fields) where the mapping $p \mapsto X_p$ is differentiable (resp. smooth).

[7] i.e. $\pi \circ X = \mathrm{id}_M$ with $\pi : T(M) \to M$ the projection on the base manifold $\pi(p, \dot{p}) = p$.

From the geometric viewpoint towards the algebraic viewpoint:

Now let $\gamma \in C^1(\mathbb{R}, M)$ be a differentiable curve containing a point $\gamma(t_0) = p \in M$, then the tangent vector $\dot{\gamma}(t_0)$ at some $t_0 \in \mathbb{R}$ gives rise to a differentiable operator

$$\dot{\gamma}(t_0)f := \frac{d}{dt}(f \circ \gamma)(t_0) = \langle \mathrm{d}f(\gamma(t_0)), \dot{\gamma}(t_0)\rangle = \langle \mathrm{d}f(p), \dot{p}\rangle \in \mathbb{R} \tag{43}$$

that takes a directional derivative of a locally defines smooth function $f : \Omega_p \to \mathbb{R}$ in the direction $\dot{p} = \dot{\gamma}(t_0) \in T_p(M)$. This brings us to an alternative *algebraic* definition of tangent vectors that we will explain right after the following exercise and remark.

Example 4 Take $M = \mathbb{R}^2$, $t_0 = 0$ take $\gamma(t) = (x, y) + t(\cos x, y)$. Then

$$\dot{\gamma}(0) = (\cos x, y) \text{ as tangent to the curve } \gamma,$$

which we can identify with

$$\dot{\gamma}(0) = \cos x \frac{\partial}{\partial x} + y\partial_y \text{ as local differential operator .}$$

Indeed a quick computation gives us

$$\dot{\gamma}(0)f = \frac{d}{dt} f(x + t\cos x, y + ty)\Big|_{t=0} = \cos x \frac{\partial f}{\partial x}(x, y) + y\frac{\partial f}{\partial y}(x, y)$$

for all locally defined, real-valued and smooth functions f.

Exercise 15 Consider the manifold $M = \mathbb{R}^n$ and denote its elements $p = \mathbf{x} = (x^1, \ldots, x^n)$. Consider the vector field $\mathbf{x} \mapsto \mathbf{v}(\mathbf{x}) \in T_{\mathbf{p}}(\mathbb{R}^n)$ where $\mathbf{v}(\mathbf{x}) = \gamma'(0)$ where γ is the smooth curve in $\mathbb{R}^n$ given by

$$\gamma(t) = \mathbf{x} + t(v^1(\mathbf{x}), \ldots, v^n(\mathbf{x})).$$

with differentiable $v^i : C^1(\mathbb{R}^n) \to \mathbb{R}$ for $i = 1, \ldots, n$.

a.) Give a general expression (in terms of components v^i and partial derivatives ∂_{x^i}) for the differential operator $\gamma'(0) = \mathbf{v}(\mathbf{x})$ at fixed location $\mathbf{x} \in M$, which arises by applying (43).

b.) Explain that there is an important difference between vector fields and vector functions on $\mathbb{R}^n$. In particular, explain the difference between vector function $\tilde{\mathbf{v}} : \mathbb{R}^n \to \mathbb{R}^n$ given by $\tilde{\mathbf{v}}(\mathbf{x}) = (v^1(\mathbf{x}), \ldots, v^n(\mathbf{x}))$ and the vector field $\mathbf{v}$.

Remark 11 **(Derivatives of real-valued functions on manifolds and Riemannian gradients)** Throughout this lecture notes we reserve the straight brackets $\langle \hat{a}, b\rangle$ to indicate function evaluation $\hat{a}(b)$. In some cases the manifold M is equipped with a so-called Riemannian metric tensor that assigns to each $p \in M$ an inner product defined on $T_p(M)$:

$$(\dot{p}_1, \dot{p}_2) \mapsto G_p(\dot{p}_1, \dot{p}_2) \text{ with } \dot{p}_1, \dot{p}_2 \in T_p(M).$$

This is a useful tool to define Riemannian distances on M:

$$d_M(p, g) = \inf_{\substack{\gamma \in Lip((0,1), M),\\ \gamma(0) = p, \gamma(1) = q}} \int_0^1 G_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))\, \mathrm{d}t$$

as we will see later. In such a case one may apply the Riesz representation theorem which uniquely relates the *derivative at $p \in M$ fixed* (a linear functional) $\dot{p} \mapsto \langle \mathrm{d}f(p), \dot{p}\rangle$ to the gradient at p, by the well-known Riesz representation theorem

$$\langle \mathrm{d}f(p), \dot{p}\rangle = G_p(\nabla f(p), \dot{p}).$$

Now as the bilinear form $G_p : T_p(M) \times T_p(M) \to \mathbb{R}$ can be identified with the linear map given by $T_p(M) \ni \dot{p} \mapsto G_p(\dot{p}, \cdot) \in T_p^*(M)$ one commonly denotes

$$\nabla f(p) = G_p^{-1}\mathrm{d}f(p) \in T_p(M).$$

However, we stress that (43) holds in the general setting of differentiable manifolds and does not need an inner product structure nor does it need a gradient.

The algebraic viewpoint:

One can consider tangent vectors X_p as differential operators acting on locally defined smooth real-valued functions f on the smooth manifold. Then X_p is a tangent vector if

$$\begin{aligned} X_p(fg) &= f(p)X_p(g) + g(p)X_p(f), \\ X_p(\alpha f + \beta g) &= \alpha X_p(f) + \beta X_p(g), \end{aligned}$$

for all $\alpha, \beta \in \mathbb{R}$ and for all $f, g : \Omega_p \to \mathbb{R}$ locally defined real-valued smooth functions, defined on an open set Ω_p with (internal point) $p \in \Omega_p$,

Now we are going to express everything in the coordinate system $\{x^i\}_{i=1}^n$ associated to a local chart (diffeomorphism) $\varphi : \Omega_p \to \mathbb{R}^n$ around p, i.e. $\varphi(p) = (x^1(p), \dots, x^n(p))^T$. Then we define

$$(\partial_{x^i})_p : C^1(\Omega_p) \to \mathbb{R} \text{ by } (\partial_{x^i})_p f = \frac{\partial}{\partial x^i}(f \circ \varphi^{-1})\Big|_{\varphi(p)}.$$

Now by the chain-law the function

$$f(\cdot) - \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x^i}\right)_p x^i(\cdot)$$

is locally flat at p so that for every X_p tangent vector at p one has

$$X_p(f) = \sum_{i=1}^{n} X_p(x^i)\ (\partial_{x^i})_p (f). \tag{44}$$

So in other words the tangent space $T_p(M)$ is an n dimensional vector space for which $\{(\partial_{x^i})_p\}_{i=1}^n$ is a basis, and the coefficients of a tangent vector $X_p \in T_p(M)$ is found by applying the vector fields to each of the coordinate functions.

Example 5 Consider the 2-sphere $M = S^2 = \{\mathbf{n} = (n^1, n^2, n^3) \in \mathbb{R}^3 \mid \|\mathbf{n}\| = 1\}$. Note that the 2-sphere needs more[8] than one chart for a complete atlas. Now let us take a point on the sphere $p = \mathbf{n} \neq (0, 0, \pm 1)$. Introduce the *local* Euler angle charts $\varphi(\mathbf{n}) = (\beta(\mathbf{n}), \gamma(\mathbf{n}))^T$ where

$$\begin{aligned} x^1(p) &= \beta(\mathbf{n}) = \arccos(n^3) \in (0, \tfrac{\pi}{2}), \\ x^2(p) &= \gamma(\mathbf{n}) = \arg(n^1 + i\, n^2) \in [0, 2\pi), \end{aligned}$$

so that we have a local parametrization around p:

$$p = \mathbf{n} = \mathbf{n}(\beta, \gamma) = \varphi^{-1}(\beta, \gamma) = (\cos\gamma\, \sin\beta\,,\, \sin\gamma\, \sin\beta\,,\, \cos\beta)^T.$$

Let us consider the vector field on $S^2 \setminus \{0, 0, \pm 1\}$:

$$X = n^1 \partial_{n^2} - n^2 \partial_{n^1}$$

i.e. the field that maps $p = (n^1, n^2, n^3)$ onto $X_p = n^1(p)\ \partial_{n^2}|_p - n^2(p)\ \partial_{n^1}|_p \in T_p(S^2)$. Then we see that $X_p(x^1) = 0$ and $X_p(x^2) = 1$ and by (44) it follows that

$$X_p = 0 \cdot (\partial_\beta)_p + 1 \cdot (\partial_\gamma)_p \text{ for all } p \in S^2 \setminus \{0, 0, \pm 1\} \text{ and } X = \partial_\gamma.$$

[8] Due to Brouwer's fixed point theorem, every tangent field on the sphere vanishes vanishes at some point on the sphere

From the algebraic viewpoint towards the geometric viewpoint:

When a smooth vector field $p \mapsto X_p$ is given one can obtain integral curves $\gamma(t)$ passing through $p \in M$ via

$$\begin{cases} \dot{\gamma}(t) = X_{\gamma(t)}, t \in I \subset \mathbb{R} \\ \gamma(0) = p. \end{cases} \tag{45}$$

where I is some open interval containing 0. In fact the above system boils down to integrating an ODE on $\mathbb{R}^n$. To this end we recall the notation for a chart $\varphi(p) = (x^1(p), \ldots, x^n(p)) \in \mathbb{R}^n$ for a locally defined diffeomorphism φ. Indeed if we apply both the left-hand side and right-hand side of the ODE in (45) to each of the coordinate functions we get an ODE on $\mathbb{R}^n$:

$$\begin{cases} \dot{\gamma}(t)(x^i) = X_{\gamma(t)}(x^i) & \text{for all } i = 1, \ldots, n \\ x^i(\gamma(0)) = x^i(p) & \text{for all } i = 1, \ldots, n \end{cases} \Leftrightarrow \begin{cases} \frac{d}{dt}(x^i \circ \gamma)(t) = X_{\gamma(t)}(x^i) & \text{for all } i = 1, \ldots, n \\ x^i(\gamma(0)) = x^i(p) & \text{for all } i = 1, \ldots, n. \end{cases}$$

This ODE typically has a unique solution (recall from ODE-theory the mild conditions of Picard's theorem where the right-hand side is required to be Lipschitz continuous w.r.t. variable $(\varphi \circ \gamma)(\cdot) : \mathbb{R} \to \mathbb{R}^n$).

In the next exercise we start off by taking the algebraic view point and then we take a tour towards the geometrical view point.

Exercise 16 Consider the disk $M = \{\mathbf{x} \in \mathbb{R}^2 \mid 0 < \|\mathbf{x}\| < 2\}$. Consider the chart $p = \varphi^{-1}(r, \phi) = (r\cos\phi, r\sin\phi)$, i.e. $\varphi(x, y) = (\sqrt{x^2 + y^2}, \arg(x + i\, y))$.

a.) Express the vector field $p \mapsto X_p$ given by

$$X = \frac{1}{\sqrt{x^2 + y^2}} (x\partial_y - y\partial_x) + x\partial_x + y\partial_y$$

into the basis $\{\partial_r, \partial_\phi\} := \{p \mapsto (\partial_r)_p \ , \ p \mapsto (\partial_\phi)_p\}$.

b.) Compute the flow-line through $p_0 := (\frac{1}{2}\sqrt{2}, \frac{1}{2}\sqrt{2})$.
In other words, find the curve $\gamma(t) = (r(t)\cos\phi(t), r(t)\sin\phi(t)) \in M$ such that

$$\begin{cases} \dot{\gamma}(t) = X_{\gamma(t)}, \text{ for } t < t_{max}. \\ \gamma(0) = p_0. \end{cases}$$

c.) Determine the maximum value of $t_{max} > 0$, such that $\gamma(t) \in M$ for all $t < t_{max}$, and determine the point $\gamma(t_{max})$ where the flow-line hits ∂M. What happens for $t \to -\infty$?

Pushforward and Vector Fields

The geometric and algebraic viewpoint on tangent vectors is visually summarized in Fig. 8.

In this subsection we will explain a technique to 'pushforward' a tangent vector on a manifold M to a tangent vector at a different point on another manifold M'. In the particular case where this other manifold is again the same manifold (i.e. $M' = M$), this 'pushforward' technique can even be used to generate an entire vector field from a single tangent vector. The latter is particular useful in case the manifold M is a Lie group G. In that case one can construct a basis for the Lie algebra $\mathcal{L}(G)$ of left-invariant vector fields on G from a basis in the tangent space $T_e(G)$ at the unity element e of the group G.

Next we formalize these statements mathematically, where we consider both the geometric and the algebraic viewpoint on tangent vectors. For a preview of what is coming see Fig. 8, Fig. 9 and Fig. 10.

Consider a differentiable, bijective mapping $\Phi : M \to M'$. Note that $M' = \Phi(M)$. Then such a mapping can be used to transfer a differentiable curve γ on M passing through p to a differentiable curve γ_{new} on M' passing through $\Phi(p)$ in the following way:

$$\gamma_{new} : \mathbb{R} \to M' \text{ given by } \gamma_{new}(t) := (\Phi \circ \gamma)(t) = \Phi(\gamma(t)). \tag{46}$$

It can also be used to transfer a smooth realvalued function $f : \Omega_p \to \mathbb{R}$ locally defined on an open set Ω_p around p to a smooth realvalued function $f_{new} : \Phi(\Omega_p) \to \mathbb{R}$ locally defined on an open set around $\Phi(p)$:

$$f_{new} : \Phi(\Omega_p) \to \mathbb{R} \text{ given by } f_{new} = f \circ \Phi^{-1}. \tag{47}$$

Now (46) allows us to pushforward tangent vectors geometrically:

$$(\Phi_* X_p) := \gamma'_{new}(0) = (D\Phi)(p)\,\gamma'(0) = D\Phi(p)\,X_p. \tag{48}$$

This defines pushforward operator

$$\Phi_* : T_p(M) \to T_{\Phi(p)}(\Phi(M)).$$

that boils down to an application of the total derivative $D\Phi(p)$ as can be seen in (48).

Similarly (47) allows us to pushforward tangent vectors algebraically:

$$((X_{new})_{\Phi(p)})(g) = (\Phi_* X_p)g = X_p(g \circ \Phi),$$

for all smooth locally defined smooth functions $g : \Phi(\Omega_p) \to \mathbb{R}$. Indeed if we set $g = f_{new}$ then we see that $((X_{new})_{\Phi(p)})(f_{new}) = X_p(f)$ (verify this!).

See Fig. 9, where the concept of pushforward is visualized both geometrically (in green) and algebraically (in blue)

Now let us consider the case where the manifold equals a Lie group $M = G$ of dimension n, recall Definition 1.

Then we consider the mapping

$$L_g h = gh$$

that applies a left-multiplication with a fixed element $g \in G$. Then for each $g \in G$ fixed operator $L_g : G \to G$ is a differentiable bijection, and one has $g = L_g e$. Thereby $\Phi = (L_g)_*$ maps any element $A \in T_e(G)$ to an element $(L_g)_* A \in T_g(G)$. Let us choose a basis for $\{A_i\}_{i=1}^n$ for $T_e(g)$ then we obtain the corresponding vector fields

$$G \ni g \mapsto \mathcal{A}_i|_g := (L_g)_* A_i = (L_g)_*\, \mathcal{A}_i|_e$$

There is a fundamental property that make these vector fields rather special and it is the main reason why we will use them frequently in the remainder of the lecture notes.

$$\mathcal{A}_i|_{gh} = (L_g)_*\, \mathcal{A}_i|_h \;\Leftrightarrow\; \mathcal{A}_i|_{gh}\, U = (L_g)_*\, \mathcal{A}_i|_h\, U = \mathcal{A}_i|_h\, (U \circ L_g) \in \mathbb{R} \tag{49}$$

for all $h, g \in G$ and all $U : G \to \mathbb{R}$ smooth. As such it gives (see the exercise below) an important local moving frame of differential operators in the Lie group that commutes with left-actions (roto-translations of functions on the group).

Now let $U : G \to \mathbb{R}$ be globally defined on G and assume it is smooth, i.e. $U \in C^\infty(G, \mathbb{R})$. Then we define the smooth function $\mathcal{A}_i U : G \to \mathbb{R}$ by

$$\mathcal{A}_i U(g) = \mathcal{A}_i|_g\, (U) \in \mathbb{R}. \tag{50}$$

The linear operator that maps $U \to \mathcal{A}_i U$ is denoted by $\mathcal{A}_i$, and it is a globally defined vector *field*. In the next exercise you will show that it is a *left-invariant vector field*. This means that it satisfies (52).

The left-invariant vector fields again form a Lie algebra where the commutator is the usual Lie-bracket of vector fields. It turns out that

$$[\mathcal{A}_i, \mathcal{A}_j] = \mathcal{A}_i \circ \mathcal{A}_j - \mathcal{A}_j \circ \mathcal{A}_i = \sum_{k=1}^n c_{ij}^k \mathcal{A}_k \text{ and } [A_i, A_j] = \sum_{k=1}^n c_{ij}^k A_k,$$

where $c_{ij}^k \in \mathbb{R}$ are the so-called structure constants of the Lie algebra. In short $(L_g)_* : T_h(G) \to T_{gh}(G)$ is a Lie-algebra isomorphism for all $h \in G$. In particular for $h = e$. Left-invariant vector fields in the special cases $G = \mathbb{R}^2$ and $G = SE(2)$ (the group of roto-translations in the plane) are shown in Fig. 10.

Exercise 17 (left-invariant vector fields: a frame of left-invariant derivative operators)
Define $\mathcal{L}_g : \mathbb{L}_2(G) \to \mathbb{L}_2(G)$ by $\mathcal{L}_g U(h) = U(g^{-1}h)$.

a.) Show that $L_g \circ L_h = L_{gh}$ for all $g, h \in G$.
b.) Show that $\mathcal{L}_g U = U \circ L_{g^{-1}}$ for all $U \in \mathbb{L}_2(G)$ and all $g \in G$ and $\mathcal{L}_g^{-1} = \mathcal{L}_{g^{-1}}$ for all $g \in G$.
c.) Show that $\mathcal{L}_g \circ \mathcal{L}_h = \mathcal{L}_{gh}$ for all $g, h \in G$.
d.) Use Eq. (49) to show that

$$\mathcal{A}_i|_{gh}\,(\mathcal{L}_g U) = \mathcal{A}_i|_h\, U \text{ for all } g, h \in G.$$

for all $U \in C^1(G, \mathbb{R})$ and all $g \in G$.
e.) Show that the left-invariant vector fields provide a moving frame of differential operators in the Lie group that commutes with roto-translations acting from the left. I.e. show that:

$$(\mathcal{A}_i|_h \circ \mathcal{L}_g)(U) = (\mathcal{L}_g \circ \mathcal{A}_i|_h)(U), \text{ for all } g, h \in G, \text{ and all } U \in C^\infty(G, \mathbb{R}). \tag{51}$$

hint: Apply $\mathcal{L}_g^{-1}$ from the left to the above equality.
f.) Recall (50). Verify that (51) can be rewritten as

$$\mathcal{A}_i \circ \mathcal{L}_g = \mathcal{L}_g \circ \mathcal{A}_i \text{ for all } g \in G. \tag{52}$$

Example 6 Let us consider an example to illustrate how the abstract notions above apply to a specific case of interest, and to show the corresponding explicit computations. Consider to this end $G = SE(2) = \mathbb{R}^2 \rtimes SO(2) \equiv \mathbb{R}^2 \rtimes S^1$ with the group product given in (30) which under the identification $SO(2) \equiv S^1$ becomes (verify this!)

$$(x, y, \theta)(x', y', \theta') = (x + x' \cos\theta - y' \sin\theta, y - x' \sin\theta + y' \cos\theta, \theta + \theta' \,\mathrm{Mod} 2\pi).$$

Then we compute for all $g = (x, y, \theta)$ and all $U \in C^1(SE(2), \mathbb{R})$:

$$\begin{aligned}((L_g)_*\, \partial_x|_{e=(0,0,0)}\, U)(x, y, \theta) &= \partial_x|_{e=(0,0,0)}\,(U \circ L_g)\\ &= \lim_{t \downarrow 0} \tfrac{U((x,y,\theta)(t,0,0)) - U(x,y,\theta)}{t}\\ &= \lim_{t \downarrow 0} \tfrac{U(x + t\cos\theta, y + t\sin\theta, \theta, \theta) - U(x,y,\theta)}{t}\\ &= \cos\theta\, \partial_x U(x, y, \theta) + \sin\theta\, \partial_y U(x, y, \theta),\end{aligned}$$

so –on the level of vector fields– we have $(L_g)_* \partial_x = \cos\theta\, \partial_x + \sin\theta\, \partial_y$ for all $g = (x, y, \theta) \in G = SE(2)$.

Exercise 18 Consider $G = SE(2)$.

a.) Verify the computation of the other left-invariant vector fields (depicted in Fig. !10))

$$\begin{aligned}((L_g)_*\, \partial_y|_{e=(0,0,0)}\, U)(x, y, \theta) &= -\sin\theta\, \partial_x U(x, y, \theta) + \cos\theta\, \partial_y U(x, y, \theta)\\ ((L_g)_*\, \partial_\theta|_{e=(0,0,0)}\, U)(x, y, \theta) &= (\partial_\theta U)(x, y, \theta)\end{aligned}$$

b.) Show that every left-invariant vector field $\mathcal{A}$ on SE(2) is of the form

$$\mathcal{A} = c^1\, \mathcal{A}_1 + c^2\, \mathcal{A}_2 + c^3\, \mathcal{A}_3,$$

with c^1, c^2, c^3 constant.

Exercise 19 (Differential Geometry on the Graph of an Image and the Hessian and Structure Tensor)
Let $f : \mathbb{R}^2 \to \mathbb{R}$ be analytic. Consider the surface

$$\Gamma_f := \{(x, y, f(x, y)) \mid (x, y) \in \mathbb{R}^2\} \subset \mathbb{R}^3,$$

which is the graph of function f. Define $X : \mathbb{R}^2 \to \Gamma_f$ by

$$X(x, y) = (x, y, f(x, y))^T =: \mathbf{p} \in \Gamma_f$$

Suppose $\mathbf{x}_0 := (x_0, y_0) \in \mathbb{R}^2$ is a point such that $\nabla f(x_0, y_0) = (0, 0)^T$. A second order Taylor expansion gives

$$f(\mathbf{x}) = f(\mathbf{x}_0) + 0 + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T Hf(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + O(\|\mathbf{x} - \mathbf{x}_0\|^3).$$

a.) What does the Hessian $Hf(\mathbf{x}_0)$ tell us about the local shape of the isolines around $\mathbf{x}_0$?
b.) Show that $\mathbf{x}_0$ is a strict local minimum if $Hf(\mathbf{x}_0)$.

Next we relate the structure tensor and Hessian of f to the 1st resp. 2nd fundamental form on surface Γ_f.

c1.) Compute the 1st fundamental form on the surface Γ_f:

$$\mathcal{G}_{\mathbf{p}} = \sum_{i,j=1}^{2} g_{ij}(\mathbf{p}) \; \mathrm{d}x^i\Big|_{\mathbf{p}} \otimes \mathrm{d}x^j\Big|_{\mathbf{p}}.$$

with duals given by

$$\langle \mathrm{d}x^i, \partial_{x^j} \rangle = \delta^i_j$$

and with

$$g_{ij}(\mathbf{p}) = \partial_{x^i} X(x,y) \cdot \partial_{x^j} X(x,y), \text{for all } \mathbf{p} = X(x,y) \in \Gamma_f.$$

c2.) What is the relation of this form to the structure matrix field $\mathbf{p} \mapsto \nabla f(\mathbf{p})(\nabla f(\mathbf{p}))^T$ or rather to the structure tensor field $\mathbf{p} \mapsto \nabla f(\mathbf{p}) \otimes \nabla f(\mathbf{p})$ of f?
d1.) Compute the 2nd fundamental form on this surface Γ_f and the Christoffel symbols $\Gamma^i_{kj}(\mathbf{p})$ given by

$$\begin{aligned} &\mathcal{H}_{\mathbf{p}} = \sum_{i,j=1}^{2} h_{ij}(\mathbf{p}) \; \mathrm{d}x^i\big|_{\mathbf{p}} \otimes \mathrm{d}x^j\big|_{\mathbf{p}}, \\ &\partial_{x^i}\partial_{x^j} X(x,y) = \sum_{k=1}^{2} \Gamma^k_{ij}(\mathbf{p}) \; \partial_{x^k} X(x,y) + h_{ij}(X(x,y)) \; \mathbf{N}_{\mathbf{p}}, \text{ for all } \mathbf{p} = X(x,y) \in \Gamma_f, \end{aligned} \tag{53}$$

with normal vector

$$\mathbf{N}(X(x,y)) = \frac{\partial_x X(x,y) \times \partial_y X(x,y)}{\|\partial_x X(x,y) \times \partial_y X(x,y)\|}.$$

d2.) What is the relation of this form to the Hessian tensor field $\mathbf{p} \mapsto HF(\mathbf{p})$?

Let the C^2 curve $K \subset \Gamma_f$ be given by $s \mapsto \mathbf{x}(s)$, with

$$\dot{\mathbf{x}}(s) = \sum_{i=1}^{2} \dot{x}^i(s) \; \partial_{x^i} X|_{\mathbf{x}(s)},$$

where s is assumed to be spatial arc-length parameter (i.e. $\|\dot{\mathbf{x}}(s)\| = 1$), for some continuously differentiable functions $\dot{x}^i(\cdot)$.

e1.) Show that the geodesic curvature (by definition obtained by projecting the curvature vector $\ddot{X}(s)$ orthogonally to tangent space $T_{X(\mathbf{x}(s))}(\Gamma_f)$) equals

$$\|P_{T_{\mathbf{p}}(\Gamma_f)} \ddot{X}(s)\| = \sqrt{\sum_{i,j=1}^{2} g_{ij}(\mathbf{p}) \left(\ddot{x}^i(s) + \sum_{l,k=1}^{2} \Gamma^i_{lk}(\mathbf{p}) \, \dot{x}^l(s) \dot{x}^k(s) \right) \left(\ddot{x}^j(s) + \sum_{p,q=1}^{2} \Gamma^j_{pq}(\mathbf{p}) \, \dot{x}^p(s) \dot{x}^q(s) \right)} \tag{54}$$

with $\mathbf{p} = X(\mathbf{x}(s)) \in K$.
e2.) Provide the equations for the geodesics on Γ_f.

f.) Show that the principal curvature (by definition obtained by projecting the curvature vector $\ddot{X}(s)$ to the span of $\mathbf{N}_{X(\mathbf{x}(s))}$):

$$\|P_{\langle N_{\mathbf{p}} \rangle} \ddot{X}(s)\| = \left| \sum_{j,k=1}^{2} \dot{x}^j(s) \dot{x}^k(s) h_{jk}(\mathbf{p}) \right| \text{ with } \mathbf{p} = X(\mathbf{x}(s)) \in K. \tag{55}$$

g.) Use your results of items f. and d. to obtain a geometric interpretation of the eigensystem of the Hessian $Hf(x,y)$.

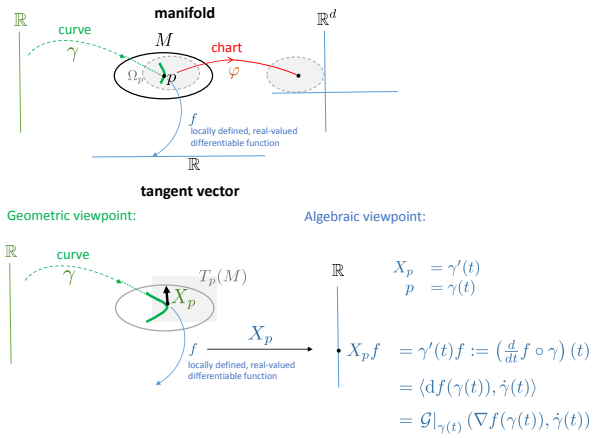

Fig. 8 Illustration of the geometric viewpoint (in green) and the algebraic viewpoint on a tangent vector on a manifold.

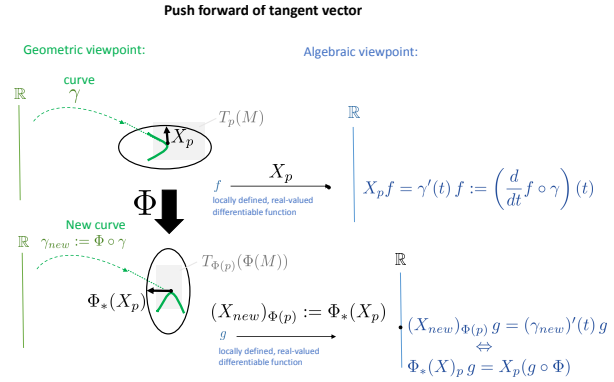

Fig. 9 Tangent vectors can be transferred via the pushforward operator.

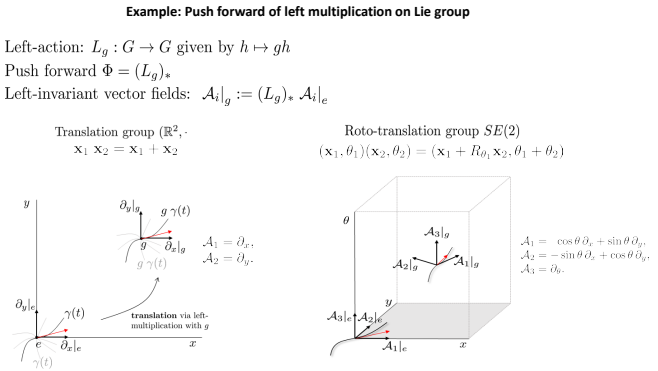

Fig. 10 Application of a pushforward operator in case the manifold M is a Lie group G of dimension n: The construction of the Lie-algebra $\mathcal{L}(G) = \text{span}\{\mathcal{A}_i\}_{i=1}^n$ of left-invariant vector fields from elements of $T_e(G) = \text{span}\{A_i = \mathcal{A}_i|_e\}_{i=1}^n$. Illustrated for $G = \mathbb{R}^2$ and $G = SE(2)$.

8 Diffusion Algorithms on Images

8.1 Linear Diffusion Algorithms

The linear diffusion equation (3) can be numerically solved by convolution. Such a convolution can be either implemented in the spatial domain via direct (equidistant) sampling of a Riemann sum approximation of

the convolution integral, or by a sampling in the Discrete Fourier Transform (DFT) domain (where again finite periodic convolutions related to a direct product in the DFT domain). For $s > 0$ small, the spatial Gaussian kernel is relatively narrow and if it is truncated close to the center, truncation errors are small. For $s > 0$ large, the Fourier transform of the Gaussian kernel is relatively narrow and if it is truncated close to the center, truncation errors are small. So for small scales $s > 0$ one applies a spatial convolution implementation and for large $s > 0$ one applies a discrete Fourier implementation (where one can take advantage of the fast Fourier transform FFT).

Unfortunately, such convolution-based implementations, lack the possibility to transparently include geometric data-adaptation via a nonlinear diffusion where one steers the translation-invariant vector fields with constant coefficients towards the position-dependent gauge-frames (or locally adaptive frames):

$$\boxed{\begin{array}{l}\text{shift-invariant frame } \rightarrow \text{ locally adaptive frame} \\ \{\partial_{x^1}, \cdots, \partial_{x^d}\} \rightarrow \left\{ \mathcal{B}_1(\mathbf{x}) := \sum_{i=1}^{d} c_1^i(\mathbf{x})\, \partial_{x^i},\ \cdots,\ \mathcal{B}_d(\mathbf{x}) := \sum_{i=1}^{d} c_d^i(\mathbf{x})\, \partial_{x^i} \right\}.\end{array}} \tag{56}$$

with $\mathbf{c}_j(\mathbf{x}) = (c_j^1(\mathbf{x}), \ldots, c_j^d(\mathbf{x}))^T$ the normalized[9] eigenvectors of the structure tensor of the image or the Hessian of the image, with indexing ordered by magnitude of the eigenvalues with $\mathbf{c}_1(\mathbf{x}) = \mathbf{c}^*(\mathbf{x})$ and minimal absolute eigenvalue $|\lambda_1(\mathbf{x})| = \min_{i \in \{1,\ldots,d\}} |\lambda_i(\mathbf{x})|$. Recall from Section 7.2 that this relates to the exponential curve fit

$$t \mapsto \mathbf{x} + t\, \mathbf{c}^*(\mathbf{x})$$

to the image data $\Gamma_f := \{(\mathbf{x}, f(\mathbf{x}))\ \ \mathbf{x} \in \mathrm{supp}(f)\}$ at location $\mathbf{x} \in \mathbb{R}^d$.

Therefore, in order to include locally adapative frames (56) in our diffusions, we resort to basic finite difference implementations that we will explain next.

Finite Difference Schemes for Linear Diffusions

As a popular and basic example we consider the explicit scheme with a central difference approximation of the Laplacian which boils down do discrete convolutions with stencil $[1, -2, 1]$:

$$\begin{array}{l} u_{i,j}^{n+1} - u_{i,j}^n = \frac{\Delta s}{|\Delta x|^2}(1u_{i-1,j}^n - 2u_{i,j}^n + 1u_{i,j}^n) + \frac{\Delta t}{|\Delta y|^2}(1u_{i,j-1}^n - 2u_{i,j}^n + 1u_{i,j+1}^n), \\ u_{i,j}^0 = f_{i,j}. \end{array} \tag{57}$$

For $n = 0, 1 \ldots$ and $i \in \{1, \ldots, N_x\}, j \in \{1, \ldots, N_y\}$.
Another popular and basic example is the corresponding implicit scheme given by

$$\begin{array}{l} u_{i,j}^{n+1} - u_{i,j}^n = \frac{\Delta t}{|\Delta x|^2}(1u_{i-1,j}^{n+1} - 2u_{i,j}^{n+1} + 1u_{i,j}^{n+1}) + \frac{\Delta t}{|\Delta y|^2}(1u_{i,j-1}^{n+1} - 2u_{i,j}^{n+1} + 1u_{i,j+1}^{n+1}), \\ u_{i,j}^0 = f_{i,j}. \end{array} \tag{58}$$

In both cases we have $u_{i,j}^n \approx u(i\Delta x, j\Delta y, n\Delta s)$, with u the solution to diffusion system (3) in the continuous setting. For simplicity, we assume $\Delta y = \Delta x$ in the remainder of this section.

When storing the components in one vector $\mathbf{u}^n := (u_{i,j}^n)_{i,j=1}^{N_x,N_y}$ and when storing the convolutions with stencil $[1, -2, 1]$ (where one may choose to work with zero padding outside the discrete domain, or to include periodic, Dirichlet or Neumann boundary conditions) as linear matrix operator $\Delta_D \in \mathbb{R}^{N_x N_y \times N_x N_y}$, we may rewrite the explicit scheme (57) as

$$\mathbf{u}^{n+1} = (1 + \Delta s\, \Delta_D)\, \mathbf{u}^n = (e^{\Delta s\, \Delta_D} + O((\Delta s/\Delta x)^2)\ )\, \mathbf{u}^n,$$

and the implicit scheme (58) as

$$\mathbf{u}^{n+1} = (1 - \Delta s\, \Delta_D)^{-1}\, \mathbf{u}^n = (e^{\Delta s\, \Delta_D} + O((\Delta s/\Delta x)^2)\ )\, \mathbf{u}^n,$$

where in the case of the implicit scheme sparse matrix inversion can be done (numerically). It reveals that the implicit scheme actually implements a small step resolvent operator and it is therefore unconditionally

[9] Normalization means here that $\|\mathbf{c}_j(\mathbf{x})\| = \sqrt{\sum_{k=1}^{d} |c_j^k(\mathbf{x})|^2} = 1$ for all $j = 1, \ldots, d$.

stable. This is in contrast to the explicit scheme where a sufficient condition for stability (regardless the initial condition $f_{i,j}$) is

$$\frac{\Delta s}{(\Delta x)^2} \leq \frac{1}{4}. \tag{59}$$

Exercise 20 Apply the well-known Gershgorin circle theorem to the matrix $(1 + \Delta s\, \Delta_D)$ to derive the stability bound (59) for the explicit scheme.

As a result, the advantage of using an implicit scheme is that one can take larger stepsizes to gain speed (while losing accuracy). The advantage of using an explicit scheme is that no (numerical) matrix-inversion is required which makes the method simple with good accuracy for small step sizes.

In comparison to the true heat equation the errors of both schemes are $O(\Delta s)+O((\Delta x)^2)$ which follows by basic Taylor expansion around the grid-points. As a result the schemes only have a first order accuracy in time. Of course, one can extend such discretization schemes to higher order schemes. When schemes have a higher order in the spatial direction, this typically comes with the price of larger discrete convolution stencils and a possible loss of rotation-covariance. Therefore it is popular to stick to the central differences in space (for second order accuracy in space). A popular scheme that is both second order in space and in scale is the Crank-Nikolson method:

$$\mathbf{u}^{n+1} = \left(I - \frac{\Delta_D\, \Delta s}{2}\right)^{-1} \left(I - \frac{\Delta_D\, \Delta s}{2}\right)\, \mathbf{u}^n.$$

Exercise 21 Explain why the Crank-Nikolson method creates only second order errors in time, whereas the Euler forward explicit and implicit methods give first order errors in time.
hint: Compute the Taylor series of $e^x, 1+x, (1-x)^{-1}$ *and* $\left(1-\frac{x}{2}\right)^{-1}\left(1+\frac{x}{2}\right)$ *around* $x=0$ *up to order* 3.

In the next two paragraphs we address two standard versions of nonlinear diffusions:

- *nonlinear diffusions with adaptive scalar diffusivity*, where the scalar diffusivity of 1 in between $\Delta = \mathrm{div} \circ 1 \circ \mathrm{grad}$ is replaced by a scalar field that strongly dampens diffusion across edges. It only accounts for the strength of edges, not their directions.
- *nonlinear diffusions with adaptive tensorial diffusivity*, where the scalar diffusivity of 1 in between $\Delta = \mathrm{div} \circ 1 \circ \mathrm{grad}$ is replaced by a symmetric field that steers the diffusion in the directions of the locally adaptive frame of vector fields (56). This allows us to integrate both the strength and direction of
  1. edges (in case the structure tensor of the image is used for exp curve fits of order 1).
  2. lines (in case the Hessian of the image is used for exp curve fits of order 2).

In Figure 11 (top-row) we observe the difference between a linear diffusion, an non-linear diffusion (with locally isotropic scalar diffusivity) and an nonlinear diffusion with a tensorial adaptive diffusivity, for a fixed large diffusion time. In Figure 11 (top-row) we observe the output of an anisotropic, tensorial, nonlinear diffusion (CED [22]) for several scales/diffusion times. The method works well on van Gogh paintings as there are very few crossings structures/lines in such paintings.

8.2 Nonlinear Diffusion with Adaptive Scalar Diffusivity

A first step to include a data-driven nonlinearity is to include a scalar diffusivity. The basic idea (by Perona & Malik [23]) was to stop the diffusion at edge locations which are typically locations where the (Gaussian)-gradient is high. This means that one obtains the diffusion PDE:

$$\begin{cases} \frac{\partial u}{\partial s}(\mathbf{x}, s) = \left(\mathrm{div} \circ e^{-\frac{\|\nabla^t f(\mathbf{x})\|^2}{c^2}} \circ \nabla u\right)(\mathbf{x}, s), & \text{for all } \mathbf{x} \in \mathbb{R}^2, s > 0, \\ u(\mathbf{x}, 0) = f(\mathbf{x}) & \text{for all } \mathbf{x} \in \mathbb{R}^2, \end{cases}$$

where we use the Gaussian gradient $\nabla^t f = \nabla G_t * f$ of the initial condition f as our diffusivity multiplier operator with multiplier $e^{-\frac{\|\nabla^t f\|^2}{c^2}}$ must be differentiable. Here one typically chooses $t > 0$ such that the standard deviation $\sigma = \sqrt{2t}$ is in the order of 1 or 2 pixel-sizes, to avoid edge dislocations over scale.

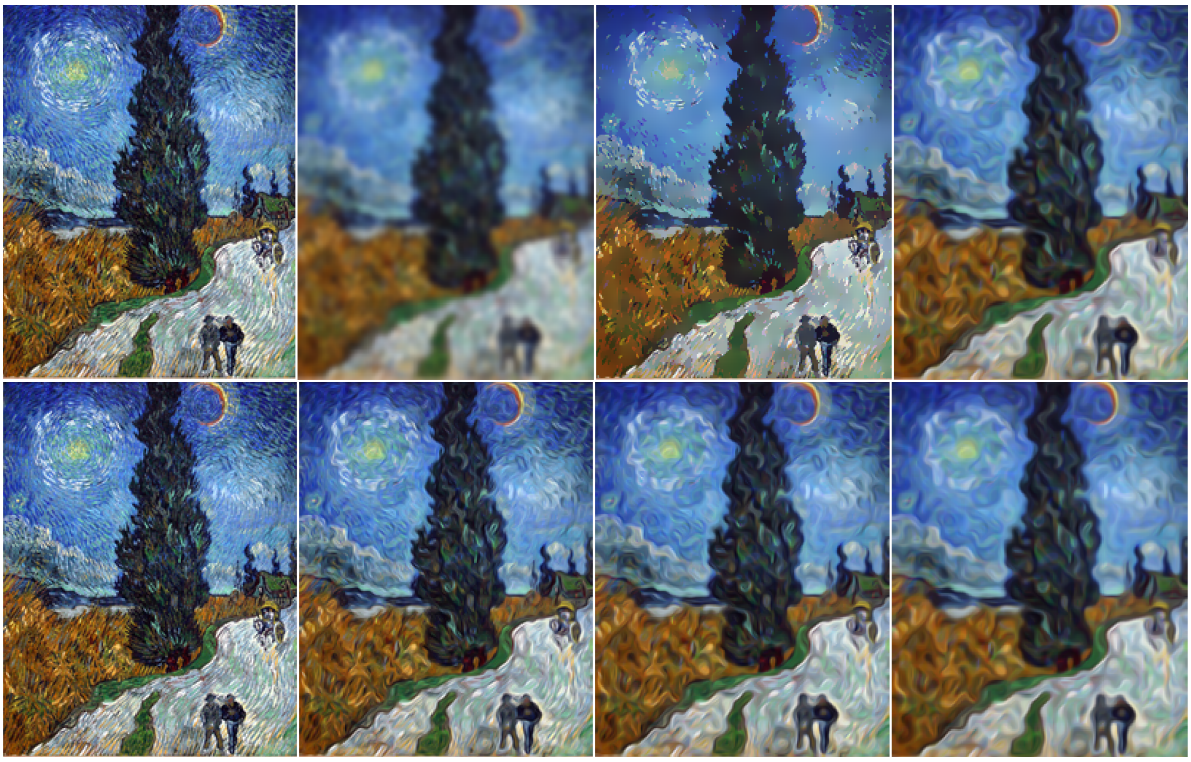

Fig. 11 Top row: from left to right, 1. input RGB-colour image $f : \Omega \to \mathbb{R}^3$, 2. output of linear diffusion, 3. output of nonlinear diffusion with scalar diffusivity [23], 4. output of anisotropic nonlinear diffusion (CED: Coherence Enhancing Diffusion [22]). Diffusions are applied to each of the RGB channels, and a stopping time (scale) of diffusion was set to $s = 4$ in all cases. Bottom row: from left to right, 1. input image, 2.-4. output CED diffused image at times, $s = 1$ $s = 3$ and $s = 4$. CED performs well on these images since they contain very little crossings. See also Fig. 12.

This can be implemented via finite differences using zero-padding. Here within the finite schemes (57) and (58) one must now take care of appropriate sampling of the diffusivity. We set the short notation $d(x,y) := e^{-\frac{\|\nabla^t f(x,y)\|^2}{c^2}}$ and $h = \Delta x$, and approximate the scalar diffusivity in say x-direction[10] as follows

$$\begin{aligned}
&\partial_x \left(d(x,y) \partial_x u(x,y,s) \right) \overset{CD.\frac{h}{2}}{\approx} \frac{d(x+\frac{h}{2},y)\ \partial_x u(x+\frac{h}{2},y,s) - d(x-\frac{h}{2},y)\ \partial_x u(x+\frac{h}{2},y,s)}{h} \\
&\overset{FD\&BD.\frac{h}{2}}{\approx} d(x+\tfrac{h}{2},y) \left(\tfrac{u(x+h,y,s)-u(x,y,s)}{h^2} \right) - d(x-\tfrac{h}{2},y) \left(\tfrac{u(x,y,s)-u(x-h,y,s)}{h^2} \right) \\
&\overset{lin.int.d}{\approx} \tfrac{d(x,y)+d(x+h,y)}{2} \left(\tfrac{u(x+h,y,s)-u(x,y,s)}{h^2} \right) - \tfrac{d(x-h,y)+d(x+h,y)}{2} \left(\tfrac{u(x,y,s)-u(x-h,y,s)}{h^2} \right) = \\
&d(x,y) \left(\tfrac{u(x+h,y,s)-2u(x,y,s)+u(x-h,y,s)}{2h^2} \right) + d(x+h,y) \left(\tfrac{u(x+h,y,s)-u(x,y,s)}{2h^2} \right) + d(x-h,y) \left(\tfrac{u(x-h,y,s)-u(x,y,s)}{2h^2} \right)
\end{aligned}$$

for each point $(x, y, s) = (i\Delta x, j\Delta y, k\Delta s)$ on the grid. Note we applied central differences (CD) for the last derivative, forward and backward differences for the first derivative, all at stepsize $h/2$ so that concatenation stays on the grid, with relatively small stencils.

Remark 12 In the final step we applied a linear interpolation which decreases the accuracy order in spatial direction from 2 to 1. One could instead use quadratic interpolation in between the 3 grid points $d(x+h, y)$, $d(x - h, y)$ and $d(x, y)$ to preserve second order accuracy in space, but this increases the stencil size for d and leads to a less sharp diffusion-reduction across edges in practice.

[10] The same procedure is applied in y-direction.

Remark 13 Above we considered the unbounded case, where we assumed f has compact support on say $\Omega = [0,1] \times [0,1]$. Alternatively, one can work on a bounded domain with Neumann boundary conditions:

$$\begin{cases} \frac{\partial u}{\partial s}(\mathbf{x},s) = \left(\mathrm{div} \circ e^{-\frac{\|\nabla^t f(\mathbf{x})\|^2}{c^2}} \circ \nabla u \right)(\mathbf{x},s) & \text{for all } \mathbf{x} \in \Omega,\ s \geq 0, \\ \frac{\partial u}{\partial n}(\mathbf{x},s) = 0 & \text{for all } \mathbf{x} \in \partial\Omega,\ s \geq 0, \\ u(\mathbf{x},0) = f(\mathbf{x}) & \text{for all } \mathbf{x} \in \Omega. \end{cases}$$

where the Neumann boundary condition can (with some book-keeping) be easily included in the finite difference discretization matrix.

Remark 14 (the parameter $c > 0$).
Visually appealing results can be obtained, but the result is very sensitive to the parameter $c > 0$. An automatic setting for parameter $c > 0$ that works reasonably well is to make a histogram of gradients and to set c equal to the 50% quantile of all gradient norms $\{\|\nabla^t f(\mathbf{x})\| \mid \mathbf{x} \in \mathrm{supp}(f)\}$ in the image.

Remark 15 (solution dependent diffusivity and the illposedness issues that come along).
It is also common in the applications to use the Gaussian gradient of the evolving solution instead of the initial condition in the discretized PDE, i.e. to replace the scalar diffusivity

$$e^{-\frac{\|\nabla^t u(\cdot,0)\|^2}{c^2}} = e^{-\frac{\|\nabla^t f\|^2}{c^2}} \rightarrow e^{-\frac{\|\nabla^t u(\cdot,s)\|^2}{c^2}} \ . \tag{60}$$

In practice, in the finite difference scheme one can for example skip the updating of the diffusivity, and apply it only in the beginning and then after each 10 iterations. The continuous setting of the Perona-Malik PDE, which actually uses (60), is not well-posed. In regions where the norm of the gradient is smaller than c, the equation is locally parabolic (and well-posed), but in regions where the norm of the gradient is larger than c one observes an ill-posed backward diffusion effect! In general one has nonexistence of global-in-time weak C^1 solutions whose initial data have regions with slope larger than the parabolicity threshold. Even for analytic initial data f such solutions break down in finite time/scale s. Nevertheless, by a rescaling procedure (while relating $c > 0$ to the grid size Δx) it is possible to obtain a continuum limit of the actual Perona & Malik PDE discretization scheme. For details, see [24]. The ill-posedness (induced by the choice (60)) is not just a problem in theory, in practice it can manifest itself by the so-called *"staircasing effect"*, where staircases arise in linear slope profiles in the image. An adequate practical fix to this problem (via regression filtering) is proposed in [25].

Remark 16 Instead of using a Gaussian function $g \mapsto e^{-(g/c)^2}$ acting on the norm gradient $g = \|\nabla^t f\|$, there are also other options like

$$g \mapsto \begin{cases} \frac{1}{\left(q + \left(\frac{g}{c}\right)^n\right)^p} & \text{if } g \geq 0 \\ 0 & \text{else} \end{cases}$$

with $n \in \mathbb{N}$, $p > 0$, $c > 0$ and $q > 0$. The case $q = 1$, $n = 2$ and $p = 1$ coincides with other type of diffusivities proposed by [23]. The case $q = 1$, $n = 2$ and $p = \frac{1}{2}$ coincides with diffusivity proposed by Charbonnier et al.[26]. The case ($q \downarrow 0$ and $n = 1$ and $p = 1$) coincides with total variation flow. In particular the total variation flow [27] is of interest, and we shall address its convergence in Subsection 9.2.

8.3 Nonlinear Anisotropic Diffusion steered by Locally Adaptive Frames

Given the locally adaptive frame (56) we construct an anisotropic tensorial diffusivity:

$$\begin{cases} \frac{\partial u}{\partial s}(\mathbf{x},s) = \left(\mathrm{div} \circ \mathbf{D}_f \circ \nabla u\right)(\mathbf{x},s) & \text{for all } \mathbf{x} \in \Omega,\ s \geq 0, \\ \frac{\partial u}{\partial n}(\mathbf{x},s) = 0 & \text{for all } \mathbf{x} \in \partial\Omega,\ s \geq 0, \\ u(\mathbf{x},0) = f(\mathbf{x}) & \text{for all } \mathbf{x} \in \Omega, \end{cases} \tag{61}$$

where the diffusion matrix $\mathbf{D}_f(\mathbf{x}) \in \mathbb{R}^{d \times d}$ is symmetric and diagonal along the eigenvectors of the structure-tensor or Hessian:

$\mathbf{D}_f(\mathbf{x}) = \sum_{k=1}^{d} \lambda^k_{NEW,f}(\mathbf{x})\ \mathbf{c}_k(\mathbf{x})(\mathbf{c}_k(\mathbf{x}))^T$ as a matrix-representation w.r.t. standard basis of tensor $\sum_{k=1}^{d} \lambda^k_{NEW,f}(\mathbf{x})\ \mathbf{c}_k(\mathbf{x}) \otimes \mathbf{c}_k(\mathbf{x})$,

and where $\lambda^k_{NEW,f}(\mathbf{x})$ are computed in a nonlinear way from the eigenvalues of the structure matrix $\mathbf{S}^{s,\varrho}f(\mathbf{x})$ or the Hessian $\mathbf{H}_s f(\mathbf{x})$ of image f, where one typically keeps the diffusion small along the eigenvector with largest eigne value, and where one amplifies the anisotropy in the other directions, to enhance coherent structures in the images. A popular choice is first proposed by J.Weickert in his coherence enhancing diffusion scheme [22], with for $d = 2$:

$$\lambda^1_{NEW}(\mathbf{x}) = \alpha, \quad \lambda^2_{NEW}(\mathbf{x}) = \alpha + (1-\alpha)\, e^{-\frac{c^2}{\lambda_1(\mathbf{x})-\lambda_2(\mathbf{x}))^2}},$$

with $c > 0$ an anisotropy parameter, that likewise in the previous section can be set automatically by histogram techniques, and an $0 < \alpha \ll 1$ to preserve positivity. In terms of the locally adaptive vector fields ('gauge frames') given by (56) this means that the diffusion generator is diagonal:

$$\begin{cases} \frac{\partial u}{\partial s}(\mathbf{x},s) = Qu(\mathbf{x},s) & \text{for all } \mathbf{x}\in\Omega,\ s\geq 0, \quad \text{with generator } Q = \sum_{k=1}^{d} \mathcal{B}_k \circ \lambda^k_{NEW,f} \circ \mathcal{B}_k. \\ \frac{\partial u}{\partial n}(\mathbf{x},s) = 0 & \text{for all } \mathbf{x}\in\partial\Omega,\ s\geq 0, \\ u(\mathbf{x},0) = f(\mathbf{x}) & \text{for all } \mathbf{x}\in\Omega. \end{cases} \tag{62}$$

Remark 17 From a numerical finite difference point of view, we observe two possibilities for the finite difference implementations. Either one expresses the system in the fixed frame of reference (61) and then applies finite differences along the grid. This is commonly done. Or one expresses the system in the moving frame of reference (62) and applies finite differences with interpolation (e.g. B-spline interpolation (21)), e.g. to approximate $\mathcal{B}_i u(\mathbf{x},s) = \lim_{t\to 0} \frac{u(\mathbf{x}+t\mathbf{c}^i(\mathbf{x}))-u(\mathbf{x})}{t} \approx \frac{u(\mathbf{x}+h\mathbf{c}^i(\mathbf{x}))-u(\mathbf{x})}{h}$, with stepsize $h > 0$.

8.4 Limitations of assigning a Single Frame per Position: Problems at Crossings

The next Mathematica exercise shows the limitations of assigning a single frame $\{\mathcal{B}_k(\mathbf{x})\}_{k=1}^d$ per position $\mathbf{x}$ in the image domain $\mathbb{R}^d$. In essence, the problem is that Gaussian gradients are computed over regions with scale $s = \frac{1}{2}\sigma^2 > 0$. If in such regions multiple edges or lines pass the direction of

- the Gaussian gradient of the image,
- the smallest eigenvector of the structure tensor of the image,
- the smallest eigenvector of the Hessian image,

is ill-defined and the gauge frame become more or less randomly oriented. This can also be seen in Fig. 6. Consider for example the huge fluctuations of the ellipsoids at the nose-tip in that image This

We will tackle this in the next section of the lecture notes, where the notion of invertible orientation scores allows us to assign per position a whole family of oriented frames. Moreover, such frames will live in the joint space of positions and orientations and allow us to account for curvature and orientation uncertainty of line detectors. The details on such locally adaptive frames in the joint space of positions and orientations will follow in the final section of these lecture notes.

Exercise-Mathematica 3. *Download the Mathematica notebook "Lecture1_DiffusiononR2.nb" from* http://www.lieanalysis.nl.201-2/.

You can run the notebook by selecting the cells and pressing 'shift' and 'enter'.
The images test images ('fingerprints' and 'van Gogh paintings') are special in the sense that there are very few complex structures such as crossings, corners and bifurcations. Therefore the locally adaptive frames and non-linear diffusions in the image domain work well.

In order to see the limitations of assigning a single frame to a spatial position (in scale space) you can download "testimage-Diffusion.png" from http://www.lieanalysis.nl.201-2/ *and run the experiments on this testimage and observe the problems for large diffusion times yourself.*

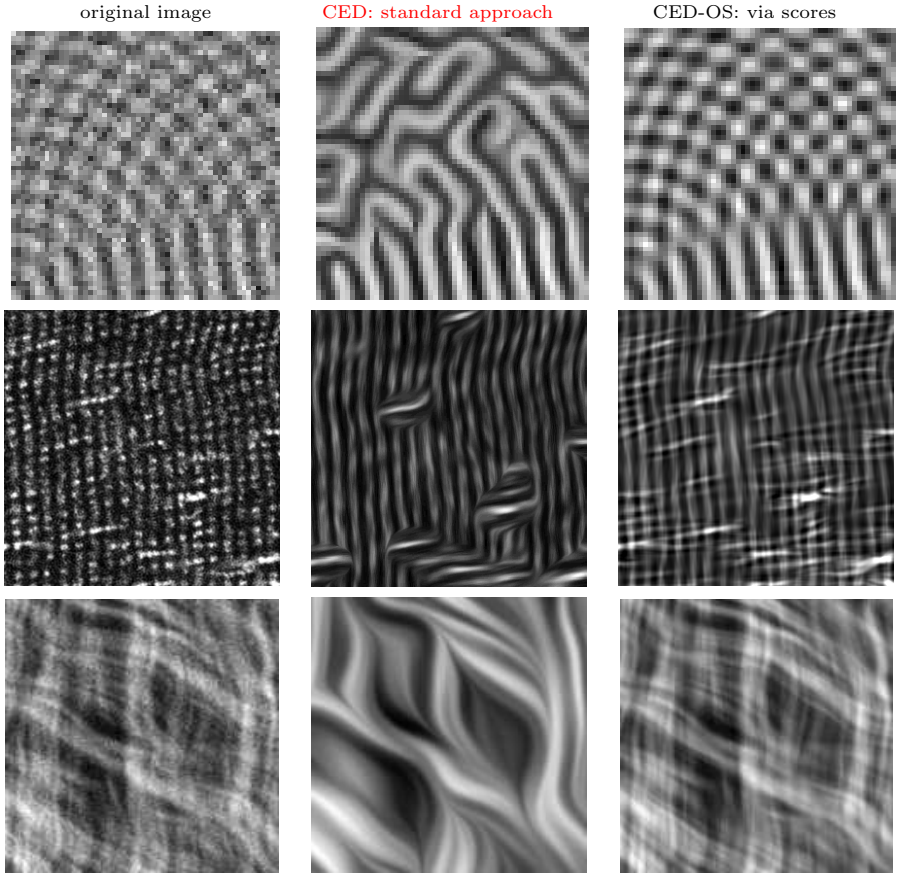

Fig. 12 *The problem of assigning a single frame per position in the vicinity of crossings, is clearly visible when applying the nonlinear diffusions (CED: Coherence Enhancing Diffusions) in the middle column. In a later section of the lecture notes, we will construct locally adaptive frames in orientation scores (multi-orientation representations of images) which allows us to make crossing-preserving coherence-enhancing diffusions. Left : original image. Middle: result of CED [22], see Eq. (61) and expressed in the locally adaptive frame in (62). Right: CED via invertible orientation score (CED-OS) cf. [28], [29, part II]. CED-OS is capable of handling crossings and bifurcations, whereas CED produces spurious artifacts at such junctions. For extensions of CEDOS to 3D images see [30], for extensions to multiple scales see [31].*

9 Variational Techniques for Image-Denoising

9.1 A Variational Method Connected to Linear Scale Space Diffusion via Resolvent Operators: Tikhonov regularization

Let us consider the following popular variational optimization for image denoising:

$$p_\alpha(f) := \underset{p\in\mathbb{L}_2(\mathbb{R}^d)}{\arg\min} \int_{\mathbb{R}^d} \alpha|f(\mathbf{x}) - p(\mathbf{x})|^2 + \|\nabla p(\mathbf{x})\|^2 \, \mathrm{d}\mathbf{x}, \tag{63}$$

with $\alpha > 0$ some a priori constant. Typically, the first term is referred to as the 'data term' as it includes the input data f and the second term is referred to as the 'smoothness prior' as in the minimization it takes care of a smoothness assumption on the output. Parameter $\alpha > 0$ below balances between data-fidelity and smoothness of the minimizer $p_\alpha(f)$.

This popular variational regularization method is a special case of 'Tikhonov regularization' (which in its general form also allows for other operators in the 'smoothness prior' rather than first order derivative operators).

Lemma 3 *Problem (63) has the following solution*

$$p_\alpha(f) = -\alpha(\Delta - \alpha I)^{-1} f = R_\alpha * f \tag{64}$$

with $R_\alpha \in \mathbb{L}_1(\mathbb{R}^d)$ the positive probability kernel (with $\mathbb{L}_1$-norm equal to 1) that arises by taking the Laplace transform of the Gaussian kernel w.r.t. scale, i.e.

$$R_\alpha(\mathbf{x}) = \alpha \int_0^\infty G_s(\mathbf{x})\, e^{-\alpha s}\, \mathrm{d}s. \tag{65}$$

Proof. The problem has a unique solution due to convexity of the mapping $f \mapsto p_\alpha(f)$, cf. [32]. It is thereby sufficient to study the Euler-Lagrange (EL) equations where the first order variation vanishes. We compute for arbitrary $\delta \in C_c^\infty(\mathbb{R}^d)$ (i.e. compactly supported smooth functions on $\mathbb{R}^d$)

$$\lim_{\epsilon \downarrow 0} \tfrac{p_\alpha(f+\epsilon\delta)-p_\alpha(f)}{\epsilon} = 2 \int_{\mathbb{R}^d} (\alpha(p-f) - \Delta p)\ \delta\, \mathrm{d}\mathbf{x} = 0$$

where we applied integration by parts in $\mathbb{R}^d$(Green's 1st identity). Now the above equation holds for all $\delta \in C_c^\infty(\mathbb{R}^d)$, and the latter function space is dense in $\mathbb{L}_2(\mathbb{R}^d)$, so we find EL-equation:

$$(\alpha I - \Delta)p = \alpha f$$

and the rest of the result follows by application of Laplace transform w.r.t. scale in the Fourier domain:

$$\begin{aligned}\mathcal{F}(p_\alpha f)(\boldsymbol{\omega}) = \mathcal{F}p(\boldsymbol{\omega}) &= \alpha(\alpha + \|\boldsymbol{\omega}\|^2)^{-1}\mathcal{F}f(\boldsymbol{\omega}) \\ &= \alpha(2\pi)^{\frac{d}{2}} \int_0^\infty \tfrac{e^{-s\|\boldsymbol{\omega}\|^2}}{(2\pi)^{\frac{d}{2}}}\, e^{-s\alpha}\, \mathrm{d}s\, \mathcal{F}f(\boldsymbol{\omega}) \\ &= (2\pi)^{\frac{d}{2}}\, \alpha \int_0^\infty (\mathcal{F}G_s)(\boldsymbol{\omega})\, e^{-s\alpha}\, \mathrm{d}s\, \mathcal{F}f(\boldsymbol{\omega})\end{aligned}$$

from which both (64), (65) and $\|R_\alpha\|_{\mathbb{L}_1(\mathbb{R}^d)} = 1$ and $R_\alpha \geq 0$ follows, by application of inverse Fourier transform. where we recall (7) and (8). □

Remark 18 There is a basic stochastic interpretation to the above formula. Let $(\mathbf{X}_S)_{S\geq 0}$ denote a Wiener process with zero mean and variance $S = \frac{1}{2}\sigma^2$. At time 0 we have a grey-scale particle at the origin $\mathbf{0} \in \mathbb{R}^d$ as $X_0 \sim \delta_{\mathbf{0}}$, then as scale/time $S > 0$ increases the probability density of finding the particle at position $\mathbf{x}$ at time $S = s$ equals

$$P(\mathbf{X}_S = \mathbf{x} \mid S = s) = G_s(\mathbf{x}).$$

Now in such a Markov traveling time is memoryless and the only continuous memoryless distribution is the negatively exponential distribution $S \sim NE(\alpha)$, i.e. $P(S = s) = \alpha e^{-\alpha s}$ for all $s \geq 0$. Now the probability density of finding a greyscale particle that started at the origin $\mathbf{0}$ at scale/time 0 regardless its traveling time at location $\mathbf{x}$ can be rewritten as as

$$P(\mathbf{X}_S = \mathbf{x} \mid S \sim NE(\alpha)) = \int_{\mathbb{R}^+} P(\mathbf{X}_S = \mathbf{x} \mid S = s)\, P(S = s)\, \mathrm{d}s.$$

where we consider a sub-ordinated random process where traveling time is random: $S \sim NE(\alpha)$.

Now assume $X_0 \sim f$ (like in Figure 1) instead of $X_0 \sim \delta_{\mathbf{0}}$ while retaining the same Gaussian transition density. Then we find by linearity and shift-invariance of the mapping $f \mapsto P_\alpha(f)$ that:

$$\begin{aligned}(p_\alpha(f))(\mathbf{x}) &= P(\mathbf{X}_S = \mathbf{x} \mid \mathbf{X}_0 \sim f, S \sim NE(\alpha)) = \int_0^\infty P(\mathbf{X}_S = \mathbf{x} \mid \mathbf{X}_0 \sim f, S = s)\ P(S = s)\, \mathrm{d}s \\ &= \int_0^\infty \left(\int_{\mathbb{R}^d} G_s(\mathbf{x} - \tilde{\mathbf{x}}) f(\tilde{\mathbf{x}})\, \mathrm{d}\tilde{\mathbf{x}} \right) \alpha e^{-\alpha s} \mathrm{d}s = \int_0^\infty ((G_s * f)(\mathbf{x}))\ \alpha e^{-\alpha s} \mathrm{d}s = \alpha \int_0^\infty u(\mathbf{x}, s) e^{-\alpha s} \mathrm{d}s \\ &= \left(\left(\int_0^\infty \alpha e^{-\alpha s} G_s \mathrm{d}s \right) * f \right)(\mathbf{x}) = (R_\alpha * f)(\mathbf{x}).\end{aligned} \tag{66}$$

Remark 19 Although not part of this course, we would like to mention that automatic choice of $\alpha > 0$ can be done in many ways, such as generalized cross-validation [20] or L-curve optimizations [33] or discrepancy-principles [34].

Remark 20 It is for application sometimes wise to take a Gamma-distribution $S \sim \Gamma(s;k,\alpha)$ with expectation $E(S) = \frac{k}{\alpha}$ instead of a negatively exponential distribution for S, in order to 'push more mass/ink' forward cf. [35–37]. This means that

$$P(S=s) = \Gamma(s;k,\alpha) := \frac{\alpha^k\, s^{k-1}}{\Gamma(k)} e^{-\alpha s} \Leftrightarrow S = \sum_{i=1}^{k} S_i, \text{ with independent } S_i \sim NE(\alpha). \tag{67}$$

Exercise 22 Check the equivalence in (67) and show that (67) boils down to iteration of (66), i.e. a k-fold iteration of the kernels:

$$(p_\alpha)^k(f) = (p_\alpha \circ \ldots \circ p_\alpha)(f) = (R_\alpha *^{(k-1)} R_\alpha) * f.$$

What does this mean for the probabilistic interpretation?

Remark 21 (discretization) If we constrain the constrain the set of continuous images from $\mathbb{L}_2(\mathbb{R}^d)$ to the span V of B-spline functions given by (23), and we replace the continuous data-term by its sampled version

$$\|\alpha - f\|^2_{\mathbb{L}_2(\mathbb{R}^d)} \to \sum_{i=1}^{N}\sum_{j=1}^{N} |\alpha(\mathbf{x}_{ij}) - f(\mathbf{x}_{ij})|$$

then the matrix-representation of the corresponding linear operator (where both domain and range are constrained to V) is given by Eq. (24).

What is also interesting is to consider the problem where the gradient norm is minimized rather than the gradient norm squared. This gives the following optimization problem:

$$\text{“}q_\alpha(f) := \arg\min_{q\in V} \int_\Omega \sqrt{\alpha}|f(\mathbf{x}) - q(\mathbf{x})|^2 + \|\nabla q(\mathbf{x})\|\, \mathrm{d}\mathbf{x}'' \tag{68}$$

over a suitable space V consisting of regular enough images f, with $\Omega \subseteq \mathbb{R}^d$. To ensure the same physical dimensions as with α in (63), we apply the square root on α. In Figure 13 we compare $q_\alpha(f)$ with $p_\alpha(f)$ for a fixed choice of $\alpha > 0$ and one typically observes that using an $\mathbb{L}_1$-norm for the gradient in the regularization term is more effective to remove salt& pepper noise while preserving the actual larges structures in the image. This is due to the fact that an $\mathbb{L}_1$-norm is less sensitive to single outliers than an $\mathbb{L}_2$-norm, and the fact that Total Variation (TV) with the $\mathbb{L}_1$-norm allows for a much wider class of functions including plateau functions with step-edges, which often occur in regular images.

The formulation of TV optimization with an $\mathbb{L}_1$-norm is therefore subtle, and in fact it requires a reformulation of (75), as gradients may not exist for elements from the space V that we will characterize next. Therefore we wrote quotes in (75) as formally it is not a valid formulation when using regular derivatives.

9.2 Total Variation and Image Denoising

Let $d \in \{2,3\}$. In contrast to the previous section we will not consider the domain $\mathbb{R}^d$, but rather a bounded and open domain $\Omega \subset \mathbb{R}^d$. For simplicity we assume Ω is the interior of a rectangular domain if $d = 2$ or the interior of a cubic domain if $d = 3$.

Let $\mathbf{C}_c^\infty(\Omega)$ denote the space vector fields on Ω which are infinitely differentiable with compact support within Ω. Then the total variation on Ω is given by

$$TV(f) := \sup_{\mathbf{v}\in C_c^\infty(\Omega)} \left\{ \int_\Omega f(\mathbf{x}) \operatorname{div} \mathbf{v}(\mathbf{x})\, \mathrm{d}\mathbf{x} \mid \forall_{\mathbf{x}\in\Omega} \|\mathbf{v}(\mathbf{x})\| \leq 1 \right\}$$

By definition the space of bounded variation functions are all elements in $\mathbb{L}_1(\Omega)$ such that $TV(f) < \infty$:

$$BV(\Omega) := \{f \in \mathbb{L}_1(\Omega) \mid TV(f) < \infty\}$$

The space $BV(\Omega)$ is standardly equipped with the bounded variation norm given by

$$\|f\|_{BV} = \|f\|_{\mathbb{L}_1(\Omega)} + TV(f).$$

Note that for constant $f = C$ one has that $TV(f) = 0$ and $\|f\|_{\mathbb{L}_1(\Omega)} = C$.

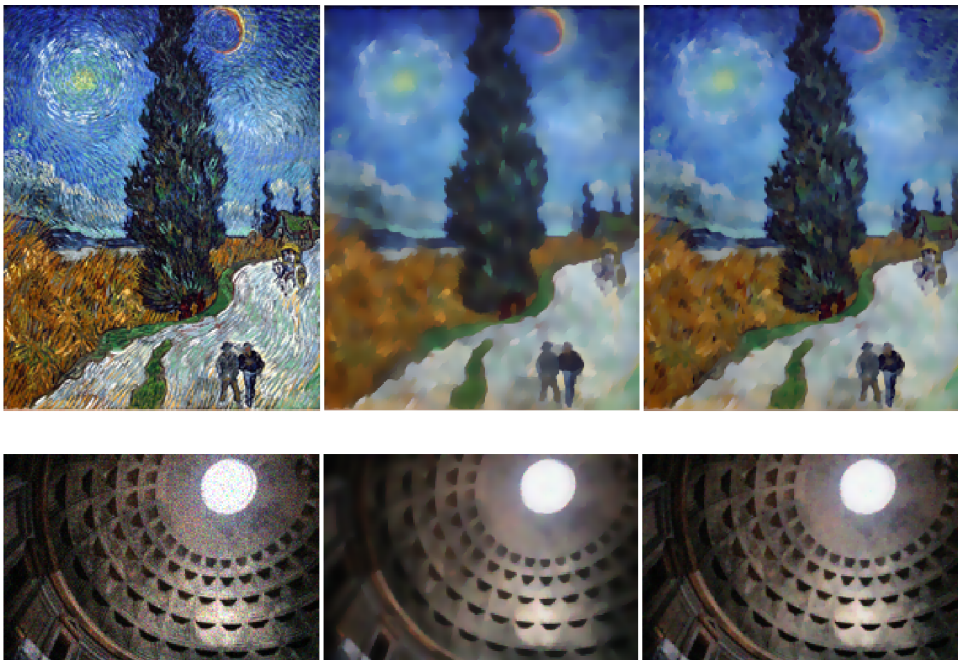

Fig. 13 Left: original images f, middle $p_\alpha(f)$ given by (63), right $q_\alpha(f)$ given by (75) or rather (71) for a fixed large value of $\alpha > 0$, so that the regularization effect is not too strong.

Intermezzo: Basic Properties of functions of bounded variations

Next we mention some basic properties on functions of bounded variations that are proven in standard works such as in [38]. We will use these properties implicitly, and when needed you can use these properties without proof in exercises if needed. It can be shown that $(BV(\Omega), \|\cdot\|_{BV})$ is a Banach space. Even a Banach algebra as one has $f, g \in BV(\Omega) \Rightarrow f \cdot g \in BV(\Omega)$.

Regular derivatives of functions of bounded variation may not exist (they may have 'jump-type' or 'removable' discontinuities), but their derivative [11] does exist almost everywhere.

Furthermore, $C_1(\Omega)$ is dense within $BV(\Omega)$, and every function $f \in BV(\Omega)$ of bounded variation can even be written as a limit (in $\mathbb{L}_1$-sense) of a sequence of infinitely differentiable compactly supported functions, cf. [38, ch:5.2, Thm.2].

Finally, a function of bounded variation $f \in BV(\Omega)$ can be restricted $f|_{\partial\Omega}$ in a well-posed way[12] to the boundary $\partial\Omega$.

Exercise 23 (Total Variation for continuously differentiable functions)
Let $d = 3$. Show that

$$f \in C^1(\Omega) \Rightarrow 0 \leq TV(f) = \int_\Omega \|\nabla f(\mathbf{x})\| \, \mathrm{d}\mathbf{x} = \int_\Omega (\nabla f)(\mathbf{x}) \cdot \frac{\nabla f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|} \, \mathrm{d}\mathbf{x}. \tag{69}$$

hints: In order to show the above equality you should rely on

1. *the standard formula* $\mathrm{div}\,(f\mathbf{v}) = f\mathrm{div}\,\mathbf{v} + \nabla f \cdot \mathbf{v}$.
2. *the fact that the Cauchy-Schwarz inequality holds with equality if and only if the vectors are linearly dependent.*
3. *the Gauss divergence theorem.*

[11] More formally, for all $f \in BV(\Omega)$ one has $\int_\Omega f \mathrm{div}\,\mathbf{v}\,\mathrm{d}\mathbf{x} = -\int_\Omega \mathbf{v}\cdot\boldsymbol{\sigma}\mathrm{d}\mu_f$ for all $\mathbf{v} \in \mathbf{C}_c^1(\mathbb{R}^d)$ for a Radon measure μ_f on Ω and a μ-measurable function $\boldsymbol{\sigma} : \Omega \to \mathbb{R}^d$ with $\|\boldsymbol{\sigma}\| = 1$. The measure μ_f is often denoted by $\|Df\|$. See [38, Thm1, Ch:5]. If $f \in C^1(\Omega)$ then $\boldsymbol{\sigma}\mathrm{d}\mu_f = \nabla f(\mathbf{x})\,\mathrm{d}\mathbf{x}$.

[12] As there exist trace theorems for functions of bounded variations [38, ch:5.3, Thm.1]. Recall, that in general $\mathbb{L}_1(\Omega)$-elements cannot be constrained to $\partial\Omega$ as it is a set with zero measure.

Let $C_c^\infty(\Omega)$ denote the vector space of real-valued functions on Ω which are infinitely differentiable with compact support (strictly contained within Ω). For the sake of simplicity (application of Gauss theorem) we restrict ourselves to the case $d = 3$, although the result does not change for $d = 2$.

Lemma 4 *Let $d = 3$: $\Omega \subset \mathbb{R}^3$. For all $f \in C^1(\Omega)$ and all $\delta \in C_c^\infty(\Omega)$ we have that*

$$\lim_{\epsilon \downarrow 0} \frac{TV(f+\epsilon\delta) - TV(f)}{\epsilon} = -\int_\Omega \delta(\mathbf{x}) \operatorname{div}\left(\frac{\nabla f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|}\right) \mathrm{d}\mathbf{x} =: \langle (\partial TV)(f), \delta \rangle.$$

Proof. We apply the previous result (69) and start some direct computations where in the numerator only the $O(\epsilon)$-term is relevant (and where we apply $\sqrt{1+\epsilon} = 1 + \frac{1}{2}\epsilon + O(\epsilon^2)$):

$$\begin{aligned}
\lim_{\epsilon \downarrow 0} \int_\Omega \tfrac{\|\nabla(f+\epsilon\delta)\| - \|\nabla f\|}{\epsilon} \,\mathrm{d}\mathbf{x} &= \lim_{\epsilon \downarrow 0} \int_\Omega \tfrac{\sqrt{\nabla f \cdot \nabla f + 2\epsilon \nabla\delta \cdot \nabla f + O(\epsilon^2)} - \sqrt{\nabla f \cdot \nabla f}}{\epsilon} \,\mathrm{d}\mathbf{x} \\
&= \lim_{\epsilon \downarrow 0} \int_\Omega \tfrac{\|\nabla f\| \left(\sqrt{1 + 2\epsilon \nabla\delta \cdot \frac{\nabla f}{\|\nabla f\|}} - 1\right)}{\epsilon} \mathrm{d}\mathbf{x} \\
&\overset{*}{=} \int_\Omega \nabla\delta \cdot \tfrac{\nabla f}{\|\nabla f\|} \,\mathrm{d}\mathbf{x} \\
&= \int_\Omega \operatorname{div}(\delta \tfrac{\nabla f}{\|\nabla f\|}) \,\mathrm{d}\mathbf{x} - \int_\Omega \delta \operatorname{div}\left(\tfrac{\nabla f}{\|\nabla f\|}\right) \mathrm{d}\mathbf{x} \\
&= \int_{\partial\Omega} \delta \|\nabla f\|^{-1} \tfrac{\partial f}{\partial n} \,\mathrm{d}\sigma + - \int_\Omega \delta \operatorname{div}\left(\tfrac{\nabla f}{\|\nabla f\|}\right) \mathrm{d}\mathbf{x} \\
&= 0 - \int_\Omega \delta \operatorname{div}\left(\tfrac{\nabla f}{\|\nabla f\|}\right) \mathrm{d}\mathbf{x}.
\end{aligned}$$

where in the third step we rely on Lebesgue dominated convergence principle when switching the limit and the integration. □

Corollary 1 *Let $f \in \mathbb{L}_1(\Omega) \cap \mathbb{L}_2(\Omega)$. Then the nonlinear PDE-system associated with a gradient flow of total variation (known as "Total Variation Flow") is given by*

$$\begin{cases} \frac{\partial v}{\partial s} = \operatorname{div}\left(\frac{\nabla v}{\|\nabla v\|}\right), \\ v(\cdot, 0) = f(\cdot)\ , \\ v(\mathbf{x}, 0) = 0 \text{ for all } \mathbf{x} \in \partial\Omega. \end{cases} \tag{70}$$

Her v denotes the unique "entropy solution" (for the precise details and definition see [39]) of the PDE. The corresponding unique[13] *solution of the denoising problem*

$$q_\alpha = \underset{q \in BV(\Omega),\, q|_{\partial\Omega} = 0}{\arg\min} \frac{\sqrt{\alpha}}{2} \int_\Omega |q(\mathbf{x}) - f(\mathbf{x})|^2 \,\mathrm{d}\mathbf{x} + TV(q) \tag{71}$$

satisfies the (non-linear) Euler Lagrange equation:

$$q_\alpha(\mathbf{x}) - \alpha^{-\frac{1}{2}} \operatorname{div}\left(\frac{\nabla q_\alpha(\mathbf{x})}{\|\nabla q_\alpha(\mathbf{x})\|}\right) = f. \tag{72}$$

for almost every $\mathbf{x} \in \Omega$ where $\nabla q_\alpha(\mathbf{x}) \neq 0$.

Exercise 24 Explain why $q_\alpha(\mathbf{x})$ is *not proportional to* the Laplace tranform w.r.t. $s > 0$ of $v(\mathbf{x}, \cdot)$, whereas $p_\alpha(\mathbf{x})$ is proportional to the Laplace tranform w.r.t. $s > 0$ of $u(\mathbf{x}, \cdot)$, recall (66).

Exercise 25 Show that an implicit finite difference scheme of (70) requires to solve at each time step an equation of the type (72).

[13] The mapping $q \mapsto (\sqrt{\alpha}/2) \int_\Omega |q(\mathbf{x}) - f(\mathbf{x})| \mathrm{d}\mathbf{x} + TV(q)$ is convex.

Remark 22 In order to avoid numerical problems at singular points $\mathbf{x}_0 \in \mathbb{R}^d$ where $\|\nabla u(\mathbf{x}_0)\| = 0$ and where in the PDE both the numerator and denominator vanish it is common to apply the following adaptations

$$\begin{cases} \frac{\partial u_\sigma}{\partial s} = \operatorname{div}\left(\frac{\nabla u_\sigma}{\sqrt{\|\nabla u_\sigma\|^2+\sigma^2}} \right), \\ u_\sigma(\cdot,0) = f(\cdot). \end{cases}$$

for $0 < \sigma \ll 1$, and to take the pointwise limit $u(\mathbf{x},s) = \lim_{\sigma \downarrow 0} u_\sigma(\mathbf{x},s)$ afterwards. Similarly one considers,

$$q_{\alpha,\sigma} - \alpha^{-\frac{1}{2}} \operatorname{div}\left(\frac{\nabla q_{\alpha,\sigma}}{\sqrt{\|\nabla q_{\alpha,\sigma}\|^2 + \sigma^2}} \right) = f. \tag{73}$$

For $f \in \mathbb{C}^1(\mathbb{R})$ equation (73) can be rewritten as

$$\lim_{\epsilon \downarrow 0} \frac{\widetilde{TV}_\sigma(q+\epsilon\delta) + (\sqrt{\alpha}/2)\|(q+\epsilon\delta) - f\|^2_{\mathbb{L}_2(\mathbb{R}^d)} - (\widetilde{TV}_\sigma(q) + (\sqrt{\alpha}/2)\|q-f\|^2_{\mathbb{L}_2(\mathbb{R}^d)})}{\epsilon} = 0.$$

More formally, Eq. (73), is an Euler-Lagrange equation for the optimization:

$$q_{\alpha,\sigma} = \underset{q \in BV(\Omega)}{\arg\min} \left(\sqrt{\alpha}\|q-f\|^2 + \widetilde{TV}_\sigma(q) \right),$$

with $\widetilde{TV}_\sigma(f)$ given by

$$\widetilde{TV}_\sigma(f) := \sup_{\substack{\mathbf{w} \in \mathbf{C}_c^\infty(\Omega), \\ \sigma_2 \in C_c^\infty(\Omega)}} \left\{ \int_\Omega \begin{pmatrix} f(\mathbf{x}) \\ \sigma \end{pmatrix} \cdot \begin{pmatrix} \operatorname{div} \mathbf{w}(\mathbf{x}) \\ \sigma_2(\mathbf{x}) \end{pmatrix} \mathrm{d}\mathbf{x} \mid \forall_{\mathbf{x}\in\Omega} \|\mathbf{w}(\mathbf{x})\|^2 + |\sigma_2(\mathbf{x})|^2 \leq 1 \right\}, \tag{74}$$

which for $f \in C^1(\Omega)$ simplifies to

$$f \in C^1(\Omega) \Rightarrow \widetilde{TV}_\sigma(f) = \int_\Omega \sqrt{\|\nabla f(\mathbf{x})\|^2 + \sigma^2}\, \mathrm{d}\mathbf{x}. \tag{75}$$

Exercise 26 Use definition (74) to show that (75) indeed holds.
hint: Use the hints/techniques of Exercise 23.

Exercise 27 Recall that functions of bounded variation are $f \in L_1(\Omega)$ with $TV(f) < \infty$. Let $0 \leq \sigma \leq 1$. Show that

$$TV(f) < \infty \Leftrightarrow \widetilde{TV}_\sigma(f) < \infty$$

for all $f \in C^1(\Omega)$.
hint: for $\|\nabla f\| \gg \sigma^2$ one has $\sqrt{\|\nabla f\|^2 + \sigma^2} = \|\nabla f\| + O\left(\frac{\sigma^2}{2\|\nabla f\|}\right)$.

Exercise-Mathematica 4. *Download the Mathematica notebook "Lecture1_TVdenoising.nb" from* http://www.lieanalysis.nl.201-2/.

You can run the notebook by selecting the cells and pressing 'shift' and 'enter'.
Check the behavior for $\sigma \downarrow 0$ and compare the TV-based denoising with the (adaptive) diffusion notebooks from the previous section.
What is the main difference in practice?

The next exercise (on the next page) exceeds the scope of this lecture, but justifies taking the limit $\sigma \downarrow 0$ in the total variation regularization (and in the PDE-algorithms implementing it), as it shows that

$$\lim_{\sigma \downarrow 0} q_{\alpha,\sigma} = q_\alpha \text{ in } \mathbb{L}_1(\Omega).$$

Exercise 28 (an alternative TV-formulation of $\widetilde{TV}_\sigma(f)$).
Consider $f^\sigma : \Omega \times [0,1] \to \mathbb{R}$ given by

$$f^\sigma(\mathbf{x}, x^{d+1}) := f(\mathbf{x}) + \sigma x^{d+1},$$

with $\mathbf{x} \in \Omega \subset \mathbb{R}^d$ and $0 < x^{d+1} < 1$, and consider

$$TV(f^\sigma) = \sup_{\tilde{\mathbf{w}} \in C_c^\infty(\Omega \times [0,1])} \left(\int_0^1 \int_\Omega f^\sigma(\mathbf{x}, x^{d+1}) \text{ div } \tilde{\mathbf{w}}(\mathbf{x}, x^{d+1}) \, \mathrm{d}\mathbf{x}\mathrm{d}x^{d+1} \mid \forall_{\mathbf{x} \in \Omega} \ : \ \tilde{\mathbf{w}}(\mathbf{x}) = 1 \right). \tag{76}$$

a.) Show that $TV(f^\sigma) \geq \widetilde{TV}_\sigma(f)$.

hint: restrict to specific $\mathbf{w}$ *(which are essentially 1-homogenous in the second entry) as below:*

$$\tilde{\mathbf{w}}(\mathbf{x}, x^{d+1}) = (w^1(\mathbf{x}, x^{d+1}), \cdots, w^d(\mathbf{x}, x^{d+1}) \ , \ w^{d+1}(\mathbf{x}, x^{d+1})) = \ \eta(x^{d+1}) \, (\mathbf{v}(\mathbf{x}), \sigma_2(\mathbf{x})), \tag{77}$$

with $\eta(1) = 0$, $\eta(0) = 0$, *and* $\eta \in C^\infty([0,1])$ *in such a way that* η *approximates* $1_{(0,1)}$ *uniformly on the open interval* $(0,1)$.

b.) Show that $TV(f^\sigma) = \widetilde{TV}_\sigma(f)$.

hint: Consider $(\mathbf{w}(\mathbf{x}), \sigma_2(\mathbf{x})) := \int_0^1 \tilde{\mathbf{w}}(\mathbf{x}, s) \, \mathrm{d}s$.

Exercise 29 (convergence)

a.) Show that if a sequence $(f_n)_{n \in \mathbb{N}}$ in $\mathbb{L}_1(\Omega)$ has the property that *every* subsequence $(f_{n_k})_{k \in \mathbb{N}}$ has a converging sub-subsequence $(f_{n_{k_l}})_{l \in \mathbb{N}}$ with the same limit $f^* \in \mathbb{L}_1(\Omega)$ then the sequence itself converges to this limit, i.e. then $f_n \to f$ in $\mathbb{L}_1(\Omega)$ as $(n \to \infty)$.

hint: this applies to metric spaces in general and the proof follows by contradiction.

b.) Recall that q_α was the solution to the total variation regularization problem (71). Now consider its approximation

$$q_{\alpha,\sigma} := \underset{q \in BV(\Omega)}{\arg\min} \ \frac{1}{2} \int_\Omega |q(\mathbf{x}) - f(\mathbf{x})|^2 \, \mathrm{d}\mathbf{x} + \alpha^{-\frac{1}{2}} \, \widetilde{TV}_\sigma(q)$$

Show that

$$\boxed{\lim_{\sigma \downarrow 0} q_{\alpha,\sigma} = q_\alpha \ \text{ in } \mathbb{L}_1(\Omega).}$$

where the limit is taken in $\mathbb{L}_1$-sense.

hint1: there exist an Arzéla-Ascoli type of theorem for functions of bounded variation:
Known as "Helly's selection criterium" (Wikipedia) or [38, Thm.4] which states the following when applied to our choice $\Omega \subset \mathbb{R}^n$*:*

Theorem 1 *Assume* $(f_k)_{k \in \mathbb{N}}$ *is a sequence in* $BV(\Omega)$ *satisfying*

$$\sup_{k \in \mathbb{N}} TV(f_k) < \infty$$

Then there exists a subsequence $(f_{k_j})_{j \in \mathbb{N}}$ *and a function* $f \in BV(\Omega)$ *such that*

$$\lim_{j \to \infty} f_{k_j} = f \in \mathbb{L}_1(\Omega) \ \textit{i.e.} \ \lim_{j \to \infty} \|f_{k_j} - f\|_{\mathbb{L}_1(\Omega)} = 0.$$

hint2: Use the previous exercise, and apply the standard result that a total variation function $TV(\cdot)$ *is lower-semi-continuous on the set of bounded variation functions, to deduce*

$$\lim_{n \to \infty} f_n = f \Rightarrow \liminf_{n \to \infty} \widetilde{TV}_\sigma(f_n) \geq \widetilde{TV}_\sigma(f).$$

hint3: Use the previous item 29a, where you may use the fact that Problem (71) has a unique solution.

References

1. B. J. Janssen, R. Duits, and B. M. ter Haar Romeny, "Linear image reconstruction by sobolev norms on the bounded domain.," in *Proc. of the first International Conference on Scale Space and Variational Methods in Computer Vision, Lecture Notes in Computer Science, Vol. 4485, Springer* (M. Sgallari and Paragios, eds.), pp. 461–472, June 2007.
2. R.Duits, *Percptional Organization in Image Analysis: A Mathematical Approach Based on Scale, Orientation and Curvature.* PhD thesis, TU/e, 2005. *http://bmia.bmt.tue.nl/people/RDuits/THESISRDUITS.pdf.*
3. B. J. Janssen, *Representation and Manipulation of Images Based on Linear Functionals.* PhD thesis, Eindhoven University of Technology, Department of Biomedical Engineering and Department of Computer Science and Mathematics, Eindhoven, The Netherlands, 2010. *http://www.win.tue.nl/casa/research/casaprojects/janssen.html.*
4. T. Iijima, "Basic theory of pattern observation." Papers of Technical Group on Automata and Automatic Control, 1959. (in Japanese).
5. T. Iijima, "Basic theory on normalization of a pattern (in case of typical one-dimensional pattern)," *Bulletin of Electrical Laboratory*, vol. 26, pp. 368–388, 1962. (in Japanese).
6. R. Duits, L. Florack, J. de Graaf, and B. ter Haar Romeny, "On the axioms of scale space theory," *Journal of Mathematical Imaging and Vision*, vol. 20, pp. 267–298, May 2004.
7. M. Felsberg and G. Sommer, "The Poisson scale-space: A unified approach to phase-based image processing in scale-space," *Journal of Mathematical Imaging and Vision*, vol. 21, no. 1-2, pp. 5–26, 2004.
8. M. Felsberg and G. Sommer, "Scale adaptive filtering derived from the Laplace equation," in *23. DAGM Symposium Mustererkennung, München* (B. Radig and S. Florczyk, eds.), vol. 2191 of *Lecture Notes in Computer Science*, pp. 124–131, Springer, Heidelberg, 2001.
9. M. Felsberg, R. Duits, and L. Florack, "The monogenic scale space on a bounded domain and its applications," *Proceedings Scale Space Conference, volume 2695 of Lecture Notes of Computer Science, Springer, Isle of Skye, UK.*, pp. 209–224, June 2003.
10. E. J. Pauwels, L. J. V. Gool, P. Fiddelaers, and T. Moons, "An extended class of scale-invariant and recursive scale space filters," vol. 17, pp. 691–701, July 1995.
11. R. Duits, M. Felsberg, and L. M. J. Florack, "α scale spaces on a bounded domain," *Proceedings Scale Space Conference, Isle of Skye,UK.*, pp. 494–510, June 2003.
12. W. Feller, *An Introduction to Probability Theory and its Applications*, vol. 2. New York, London, Sydney: John Wiley and sons, Inc., 1966.
13. J. J. Koenderink, "The structure of images," *Biological Cybernetics*, vol. 50, pp. 363–370, 1984.
14. T. Lindeberg, *Scale-Space Theory in Computer Vision.* Kluwer Academic Publishers, 1993.
15. J. Damon, "Local Morse theory for solutions to the heat equation and Gaussian blurring," vol. 115, pp. 368–401, January 1995.
16. R. Duits, *Perceptual Organization in Image Analysis.* PhD thesis, Eindhoven University of Technology, Dep. of Biomedical Engineering, The Netherlands, 2005. *http://yp.wtb.tue.nl/pdfs/5474.pdf* , cum laude and selected for promotion prize ($\leq$ 2% best at TU/e).
17. T. Lindeberg, "Edge detection and ridge detection with automatic scale selection," *IJCV*, vol. 30, 1998.
18. M. Unser, "Splines: A perfect fit for signal and image processing," vol. 16, pp. 22 – 38, 12 1999.
19. E. Bekkers, *Retinal Image Analysis using Sub-Riemannian Geometry in $SE(2)$.* PhD thesis, Eindhoven University of Technology, 2017. cum laude ($\leq$ 5% best at TU/e), URL: *https://pure.tue.nl/ws/files/52750592/20170123_Bekkers.pdf.*
20. E. Bekkers, M. Loog, B. ter Haar Romeny, and R. Duits, "Template matching via densities on the roto-translation group," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, pp. 452–466, februari 2018.
21. B. Hall, *Lie Groups, Lie Algebras, and their Representations: An elementary Introduction.* Springer, 2003.
22. J. Weickert, "Coherence-enhancing diffusion filtering.," *International Journal of Computer Vision*, vol. 31, no. 2/3, pp. 111–127, 1999.
23. P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, 1990.
24. S. Esedoglu, "An analysis of the perona-malik scheme," *Communications on Pure and Applied Mathematics*, vol. 54, pp. 1442–1487, 12 2001.
25. A. Buades, B. Coll, and J. M. Morel, "The staircasing effect in neighborhood filters and its solution," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1499–1505, 2006.
26. P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud, "Two deterministic half-quadratic regularization algorithms for computed imaging," in *Proceedings of 1st International Conference on Image Processing*, vol. 2, pp. 168–172, 1994.
27. T. Brox, J. Weickert, B. Burgeth, and P. Mrzek, "Nonlinear structure tensors," *Image and Vision Computing*, vol. 24, no. 1, pp. 41 – 55, 2006.
28. E. M. Franken and R. Duits, "Crossing-preserving coherence-enhancing diffusion on invertible orientation scores," *International Journal of Computer Vision (IJCV)*, vol. 85, no. 3, pp. 253–278, 2009.
29. R. Duits and E. M. Franken, "Left invariant parabolic evolution equations on $SE(2)$ and contour enhancement via invertible orientation scores, part II: Nonlinear left-invariant diffusion equations on invertible orientation scores," *Quarterly of Applied mathematics, AMS*, vol. 68, pp. 293–331, June 2010.
30. M. H. J. Janssen, A. J. E. M. Janssen, E. J. Bekkers, J. O. Bescós, and R. Duits, "Design and processing of invertible orientation scores of 3d images," *Journal of Mathematical Imaging and Vision*, Mar 2018.
31. U. Sharma and R. Duits, "Left-invariant evolutions of wavelet transforms on the similitude group," *ACHA*, vol. 39, pp. 110–137, 2015.
32. L. Evans, *Partial Differential Equations (2nd edition).* Graduate Studies in Mathematics vol 19 AMS, 2010.
33. P. C. Hansen, "The l-curve and its use in the numerical treatment of inverse problems," in *in Computational Inverse Problems in Electrocardiology, ed. P. Johnston, Advances in Computational Bioengineering*, pp. 119–142, WIT Press, 2000.

34. O. Scherzer, "The use of morozov's discrepancy principle for tikhonov regularization for solving nonlinear ill-posed problems," *Computing*, vol. 51, pp. 45–60, Mar 1993.
35. J. Zhang, B. Dashtbozorg, E. Bekkers, J. Pluim, R. Duits, and B. ter Haar Romeny, "Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores," *IEEE-TMI*, vol. 35, no. 12, pp. 2631–2644, 2016.
36. R. Duits, E. Creusen, A. Ghosh, and T. Dela Haije, "Morphological and linear scale spaces for fiber enhancement in dw-mri," *Journal of Mathematical Imaging and Vision*, vol. 46, pp. 326–368, July 2013.
37. J. M. Portegies and R. Duits, "New exact and numerical solutions of the (convection-) diffusion kernels on SE(3)," *Differential Geometry and its Applications*, vol. 53, pp. 182–219, 2017.
38. L. Evans and R. Gariepy, *Measure Theory and Fine Properties of Functions*. Studies in Advanced Mathematics, CRC Press, 1992.
39. F. Andreu, C. Ballester, V. Caselles, and M. Mazon, "The dirichlet problem for the total variation flow," *Journal of Functional Analysis*, vol. 180, no. 2, pp. 347 – 403, 2001.