Dissertation
submitted to the
Combined Faculties for the
Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Put forward by

Diplom-Physicist: Gerald Stephen Kigathi Mwangi
Born in: Nairobi, Kenya

Oral examination: 26.07.2017

An Extension To The Principle Of Least Action For Gibbs Random Fields

Referees: PD. Dr. Christoph S. Garbe

Prof. Dr.   Jürgen Hesser

for Helen ...

Zusammenfassung

In dieser Arbeit untersuchen wir Methoden zur Modellierung von Low-Level-Computer-Vision-Problemen mit Gibbs'sche Zufalls Modelle. In solchen Modellen ist die Aufgabe, den Wert eines nicht beobachtbaren Variablen, gennant Gibbs Zufallsfeld (GRF), abzuleiten gegeben beobachtete Daten und eine an die GRF verknüpfte Energiefunktion. Dieser Vorgang wird als Prinzip der kleinsten Wirkung (PKW) bezeichnet.

GRF-Energiefunktionale bestehen im Allgemeinen aus zwei additiven Terme, dem Datenterm und der prior Energie. Es wird argumentiert, dass der Datenterm die GRF zu wenigen eindeutigen Werte während des Inferenzprozesses zwingen muss. Auf der anderen Seite soll die prior Energie weniger restriktiv und invariant zu einer breiten Klasse von Transformationen sein, einer sogenannten Lie-Gruppe, die die codierten intrinsischen Einschränkungen erhält, um einen guten Kompromiss mit dem Datenterm während des Inferenzprozesses zu ermöglichen.

Wir besprechen den bekannten Total Variation Prior (TV) und führen einen neuartigen Struktur Tensor (ST) basierten Prior ein. Beide Priors bestrafen Level-Sets mit von Null verschiedenen Krümmung aufgrund ihrer Invarianz unter lokalen Rotationen und Translationen.

Wir wenden den TV und den ST Prior im Rahmen eines uni-modalen und multi-modalen optischen Flusses an, für den wir auch einen neuartigen Ähnlichkeitsmaß entwickelt haben. Die Ergebnisse zeigen, dass der TV Prior numerisch zu besseren Flussschätzungen im uni-modalen Fall und im multimodalen Fall zu optisch höherwertigen optischen Flüsse führt.

Schließlich entwickeln wir eine Erweiterung der PLA, die Variationen der räumlichen Domäne berücksichtigt, die den erweiterten Least Action Algorithm (ELAA) genannt wird. Die Ergebnisse zeigen, dass das TV-basierte Bildentrauschung mit dem ELAA deutlich schneller und robuster als herkömmliche Methoden ist. Auf der anderen Seite weist die ELAA beim Bildentrauschung mit dem ST Prior nur bescheidene Beschleunigungen auf. Wir zeigen, dass die schwache Leistung des ST Priors mit seiner Unfähigkeit verbunden ist, die Krümmung der Level-Sets korrekt zu messen.

Abstract

In this thesis we investigate methods of modeling low level computer vision problems with Gibbs Random Models. In such models the task is to infer the value of an unobservable variable called a Gibbs Random Field (GRF) given observed data and an energy functional asserted to the GRF. This process is called the principle of least action (PLA).

GRF energy functionals generally consist of two additive terms, the data term energy and the prior energy. It is argued the data term must penalize the GRF to few unique minima during the inference process. On the other hand the prior energy must be less restrictive and invariant to a wide class of transformations called a Lie group, which preserve the encoded intrinsic constraints to allow for a good trade-off with the data term during the inference process.

We review the established Total Variation (TV) prior and introduce a novel structure tensor (ST) based prior. Both priors penalize level-sets with non-zero curvature due to their invariance under local rotations and translations.

We apply the TV and the ST prior in the context of uni-modal and multi-modal optical flow for which we also developed a novel similarity measure. The results show that the TV prior leads to better end-point-errors (EPE) in the uni-modal case then the ST prior and visually more consistent optical flows in the multi-modal case.

Finally we develop an extension to the PLA which takes variations of the spacial domain into account called the Extended Least Action Algorithm (ELAA). The results show that TV-based de-noising with the ELAA is significantly faster and robuster then conventional methods. On the other hand in de-noising with the ST prior the ELAA provides only modest speedups. We show that the weak performance of the ST prior in the ELAA is due to its inability to correctly measure level-sets with high curvature.

Acknowledgements

I would like to thank my first supervisor PD. Dr. Christoph S. Garbe for his scientific support and substantial engagement in securing the funds for my PhD research. Especially valuable were his remarks on low level image processing ('see how you can get those level-sets straightened out') which helped grounding the rather abstract ideas presented. I would also like to thank the second referee Prof. Dr. Jürgen Hesser for accepting the duty of reviewing the presented thesis.

The quality of the arguments in this thesis would not have been possible without the vital discussions with Prof. Paul Fieguth and his remarks on the individual sections. Prof. Fieguth extensively helped me shape my thoughts and the rigor of my arguments through pinpointing the errors therein for which I am very thankful to him. I would also like to thank Dr. Rahul Nair for proof-reading section 4 and giving valuable tips on the structure of that section.

An endeavor in scientific research of the scope of a PhD demands perseverance and sacrifice not only from the author but from those near to him. Without the endurance and psychological support of my family I would have had a hard time dealing with the numerous setbacks throughout the course of the PhD research. I am therefore especially thankful for the patience and love towards me by my wife $\mathcal{B}$abette, my two daughters $\mathbb{A}$ranka and Emy (thanks for the welcome distraction), my mother Marga and my sisters $\mathcal{S}$andra and $\mathcal{I}$ngrid, my parents in law $\mathcal{G}$erhard and $\mathcal{E}$lisabeth.

Contents

1. Introduction

The main topic of this thesis is concerned about symmetries in the mathematical modeling of computer vision problems. Many objects in nature posses among others the notable characteristic of symmetry regarding their attributes such as their form and color. A symmetry of an object $\mathcal{O}$ is such that if $\mathcal{O}$ undergoes a specific transformation g, then it appears for an observer to be unchanged. Say we have a computer vision problem involving the object $\mathcal{O}$, modeled with a mathematical model M. It is natural to reflect the symmetry of the object $\mathcal{O}$ within the model M, such that M is invariant in some sense under the transformation g. The goal of this theses is to analyze the structure of the symmetries of a mathematical model M. We will prove that knowledge of the symmetries of M may lead to significant speed ups of any algorithm using M.

Symmetries generally fall into two categories: global and local symmetries. A ball of uniform color for instance does not change its appearance to an observer upon rotation around an arbitrary axis through the center of the ball. This example is one of *global* symmetry since the ball as a whole is transformed (rotated). We can formally describe the global symmetry of the object $\mathcal{O}$ in the following way: If the surface of the object is described by the functional relationship $\phi_{\mathcal{O}}(\boldsymbol{x}) = const$ (e.g. $\phi_{\mathcal{O}}(\boldsymbol{x}) = x^2 + y^2 + z^2 = 1$ for a ball of unit radius) then our intuition of global symmetry is equal to the statement that $\phi_{\mathcal{O}}(\boldsymbol{x}) = const$ is invariant under the global transformation $\boldsymbol{x}' = g \circ \boldsymbol{x}$

$$\phi_{\mathcal{O}}(g \circ \boldsymbol{x}) = \phi_{\mathcal{O}}(\boldsymbol{x}) \tag{1.1}$$

Not all objects in nature are symmetric with respect to global transformations. For example in figure 1.1 an image of a leaf is shown. Since the leaf is not symmetric with respect to any global transformation g, its projection onto the image plane Ω is not symmetric with respect any global transformation g^{Ω} on Ω. However if we inspect local regions of the leaf, that is we zoom into those regions at various locations on the leaf, we see that the features of the leaf within the regions do posses symmetries. Figure 1.1b shows a close up of the region highlighted in figure 1.1a through which a vein of the leaf runs. The vein appears to be linear and thus symmetric towards translations along its tangential direction. This symmetry is reflected by the vectors at each position of the vein. They indicate
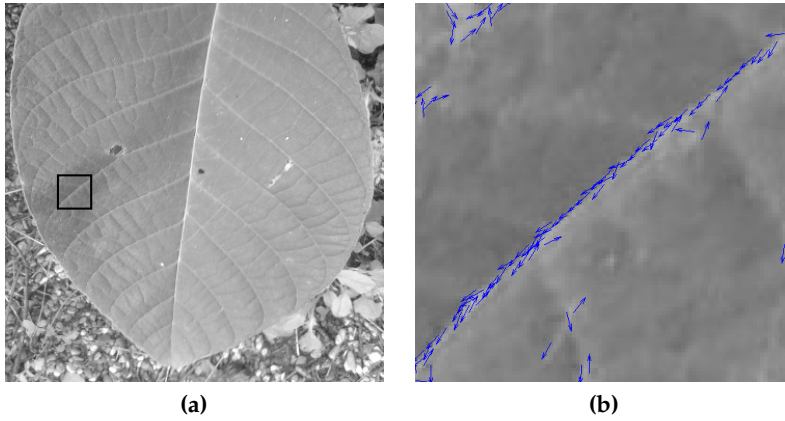
(a) **(b)**

Figure 1.1.: Figure 1.1a shows an image of a leaf. The leaf clearly has no global symmetry. Figure 1.1b shows a close-up of the region around a vein of the leaf, indicated by the box in figure 1.1a. The vectors in figure 1.1b along the vein indicate local translations which leave the vein invariant.

local translations, which leave the vein invariant. A local transformation as indicated by the vectors in figure 1.1b may be represented by the vector field $\boldsymbol{b}(\boldsymbol{x})$ such that the local transformation $g^B(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{b}(\boldsymbol{x})$ leaves the image ϕ invariant

$$\phi(\boldsymbol{x} + \boldsymbol{b}(\boldsymbol{x})) = \phi(\boldsymbol{x}) \tag{1.2}$$

In general we cannot assume that $g^B(\boldsymbol{x})$ in eq. (1.2) is unique since there can always exist a vector field $\boldsymbol{b}'(\boldsymbol{x}) \neq \boldsymbol{b}(\boldsymbol{x})$ which satisfies eq. (1.2). On the other side any transformation g^B satisfying eq. (1.2) uniquely determines the geometry of ϕ for if we were to draw lines along the tangential vectors $\boldsymbol{b}(\boldsymbol{x})$ by connecting $\boldsymbol{x}$ with $\boldsymbol{x} + \boldsymbol{b}(\boldsymbol{x})$ we would reconstruct the object $\mathcal{O}$ from $g^B(\boldsymbol{x})$.

Computational modeling in general is concerned about acquiring information about our physical reality given a set of observations. For instance a typical computer vision problem is the classification of the contents of a given image $\phi_{\mathcal{O}}$. Let $\mathbb{P}$ represent the classification problem and $M_{\mathbb{P}}$ a mathematical model (a set of formulas) describing $\mathbb{P}$. If $\phi_{\mathcal{O}}$ is the image of the leaf in figure 1.1 then an algorithm deploying $M_{\mathbb{P}}$ should indeed verify that $\phi_{\mathcal{O}}$ is the image of a leaf and additionally report some information about the leaf like the size, the name of the plant etc. The geometric invariance of $\phi_{\mathcal{O}}$ towards the transformation g^B in eq. (1.2) encodes information about the geometry of the leaf. The classification model $M_{\mathbb{P}}$ can be made aware of the geometric invariance (eq. (1.2)) and the local geometric structures of the leaf (e.g the veins) by constructing the mathematical formulas of $M_{\mathbb{P}}$ such that they themselves are invariant under the same spacial

(a)　　　　　　**(b)** Y　　　　　　**(c)** I
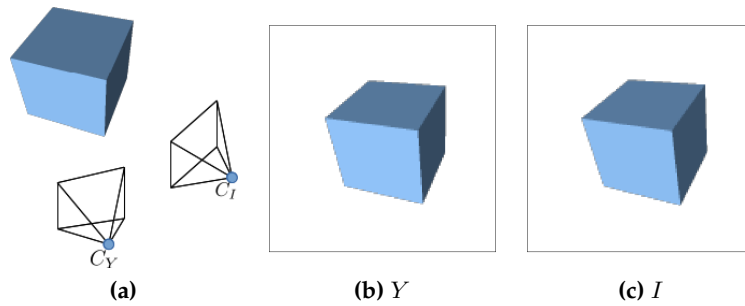
Figure 1.2.: Figure 1.2a: Two cameras are shown recording a scene from different positions. The scene could could be a rigid scene or a dynamic scene with moving objects. Figure 1.2b shows the image Y captured from the camera C_Y and figure 1.2c the image I from the camera C_I. One possible question is: How can the pixels of the image I be mapped to those of the image Y? Such a mapping can be used to deduce the 3-dimensional structure of the box similar to how the human brain constructs a 3-dimensional image given the 2-dimensional images obtained by the left and the right eye.

transformations g^B. This invariance of $M_{\mathbb{P}}$ under g^B would make it possible for $M_{\mathbb{P}}$ to identify different regions of the image $\phi_{\mathcal{O}}$ with dominant linear structures and furthermore identifying these linear structures as belonging to one global structure, e.g. the vein of the leaf. In general the concept of constructing computational models integrating known geometric invariances of the data for the aid of acquiring information in the data is at the core of this thesis.

The process of acquiring information from our physical reality via mathematical modeling is problematic itself in many ways. For one, the information which we may wish to gather may lay hidden in the data we can possibly acquire from a physical system. One such problem is called stereography ([48, 124, 77, 102]), depicted in figure 1.2. The statement of the problem goes as follows: given two images Y and I (figures 1.2b and 1.2c) obtained from the cameras C_Y and C_I (figure 1.2a) recording an object $\mathcal{O}$ (the box in figure 1.2a) how can we infer the 3-dimensional structure of $\mathcal{O}$ (the width, height and depth of the box)? This problem has already been solved by nature since the human brain is capable of reconstructing a 3-dimensional image given the 2-dimensional images obtained by the left and the right eye.

Besides the problem of hidden information described above there is another problem in the process of information acquisition. The means we use to acquire the data have technical limitations. For instance the cameras C_Y and C_I in figure 1.2a in general produce images of limited resolutions which may also be subject to noise.

Both problems in the process of information acquisition may be sub-summed as

the problem of *inference*: Given some possibly corrupted data D of a physical system we wish to infer some information stored in the unknown latent variable ϕ. In general D and ϕ may be discrete variables, continuous functions over some domain Ω or a combination of both. In this thesis we will only handle problems for which D and ϕ are continuous functions over Ω

$$D, \phi : \Omega \to \mathbb{R}^n \tag{1.3}$$

The inference problem then becomes the problem of mapping D to ϕ

$$D\left(\boldsymbol{x}\right) \xrightarrow{T} \phi\left(\boldsymbol{x}\right) \tag{1.4}$$

where T denotes a process or an algorithm which is depends the data D and ϕ is the *result* of the process T. Since ϕ is unknown we have to for one make assumptions on its geometric properties and furthermore model how it is linked to the data D. These aspects of ϕ are then embedded in the inference process T. For now we want to motivate how the geometrical properties of ϕ can be taken into account by T. Consider a local transformation g such that ϕ is transformed to an alternative ϕ'

$$\phi'\left(\boldsymbol{x}\right) = \phi\left(g \circ \boldsymbol{x}\right) \tag{1.5}$$

We can regard ϕ' as being inferred from the data D via the inference process T' similar to eq. (1.4). If ϕ is symmetric under g in the sense of eq. (1.2) then this implies that the two inference processes T and T' are equal and thus the inference process T is itself symmetric under the action of g. We conclude that knowledge of the set of local transformations $\{g\}$ which satisfy eq. (1.2) allows us to identify those inference processes T which are equal to each other upon action of $\{g\}$. This has two consequences. The first is that we can *design* an inference process T which is invariant upon the action of the set $\{g\}$. As a result this guarantees the invariance of ϕ upon the action of $\{g\}$. The second consequence is more subtle. If we split the inference process T into n intermediate steps

$$D \xrightarrow{T} \phi = D \xrightarrow{T^1} \phi^1 \xrightarrow{T^2} \phi^2 \cdots \xrightarrow{T^{n-1}} \phi^{n-1} \xrightarrow{T^n} \phi \tag{1.6}$$

the intermediate steps T^i and ϕ^i need *not* be invariant under the set $\{g\}$. However for particularly well chosen $g' \in \{g\}$ such that

$$g' \circ T^i = T^{i+k} \tag{1.7}$$

we may minimize the number steps in eq. (1.6) und thus obtain the *shortest path* in the inference problem.

The overall structure of this thesis is as follows: In section 2.1 we introduce the latent variable ϕ as a Gibbs Random Field (GRF). The main property of GRFs is that they are associated with an energy functional $E_D(\phi)$. The inference process T is explicitly formulated as the minimization problem

$$\phi^\star = \text{argmin}_\phi E_D(\phi) \quad \leftrightarrow \quad D \xrightarrow{T} \phi^\star \tag{1.8}$$

In section 2.6 we will introduce the definition of an r-dimensional Lie group $\mathbb{G}$ and its corresponding Lie algebra $\mathcal{G}$. This facilitates the formally correct definition of the local symmetry in eq. (1.2) in the form of the level-set equation

$$X\phi = 0 \quad if \quad \phi\,(g \circ \boldsymbol{x}) = \phi\,(\boldsymbol{x})\,, \quad g = \exp\,(tX) \in \mathbb{G}, \quad X \in \mathcal{G} \tag{1.9}$$

Sections 2.1 and 2.6 prepare the stage for the introduction of Emmy Noethers celebrated first theorem in section 2.7. In a nutshell this theorem states that if an energy functional $E_D(\phi)$ is invariant upon the action of an r-dimensional Lie group $\mathbb{G}$, then there exists r *divergence-free* vector fields $\boldsymbol{W}_m$

$$g \circ E_D\,(\phi) = E_D\,(\phi) \quad \forall g \in \mathbb{G} \quad \leftrightarrow \quad \exists \boldsymbol{W}_m, \quad \text{div}\,(\boldsymbol{W}_m) = 0 \quad \forall \quad 1 \le m \le r \tag{1.10}$$

Since its first publication in 1918, Noether's first theorem has had far reaching implications in our understanding of the fundamental laws of motion in physics as well as the deep connection between the symmetries of a physical system and its conservation laws. For instance the time invariance of the laws of motion in the universe reveals the conservation of energy. In layman words: It does not matter if we carry out an experiment now or next week, the results will be the same since the energy of the universe does not vanish! Building on section 2.7 we demonstrate in section 3 the construction of a prior energy functional $E^{prior}(\phi)$ which is invariant under the Lie group $\mathbb{T} \times SO(2)$ which is the group of local translations and rotations. In section 3.3 we will use the prior developed in section 3 in the context of optical flow [107]. In section 4 we will introduce a generalization of the Newton approach for solving the inference problem in eq. (1.8) which takes local transformations of the spatial coordinates $\boldsymbol{x}$ in Ω (see eq. (1.2)) into account to facilitate the search for the shortest path in the inference problem in eq. (1.6).

2. Background

2.1. Gibbs Random Fields

A physical system C is a dynamical composite of elements which interact with each other as well as with the environment the system C is embedded in. The elements are described by a vector of parameters $\phi = (\phi_1, \ldots, \phi_n)$. The physical system C relates a specific value $\phi^\star$ of the vector ϕ to a set of observables $Y = \{Y_1, \ldots, Y_k\}$

$$Y = C\left(\phi^\star\right) \tag{2.1}$$

In the case that the elements of the system C are continuously distributed over a finite space Ω, the parameter vector ϕ is a function on Ω

$$\phi(\boldsymbol{x}) \in \mathbb{R}^n \quad \boldsymbol{x} \in \Omega \tag{2.2}$$

called a Gibbs-Random-Field (GRF) [39]. In this thesis we will generally assume that the GRF ϕ is infinitely differentiable, $\phi\left(\boldsymbol{x}\right) \in \mathcal{C}^\infty\left(\Omega\right)$, unless otherwise explicitly stated. The interactions of the elements of the system C with the environment are characterized by an energy functional $E_Y^{data}(\phi)$ called the data term, which couples the GRF $\phi(\boldsymbol{x})$ to the observables Y. There is another energy form $E^{prior}(\phi, \nabla\phi)$ within the system C called the prior. $E^{prior}(\phi, \partial_j\phi)$ describes how the elements of C interact with each other. Together both energy functionals form the total energy of the system C

$$E_Y\left(\phi, \nabla\phi\right) = E_Y^{data}(\phi) + E^{prior}(\phi, \nabla\phi) \tag{2.3}$$

which is related to the probability distribution

$$p\left(\phi, \nabla\phi | Y\right) = p\left(Y | \phi\right) \cdot p(\phi, \nabla\phi) \sim \exp\left(-E_Y\left(\phi, \nabla\phi\right)\right) \tag{2.4}$$

$$p\left(Y | \phi\right) = \exp\left(-E_Y^{data}\left(\phi\right)\right) \tag{2.5}$$

$$p\left(\phi, \nabla\phi\right) = \exp\left(-E^{prior}\left(\phi, \nabla\phi\right)\right) \tag{2.6}$$

The value of the probability distribution $p(\phi, \nabla\phi|Y)$ evaluated at the values $\hat{\phi}(\boldsymbol{x})$ describes the probability that the GRF $\phi(\boldsymbol{x})$ assumes the values $\hat{\phi}(\boldsymbol{x})$ at each point $\boldsymbol{x} \in \Omega$. The set of values $\hat{\phi}(\boldsymbol{x})$ is what is called a *configuration* of the GRF ϕ.

$E_Y(\phi, \nabla\phi)$ is designed such that it is minimal once the GRF $\phi(\boldsymbol{x})$ fulfills the forward problem in eq. (2.1)

$$\phi^\star = \text{argmin}_\phi \left(E_Y(\phi, \nabla\phi) \right) \tag{2.7}$$

The particular value $\phi^\star(\boldsymbol{x})$ of the GRF ϕ is the most probable configuration of the distribution $p(\phi|Y)$ due to eq. (2.4) and the solution to the inverse problem

$$\phi^\star = C^{-1}(Y) \tag{2.8}$$

In this thesis we will not be concerned with the probability interpretation of GRF models via eq. (2.4) but rather directly with the formulation of a GRF model via the energy functional $E_Y(\phi, \nabla\phi)$ in eq. (2.3). The focus lies on the prior E^{prior} and we will show how it incorporates the geometrical invariance in eq. (1.2).

2.2. Convex Optimization

Inverse problems as in eq. (2.7) generally fall into two categories: convex and non-convex problems. Convex energy functionals [97, 16, 17] are generally easier to minimize and the minimization algorithms are robust to alterations of the initial conditions. On the other hand non-convex energy functionals generally model the dependency of the GRF $\phi(\boldsymbol{x})$ on the data more precisely. However they are also harder to minimize and the minimization algorithms are more sensitive to initialization. We will give a short introduction into convex analysis following [97, 103]. We will first start with describing convex, lower semi-continuous (l.s.c) and proper functions $f : X \to \mathbb{R}$ which map any Hilbert space X onto the real numbers. The goal of this section is to introduce variational methods for finding the minimum of such functions. Since convex functions need not be differentiable, traditional optimization methods such as steepest descent fail and we will need a generalization of the concept of differentiability called the subdifferential. The subdifferential $\partial f(\phi)$ at a specific point $\phi \in X$ at which f is continuous but not differentiable is basically a set of lines tangent to f at ϕ. This extended concept of differentiability is important for the generalization of steepest descent, the so-called proximal operator $\text{prox}(\phi|\tau f)$ [74, 96]. The constant τ is similar to the step size in steepest descent methods and the stationary points $\phi^\star = \text{prox}(\phi^\star|\tau f)$ coincide with the minimizers of f

$$f(\phi) \geq f(\phi^\star), \quad \forall \phi \in X \tag{2.9}$$

Our general goal is to introduce the method of Pock et. al. [89, 17] for minimizing a class of convex functions of the form $E(\phi, A\phi) = g(A\phi) + f(\phi)$ where A is a linear operator, since the energy functional $E_Y(\phi, \nabla\phi)$ in eq. (2.4) is in this class.

Central to the method of [89] is the concept of *duality*. The dual space of the Hilbert space X is the set $X^\star$ such that for each $p \in X^\star$ the scalar product $\langle \cdot, \cdot \rangle$ yields a linear operator $\langle p, \cdot \rangle : X \to \mathbb{R}$ on X. It turns out that every convex function $E(\phi, A\phi)$ on X has a unique *concave* dual conjugate $E^\star(p, A^\star p)$ on $X^\star$ such that the minimum of E equals the maximum of the conjugate $E^\star$

$$E(\phi^\star, A\phi^\star) = \min_{\phi \in X} E(\phi, A\phi) = \max_{p \in X^\star} E^\star(p, A^\star p) = E^\star(p^\star, A^\star p^\star) \tag{2.10}$$

Eq. (2.10) is called a saddle-point problem with the pair $(\phi^\star, p^\star)$ being the saddle-point. Thus to obtain the minimizer $\phi^\star$ of E the scheme devised in [89], based on the work of Arrow and Hurrwicz in [2] and Popov in [91], starts from an initial guess (ϕ^0, p^0) and progresses to the saddle-point $(\phi^\star, p^\star)$ through alternating steps of *descent* in the primal variable ϕ and *ascent* in the dual variable p. The benefits of such a primal-dual approach are that the evaluation of the function $E(\phi, A\phi)$ and its differentials may pose numerical problems which are avoidable in the dual picture. Indeed for the case of total variation denoising with the ROF model [100] it was shown in [16] the transformation of the ROF model to the dual frame leads to a numerically vastly more stable and high-performance algorithm.

What follows are some basic definitions of convex functions.

Definition 1 (Proper function)**.** *Let $f : X \to \mathbb{R} \cup \infty$ be a mapping from a finite dimensional real vector space X to the real numbers. f is a proper function if its domain*

$$dom(f) = \{\phi \in X \,|\, f(\phi) < \infty\} \tag{2.11}$$

is not empty.

Definition 2 (Lower Semi-Continuous Function)**.** *Let $f : X \to \mathbb{R} \cup \infty$. f is lower semi-continuous if the level-sets*

$$S_\alpha^f = \{\phi \in X \,|\, f(\phi) \leq \alpha\} \tag{2.12}$$

are closed for all $\alpha \in \mathbb{R}$

Definition 3 (Convex function)**.** *Let $f : X \to \mathbb{R} \cup \infty$ be a proper function, $\phi_0 \in X$ and $\phi_1 \in X$ and $\lambda \in [0, 1]$. f is called a convex function if it satisfies*

$$\lambda f(\phi_0) + (1 - \lambda)f(\phi_1) \geq f(\lambda\phi_0 + (1 - \lambda)\phi_1) \tag{2.13}$$

In the case of strict inequality f is called strictly convex. If $g : X \to \mathbb{R} \cup \{-\infty\}$ is such that $-g$ is a proper convex function, then g is a proper concave function.

The domain dom(f) of any convex function f is convex itself, since for any $\phi_{0,1} \in$ dom(f) the image $f(\phi_\lambda)$ of the line segment $\phi_\lambda = \lambda\phi_0 + (1 - \lambda)\phi_1$ is bounded by virtue of eq. (2.13). Hence ϕ_λ is contained in dom(f) for all $\lambda \in [0, 1]$ and dom(f) is convex.

Definition 4 (Epigraph). *Let $f : X \to \mathbb{R} \cup \infty$. The epigraph of f is defined as the set*

$$epi(f) = \{(\phi, \mu) \in X \times \mathbb{R} \mid f(\phi) \leq \mu\} \tag{2.14}$$

If f is convex then epi(f) also convex.

We will denote by Γ_0 the class of proper, l.s.c and convex functions $f : X \to \mathbb{R} \cup \infty$.

In traditional optimization techniques such as steepest descent the minimum of a function $f \in \Gamma_0$ is characterized by the gradient with respect to $\phi \in X$

$$f(\phi) \geq f(\phi^\star), \quad \forall \phi \in X \Leftrightarrow \quad \nabla f(\phi^\star) = 0 \tag{2.15}$$

However eq. (2.15) assumes that the gradient exists at $\phi^\star$, and thus that f is smooth. In general this may not be the case. For instance the function $f(\phi) = \|\phi\|$ is convex, proper and l.s.c but not smooth at its minimum $\phi = 0$. Hence $\nabla f(\phi) = \frac{\phi}{\|\phi\|}$ does not exist at $\phi = 0$ and we need a broader definition of differentiability which also encompasses the case of non-smooth functions.

Definition 5 (Subdifferential). *Let X a Hilbert space, $X^\star$ be its dual with respect to the scalar product $\langle \cdot, \cdot \rangle$ and $f : X \to \mathbb{R}$ be convex. The subdifferential $\partial f(\phi) \subset X^\star$ of f at $\phi \in X$ is defined as the set of elements $p \in X^\star$ such that*

$$f(v) - f(\phi) \geq \langle p, v - \phi \rangle, \quad \forall v \in X \tag{2.16}$$

The subdifferential $\partial f(\phi)$ of a convex function f can be understood as a map ∂f whose image of a point ϕ is a set $\partial f(\phi)$. Whenever f is smooth at ϕ the subdifferential reduces to the ordinary gradient, $\partial f(\phi) = \{\nabla f(\phi)\}$. The characterization of the minimizers of a convex function f by means of its subdifferential is summarized by the following lemma

Lemma 1. *Let $f : X \to \mathbb{R}$ be a convex function. An element $\phi^\star \in X$ is a global minimizer of f if it satisfies*

$$\mathbf{0} \in \partial f(\phi^\star) \tag{2.17}$$

Proof. If eq. (2.17) holds then by setting $p = \mathbf{0}$ in the definition of the subdifferential ∂f in eq. (2.16) we get $f(v) - f(\phi^\star) \geq 0$ for all $v \in X$ hence $\phi^\star$ minimizes f. The reversed argument also holds. $\qquad\square$

2.2.1. The Proximal Operator

One of the basic algorithms of non-smooth convex optimization is the Projection Onto Convex Sets (POCS) algorithm ([120, 23, 20]). The problem is stated as follows:

Definition 6 (Projection Onto Convex Sets). *Let X be a Hilbert space, $C \subset X$ a convex subset and $\phi \in X$. The projection of ϕ onto C is obtained by the constrained optimization problem*

$$z = P_C(\phi) = \underset{y \in C}{argmin} \left\{ \frac{\|\phi - y\|^2}{2} \right\} \tag{2.18}$$

If $\phi \in C$ then $\phi = P_C(\phi)$ holds.

Eq. (2.18) can be rewritten in an unconstrained form

$$z = P_C(\phi) = \underset{y \in X}{argmin} \left\{ D_C(y, \phi) \right\}, \quad D_C(y, \phi) = \delta_C(y) + \frac{\|\phi - y\|^2}{2} \tag{2.19}$$

where the convex *indicator function* $\delta_C(\phi) \in \Gamma_0$ is defined by

$$\delta_C(\phi) = \begin{cases} 0 & \phi \in C \\ \infty & otherwise \end{cases} \tag{2.20}$$

The minimization in eq. (2.19) is carried over the entire set X however the indicator function δ_C constrains the minimization to the set C. The projection $z = P_C(\phi)$ of ϕ can be seen as a minimizer of δ_C. z is unique since $D_C(y, \phi)$ is strictly convex in y for fixed ϕ.

In [74] a generalization of the projection operator in eq. (2.19) for the whole class of convex lower semi-continuous functions $f \in \Gamma_0$ was proposed called the *proximal operator*

Definition 7 (Proximal Operator). *Let $f \in \Gamma_0$. The proximal operator is defined as*

$$prox(\phi|f) = \underset{y \in X}{argmin} \left\{ f(y) + \frac{\|\phi - y\|^2}{2} \right\} \tag{2.21}$$

10

The proximal operator has 3 important properties described in the following lemmas

Lemma 2 (Uniqueness). *For any $\phi \in X$ the mapping $z = prox(\phi|f)$ is unique*

Proof. Since f is convex and $\|\phi - y\|^2$ is strictly convex the sum $f(y) + \frac{\|\phi - y\|^2}{2}$ is strictly convex and hence $prox(\phi|f)$ is the unique minimizer. $\square$

Lemma 3 (Stationary Points). *$\phi^\star \in X$ is a stationary point*

$$\phi^\star = prox(\phi^\star|f) \tag{2.22}$$

if and only if $\phi^\star$ is a minimizer of f

Proof. We show first the $'\Rightarrow'$ argument. Since the proximal point $z = prox(\phi^\star|f)$ minimizes $f(y) + \frac{\|\phi - y\|^2}{2}$ by definition, it satisfies the subgradient equation

$$\mathbf{0} \in \partial f(z) + z - \phi \tag{2.23}$$

Hence for $z = \phi = \phi^\star$ it follows that $\mathbf{0} \in \partial f(\phi^\star)$.

$'\Leftarrow'$: Let $\phi^\star$ be a minimizer of f, $\mathbf{0} \in \partial f(\phi^\star)$. We can unambiguously reconstruct eq. (2.22) by adding a zero to $\partial f(\phi^\star)$ to obtain $\mathbf{0} \in \partial f(\phi^\star) + \phi^\star - \phi^\star$ and then reverse the $'\Rightarrow'$ argument. $\square$

Proposition 1 (Non-Expansive). *For any pair $\phi, y \in X$ the proximal operator $prox(\cdot|f)$ is non-expansive*

$$\|prox(\phi|f) - prox(y|f)\| \leq \|\phi - y\| \tag{2.24}$$

For the proof of proposition 1 we need the following lemma

Lemma 4 (Monotonic Subdifferential). *Let $f : X \to \mathbb{R}$ be a convex function. For any $\phi_1 \in X$, $\phi_2 \in X$, $y_1 \in \partial f(\phi_1)$ and $y_2 \in \partial f(\phi_2)$ the monotonicity condition*

$$\langle y_1 - y_2, \phi_1 - \phi_2 \rangle \geq 0 \tag{2.25}$$

holds.

Proof. By the definition of the subdifferential in eq. (2.16) we have

$$\begin{aligned} y_1 \in \partial f(\phi_1) &\Rightarrow f(v) - f(\phi_1) \geq \langle y_1, v - \phi_1 \rangle \, \forall v \in X \\ y_2 \in \partial f(\phi_2) &\Rightarrow f(v') - f(\phi_2) \geq \langle y_2, v' - \phi_2 \rangle \, \forall v' \in X \end{aligned} \tag{2.26}$$

Since f is convex the inequalities in eq. (2.26) hold also for all $\phi_1, \phi_2 \in X$. After some manipulations we get

$$\langle y_1, \phi_1 - v \rangle \geq f(\phi_1) - f(v) \tag{2.27}$$
$$\langle -y_2, v' - \phi_2 \rangle \geq f(\phi_2) - f(v') \tag{2.28}$$

We set $v = \phi_2$ and $v' = \phi_1$ and sum both inequalities and get

$$\langle y_1 - y_2, \phi_1 - \phi_2 \rangle \geq 0 \tag{2.29}$$

$\square$

proof of proposition 1. Let $\phi_1 = \text{prox}(y_1|f)$ and $\phi_2 = \text{prox}(y_2|f)$. From eq. (2.23) it follows that

$$y_1 - \phi_1 \in \partial f(\phi_1), \quad y_2 - \phi_2 \in \partial f(\phi_2) \tag{2.30}$$

By lemma 4 we have the inequality

$$\langle y_1 - \phi_1 - (y_2 - \phi_2), \phi_1 - \phi_2 \rangle \geq 0 \quad \Rightarrow \quad \langle y_1 - y_2, \phi_1 - \phi_2 \rangle \geq \|\phi_1 - \phi_2\|^2 \tag{2.31}$$

Using Schwartz's inequality $\|y_1 - y_2\| \, \|\phi_1 - \phi_2\| \geq \langle y_1 - y_2, \phi_1 - \phi_2 \rangle$ we get to

$$\|y_1 - y_2\| \geq \|\phi_1 - \phi_2\| \tag{2.32}$$

which proves the proposition. $\square$

The proximal operator is at the center of the proximal point algorithm which was proposed in [69]. Since the proximal point algorithm predates the splitting algorithm in [89] which is partly the focus of this thesis, we will briefly discuss it. The proximal point algorithm for minimizing a convex function $f \in \Gamma_0$ generates a sequence $\{\phi_i\}, 0 \leq i \leq \infty$ by computing the iteration

$$\phi_{k+1} = \text{prox}(\phi_k|\tau f) \tag{2.33}$$

where $\tau > 0$ is a fixed step parameter. We define the difference $y_k = \phi_k - \phi_{k-1}$. Each iteration y_{k+1} is bounded by the previous y_k, $\|y_{k+1}\| \leq \|y_k\|$ by virtue of the non-expansive property in proposition 1 and thus in [43] it was shown that the function values $f(\phi_n)$ satisfy

$$f(\phi_n) - f(\phi^\star) \leq \frac{\|\phi_0 - \phi^\star\|^2}{\sigma_n}, \quad \lim_{n \to \infty} \sigma_n = \infty \tag{2.34}$$

If f is smooth such that $\partial f = \{\nabla f\}$ the proximal point iteration in eq. (2.33) reduces to

$$\phi_{k+1} = \phi_k - \tau_k \nabla f(\phi_{k+1}) \tag{2.35}$$

by setting $z = \phi_{k+1}$ and $\phi = \phi_k$ in eq. (2.23). Eq. (2.35) is called the *backward gradient step* (see [86]) and is similar to the *forward gradient step* defined as

$$\widetilde{\phi}_{k+1} = \phi_k - \tau_k \nabla f(\phi_k) \tag{2.36}$$

also known as steepest descent [39, 86]. The backward gradient step in eq. (2.35) is an equation which must be solved to obtain the upgrade ϕ_{k+1}. On the other side in the steepest descent step in eq. (2.36) the upgrade $\widetilde{\phi}_{k+1}$ is directly computed from the previous step ϕ_k. Hence the backward gradient step is also called an *implicit* and the forward gradient step an *explicit* step. In [86] it was pointed out that although steepest descent methods are simpler to compute then backward gradient methods, they often converge slower. On the other hand a single iteration in the backward gradient algorithm may be numerical as involved the original minimization problem. To overcome this issue splitting schemes [10, 108, 19, 22, 21, 26, 29, 64, 72, 87] have been devised which split the function f into components for which the proximal operators are numerically easier to compute. The method of Pock et. al. [89] is an example of a splitting scheme. Its fundamental ingredient is the concept of *duality* which we shall introduce now.

2.2.2. Fenchel Duality

The traditional description of a (convex) function $f : X \rightarrow \mathbb{R}$ is that of a mapping of *points* $\phi \in X$ to values $f(\phi) \in \mathbb{R}$. In many cases the proximal operator $\mathrm{prox}\,(\cdot|\tau f)$ can evaluated in an elegant manner by transforming f to a description in which the *slopes* of lines tangent to the epigraph $\mathrm{epi}(f)$ are mapped to the image of f. In figure 2.1a a convex function f and its epigraph $\mathrm{epi}(f)$ are depicted. Since $f \in \Gamma_0$ the epigraph is convex itself. In figure 2.1b the half-planes $H_{p1,2}$ are the epigraphs of the tangent lines with slopes $p_{1,2}$ at the points $\phi_{1,2} \in X$. The epigraph of f is clearly a subset of the intersection of H_{p_1} and H_{p_2}, $\mathrm{epi}(f) \subset H_{p_1} \cap H_{p_2}$. Naturally the half-plane H_p is a continuous function of the slope p. If we consider *all* half-planes H_p which contain the epigraph $\mathrm{epi}(f)$ then $\mathrm{epi}(f)$ is identical to the intersection of all such half-planes

$$\mathrm{epi}(f) = \cap H_p, \quad \mathrm{epi}(f) \subset H_p \tag{2.37}$$
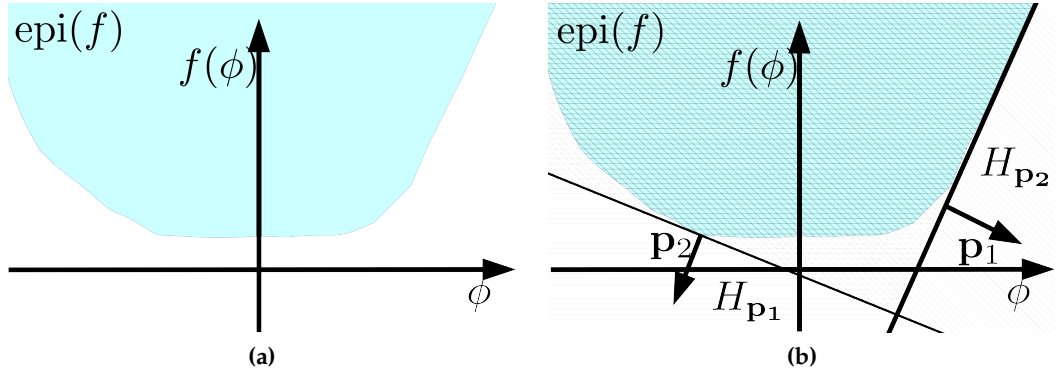
Figure 2.1.: Figure 2.1a shows a convex function $f \in \Gamma_0$ and its epigraph epi(f) and figure 2.1b shows two additional tangents at the points $\phi_{1,2}$, their normals $p_{1,2}$ and their half-planes $H_{p_{1,2}}$. The epigraph is clearly contained in the intersection of the half-planes, epi$(f) \subset H_{p_1} \cap H_{p_2}$. By adding more half-planes we can substitute epi(f) by the intersection of all the half-planes thus obtaining a picture in which epi(f) and f are fully described in terms of the normal vector $p \in X^\star$.

Each half-plane H_p is the epigraph of a linear function h_p

$$h_p(\phi) = \langle p, \phi \rangle - b \tag{2.38}$$

In eq. (2.38) it is evident that the slope p of h_p is an element of the dual Hilbert space $X^\star$. Since epi(f) is contained in each H_p the function f majorizes the corresponding linear function h_p

$$f(\phi) \geq \langle p, \phi \rangle - b \quad \forall \phi \in X \tag{2.39}$$

Eq. (2.39) imposes a bound on the possible values for the offset b since

$$b \geq \langle p, \phi \rangle - f(\phi) \tag{2.40}$$

The largest lower bound of b is a function of p called the *conjugate* of f.

Definition 8 (Conjugate Function). *Let $f : X \to \mathbb{R}$ be a convex, l.s.c and proper function on a Hilbert space X. Then $f^\star : X^\star \to \mathbb{R}$ defined by*

$$f^\star(p) = \max_{\phi \in X} \{\langle p, \phi \rangle - f(\phi)\} \tag{2.41}$$

is called the convex conjugate *function of f.*

If $g : X \to \mathbb{R}$ is a concave function then its conjugate $g_\star : X\star \to \mathbb{R}$ is defined by

$$g_\star(p) = \min_\phi \left\{ \langle p, \phi \rangle - g(\phi) \right\} \tag{2.42}$$

The following lemma relates convex and concave conjugates to each other.

Lemma 5. *Let $g : X \to \mathbb{R}$ be a concave function on the Hilbert space X and let $h = -g$ then*

$$g_\star(p) = -h^\star(-p) \tag{2.43}$$

Proof. By the definition of the conjugate for concave function in eq. (2.42) we have

$$g_\star(p) = \min_x \left\{ \langle p, \phi \rangle - g(\phi) \right\} = \min_x \left\{ \langle p, \phi \rangle + h(\phi) \right\} \tag{2.44}$$

$$= \min_x \left\{ - \left(\langle -p, \phi \rangle - h(\phi) \right) \right\} = - \max_x \left\{ \langle -p, \phi \rangle - h(\phi) \right\} \tag{2.45}$$

$$= -h^\star(-p) \tag{2.46}$$

$\square$

The half-plane defined by $\widetilde{h}_p(\phi) = \langle p, \phi \rangle - f^\star(p)$ is tangential to f at the supremum $\phi^\star$ of eq. (2.41). From the definition of the convex conjugate $f^\star$ in eq. (2.41) it directly follows that

$$f(\phi) + f^\star(p) \geq \langle p, \phi \rangle \tag{2.47}$$

Eq. (2.47) is called *Fenchel's inequality* [97]. for a concave function g a similar inequality holds:

$$\langle p, \phi \rangle \geq g(\phi) + g_\star(p) \tag{2.48}$$

The following lemma shows an important consequence of the special case of equality in eq. (2.47)

Lemma 6. *Let $f \in \Gamma_0$. Then*

1. $f(\phi_0) + f^\star(p_0) = \langle p_0, \phi_0 \rangle \quad \Leftrightarrow \quad p_0 \in \partial f(\phi_0)$
2. $f(\phi_0) + f^\star(p_0) = \langle p_0, \phi_0 \rangle \quad \Leftrightarrow \quad \phi_0 \in \partial f^\star(p_0)$

Proof. "$\Rightarrow$": Case 1. From the definition of $f^\star$ in eq. (2.41) we have

$$f^\star(p_0) \geq \langle p_0, \phi \rangle - f(\phi), \quad \forall \phi \in X \tag{2.49}$$

Let $f(\phi_0) + f^\star(p_0) = \langle p_0, \phi_0 \rangle$ hold. Then from eq. (2.49)

$$\langle p_0, \phi_0 \rangle = f(\phi_0) + f^\star(p_0) \geq f(\phi_0) - f(\phi) + \langle p_0, \phi \rangle, \quad \forall \phi \in X$$
$$f(\phi) - f(\phi_0) \geq \langle p_0, \phi - \phi_0 \rangle \quad \forall \phi \in X \tag{2.50}$$

By the definition of the subdifferential in eq. (2.16) we get $p_0 \in \partial f(\phi_0)$.

Case 2: Since f is convex, proper and l.s.c we have $f^{\star\star} = f$ (see Corollary 7.4.2 in [97]) and thus

$$f(\phi_0) \geq \langle p, \phi_0 \rangle - f^\star(p), \quad \forall p \in X^\star \tag{2.51}$$

Similar to eq. (2.50) we obtain

$$\langle p_0, \phi_0 \rangle = f(\phi_0) + f^\star(p_0) \geq f^\star(p_0) - f^\star(p) + \langle p, \phi_0 \rangle, \quad \forall p \in X^\star$$
$$f^\star(p) - f^\star(p_0) \geq \langle p - p_0, \phi_0 \rangle \quad \forall p \in X^\star \tag{2.52}$$

and thus $\phi_0 \in \partial f^\star(p_0)$.

"$\Leftarrow$": Case 1: From $p_0 \in \partial f(\phi_0)$ we have

$$f(\phi) - f(\phi_0) \geq \langle p_0, \phi - \phi_0 \rangle \quad \forall \phi \in X \tag{2.53}$$
$$\Rightarrow \langle p_0, \phi_0 \rangle \geq \langle p_0, \phi \rangle - f(\phi) + f(\phi_0) \quad \forall \phi \in X \tag{2.54}$$

We apply the definition of the conjugate function $f^\star$ in eq. (2.41) to eq. (2.54)

$$\langle p_0, \phi_0 \rangle \geq f^\star(p_0) + f(\phi_0) \tag{2.55}$$

Combining eq. (2.55) with Fenchel's inequality in eq. (2.47) we deduce $f(\phi_0) + f^\star(p_0) = \langle p_0, \phi_0 \rangle$. The proof for the "$\Leftarrow$" in Case 2 is similar to the above argument.
$\square$

In lemma 1 we showed that $\phi^\star$ is a minimizer of the convex function f if and only if it satisfies the subdifferential equation

$$\mathbf{0} \in \partial f(\phi^\star) \tag{2.56}$$

By lemma 6 the condition in eq. (2.56) is equivalent to $\phi^\star \in \partial f^\star(\mathbf{0})$. Thus the subdifferential $\partial f^\star(\mathbf{0})$ is the set of all minimizers of f and

$$f(\phi^\star) = -f^\star(\mathbf{0}) \tag{2.57}$$

must hold. Since the conjugate function $f^\star$ is convex its negative is concave. In the following we want to show that eq. (2.57) can be transformed into a saddle-

point equation which equates the *minimum* value of f, $f(\phi^\star)$ to the *maximum* of a *concave* function involving the conjugate function $f^\star$ thus setting the stage for Fenchel's duality theorem, to be introduced soon.

Similar to the definition of the indicator function in eq. (2.20) we can define a conjugate indicator function

$$\delta_{C^\star}^{\alpha}(p) = \begin{cases} -\alpha & p \in C^\star \\ \infty & otherwise \end{cases} \quad C^\star = \{\mathbf{0}\} \tag{2.58}$$

with finite α and reformulate the right hand side of eq. (2.57) as a maximization over $p \in X^\star$

$$-f^\star(\mathbf{0}) + \alpha = \max_{p \in X^\star} \left\{ -\delta_{C^\star}^{\alpha}(p) - f^\star(p) \right\} \tag{2.59}$$

such that eq. (2.57) becomes

$$\min_{\phi \in X} \left\{ f(\phi) + \alpha \right\} = \max_{p \in X^\star} \left\{ -\delta_{C^\star}^{\alpha}(p) - f^\star(p) \right\} \tag{2.60}$$

It is easy to see that $g_\star(p) = \delta_{C^\star}^{\alpha}(p)$ is the conjugate of the constant function $g(\phi) = \alpha$ and thus eq. (2.60) relates the minimizers of $f(\phi) + \alpha$ to the maximizers of the concave dual function $-\delta_{C^\star}^{\alpha}(p) - f^\star(p)$. In fact as the next theorem will show this duality between the minimizers of the convex $f - g$, where g is an arbitrary concave function, and the maximizers of the concave conjugate $g_\star - f^\star$ always exists.

Theorem 1 (Fenchel's Duality Theorem). *Let X and Y be two Hilbert spaces, $f : X \to \mathbb{R} \cup \{\infty\}$ a proper convex function, $g : Y \to \mathbb{R} \cup \{-\infty\}$ a proper concave function and $A : X \to Y$ a linear map. Furthermore let*

$$E(\phi, A\phi) = f(\phi) - g(A\phi) \tag{2.61}$$
$$E^\star(p, A^\star\phi) = g_\star(p) - f^\star(A^\star p) \tag{2.62}$$

Then

$$\min_{\phi \in X} E(\phi, A\phi) = \max_{p \in Y^\star} E^\star(p, A^\star p) \tag{2.63}$$

must hold.

Eq. (2.63) shows that the convex function $E(\phi, A\phi)$ is bounded from below by the maximum of the concave conjugate function $E^\star(p, A^\star p)$. One could take the dual point of view and say that the conjugate function $E^\star$ is bounded from above

by the minimum of E. This makes eq. (2.63) a saddle-point equation and the point $(\phi^\star, p^\star) \in X \times Y^\star$ for which eq. (2.63) is fulfilled is a saddle-point.

Proof. To begin with, we first prove eq. (2.63) for the case $A = \mathbb{1}$

$$\min_{\phi \in X} \{f(\phi) - g(\phi)\} = \max_{p \in X^\star} \{g_\star(p) - f^\star(p)\} \tag{2.64}$$

By Fenchel's inequality in eq. (2.47) we have

$$f(\phi) + f^\star(p) \geq \langle p, \phi \rangle \geq g(\phi) + g_\star(p) \quad \Rightarrow f(\phi) - g(\phi) \geq g_\star(p) - f^\star(p) \tag{2.65}$$

and thus

$$\alpha = \min_x \{f(\phi) - g(\phi)\} \geq \max_p \{g_\star(p) - f^\star(p)\} \tag{2.66}$$

We have to prove that

$$\max_p \{g_\star(p) - f^\star(p)\} \geq \alpha \tag{2.67}$$

Let the sets F and G be the epigraphs

$$F = \{(\phi, \mu) \,|\, \phi \in X, \mu \geq f(\phi)\} \tag{2.68}$$
$$G = \{(\phi, \mu) \,|\, \phi \in X, \mu \leq g(\phi) + \alpha\} \tag{2.69}$$

Since F and G are convex sets and $f(\phi) \geq g(\phi) + \alpha$ by the definition of α in eq. (2.66) G and the relative interior of F are disjoint, $\mathrm{ri}F \cap G = \emptyset$. There then exists a half plane defined by the affine function $h_p(\phi) = \langle p, \phi \rangle - \beta$

$$H = \{(\phi, \mu) \,|\, \phi \in X, \mu \geq h_p(\phi)\} \tag{2.70}$$

which separates F and G in the sense that $\mathrm{ri}F \subset H$ and $H \cap \mathrm{ri}D = \emptyset$. It follows that

$$f(\phi) \geq h_p(\phi) \geq g(\phi) + \alpha \tag{2.71}$$

The left inequality yields

$$\beta \geq \langle p, \phi \rangle - f(\phi) \quad \Rightarrow \beta \geq f^\star(p) \tag{2.72}$$

where eq. (2.72) follows from the definition of the conjugate $f^\star$ in eq. (2.41). The

right inequality in eq. (2.71) yields in combination with eq. (2.72)

$$\langle p, \phi \rangle - g(\phi) \geq \beta + \alpha \tag{2.73}$$
$$\Rightarrow g_\star(p) \geq \beta + \alpha \geq f^\star(p) + \alpha \tag{2.74}$$

From eq. (2.74) we conclude

$$\max_p \left\{ g_\star(p) - f^\star(p) \right\} \geq \alpha \tag{2.75}$$

which in combination with eq. (2.66) proves eq. (2.64).

For the general case in eq. (2.63) let $\widetilde{g}(\phi) = g(A\phi)$. Then we can compute the conjugate

$\widetilde{g}_\star(p)$ by finding an alternative representation of $-g(A\phi)$

$$-g(A\phi) = -\max_{y \in Y} \left\{ g(y) + \min_{p_y \in Y^\star} \langle p_y, A\phi - y \rangle \right\} \tag{2.76}$$

$$= \min_{y \in Y} \left\{ -g(y) - \min_{p_y \in Y^\star} \langle p_y, A\phi - y \rangle \right\} \tag{2.77}$$

$$= \min_{y \in Y} \max_{p_y \in Y^\star} \left\{ -g(y) - \langle p_y, A\phi - y \rangle \right\} \tag{2.78}$$

then

$$\widetilde{g}_\star(p) = \min_{\phi'} \left\{ \langle p, \phi' \rangle - \widetilde{g}(\phi') \right\} = \min_{\phi'} \left\{ \langle p, \phi' \rangle - g(A\phi') \right\} \tag{2.79}$$

$$= \min_{y \in Y} \max_{p_y \in Y^\star} \min_{\phi' \in X} \left\{ \langle p, \phi' \rangle - g(y) - \langle p_y, A\phi' - y \rangle \right\} \tag{2.80}$$

$$= \min_{y \in Y} \max_{p_y \in Y^\star} \min_{\phi' \in X} \left\{ \langle p - A^\star p_y, \phi' \rangle - g(y) + \langle p_y, y \rangle \right\} \tag{2.81}$$

$$= \max_{p_y \in Y^\star} \min_{\phi' \in X} \left\{ \langle p - A^\star p_y, \phi' \rangle + g_\star(p_y) \right\} \tag{2.82}$$

The dual operator $A^\star : Y^\star \to X^\star$ defined by $\langle A^\star p_y, \phi' \rangle = \langle p_y, A\phi' \rangle$ was used in eq. (2.81). We insert $\widetilde{g}_\star(p)$ into eq. (2.64)

$$\min_{\phi \in X} \left\{ f(\phi) - \widetilde{g}(\phi) \right\} = \max_{p \in X^\star} \max_{p_y \in Y^\star} \min_{\phi' \in X} \left\{ \langle p - A^\star p_y, \phi' \rangle + g_\star(p_y) - f^\star(p) \right\} \tag{2.83}$$

and carry out the maximization over p and minimization over ϕ'. These two operations fix the constraint $p = A^\star p_y$ thus we get

$$\min_{\phi \in X} \left\{ f(\phi) - g(A\phi) \right\} = \max_{p_y \in Y^\star} \left\{ g_\star(p_y) - f(A^\star p_y) \right\} \tag{2.84}$$

$\square$

For the case when *both* f and g are convex, Fenchel's duality theorem translates by lemma 5 to

$$\min_{\phi \in X} \{f(\phi) + g(A\phi)\} = \max_{p_y \in Y^\star} \{-g^\star(-p_y) - f^\star(A^\star p_y)\} \qquad (2.85)$$

The next proposition shows that the saddle-points $(\phi^\star, p^\star)$ of the saddle-point equation in eq. (2.63) relate the subdifferentials of f and $g_\star$ to one another.

Proposition 2 (Kuhn-Tucker conditions). *Let eq. (2.63) be satisfied for a saddle-point $(\phi^\star, p^\star)$. Then the subdifferential equations*

$$A^\star p^\star \in \partial f(\phi^\star), \quad A\phi^\star \in \partial g_\star(p^\star) \qquad (2.86)$$

hold.

Proof. Let $(\phi^\star, p^\star)$ be a saddle-point of Fenchel's duality theorem in eq. (2.63). It follows that

$$f(\phi^\star) + f^\star(A^\star p^\star) = g_\star(p^\star) + g(A\phi^\star) \qquad (2.87)$$

Eq. (2.87) simultaneously imposes equality in eqns. (2.47) and (2.48)

$$f(\phi^\star) + f^\star(A^\star p^\star) = \langle p^\star, A\phi^\star \rangle \qquad (2.88)$$
$$g(A\phi^\star) + g_\star(p^\star) = \langle p^\star, A\phi^\star \rangle \qquad (2.89)$$

and the conditions in eq. (2.86) follow by applying lemma 6 to eqns. (2.88) and (2.89). $\qquad \square$

With the help of the Kuhn-Tucker conditions in eq. (2.86) we can formulate a condition for the minimizer $\phi^\star$ alone which resembles the condition in $0 \in \partial f(\phi^\star)$ (eq. (2.17)) but for the function $E(\phi, A\phi)$: Since the Kuhn-Tucker conditions in eq. (2.86) hold for the saddle-point $(\phi^\star, p^\star)$ and $p^\star \in \partial g(A\phi^\star)$ (by lemma 6) we have

$$0 = A^\star p^\star - A^\star p^\star \in \partial f(\phi^\star) - A^\star \partial g(A\phi^\star) = \partial E(\phi^\star, A\phi^\star) \qquad (2.90)$$

where $A^\star \partial g(A\phi^\star)$ implies that $A^\star$ is applied to every element of the set $\partial g(A\phi^\star)$. Eq. (2.90) states that the minimizer $\phi^\star$ of $E(\phi, A\phi)$ can be characterized as the point where the subdifferentials ∂f and ∂g balance each other.

Theorem 2 (Morreau's Theorem). *Let $f : X \to \mathbb{R}$ be a closed, proper and convex function on the Hilbert space X and let $w(z) = \frac{1}{2} \langle z, z \rangle$ for any $z \in X$. Then*

$$\min_{\phi \in X} \{f(\phi) + w(z - \phi)\} + \min_{p \in X^\star} \{f^\star(p) + w(z - p)\} = w(z) \qquad (2.91)$$

holds. The pair of minimizers $(\bar{\phi}, \bar{p}) \in X \times X^\star$ is unique for every $z \in X$ and provides a decomposition of z, $z = \bar{\phi} + \bar{p}$.

Proof. We fix $z \in X$ and define $g(\phi) = -w(z - \phi)$. The conjugate of g is

$$g_\star(p) = -w(z - p) + w(z) \tag{2.92}$$

We insert f, g and eq. (2.92) into eq. (2.63) of Fenchel's Duality theorem and directly verify eq. (2.91)

$$\min_{\phi \in X} \{f(\phi) + w(z - \phi)\} = \max_{p \in X^\star} \{-w(z - p) - f^\star(p)\} + w(z) \tag{2.93}$$

$$= -\min_{p \in X^\star} \{w(z - p) + f^\star(p)\} + w(z) \tag{2.94}$$

$$\min_{\phi \in X} \{f(\phi) + w(z - \phi)\} + \min_{p \in X^\star} \{w(z - p) + f^\star(p)\} = w(z) \tag{2.95}$$

The pair of minimizers $(\bar{\phi}, \bar{p})$ is unique since the function $w(z)$ is strictly convex. Lastly we have to show that $(\bar{\phi}, \bar{p})$ decomposes z, $z = \bar{\phi} + \bar{p}$: The Kuhn-Tucker conditions are

$$\bar{p} \in \partial f(\bar{\phi}), \quad \bar{\phi} \in \partial g_\star(\bar{p}) \tag{2.96}$$

However since $g_\star$ is smooth we have

$$\bar{\phi} = \nabla g_\star(p)|_{p=\bar{p}} = z - \bar{p} \tag{2.97}$$

from which follows that $z = \bar{\phi} + \bar{p}$. $\qquad\square$

From from the decomposition eq. (2.97) and the definition of the proximal operator in eq. (2.21) it follows that

$$z = \mathrm{prox}\,(z|f) + \mathrm{prox}\,(z|f^\star) \tag{2.98}$$

Hence $\mathrm{prox}\,(z|f)$ can be computed purely in terms of the conjugate function $f^\star$ in cases where $\mathrm{prox}\,(z|f^\star)$ is simpler to compute.

2.2.3. Primal Dual Splitting

In this section we will introduce the splitting method first described by Pock et. al in [89] and refined in [17] which is a primal-dual method for solving the saddle-point problem in eq. (2.63) for a given convex function $E(\phi, A\phi) = g(A\phi) + f(\phi)$. Splitting methods generally include an iterative strategy in which

---

Algorithm 1 Chambolle Pock Primal-Dual Splitting

Choose parameters $\tau, \sigma > 0$, $\theta \in (0,1)$ and initial guesses $\bar{\phi}^0$, p^0 and $\phi^0 = \bar{\phi}^0$
while $G(\phi^n, p^n) > \epsilon$ or $n < M$ **do**
    Ascending step: $p^{n+1} = \text{prox}(p^n + \sigma A\phi^n | \sigma f^\star)$
    Descending step: $\bar{\phi}^{n+1} = \text{prox}(\bar{\phi}^n - \tau A^\star p^{n+1} | \tau g)$
    Interpolation step: $\phi^{n+1} = \bar{\phi}^{n+1} + \theta(\bar{\phi}^{n+1} - \bar{\phi}^n)$
end while

The ascending step moves the dual variable p in the direction which maximizes the dual function $E^\star(p, A^\star p)$ and the descending step moves the primal variable ϕ in the direction which minimizes the primal function $E(\phi, A\phi)$. Since $\bar{\phi}^{n+1}$ is computed from p^{n+1} and thus runs one step ahead, the interpolation set is needed to pull it back or else ascending and descending step may begin to periodically oscillate.

---

the convex functions g and f are deployed independently in two successive iterations. Among the first widely used splitting algorithms is the forward-backward algorithm introduced in [64] and further developed in [10, 21, 87]. The forward-backward algorithm combines a forward-gradient step involving g (eq. (2.36)) with a proximal step (eq. (2.33)) involving f

$$\phi^{n+1} = \text{prox}(\phi^n - \tau \nabla g(A\phi^n) | \sigma f) \tag{2.99}$$

The iteration in eq. (2.99) however is based upon the assumption that the function g is smooth while the function f is possibly non-smooth. The Douglas-Rachford algorithm [64, 26, 22, 108] relaxes the smoothness criterion on g by introducing an auxiliary variable v and the update rules

$$\lambda^n \in [\epsilon, 2 - \epsilon] \tag{2.100}$$
$$v^{n+1} = v^n + \lambda^n \left(\text{prox} \left(2\phi^n - v^n | \tau g \right) - \phi^n \right)$$
$$\phi^{n+1} = \text{prox} \left(v^{n+1} | \tau f \right) \tag{2.101}$$

where $\epsilon \in]0, 1[$. Despite being more general then the forward-backward algorithm, the Douglas-Rachford algorithm can be computationally more demanding, depending on the functions g and f, since it contains two proximal operators. See [23] for an overview of both algorithms and their applications.

The splitting algorithm which was proposed by Pock et. al. in [89] is based upon the following observation: Since g is assumed to be convex, proper and lower semi-continuous it is closed under the conjugate operation in eq. (2.41), $g = g^{\star\star}$

and thus

$$g(A\phi) = \max_{p \in Y^\star} \left\{ \langle p, A\phi \rangle - g^\star(p) \right\} \tag{2.102}$$

Thus the minimization problem

$$\min_{\phi \in X} E(\phi, A\phi) = \min_{\phi \in X} \left\{ g(A\phi) + f(\phi) \right\} \tag{2.103}$$

is equivalent to the saddle-point point problem

$$\min_{\phi \in X} E(\phi, A\phi) = \min_{\phi \in X} \max_{p \in Y^\star} \left\{ \langle p, A\phi \rangle - g^\star(p) + f(\phi) \right\} \tag{2.104}$$

Hence the authors in [89] propose an algorithm (see alg. 1) which beginning with a pair of initial guesses (ϕ^0, p^0) consists of an *ascending* iteration in the dual variable p and a *descending* iteration and subsequent interpolation step in the primal variable ϕ. After computing the dual update p^{n+1} from p^n and ϕ^n the update $\bar{\phi}^{n+1}$ is computed from p^{n+1}. Thus $\bar{\phi}^{n+1}$ is always *one step ahead* of p^{n+1}. To mitigate this effect an interpolation (ϕ^{n+1}) of $\bar{\phi}$ between the updated state and the current state is saved for the computation of the next update of the dual variable p. In the case of interpolation with $\theta = 1$ and $\tau\sigma \|A\|^2 \leq 1$ convergence of the algorithm is guaranteed [89, 17].

To access the convergence of the algorithm the authors in [89] introduced the *partial primal-dual Gap*

$$G_{X \times Y^\star}(\phi, p) = E(\phi, A\phi) - E^\star(A^\star p, p) \tag{2.105}$$

$$E(\phi, A\phi) = \max_{p' \in Y^\star} \left\{ \langle p', A\phi \rangle - g^\star(p') + f(\phi) \right\} \tag{2.106}$$

$$E^\star(p, A^\star p) = \min_{\phi' \in X} \left\{ \langle A^\star p, \phi' \rangle - g^\star(p) + f(\phi') \right\} \tag{2.107}$$

From Fenchel's theorem in eq. (2.85) we have

$$\min_{\phi \in X} E(\phi, A\phi) = \max_{p \in Y^\star} E^\star(p, A^\star p) \tag{2.108}$$

from which it follows that the primal-dual gap $G_{X \times Y^\star}$ is positive semi-definite

$$G_{X \times Y^\star}(\phi, p) \geq 0 \tag{2.109}$$

The points $(\phi^\star, p^\star)$ for which the primal-dual gap vanishes, $G_{X \times Y^\star}(\phi^\star, p^\star) = 0$ are saddle-points. Given the initial guesses ϕ^0 and p^0 the authors in [17] proved that for any pair of Hilbert spaces X and Y, $G_{X \times Y^\star}$ is bound from above and

that alg. 1 has $O(\frac{1}{N})$ convergence

$$G_{X \times Y^\star}(\phi^N, p^N) \leq \frac{D(\phi^0, p^0)}{N}, \quad \phi^N = \frac{1}{N} \sum_{i=1}^{N} \phi^i, \; p^N = \frac{1}{N} \sum_{i=1}^{N} p^i \qquad (2.110)$$

And in the limit $N \to \infty$ the cluster point (ϕ^N, p^N) converges to a saddle-point $(\phi^\star, p^\star)$ of the saddle-point problem in eq. (2.63). Furthermore in the case when X and Y are finite dimensional, $X \times Y \subset \mathbb{R}^q \times \mathbb{R}^r$ the sequence (ϕ^n, p^n) itself converges to the same saddle-point $(\phi^\star, p^\star)$. Hence the iterations (ϕ^n, p^n) of any implementation of alg. 1 on a computer necessarily converge to an approximate solution of the saddle-point problem in eq. (2.63). An improved version of alg. 1 was additionally proposed in [17] which dynamically adapts the step parameters τ and σ at every iteration n of the algorithm and achieves a convergence rate of the order $O(\frac{1}{N^2})$. However we will use alg. 1 in this thesis due to its simplicity and in chapter 4 we will propose an extension which yields $O(\frac{1}{2N})$ convergence rate.

2.3. Principle of Least Action

Eq. (2.7) is the manifestation of the principle of least action (PLA) [61, 24, 88, 36, 49]. The origins of the PLA date back to the year 1740. In that year Pierre-Louis Maupertuis introduced the action E of a body of mass m as a quantity which is minimal when the body is at rest [24]. He further went on in 1746 to extend his principle to systems of non interacting point masses m_i which follow trajectories $\boldsymbol{q}_i(t)$ [25]. In that paper he showed that the action E of the many body system has the form of a time integral

$$E(\dot{\boldsymbol{q}}) = \int \frac{1}{2} \sum_i m_i \left\| \dot{\boldsymbol{q}}_i(t) \right\|^2 dt \qquad (2.111)$$

Maupertuis showed that the trajectories $\boldsymbol{q}_i^\star(t)$ which minimize eq. (2.111) are also the solutions to Newtons equations

$$m_i \ddot{\boldsymbol{q}}_i^\star(t) = 0 \qquad (2.112)$$

realizing Newtons first axiom *A body of mass m remains in the state of rest or uniform linear motion unless a force is acted upon it to change its state* [79].

The PLA was extended in 1750 by Leonhard Euler and Joseph Lagrange to incorporate interactions between the point masses m_i. They defined the action $E(\dot{\boldsymbol{q}}, \boldsymbol{q})$ as the time integral over a function $\mathcal{L}(\dot{\boldsymbol{q}}, \boldsymbol{q}, t)$, now called the Lagrangian,

and showed that $E(\dot{q}, q)$ is minimal when the trajectories fulfill the equation

$$\frac{\delta \mathcal{L}}{\delta q_i(t)} - \frac{d}{dt}\left(\frac{\delta \mathcal{L}}{\dot{q}_i}(t)\right) = 0 \tag{2.113}$$

The differential equations in eq. (2.113) are named in honor of the originators the Euler-Lagrange equations.

In the context of Gibbs random field theory and the minimization problem in eq. (2.7) the energy $E(\phi, \nabla\phi)$ can be written in terms of an integral over the Lagrangian $\mathcal{E}(\phi(\boldsymbol{x}), \nabla\phi(\boldsymbol{x}))$

$$E(\phi, \nabla\phi) = \int_{\Omega} \mathcal{E}(\phi(\boldsymbol{x}), \nabla\phi(\boldsymbol{x}))d^n x \tag{2.114}$$

The Euler-Lagrange equations in eq. (2.113) then generalize to the GRF field equations

$$\frac{\delta \mathcal{E}}{\delta \phi(\boldsymbol{x})} - \sum_{\mu} \frac{d}{dx_{\mu}}\left(\frac{\delta \mathcal{E}}{\partial_{\mu}\phi(\boldsymbol{x})}\right) = 0 \tag{2.115}$$

The Euler-Lagrange GRF equations in eq. (2.115) are the central equation in many physical theories which can be modeled in the context of Gibbs random fields. For instance in classical electrodynamics [62, 38] the GRF of the theory is the vector $\boldsymbol{A}(\boldsymbol{x}, t)$ which generates the electric field $\boldsymbol{E}(\boldsymbol{x}, t)$ and the magnetic field $\boldsymbol{B}(\boldsymbol{x}, t)$. The Euler-Lagrange equations of the density $\mathcal{E}(\boldsymbol{A}(\boldsymbol{x}, t), \nabla\boldsymbol{A}(\boldsymbol{x}, t))$ are the celebrated Maxwell-Equations [70], which are dynamical equations relating $\boldsymbol{E}(\boldsymbol{x}, t)$ and $\boldsymbol{B}(\boldsymbol{x}, t)$. Through the Maxwell equations James Clerk Maxwell was able to predict the electromagnetic waves, the existence of which was proved by the physicist Heinrich Hertz [50].

2.4. Image De-Noising

An example of a physical system containing a GRF is a camera C recording an object O. The domain $\Omega \subset \mathbb{R}^2$ is the focal plane of the camera C and the object O is naturally projected onto the focal plane Ω producing the projection I_O. In theory the projection I_O is a continuous function in the coordinate frame of the plane O where the particular function value $I_O(\boldsymbol{x})$ is the light intensity the object O reflects to the point $\boldsymbol{x}$ on the focal plane Ω. At the heart of the image acquisition process of basically all modern camera systems lies the concept of a
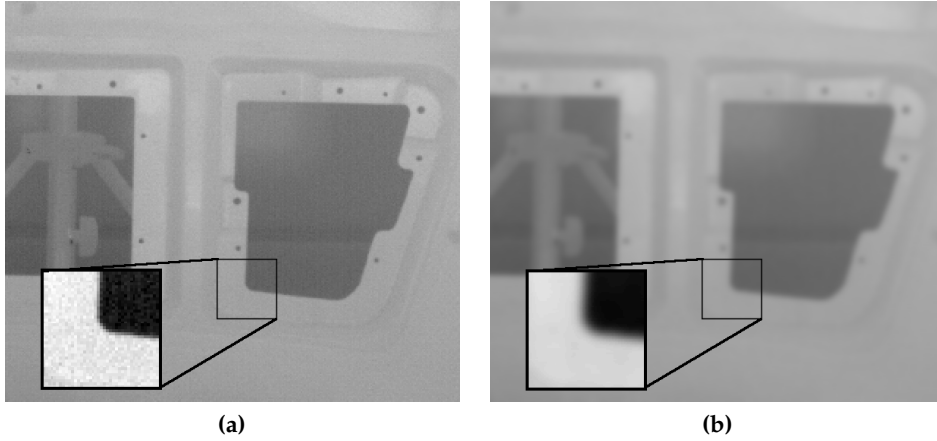
(a)	(b)

Figure 2.2.: Figure 2.2a shows an image I^c taken of an object O with a thermographic camera. A region of interest is shown where the contrast was enhanced to visualize the noise corruption. Figure 2.2b shows the result $I_O^\star$ of the minimization problem eq. (2.121) with the prior in eq. (2.122). The noise is removed but the boundaries of O are over smoothed

CCD collecting the photons of the light at discrete positions $\boldsymbol{x}_{i,j}$ called pixels

$$I_{ij}^c = I^c(\boldsymbol{x}_{i,j}) \in \mathbb{R}, \quad \boldsymbol{x}_{i,j} \in \Omega \quad 1 < i < n, 1 < j < m \tag{2.116}$$

The observables Y are the recorded intensities I_{ij}^c at the pixels $\boldsymbol{x}_{i,j}$. In this sense the camera C is a function which maps the continuous projection $I_O(\boldsymbol{x})$ to the discretely sampled intensities I_{ij}^c

$$I_{ij}^c = C_{ij}\left(I_O\right) \tag{2.117}$$

The intensity I_{ij}^c is basically a function of the number of photons collected by the CCD at the pixel $\boldsymbol{x}_{i,j}$. This number cannot be acquired deterministically, it is rather the result of a stochastic process described as independently identically distributed (iid) noise

$$\hat{I}_{ij}^c = I_O\left(\boldsymbol{x}_{i,j}\right) + n \quad n \sim p\left(I_{ij}^c | I_O\left(\boldsymbol{x}_{i,j}\right)\right) \tag{2.118}$$

$p(I_{ij}^c | I_O\left(\boldsymbol{x}_{i,j}\right))$ is the likelihood that I_{ij}^c assumes the value $\hat{I}_{ij}^c$ given the incoming intensity $I_O(\boldsymbol{x}_{i,j})$ at the pixel $\boldsymbol{x}_{i,j}$. Like in eq. (2.5) it is mapped to the data term energy $E_{I^c}(I_O)$.

In order to infer the values of $I_O(\boldsymbol{x}_{i,j})$ at the pixels $\boldsymbol{x}_{i,j}$ from the noisy data I_{ij}^c we need to pose some form of regularity on the values $I_O(\boldsymbol{x})$ to counter the pixel-wise noise imposed by the CCD in eq. (2.118). Such regularity can be achieved

by correlating the intensities $I_O(\boldsymbol{x})$ at all pixels with each other in the prior

$$p\left(I_O\right) = \exp\left(-E^{prior}\left(I_O\right)\right) \tag{2.119}$$

$$E^{prior}\left(I_O\right) = \int_\Omega \mathcal{E}\left(I_O\left(\boldsymbol{x}\right), I_O\left(\Omega/\left\{\boldsymbol{x}\right\}\right)\right) dx \tag{2.120}$$

where the integrand correlates the intensity $I_O(\boldsymbol{x})$ at the point $\boldsymbol{x} \in \Omega$ with the intensities at all other points $\Omega/\left\{\boldsymbol{x}\right\}$ so that the problem of inferring I_O from the data I^c becomes the minimization problem

$$I_O^\star = \operatorname{argmin}_{I_O}\left(E_{I^c}\left(I_O\right)\right), \quad E_{I^c}\left(I_O\right) = E_{I^c}^{data}(I_O) + E^{prior}\left(\nabla I_O\right) \tag{2.121}$$

However in practice for a $n \times n$ dimensional image I^c the minimization in eq. (2.121) achieves a complexity of the order $\mathcal{O}(n^4)$ since every pixel is correlated to $n^2 - 1$ other pixels. Even for medium sized images with $n = 500$ the computations involved in eq. (2.121) are practically infeasible.

To reduce the complexity we want the integrand $\mathcal{E}$ in eq. (2.120) only to correlate the values $I_O(\boldsymbol{x})$ within a neighborhood $U_{\boldsymbol{x}_{i,j}} \subset \Omega$ with each other. One possible and very simple way to implement $\mathcal{E}$ is to have it penalize the L_2 norm of the gradient $\nabla I_O(\boldsymbol{x})$

$$E_{L_2}^{prior}\left(\nabla I_O\right) = \int_\Omega \left\|\nabla\, I_O\left(\boldsymbol{x}\right)\right\|^2 dx \tag{2.122}$$

where the gradient operation ∇ can be realized by finite differences. While the prior in eq. (2.122) can be implemented in a very efficient manner, it has an important drawback. It isotropically smooths the GRF I_O regardless of the underlying geometry of the object O being recorded. In figure 2.2a the image I^c of an object O recorded by a thermographic camera is shown. A region of interest with enhanced contrast is shown to visualize the noise corruption due to the image measuring process in eq. (2.118). Figure 2.2b shows the result of the minimization in eq. (2.121) with the L_2 prior in eq. (2.120). $E_{L_2}^{prior}$ reduces the noise in I_O but due to its isotropic nature it over-smooths the boundaries of O. In section 2.5 and following we will introduce a methodology aimed at designing prior energies E^{prior} which incorporate information about the geometry of the objects recorded in order to avoid the over-smoothing across their boundaries.
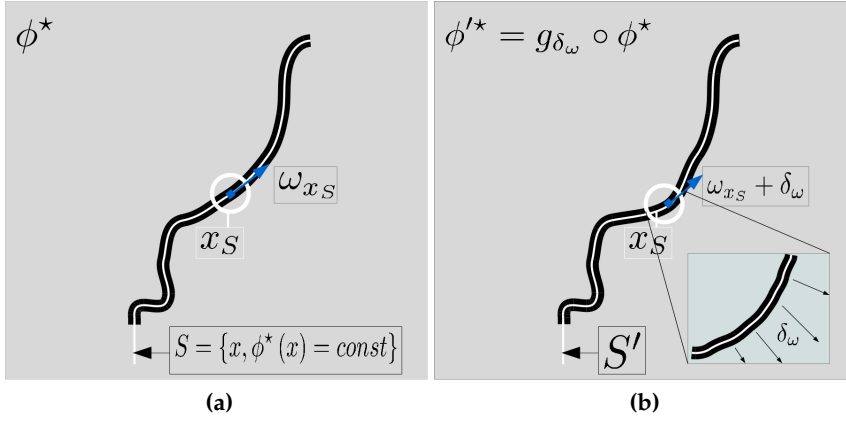
Figure 2.3.: Local transformation of an image ϕ with a level-set S. Figure 2.3a shows an image $\phi(\boldsymbol{x})$ with a line S along which the intensity values are constant. At each point $\boldsymbol{x}_S$ the vector $\boldsymbol{\omega}_S$ is the tangential vector on S. Figure 2.3b shows the result of the local distortion of S under the action of the operator g_{δ_ω}. g_{δ_ω} acts on S by adding to $\boldsymbol{\omega}_S$ a spacial dependent vector $\boldsymbol{\delta}_\omega(\boldsymbol{x})$

2.5. Lie Groups and the Noether Theorem

2.5.1. Motivation 1, the problem

In section 2.1 we had claimed that the problem with the L_2 prior

$$E_{L_2}(\phi) = \int_\Omega \|\nabla\phi\|^2 \tag{2.123}$$

over-smooths the GRF ϕ over the boundaries of the object recorded by the camera C. In general the minimizers $\phi^\star$ of the energy E_{L_2} are the constant functions $\phi = const$

$$A_c = \left\{ \phi_c^\star \,\middle|\, \phi_c^\star = \operatorname{argmin}_\phi(E_{L_2}(\nabla\phi)) = c, \quad c \in \mathbb{R} \right\} \tag{2.124}$$

In the context of the minimization problem in eq. (2.121) the minimizer set A_c in eq. (2.124) emphasizes that the prior E_{L_2} does not allow for the solution $I_O^\star$ (eq. (2.121)) to have discontinuities. Thus E_{L_2} is completely unaware of the geometry in the data I^c (figure 2.2a). However E_{L_2} has a advantageous property. Consider the set of rotations $SO(2)$ of the coordinate frame Ω

$$\boldsymbol{x}' = \mathbf{R}_\theta \boldsymbol{x}, \quad \mathbf{R}_\theta = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \in SO(2) \tag{2.125}$$

The gradient $\nabla\phi$ transforms under the rotation in eq. (2.125) like a vector, $\nabla'\phi = \mathbf{R}_\theta \nabla\phi$ and the matrix $\mathbf{R}_\theta$ satisfies $\mathbf{R}_\theta^T \mathbf{R}_\theta = \mathbb{1}$. Thus the L_2 energy is also invariant towards the rotations in eq. (2.125)

$$E'_{L_2} = \int_\Omega \nabla^T \phi \mathbf{R}_\theta^T \mathbf{R}_\theta \nabla\phi d^2x = \int_\Omega \|\nabla\phi\|^2 \, d^2x \qquad (2.126)$$

In general the invariance of the prior energy $E^{prior}(\nabla\phi)$ of a GRF ϕ with respect to the rotations in eq. (2.125) is a favorable feature since the gradient $\nabla\phi$ should not be penalized to a specific orientation. In the context of the minimization problem in eq. (2.121) rotational invariance of the prior $E^{prior}(\nabla I_O)$ ensures the gradient $\nabla I_O^\star$ is not affected by the orientation of the camera system C.

Several methods have been introduced which allow for the construction of anisotropic priors. These methods include TV-Regularization [100, 16] which will be introduced in section 2.9 , anisotropic difusion guided by directional operators like the structure tensor [80, 116, 117] and level set methods of the Mumford-Shah type [75, 1, 113]. Among the earliest attempts for anisotropic regularization was the work of Nagel et. al. [76]. They introduced a quadratic prior

$$E_D^{prior} (\nabla\phi) = \int (\nabla\phi(\boldsymbol{x}))^T D(\boldsymbol{x}) (\nabla\phi(\boldsymbol{x})) \, d^2x \qquad (2.127)$$

The operator $D(\boldsymbol{x})$ is a local 2×2 symmetric valued matrix estimated within a local window around each point $\boldsymbol{x}$. $D(\boldsymbol{x})$ is precomputed and assumed to be fixed under variation of ϕ. Thus it's eigenvectors function as a guide for the gradient ϕ. For instance in eq. (2.121) we can insert eq. (2.127) for E^{prior}. Computing D such that it has only one non-zero eigenvalue λ and an eigenvector $\boldsymbol{b}$ oriented perpendicular to the weighted gradient of the data I^c

$$D(\boldsymbol{x}) = \lambda\boldsymbol{b}(\boldsymbol{x})\boldsymbol{b}(\boldsymbol{x})^T, \quad \boldsymbol{b}(\boldsymbol{x}_0) \perp \langle \nabla I^c(\boldsymbol{x}) \rangle (\boldsymbol{x}_0) \qquad (2.128)$$

$$\langle \nabla I^c(\boldsymbol{x}) \rangle (\boldsymbol{x}_0) = \int_A w(\|\boldsymbol{x} - \boldsymbol{x}_0\|) \nabla I^c(\boldsymbol{x}) \, d^2x \qquad (2.129)$$

the prior E_D^{prior} penalizes the tangential component of ∇I_O along $\boldsymbol{b}$ in the minimization in eq. (2.121). Thus the solution $I_O^\star$ can have discontinuities perpendicular to $\boldsymbol{b}$. The drawback of E_D^{prior} is that we do not know if $\boldsymbol{b}$ is the true tangential vector in the unbiased projection of the object O. And since D is fixed E_D^{prior} can not be invariant under the rotations in eq. (2.125). Thus the minimization in eq. (2.121) can produce a solution $I_O^\star$ in eq. (2.121) that has discontinuities which do not reflect the true boundaries of the object O. We conclude that prior energies E_D^{prior} which are *not* rotation invariant are a source of error for the orientation of

$\nabla I_O^\star$ in eq. (2.121).

On the other side a potential anisotropic prior E^{prior} which *is* rotation invariant would lead to a solution $I_O^\star$ in eq. (2.121) for which the orientation of its structures is only determined by the data term E^{data}.

In the following we will introduce a methodology which allows us to characterize prior energies $E^{prior}(\nabla\phi)$ which allow for discontinuities in their minimizers $\phi^\star = \mathrm{argmin}_\phi E^{prior}(\nabla\phi)$ while remaining invariant to a specified but more general set of spacial transformations $\mathbb{G}^\Omega$.

2.5.2. Motivation 2, the solution

Another way to state the problem that the prior energy E_{L_2} only allows for constant minimizers $\phi^\star = const$ (eq. (2.124)) goes as follows. The energy $E_{L_2}(\nabla\phi)$ is invariant upon the transformation $\phi'(\boldsymbol{x}) = \phi(\boldsymbol{x}) + d$ where d is a constant over Ω. Thus if $\phi_0^\star = c'$ is a minimizer of E_{L_2}, $c' \in A_c$ then so is $\phi'^\star = c' + d$ since $c' + d \in \mathbb{R}$ and by the definition of A_c in eq. (2.124) we have $\phi'^\star \in A_c$. We would like to think of the operation of addition with constants d as a set $\mathbb{G}_{const}$ of operators g_d

$$g. : \mathbb{R} \to \mathbb{G}_{const}, \quad g_d = \cdot + d, \quad g_d \in \mathbb{G}_{const} \tag{2.130}$$

With the help of the construction in eq. (2.130) we can restate the invariance of E_{L_2} in the following way

$$g_d \circ E_{L_2}(\nabla\phi) = E_{L_2}(\nabla(\phi + d)) = E_{L_2}(\nabla\phi) \tag{2.131}$$

and A_c in eq. (2.124) can be viewed as being spun by *one* constant function $\phi_0^\phi(\boldsymbol{x}) = c$ and the set $\mathbb{G}_{const}$

$$A_c = \{\phi^\star \,|\, \phi^\star = g_d \circ \phi_0^\star, \quad g_d \in \mathbb{G}_{const}\} \tag{2.132}$$

With the constructions in eq. (2.130) and eq. (2.132) the problem statement that the prior E_{L_2} only allows for constant minimizers is transfered to the statement that the set $\mathbb{G}_{const}$ under which E_{L_2} is invariant is *too small* in some sense.

A more flexible prior energy E^{prior} should be invariant to a more general set of transformations $\mathbb{G}^\phi$. At the same time E^{prior} should also be invariant to a spacial set of transformations $\mathbb{G}^\Omega$ in order for it not to impede the orientation of the gradient $\nabla\phi$ as motivated in section 2.5.1. Hence E^{prior} is assumed to be

invariant to the set $\mathbb{G} = \mathbb{G}^\phi \times \mathbb{G}^\Omega$ with the actions

$$g_{\omega^\phi} \circ \phi(\boldsymbol{x}) = \phi(\boldsymbol{x}) + \omega^\phi(\boldsymbol{x}), \quad g_{\omega^\phi} \in \mathbb{G}^\phi \tag{2.133}$$

$$g_{\omega^\Omega} \circ \boldsymbol{x} = \boldsymbol{x} + \boldsymbol{\omega}^\Omega(\boldsymbol{x}), \quad g_{\omega^\Omega} \in \mathbb{G}^\Omega \tag{2.134}$$

The transformations in eq. (2.133) and eq. (2.134) formally capture all the possible transformations the prior energy E^{prior} is invariant to. In this sense $\mathbb{G}$ is maximal and E^{prior} is invariant upon the action the entire set $\mathbb{G}$

$$g \circ E^{prior} = E^{prior}, \quad \forall g \in \mathbb{G} \tag{2.135}$$

For instance the prior energy E_{L_2} is conditionally invariant to the set $\mathbb{G} = \mathbb{G}_{const} \times SO(2)$, the set of addition of the variable ϕ with constants (eq. (2.130)) and the set of rotations in Ω (see eq. (2.125)).

Similar to the definition of A_c in eq. (2.132) we can describe the minimizers of E^{prior} as being related to each other by the elements of $\mathbb{G}$

$$A = \{\phi^\star \,|\, \phi^\star = g \circ \phi_0^\star \quad g \in \mathbb{G}\} \tag{2.136}$$

The set $\mathbb{G}^\Omega$ contains operators which are purely geometric. The idea is to show that A may be split into sub sets $A_\Omega(\phi_c^\star)$ whose elements are related to each other by the elements $g_{\omega^\Omega} \in \mathbb{G}^\Omega$

$$A_\Omega(\phi_c^\star) = \left\{\phi^\star \,\middle|\, \phi^\star(\boldsymbol{x}) = \phi_c^\star(g_{\omega^\Omega} \circ \boldsymbol{x}), \quad g_{\omega^\Omega} \in \mathbb{G}^\Omega\right\} \tag{2.137}$$

$$A = \left\{A_\Omega(\phi_c^\star) \,\middle|\, \phi_c^\star = g_{\omega^\phi} \circ \phi_0^\star, \quad g_{\omega^\phi} \in \mathbb{G}^\phi\right\} \tag{2.138}$$

This is significant for the following reason: knowledge of the geometric set of transformations $\mathbb{G}^\Omega$ under which E^{prior} is invariant allows for a reduction of the set of maximizers A to a set A_{red} such that the elements $\phi_c^\star \in A_{red}$ are not related to each other by $\mathbb{G}^\Omega$

$$A_{red} = \left\{\phi_c^\star \,\middle|\, \phi_c^\star = g_{\omega^\phi} \circ \phi_0^\star, \quad g_{\omega^\phi} \in \mathbb{G}^\phi\right\} \tag{2.139}$$

$$\phi_d^\star(\boldsymbol{x}) \neq \phi_c^\star(g_{\omega^\Omega} \circ \boldsymbol{x}) \quad \forall g_{\omega^\Omega} \in \mathbb{G}^\Omega, \phi_{c,d}^\star \in A_{red} \tag{2.140}$$

We may also turn the argument around: we could specify the geometric set of transformations $\mathbb{G}^\Omega$ and design a prior $E^{prior}(\nabla\phi)$ which is conditionally invariant under $\mathbb{G}^\Omega$, thus having a reduced maximizer set A_{red}. To give a hint of how the prior $E^{prior}(\nabla\phi)$ could be designed we need the definition of a level-set. A level-set of an image $\phi_0^\star$ is a sub set $S_c \subset \Omega$ defined by

$$S_c = \{\boldsymbol{x} \,|\, \phi_0^\star(\boldsymbol{x}) = c\} \tag{2.141}$$

By the definition of the action of $g_{\omega\Omega}$ in eq. (2.134) we see that $g_{\omega\Omega}$ is a geometrical transformation that deforms the level-sets S_c (see figure 2.3). We are free to define $g_{\omega\phi}$ so that it is orthogonal to $g_{\omega\Omega}$ in the sense that the level-sets S_c are invariant under $g_{\omega\phi}$

$$S_{c'} = g_{\omega\phi} \circ S_c = S_c \tag{2.142}$$

since a transformation of S_c is purely geometrical. Now the level-set S_c may alternatively be defined with the help of the vector-field $\boldsymbol{\omega}_\delta(\boldsymbol{x})$ which (see figure 2.3) is the set of vectors *tangent* to S_c

$$S_c = \{\boldsymbol{x} \,|\, \boldsymbol{\omega}_\delta(\boldsymbol{x}) \cdot \nabla\phi_0^\star(\boldsymbol{x}) = 0\} \tag{2.143}$$

In figure 2.3b we show an example of a level-set S which is distorted by the operator $g_{\omega_\delta} \in \mathbb{G}^\Omega$. The resulting level-set S' has the vector-field $\boldsymbol{\omega}_\delta'(\boldsymbol{x}) = \boldsymbol{\omega}_\delta(\boldsymbol{x}) + \boldsymbol{\delta}(\boldsymbol{x})$ as tangent vectors.

$$S_c' = \{\boldsymbol{x} \,|\, (\boldsymbol{\omega}_\delta(\boldsymbol{x}) + t\boldsymbol{\delta}(\boldsymbol{x})) \cdot \nabla\phi_0^\star(\boldsymbol{x}) = 0\} \tag{2.144}$$

However it also possible to represent S_c' with the help of a deformation of the gradient operator ∇ itself

$$S_c' = \{\boldsymbol{x}' \,|\, \boldsymbol{\omega}_\delta(\boldsymbol{x}') \cdot \nabla_{t\boldsymbol{\delta}}\phi_0^\star(\boldsymbol{x}') = 0\} \tag{2.145}$$

The operator $\nabla_{t\boldsymbol{\delta}}$ loosely speaking encodes a reversal of the action of $g_{\omega\Omega}$ on $\boldsymbol{x}$ so that S_c' can be represented with the same tangential vector-field as S_c but in the new frame $\boldsymbol{x}' = g_{\omega_\delta} \circ \boldsymbol{x}$. The operator $\nabla_{t\boldsymbol{\delta}}$ is called a *push forward* of the gradient ∇. With the help of the push forward $\nabla_{t\boldsymbol{\delta}}$ it is possible to translate the notion of invariance with respect to $\mathbb{G}^\Omega$ to the requirement that $E^{prior}(\nabla_{t\boldsymbol{\delta}}\phi)$ must be constant with respect to variations of the vector-field $\boldsymbol{\delta}(\boldsymbol{x})$

$$\frac{d}{dt}E^{prior}(\nabla_{t\boldsymbol{\delta}}\phi)\bigg|_{t=0} = 0 \tag{2.146}$$

Given a specific form of the operators in $\mathbb{G}^\Omega$, eq. (2.146) poses constraints on the form of the differential operators in the prior $E^{prior}(\nabla_{t\boldsymbol{\delta}}\phi)$. Eq. (2.146) also ensures that $E^{prior}(\nabla_{t\boldsymbol{\delta}}\phi)$ is indifferent to a large class of level-sets $\{S\}$, which are generated by $\mathbb{G}^\Omega$ acting on S (see eq. (2.145)).

2.6. Lie Groups

We will now give a more formal description of the transformation group $\mathbb{G} = \mathbb{G}^{\phi} \times \mathbb{G}^{\Omega}$ represented by the transformations in eq. (2.133) and eq. (2.134). Following this section we will make the invariance property of E^{prior} in eq. (2.146) formally more explicit in section 2.7. In this section the set of operators $\mathbb{G}$ is taken to act on a vector space $\mathcal{M}$.

Definition 9 (Lie Group). *Let $\mathbb{G}$ be a smooth manifold. $\mathbb{G}$ is called a Lie Group if there exists a multiplication operation $'\cdot : \mathbb{G} \times \mathbb{G} \to \mathbb{G}'$ such that the following elements exist*

- *Neutral element e: $e \cdot g = g$ for all $g \in \mathbb{G}$*

- *Inverse element: $i(g) = g^{-1} \in \mathbb{G}$ for all $g \in \mathbb{G}$*

- *Product: $g_1 \cdot g_2 \in \mathbb{G}$ for all $g_{1,2} \in \mathbb{G}$*

The group operation $'\cdot'$ can also be used to define the *left action* l_g and the *right action* r_g on $\mathbb{G}$

Definition 10 (Left and Right Action). *The left action l_g and the right action r_g on a Lie group $\mathbb{G}$ is the mapping*

$$l_g : \mathbb{G} \to \mathbb{G} \quad l_g(h) = g \cdot h \quad g, h \in \mathbb{G} \tag{2.147}$$
$$r_g : \mathbb{G} \to \mathbb{G} \quad r_g(h) = h \cdot g \quad g, h \in \mathbb{G} \tag{2.148}$$

Since $\mathbb{G}$ is assumed to be a smooth manifold the following lemma holds

Lemma 7 (Smoothness of $l_g(h)$, $r_g(h)$ and $i(h)$). *The left action $l_g(h)$, the right action $r_g(h)$ and the inverse $i(h)$ are smooth diffeomorphisms with respect to $h \in \mathbb{G}$*

The proof is given in appendix **??**. A classical example of a Lie group is the group of invertible n-dimensional Matrices $GL(\mathbb{R}, n)$ over the vector space $\mathcal{M} = \mathbb{R}^n$ [40]. The dimension of $GL(\mathbb{R}, n)$ is n^2 and the group operation $'\cdot'$ is the matrix multiplication.

In section 2.5.2 we argument that the set $\mathbb{G}$ acts in a two-fold manner on the functions $\phi(\boldsymbol{x}) \in \mathcal{C}^{\infty}(\Omega)$, namely by acting on the spacial coordinates $\boldsymbol{x} \in \Omega$ in eq. (2.134) and on the function values ϕ them selves in eq. (2.133). The spaces Ω and $\mathcal{C}^{\infty}(\Omega)$ are both vector spaces, that is the addition operation $'+'$ and multiplication with a factor $\lambda \in \mathbb{R}$ are defined in both spaces. It is thus natural to combine both Ω and $\mathcal{C}^{\infty}(\Omega)$ to one single vector space $\mathcal{M} = \Omega \times \mathcal{C}^{\infty}(\Omega)$. However since the functions $\phi(\boldsymbol{x})$ are unknown and we would also like to place constraints on their derivatives $\phi_{,K}$ (K is a multi-index, e.g. $K = [xxx, yy]$), we

combine Ω together with the *Jet space* $J^k\left(\mathcal{C}^\infty\left(\Omega\right)\right)$ [68], $\mathcal{M} = \Omega \times J^k\left(\mathcal{C}^\infty\left(\Omega\right)\right)$. $J^k\left(\mathcal{C}^\infty\left(\Omega\right)\right)$ is the set of smooth differentiable functions and their derivatives up to order k with compact support in Ω. The points $\boldsymbol{z} \in \mathcal{M}$ are vectors of the *independent* variables $\boldsymbol{x}$, the *dependent* variable $\phi\left(\boldsymbol{x}\right)$ and its derivatives $\phi_{,K}$

$$\boldsymbol{z} = \left(\boldsymbol{x}, \phi\left(\boldsymbol{x}\right), \phi_{,K}\left(\boldsymbol{x}\right)\right) \tag{2.149}$$

For this work we will focus only on first order derivatives, $k = 1$ so that the vectors $\boldsymbol{z}$ have the form

$$\boldsymbol{z} = \left(\boldsymbol{x}, \phi\left(\boldsymbol{x}\right), \nabla\phi\left(\boldsymbol{x}\right)\right) \tag{2.150}$$

The action of $\mathbb{G}$ on $\mathcal{M}$ is straightforward

$$\widetilde{\boldsymbol{z}} = \left(\widetilde{\boldsymbol{x}}, \widetilde{\phi}\left(\widetilde{\boldsymbol{x}}\right), \widetilde{\nabla}\widetilde{\phi}\left(\widetilde{\boldsymbol{x}}\right)\right) \tag{2.151}$$

$$\widetilde{\boldsymbol{x}} = g_{a_1\ldots a_r} \circ \boldsymbol{x} \tag{2.152}$$

$$\widetilde{\phi} = g_{a_1\ldots a_r} \circ \phi \tag{2.153}$$

$$\widetilde{\nabla} = J^{-T}\nabla, \quad J_{\mu\nu} = \frac{d\widetilde{x}_\mu}{dx_\nu} \tag{2.154}$$

Since the elements $g_{a_1\ldots a_r}$ are continuous in the parameters a_i we are free define to a smooth path γ in the parameter space $\mathcal{A}$

$$\gamma : t \rightarrow \left(a_1\left(t\right)\ldots a_r\left(t\right)\right) \tag{2.155}$$

$$g_{\gamma(0)} = e \tag{2.156}$$

The derivative of $g_{\gamma(t)}$ with respect to t at $t = 0$ is an element of the tangential space of $\mathbb{G}$ at the neutral element $e \in \mathbb{G}$, $T_e\mathbb{G}$

$$\left.\frac{d}{dt}g_{\gamma(t)}\right|_{t=0} = X_e \in T_e\mathbb{G} \tag{2.157}$$

The subscript on the vector X_e denotes that it belongs to $T_e\mathbb{G}$. The coordinates of X_e relative to the space $\mathcal{M}$ can be computed when we look at the derivative of the induced action of $g_{\gamma(t)}$ on the space of smooth functions with support on $\mathcal{M}$, $\mathcal{F}\left(\mathcal{M}\right)$. The action of X_e on $\mathcal{F}\left(\mathcal{M}\right)$ can be computed by evaluating $F \in \mathcal{F}\left(\mathcal{M}\right)$ on the tranformed vector $\widetilde{\boldsymbol{z}} = g_{\gamma(t)} \circ \boldsymbol{z}$ and the taking the derivative with respect to t at the neutral element e

$$X_e F\left(\boldsymbol{z}\right) = \left.\frac{d}{dt}F\left(\widetilde{\boldsymbol{z}}\right)\right|_{t=0} = \sum_{i=1}^r (\omega_i^\mu \partial_\mu F\left(\boldsymbol{z}\right) + \omega_i^\phi \frac{d}{d\phi}F\left(\boldsymbol{z}\right) + D\phi_i^\nu \frac{d}{d\partial_\nu\phi}F\left(\boldsymbol{z}\right))\alpha_i \tag{2.158}$$

where we have

$$
\omega_i^\mu\left(\boldsymbol{x}\right) = \left.\frac{d\widetilde{x}_\mu}{da_i}\right|_{t=0} \qquad \omega_i^\phi\left(\boldsymbol{x}, \phi\right) = \left.\frac{d\widetilde{\phi}}{da_i}\right|_{t=0} \qquad \alpha_i = \left.\frac{da_i}{dt}\right|_{t=0} \tag{2.159}
$$

$$
D\phi_i^\nu = \frac{d\omega_i^\phi}{dx_\nu} - \sum_\mu \frac{d\omega_i^\nu}{dx_\mu}\partial_\mu\phi \tag{2.160}
$$

The function $D\phi_i^\nu$ is called the prolonged action of $g_{\gamma(t)}$ on the gradient operator ∇ [68, 83] (see appendix B.1 for the derivation). Notice that while ω_i^μ and ω_i^ϕ are functions defined on $\mathcal{M}$, the coefficients α_i are independent of $\mathcal{M}$. They are the components of the vector X_e with respect to the r basis operators

$$
X_{e,i} = X_{e,i}^\Omega + \omega_i^\phi\frac{d}{d\phi} + D\phi_i^\nu\frac{d}{d\partial_\nu\phi}, \qquad X_{e,i}^\Omega = \omega_i^\mu\partial_\mu \tag{2.161}
$$

so that X_e has the operator form

$$
X_e = \sum_i \alpha_i X_{e,i} \tag{2.162}
$$

An important point about X_e is that it is an *operator valued* function over Ω since the coefficients ω_i^μ and ω_i^ϕ in eq. (2.159) are functions over Ω. We will refer to X_e as a vector at the unit element e, to $\boldsymbol{\omega}_i^\mu$ as a vector valued function (VVF) and to ω_i^Ω as a scalar valued function (SVF). The vector X_e only exists in the tangential space at $e \in \mathbb{G}$, $X_e \in T_e\mathbb{G}$. However it is possible to construct a vector Y_h at a location $h \in \mathbb{G}$ by relating it to X_e with a map $l_{h\star}$ called the *push-forward*

$$
Y_h F\left(\boldsymbol{z}\right) = \left(l_{h\star}X_e\right) F\left(\boldsymbol{z}\right) = \left.\frac{d}{dt}F\left(l_h\left(g_{\gamma(t)}\right)\circ\boldsymbol{z}\right)\right|_{t=0} \tag{2.163}
$$

The vector X_e operates on the function F in eq. (2.158) as a differential operator at the point $e \circ \boldsymbol{z} = \boldsymbol{z}$. The effect of $l_{h\star}$ is that it transports the vector X_e to the vector Y_h which operates on F at the point $l_h\left(e\right) \circ \boldsymbol{z} = h \circ \boldsymbol{z}$. As Y_h is a smooth function with respect to h which is defined everywhere in $\mathbb{G}$ it is called a *vector field over* $\mathbb{G}$. This means the coordinates of the vector field Y_h are the operators $h \in \mathbb{G}$ and *not* the points $\boldsymbol{z} \in \mathcal{M}$. The set of vector fields is the union of all the tangential spaces over $\mathbb{G}$

$$
T\mathbb{G} = \bigcup_{h\in\mathbb{G}} T_h\mathbb{G} \tag{2.164}
$$

Similar to X_e in eq. (2.162) the vector Y_h has a coordinate representation with

respect to the tangential space $T_h \mathbb{G}$

$$Y_h F(\boldsymbol{z}) = \sum_i \alpha'_i Y_{h,i} \tag{2.165}$$

$$Y_{h,i} = \omega'^\mu_i \partial_\mu + \omega'^\phi_i \frac{d}{d\phi} + D'\phi^\nu_i \frac{d}{d\partial_\nu \phi} \tag{2.166}$$

There exists a unique sub set $\mathcal{G} \subset T\mathbb{G}$ called the *Lie algebra*. It defined as the set of all vector fields $X_h \in T\mathbb{G}$ which are invariant under the left action l_g for any $g \in \mathbb{G}$

$$l_{g\star} X_h = X_{g\cdot h} = \sum_i \alpha_i X^i_{g\cdot h} \quad \forall g \in \mathbb{G}, X_h \in \mathcal{G} \tag{2.167}$$

From eq. (2.167) we see that a consequence of left invariance is that the coordinate vector $\boldsymbol{\alpha}$ is constant under the transformation l_g. This is what is referred to as the *parallel transport* of $\boldsymbol{\alpha}$ along the transformation l_g. The Lie algebra $\mathcal{G}$ has the property that it is closed under the antisymmetric commutator $[\cdot, \cdot]$

$$[Y_h, X_h] = M_h \in \mathcal{G} \quad \forall X_h, Y_h \in \mathcal{G} \tag{2.168}$$

Eq. (2.168) also implies that the commutator $[Y_h, X_h]$ is also left invariant [57]. If we consider the basis of the Lie algebra $\mathcal{G}$, the r vector fields $X^1_h, \cdots, X^r_h$ then the closed-ness of the commutator $[\cdot, \cdot]$ implied by eq. (2.168) translates to the fact that the commutator $[\cdot, \cdot]$ of two basis elements X^i_h and X^j_h is a linear combination of the basis $X^1_h, \cdots, X^r_h$ itself

$$\left[X^i_h, X^j_h\right] = \sum_{l=1}^r C^l_{i,j} X^l_h \tag{2.169}$$

The constants $C^l_{i,j}$ are called the *structure constants* of the Lie algebra $\mathcal{G}$. The commutator in eq. (2.168) has a geometric meaning. Suppose Y_e is the vector in the tangent space T_e of the one parameter group g^Y_t in the sense of eq. (2.157). It is easy to show (see appendix B.2) that the rate of change of the vector field $X_{g^Y_t}$ at the unit element e is equal to the commutator between X_e and Y_e

$$\frac{d}{dt} X_{g^Y_t} \bigg|_{t=0} = [Y_e, X_e] \tag{2.170}$$

Since $[Y_h, X_h]$ is left invariant, eq. (2.170) may be translated to any point $g \in \mathbb{G}$

$$h \circ \frac{d}{dt} X_{g^Y_t} \bigg|_{t=0} = \frac{d}{dt} X_{h \cdot g^Y_t} \bigg|_{t=0} = [Y_h, X_h] \tag{2.171}$$

Essentially eq. (2.171) tells us that once we can tell how the vector field X_g changes along any path g_t^Y which goes through unity, $g_t^Y|_{t=0} = e$, we can compute its rate of change along *any other* path in $\mathbb{G}$.

2.6.1. The Group $\mathbb{G} = \mathbb{T} \times SO(2)$

The group $\mathbb{G} = \mathbb{T} \times SO(2)$ is the group of translations and rotations in the plane $\mathbb{R}^2$. Its algebra is the algebra $\mathcal{G} = \mathfrak{t} \times \mathfrak{so}(2)$ which has the basis $\{X_e^{\Omega,x}, X_e^{\Omega,y}, X_e^{\Omega,\theta}\}$. The subset $\mathfrak{t} = \{X_e^{\Omega,x}, X_e^{\Omega,y}\}$ is the set generators of infinitesimal translations

$$X_e^{\Omega,x} = \partial_x, \quad X_e^{\Omega,y} = \partial_y \tag{2.172}$$

$\mathfrak{t}$ is a commutative basis since $[\partial_x, \partial_y] = 0$. The basis for $\mathfrak{so}(2)$ is the single operator $X_e^{\Omega,\theta}$ which is the generator of infinitesimal rotations. With respect to the Cartesian coordinate frame ∂_θ it has the following representation

$$X_e^{\Omega,\theta} = -y\partial_x + x\partial_y \tag{2.173}$$

From eq. (2.173) we can see that ∂_θ does not commute with $\mathfrak{t}$ and the commutators for the basis $\left\{X_e^{\Omega,x}, X_e^{\Omega,y}, X_e^{\Omega,\theta}\right\}$ are easily computed

$$\left[X_e^{\Omega,\theta}, X_e^{\Omega,x}\right] = -X_e^{\Omega,y} \quad \left[X_e^{\Omega,\theta}, X_e^{\Omega,y}\right] = X_e^{\Omega,x} \quad \left[X_e^{\Omega,x}, X_e^{\Omega,y}\right] = 0 \tag{2.174}$$

We note that the group $SO(2)$ generates the unit circle S^1 by rotating the point $\boldsymbol{x}_0 = (x, y)$

$$\boldsymbol{x}(\theta) = g_\theta \circ \boldsymbol{x}_0 = R_\theta \boldsymbol{x}_0, \quad R_\theta = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \tag{2.175}$$

The meaning of the first two commutators in eq. (2.174) is that the gradient operator ∇ is rotated by $90°$ counter clockwise under the action of $X_e^{\Omega,\theta}$

$$\frac{d}{d\theta}\nabla_{\boldsymbol{x}(\theta)}\bigg|_{\theta=0} = \left[X_e^{\Omega,\theta}, \nabla\right] = \mathbf{M}_\theta \cdot \nabla, \quad \mathbf{M}_\theta = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{2.176}$$

The matrix $\mathbf{M}_\theta$ is one of the Pauli matrices [59]. The Pauli matrices are a subset of the basis of the Lorentz algebra, the algebra defining the Lorentz group of special relativity which is an important symmetry for many quantum field theories for instance quantum electrodynamics [59, 88, 37].

2.7. Noether's First Theorem

Embedding Geometrical Constraints into Prior Energies

In section 2.5.1 argued that in order for a prior $E^{prior}(\nabla\phi)$ needs to be invariant to a large group of transformations $\mathbb{G}$ in order for it's minimizers

$$A = \left\{ \phi^\star \middle| \phi^\star = \operatorname{argmin}_\phi \left(-E^{prior}(\nabla\phi) \right) \right\} \tag{2.177}$$

to be non trivial, that is $\phi^\star \neq const$. Invariance of E^{prior} was linked to the requirement that the minimizer set A in eq. (2.177) be generated by the group $\mathbb{G}$

$$A = \{ \phi^\star \mid \phi^\star = g_\omega \circ \phi_0^\star \quad g_\omega \in \mathbb{G} \} \tag{2.178}$$

In eq. (2.138) we explained that a transformation $g_\omega \in \mathbb{G}$ may partition the set A into subsets A_Ω whose elements are related to each other through geometrical transformations $g_{\omega\Omega}$ on the coordinate frame Ω. We motivated the introduction of deformations to the gradient operator ∇ such that the level-sets S' in eq. (2.145) have the same functional form in the transformed coordinates $\boldsymbol{x}' = g_{\omega\Omega} \circ \boldsymbol{x}$ as in the original coordinates (see eq. (2.143)). With the help of the machinery introduced in section 2.6 we can express a level-set S_V of $\phi^\star \in \mathcal{A}$ in terms of a left invariant vector field V_h operating on $\phi^\star$ at the identity $e \in \mathbb{G}$

$$S_V = \left\{ \boldsymbol{x} \middle| V_e^\Omega \phi^\star(\boldsymbol{x}) = 0 \right\} \quad V_e^\Omega = \sum_i \alpha_i X_e^{\Omega,i} \quad X_e^{\Omega,i} \in \boldsymbol{X}_e^\Omega \tag{2.179}$$

for a particular sub algebra $\boldsymbol{X}_e^\Omega \subset \mathcal{G}$. Through eq. (2.179) the vector field V_h defines the particular geometry of S_V. For instance if the set of differential operators $\{X_e^{\Omega,i}\}$ are the Cartesian operators $\{\partial_x, \partial_y\}$ then the corresponding level-sets $S_{\boldsymbol{v}}^{lin}$ are the straight lines with tangential vector $\boldsymbol{v}$

$$S_{\boldsymbol{v}}^{lin} = \left\{ \boldsymbol{x} \middle| \boldsymbol{v}^T \cdot \nabla\phi^\star(\boldsymbol{x}), \boldsymbol{x} \in \Omega \right\} = 0 \tag{2.180}$$

The tangential vector $\boldsymbol{v}$ is a constant function in Ω so that the level-set $S_{\boldsymbol{v}}^{lin} \subset \Omega$ is a line oriented in the direction $\boldsymbol{v}$ which is invariant upon the transformation $\boldsymbol{x} \to \boldsymbol{x} + \tau\boldsymbol{v}$. Different priors $E^{prior}(\nabla\phi)$ may be constructed for which the minimizers $\phi^\star \in \mathcal{A}$ have linear level-sets $S_{\boldsymbol{v}}^{lin}$ of all orientations $\boldsymbol{v}$. For instance in section 2.9 we will introduce the total variation prior E_{TV}^{prior} and in section 3 the structure tensor based prior E_{ST}^{prior}. Both priors penalize level-sets with non-zero curvature so that their minimizers $\phi^\star$ only contain linear level-sets $S_{\boldsymbol{v}}^{lin}$. However the orientation vector $\boldsymbol{v}$ is not constrained.

A generalization to priors $E^{prior}(\boldsymbol{X}_e^\Omega \phi)$ whose minimizers $\phi^\star$ have level-sets S_V of the generic form in eq. (2.179) is possible if the sub algebra $\boldsymbol{X}_e^\Omega \subset \mathcal{G}$ is a commutative algebra. In [6, 34] it was shown that only for pair-wise commuting differential operators $X_e^{\Omega,1}, X_e^{\Omega,2} \in \boldsymbol{X}_e^\Omega$, $\left[X_e^{\Omega,1}, X_e^{\Omega,2}\right] = 0$ may there exist images $\phi^\star$ which have level-sets S_V of the form in eq. (2.179). Furthermore for the commuting operators $X_e^{\Omega,1}, X_e^{\Omega,2} \in \boldsymbol{X}_e^\Omega$ there exist pairs of functions $\epsilon_1(\boldsymbol{x})$ and $\epsilon_2(\boldsymbol{x})$ which obey the conditions

$$X_e^{\Omega,1}\epsilon_1\left(\boldsymbol{x}\right) = 1, \quad X_e^{\Omega,1}\epsilon_2\left(\boldsymbol{x}\right) = 0$$
$$X_e^{\Omega,2}\epsilon_1\left(\boldsymbol{x}\right) = 0, \quad X_e^{\Omega,2}\epsilon_2\left(\boldsymbol{x}\right) = 1 \tag{2.181}$$

The function $\boldsymbol{\epsilon}(\boldsymbol{x}) = (\epsilon_1(\boldsymbol{x}), \epsilon_2(\boldsymbol{x}))$ represents a coordinate transformation from the Cartesian frame Ω to the curvi-linear frame Ω^ϵ with the coordinates (ϵ_1, ϵ_2)

$$\boldsymbol{\epsilon}(\boldsymbol{x}) : \Omega \to \Omega^\epsilon \tag{2.182}$$

Within the coordinate frame Ω^ϵ the commutating operators $X_e^{\Omega,1}, X_e^{\Omega,2} \in \boldsymbol{X}_e^\Omega$ take on the form $X_e^{\Omega,1/2} = \partial_{\epsilon_{1/2}}$ and the level-sets S in eq. (2.179) are represented by the inverse transformation $\boldsymbol{x}(\boldsymbol{\epsilon})$

$$S_{\boldsymbol{\alpha}}^\epsilon = \left\{ \boldsymbol{x}(\boldsymbol{\epsilon}) \,\middle|\, \sum_i \alpha_i \partial_{\epsilon_i} \phi(\epsilon_1, \epsilon_2) = 0 \right\} \tag{2.183}$$

By comparison of $S_{\boldsymbol{\alpha}}^\epsilon$ in eq. (2.183) with $S_{\boldsymbol{v}}^{lin}$ in eq. (2.180) we can see that $S_{\boldsymbol{\alpha}}^\epsilon$ is a line in the domain Ω^ϵ with orientation $\boldsymbol{\alpha}$.

The availability of the transform $\boldsymbol{\epsilon}(\boldsymbol{x})$ and its inverse is guaranteed if the differential operators $X_e^{\Omega,1}, X_e^{\Omega,2} \in \mathcal{G}$ commute. Conversely if a transform $\widetilde{\boldsymbol{\epsilon}}(\boldsymbol{x})$ is available for which the Cauchy-Riemann equations

$$\partial_x\widetilde{\epsilon}_1 = -\partial_y\widetilde{\epsilon}_2, \quad \partial_x\widetilde{\epsilon}_2 = \partial_y\widetilde{\epsilon}_1 \tag{2.184}$$

then differential operators $\widetilde{X}_e^{\Omega,1}, \widetilde{X}_e^{\Omega,2} \in \boldsymbol{X}_e^\Omega$ can be computed and for which the level-sets $S_{\boldsymbol{\alpha}}^{\widetilde{\epsilon}}$ are linear in the domain $\Omega^{\widetilde{\epsilon}}$. Hence the specification of a transform $\boldsymbol{\epsilon}(\boldsymbol{x})$ which obeys eq. (2.184) is sufficient to construct a prior $E^{prior}(\boldsymbol{X}_e^\Omega \phi)$ for which the minimizers $\phi^\star$ have level-sets $S_{\boldsymbol{\alpha}}^\epsilon$ which are linear in the domain Ω^ϵ and we will call the domain Ω^ϵ the linear domain of $E^{prior}(\boldsymbol{X}_e^\Omega \phi)$.

Now we have shown that a prior $E^{prior}(\boldsymbol{X}_e^\Omega \phi)$ may be constructed using the operators of a commutative sub algebra $\boldsymbol{X}_e^\Omega \subset \mathcal{G}$. However in eq. (2.178) we explained that the minimizer set $\mathcal{A}$ can be described as being spun up by the entire group G. Thus the minimizer set $\mathcal{A}$ itself is by definition invariant under

the action of $\mathbb{G}$

$$h \circ \mathcal{A} = \mathcal{A} \quad \forall h \in \mathbb{G} \tag{2.185}$$

Eq. (2.185) is equivalent to the restriction that the prior energy E^{prior} itself must be invariant under the action of $\mathbb{G}$

$$E^{prior}\left(\mathcal{X}^{\Omega}_{g_{\gamma(t)}}\phi\right) = const, \quad \text{with respect to } t \tag{2.186}$$

for any one-parameter sub group $g_{\gamma(t)} \in \mathbb{G}$. Eq. (2.186) poses a constraint on *how* the prior E^{prior} is constructed. For instance the energy $E^{prior}(\nabla) = \int (\partial_x \phi)^2 d^2 x$ is not invariant under the rotation group $SO(2)$, and does not obey eq. (2.186) if $\mathbb{G}$ contains $SO(2)$ as a subgroup. Hence its minimizers $\phi^\star$ have level-sets oriented only in x-direction rendering it useless for anisotropic regularization.

The following section is devoted to the consequences of eq. (2.186) for models

$$E\left(\phi, \boldsymbol{X}^{\Omega}_{e}\phi\right) = E^{data}(\phi) + E^{prior}\left(\boldsymbol{X}^{\Omega}_{e}\phi\right) \tag{2.187}$$

For which the prior E^{prior} obeys the invariance relation in eq. (2.186) for a general Lie group $\mathbb{G}$.

2.7.1. Noethers Theorems

In her original paper [81, 82] Emmy Noether handles the question: Given a model of a physical system, encoded in an action

$$E = \int_{\Omega} \left(\mathcal{E}\left(\boldsymbol{x}, \{\phi_\rho\}, \{\nabla_K \phi_\rho\}\right)\right) d^n x \tag{2.188}$$

which depends on ρ fields $\phi_1 \dots \phi_\rho$ and their derivatives to order K, and knowledge of a set of smooth transformations $\mathbb{G}$ under which the action E is invariant

$$E' = g_\gamma \circ E = E \quad \forall g_\gamma \in \mathbb{G} \tag{2.189}$$

what are the special properties hidden in the model that invoke the symmetry?

To answer this question she deals with two cases:

- Finite dimensional Lie groups $\mathbb{G}$, which we introduced in section 2.6. For now it is sufficient to think of $\mathbb{G}$ as the set of smooth functions g_γ defined on an r dimensional space, $\gamma = (\alpha_1, \dots, \alpha_r)$.

- Infinite dimensional Lie groups $\mathbb{G}_\infty$, which are generalizations of the finite dimensional groups in the sense that the r parameters $\alpha_1, \ldots, \alpha_r$ are functions over the Cartesian coordinate frame Ω. We will not handle this case.

In the case of the finite dimensional group Emmy Noether took g_γ to be the smooth *infinitesimal* transformation, encoding both variations of the fields and of the coordinates

$$\phi'_\rho(\boldsymbol{x}) = \phi_\rho(\boldsymbol{x}) + \sum_{m=1}^r \alpha_m \omega_m^{\phi_\rho}(\boldsymbol{x}) \quad \boldsymbol{x}' = \boldsymbol{x} + \sum_{m=1}^r \alpha_m \boldsymbol{\omega}_m^\Omega(\boldsymbol{x}) \tag{2.190}$$

She proved that if the action E is invariant under g_γ in eq. (2.189), then there exists r vectors $\boldsymbol{W}_m$ such the integral relationship

$$E - E' = \int_\Omega \sum_{m=1}^r \alpha_m \left[\sum_\rho \bar{\omega}_m^{\phi_\rho} [\mathcal{E}]_\rho + \operatorname{div}(\boldsymbol{W}_m) \right] = 0 \tag{2.191}$$

$$\bar{\omega}_m^{\phi_\rho} = \left(\omega_m^{\phi_\rho} - \omega_m^{\mu\Omega} \partial_\mu \phi_\rho \right), \quad [\mathcal{E}]_\rho = \frac{\delta \mathcal{E}}{\delta \phi_\rho} - \frac{d}{dx^\mu} \left(\frac{\delta \mathcal{E}}{\delta \phi_{\rho,\mu}} \right) \tag{2.192}$$

where $[\mathcal{E}]_\rho$ are the Euler-Lagrange differentials of the fields ϕ_ρ and the divergences $\operatorname{div}(\boldsymbol{W}_m)$ appear by carefully collecting all terms which occur as a result of the integral product rule

$$\int f \cdot \partial_\mu g \, d^n x = \int \partial_\mu (f \cdot g) \, d^n x - \int \partial_\mu f \cdot g \, d^n x \tag{2.193}$$

when computing the symbolic form of $[\mathcal{E}]_\rho$. The main result is the argument that since the α_m, $\omega_m^{\phi_\rho}$ and the ω_m^μ are assumed to linearly independent, the r equations

$$\sum_\rho \bar{\omega}_m^{\phi_\rho} [\mathcal{E}]_\rho + \operatorname{div}(\boldsymbol{W}_m) = 0 \quad m = 1, \ldots, r \tag{2.194}$$

relate r of the ρ Euler-Lagrange equations $[\mathcal{E}]_\rho$ so that the physical system only has $\rho - r$ independent Euler-Lagrange equations $[\mathcal{E}]_\rho$ and thus only $\rho - r$ independent fields ϕ_ρ. In the case $\rho \le r$ the system of equations in eq. (2.194) is overdetermined, eq. (2.191) can only hold if *all* the divergences and *all* the Euler-Lagrange equations vanish

$$[\mathcal{E}]_\rho \left(\phi_1^\star, \ldots, \phi_\rho^\star \right) = 0, \quad \operatorname{div}(\boldsymbol{W}_m) \left(\phi_1^\star, \ldots, \phi_\rho^\star \right) = 0, \quad \rho \le r \tag{2.195}$$

Eq. (2.195) implies that only at the minima of the fields, $\phi_\rho^\star$ the r vectors $\boldsymbol{W}_m$ are

conserved.

Kepler's Two Body Problem

Kepler's two body problem is the problem of calculating the problem of estimating the trajectory of a body of mass m_e (the earth) which is moving within the vicinity of another body with mass m_s (the sun). According to Newton there exists a gravitational force between the masses coming from the energy $V(r)$ of the gravitational field surrounding the mass m_s at the origin in $\mathbb{R}^3$

$$V(\mathbf{r}_e(t)) = -\frac{m_e \cdot m_s}{r} \quad r = \|\mathbf{r_e} - \mathbf{r_s}\| \tag{2.196}$$

The kinetic energy of the mass m_e is $\frac{1}{2}m_e\dot{r}^2$ so that the Lagrangian of the path $\mathbf{r}_e(t)$ is

$$L(\mathbf{r}_e(t)) = \frac{1}{2}m_e\dot{r}_e^2 + \frac{1}{2}m_e\dot{r}_s^2 - V(\mathbf{r}_e(t)) \tag{2.197}$$

The Euler-Lagrange equations are easily computed

$$\ddot{r} - \frac{m_s + m_e}{r^2} = 0 \tag{2.198}$$

The parameter t is the time parameter of the two body system and eq. (2.198) describes the motion of the total mass around the center of mass of the two body system. The Kepler Lagrangian in eq. (2.197) exhibits a symmetry under four different one parameter Lie group actions, namely the action of time shift and rotations around the three spacial axis (the group $SO(3) \times \mathbb{R}$)

$$t' = t + \delta t \tag{2.199}$$
$$\mathbf{r}' = \mathbf{r} + \partial_{\theta_i}\mathbf{r}'\delta\theta_i \quad i = x, y \text{ or } z \tag{2.200}$$

where θ_i are rotation around the x-,y- or z-axis. From Noether's theorem there exist four corresponding conserved quantities:

$$W_t = \mathcal{H} = \frac{1}{2}m_e\dot{r}^2 + V(\mathbf{r}_e(t)) \quad \textit{time shift} \tag{2.201}$$

$$W_x = z\dot{y} - y\dot{z} \quad \textit{Rotation around x-axis} \tag{2.202}$$
$$W_y = z\dot{x} - x\dot{z} \quad \textit{Rotation around y-axis} \tag{2.203}$$
$$W_z = x\dot{y} - y\dot{x} \quad \textit{Rotation around z-axis} \tag{2.204}$$

The conserved quantity $\mathcal{H}$ in eq. (2.201) is the *Hamiltonian Energy* of the two body system. It constant time and thus manifests that the total energy of the two

body system does not dissipate away since there are no external forces interacting with the two masses m_e and m_s, that is the two body system is a *closed system*. The vector $\boldsymbol{W} = (W_x, W_y, W_z)$ (Eqs. eq. (2.202) to eq. (2.204)) is the total *angular momentum* the masses m_e and m_s have as they rotate around each other. The solutions to the Euler-Lagrange equations in eq. (2.198) are elliptic curves in the surface $S_{\boldsymbol{W}}$ orthogonal to $\boldsymbol{W}$. The constancy of $\boldsymbol{W}$ with respect to the special orthogonal group $SO\,(3)$ comes the fact that $S_{\boldsymbol{W}}$ is actually a flat Euclidean plane embedded in a 3-dimensional Euclidean space.

2.7.2. Noether's First Theorem: A Modern Version

In this section we explicitly derive Noether's first theorem for models with one field ϕ and its first derivatives $\boldsymbol{X}_e^\Omega \phi$ using the Lie algebra introduced in section 2.6. We consider the total energy

$$
E = \int_\Omega \mathcal{E}\left(\phi, \boldsymbol{X}_e^\Omega \phi\right) \mathcal{N}(\boldsymbol{x}) d^2 x = \int_\Omega \mathcal{E}^{data}\left(\phi\right) \mathcal{N}(\boldsymbol{x}) d^2 x + \int_\Omega \mathcal{E}^{prior}\left(\boldsymbol{X}_e^\Omega \phi\right) \mathcal{N}(\boldsymbol{x}) d^2 x
$$

$$(2.205)$$

The explanation of the constant $\mathcal{N}(\boldsymbol{x})$ will shortly follow. We apply a one parameter group $g_{\gamma(t)}$ to E^{prior} and according to eq. (2.157) we can compute the vector V_e in the tangent space of $g_{\gamma(t)}$ at $t = 0$.

$$
\left.\frac{d}{dt} g_{\gamma_t} \circ E\right|_{t=0} = \int_\Omega \left(\sum_{i=1}^q P_i \left[V_e^\Omega, X_e^{i,\Omega}\right] \phi + v^\phi \left[\mathcal{E}\right]\right) \mathcal{N}(\boldsymbol{x}) d^2 x \tag{2.206}
$$

$$
P_i = \frac{\delta \mathcal{E}}{\delta X_e^{i,\Omega} \phi}, \quad [\mathcal{E}] = \frac{\delta \mathcal{E}}{\delta \phi} - \sum_i \frac{d}{dx^\mu}\left(w_i^\mu P_i\right) \tag{2.207}
$$

For the derivation of eq. (2.206) see eq. (B.47) in appendix B.3. Eq. (2.206) will play an important role in section 4 which is why we will discuss it in detail. The differentials $[\mathcal{E}]$ are called the Euler-Lagrange differentials [82, 68] and the vector $\boldsymbol{P}$ is called the canonical momentum [60]. Eq. (2.206) is the most general form of variation. It contains two components, namely one component proportional to intensity variations of the field ϕ, v^ϕ and one component proportional to variations of the coordinate frame Ω encoded in $V_e^\Omega = v^\mu \partial_\mu$. Just as in eq. (2.194) the variations v^ϕ and v^μ are taken to be independent from each other. This means that the Lie group $\mathbb{G}$ splits into two *independent* sub groups $\mathbb{G}^\phi$ and $\mathbb{G}^\Omega$ and we can write $\mathbb{G}$ as the composition $\mathbb{G} = \mathbb{G}^\phi \times \mathbb{G}^\Omega$. The integral volume $d^2 x$ also

transforms under the action of g_t^V

$$\frac{d}{dt}\left(g_t^V \circ d^2x\right)|_{t=0} = \frac{dv^\mu}{dx^\mu}d^2x \tag{2.208}$$

However we are only interested in the Euler-Lagrange differentials $[\mathcal{E}]$ and the canonical momentum vector $\boldsymbol{P}$ since only they depend on the particular Lagrangian $\mathcal{E}$. Thus the normalization factor $\mathcal{N}(\boldsymbol{x})$ in the integral in eq. (2.205) is chosen such that the volume element $\mathcal{N}(\boldsymbol{x})d^2x$ is invariant under the transformation $g_{\gamma(t)}$

$$\frac{d}{dt}\left(g_t^V \circ \mathcal{N}(\boldsymbol{x})d^2x\right)\Big|_{t=0} = \left(\mathcal{N}(\boldsymbol{x})\frac{dv^\mu}{dx^\mu} + \frac{d\mathcal{N}(\boldsymbol{x})}{dt}\Big|_{t=0}\right)d^2x = 0 \tag{2.209}$$

In the following we will drop the normalization $\mathcal{N}(\boldsymbol{x})$ and assume d^2x to invariant under any transformation g_t^V. Since V_e is an element of the Lie algebra $\mathcal{G}$ we can expand it in terms of the r basis elements X_e^i

$$V_e = \sum_{i=1}^{r} \alpha_i X_e^i \tag{2.210}$$

Under the expansion in eq. (2.210) the eq. (2.206) becomes

$$\frac{d}{dt}g_{\gamma t} \circ E\Big|_{t=0} = \sum_{m=1}^{r} \alpha_m \int_\Omega \left(\widetilde{\boldsymbol{B}}_m \phi + \omega_m^\phi[\mathcal{E}]\right)d^2x \tag{2.211}$$

where the operator $\widetilde{\boldsymbol{B}}_m$ is the scalar product

$$\widetilde{\boldsymbol{B}}_m = \sum_{i=1}^{q} P_i \left[X_e^{m,\Omega}, X_e^{i,\Omega}\right] \tag{2.212}$$

The operator $\widetilde{\boldsymbol{B}}_m$ in eq. (2.212) will play an important role in section 4 where we propose an extension to the principle of least action. In appendix B.3 we show how eq. (2.211) can be transformed in to the original version in eq. (2.191)

$$\frac{d}{dt}g_{\gamma t} \circ E\Big|_{t=0} = \int_\Omega \sum_m \alpha_m \left(\text{div}\left(\boldsymbol{W}_m\right) + \widetilde{\omega}_m^\phi[\mathcal{E}]\right)d^2x \tag{2.213}$$

$$\widetilde{\omega}_m^\phi = \omega_m^\phi - X_e^{m,\Omega}(\phi) \tag{2.214}$$

with the vector valued functions (VVF) $\boldsymbol{W}_m$

$$W_m^\mu = \omega_m^\mu \mathcal{E} + \sum_i \omega_i^\mu \left(\omega_m^\phi - X_e^m(\phi)\right) P_i \tag{2.215}$$

Thereby we prove the following relation

$$\text{div}\left(\boldsymbol{W}_m\right) - X^{\Omega}_{e,m}\left(\phi\right)\left[\mathcal{E}\right] = \widetilde{\boldsymbol{B}}_m\phi \tag{2.216}$$

If the energy E is assumed to be invariant with respect to any one parameter group $g_t^V \subset \mathbb{G}$

$$\left.\frac{d}{dt}g_{\gamma_t} \circ E\right|_{t=0} = 0 \tag{2.217}$$

then by the argumentation in section 2.7.1 the divergences of the vectors $\boldsymbol{W}_m$ in eq. (2.215) *and* the Euler-Lagrange differentials must vanish

$$\left[\mathcal{E}\right]\left(\phi^\star\right) = 0, \quad \text{div}\boldsymbol{W}_m = 0 \quad \forall 1 \leq m \leq r \tag{2.218}$$

and by eq. (2.216) the equations in eq. (2.218) imply

$$\widetilde{\boldsymbol{B}}_m\phi^\star = 0 \tag{2.219}$$

2.7.3. Pure Spacial Symmetries

Since we assumed the invariance of the energy E under the group $\mathbb{G} = \mathbb{G}^\phi \times \mathbb{G}^\Omega$ in eq. (2.217) as well as the independence of the sub groups $\mathbb{G}^\phi$ and $\mathbb{G}^\Omega$ it follows that E must also be invariant with respect to pure spacial one dimensional sub groups $g_t^{V^\Omega} \subset \mathbb{G}^\Omega$

$$\left.\frac{d}{dt}\left(g_t^{V^\Omega} \circ E\right)\right|_{t=0} = 0 \tag{2.220}$$

The Lie algebra element V_e^Ω corresponding to $g_t^{V^\Omega}$ does not contain any variations to the field ϕ thus we can obtain an expression for eq. (2.220) simply by setting $v^\phi = 0$ and $\omega_i^\phi = 0$ in eqs. (2.206) and (2.211)

$$\left.\frac{d}{dt}\left(g_t^{V^\Omega} \circ E\right)\right|_{t=0} = \int_\Omega \sum_{m=1}^r \alpha_m \widetilde{\boldsymbol{B}}_m\phi \, d^2x = 0 \tag{2.221}$$

It follows that if eq. (2.221) holds for any one parameter sub group $g_t^{V^\Omega} \subset \mathbb{G}$ (any coefficient vector $\boldsymbol{\alpha}$) then

$$\widetilde{\boldsymbol{B}}_m\phi = 0 \tag{2.222}$$

must hold for *any* field configuration ϕ. Eq. (2.222) is specifically a constraint on the prior energy E^{prior} since the data term E^{data} does not contain any derivatives $X_e^{\Omega,i}\phi$ and thus the canonical momentum $\boldsymbol{P}$ (eq. (2.207)) only depends on the prior energy density $\mathcal{E}^{prior}$. There are three cases to consider such that eq. (2.222) can hold:

- Case a: The Lie algebra $\mathcal{G}$ is commutative, $[X_{e,i}^{\Omega}, X_{e,m}^{\Omega}] = 0$ for all $1 \leq i, m \leq r$

- Case b: $P_i = 0$ for all $1 \leq i \leq r$

- Case c: If we have $[X_{e,i}^{\Omega}, X_{e,m}^{\Omega}] \neq 0$ for some i and m the canonical momentum $\boldsymbol{P}$ if non-vanishing must be orthogonal to the vector $\mathbf{M}_m$, which is a vector for fixed m defined as $(\mathbf{M}_m)_i = \left[X_{e,i}^{\Omega}, X_{e,m}^{\Omega}\right](\phi)$ over Ω

We call cases a and b trivial symmetries and case c a non-trivial symmetry. In chapter 3 we will introduce a prior E^{prior} which is conditionally invariant to the group $\mathbb{G}^{\Omega} = \mathbb{T} \times SO(2)$ which is the group of local translations and rotations. Its algebra $\mathcal{G} = \mathfrak{t} \times \mathfrak{so}(2)$ is 3-dimensional and although it is not a commutative algebra we will show that eq. (2.222) still holds for any field ϕ.

2.8. Current usage of Noethers Theorems

Up to the present day, Emmy Noethers theorems serve as guiding principles for the construction of models of our physical reality which must admit observed symmetries. For instance the gauge invariance of Maxwells equations results via Noethers theorems in the conservation of the electrical charge density ρ. This conservation has been experimentally observed, and thus modern theories of particle physics such as the standard model [88] directly implement electro-dynamical gauge invariance at the core.

However the case may arrive where a GRF model $E_C(\phi_i, \nabla\phi_i)$ with multiple GRFs ϕ_i for a physical system C is constructed *prior* to observing that the system C has a symmetry group $\mathbb{G}_C$. In this case the energy E_C is invariant under the action of $\mathbb{G}_C$ but the invariance is not apparent in the original formulation of the energy E_C. For such cases the method of Moving Frames [68, 32, 33] was introduced which builds on Noethers theorems. At the center of the Moving Frame theory is the Re-parameterization theorem, which states that the invariant energy $E_C(\phi_i, \nabla\phi_i)$ may be re-parameterized in terms of r $\mathbb{G}_C$-invariant GRFs κ_i

$$\widetilde{E}_C(\kappa_1, \ldots, \kappa_r) = E_C(\phi_i(\kappa_1, \ldots, \kappa_r), \nabla\phi_i(\kappa_1, \ldots, \kappa_r)) \tag{2.223}$$

The mathematical expression of the re-parameterized energy $\widetilde{E}_C(\kappa_1, \ldots, \kappa_r)$ is then readily invariant since the new GRFs κ_i are themselves invariant under $\mathbb{G}_C$. This method is also called the method of invariantization.

2.9. Total Variation

In this section we will introduce a widely used method for anisotropic regularization of the GRF ϕ called Total Variation (TV) [100, 16, 115, 29, 11]. In the context of shock-filtering [100, 84, 99] it was shown that the functional

$$E_{L_1}(\phi) = \int |\nabla\phi| dx \tag{2.224}$$

has the appealing property that it does not penalize large discontinuities. However its functional derivative with respect to ϕ is ill conditioned in the case $\nabla\phi \approx 0$. To alleviate the case, [100] chose the approximative prior

$$E_{L_1 approx}(\phi) = \int \sqrt{|\nabla\phi|^2 + \epsilon} dx \tag{2.225}$$

which is well behaved for $\epsilon > 0$. They were able to achieve good results with relatively sharp preserved discontinuities with data ϕ^0 having low SNRs. Nevertheless in the limit $\epsilon \to 0$ the Euler-Lagrange equations become more and more computationally instable. A theoretically more well conditioned form of TV is needed which we will outline, following [11, 103]. To do this we need to explore the function-space the minimizers of eq. (2.224) might belong to. Smooth functions ϕ_{smooth} are functions for which $\nabla\phi$ exists everywhere, thus they may be minimizers of eq. (2.224). But functions $\phi_{discont}$ containing discontinuities do *not* have finite L_1 norm of their gradients, $E_{L_1}(\phi_{discont}) = \infty$ since the gradient $\nabla\phi_{discont}$ does not exist at the discontinuities. A generalization of eq. (2.224) is possible if one assumes $\nabla\phi$ to be a distribution, more precisely a Radon measure [11, 103] in the space $\mathcal{M}(\Omega)$. If there exists a Radon measure $\boldsymbol{\mu} \in \mathcal{M}(\Omega)$, such that for every $\boldsymbol{p} \in \mathcal{C}_0(\Omega)$ with compact domain, the following equality holds

$$\int_\Omega \phi \cdot \mathrm{Div}\boldsymbol{p} dx = -\int \boldsymbol{p}^T d\boldsymbol{\mu} < \infty \tag{2.226}$$

then $\boldsymbol{\mu}$ is called the weak derivative of ϕ and we can identify $\nabla\phi = \boldsymbol{\mu}$. It is then possible to define the function-space of bounded variation

$$BV = \{\phi \in L_1(\Omega) \,|\, \nabla\phi \in \mathcal{M}(\Omega)\} \tag{2.227}$$

Now it is possible to define a norm on BV. By virtue of the Hölder inequality [42] there exists a scalar C for which we can determine the upper bound of eq. (2.226)

$$\int_\Omega \phi \cdot \text{Div} \boldsymbol{p} dx \leq C \|\phi\|_\infty \tag{2.228}$$

The scalar C is the norm of the Radon measure $\nabla \phi$ and is called the total variation of ϕ

$$TV(\phi) = \sup \left\{ \int_\Omega \phi \cdot \text{Div} \boldsymbol{p} d^2 x \mid \quad \|\boldsymbol{p}\|_\infty \leq 1 \right\} \tag{2.229}$$

As was discussed in [11] the functions ϕ are geometrically piecewise smooth, meaning there exists a partitioning $\{\Omega_k\}$ of Ω such that $(\nabla \phi)_{\Omega_k}$ are L_1 integrable. If dl_{mk} is a line segment in the intersection $\Omega_m \cap \Omega_k$ then $TV(\phi)$ can be written in the form

$$TV(\phi) = \sum_k \|\nabla \phi_{\Omega_k}\|_{L_1} + \sum_{k<m} L_{lm} \tag{2.230}$$

$$L_{lm} = \int_{\Omega_l \cap \Omega_m} |\phi_l - \phi_m| \, dl_{lm} \tag{2.231}$$

where ϕ_l is the value of ϕ on the portion of $\partial \Omega_l$ which is interfacing with Ω_m and vice versa for ϕ_m. The first term in eq. (2.230) penalizes the smooth parts of ϕ (the gradients $(\nabla \phi)_{\Omega_k}$). Similar to eq. (2.131) $\|\nabla \phi_{\Omega_k}\|_{L_1}$ is invariant to shifts of ϕ_{Ω_k} by constants c_k'

$$\|\nabla (\phi_{\Omega_k} + c_k')\|_{L_1} = \|\nabla \phi_{\Omega_k}\|_{L_1} \tag{2.232}$$

Thus due to eq. (2.232) we can view the smooth functions ϕ_{Ω_k} as being centered around constants c_k

$$\phi_{\Omega_k}(\boldsymbol{x}) = c_k + \widetilde{\phi}_{\Omega_k}(\boldsymbol{x}) \tag{2.233}$$

where the c_k are determined by the data ϕ_d. For instance if we combined the TV functional with a data term $E_{data} = \sum_k \int_{\Omega_k} (\phi_d - \phi_{\Omega_k})^2 dx$

$$E(\phi, \nabla \phi) = \sum_k \int_{\Omega_k} (\phi_d - \phi_{\Omega_k})^2 dx + \lambda TV(\phi) \tag{2.234}$$

then in [75, 113] it was shown that the c_k can be computed to be the *mean* of the data ϕ_d within the area Ω_k, $c_k = \int_{\Omega_k} \phi_d d^2 x$ given that the deviations $\widetilde{\phi}_{\Omega_k}$ are penalized by the first term in eq. (2.230).

The second term in eq. (2.230) penalizes the length of the section $\Omega_m \cap \Omega_k$ while maintaining the values $\phi_{k,m}$ and thus the *jump* $|\phi_k - \phi_m|$. It essentially penalizes the curvature of the line interfacing with both Ω_k and Ω_m. We will make this point clear in the following section. For now we remark that if we set $\phi_{\Omega_k} = c_k$ in the data term in eq. (2.234) then we obtain

$$\widetilde{E}\left(\phi, \nabla\phi\right) = \sum_k \int_{\Omega_k} (\phi_d - c_k)^2 dx + \lambda TV\left(\phi\right) \tag{2.235}$$

which is of course only an approximation to eq. (2.234). The data term in eq. (2.235) is a measure for the variance of ϕ_d in Ω_k. The two terms in eq. (2.230) together with the data term $\sum_k \int_{\Omega_k} (\phi_d - c_k)^2 dx$ in eq. (2.235) balance the size of the partitions Ω_k since the boundaries of *small* partitions Ω_k have *high* curvature and thus high TV values, but *low* variances in the data term. On the other side *large* partitions Ω_k have boundaries of *low* curvature and thus low TV values, but *high* variances. The parameter λ in eqs. (2.234) and (2.235) marks the trade-off between the TV term and the data term in eqs. (2.234) and (2.235) and thus it determines the size of the partitions Ω_k.

2.9.1. The Mean Curvature of Total Variation

In eq. (2.230) we had argued that the TV measure can be split into a smooth part $\|\nabla\phi_{\Omega_k}\|_{L_1}$ measuring the deviation of the smooth functions ϕ_{Ω_k} from the constants c_k. We had claimed that the second term in eq. (2.230), the boundary term L_{lm} measures the curvature of the boundary between Ω_l and Ω_l. The line integral in L_{lm} in eq. (2.231) can be rewritten essentially as a measure for the length of the level-set S_{lm} interfacing Ω_l and Ω_m

$$L_{lm} = |\phi_k - \phi_m| \, \|S\|_{lm} \, , \quad \|S\|_{lm} = \int_0^T \left\| \frac{d}{dt}\left(\boldsymbol{x}(t)\right) \right\| dt \tag{2.236}$$

The path $\boldsymbol{x}(t)$ can be considered as being generated by a one parameter Lie group $g_t^{V^\Omega}$ acting on the point $\boldsymbol{x}_0$ which is on the interfacing boundary between Ω_l and Ω_m

$$\boldsymbol{x}(t) = g_t^{V^\Omega} \circ \boldsymbol{x}_0, \quad \boldsymbol{x}_0 \in \Omega_l \cap \Omega_m \tag{2.237}$$

so that the length $\|S\|_{lm}$ is controlled by the Lie algebra element $V_e^\Omega = v(\boldsymbol{x})^\mu \partial_\mu$

$$\|S\|_{lm} = s\left(T\right), \quad s\left(t\right) = \int_0^t \left\| \boldsymbol{v}\left(\boldsymbol{x}(t')\right) \right\| dt' \tag{2.238}$$

The function $s\left(t\right)$ in eq. (2.238) is called the *arc length* of the curve $\boldsymbol{x}\left(t\right)$. By virtue of the definition of the arc length $s(t)$ in eq. (2.238) we can express derivatives with respect to s by

$$\frac{d}{ds} = \frac{1}{\|\boldsymbol{v}\|}\frac{d}{dt} \tag{2.239}$$

The *curvature* of $\boldsymbol{x}_{lm}(t)$, $\kappa(\boldsymbol{x}_{lm}(t))$ is obtained by re-parameterizing $\boldsymbol{x}(t)$ in terms of its arc length s, $\boldsymbol{x}(t) \rightarrow \boldsymbol{x}(t(s))$ and taking the second derivative of $\boldsymbol{x}(s)$ using eq. (2.239)

$$\kappa\left(\boldsymbol{x}_0\right) = \left\|\frac{d^2}{ds^2}\boldsymbol{x}\left(s\right)\right\|_{s=0} \tag{2.240}$$

$$= \frac{1}{\|\boldsymbol{v}\|^3}\left(v_x \cdot \frac{dv_y}{dt} - v_y \cdot \frac{dv_x}{dt}\right)\bigg|_{t=0} \tag{2.241}$$

In [11] it is shown that the expression for the curvature $\kappa(\boldsymbol{x}_0)$ at the point $\boldsymbol{x}_0$ in eq. (2.241) is equivalent to the *mean curvature* [28, 15, 55]

$$\kappa = \text{Div}\left(\frac{\nabla\,\phi}{|\nabla\phi|}\right) \tag{2.242}$$

which is the functional derivative of the TV norm in eq. (2.224) with respect to ϕ

$$\kappa = -\partial TV\left(\phi\right) \tag{2.243}$$

For a thorough derivation of the mean curvature κ in terms of weak derivatives in BV spaces see [103]. The Euler-Lagrange equations of any energy function $E = \int \mathcal{E}d^2x$ including the TV functional in eq. (2.224) as a prior

$$[\mathcal{E}]\left(\boldsymbol{x}\right) = \frac{\delta\mathcal{E}}{\delta\phi}\left(\boldsymbol{x}\right) + \lambda\kappa\left(\boldsymbol{x}\right) = 0 \tag{2.244}$$

pose a bound on the value of the curvature $\kappa(\boldsymbol{x})$. Thus the TV functional penalizes the curvature κ of the S_{lm} interfacing Ω_l and Ω_m. As κ is an invariant of the Lie group $SE\left(2\right)$, the group of rotations and translations, TV is also an invariant of that group.

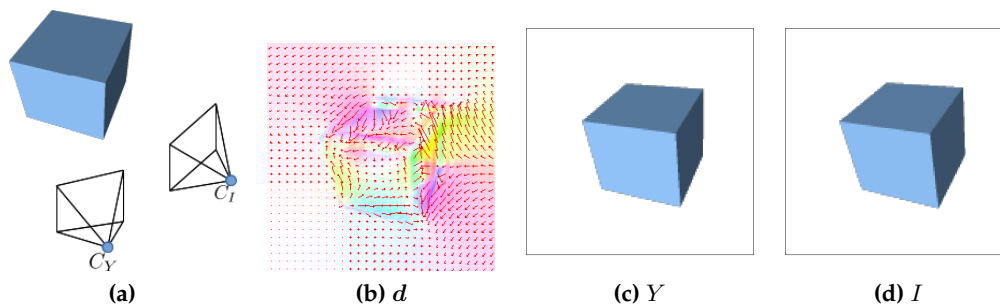| **(a)** | **(b)** d | **(c)** Y | **(d)** I |

Figure 2.4.: Figure 2.4a: Two cameras C_Y and C_I are shown recording a scene from different positions. The scene could could be a rigid scene or a dynamic scene with moving objects. Figure 2.4c shows the image Y captured from the camera C_Y and figure 2.4d the image I from the camera C_I. Figure 2.4b shows the optical flow d. The vectors in figure 2.4b indicate which pixels x' in I and x in Y are mapped to each other.

2.10. Convex Optimization

2.11. Optical Flow

In section 2.1 we had introduced the notion of an inverse problem, namely that given some data Y and a model C we would like to find the GRF ϕ which is mapped to the data Y by the model C, see eq. (2.8). A prime example of an inverse problem in computer vision is optical flow [52, 66, 107, 9, 12, 58, 115]. Optical Flow labels the task of densely measuring the motion between two or more frames captured by a camera, or the dense registration of two or more cameras on a pixel-by-pixel basis. Optical flow is a crucial step in many areas of computer vision. For instance optical flow estimation is a part of video compression [45, 44] used to detect areas of the video in which the rate brightness change is small. For example during the recording of a rigid scene optical flow can be used to determine when the camera motion stalls. During such periods the frames of the video can be stored in a memory efficient manner. In recent years structure from stereography and structure from motion (3D from a single camera) have gained popularity as a means to capture 3D models for film productions and also due to the availability of low cost 3D printing [66, 90, 92, 109, 105, 104, 56, 102]. In both the stereography and the structure from motion pipelines optical flow is used for the triangulation of the dense point cloud, prior to generation of the final 3D mesh. In the case of a dual-modal setup both cameras may be of different types. For instance in medical imaging multi-modal dense image registration is used to fuse image information from CT and MR modalities of the human brain [5] and of the human spine [119].

In optical flow modeling the task at hand is to estimate the disparity between two images Y and I recorded by two cameras C_Y and C_I (see figure 2.4). Each image is a map between the coordinate space $\Omega \subset \mathbb{R}^2$ and the real numbers $\mathbb{R}$. Thus $Y(\boldsymbol{x})$ is the intensity recorded by the camera C_Y at the pixel location $\boldsymbol{x} \in \Omega$ while $I(\boldsymbol{x}')$ is the intensity recorded by C_I at the location $\boldsymbol{x}' \in \Omega$. In figure 2.4a we have depicted a multi-modal setup in which the two cameras C_Y and C_I are recording images (figures 2.4c and 2.4d) from different angles. In this context the optical flow field is the unknown variable $\boldsymbol{d}$ which maps the location $\boldsymbol{x}'$ in the image I to the location $\boldsymbol{x}$ in the image Y

$$\boldsymbol{x}' = \boldsymbol{x} + \boldsymbol{d}(\boldsymbol{x}) \tag{2.245}$$

The optical field $\boldsymbol{d}$ is shown in figure 2.4b as a set of vectors at every pixel $\boldsymbol{x}' \in \Omega$, whose magnitude and orientation reflect the motion of the pixel $\boldsymbol{x}'$.

The standard methodology [52, 66, 111, 18] for the estimation of the optical flow $\boldsymbol{d}$ is to model $\boldsymbol{d}$ as a GRF with a given data term $E_{Y,I}^{data}(\boldsymbol{d})$. Without further information of the mapping between Y and I from another source (e.g. sparse feature mapping with SIFT features [65]), the data term $E_{Y,I}^{data}(\boldsymbol{d})$ cannot depend directly on $\boldsymbol{d}$ but can only be defined as a *similarity measure* between the image $Y(\boldsymbol{x})$ and the *warped* image $I_{\boldsymbol{d}}(\boldsymbol{x}) = I(\boldsymbol{x} + \boldsymbol{d}(\boldsymbol{x}))$ [111, 18]

$$E_{Y,I}^{data}(\boldsymbol{d}) = F(Y, I_{\boldsymbol{d}}), \quad I_{\boldsymbol{d}}(\boldsymbol{x}) = I(\boldsymbol{x} + \boldsymbol{d}(\boldsymbol{x})) \tag{2.246}$$

In general the mapping of $Y(\boldsymbol{x})$ and $I_{\boldsymbol{d}}(\boldsymbol{x})$ via $\boldsymbol{d}(\boldsymbol{x})$ is ill-determined: According to [58] the regions $\mathcal{A} \subset \Omega$ of an image ϕ can be given an *intrinsic dimension* [8, 122] iD which depends on their content

- $iD = 0$ if the image patch $\phi_{\mathcal{A}}$ is homogeneous

- $iD = 1$ if the image patch $\phi_{\mathcal{A}}$ contains an edge

- $iD = 2$ otherwise (e.g corners and/or textures)

If we consider two image patches $Y_{\mathcal{A}}$ and $I_{\mathcal{A}}$ with equal intrinsic dimension iD then the number of components of the optical flow $\boldsymbol{d}$ between $Y_{\mathcal{A}}$ and $I_{\mathcal{A}}$ which can be *uniquely* determined by $E_{Y,I}^{data}(\boldsymbol{d})$ is identical to iD. For instance if $Y_{\mathcal{A}}$ and $I_{\mathcal{A}}$ both display a corner of an object ($iD = 2$), the optical flow in $\mathcal{A}$, $\boldsymbol{d}_{\mathcal{A}}$ can be uniquely determined by the similarity measure $E_{Y,I}^{data}(\boldsymbol{d})$. However for edges ($iD = 1$) only one component of the optical flow $\boldsymbol{d}_{\mathcal{A}}$ can be determined and for homogeneous patches $\boldsymbol{d}_{\mathcal{A}}$ is completely undefined. Thus globally, that is over all $\boldsymbol{x} \in \Omega$ the similarity measure $E_{Y,I}^{data}(\boldsymbol{d})$ cannot determine $\boldsymbol{d}(\boldsymbol{x})$ uniquely. For this reason optical flow models deploy a prior energy $E^{prior}(\nabla \boldsymbol{d})$ which smooths $\boldsymbol{d}(\boldsymbol{x})$ such that information of the components of $\boldsymbol{d}$ in regions with $iD = 2$ is

carried on to neighboring regions with $iD \neq 2$ such that $d(x)$ is well defined over all Ω. The total energy

$$E_{Y,I}(d) = E_{Y,I}^{data}(d) + E^{prior}(\nabla d) \tag{2.247}$$

is then a trade-off between the similarity of Y and I_d and the smoothness of the optical flow $d(x)$.

2.11.1. Uni-Modal Optical Flow

Among the earliest methods for optical flow estimation are the methods described in the seminal papers of Horn and Schunck [52] and Lukas and Kanade [66]. In [52] the following model for computing the flow between two frames of a video was proposed

$$E_{Y,I}(d) = E_{Y,I}^{data}(d) + E^{prior}(d) \tag{2.248}$$

$$E_{Y,I}^{data}(d) = \int_{\Omega} \Big(Y(x) - I_d(x) \Big)^2 dx, \quad I_d(x) = I(x + d(x)) \tag{2.249}$$

$$E^{prior}(d) = \lambda \int_{\Omega} \sum_i \|\nabla d_i\|^2 dx \tag{2.250}$$

In eq. (2.249) the frame I is warped back to the frame Y by the field $d(x)$. The prior energy $E^{prior}(d)$ in eq. (2.250) imposes an isotropic smoothness constraint on the flow field d. As we discussed in section 2.5.1 the main limitation of the L_2 prior in eq. (2.248) is that it does not preserve edges in the flow field $d(x)$. To overcome this limitation [85] and [121] used the TV prior in eq. (2.224) as a smoothing term for each of the components of d

$$E_{Y,I}(d) = E_{Y,I}^{data}(d) + \lambda \int_{\Omega} \sum_i \|\nabla d_i\| dx \tag{2.251}$$

According to section 2.9 the level-sets of each component $d_i(x)$ are smoothed while the discontinuities are preserved.

2.11.2. Multi-Modal Optical Flow

The next issue with the model in eq. (2.248) is that the likelihood $E_{Y,I}^{data}(d)$ in eq. (2.249) makes the assumption that the cameras C_Y and C_I are sensitive to the same physical light spectrum. For instance in figure 2.4 the image Y recorded by the camera C_Y in figure 2.4c has the same intensity spectrum as the image I

recorded by the camera C_I (figure 2.4d) and we say that Y and I are *equal by distribution*

$$Y \stackrel{d}{\approx} I \qquad (2.252)$$

Thus it is possible to find an optical flow field $d^\star$ such that for each pixel $x \in \Omega$ the warped image I_d approximates the image Y, $Y(x) \approx I_d(x)$. However there exists multi-modal setups where the cameras are *not* sensitive to the same spectra. In section 3.7.5 we will show an instance of a multi-modal setup consisting of a visual spectrum camera C_{vsc} and a thermographic camera C_{tc} recording the same scene. Due to the different spectra in which the cameras are responsive, the recorded images I_{vsc} and y_{tc} fail to admit the correspondence $I_{vsc} \stackrel{d}{\approx} y_{tc}$.

We will now discuss three statistical similarity measures for arbitrary images Y and I which avoid the assumption of brightness constancy. For this we will take the two images Y and I to be random variables with the marginal distributions $p(Y)$ and $p(I)$. Then the mean and the variance are defined as

$$\mathbb{E}(X) = \int X \cdot p(X) \qquad (2.253)$$

$$\mathbb{V}ar(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) \qquad (2.254)$$

The three similarity measures all avoid the brightness constancy assumption implied by the data term in eq. (2.249) by only taking into account the *statistical* features of the images Y and I such as their joint entropy and joint covariance.

Mutual Information

Mutual Information (MI) [67, 95, 111, 51] is a popular similarity measure used mainly in medical imaging where images from different modalities including Magnetic Resonance Imaging (MRI) [27], Computed Tomography (CT) [53] and Positron Emission Tomography (PET) [112] are registered against each other. For images Y and I from two different modalities capturing the same scene, MI is defined with the joint distribution $p(Y, I)$ by

$$MI(Y, I) = \int p\left(\hat{Y}, \hat{I}\right) \ln \frac{p\left(\hat{Y}, \hat{I}\right)}{p\left(\hat{Y}\right) \cdot p\left(\hat{I}\right)} d\hat{Y} d\hat{I} \qquad (2.255)$$

MI measures how strongly the images Y and I *statistically* depend on each other. In the case that Y and I are statistically independent, $p(Y, I) = p(Y) \cdot p(I)$, then by eq. (2.255) MI is zero. On the other side, MI is maximal when I completely

determinates Y or vice versa. In the context of optical flow MI is used to measure the similarity between Y and I_d

$$E_{Y,I}^{data}\left(\boldsymbol{d}\right) = -MI\left(Y, I_{\boldsymbol{d}}\right) \tag{2.256}$$

However, as [93] puts it, MI does not explain the kind of dependency between images Y and I, its maxima are statistically but not geometrically meaningful, since it disregards any spatial information, which is essential for optical flow. Thus optical flow likelihoods based on MI usually tend to have many local minima rendering MI too unconstrained for optical flow.

Correlation Ratio

To alleviate the problems with MI, [93] arguments that a better similarity measure would be one that measures the *functional* relation between the images Y and I. The key ingredient for their proposal is that the pixel values $I\left(\boldsymbol{x}\right)$ and $Y\left(\boldsymbol{x}\right)$ are assumed to be the realizations of random variables, which by abuse of notation we denote by $\hat{I}$ and $\hat{Y}$. Then the normalized joint histogram of the images I and Y can be interpreted as the joint probability distribution $p\left(\hat{Y}, \hat{I}\right)$, and the conditional distribution

$$p\left(\hat{Y}\,\middle|\,\hat{I} = I\right) = \frac{p\left(\hat{Y}, \hat{I} = I\right)}{p\left(\hat{I} = I\right)} \tag{2.257}$$

encodes the spatial functional relationship between Y and I. They introduced the Correlation Ratio (CR) [94, 95, 111, 18]

$$\eta_{CR}\left(I|Y\right) = \frac{\mathbb{V}ar\left(\phi^{\star}\left(Y\right)\right)}{\mathbb{V}ar\left(I\right)}, \quad E_{Y,I}^{data}\left(\boldsymbol{d}\right) = -\eta_{CR}\left(I_{\boldsymbol{d}}|Y\right) \tag{2.258}$$

The optimal function $\phi^{\star}$ was shown to be the expectation value of $\hat{I}$, conditioned on a realization of $\hat{Y}$

$$\phi^{\star}\left(Y\right) = \mathbb{E}\left(\hat{I}\,\middle|\,\hat{Y} = Y\right) = \int Ip\left(I|Y\right)dI \tag{2.259}$$

The function $\phi(\hat{Y})$ maps any realization of $\hat{Y}$ to an expectation value of $\hat{I}$. As $\hat{Y}$ is a random variable, $\phi(\hat{Y})$ is also at random. Its variance measures how well $\hat{I}$ is *functionally explained* by a realization of $\hat{Y}$. The measure in eq. (2.258) is bounded between 0 and 1, 0 indicating that Y and I are independent, 1 indicating a functional relationship $I = \phi^{\star}\left(Y\right)$. The function $\phi^{\star}$, although not necessarily

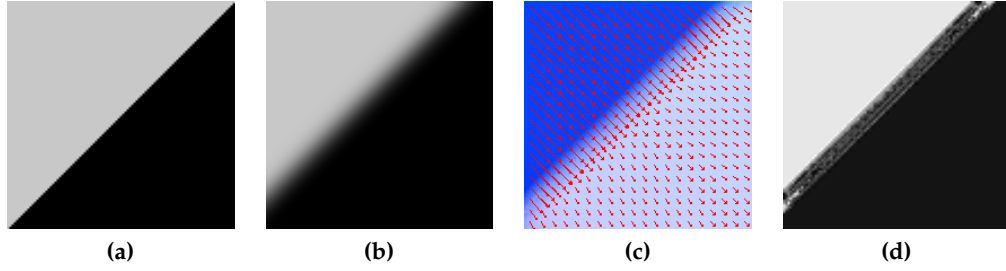**(a)**   **(b)**   **(c)**   **(d)**

Figure 2.5.: Figure 2.5a shows a synthetic high resolution image I^{syn}. In figure 2.5b we show a low resolution image Y^{syn}. Y^{syn} is computed by down sampling I^{syn} by a factor $\sigma^{sc} = 5$, creating y^{syn} followed by cubic interpolation and translated by 10 pixels relative to I^{syn}. An optical flow model which incorporates knowledge of the scale difference between Y^{syn} and I^{syn} should produce a flow $\widetilde{d}$, such that the warped image $I^{syn}_{\widetilde{d}}$ matches the image Y^{syn} *up to scale* σ^{sc}, thereby preserving the features of I^{syn}. Figure 2.5c shows the flow d computed with the model in eq. (2.248). Since the model in eq. (2.248) does not incorporate knowledge of the scale difference σ^{sc}, the features of the warped image I^{syn}_{d} (figure 2.5d) are heavily distorted

continuous, is measurable in the L_2-sense. Thus CR is a much stronger constraint then MI and has fewer, but more meaningful minima [93].

Cross Correlation

Cross Correlation [31, 78, 13, 111] is the strongest constrained similarity measure. It is basically an additional constraint to CR, namely that the functional relationship in eq. 2.258 must be linear. Then η reduces to

$$\eta_{CC}(I|Y) = \frac{Cov(Y, I)}{\mathbb{V}ar(I) \cdot \mathbb{V}ar(Y)} \quad Y = f \cdot I + \beta \tag{2.260}$$

As we will see in section 3.4 a measure similar to eq. (2.260) will be computed based on the assumption that both Y and I are Gaussian. The Gaussian assumption is valid when both cameras Y and I produce Gaussian noise and the joint histogram is predominantly linear. Linearity in the joint histogram occurs when the recorded scene contains materials with uniform luminosity in the frequency bands of the cameras Y and I.

All three similarity measures have in common that the images Y and I must have the same spatial resolution in order to compute the measure. For instance in eq. (2.256) [67] computed the joint probability $p(Y, I)$ as a normalized histogram $h(\hat{Y}, \hat{I})$ created from the samples $\hat{Y} = Y(\boldsymbol{x})$ and $\hat{I} = I(\boldsymbol{x})$ drawn from all locations $\boldsymbol{x} \in \Omega$.

The problem that we want to attack is the determination of the optical flow between a *low resolution image* y obtained from a low resolution camera C_y and a high resolution image I from a camera C_I. From now on lower case letters stand for low resolution and higher case letters for high resolution images. In an ad-hoc fashion we could first filter and down-sample the image I with a convolution filter G to obtain an image i with the same spatial resolution of y

$$i\left(\boldsymbol{x}\right) = \left(G \star I\right)\left(\boldsymbol{x}\right), \quad \boldsymbol{x} \in \Omega \tag{2.261}$$

and evaluate the similarity measures on the image pair y and i. The negative impact is that we could only estimate an optical flow $\boldsymbol{d}$ with the same low resolution as the image y. Conversely we could up-sample the image y with some interpolation scheme to produce a high resolution image Y and evaluate the similarity measures on the pair Y and I. This situation is shown in figure 2.5. I^{syn} in figure 2.5a shows a sharp linear boundary. We down sampled I^{syn} by a factor $\sigma^{sc} = 5$ to create y^{syn}. y^{syn} was translated by 10 pixels and the high resolution image Y^{syn} (figure 2.5b) was computed by a cubic interpolation of y^{syn} again by a factor $\sigma^{sc} = 5$. The images Y^{syn} and I^{syn} thus differ in optical scale, and the scale difference is the parameter σ^{sc}. We used the model of [52]

$$E\left(\boldsymbol{d}^{syn}\right) = \frac{1}{2} \int_{\Omega} \left(Y^{syn}\left(\boldsymbol{x}\right) - I^{syn}_{\boldsymbol{d}^{syn}}\left(\boldsymbol{x}\right)\right)^2 dx + \frac{\lambda}{2} \sum_i \int_{\Omega} \left\|\nabla d^{syn}_i\left(\boldsymbol{x}\right)\right\|^2 dx \tag{2.262}$$

(see eq. (2.248)) to compute the optical flow $\boldsymbol{d}^{syn}$ mapping I^{syn} to Y^{syn} (see figure 2.5c). Figure 2.5d shows the image $I^{syn}_{\boldsymbol{d}^{syn}}\left(\boldsymbol{x}\right) = I^{syn}\left(\boldsymbol{x} + \boldsymbol{d}^{syn}\left(\boldsymbol{x}\right)\right)$. We can see that the optical flow $\boldsymbol{d}$ corrupts the sharp boundary of I^{syn} in order to match it to the varying gray levels of the blurred boundary in Y^{syn} (figure 2.5b). The problem is that the model in eq. (2.262) can account for the *difference in size* of the images y^{syn} and I^{syn} but it does not take the *difference in optical scale* σ^{sc} into account. Thus we need a model that can account for the optical scale σ^{sc}.

2.12. Image Fusion

In this section we will introduce the image fusion method of Hardie et. al. [47]. In that paper the authors solved the problem of refining the low optical resolution of an image y_{tc} obtained by a thermographic camera (TC) using the image I_{vsc} obtained by a visual spectrum camera (VSC). The result of their method is a thermographic image Y_{tc} with improved optical resolution (see figure 2.6a for a schematical overview). They used this method for the subject of remote sensing [14] where the TC and the VSC are built in a co-aligned fashion within the body of a satellite. In section 3.3 the goal is to extend this method to the case where
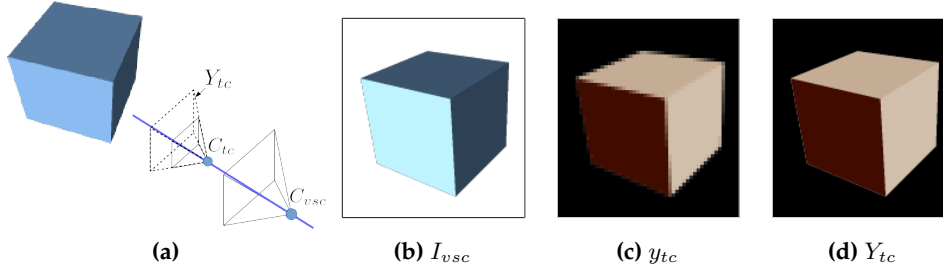
(a)   **(b)** I_{vsc} **(c)** y_{tc} **(d)** Y_{tc}

Figure 2.6.: Figure 2.6a shows a schematic setup of the camera configuration considered by Hardie et. al. [47]. The blue line indicates the orthogonal direction to both the TC and the VSC image planes. The camera centers (indicated by the blue circles) are aligned along the dashed line. Figure 2.6b shows the image I_{vsc} captured by the VSC and figure 2.6c the image y_{tc} captured by the TC. The image y_{tc} has a smaller optical resolution then the image I_{vsc}. The method in [47] takes the data y_{tc} and I_{vsc} to produce a higher resolution thermographic image Y_{tc}, shown in figure 2.6d

the TC and the VSC are not co-aligned. In this case the low resolution image y_{tc} and thus the high resolution Y_{tc} have a natural separation from the VSC image I_{vsc} similar to figure 2.4. We will show that it is possible to jointly estimate the image Y_{tc} and the optical flow $\boldsymbol{d}(\boldsymbol{x})$ between Y_{tc} and I_{vsc}.

The method of [47] goes as follows: In figure 2.7a a model of the CCD of the low resolution TC is shown overlaid with a higher resolution grid representing the VSC. The gray region in figure 2.7a symbolizes one pixel of the TC and it can be seen that each pixel of the TC covers a group of pixels of the VSC. Since the TC pixel has a finite surface, we need to specify how this pixel absorbs photons landing at different points in its area in order to relate the covered pixels of the VSC to it. The response of each individual pixel in the TC is called the point spread function (PSF), $W_{\sigma^{sc}}(x,y)$, the vector (x,y) being the location on the surface of the TC pixel with respect to the VSC coordinate frame. Figure 2.7b is the result of a theoretical model of a FLIR TC [41]. The model, obtained by Hardie et al. [46], combines absorption properties of the CCD pixel with physical properties of the camera lens. We can see that each TC pixel has a non uniform response to incoming photons and the PSF $W_{\sigma^{sc}}$ is approximately Gaussian with standard deviation σ^{sc}. Using this information we can model a super-resolved version Y_{tc} of the TC image y_{tc} with the help of the PSF $W_{\sigma^{sc}}$, by stating that y_{tc} is the result of the convolution of Y_{tc} with $W_{\sigma^{sc}}$

$$y_{tc} = \langle Y_{tc} \rangle_{\sigma^{sc}} + n \quad n \sim \mathcal{N}(0|C_n), \quad \langle Y_{tc} \rangle_{\sigma^{sc}} = W_{\sigma^{sc}} Y_{tc} \tag{2.263}$$

The problem of estimating Y_{tc} is that there is an infinite amount of high resolution TC images $Y_{tc}^\star$ which relate to y_{tc} via eq. (2.263) since the high spacial
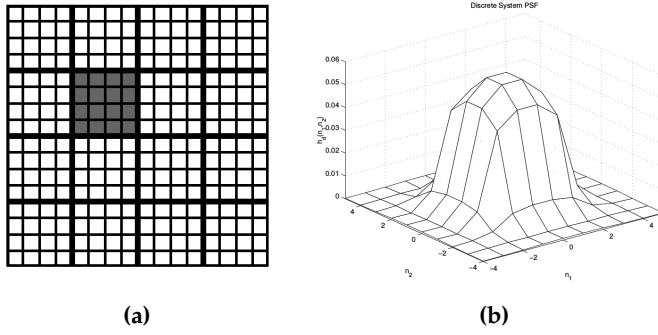
(a) **(b)**

Figure 2.7.: Figure 2.7a The thick grid depicts the CCD of the low resolution thermographic camera. The finer grid a virtual super-resolved version of the pixels in the TC. Figure 2.7b shows the point spread function $W_{\sigma^{sc}}(x, y)$ of the gray pixel in figure 2.7a, taken from Hardie et al. [46]. It shows that each pixel in the TC image has a non uniform response over its surface to incoming photons.

frequency components of Y_{tc} are filtered out by $W_{\sigma^{sc}}$. In [47] Hardie suggested the use of a high resolution imager I_{vsc} whose camera center is co-aligned (see figure 2.6a) with the TC image y_{tc} and correlated with Y_{tc}. The rationale behind their approach is to combine the desired features such as sharp edges and corners of I_{vsc} with the intensity spectrum of y_{tc} into the super-resolved image Y_{tc}, while avoiding limitations such as the noise model of y_{tc}.

The key ingredient in the model of [47] is that the intensities of Y_{tc} and I_{vsc} are assumed to be samples drawn from the joint Gaussian $p(Y_{tc}, I_{vsc})$. As I_{vsc} is already fixed as input data we can derive a conditional distribution for Y_{tc} via the Bayesian rule [101]

$$p(Y_{tc}|I_{vsc}) = \frac{p(Y_{tc}, I_{vsc})}{p(I_{vsc})} \sim \mathcal{N}\left(\mu_{Y_{tc}|I_{vsc}}|C_{Y_{tc}|I_{vsc}}\right) \tag{2.264}$$

$$C_{Y_{tc}|I_{vsc}} = C_{Y_{tc},Y_{tc}} - C_{Y_{tc},I_{vsc}}^2 \cdot C_{I_{vsc},I_{vsc}}^{-1} \tag{2.265}$$

$$\mu_{Y_{tc}|I_{vsc}}(\mathbf{x}) = \mu_{Y_{tc}} + C_{Y_{tc},I_{vsc}} \cdot C_{I_{vsc},I_{vsc}}^{-1}(I_{vsc}(\boldsymbol{x}) - \mu_{I_{vsc}}) \tag{2.266}$$

where the variances and means are computed globally

$$C_{u,v} = \int_{\Omega}(u(\boldsymbol{x}) - \mu_u) \cdot (v(\boldsymbol{x}) - \mu_v) d^2x \tag{2.267}$$

$$\mu_u = \int_{\Omega} u(\boldsymbol{x}) d^2x$$

We see that the mean of Y_{tc} conditioned on I_{vsc}, $\mu_{Y_{tc}|I_{vsc}}$ (eq. (2.266)) is linear in the values of I_{vsc}, thus in this model the intensities of Y_{tc} are assumed to be

globally linearly related to the intensities of I_{vsc}. We combine eq. (2.264) with the Gaussian likelihood in eq. (2.263) to the posterior

$$p\left(Y_{tc}|y_{tc}, I_{vsc}\right) \sim p\left(y_{tc}|Y_{tc}\right) \cdot p\left(Y_{tc}|I_{vsc}\right) = \exp(-E_{y_{tc},I_{vsc}}\left(Y_{tc}\right)) \qquad (2.268)$$

with the associated energy

$$E_{y_{tc},I_{vsc}}\left(Y_{tc}\right) = \frac{1}{2}\int_{\Omega}\left(y_{tc}\left(\boldsymbol{x}\right) - W_{\sigma^{sc}}Y_{tc}\left(\boldsymbol{x}\right)\right)^{2} \cdot C_{n}^{-1}d^{2}x$$
$$+ \frac{1}{2}\int_{\Omega}\left(Y_{tc}\left(\boldsymbol{x}\right) - \mu_{Y_{tc}|I_{vsc}}\left(\boldsymbol{x}\right)\right)^{2} \cdot C_{Y_{tc}|I_{vsc}}^{-1}d^{2}x \qquad (2.269)$$

The minimization of eq. (2.269) and thus maximization of (2.268) with respect to Y_{tc} gives the analytical solution [47] in lexicographic reordering (appendix D)

$$Y_{tc}^{\star} = \mu_{Y_{tc}|I_{vsc}} + C_{Y_{tc}|I_{vsc}} \cdot W_{\sigma^{sc}}^{T}\mathbf{H}^{-1} \cdot \left(y_{tc} - \langle\mu_{Y_{tc}|I_{vsc}}\rangle_{\sigma^{sc}}\right) \qquad (2.270)$$
$$\langle\mu_{Y_{tc}|I_{vsc}}\rangle_{\sigma^{sc}} = W_{\sigma^{sc}}\mu_{Y_{tc}|I_{vsc}}$$
$$\mathbf{H} = \left(W_{\sigma^{sc}} \cdot C_{Y_{tc}|I_{vsc}} \cdot W_{\sigma^{sc}}^{T} + C_{n}\right)$$

Eq. (2.270) is computationally expensive due to the dense operator $W_{\sigma^{sc}}$ and the matrix-inverse operation. However if $W_{\sigma^{sc}}$ is approximately Gaussian then the diagonal entries of the matrix $\mathbf{H}$ are larger than the off-diagonal entries. In [123] a computationally tractable approximation was introduced

$$\hat{Y}_{tc} = \mu_{Y_{tc}|I_{vsc}} + C_{\langle Y_{tc}\rangle_{\sigma^{sc}}|\langle I_{vsc}\rangle_{\sigma^{sc}}} \cdot \widetilde{\mathbf{H}}^{-1}\left(y_{tc} - \langle\mu_{Y_{tc}|I_{vsc}}\rangle_{\sigma^{sc}}\right) \qquad (2.271)$$
$$\langle I_{vsc}\rangle_{\sigma^{sc}} = W_{\sigma^{sc}}I_{vsc}, \quad \langle Y_{tc}\rangle_{\sigma^{sc}} = W_{\sigma^{sc}}Y_{tc} \approx y_{tc} \qquad (2.272)$$
$$\widetilde{\mathbf{H}} = \left(C_{\langle Y_{tc}\rangle_{\sigma^{sc}}|\langle I_{vsc}\rangle_{\sigma^{sc}}} + C_{n}\right) \qquad (2.273)$$
$$\mu_{\langle Y_{tc}\rangle_{\sigma^{sc}}|I_{vsc}}\left(\mathbf{x}\right) = \mu_{\langle Y_{tc}\rangle_{\sigma^{sc}}} + C_{\langle Y_{tc}\rangle_{\sigma^{sc}},I_{vsc}} \cdot C_{I_{vsc},I_{vsc}}^{-1}\left(I_{vsc}\left(\boldsymbol{x}\right) - \mu_{I_{vsc}}\right)$$
$$(2.274)$$

where the matrix $\widetilde{\mathbf{H}}$ is a diagonal matrix and thus easily invertible. The approximated conditional mean $\mu_{\langle Y_{tc}\rangle_{\sigma^{sc}}|I_{vsc}}$ is a transformation of the intensities of the VSC image I_{vsc} to the spectrum of the TC image y_{tc}.

The key issue is that eq. (2.271) requires both modalities, I_{vsc} and y_{tc}, to be co-aligned. Since we are dealing with an optical flow problem y_{tc} and thus Y_{tc} is shifted by a disparity $\boldsymbol{d}\left(\boldsymbol{x}\right)$ from I_{vsc}. This disparity has to be taken in to account by our model in chapter 3.3. The second issue is that the assumption that Y_{tc} and I_{vsc} are *globally* joint Gaussian is not supported by our data. However by computing $C_{Y_{tc}|I_{vsc}}$ in *local* sub-domains of the space Ω we can show that Y_{tc} and I_{vsc} are *locally* joint Gaussian. This will also be shown in chapter 3.3.

3. Linearized Priors

3.1. The Linear Structure Tensor

We shall now proceed to introduce a prior based on the considerations made in chapter 2.5. We will concentrate on the translation group $\mathbb{T}$ for which the Lie algebra $\mathfrak{t}$ is characterized by the set of vectors $\boldsymbol{v}$ which are constant within a sub domain $A \subset \Omega$. The basis operators $X_e^i \in \mathfrak{t}$ are the Cartesian differential operators $\{\partial_x, \partial_y\}$, and the spatial component V_e^Ω of a vector $V_e \in \mathfrak{t}$ has the representation

$$V_e^\Omega = v_x(\boldsymbol{x})\,\partial_x + v_y(\boldsymbol{x})\,\partial_y \in \mathfrak{t} \quad \boldsymbol{v}(\boldsymbol{x})|_A = const \tag{3.1}$$

The translation group $\mathfrak{t}$ is a commutative algebra so the basis $\{\partial_x, \partial_y\}$ is commutative and any vector V_e^Ω commutes with the $\{\partial_x, \partial_y\}$

$$\left[V_e^\Omega, \partial_{x,y}\right] = 0 \tag{3.2}$$

Thus V_e^Ω is translation invariant. Consider an image $\phi(\boldsymbol{x})$. The level-sets S_X corresponding to the vector V_e^Ω are are defined by

$$S_X = \left\{\boldsymbol{x} \,\middle|\, \boldsymbol{v}^T \cdot \nabla\phi(\boldsymbol{x}) = 0\right\} \tag{3.3}$$

We would like to characterize the dominant strength and the orientation of $\nabla\phi$ within the sub domain $A \subset \Omega$. In [7] it was suggested that the tangential vector $\boldsymbol{v}$ of the level sets S_X can be computed by minimizing the energy

$$J(\boldsymbol{v}) = \frac{1}{2}\int_A w\left(\|\boldsymbol{x}\|\right)\boldsymbol{v}^T \cdot \left(\nabla\phi(\boldsymbol{x})\,\nabla^T\phi(\boldsymbol{x})\right)\boldsymbol{v} = \frac{1}{2}\boldsymbol{v}^T S\boldsymbol{v} \tag{3.4}$$

$$S = \int_A w\left(\|\boldsymbol{x}\|\right)\left(\nabla\phi(\boldsymbol{x})\,\nabla^T\phi(\boldsymbol{x})\right)d^2x = \langle\nabla\phi\nabla^T\phi\rangle \tag{3.5}$$

The weight function $w(\|x\|)$ is normalized and weights the contributions of the gradient $\nabla\phi(\boldsymbol{x})$ at various points $\boldsymbol{x} \in A$. Typically a Gaussian is deployed for the weight function, $w(\|x\|) = G_\kappa(\boldsymbol{x} - \boldsymbol{x}_0)$ where $\boldsymbol{x}_0$ is the center pixel of A. This way the gradient $\nabla\phi(\boldsymbol{x}_0)$ is the dominant contribution to the integral in eq. (3.5).

The matrix S is called the structure tensor. Since S is a symmetric matrix there exists an orthogonal decomposition

$$S = V^T DV \quad D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \quad V = (\boldsymbol{V}_1, \boldsymbol{V}_2) \tag{3.6}$$

The eigenvalues give of the squared strength of the gradient in the basis defined by the columns of V. They characterize the structure in A in the following way

- $\lambda_1 > \lambda_2$: Strong linear level set with tangential vector $\boldsymbol{v} = \boldsymbol{V}_2$

- $\lambda_1 \approx \lambda_2 \approx 0$: No strong gradient, image is approximately constant

- $\lambda_1 \approx \lambda_2 \gg 0$: No *linear* level sets, level sets have strong curvature

We want to study the variation of the structure tensor S under the $SO(2)$ at the unit element e. Let S_θ be the structure tensor where the local coordinate frame A is rotated by the $SO(2)$ (see eq. (2.175))

$$S_\theta = \int_A w \left(\|\boldsymbol{x}(\theta)\| \right) \left(\nabla_{\boldsymbol{x}(\theta)} \phi \left(\boldsymbol{x}(\theta) \right) \nabla^T_{\boldsymbol{x}(\theta)} \phi \left(\boldsymbol{x}(\theta) \right) \right) d^2 x_\theta \tag{3.7}$$

The $SO(2)$ only rotates the domain A and does not deform it otherwise, thus the integral measure $d^2 x_\theta$ is independent of θ, $d^2 x_\theta = d^2 x$. Since the weighting function w only depends on the norm $\|\boldsymbol{x}(\theta)\|$ which is preserved by the $SO(2)$, it is also invariant. The only component which changes is the gradient $\nabla_{\boldsymbol{x}(\theta)}$. Using eq. (2.176) and the product rule we can compute the derivative of S_θ at $\theta = 0$

$$\left. \frac{d}{d\theta} S_\theta \right|_{\theta=0} = \int_A w \left(\|\boldsymbol{x}\| \right) \left(\mathbf{M}_\theta \nabla \phi \nabla^T \phi + \nabla \phi \nabla^T \phi \mathbf{M}_\theta^T \right) d^2 x \tag{3.8}$$

$$= \mathbf{M}_\theta \cdot S - S \cdot \mathbf{M}_\theta = [\mathbf{M}_\theta, S] \tag{3.9}$$

In eq. (3.9) we used $\mathbf{M}_\theta^T = -\mathbf{M}_\theta$. We can get some information on the magnitude of the rate of change $\left. \frac{d}{d\theta} S_\theta \right|_{\theta=0}$ by multiplying the commutator in eq. (3.9) with the eigenvectors $\boldsymbol{v}_{1,2}$

$$\widetilde{\boldsymbol{v}}_{1,2} = [\mathbf{M}_\theta, S] \, \boldsymbol{v}_{1,2} \tag{3.10}$$

It is easy to show that both projections $\widetilde{\boldsymbol{v}}_{1,2}$ in eq. (3.10) have the same norm

$$\|\widetilde{\boldsymbol{v}}_{1,2}\| = |\lambda_1 - \lambda_2| \tag{3.11}$$

With the help of eq. (3.11) we can reformulate our characterization of the eigenvalues $\lambda_{1,2}$ in the following way

- $\lambda_1 > \lambda_2$: Structure tensor S has strong change under $SO(2)$

- $\lambda_1 \approx \lambda_2$: Structure tensor S is largely invariant under the $SO(2)$ and approximately diagonal, $S \sim \mathbb{1}_{2\times 2}$ where $\mathbb{1}_{2\times 2}$ is the unity element of $\mathbb{R}^2$

3.2. Structure Tensor Based Prior

Since the vector field V_e^{Ω} in eq. (3.1) is translation invariant the structure tensor S is also translation invariant. Under the rotation group $SO(2)$ the structure tensor is *not* invariant. Nonetheless it has an important transformation property: the transformed structure tensor S' may be written in terms of the old matrix S and the rotation matrix $R_\theta \in SO(2)$

$$S' = R_\theta^T S R_\theta \tag{3.12}$$

We would like to construct a prior E_{ST}^{prior} which is invariant under the combined group $\mathbb{G}_\Omega = \mathbb{T} \times SO(2)$. Since the eigenvalues λ_i of the structure tensor S are positive definite we propose as an energy prior for ϕ the integral over the determinant of S

$$E_{ST}^{prior} = \int_\Omega \mathcal{E}_{ST}(S) \, d^2x \tag{3.13}$$

$$\mathcal{E}_{ST}(S) = \frac{\lambda}{2} \det(S) \tag{3.14}$$

We want to show that E_{ST}^{prior} is invariant under the $SO(2)$. We insert S_θ from eq. (3.7) into the determinant in eq. (3.14) and evaluate the derivative of E_{ST}^{prior} with respect to θ

$$\left. \frac{d}{d\theta} E_{ST}^{prior}(S_\theta) \right|_{\theta=0} = \int_\Omega \mathrm{Tr}\left(\mathbf{P}^{ST} \cdot [\mathbf{M}_\theta, S] \right) d^2x, \quad P_{ij}^{ST} = \frac{\delta \mathcal{E}_{ST}}{\delta S_{ij}} \tag{3.15}$$

The matrix $\mathbf{P}^{ST}$ is the canonical momentum with respect to the structure tensor S, thus $\mathbf{P}^{ST}$ has the same transformation properties under the $SO(2)$ as S. The trace in eq. (3.15) can be further transformed

$$\mathrm{Tr}\left(\mathbf{P}^{ST} \cdot [\mathbf{M}_\theta, S] \right) = 2 \cdot \mathrm{Tr}\left(\mathbf{P}^{ST} \cdot \mathbf{M}_\theta \cdot S \right) \tag{3.16}$$

The matrix under the trace on the right hand side of eq. (3.16) is a product of a symmetric and an anti-symmetric matrix, and thus itself anti-symmetric. Since traces over anti-symmetric matrices vanish, it follows that the prior E_{ST}^{prior} is

invariant under the $SO(2)$

$$\left.\frac{d}{d\theta}E_{ST}^{prior}(S_\theta)\right|_{\theta=0} = 2\int_\Omega \text{Tr}\left(\mathbf{P}^{ST}\cdot\mathbf{M}_\theta\cdot S\right)d^2x = 0 \tag{3.17}$$

We note that the symmetry expressed by eq. (3.17) is a non-trivial symmetry, since only the trace as a whole vanishes.

3.3. Geometrical Optical Flow Model

In section 2.11 we introduced the notion of an optical flow field $\boldsymbol{d}(\boldsymbol{x})$ which maps the domain Ω_I of an image $I(\boldsymbol{x})$ recorded by the camera C_I to the domain Ω_y of the image $Y(\boldsymbol{x})$ recorded by the camera C_Y (see figure 2.4). The basic variational method outlined for obtaining $\boldsymbol{d}(\boldsymbol{x})$ was: The optical flow $\boldsymbol{d}(\boldsymbol{x})$ is computed by minimizing the energy functional $E_{Y,I}(\boldsymbol{d})$ (see eq. (2.247)) which contains a data term (also called similarity measure in this context) $E_{Y,I}^{data}(\boldsymbol{d})$ and a prior energy for the gradient of $\boldsymbol{d}$, $E^{prior}(\nabla\boldsymbol{d})$. The similarity measure $E_{Y,I}^{data}\left(\hat{\boldsymbol{d}}\right)$ basically tells us how similar the images $Y(\boldsymbol{x})$ and $I_{\hat{d}}(\boldsymbol{x})$ (defined in eq. (2.246)) are given an estimate $\hat{\boldsymbol{d}}$ of the optical flow field $\boldsymbol{d}$.

In the uni-modal case in section 2.11.1 the similarity measure $E_{Y,I}^{data}$ was defined as the pixel difference between the $Y(\boldsymbol{x})$ and $I_d(\boldsymbol{x})$ (see eq. (2.249)). In section 2.11.2 it was explained that the measure $E_{Y,I}^{data}$ defined in eq. (2.249) is generally insufficient in the multi-modal case where the cameras C_I and C_Y may be sensitive to different light spectra. We introduced alternative similarity measures based on Mutual Information (MI) (eq. (2.255), [67]), Correlation Ratio (CR) (eq. (2.258), [94]) and Cross Correlation (CC) (eq. (2.260), [95]). The similarity measure based on MI, CR and CC pose constraints on the similarity $Y(\boldsymbol{x})$ and $I_{\hat{d}}(\boldsymbol{x})$ in ascending order with CC posing the strongest constraint $Y(\boldsymbol{x}) = f \cdot I(\boldsymbol{x}) + \beta$ (see eq. (2.260)). However all three measures assume that the images $Y(\boldsymbol{x})$ and $I_d(\boldsymbol{x})$ have the same spacial and optical resolution such in the case of CC an incorrect optical flow $\boldsymbol{d}^\star$ is estimated (see figure 2.5).

In section 2.12 we outlined the method of Hardie et. al. [47] in which given a low resolution image y the goal is to estimate an image Y with higher resolution with the aid of a high resolution image I from an external camera C_I. The virtue of their method is that it integrates the physical relationship between the CCD of C_y and that of C_I. This relationship is embodied by the PSF $W_{\sigma^{sc}}$ in eq. (2.263). However in their paper [47] they assumed C_y and C_I to co-aligned (see figure 2.6). In this section we will extend this method to incorporate an optical flow field mapping Ω_I to Ω_Y (the domain of the unknown high resolution
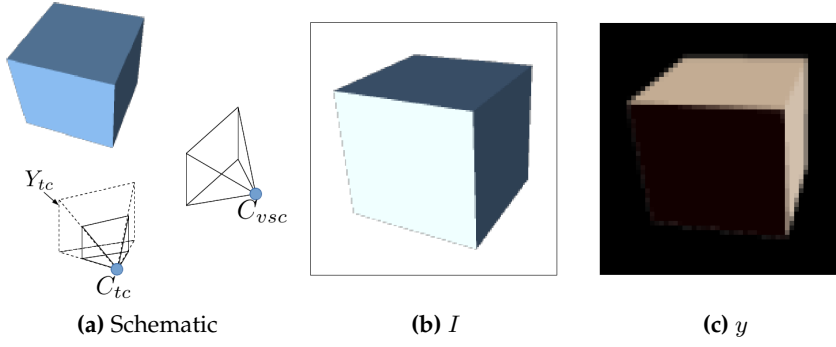
(a) Schematic   **(b)** I **(c)** y

Figure 3.1.: Figure 3.1a shows the setup of a thermographic camera (TC), C_{tc}, and a visual spectrum camera (VSC), C_{vsc}, recording an object O. Figure 3.1b shows the image I which is recorded by C_{vsc} and figure 3.1c the lower resolution image y recorded by C_{tc}. The solid line cone of C_{tc} in figure 3.1a which is small compared to the cone of C_{vsc} indicates the low resolution of the TC compared to that of the VSC. The dotted cone indicates the high resolution of the image Y, which is jointly estimated together with the optical flow d (the mapping between I and y) by the model in eq. (3.25)

Y). The result, a model capable of jointly estimating $d(x)$ and Y given y and I is basically a CC-type similarity measure encoding the difference in optical resolution between the C_y and the C_I camera.

3.4. Multi-Modal Optical Flow with Differing Resolutions

We consider the camera setup in figure 3.1a with the low resolution thermographic camera C_{tc} and the high resolution visual spectrum camera C_{vsc}. As opposed to the setup in Hardie et. al. [47] (figure 2.6) the cameras C_{tc} and C_{vsc} are not co-aligned. The goal in this section is to extend the method introduced in section 2.12 to include the unknown optical flow field $d(x)$ representing the separation of C_{tc} and C_{vsc}. In a nutshell we assume the low resolution image y to be co-aligned with the image I_d

$$I_d(x) = I(x + d(x)) \tag{3.18}$$

for a given optical flow field $d(x)$. From eq. (2.270) we can compute the super resolved image $\hat{Y}_d(x)$ which is a function of d. In principle the energy in eq. (2.269) with the image $\hat{Y}_d(x)$ then serves as a similarity measure between y and I_d

$$E_{y,I}^{data}(d) = E_{y,I_d}\left(\hat{Y}_d\right) \tag{3.19}$$

$\hat{Y}_d$ is an implicit function which does not need to be estimated directly since by virtue of eq. (2.270) it is easily computed when needed. However since we are only interested in d we do not need to explicitly evaluate $\hat{Y}_d$.

Computation of the similarity measure $E_{y,I}^{data}$

We will now briefly compute the exact form of $E_{y,I}^{data}(d)$ in eq. (3.19). A detailed description can be found in appendix D. In section 2.12 the model for computing the super- resolved image Y is given by the energy $E(Y)$ (see eq. (2.269))

$$E_{y,I}(Y) = \frac{1}{2}\int_\Omega \left(y(\boldsymbol{x}) - W_{\sigma^{sc}}Y(\boldsymbol{x})\right)^2 \cdot C_n^{-1}d^2x \tag{3.20}$$

$$+\frac{1}{2}\int_\Omega \left(Y(\boldsymbol{x}) - \mu_{Y|I}(\boldsymbol{x})\right)^2 \cdot C_{Y|I}^{-1}d^2x \tag{3.21}$$

with the conditional variance and mean

$$C_{Y|I} = C_{Y,Y}^{\sigma^{sc}} - C_{Y,I}^{\sigma^{sc},2} \cdot C_{I,I}^{\sigma^{sc},-1} \tag{3.22}$$

$$\mu_{Y|I}(\boldsymbol{x}) = \mu_Y + f \cdot (I(\boldsymbol{x}) - \mu_I), \quad f = C_{Y,I}^{\sigma^{sc}} \cdot C_{I,I}^{\sigma^{sc},-1} \tag{3.23}$$

The first integrand of $E_{y,I}(Y)$ in eq. (3.20) models the relationship between the low resolution image y of the camera C_{tc} and the unknown high resolution image Y, namely that one pixel in y is mapped to a window of pixels in Y via the PSF $W_{\sigma^{sc}}$ (figure 2.7a). Essentially it couples the low resolution component of Y, $W_{\sigma^{sc}}Y$ to the C_{tc} image y. On the other side the second integrand of $E_{y,I}(Y)$ in eq. (3.21) models the relationship between the *intensities* of Y and I. This done by transforming the spectrum of I via the factor f (eq. (3.23)) to match that of Y. Since this is done on a pixel-by-pixel basis, eq. (3.21) pins down the high resolution component of Y.

At this point we incorporate the optical flow $d(\boldsymbol{x})$ which separates the cameras C_{tc} and C_{vsc} by assuming Y to be co-aligned the warped image I_d (eq. (3.18)). Thus we substitute I for I_d in the integrand in eq. (3.21)

$$E(Y,d) = \frac{1}{2}\int_\Omega \left(y(\boldsymbol{x}) - \langle Y\rangle_{\sigma^{sc}}(\boldsymbol{x})\right)^2 \cdot C_n^{-1}d^2x \tag{3.24}$$

$$+\frac{1}{2}\int_\Omega \left(Y(\boldsymbol{x}) - \mu_{Y|I_d}(\boldsymbol{x})\right)^2 \cdot C_{Y|I_d}^{-1}d^2x \tag{3.25}$$

with expressions for $C_{Y|I_d}$ and $\mu_{Y|I_d}$ similar to those in eqs. 3.22 and 3.23.

While keeping d fixed we minimize $E(Y,d)$ with respect to Y and obtain a
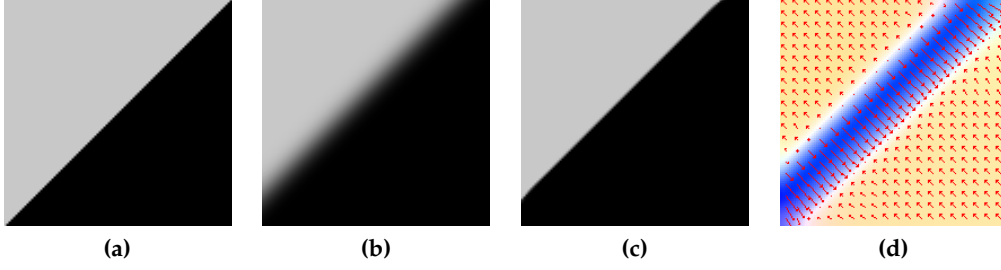
Figure 3.2.: Figure 3.2a shows a synthetic high resolution image I^{syn}. In figure 3.2b we show a low resolution image y^{syn}. y^{syn} is computed by convolution of I^{syn} with Gaussian $G_{\sigma^{sc}}$ with standard deviation $\sigma^{sc} = 5$ and translated by 10 pixels relative to I^{syn}. Figure 3.2d shows the flow d computed with the model in eq. (3.27), which incorporates knowledge of the scale difference between y^{syn} and I^{syn} and figure 3.2c show the warped image I_d

simplified closed form solution for Y similar to eq. (2.271)

$$\hat{Y}_d = \mu_{Y|I_d} + C^{\sigma^{sc}}_{\langle Y \rangle_{\sigma^{sc}}|\langle I_d \rangle_{\sigma^{sc}}} \cdot \left(C^{\sigma^{sc}}_{\langle Y \rangle_{\sigma^{sc}}|\langle I_d \rangle_{\sigma^{sc}}} + C_n \right)^{-1} \left(y - \tilde{\mu}_{Y|I_d} \right) \qquad (3.26)$$

We insert the closed form expression for $\hat{Y}$ from eq. (3.26) into $E(Y, d)$ and obtain an energy measuring the similarity between y and $\langle I \rangle_{\sigma^{sc},d} = W_{\sigma^{sc}} I_d$

$$E^{data}_{y,I}(d) = E_{y,I_d}\left(\hat{Y}_d\right)$$
$$= \frac{1}{2} \int_\Omega \left(y(x) - \mu_Y - f \cdot (\langle I \rangle_{\sigma^{sc},d}(x) - \mu_I) \right)^2 \cdot U^{\sigma^{sc}} \qquad (3.27)$$
$$f = C^{\sigma^{sc}}_{y,\langle I_d \rangle_{\sigma^{sc}}} C^{\sigma^{sc},-1}_{\langle I_d \rangle_{\sigma^{sc}},\langle I_d \rangle_{\sigma^{sc}}} \qquad (3.28)$$
$$U^{\sigma^{sc}} = C^{\sigma^{sc}}_{\langle Y \rangle_{\sigma^{sc}}|\langle I_d \rangle_{\sigma^{sc}}} \left(C^{\sigma^{sc}}_{\langle Y \rangle_{\sigma^{sc}}|\langle I_d \rangle_{\sigma^{sc}}} + C_n \right)^{-2} \qquad (3.29)$$

$E^{data}_{y,I}(d^\star)$ in eq. (3.27) is minimal when the specific optical flow $d^\star$ satisfies

$$y(x) \approx f \cdot \langle I \rangle_{\sigma^{sc},d^\star} + (\mu_Y - f \cdot \mu_I) \qquad (3.30)$$

Comparing eq. (3.30) to the CC eq. (2.260) we can see that with $\beta = (\mu_Y - f \cdot \mu_I)$ eq. (3.30) is also a CC-type equation, however it correlates the low resolution images y and $\langle I \rangle_{\sigma^{sc},d^\star}$.

It is easy to show that

$$C^{\sigma^{sc}}_{\langle Y \rangle_{\sigma^{sc}}|\langle I_d \rangle_{\sigma^{sc}}} = C^{\sigma^{sc}}_{Y,Y} \left(1 - \eta_{CC}(Y|I_d)\right) \qquad (3.31)$$

This causes $C^{\sigma^{sc}}_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}\left(C^{\sigma^{sc}}_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}+C_n\right)^{-2}$ to additionally prune the optical flow $\boldsymbol{d}(\boldsymbol{x})$, such that $E^{data}_{y,I}$ is a more robust similarity measure then CC.

To demonstrate that our likelihood $E^{data}_{y,I}$ in eq. (3.27) respects the difference in scale between y and I we have estimated the flow with $E^{data}_{Y,I_{\boldsymbol{d}}}$ as the similarity measure for the data y^{syn} and I^{syn} in figure 2.5. The standard deviation σ^{sc} in $E^{data}_{y,I}$ was set to $\sigma^{sc}=5$ and the factor f is automatically computed as $f\approx 1$ since the intensity distributions of y^{syn} and I^{syn} are aproximately the same. The image $I^{syn}_{\boldsymbol{d}}$ is convolved with $W_{\sigma^{sc}}$. The resulting image $\widetilde{I}^{syn}$ has the same scale as y^{syn}. The resulting optical flow $\boldsymbol{d}^{syn}$ is shown in figure 3.2d. Notice the blurred boundary $\boldsymbol{d}^{syn}$ around the linear feature in I^{syn} (figure 3.2a). This is the result of $E^{data}_{y,I}$ in eq. (3.27) measuring the difference between y^{syn} and the blurred image $\widetilde{I}^{syn}_{\boldsymbol{d}}=W_{\sigma^{sc}}I_{\boldsymbol{d}}$. In figure 3.2c we see $I_{\boldsymbol{d}^{syn}}$. The linear boundary has been warped by $\boldsymbol{d}^{syn}$ without being corrupted like in figure 2.5d.

3.5. Localization

The similarity measure $E^{data}_{y,I}$ in eq. (3.27) basically compares the image y with the transformed image $f\cdot\langle I\rangle_{\sigma^{sc},\boldsymbol{d}}$, where f is defined in eq. (3.28). Thus for some minimizer of $E^{data}_{y,I}(\boldsymbol{d})$, $\boldsymbol{d}^{\star}$ the image $f\cdot\langle I\rangle_{\sigma^{sc},\boldsymbol{d}}$ is assumed to approximate y

$$y\left(\boldsymbol{x}\right)\approx f\cdot\widetilde{I}_{\boldsymbol{d}^{\star}}\left(\boldsymbol{x}\right) \tag{3.32}$$

Eq. (3.32) manifests the assumption of a *global* linear relationship between the intensities of the image y and and those of I, with f being the linear factor. There are many situations in multi modal optical flow where the linearity relation expressed by eq. (3.32) is not valid. For instance in figure 3.9 in section 3.7.5 the images I_{vsc} and y_{tc} of a visual spectrum camera (VSC) and a thermographic camera (TC) recording the same scene is shown for which the assumption of global linearity between I_{vsc} and y_{tc} fails, since objects that have similar reflectance properties in the visual spectrum can have differing properties in the infra-red spectrum. However if we focus on small regions $\mathcal{A}\subset\Omega$ then a local relation similar to eq. (3.32)

$$y\left(\boldsymbol{x}\right)|_{\mathcal{A}}\approx f\cdot\langle I\rangle_{\sigma^{sc},\boldsymbol{d}^{\star}}\left(\boldsymbol{x}\right)|_{\mathcal{A}} \tag{3.33}$$

might hold provided that the region $\mathcal{A}$ is occupied by the same object in both images which will be the case in section 3.7.5. Therefore we propose a local

version of the variance and the mean in eq. (2.267)

$$C_{u,v}(\boldsymbol{x}_0) = \int_\Omega \omega(\boldsymbol{x} - \boldsymbol{x}_0)(u(\boldsymbol{x}) - \mu_u(\boldsymbol{x}_0)) \cdot (v(\boldsymbol{x}) - \mu_v(\boldsymbol{x}_0)) d^2x \qquad (3.34)$$

$$\mu_u(\boldsymbol{x}_0) = \int_\Omega \omega(\boldsymbol{x} - \boldsymbol{x}_0) u(\boldsymbol{x}) d^2x$$

where ω is a window function which we take to be constant within a square window $U_{\boldsymbol{x}_0} \subset \Omega$ centered around the point $\boldsymbol{x}_0 \in \Omega$ with the window size a

$$\omega(\boldsymbol{x}) = \begin{cases} \frac{1}{|U_{\boldsymbol{x}_0}^a| - 1} & \boldsymbol{x} \in U_{\boldsymbol{x}_0}^a, \quad width(U_{\boldsymbol{x}_0}^a) = height(U_{\boldsymbol{x}_0}^a) = a \\ 0 & else \end{cases} \qquad (3.35)$$

With this definition the conditional variance $C_{Y|I_{\boldsymbol{d}}}$ and the factor f (eq. (3.28)) become functions of the coordinates $\boldsymbol{x} \in \Omega$ and the parameter a

$$C_{\langle Y \rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}} \rangle_{\sigma^{sc}}}^{\sigma^{sc},a}(\boldsymbol{x}) = C_{Y,Y}^{\sigma^{sc},a}(\boldsymbol{x}) - C_{y,\langle I_{\boldsymbol{d}} \rangle_{\sigma^{sc}}}^{\sigma^{sc},a,2}(\boldsymbol{x}) \cdot C_{\langle I_{\boldsymbol{d}} \rangle_{\sigma^{sc}},\langle I_{\boldsymbol{d}} \rangle_{\sigma^{sc}}}^{\sigma^{sc},a,-1}(\boldsymbol{x}) \qquad (3.36)$$

$$f^{\sigma^{sc},a}(\boldsymbol{x}) = C_{y,\langle I_{\boldsymbol{d}} \rangle_{\sigma^{sc}}}^{\sigma^{sc},a}(\boldsymbol{x}) \cdot C_{\langle I_{\boldsymbol{d}} \rangle_{\sigma^{sc}},\langle I_{\boldsymbol{d}} \rangle_{\sigma^{sc}}}^{\sigma^{sc},a,-1}(\boldsymbol{x}) \qquad (3.37)$$

We substitute $C_{\langle Y \rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}} \rangle_{\sigma^{sc}}}^{\sigma^{sc},a}$ and $f^{\sigma^{sc},a}$ in eq. (3.27) with the local versions from eqs. 3.36 and 3.37 and obtain

$$E_{y,I_{\boldsymbol{d}}}^{data,l}(\sigma^{sc}, a, \boldsymbol{d}) = \frac{1}{2} \int_\Omega \left(y(\boldsymbol{x}) - \langle f^{\sigma^{sc},a} I_{\boldsymbol{d}} \rangle_{\sigma^{sc}}(\boldsymbol{x}) \right)^2 \cdot U^{\sigma^{sc},a}(\boldsymbol{x}) d^2x$$

$$U^{\sigma^{sc},a}(\boldsymbol{x}) = C_{\langle Y \rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}} \rangle_{\sigma^{sc}}}^{\sigma^{sc},a}(\boldsymbol{x}) \left(C_{\langle Y \rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}} \rangle_{\sigma^{sc}}}^{\sigma^{sc},a}(\boldsymbol{x}) + C_n \right)^{-2} \qquad (3.38)$$

Notice that the PSF $W_{\sigma^{sc}}$ is now convolved with the product $f^{\sigma^{sc},a}(\boldsymbol{x}) \cdot I_{\boldsymbol{d}}(\boldsymbol{x})$ since $f^{\sigma^{sc},a}$ is now a function.

3.6. The Multigrid Newton algorithm

We combine now our global similarity measures in eq. (3.27) and together with the structure tensor prior E_{ST}^{prior} (eq. (3.14)) and the TV prior E_{TV}^{prior} (eq. (2.229)) to the following two models

$$E_{ST}^g(\boldsymbol{d}) = E_{Y,I}^{data}(\sigma^{scale}, \boldsymbol{d}) + \lambda_{ST} \sum_{i=1}^2 E_{ST}^{prior}(\nabla d_i) \qquad (3.39)$$

$$E_{TV}^g(\boldsymbol{d}) = E_{Y,I}^{data}(\sigma^{scale}, \boldsymbol{d}) + \lambda_{TV} \sum_{i=1}^2 E_{TV}^{prior}(\nabla d_i) \qquad (3.40)$$

Algorithm 2 Multigrid Optical Flow (MOF)

Initialize $\boldsymbol{d}_0 = 0$, $k = 0$, $\sigma_{MOF} = 0.7$
Set $\boldsymbol{r}_0 = \frac{\delta E(\boldsymbol{d})}{\delta \boldsymbol{d}}(\boldsymbol{d}_0)$
scale $s = s_{\text{Max}}$
while $s > 1$ **do**
    downsample $y_s = G_{s \cdot \sigma_{MOF}} \star y_0$, $I_s = G_{s \cdot \sigma_{MOF}} \star I_0$
    while $\|r\| > \epsilon$ or $k < N$ **do**
        set $\boldsymbol{d}_{k+1} = \boldsymbol{d}_k + \alpha \boldsymbol{\delta}$
        expand $E(\boldsymbol{d}_{k+1}) = E(\boldsymbol{d}_k) + \alpha \boldsymbol{b}_k^T \boldsymbol{\delta} + \frac{\alpha^2}{2} \boldsymbol{\delta}^T Q_k \boldsymbol{\delta}$
        solve $Q_k \boldsymbol{\delta} = \boldsymbol{b}_k$ for $\boldsymbol{\delta}$ with conjugate gradients
        compute $\boldsymbol{d}_{k+1} = \boldsymbol{d}_k + \alpha \boldsymbol{\delta}$, $k \to k + 1$
    end while
    upsample $\boldsymbol{d}_N$, set $\boldsymbol{d}_0 = \boldsymbol{d}_N$, $k = 0$
    $s = s - 1$
end while
set optimal solution $\boldsymbol{d}^\star = \boldsymbol{d}_N$

To minimize the models in eqs. 3.39 to 3.40 and obtain the optimum flow field $\boldsymbol{d}^\star$ we deploy a simple newton scheme with a nested linearization (see alg. 2 (MOF) where $E(\boldsymbol{d})$ is to be substituted with the models in eqs. 3.39 to 3.40). The linearized model is solved by a conjugate gradients algorithm with block Jacobi preconditioning.

When minimizing the structure tensor based model E_{ST}^g (eq. (3.39)) the following major numeric problem occurs: The problem arises in step 9 of the MOF algorithm. The second functional derivative $Q_{k,ST}^g$ of the energy model E_{ST}^g consists of one part comming from the likelihood and one part coming from the prior, $Q_k^g = Q_k^{data,g} + \lambda_{ST} Q_{k,ST}^{prior}$. The matrix $Q_{k,ST}^{prior}$ is the second derivative of the prior E_{ST}^{prior} with respect to the individual components of $\boldsymbol{d}$

$$Q_{k,ST}^{prior,ii} = \frac{\delta^2}{\delta d_i^2} E_{ST}^{prior}(\nabla d_i) \quad Q^{prior,ij} = 0 \tag{3.41}$$

Since $E_{ST}^{prior}(\nabla d_i)$ is purely quartic in d_i by definition (eq. (3.14)) the matrix $Q_{k,ST}^{prior}$ in eq. (3.41) a purely quadratic functional of d_i. At small k in alg 2 its eigenvalues equal zero due to the initial guess $\boldsymbol{d}_0 = 0$. The matrix Q_k^{data} is the second derivative of the data term $E_{Y,I}^{data}$ in eq. (3.39) and eq. (3.40). In regions where there is no motion the eigen values of Q_k^{data} are also small. This makes the linearized solution in step 9 numerically instable and the instability carries on to all stages s of the MOF alg. 2. Our solution to this problem is to extend

$E_{ST}^{prior}(\nabla d_i)$ to include an L_2 prior on the flow field $\boldsymbol{d}$ but with a small lagrange multiplier λ_2

$$\widetilde{E}_{ST}^{prior}(\nabla d_i) = E_{ST}^{prior}(\nabla d_i) + \lambda_2 \left\| \nabla d_i \right\|^2 \tag{3.42}$$

With the L_2 prior in 3.42 the linearized solution in step 9 becomes numerically stable. Thus we substitute E_{ST}^{prior} for $\widetilde{E}_{ST}^{prior}$ (eq. (3.42)) in eq. (3.39)

$$E_{ST}^g(\boldsymbol{d}) = E_{Y,I}^{data}(\sigma^{scale}, \boldsymbol{d}) + \lambda_{ST} \sum_{i=1}^{2} \widetilde{E}_{ST}^{prior}(\nabla d_i) \tag{3.43}$$

The challenge is to find the appropriate setting for λ_2 in eq. (3.42) such that on the one side the MOF alg. 2 becomes numerically stable but on the other side the isotropic L_2 part in eq. (3.42) does not over weigh the original anisotropic prior $\widetilde{E}_{ST}^{prior}$. In section 3.7.6 we will map out a strategy for finding a suitable value for λ_2 based on the eigen values of the second order functional derivative of E_{ST}^{prior} in eq. (3.42).

3.7. Results

In this section we want to evaluate and compare the two optical flow models in eqs. 3.39 and 3.40. The methodology for this section is as follows: In section 3.7.1 we want to study the effect of the size of the structure tensor S in eq. (3.39) on the estimation of the optical flow $\boldsymbol{d}$. We evaluate the models in eqs. 3.39 and 3.40 on a sequence of two images obtained from *one* camera from the Middleburry data-set [3] (figure 3.3). The Middleburry data-set offers also the ground truth optical flow $\boldsymbol{d}_{gt}$ which we use to asses the quality of the estimated optical flow of the models in eqs. 3.39 and 3.40. Since the data in figure 3.3 is collected from one camera we use the global similarity measure in eq. (3.27) thereby setting $\sigma = 0$. Thus eq. (3.27) effectively reduces to the brightness constancy constraint in eq. (2.249). We compare the results of the models 3.39 and 3.40 for different window sizes σ_{ST} of the structure tensor S to find the best value for σ_{ST}. Having found the optimal windowsize σ_{ST} we will extend the evaluation of section 3.7.1 to include a synthesized multi-modal image set in section 3.7.4. The goal of that section is to study the effect of the PSF $W_{\sigma^{sc}}$ on the global similarity measure $E_{Y,I}^{data}(\sigma^{sc}, \boldsymbol{d})$ in a multi-modal coaxial setup similar to that in figure 2.6. Hereby we will create an artificial low resolution image y_{test}^{scale} by taking the image I in figure 3.3a, filtering it with a Gaussian filter of standard deviation σ^{sc}_{test}, thus creating an artificial scale difference and inverting the result. We will show that $E_{y_{test}^{scale}, I}^{data}(\sigma^{sc}, \boldsymbol{0})$ is minimal at $\sigma^{sc} = \sigma^{sc,\star}$.

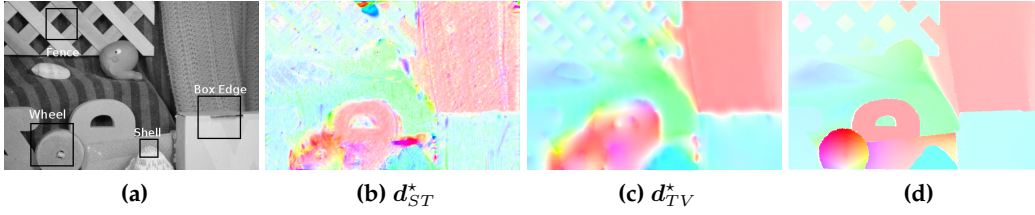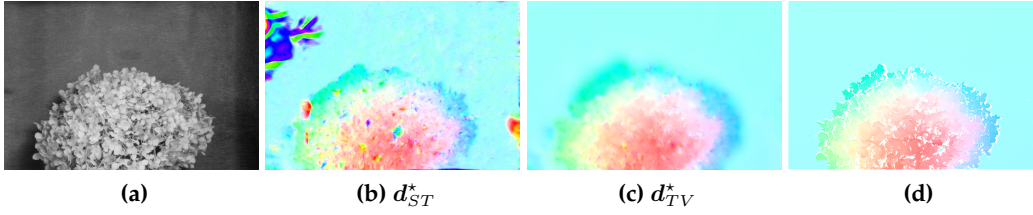| (a) | (b) $d_{ST}^{\star}$ | (c) $d_{TV}^{\star}$ | (d) |

Figure 3.3.: Rubberwhale Sequence: Figure 3.3a shows one frame of the sequence. figure 3.3b shows the estimated optical flow $d_{ST}^{\star}$, figure 3.3c the flow $d_{TV}^{\star}$ and figure 3.3d shows the provided ground truth

With the optimal windowsize σ_{ST} for the structure tensor S and a strategy for finding the correct scale difference we turn to the problem of estimating the optical flow on the real data in section 3.7.5. As we discussed in section 3.5 the similarity measure $E_{Y,I}^{data}(\sigma^{sc}, d)$ is based on the assumption that the images Y and I share the linear relationship $Y = \alpha I + \beta$ which is not supported when the cameras C_Y and C_I are sensitive to different light spectra. For this reason we will deploy the local similarity measure $E_{Y,I}^{data,l}(\sigma^{sc}, d)$ introduced in eq. (3.38) in the localized optical flow models

$$E_{ST}(d) = E_{y,I}^{data,l}(\sigma^{sc}, a, d) + \lambda_{ST} E_{ST}^{prior}(\nabla d) \tag{3.44}$$

$$E_{TV}(d) = E_{y,I}^{data,l}(\sigma^{sc}, a, d) + \lambda_{TV} E_{TV}^{prior}(\nabla d) \tag{3.45}$$

We will show that our strategy in section 3.7.4 for finding the true scale difference $\sigma^{sc,\star}$ between the images y and I also yields a strategy for finding the correct window size a for $E_{y,I}^{data,l}(\sigma^{sc}, a, d)$ such that $y = \alpha I + \beta$ *locally* holds.

3.7.1. Uni-Modal Data

We will now discuss the results of our optical flow method on the middleburry data set for which there exists ground truth (GT). As the GT is the true flow field for the data we use it to asses the quality of the computed optical flow. To do this we define the Endpoint error (EPE)

$$EPE(x) = \|d - d_{gt}\|(x) \tag{3.46}$$

The EPE meassures how well the computed optical flow d fits the true optical flow d_{gt}.

72

(a)   **(b) $d^{\star}_{ST}$**   **(c) $d^{\star}_{TV}$**   **(d)**

Figure 3.4.: Hydrangea Sequence: Figure 3.4a shows one frame of the sequence. figure 3.4b shows the estimated optical flow $d^{\star}_{ST}$, figure 3.4c the flow $d^{\star}_{TV}$ and figure 3.4d shows the provided ground truth

We will further need the definition of the mean curvature (eq. (2.242))

$$\kappa = \mathrm{Div}\left(\frac{\nabla I}{|\nabla I|}\right) \tag{3.47}$$

The curvature κ is a good measure to show which features can be reliably matched by the optical flow d and thus have small EPE values.

3.7.2. Rubber Whale Sequence

Our goal in this section is to evaluate the effect of the structure tensor window size σ_{ST} of the model E_{ST} (eq. (3.39)) on the quality of the optical flow mapping d between two images Y_C and I_C recorded by *the same camera* C. We compare the results of the model E_{ST} with those obtained by E_{TV}, thereby denoting $d^{\star}_{ST}$ the optical flow obtained by E_{ST} and $d^{\star}_{TV}$ the flow obtained by E_{TV}. In figure 3.3 the rubber whale sequence of the middleburry data set is shown, and in figure 3.3d the corresponding ground truth d_{gt}. This sequence is generated with one camera recording a dynamic scene. The reason for choosing this scene is that it contains linear level-set features as well curvy-linear features such as circular features. In figure 3.3a we have highlighted a linear level-set region of interest (ROI) labeled as *Box Edge,* a ROI partially containing linear structures labeled as *Fence,* a circular feature labeled as *Wheel* and a ROI containing a generic non-linear level-set called *Shell* . In figure 3.3b the computed flow-field $d^{\star}_{ST}$ for the energy $E_{ST}(d)$ (eq. (3.39)) is shown for a filter size of $\sigma_{ST} = 11$, while in figure 3.3c the resulting flow for the TV model is shown. We can observe from the comparison between figures 3.3b and 3.3c that the TV model produces smoother results which are closer to the ground truth (figure 3.3d). In figure 3.5 we have binned the curvature κ in to 40 bins and plotted the average EPE per curvature bin for both the ST model in eq. (3.39) (varying σ_{ST} in figures 3.5a to 3.5c) and for the TV model in eq. (3.40) (figure 3.5d). In all plots in figure 3.5
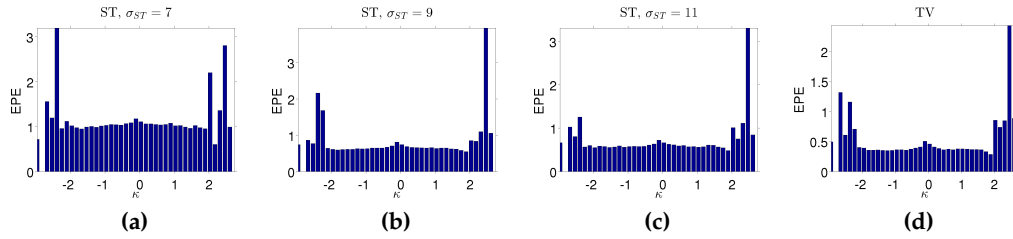
Figure 3.5.: EPE to level-set curvature: Figures 3.5a to 3.5d show plots of the EPE (eq. (3.46)) against the curvature κ (eq. (3.47)) for the rubber whale sequence (figure 3.3). Figures 3.5a to 3.5c show the results for the structure tensor model E_{ST} and figure 3.5d the result for the TV model E_{TV}. The curvature κ was split into 40 bins and the height of the bars is the average EPE per curvature bin.

ROI	Filtersize	Median EPE	ROI	Filtersize	Median EPE
	7	2.36		7	0.46
Wheel	9	1.32	Fence	9	0.39
	11	1.15		11	0.35
	TV	1.38		TV	0.18
	7	0.86		7	0.44
Shell	9	0.62	Box Edge	9	0.34
	11	0.50		11	0.30
	TV	0.17		TV	0.09

Table 3.1.: EPE for different filter-sizes σ_{ST} for the model E_{ST}^g (eq. (3.39)) and for the TV model E_{TV}^g (eq. (3.40)). The value shown in the column *Median EPE* is the median EPE per ROI. The median per ROI was chosen over the average EPE per ROI due to its robustness towards outlier EPE values. The EPE values for the model E_{ST}^g decrease with increasing structure tensor filtersizes σ_{ST}. However the general trend is that the ROI's with high curvatures κ (*Wheel* and *Shell*) tend to have higher EPE values then the ROI's with low curvatures (*Fence* and *Box Edge*).

the general tendency is that level-sets with low curvature values $\kappa \approx 0$ have high EPE values while with increasing curvature up to $\|\kappa\| \approx 2$ the EPE values fall off. This verifies the aperture problem (discussed in section 2.11) where less distinctive level-sets with intrinsic dimension $iD < 2$ (low κ) lead to less accurate optical flow estimates. However for higher curvature values $\|\kappa\| > 2$ the EPE values significantly rise. This is due to the fact that both the priors E_{ST}^{prior} (eq. (3.39)) and E_{TV}^{prior} (eq. (3.40)) penalize level-sets in the optical flow $\boldsymbol{d}$ which have higher curvature then level-sets which are more straight. In the case of E_{TV}^{prior} this is more evident since according to eqs. 2.242 and 2.243 in section 2.9 the curvature κ is the functional derivative of E_{TV}^{prior} and thus explicitly forced to vanish. Now comparing the plots in figures 3.5a to 3.5c we can see larger
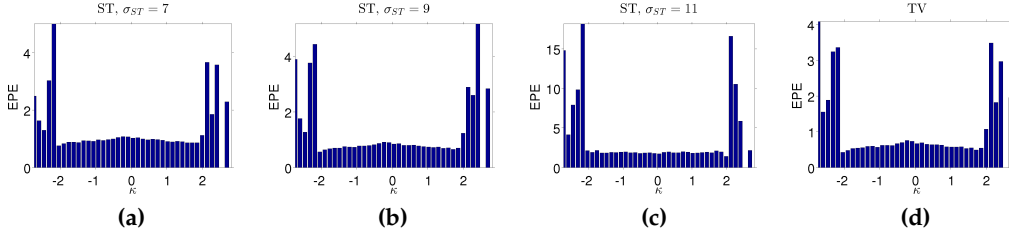
Figure 3.6.: EPE to level-set curvature: Figures 3.6a to 3.6d show plots of the EPE (eq. (3.46)) against the curvature κ (eq. (3.47)) for the hydrangea sequence (figure 3.4). Figures 3.6a to 3.6c show the results for the structure tensor model E_{ST} and figure 3.5d the result for the TV model E_{TV}. The curvature κ was split into 40 bins and the height of the bars is the average EPE per curvature bin.

window sizes σ_{ST} of the structure tensor in eq. (3.39) lead to small values for the EPE and thus more accurate optical flow estimates. In comparison with the TV model in eq. (3.40) (figure 3.5d) the EPE values structure tensor model E_{ST} with $\sigma_{ST} = 11$ in figure 3.5c are closest to those of the TV model in figure 3.5d with an average discrepancy of 0.3 between figure 3.5c and figure 3.5d. For this reason we will set $\sigma_{ST} = 11$ for the rest of this section.

Table 3.1 shows the median EPE values for different ROI's in figure 3.3a. We can see that ROI's with rather linear level-sets like the Fence and the Box Edge ROI have comparatively lower EPE values then ROI's containing level-set with larger curvature like the Wheel and the Shell for the E_{ST} model. For the E_{TV} only the Wheel has a higher EPE value.

3.7.3. Hydrangea Sequence

In figure 3.4a we show the hydrangea sequence of the middleburry dataset. In contrast to the rubber whale sequence in figure 3.3a figure 3.4a consists of a largely texture-less background and a hydrangea plant in the foreground. The hydrangea contains largely level-sets with high curvature κ which leads to erroneous optical flows $\boldsymbol{d}_{ST}^{\star}$ (figure 3.4b) and $\boldsymbol{d}_{TV}^{\star}$ (figure 3.4c) compared to the ground truth $\boldsymbol{d}_{gt}$ in figure 3.4d. In figure 3.6 we have again plotted the EPE against the curvature κ in a fashion similar to figure 3.5. Other than in figure 3.5 we can see that for increasing window sizes $\sigma_{ST} = 7$ (figure 3.6a) to $\sigma_{ST} = 11$ (figure 3.6c) for the model E_{ST} the EPE values increase, especially at higher curvatures $\|\kappa\| > 2$. Since the background in figure 3.4a is largely constant the condition defining the level-sets S_V (see eq. (2.179))

$$S_V = \{\boldsymbol{x} \,|V_e I\left(\boldsymbol{x}\right) = 0\} \quad V_e = \boldsymbol{v}\nabla \tag{3.48}$$

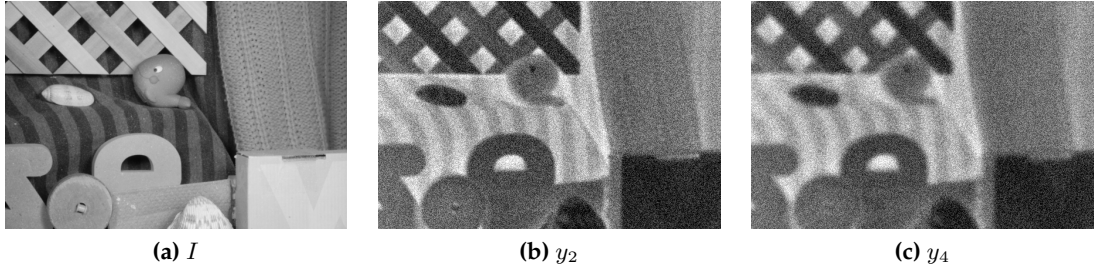**(a)** I **(b)** y_2 **(c)** y_4

Figure 3.7.: Synthesized multi-modal data. This data simulates the camera arrangement in figure 2.6. The image I in figure 3.7a is from the rubberwhale data set in figure 3.3. Figures 3.7b and 3.7c show the image $y_{\sigma^{sc}_{test}}$ (eq. (3.50)) at the scales $\sigma^{sc}_{test} = 2$ and $\sigma^{sc}_{test} = 4$

is independent of the vector V_e. In other words the background in figure 3.4a contains level-set of all possible curvatures κ. This leads to an equal distribution of the EPE values over the different curvatures κ in figure 3.6c. Comparing the E_{ST} model to the TV based model E_{TV} we see that the EPE values in figure 3.6d are comparatively smaller then those for the E_{ST} model in figures 3.6a to 3.6c . The EPE values in figure 3.6d however are larger then the corresponding values in figure 3.5d for the rubber whale sequence (figure 3.3a).

3.7.4. Estimation of the Scale Difference σ^{sc}

In this section we want to analyze the dependency of the similarity measure $E^{data}_{y,I}(\sigma^{sc}, \boldsymbol{d})$ on the scale difference parameter σ^{sc} of the PSF $W_{\sigma^{sc}}$. First we assume two co-aligned cameras C_y and C_I (see figure 2.6) with the images y and I being generated the following way (see figure 3.7): I is taken from the rubberwhale data set in figure 3.3. We simulate images $y_{\sigma^{sc}_{test}}$ of different resolutions $\sigma^{sc}_{test} = 1 \cdots 5$ first by inverting the intensities of I followed by filter operations with Gaussians of the standard deviations $\sigma^{sc}_{test} = 1 \cdots 5$ and the addition of iid noise

$$Y(\boldsymbol{x}) = -I(\boldsymbol{x}) + I_{min} + I_{max} \tag{3.49}$$

$$y^{test}_{\sigma}(\boldsymbol{x}) = \langle Y \rangle_{\sigma^{sc}_{test}}(\boldsymbol{x}) + n(\boldsymbol{x}), \quad n \sim \mathcal{N}(0, \chi_{std}) \tag{3.50}$$

I_{min} and I_{max} in eq. (3.49) are the minimum respectively maximum intensity of the image I and the standard deviation χ_{std} in eq. (3.50) was set to $\chi_{std} = 50$. In figure 3.7 the image I is shown along with the synthesized images y_2 and y_4 (y^{test}_{σ} from eq. (3.50) with $\sigma^{sc}_{test} = 2$ rsp. $\sigma^{sc}_{test} = 4$). The goal is now to show that $E^{data}_{y,I}(\sigma^{sc}, \boldsymbol{0})$ seen as a function of σ^{sc} is minimal at the test scale $\sigma^{sc,\star} = \sigma^{sc}_{test}$.
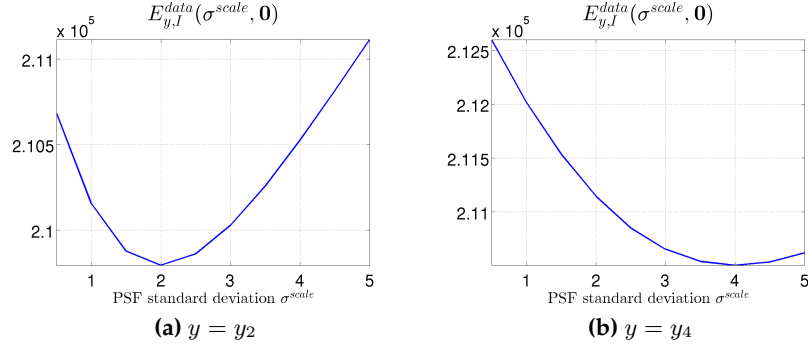
(a) $y = y_2$ **(b)** $y = y_4$

Figure 3.8.: Figures 3.8a and 3.8b show plots the similarity measure $E_{y,I}^{data}(\sigma^{sc}, \boldsymbol{d})$ for the cases $y = y_2$, $\sigma^{sc}{}_{test} = 2$, and $y = y_4$, $\sigma^{sc}{}_{test} = 4$. We can observe that $E_{y,I}^{data}(\sigma^{sc}, \boldsymbol{d})$ is minimal with respect to σ^{sc} at the correct scales $\sigma^{sc}{}_{test}$

This might seem obvious since $E_{y,I}^{data}$ which was computed in eq. (3.27)

$$E_{y,I}^{data}(\sigma^{sc}, \boldsymbol{d}) = \frac{1}{2} \int_\Omega \Big(y(\boldsymbol{x}) - f \cdot \langle I \rangle_{\sigma^{sc}}(\boldsymbol{x}) \Big)^2 \cdot F \tag{3.51}$$

$$f = C_{y,\langle I_{\boldsymbol{d}} \rangle_{\sigma^{sc}}}^{\sigma^{sc}} C_{\langle I_{\boldsymbol{d}} \rangle_{\sigma^{sc}}, \langle I_{\boldsymbol{d}} \rangle_{\sigma^{sc}}}^{\sigma^{sc},-1} \tag{3.52}$$

$$F = C_{\langle Y \rangle_{\sigma^{sc}} | \langle I_{\boldsymbol{d}} \rangle_{\sigma^{sc}}}^{\sigma^{sc}} \Big(C_{\langle Y \rangle_{\sigma^{sc}} | \langle I_{\boldsymbol{d}} \rangle_{\sigma^{sc}}}^{\sigma^{sc}} + \lambda C_n \Big)^{-2} \tag{3.53}$$

was derived from the basic assumption that y is the result of the convolution of the PSF $W_{\sigma^{sc}}$ with the high resolution image Y along with additive noise (see eq. (2.263))

$$y = \langle Y \rangle_{\sigma^{sc}} + n \quad n \sim \mathcal{N}(0|C_n) \tag{3.54}$$

which is similar to how we generated the test images y_σ^{test} in eq. (3.50). However the factor F in eq. (3.51) is highly non-linear in σ^{sc}, so we want to show that despite this non-linearity $E_{y,I}^{data}(\sigma^{sc}, \boldsymbol{0})$ has a global minimum at $\sigma^{sc,\star} = \sigma^{sc}{}_{test}$. In figure 3.8 we have plotted $E_{y,I}^{data}(\sigma^{sc}, \boldsymbol{0})$ over σ^{sc} for the cases $\sigma^{sc}{}_{test} = 2$ (figure 3.8a) and $\sigma^{sc} = 4$ (figure 3.8b). For both cases $E_{y,I}^{data}(\sigma^{sc}, \boldsymbol{0})$ is minimal at the correct scale $\sigma^{sc}{}_{test}$. From this we learn that $E_{y,I}^{data}(\sigma^{sc}, \boldsymbol{d})$ is sensitive to the scale difference σ^{sc} between y and I and we can use it to determine the true scale difference $\sigma^{sc}{}_{test}$.

3.7.5. Real Multimodal Optical Flow Data

We would now like to evaluate our optical flow model in eq. (3.39) on the data in figure 3.9. The image y_{tc} (figure 3.9b) was recorded with a thermographic

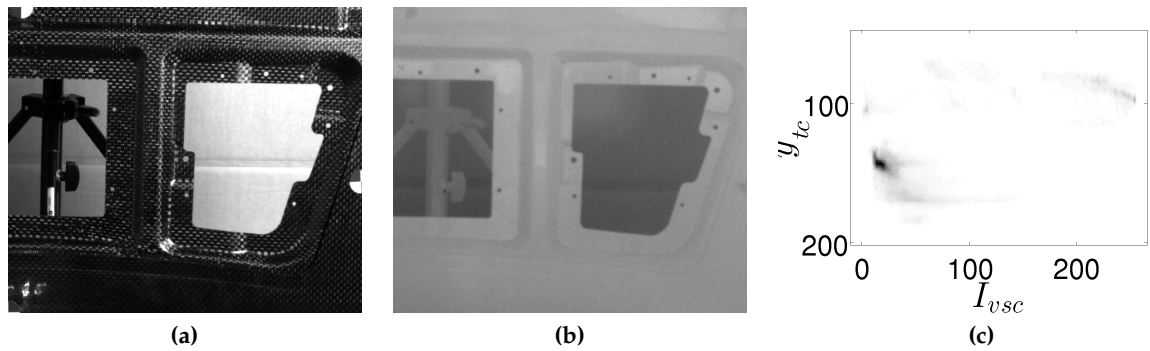| (a) | (b) | (c) |

Figure 3.9.: figure 3.9a shows an image from a visual spectrum camera (VSC). The object recorded is a carbon-fiber reinforced polymer (CFRP). Figure 3.9b shows an image of the same CFRP recorded with a thermographic camera (TC). The TC is sensitive in the infrared domain, thus higher intensities in figure 3.9b correspond to warmer objects (the CFRP) and lower intensities to colder objects (the background). As in figure 3.1a the optical centers of the VSC and the TC are physically separated so the problem that is being addressed is that of finding the optical flow field $d(x)$ (see eq. (2.245)) which maps every pixel in the TC image to the corresponding pixel in the VSC image. Figure 3.9c shows the joint histogram of the VSC and TC image. It shows a complex mapping of the intensities of figure 3.9a to those of figure 3.9b indicating that a linearity assumption between the TC and the VSC is not valid

camera (TC) C_{tc} and the image I_{vsc} with a visual spectrum camera (VSC) C_{vsc}. The recorded object is a carbon fiber reinforced polymer (CFRP). CFRP materials are becoming increasingly widespread in automotive and aerospace industries, but also in consumer goods, due to their adaptivity to different shapes, good rigidity and high strength-to-weight ratio [35, 54, 110, 30]. Improved fabrication techniques such as Injection and Double Vacuum Assisted Resin Transfer Molding [30] are reducing the production costs and time to manufacture. The properties of CFRP strongly depend on the processing of the material, thus detection of defects within the layers of the CFRP and their characterization are indispensable, especially for safety-relevant parts. Active thermal measurement methods [118, 106, 71] have become vital for the assessment of the quality of CFRP materials. These methods are based on the evaluation of a previously excited heat flow in the tested component and its disturbance by hidden defects. The heat flow is generated with a heat pulse or through sinusoidal modulation, observed with a TC, followed by a pixel-wise computation of the complex phase between the excitation signal and the reflected infrared signal. This phase information encodes the heat-loss within a penetration depth δ of the probed material, with depths of 1mm to 2mm typical for CFRPs.

Current state-of-the-art TCs possess resolutions of only 640×512 pixels and a noise equivalent temperature difference of 20mK. Nevertheless, these cameras
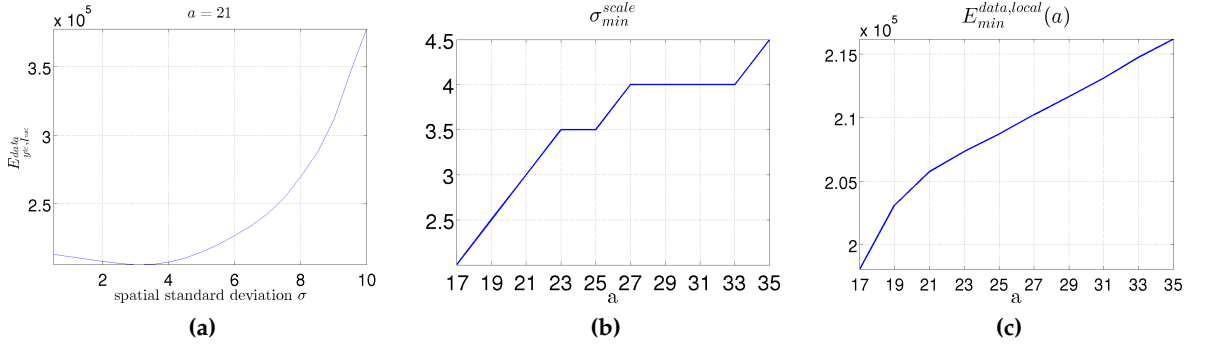
Figure 3.10.: Figure 3.10a: Plot $E_{y_{tc},I_{vsc}}^{data,l}(\sigma^{sc}, a, \mathbf{0})$ over the PSF scale difference σ^{sc} for the images y_{tc} and I_{vsc} in figure 3.9 for the window size $a = 25$. Figure 3.10b shows the minimum scale $\sigma^{sc}{}_{min}$ defined in eq. (3.59) as a function over the window size a and figure 3.10c the similarity measure $E_{min}^{data,l}$ (eq. (3.60)) over a. The minimum scale $\sigma^{sc}{}_{min}$ increases or stays constant but does not decrease for larger window sizes a. The window size $a = 21$ marks a sweet spot where $\sigma^{sc}{}_{min}(21) = \sigma^{sc,\star} = 3$ while $E_{min}^{data,l}(21)$ is comparatively minimal.

are very expensive, and the CFRP application domain requires the detection of defects at the noise limit. On the other hand, VSCs are inexpensive and easily deliver images of 10 megapixels per frame with very little noise. Thus the goal is to combine low resolution TC image y_{tc} with the high frequency information borrowed from the VSC image I_{vsc}. Since the cameras C_{tc} and C_{vsc} are a) physically separated from each other and b) have different optical resolutions we could utilize our optical flow model in eq. (3.39) to estimate the optical flow mapping between C_{tc} and C_{vsc}, thereby computing a high resolution version Y_{tc} of y_{tc}. However the similarity measure $E_{y,I}^{data}(\sigma^{sc}, \mathbf{d})$ in eq. (3.39) is based on the assumption that the cameras C_y and C_I to be registered have a linear relationship in their intensity spectrum and the images y_{tc} and I_{vsc} in figure 3.9 lack this linearity relationship. This can be seen in the joint histogram in figure 3.9c where there is no unique correspondence between the intensities of y_{tc} to those of I_{vsc}. In section 3.5 we therefore introduced the local similarity measure $E_{y,I}^{data,l}(\sigma^{sc}, a, \mathbf{d})$ which constrains the similarity assessment of the images y and I to the local square regions $\mathcal{A}_{\mathbf{x}_0}^a$ of size a centered around $\mathbf{x}_0 \in \Omega$

$$E_{y_{tc},I_{vsc,\mathbf{d}}}^{data,l}\left(\sigma^{sc}, a, \mathbf{d}\right) = \frac{1}{2} \int_\Omega \left(y_{tc} - \langle f^{\sigma^{sc},a} I_{vsc,\mathbf{d}} \rangle_{\sigma^{sc}}\right)^2 \cdot U^{\sigma^{sc},a} d^2x \quad (3.55)$$

$$U^{\sigma^{sc},a}\left(\mathbf{x}\right) = C_{\langle Y \rangle_{\sigma^{sc}}|\langle I_{\mathbf{d}} \rangle_{\sigma^{sc}}}^{\sigma^{sc},a}\left(\mathbf{x}\right) \left(C_{\langle Y \rangle_{\sigma^{sc}}|\langle I_{\mathbf{d}} \rangle_{\sigma^{sc}}}^{\sigma^{sc},a}\left(\mathbf{x}\right) + C_n\right)^{-2} \quad (3.56)$$

The problem that arises is how large to set the window size a. If it is set too small the signal to noise ratio will be too small so that not enough information of
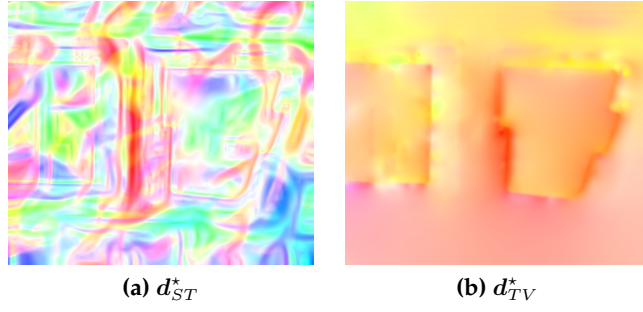
(a) $\boldsymbol{d}^\star_{ST}$ **(b)** $\boldsymbol{d}^\star_{TV}$

Figure 3.11.: Resulting optical flows of the local models E^l_{ST} ($\boldsymbol{d}^\star_{ST}$, eq. (3.44)) and E^l_{TV} ($\boldsymbol{d}^\star_{TV}$, eq. (3.45)). We can see that the structure tensor prior in the model E^l_{ST} fails to isotropically smooth the optical flow $\boldsymbol{d}^\star_{ST}$ in the regions where the images y_{tc} and I_{vsc} are predominantly homogeneous. In these regions the TV model E^l_{TV} excels due to the L_1 piecewise smoothing term in eq. (2.230).

the features in the y_{tc} and the I_{vsc} image are captured to robustly register them. On the other hand if a is set too large we eventually loose the local linearity between the y_{tc} and the I_{vsc} image. In order to find the correct value for the window size a we propose the following strategy: We first make sure that the cameras C_{vsc} and C_{tc} have the same aperture angle. This allows us to deduce the difference in optical scale, $\sigma^{scale\star}$ directly from the resolutions of the cameras. In section 3.7.4 we showed that the global similarity measure in eq. (3.27) is convex in the optical scale parameter σ^{sc}. For fixed widow sizes a the local similarity measure in eq. (3.56) is also convex in σ^{sc} but the minimum scale $\sigma^{sc}_{min}(a)$ is a function of a. The idea is to vary a such that for an optimum $a^\star$ the minimum scale $\sigma^{sc}_{min}(a^\star)$ equals the true scale difference $\sigma^{scale\star}$.

In more detail: The markers on the CFRP in figure 3.9 were used to set the zoom of the lenses such that the aperture angles of C_{tc} and C_{vsc} are approximately equal

$$\frac{w_{tc}}{f_{tc}} \approx \frac{w_{vsc}}{f_{vsc}} \qquad\qquad (3.57)$$

where $f_{tc/vsc}$ is the focal length and $w_{tc/vsc}$ the width of the CCD in the TC/ the VSC. The true difference in optical scale $\sigma^{sc,\star}$ between the cameras C_{tc} and C_{vsc} is given by the fraction of the focal lengths

$$\sigma^{sc,\star} = \frac{f_{vsc}}{f_{tc}} \approx \frac{w_{vsc}}{w_{tc}} \qquad\qquad (3.58)$$

Now the TC C_{tc} that was used to capture y_{tc} has a resolution of 640×480 and the VSC C_{vsc} is a Full HD camera (1920×1080) so that we have $\sigma^{sc,\star} \approx \frac{1920}{640} = 3$.

We make the following definitions

$$\sigma^{sc}{}_{min}(a) = \operatorname*{argmin}_{\sigma}\left(E^{data,l}_{y_{tc},I_{vsc}}\left(\sigma^{sc}, a, \mathbf{0}\right)\right) \tag{3.59}$$

$$E^{data,l}_{min}(a) = E^{data,l}_{y_{tc},I_{vsc}}\left(\sigma^{sc}{}_{min}(a), a, \mathbf{0}\right) \tag{3.60}$$

where the minimum scale $\sigma^{sc}{}_{min}(a)$ is a function of the window size a and $E^{data,l}_{min}(a)$ is the value of the similarity measure $E^{data,l}_{y_{tc},I_{vsc}}$ for that particular window size. Figure 3.10c shows the plot of $E^{data,l}_{min}$ (eq. (3.60)) over the window size a which is non-convex. Hence we cannot compute an optimal window size $a^\star$ by attempting to minimize $E^{data,l}_{min}(a)$. However $E^{data,l}_{y_{tc},I_{vsc}}$ in eq. (3.56) is convex in σ^{sc} for fixed values of a. The idea now is to find the optimal value $a^\star$ such that $\sigma^{sc}{}_{min}(a^\star) \approx \sigma^{sc,\star}$. Figure 3.10b shows $\sigma^{sc}{}_{min}(a)$ as plotted over different values of a. We can see that $\sigma^{sc}{}_{min}(a)$ increases with increasing window size a although not monotone and the true optical scale difference $\sigma^{sc}{}_{min}(a^\star) = 3$ is reached for the window size $a^\star = 21$. Thus the window size $a^\star = 21$ is the size for which $y_{tc}|_{\mathcal{A}^{a^\star}_{\mathbf{x}_0}}$ and $I_{vsc}|_{\mathcal{A}^{a^\star}_{\mathbf{x}_0}}$ are expected to be linearly dependent

$$y_{tc}|_{\mathcal{A}^{a^\star}_{\mathbf{x}_0}} \approx f \cdot \langle I_{\mathbf{d}^\star, vsc}\rangle_{\sigma^{sc}}|_{\mathcal{A}^{a^\star}_{\mathbf{x}_0}} \tag{3.61}$$

With the values $a^\star = 21$ and $\sigma^{sc,\star} = 3$ we compute the optical flow $\mathbf{d}$ for the models E^l_{ST} and E^l_{TV}

$$E^l_{ST}(\mathbf{d}) = E^{data,l}_{Y,I}(\sigma^{sc,\star}, a^\star, \mathbf{d}) + \lambda_{ST} \sum_{i=1}^{2} \widetilde{E}^{prior}_{ST}(\nabla d_i) \tag{3.62}$$

$$E^l_{TV}(\mathbf{d}) = E^{data,l}_{Y,I}(\sigma^{sc,\star}, a^\star, \mathbf{d}) + \lambda_{TV} \sum_{i=1}^{2} E^{prior}_{TV}(\nabla d_i) \tag{3.63}$$

In figure 3.11 we show the resulting optical flow $\mathbf{d}^\star_{ST}$ and $\mathbf{d}^\star_{TV}$. The model E_{TV} produces a piece-wise smooth optical flow $\mathbf{d}^\star_{TV}$ due to the piece-wise smoothing term of TV in eq. (2.230). On the other side the structure tensor model E_{ST} produces artificial motion boundaries in the regions where y_{tc} and I_{vsc} are structureless. This is due to the small weighting of the L_2 term in eq. (3.42). The value $\lambda_2 = 10^{-6}$ was chosen purely for stabilizing the structure tensor models in eq. (3.39) and eq. (3.62) in the initial iterations of the multigrid optical flow algorithm in alg. 2. However regions with structure are correctly matched by both models. To access the local linearity hypothesis in eq. (3.61) we deployed Pearson's χ^2 statistic [114].
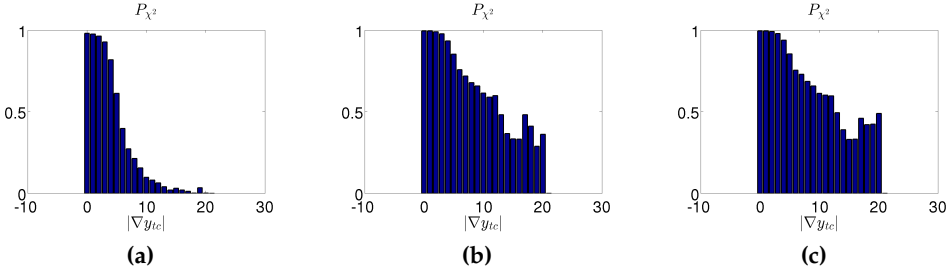
Figure 3.12.: Comparison of the p-values (eq. (3.66)) for the hypotheses (eq. (3.69)) $H_{\hat{d}=0}$ (figure 3.12a), $H_{\hat{d}=d^\star_{ST}}$ (figure 3.12b) and $H_{\hat{d}=d^\star_{TV}}$ (figure 3.12c). The p-values where computed for windows $\mathcal{A}_{x_0}$ around each pixel $x_0 \in \Omega$ and plotted over the binned values of the gradient ∇y. All three diagrams show high p-values for gradients $\nabla y \approx 0$ indicating that the structureless areas in the data in figure 3.9 obey the linear relation in eq. (3.69) regardless of the optical flow $\hat{d}$. For higher values of the gradient ∇y the hypothesis $H_{\hat{d}=0}$ in figure 3.12a fails as expected since the p-values tend to zero. The p-values at higher gradients for the hypotheses $H_{\hat{d}=d^\star_{ST}}$ (figure 3.12b) and $H_{\hat{d}=d^\star_{TV}}$ (figure 3.12c) are significantly higher then for $H_{\hat{d}=0}$ with $H_{\hat{d}=d^\star_{TV}}$ having the highest p-values meaning that the total variation model E^l_{TV} in eq. (3.45) best fulfills the linearity hypothesis in eq. (3.69).

Pearson's χ^2 statistic

Pearson's χ^2 statistic is a method of assessing whether a hypothesis H_Θ which is parameterized by the parameter set Θ is compatible a given set of observations $\hat{X}_i$, $1 \le i \le k$. . We will give a short overview of the method. Suppose the k observations $\hat{X}_i$ are realizations of the random variables X_i. For each random variable X_i we can compute an expectation value E_i given our hypothesis H_Θ. The sum V defined by

$$V = \sum_{i=1}^k \frac{(X_i - E_i)^2}{E_i} \tag{3.64}$$

is a random variable that follows the χ^2 distribution with cumulative distribution function P

$$P(\bar{V} < v) = \int_0^v f(v')dv', \quad f(v) = \frac{v^{\frac{k}{2}-1}e^{-\frac{v}{2}}}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} \tag{3.65}$$

where $\bar{V}$ is a potential value the random variable V can take. $P(\bar{V} < v)$ is the probability that $\bar{V}$ is smaller then a given number v. The probability that the

number v is smaller then any value $\bar{V}$ is computed by

$$p(v) = 1 - P(\bar{V} < v) \tag{3.66}$$

$p(v)$ is called the *p-value* of v. The method of accepting or rejecting the hypothesis H_Θ goes as follows: We can compute the observation $\hat{V}$ from the data $\hat{X}_i$ and the expectation values E_i generated by the hypothesis H_Θ

$$\hat{V} = \sum_{i=1}^{k} \frac{\left(\hat{X}_i - E_i\right)^2}{E_i} \tag{3.67}$$

If the p-value of the *particular* realization $\hat{V}$ satisfies

$$p(\hat{V}) > \alpha \tag{3.68}$$

for some given $\alpha \in [0, 1]$ then the hypothesis H_Θ is accepted as compatible with the data $\hat{X}_i$.

We want to use Pearson's χ^2 test to evaluate the optical flow results $\boldsymbol{d}_{ST}^{\star}$ and $\boldsymbol{d}_{TV}^{\star}$. The hypotheses to test are generated by the assumption that the local data term $E_{Y,I_d}^{data,l}$ in eq. (3.27) is minimal for a particular optical flow $\hat{\boldsymbol{d}}$

$$H_{\hat{\boldsymbol{d}}}: \quad y\left(\boldsymbol{x}\right) \approx \mu_{\widetilde{Y}|\widetilde{I}_{\hat{\boldsymbol{d}}}}\left(\boldsymbol{x}\right) \quad \mu_{\widetilde{Y}|\widetilde{I}_{\hat{\boldsymbol{d}}}}\left(\boldsymbol{x}\right) = \langle f_{\hat{\boldsymbol{d}}}^{\sigma^{sc,\star},a^{\star}} \cdot I_{\hat{\boldsymbol{d}}} \rangle_{\sigma^{sc}}\left(\boldsymbol{x}\right) \tag{3.69}$$

The random variables X_i are taken to be the pixels of the CCD of the TC, C_{tc}, which can record different intensities. The particular recorded image intensities $y_{tc}(\boldsymbol{x})$ are interpreted as the realizations $\hat{X}_i$ and the conditional expectation values $\mu_{\widetilde{Y}|\widetilde{I}_{\hat{\boldsymbol{d}}}}\left(\boldsymbol{x}\right)$ are the expectations E_i of the X_i. We calculate the local observations $\hat{V}(\boldsymbol{x}_0)$

$$\hat{V}_{\hat{\boldsymbol{d}}}(\boldsymbol{x}_0) = \int_{\mathcal{A}_{\boldsymbol{x}_0}} \left(y(\boldsymbol{x}) - \mu_{\widetilde{Y}|\widetilde{I}_{\hat{\boldsymbol{d}}}}\left(\boldsymbol{x}\right)\right)^2 \cdot \mu_{\widetilde{Y}|\widetilde{I}_{\hat{\boldsymbol{d}}}}^{-1}\left(\boldsymbol{x}\right) d^2x \tag{3.70}$$

and compute the p-values $p(\hat{V}_{\hat{\boldsymbol{d}}}(\boldsymbol{x}_0))$ from eq. (3.65) and eq. (3.66) which we abbreviate by $p_{\hat{\boldsymbol{d}}}(\boldsymbol{x}_0)$. In figure 3.12 we have plotted the local p-values $p_{\hat{\boldsymbol{d}}}(\boldsymbol{x}_0)$ over the gradients $\|\nabla y(\boldsymbol{x}_0)\|$ through binning by the value of the gradient for the cases $\hat{\boldsymbol{d}} = \boldsymbol{0}, \hat{\boldsymbol{d}} = \boldsymbol{d}_{TV}^{\star}$ and $\hat{\boldsymbol{d}} = \boldsymbol{d}_{ST}^{\star}$. For regions of less structure $\|\nabla y(\boldsymbol{x}_0)\| \approx 0$ all three hypothesis yield high p-values such the local linearity assumption in eq. (3.69) is valid. However the p-value $p_{\boldsymbol{d}_0}$ converges to 0 for regions with higher gradients invalidating the hypothesis $\hat{\boldsymbol{d}} = \boldsymbol{0}$ for those regions as expected since the cameras C_{vsc} and C_{tc} are not aligned. If we set the threshold for acceptance $\alpha = 0.5$ then the linearity hypotheses $H_{\boldsymbol{d}_{TV}^{\star}}$ and $H_{\boldsymbol{d}_{ST}^{\star}}$ (eq. (3.69)) are valid for regions

Joint Histogram	VSC I_{vsc}	warped VSC $I_{vsc,d}$	TC y_{tc}

(a)   (b)   (c) $d = d^\star_{ST}$ (d)

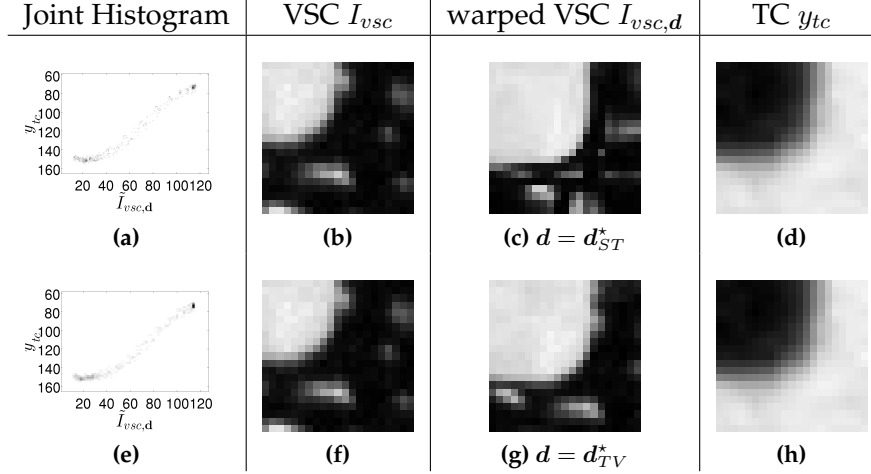(e)   (f)   (g) $d = d^\star_{TV}$ (h)

Figure 3.13.: Comparison of region of interests (ROI) of size $a^\star = 21$. Figures 3.13b and 3.13f show a ROI of I_{vsc} and 3.13d and 3.13h the corresponding ROI of the image y_{tc}. Figures 3.13c and 3.13g show figure 3.13b warped by the flows $d^\star_{ST}$ and $d^\star_{TV}$. 3.13a and 3.13e show the histograms between 3.13d and the filtered roi's $\widetilde{I}_{vsc,d} = W_{\sigma^{sc},\star} \star I_{vsc,d}$

with gradients $\|\nabla y_{tc}\| < 10$ since the corresponding p-values in figure 3.12c and figure 3.12b are exceed α. For higher valued gradients $\|\nabla y_{tc}\| > 10$ the p-value $p_{d^\star_{TV}}$ (figure 3.12c) is slightly higher than $p_{d^\star_{ST}}$ (figure 3.12b) hence the model total variation model E^l_{TV} in eq. (3.45) better fulfills the hypothesis of linearity in eq. (3.69) then the structure tensor based model E^l_{ST} in eq. (3.44) and is thus better suited for multi-modal optical flow.

Figure 3.13 shows a comparison between an ROI in the image y_{tc} (figure 3.13h) and the corresponding ROI from the image I_{vsc} (figure 3.13b) warped by $d^\star_{ST}$ (figure 3.13c) and by $d^\star_{TV}$ (figure 3.13g). The gradients within this ROI are of the order $\|\nabla y_{tc}\| \approx 10$ and hence the corresponding p-values $p_{d^\star_{TV}}$ and $p_{d^\star_{ST}}$ are in the accepted range, $p_{d^\star_{TV}} > \alpha$ and $p_{d^\star_{ST}} > \alpha$. Hence the linearity hypothesis in eq. (3.69) holds for both $\hat{d} = d^\star_{TV}$ and $\hat{d} = d^\star_{ST}$ and the histograms in figure 3.13e (hypothesis $\hat{d} = d^\star_{TV}$) and figure 3.13a (hypothesis $\hat{d} = d^\star_{ST}$) visually reflect the linear dependence of the ROI in the image y_{tc} (figure 3.13h) and the corresponding ROI from the image I_{vsc} (figure 3.13b).

(a) $\lambda_2 = 10^{-3}$ **(b)** $\lambda_2 = 10^{-6}$ **(c)** $\lambda_2 = 10^{-9}$
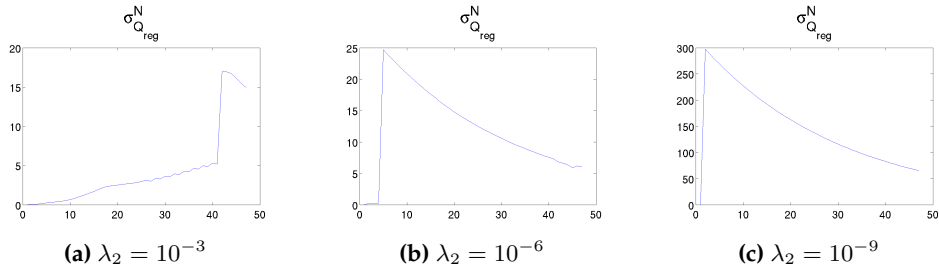
Figure 3.14.: The largest eigenvalue σ_Q^k of Q^{reg} plotted over the iterations k for three values of λ_2 in eq. (3.42). Initially we have $\sigma_Q^k \approx 8\lambda_2$ which is the eigenvalue of the L_2 term in eq. (3.42). For $\lambda_2 = 10^{-3}$ we see that σ_Q^k slowly rises for increasing iterations k until at $k \approx 40$ a sudden jump occurs and σ_Q^k begins to decrease. This is the regime where the structure tensor prior E_{ST}^{prior} begins to act an-isotropically. For smaller values of λ_2 (figures 3.14b and 3.14c) the jump occurs sooner indicating quicker an-isotropic behavior of E_{ST}^{prior}.

3.7.6. Eigenvalue analysis and the stabilization parameter λ_2

In chapter 3.3 we stated that the L_2 term in eq. (3.42) is needed to support the numerical stability of the model. We will back this statement now. Figures 3.14a, 3.14b and 3.14c show the largest eigenvalue of Q_{reg}^i, σ_Q^i at each iteration on the coarsest scale of the pyramid for different values of λ_2 in the multigrid optical flow algorithm in alg. 2. The data used was the rubber whale sequence in figure 3.3 but we found similar results for the hydrangea sequence in figure 3.4 and the CFRP data in figure 3.9. The figures all show that σ_Q^N rises to a maximum after which it decreases and converges. The initial value of σ_Q^i is of the order of λ_2 indicating that in the initial steps the L_2 term in eq. (3.42) governs the regularization. As the number of iterations increases the structure tensor determinant gets more weight, until the point where its influence over weighs that of the L_2 term. For $\lambda_2 = 10^{-9}$ (figure 3.14c) the sudden jump occurs nearly instantly. Since the influence of the L_2 is negligible, the specific decaying form of σ_Q^k in figure 3.14c is an indication that E_{ST}^{prior} is smoothing an-isotropically. On the other side Figures 3.15a, 3.15b and 3.15c show the residual vector $\boldsymbol{b}$ for different values of λ_2. Comparing the magnitude of the residual vector $\boldsymbol{b}$ in Figures 3.15a, 3.15b and 3.15c we see that for $\lambda_2 = 10^{-9}$, $\boldsymbol{b}$ is an order of magnitude larger then the other cases, which leads to longer convergence rates or numerically instable solution. This means we have a trade-off between

- $\lambda_2 \sim 10^{-3}$: Faster convergence but less influence of structure tensor (need $i > 40$ iterations for ST to act)

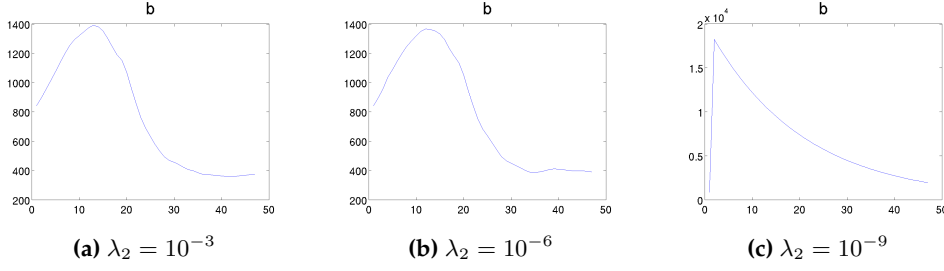- $\lambda_2 \sim 10^{-9}$: slower convergence but more influence of structure tensor

Figure 3.15.: The residual vector $\boldsymbol{b}$ plotted over the iterations k for three values of λ_2 in eq. (3.42). While the norm of $\boldsymbol{b}$ is approximately equal for $\lambda_2 = 10^{-3}$ and $\lambda_2 = 10^{-6}$, it is an order of magnitude higher for $\lambda_2 = 10^{-9}$. This indicates a numerical instability of the MOF algorithm for $\lambda_2 = 10^{-9}$

(need only $i > 1$ iterations for ST to act)

We choose $\lambda_2 = 10^{-6}$ since in this case $\boldsymbol{b}$ is of the same order of magnitude as for $\lambda_2 = 10^{-3}$ but as we see in figure 3.14b the structure tensor only needs 4 iterations until its eigenvalues over-weigh the eigenvalues of the L_2 term.

3.7.7. Summary

In this section we introduced a prior energy $E_{ST}^{prior}(\nabla \phi)$ based on the structure tensor [7, 6]. The construction was based on the principles outlined in section 2.5, namely that E_{ST}^{prior} must be invariant under the action of the group $\mathbb{G} = \mathbb{T} \times SO(2)$ in order to preserve linear level-sets of ϕ regardless of their orientation. We deployed E_{ST}^{prior} in the context of multi-modal optical flow. In Multi-modal optical flow the task is to register two images y and I recorded by the cameras C_y and C_I on a pixel-by-pixel basis with a mapping represented by the GRF $\boldsymbol{d}(\boldsymbol{x})$. Within this context a similarity measure $E_{y,I}^{data}(\sigma^{sc}, \boldsymbol{d})$ which measures how similar the image y is with the warped image $I_{\boldsymbol{d}} = I(\boldsymbol{x} + \boldsymbol{d}(\boldsymbol{x}))$ was introduced. The similarity measure $E_{y,I}^{data}$ is a cross correlation type measure with the feature that it can handle the situation in which the images y and I have differing optical resolutions. Essentially $E_{y,I}^{data}(\sigma^{sc}, \boldsymbol{d})$ measures the similarity between the low resolution image y and the filtered image $\langle I_{\boldsymbol{d}} \rangle_{\sigma^{sc}}$ and since it is a CC type measure it is minimal when

$$y(\boldsymbol{x}) \approx f \langle I_{\boldsymbol{d}^\star} \rangle_{\sigma^{sc}}(\boldsymbol{x}) + \beta \tag{3.71}$$

holds for an optimal value $\boldsymbol{d}^\star$ of the optical flow and constants f and β.

First we combined our new similarity measure $E_{y,I}^{data}$ together with our new prior

energy E_{ST}^{prior} and with the total variation prior E_{TV}^{prior} from section 2.9 to the optical flow models

$$E_{ST}^{g}\left(\boldsymbol{d}\right) = E_{Y,I}^{data}(\sigma^{scale}, \boldsymbol{d}) + \lambda_{ST}\sum_{i=1}^{2} E_{ST}^{prior}\left(\nabla d_i\right) \tag{3.72}$$

$$E_{TV}^{g}\left(\boldsymbol{d}\right) = E_{Y,I}^{data}(\sigma^{scale}, \boldsymbol{d}) + \lambda_{TV}\sum_{i=1}^{2} E_{TV}^{prior}\left(\nabla d_i\right) \tag{3.73}$$

To compare E_{ST}^{prior} with the total variation prior E_{TV}^{prior} we deployed both models in eqs. (3.72) and (3.73) on the rubber whale sequence (figure 3.3) and the hydrangea sequence (figure 3.4) from the middlebury dataset [3]. Both sequences are uni-modal ($\sigma^{sc,\star}$=0) and hence the similarity measure $E_{Y,I}^{data}$ reduces to an ordinary CC measure. Our findings is that the model E_{ST}^{g} in eq. (3.72) produces optical flows $\boldsymbol{d}_{ST}^{\star}$ with better endpoint errors (EPE) when the window size of the structure tensor σ_{ST} is sufficiently large. The best results where achieved for $\sigma_{ST} = 11$. However the TV based model in eq. (3.73) still produces an optical flow $\boldsymbol{d}_{TV}^{\star}$ with an EPE better then the best result for the model E_{ST}^{g}.

We went on to simulate a setup of co-aligned (no separating flow $\boldsymbol{d} = 0$) images y_{σ}^{test} and I^{test} which are different in optical resolution and intensity distribution. Several images y_{σ}^{test} were generated from I^{test} by inverting the intensities of I^{test} followed by Gaussian filtering with spacial standard deviations $\sigma_{test}^{sc,\star} = 1 \cdots 5$ and addition of iid noise, thus simulating the linearity relation in eq. (3.71). The goal was to find out if the similarity measure $E_{y_{\sigma}^{test}, I^{test}}^{data}(\sigma^{sc}, \boldsymbol{0})$ is capable of capturing the true scale difference $\sigma_{test}^{sc,\star}$ through variation with respect to the scale parameter σ^{sc}. We showed that $E_{y_{\sigma}^{test}, Itest}^{data}(\sigma^{sc}, \boldsymbol{0})$ is convex with respect to σ^{sc}. Furthermore the minimum is equal to the true scale difference, $\sigma^{sc,\star} = \sigma_{test}^{sc,\star}$. Thus we concluded that given the hypothesis that two images y and I with different optical resolutions are linearly dependent (eq. (3.71)), the similarity measure $E_{y,I}^{data}(\sigma^{sc}, \boldsymbol{0})$ is sensitive to the true scale difference $\sigma^{sc,\star}$.

An analysis of our optical flow method followed for a real world multi-modal setup. The setup included a camera rig containing a thermographic camera (TC) C_{tc} and a visual spectrum camera (VSC) C_{vsc}. Both cameras recorded an object made out of carbon fiber reinforced polymers (CFRP) producing the images y_{tc} and I_{vsc}, and the cameras are known to have a difference in optical scale of $\sigma^{sc,\star} = 3$. The task was to estimate the optical flow $\boldsymbol{d}^{\star}$ between y_{tc} and I_{vsc}. The problem we encountered was that y_{tc} and I_{vsc} do *not* share a global linear relationship such as eq. (3.71), since the infra-red reflectance of the CFRP can vary across different regions of equal intensity in the visual spectrum domain. Therefore we proposed a *local* CC-type similarity measure $E_{y,I}^{data,l}(\sigma^{sc}, a, \boldsymbol{d})$ which is based on

the hypothesis that y_{tc} and I_{vsc} are locally linearly dependent

$$H_{\boldsymbol{d}}^{\star} : y_{tc}|_{\mathcal{A}_{\boldsymbol{x}_0}^{a^{\star}}} \approx f \cdot \langle I_{\boldsymbol{d}^{\star},vsc} \rangle_{\sigma^{sc}} |_{\mathcal{A}_{\boldsymbol{x}_0}^{a^{\star}}} \tag{3.74}$$

where $\mathcal{A}_{\boldsymbol{x}_0}^{a^{\star}}$ is a window around any point $\boldsymbol{x}_0 \in \Omega$ with some optimal window size $a^{\star}$. We showed that the optimal scale $\boldsymbol{d}^{\star}$ of $E_{y_{tc},I_{vsc}}^{data,l}(\sigma^{sc}, a, \boldsymbol{0})$ is a function of the local window size a, $\sigma^{sc,\star} : \sigma^{sc,\star}(a)$. The value $a^{\star} = 21$ was deduced from the expectation that $\sigma^{sc,\star}(a^{\star})$ should be equal to the true optical scale difference, $\sigma^{sc,\star}(a^{\star}) = 3$. We were able to compute the optical flow between y_{tc} and I_{vsc} with the local models

$$E_{ST}(\boldsymbol{d}) = E_{y,I}^{data,l}(\sigma^{sc,\star}, a^{\star}, \boldsymbol{d}) + \lambda_{ST} E_{ST}^{prior}(\nabla \boldsymbol{d}) \tag{3.75}$$

$$E_{TV}(\boldsymbol{d}) = E_{y,I}^{data,l}(\sigma^{sc,\star}, a^{\star}, \boldsymbol{d}) + \lambda_{TV} E_{TV}^{prior}(\nabla \boldsymbol{d}) \tag{3.76}$$

To analyze the validity of the local linearity hypothesis $H_{\boldsymbol{d}}^{\star}$ in eq. (3.74) for the computed flows $\boldsymbol{d}^{\star} = \boldsymbol{d}_{ST}^{\star}$ (from eq. (3.75)) and $\boldsymbol{d}^{\star} = \boldsymbol{d}_{TV}^{\star}$ (eq. (3.76)) we deployed Pearson's χ^2 test. The p-value $p_{\boldsymbol{d}}^{\star}(\boldsymbol{x})$ in Pearson's χ^2 test is an indicator for the validity of the hypothesis $H_{\boldsymbol{d}}^{\star}$ such that $p_{\boldsymbol{d}}^{\star}(\boldsymbol{x}) = 1$ means that $H_{\boldsymbol{d}}^{\star}$ is definitely valid at the point $\boldsymbol{x}$ and $p_{\boldsymbol{d}}^{\star}(\boldsymbol{x}) = 0$ means that $H_{\boldsymbol{d}}^{\star}$ must be rejected.

The p-values $p_{\boldsymbol{d}_{ST}^{\star}}(\boldsymbol{x})$ and $p_{\boldsymbol{d}_{TV}^{\star}}(\boldsymbol{x})$ in regions with small gradients $\|y_{tc}\| < 10$ were sufficient, $p_{\boldsymbol{d}_{ST}^{\star}/\boldsymbol{d}_{TV}^{\star}} \approx 1$, to accept the linearity hypotheses $H_{\boldsymbol{d}_{ST}^{\star}}$ and $H_{\boldsymbol{d}_{TV}^{\star}}$. For higher gradients $\|y_{tc}\| > 10$ both p-values dropped off below the 50% quantile, however with $p_{\boldsymbol{d}_{TV}^{\star}}$ being slightly larger then $p_{\boldsymbol{d}_{ST}^{\star}}$. Thus although both models in eq. (3.76) and eq. (3.75) are not fully consistent with the linearity hypothesis in eq. (3.74) at the boundaries of the CFRP, the TV model in eq. (3.76) produced an optical flow $\boldsymbol{d}_{TV}^{\star}$ which is more consistent with the data y_{tc} and I_{vsc} at the boundaries then the structure tensor based model in eq. (3.75).

The good p-values in the regions $\|y_{tc}\| < 10$ do not give any information about how consistent the computed flows $\boldsymbol{d}_{ST}^{\star}$ (from eq. (3.75)) and $\boldsymbol{d}_{TV}^{\star}$ (eq. (3.76)) are themselves. This is why we visually compared $\boldsymbol{d}_{ST}^{\star}$ and $\boldsymbol{d}_{TV}^{\star}$ to each other: The flow $\boldsymbol{d}_{ST}^{\star}$ had a lot of artificial linear boundaries in regions where y_{tc} and I_{vsc} are largely constant. In contrast the flow $\boldsymbol{d}_{TV}^{\star}$ was smooth everywhere except at the physical boundary locations of the CFRP. This is due to the behavior of $E_{ST}^{prior}(\nabla \boldsymbol{d})$, namely that it only penalizes the curvature of the level-sets of $\boldsymbol{d}$. On the other hand the TV prior E_{TV}^{prior} not only penalizes the curvature of the level-sets but also enforces smoothness of the solution $\boldsymbol{d}_{TV}^{\star}$ in regions where the images y_{tc} and I_{vsc} are not discontinuous. Thus we conclude two things: First the TV prior E_{TV}^{prior} is superior to our structure tensor based prior E_{ST}^{prior} for multi-modal optical flow since the TV model in eq. (3.76) produces visually more consistent optical flows. Secondly the local linearity hypothesis in eq. (3.74)

which emerges from our proposed similarity measure $E_{y_{tc},I_{vsc}}^{data,l}(\sigma^{sc}, a, \boldsymbol{d})$ holds very well in regions where the images y_{tc} and I_{vsc} vary only gradually with no discontinuities, but must be rejected at the discontinuities of y_{tc} and I_{vsc}. This behavior is presumably due to the lack of information about the physics of the TC in the similarity measure $E_{y_{tc},I_{vsc}}^{data,l}(\sigma^{sc}, a, \boldsymbol{d})$ as we had only encoded the basic assumption that TC produces Gaussian noise (see eq. (2.263) in section 2.12). Including a more realistic noise distribution into $E_{y_{tc},I_{vsc}}^{data,l}$ will generally lead to a (possibly non-linear) relation between y_{tc} and I_{vsc} which is better suited to estimate the optical flow at the boundaries of the CFRP.

4. The Extended Least Action Algorithm

In section 2.3 we had reviewed the principle of least action (PLA) of an arbitrary Gibbs random field (GRF) theory which is modeled by the energy

$$E\left(\phi, \boldsymbol{X}_e^\Omega \phi\right) = E^{data}\left(\phi\right) + E^{prior}\left(\boldsymbol{X}_e^\Omega \phi\right) = \int_\Omega \mathcal{E}_{tot}\left(\phi\left(\boldsymbol{x}\right), \boldsymbol{X}_e^\Omega \phi\left(\boldsymbol{x}\right)\right) d^2 x \quad (4.1)$$

where the prior $E^{prior}\left(\boldsymbol{X}_e^\Omega \phi\right)$ depends on the operators $X_e^{\Omega,i} \in \mathcal{B}$ of a commutative sub algebra $\mathcal{B} \subset \mathcal{G}$ of some arbitrary Lie algebra $\mathcal{G}$. In short the PLA states that the minimizer $\phi^\star$ of the energy E in eq. (4.1) is the value of the GRF ϕ which obeys the Euler-Lagrange equations

$$\left[\mathcal{E}_{tot}\right]\left(\phi(\boldsymbol{x})\right)|_{\phi=\phi^\star} = 0, \quad \left[\mathcal{E}_{tot}\right]\left(\phi(\boldsymbol{x})\right) = \frac{\delta \mathcal{E}}{\delta \phi(\boldsymbol{x})} - \text{Div}(\boldsymbol{P}) \quad (4.2)$$

$$\text{Div}(\boldsymbol{P}) = \sum_\mu \frac{d}{dx_\mu}\left(\sum_i \omega_{\Omega,i}^\mu P_i\right), \quad P_i = \frac{\delta \mathcal{E}^{prior}}{\delta X_e^{\Omega,i}} \quad (4.3)$$

Section 2.7.2 was dedicated to the analysis of the energy $E\left(\phi, \boldsymbol{X}_e^\Omega \phi\right)$ under the action of an r-dimensional finite Lie Group $\mathbb{G} = \mathbb{G}^\phi \times \mathbb{G}^\Omega$ which contains transformations of the GRF ϕ and of the coordinate frame Ω. Noether's theorem in eq. (2.217) states that if the energy E in eq. (4.1) is invariant under the Lie group $\mathbb{G} = \mathbb{G}^\phi \times \mathbb{G}^\Omega$ and if $\mathbb{G}^\phi$ and $\mathbb{G}^\Omega$ are independent from each other, then E must be invariant under $\mathbb{G}^\Omega$. Then the canonical momentum $\boldsymbol{P}$ and hence the prior energy E^{prior} must fulfill eq. (2.222) for *any* value of the GRF ϕ

$$\widetilde{\boldsymbol{B}}\phi = 0, \quad \widetilde{\boldsymbol{B}} = \sum_i P_i\left[X_e^{\Omega,i}, X_e^{\Omega,m}\right] \quad (4.4)$$

Eq. (4.4) stems from the group $\mathbb{G}^\Omega$, the spatial transformations of the domain Ω, while the Euler-Lagrange equations in eq. (4.2) stem from the group $\mathbb{G}^\phi$, the functional transformations of the GRF ϕ. The coupling of the Euler-Lagrange differentials $[\mathcal{E}]$ and the product $\widetilde{\boldsymbol{B}}\phi$ in eq. (2.216) suggest that eq. (4.4) and thus $\mathbb{G}^\Omega$ impact the solutions space of the Euler-Lagrange equations in eq. (4.2). Indeed

(a) $\phi_0(\boldsymbol{x})$ **(b)** Ω_0 **(c)** $\widetilde{\phi}(\boldsymbol{x}) = \phi_0(T_t^B \circ \boldsymbol{x})$ **(d)** $\Omega^\epsilon = T_t^B \circ \Omega_0$
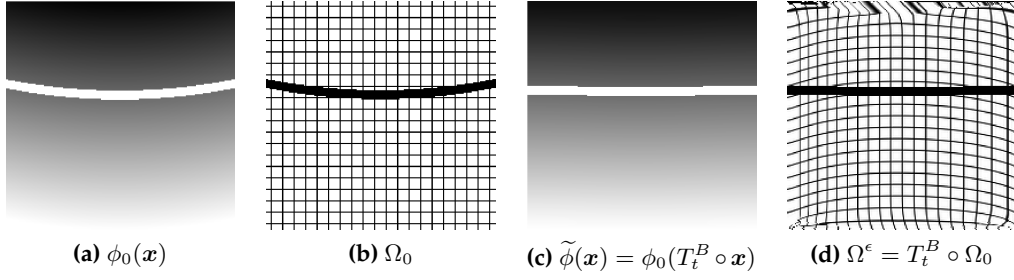
Figure 4.1.: Figure 4.1a shows an image ϕ_0 with parabolic level-sets according to eq. (4.9). The white line indicates the level-sets $S_{\phi_0,c}$ with $39 < c < 43$. In figure 4.1b the coordinate frame Ω_0 is shown together with the level-sets $S_{\phi_0,c}$. Figure 4.1c shows the warped image $\phi_0(T_t^B \circ \boldsymbol{x})$ and figure 4.1d the transformed coordinate frame $\widetilde{\Omega} = T_t^B \circ \Omega_0$. $\widetilde{\Omega}$ has been deformed by the algorithm in eq. (4.10) in such a way that the level-set $S_{\phi_0,c}$ (indicated by the black line) appears to be straight and hence it is identified with the linear domain Ω^ϵ of the TV prior $E_{TV}^{prior}(\nabla\phi)$.

we will show that there exists an optimal deformation of the coordinate frame Ω which serves as an aid to solving the Euler-Lagrange equation in eq. (4.2). The deformation algorithm will exploit the non-trivial symmetries of the energy in eq. (4.1) in order to narrow down the space of possible solutions to the Euler-Lagrange equations in eq. (4.2). We will evaluate its effectiveness for both the TV prior E_{TV}^{prior} which we reviewed in section 2.9 and the structure tensor prior E_{ST}^{prior} which we introduced in section 3.

The Basic Idea

We will start the motivation of the new algorithm in the following way: As we outlined in section 2.6 the role of the prior energy $E^{prior}\left(X_e^{\Omega,i}\phi\right)$ in eq. (4.1) is to set a constraint on the form of the level-sets of the image ϕ. By constraint we mean the particular geometrical form of the level-sets S_α^ϵ of the minimizers $\phi^\star \in A$ (A was defined in eq. (2.136)). This form is defined by the coordinate transformation $\epsilon(\boldsymbol{x})$ which obeys the Cauchy-Riemann equations in eq. (2.184). The coordinate transformation $\epsilon(\boldsymbol{x})$ simply transforms the level-set S_α^ϵ to the linear domain Ω^ϵ (the image of $\epsilon(\boldsymbol{x})$, see eq. (2.182)) where S_α^ϵ *appears* to be a line with orientation vector $\boldsymbol{\alpha}$.

Two examples of prior energies previously introduced, The structure tensor based prior energy in eq. (3.14) in section 3.3 and the total variation (TV) prior in eq. (2.229), are constructed from the Lie algebra $\mathfrak{t}$ of the translation group $\mathbb{T}$ which is spun by the Cartesian operators $\{\partial_x, \partial_y\}$. Thus the level-sets S of their

minimizers are lines (zero curvature κ (eq. (2.241))) in the Cartesian coordinate frame Ω.

Our idea is to find a method for obtaining a transformation *of the coordinate frame* Ω, T_t^B such that given *an arbitrary* image $\phi_0(\boldsymbol{x})$ the level-sets of the *warped* image $\phi_0(T_t^B \circ \boldsymbol{x})$ satisfies the geometrical constraints imposed by E^{prior}, specifically such that the level-sets

$$S = \left\{ \boldsymbol{x}' \left| \sum_i \alpha_i X_e^{\Omega,i} \phi\left(\boldsymbol{x}'\right) = 0 \right. \right\}, \quad \boldsymbol{x}' = T_t^B \circ \boldsymbol{x} \tag{4.5}$$

are linear in the linear domain Ω^ϵ of the prior $E^{prior}(X_e^{\Omega,i}\phi)$ (see figure 4.1). In general it is not possible to derive T_t^B analytically since an analytical expression for the level-sets of an arbitrary image ϕ_0 is not available. We have therefore devised a method to iteratively estimate T_t^B as the result of a flow equation on the domain Ω

$$\dot{\boldsymbol{x}}\left(t\right) = \boldsymbol{b}\left(\boldsymbol{P}, \boldsymbol{x}\left(t\right)\right) \tag{4.6}$$

where $\boldsymbol{P}$ is the canonical momentum (eq. (4.2)). Then the action of T_t^B on Ω can be understood as the integration of $\dot{\boldsymbol{x}}(t)$ in eq. (4.6) over the flow parameter t

$$T_t^B \circ \boldsymbol{x} = \int_0^T \boldsymbol{b}\left(\boldsymbol{P}, \boldsymbol{x}\left(t\right)\right) dt \tag{4.7}$$

We will later derive the explicit form of $\boldsymbol{b}$ from eq. (2.221). For now we only need to know that $\boldsymbol{b}$ only depends on the prior energy E^{prior} and *not* on the data term E^{data}. Therefore we consider the total energy $E(\phi, \nabla\phi)$ in eq. (4.1) only to consist of the TV prior E_{TV}^{prior}

$$E\left(\phi, \nabla\phi\right) = E_{TV}^{prior}\left(\nabla\phi\right) \tag{4.8}$$

What follows is a display how the flow equation in eq. (4.6) helps to (partially) minimize E_{TV}^{prior}. Consider the image $\phi_0(\boldsymbol{x})$ in figure 4.1a. It's level-sets have the shape of a parabola

$$S_{\phi_0,y_0} = \left\{ \boldsymbol{x} \left| y = (x - x_0)^2 + y_0, \quad (x,y)^T = \boldsymbol{x}, y_0 \in \mathbb{R} \right. \right\} \tag{4.9}$$

where x_0 is the x-component of the center pixel of ϕ_0 and y_0 is an offset in y-direction. The white line in figure 4.1a indicates the bundle of level-sets S_{ϕ_0,y_0} with $39 < y_0 < 43$, it is not part of the actual image. We implemented a simple algorithm that solves the integration problem in eq. (4.7) by splitting the time domain $\{0, T\}$ into N time steps t_n, $0 \le n \le N$ and iteratively computing a new

coordinate frame Ω_{n+1} given an old estimate Ω_n

$$\Omega_{n+1}: \quad \boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \tau_\Omega \boldsymbol{b}\left(\boldsymbol{P}, \boldsymbol{x}_n\right), \quad \boldsymbol{x}_n \in \Omega_n \tag{4.10}$$

where Ω_n is the transformed coordinate frame at time step t_n. The initial coordinate frame Ω_0 (figure 4.1b) is the Cartesian coordinate frame and the black line in figure 4.1b indicates the level-set S_{ϕ_0, y_0}. The simple procedure in eq. (4.10) deforms the coordinate frame Ω_0 under the influence of the prior E_{TV}^{prior} such that it assumes the form $\widetilde{\Omega} = T_t^B \circ \Omega_0$ in figure 4.1d after $N = 2500$ iterations. In figure 4.1c the image $\phi_0(T_t^B \circ \boldsymbol{x})$ is shown which is the result of transforming the original image $\phi_0(\boldsymbol{x})$ (figure 4.1a) to the frame $\widetilde{\Omega}$ in figure 4.1d. Within this frame the level-sets of ϕ_0 appear to be straight lines since the curvature κ of the level-sets of ϕ_0 are penalized by the TV prior E_{TV}^{prior}. Hence the domain $\widetilde{\Omega}$ can be identified with the linear domain Ω^ϵ of the TV prior E_{TV}^{prior}. In figure 4.1d we can see that the transformed domain $\widetilde{\Omega}$ has been curved in the opposite direction and thus $\widetilde{\Omega}$ has negative curvature $-\kappa$. Hence the curvature of ϕ_0 in figure 4.1a is not *cancelled* but merely inverted and deferred to the coordinate frame Ω.

4.0.8. Newtonian Minimization

One of the basic algorithms for solving the Euler-Lagrange equations in eq. (4.2) is the method of steepest descent (see [39] for an overview of gradient methods in image processing). The basic idea of steepest descent is to view the minimizers $\phi^\star \in A$ as the result of a flow equation driven by the Euler-Lagrange differentials $[\mathcal{E}_{tot}]$

$$\dot{\phi}(t, \boldsymbol{x}_0) = -\left[\mathcal{E}_{tot}\right]\left(\phi(t, \boldsymbol{x}_0)\right) \tag{4.11}$$

such that for $t \to t^\star$ where the limit $t^\star$ might by infinite, $\phi(t, \boldsymbol{x}_0)$ converges to $\phi^\star(\boldsymbol{x}_0)$

$$\dot{\phi}(t, \boldsymbol{x}_0)\Big|_{t \to t^\star} = 0 \implies \phi(t, \boldsymbol{x}_0)|_{t \to t^\star} = \phi^\star(\boldsymbol{x}_0) \tag{4.12}$$

In practical implementations we discretize the interval $[0, t^\star]$ into N time steps t^n and identify the GRF ϕ at the the different time steps by $\phi^n(\boldsymbol{x}_0) = \phi(t^n, \boldsymbol{x}_0)$. Starting with an initial guess ϕ^0, we compute a new estimate of the field ϕ by advancing a previous estimate ϕ^n along the negative direction of the gradient of $E(\phi, \boldsymbol{X}_e^\Omega \phi)$ which is provided by the Euler-Lagrange differentials $[\mathcal{E}_{tot}]$

$$\phi^{n+1}(\boldsymbol{x}_0) = \phi^n(\boldsymbol{x}_0) - \tau^\phi \left[\mathcal{E}_{tot}\right]\left(\phi^n(\boldsymbol{x}_0)\right) \tag{4.13}$$

The scheme is repeated (see algorithm 3, Basic Newton Algorithm (BNA)) until either the Euler-Lagrange differentials vanish or the fixed number N of iterations is reached.

Our new methodology is to combine the concept of steepest descent for the spacial coordinate frame Ω from eq. (4.6) with the concept of steepest descent for the image ϕ (in eq. (4.11))

$$\dot{\phi}(t, \boldsymbol{x}(t)) = - \left[\mathcal{E}_{tot} \right] (\phi(t, \boldsymbol{x}(t))) \tag{4.14}$$

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{b}\left(\boldsymbol{P}(t), \boldsymbol{x}(t) \right) \tag{4.15}$$

where the canonical momentum $\boldsymbol{P}$ depends on the current state $(\phi(t, \boldsymbol{x}(t)))$, hence the time dependence. The combination of the discretized eqs. (4.10) and (4.13) is straightforward

$$\phi^{n+1}(\boldsymbol{x}_n) = \phi^n(\boldsymbol{x}_n) - \tau^\phi \left[\mathcal{E}_{tot} \right] (\phi^n(\boldsymbol{x}_n)) \tag{4.16}$$

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \tau_\Omega \boldsymbol{b}\left(\boldsymbol{P}_n, \boldsymbol{x}_n \right) \tag{4.17}$$

The exact form of the vector $\boldsymbol{b}(\boldsymbol{x})$ will soon be deduced, now we wish to give an intuitive idea of $\boldsymbol{b}(\boldsymbol{x})$ should look like. For this we will discuss the influence of Noether's Theorem on the level-sets S. From this discussion it will follow that the VVF $\boldsymbol{b}(\boldsymbol{x}(t))$ is the set of tangential vectors to the level-sets of ϕ at time t.

4.0.9. The dynamics of the level-sets S

Noethers Theorem states that if the energy functional E is invariant under a Lie group $\mathbb{G}$ of dimension r, then there exists r identities relating the Euler-Lagrange differentials $[\mathcal{E}]$ (eq. (4.45)) and the divergences of r vector valued functions $\boldsymbol{W}_m$ (eq. (2.213))

$$\frac{d}{dt} g_{\gamma_t} \circ E \Big|_{t=0} = \int_\Omega \sum_m \alpha_m \left(\mathrm{div}\left(\boldsymbol{W}_m \right) + \widetilde{\omega}_m^\phi [\mathcal{E}_{tot}] \right) d^2 x = 0 \tag{4.18}$$

$$\widetilde{\omega}_m^\phi = \omega_m^\phi - X_e^{m,\Omega}(\phi) \tag{4.19}$$

for any one-parameter subgroup $g_{\gamma_t} \subset \mathbb{G}$. If $\mathbb{G}$ is a pure spacial Lie group, $\omega_m^\phi = 0$, then eq. (4.18) simplifies to

$$\frac{d}{dt} g_{\gamma_t} \circ E \Big|_{t=0} = \int_\Omega \sum_m \alpha_m \left(\mathrm{div}\left(\boldsymbol{W}_m \right) - X_e^{m,\Omega}(\phi) [\mathcal{E}_{tot}] \right) d^2 x = 0 \tag{4.20}$$
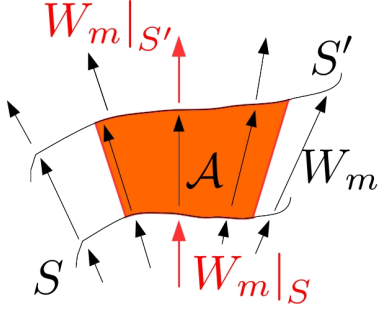
Figure 4.2.: This figure shows a transformation of the level-set S to S' along the vector $\boldsymbol{W}_m(\boldsymbol{x})$. The region $\mathcal{A} \subset \Omega$ is the region a section of S traverses as it is shifted along $\boldsymbol{W}_m$ to the end position S'. If the divergence of $\boldsymbol{W}_m$ vanishes, this means that the incoming flux of $\boldsymbol{W}_m$ equals the outgoing flux (both indicated by the red arrows), $\boldsymbol{W}_m|_S = \boldsymbol{W}_m|_{S'}$

which is independent of the integration region. Since eq. (4.20) must hold for any coefficient vector $\boldsymbol{\alpha}$, the individual divergences must satisfy

$$\int_\Omega \left(\mathrm{div}\left(\boldsymbol{W}_m \right) \right) d^2x = \int_\Omega \left(X_e^{m,\Omega}\left(\phi \right) \left[\mathcal{E}_{tot} \right] \right) d^2x \tag{4.21}$$

Thus on the set $\phi^\star \in A$ the VVFs $\boldsymbol{W}_m$ are divergence free.

Dynamics of the normal vector n_S

Eq. (4.21) must hold for any integration domain Ω which means that the integrands themselves must be equal

$$\mathrm{div}\left(\boldsymbol{W}_m \right) = X_e^{m,\Omega}\left(\phi \right) \left[\mathcal{E}_{tot} \right] \tag{4.22}$$

By Gauss' law the integrated divergence of $\boldsymbol{W}_m$ within any subset $\mathcal{A} \subset \Omega$ equals the integral of the flux of $\boldsymbol{W}_m$ over the surface $\partial\mathcal{A}$

$$\int_\mathcal{A} \mathrm{div}\left(\boldsymbol{W}_m \right) d^2x = \int_{\partial\mathcal{A}} \boldsymbol{W}_m d\boldsymbol{n}_S \tag{4.23}$$

where $\boldsymbol{n}$ is the normal vector on the surface $\partial\mathcal{A}$. Thus from eq. (4.22) we have

$$\int_{\partial\mathcal{A}} \boldsymbol{W}_m d\boldsymbol{n}_S = \int_\mathcal{A} \left(X_e^{m,\Omega}\left(\phi \right) \left[\mathcal{E}_{tot} \right] \right) d^2x \tag{4.24}$$

In figure 4.2 we have depicted the situation where a level-set S is shifted along the vector $\boldsymbol{W}_m$ with S' being the result of the shift and $\mathcal{A}$ is the region traversed by the shift of a section of S. We denote this transformation by $S \rightarrow S'$. The

boundary $\partial \mathcal{A}$ consists of two lines tangential to $\boldsymbol{W}_m$ besides the sections of S and S'. Since the flux over the tangential lines vanishes we have

$$\int_S \boldsymbol{W}_m d\boldsymbol{n}_S - \int_{S'} \boldsymbol{W}_m d\boldsymbol{n}_{S'} = \int_{\mathcal{A}} X_e^{m,\Omega}(\phi)[\mathcal{E}_{tot}] dx^2 \tag{4.25}$$

From Eq. (4.25) we see that the Euler-Lagrange differentials $[\mathcal{E}_{tot}]$ and the basis element $X_m^{m,\Omega}$ act as a *source* that drives the transformation $S \to S'$ in the sense that the level-set S propagates until it traverses a region in which the integrand of the right hand side in eq. (4.25) vanishes. More precisely eq. (4.25) can interpreted as an equation of motion for the *normal vector* on S, $\boldsymbol{n}_S$.

Dynamics of the tangential vector to S

The goal of this section is to show that eq. (4.25) does *not suffice* explain the motion of the level-sets S in the BNA (algorithm 3). We will derive an equation for the normalized *tangential* vectors of the level-sets S, which we label $\boldsymbol{b}_S(\boldsymbol{x})$. In fact the vector $\boldsymbol{b}(\boldsymbol{x}(t))$ which was introduced in eq. (4.6) happens to be equal to $\boldsymbol{b}_S$ evaluated at $\boldsymbol{x}(t)$

$$\boldsymbol{b}(\boldsymbol{x}(t)) = \boldsymbol{b}_S(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{x}(t)} \tag{4.26}$$

Since $\boldsymbol{b}_S$ is the tangential vector to S we can write the level-set equation for S in the form of an operator equation

$$B\phi = 0, \quad B = b_S^\mu(\boldsymbol{x})\partial_\mu \tag{4.27}$$

If we define B to be the basis operator of the algebra of a one parameter Lie group T_t^B (see eq. (2.157))

$$\frac{d}{dt}T_t^B \circ \phi\Big|_{t=0} = B\phi = b_S^\mu \partial_\mu \phi = 0 \tag{4.28}$$

then the image $\phi(\boldsymbol{x})$ is an invariant T_t^B. At first glance it seems that the operator B can not achieve what we set out to do in the flow equation in eq. (4.6) (and figure 4.1). And indeed we are led to believe that the energy E in eq. (4.1) is also invariant to the transformation T_t^B

$$\frac{d}{dt}T_t^B \circ E\left(\phi, \boldsymbol{X}_e^\Omega \phi\right)\Big|_{t=0} = 0 \tag{4.29}$$

since it depends on ϕ which in turn obeys eq. (4.28). Noether's theorem in eq. (2.213) holds independently of the image ϕ, hence eq. (4.29) would have to

(a)
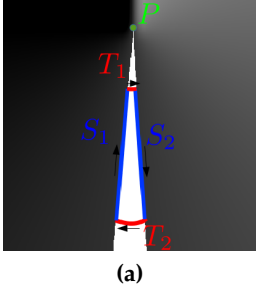
Figure 4.3.: Figure 4.3a shows an image ϕ in which a group of level-sets with $47 < \phi < 53$ (indicated by the white area) all converge into one point P at the top of the image. The line sections $S_{1,2}$ and $T_{1,2}$ enclose the region $\mathcal{R}_B$ in eq. (4.35) and eq. (4.37). The normal vectors $\boldsymbol{n}_{S_{1,2}}$ of lines $S_{1,2}$ are orthogonal to $\boldsymbol{b}_S$, hence the corresponding line integrals on the right hand side of eq. (4.35) vanish. In contrast the normal vectors $\boldsymbol{n}_{T_{1,2}}$ of lines $T_{1,2}$ are parallel to $\boldsymbol{b}_S$ so that the corresponding line integrals on the right hand side of eq. (4.35) do not vanish

hold for any image ϕ. We will show that eq. (4.29) cannot hold for at least one image ϕ if eq. (4.27) holds *uniformly* for all $\boldsymbol{x} \in \Omega$ by contradiction:

Assume eq. (4.29) to hold for T_t^B. Then by Noether's theorem there must exist a vector valued function $\boldsymbol{W}_B(\boldsymbol{x})$ for which similar to eq. (4.21) the identity

$$\int_\Omega \left(\mathrm{div}\left(\boldsymbol{W}_B \right) \right) d^2 x = \int_\Omega \left(B\left(\phi \right) \left[\mathcal{E}_{tot} \right] \right) d^2 x \tag{4.30}$$

must hold. However due to eq. (4.28) the integrand of the right hand side of eq. (4.30) vanishes uniformly on Ω, leading to the conclusion that the divergence of $\boldsymbol{W}_B(\boldsymbol{x})$ must vanish uniformly too

$$\mathrm{div}\left(\boldsymbol{W}_B \right)\left(\boldsymbol{x} \right) = 0, \quad \forall \boldsymbol{x} \in \Omega \tag{4.31}$$

The fallacy in our argumentation lies in the last statement in eq. (4.31). As we will see eq. (4.31) cannot hold: We recall the generic form of the divergence vector $\boldsymbol{W}_m$ from eq. (2.215)

$$W_m^\mu = \omega_m^\mu \mathcal{E} + \sum_i \omega_i^\mu \left(\omega_m^\phi - X_e^{m,\Omega}\left(\phi \right) \right) \frac{\delta \mathcal{E}}{\delta \left(X_e^i \phi \right)} \tag{4.32}$$

and replace $X_e^{m,\Omega}$ with B as the generating operator, thereby setting $\omega_m^\phi = 0$ (B is a purely spacial operator)

$$W_B^\mu = b_S^\mu \mathcal{E} - \sum_i \omega_i^\mu \cdot B\left(\phi \right) \frac{\delta \mathcal{E}}{\delta \left(X_e^i \phi \right)} \tag{4.33}$$

97

Again by virtue of eq. (4.28) $\boldsymbol{W}_B$ simplifies even more

$$W_B^\mu = b_S^\mu \mathcal{E} \tag{4.34}$$

To show why $\operatorname{div}(\boldsymbol{W}_B) = 0$ cannot hold we have constructed an image ϕ (figure 4.3a) in which a group of level-sets indicated by a white region converge into one point P. We construct an integral over a region $\mathcal{R}_B$ which is enclosed by two level-sets S_1 and S_2 (the blue lines in figure 4.3a) and two curves T_1 and T_2 (the red lines in figure 4.3a) connecting the level-sets

$$\int_{\mathcal{R}_B} \operatorname{div}(\boldsymbol{W}_B) d^2 x = \int_{S_1} \mathcal{E} \cdot (\boldsymbol{b}_S \cdot \boldsymbol{n}_{S_1}) \, dS_1 + \int_{S_2} \mathcal{E} \cdot (\boldsymbol{b}_S \cdot \boldsymbol{n}_{S_2}) \, dS_2$$
$$+ \int_{T_1} \mathcal{E} \cdot (\boldsymbol{b}_S \cdot \boldsymbol{n}_{T_1}) \, dT_1 + \int_{T_2} \mathcal{E} \cdot (\boldsymbol{b}_S \cdot \boldsymbol{n}_{T_2}) \, dT_2 \tag{4.35}$$

The arrows in figure 4.3a indicate the orientation of the line integral on the right hand side of eq. (4.35). Since the vector valued function $\boldsymbol{b}_S(\boldsymbol{x})$ is the tangent vector on the level-sets S_1 and S_2 the corresponding line integrals in eq. (4.35) vanish. Furthermore T_1 and T_2 have opposite orientation so we can choose the gauge

$$(\boldsymbol{b}_S|_{T_1} \cdot \boldsymbol{n}_{T_1}) = +1, \quad (\boldsymbol{b}_S|_{T_2} \cdot \boldsymbol{n}_{T_2}) = -1 \tag{4.36}$$

Thus we have for the divergence

$$\int_{\mathcal{R}_B} \operatorname{div}(\boldsymbol{W}_B) d^2 x = \int_{T_1} \mathcal{E}\left(\phi(\boldsymbol{x}), \boldsymbol{X}_e^\Omega \phi(\boldsymbol{x})\right) dT_1 - \int_{T_2} \mathcal{E}\left(\phi(\boldsymbol{x}), \boldsymbol{X}_e^\Omega \phi(\boldsymbol{x})\right) dT_2 \tag{4.37}$$

If $\operatorname{div}(\boldsymbol{W}_B) = 0$ was to be true then the two line integrals in eq. (4.37) would have to cancel. However we did not make any assumptions on the length of the curves T_i or on the distance between them so eq. (4.37) must hold for *any* configuration of the T_i. If we push T_1 to the point P where it has zero length, $\|T_1\| = 0$, then eq. (4.37) simplifies to

$$0 = -\int_{T_2} \mathcal{E}\left(\phi(\boldsymbol{x}), \boldsymbol{X}_e^\Omega \phi(\boldsymbol{x})\right) dT_2 \tag{4.38}$$

since the T_1 integral in eq. (4.37) vanishes (integrals over sets of zero measure vanish). Eq. (4.38) holds for any configuration of T_2 with non zero length. Hence we conclude that the energy density $\mathcal{E}(\phi(\boldsymbol{x}), \boldsymbol{X}_e^\Omega \phi(\boldsymbol{x}))$ must vanish on Ω. In other words: There cannot exist a non zero energy density which is invariant under T_t^B given the image ϕ in figure 4.3a. We have derived a clear contradiction. It

follows that the energy E *cannot* be an invariant of the transformation T_t^B!

$$\frac{d}{dt}T_t^B \circ E\left(\phi, \boldsymbol{X}_e^\Omega \phi\right)\Big|_{t=0} \neq 0 \tag{4.39}$$

The question that remains is that if T_t^B does not change the values of $\phi(\boldsymbol{x})$ what does it change? In eq. (4.6) we already gave a hint: T_t^B is an operator on the coordinates $\boldsymbol{x}$ themselves

$$\boldsymbol{x}(t) = T_t^B \circ \boldsymbol{x}_0 \tag{4.40}$$

The constraint in eq. (4.28) guaranties that the level-set of $\phi(\boldsymbol{x}(t))$ *move along with* the flow in eq. (4.40). The image ϕ just *appears* to be transformed if we view it from an absolute reference frame Ω_0

$$\widetilde{\phi}(\boldsymbol{x}_0) = \phi\left(T_t^B \circ \boldsymbol{x}_0\right), \quad \boldsymbol{x}_0 \in \Omega_0 \tag{4.41}$$

However the particular reference frame Ω_0 is irrelevant as two frames Ω and Ω' may always be connected by the flow in eq. (4.40).

4.1. The Extended Least Action Algorithm

In this section we will compute the exact form of the operator B. The two core aspects of the operator B as motivated in the previous section are: B must be a level-set operator (eq. (4.27)) and furthermore it must be the generator of a one parameter flow equation on the coordinate frame Ω in the sense of eq. (4.40). Thus we are looking for a level-set operator which comes from a spacial transformation. In eq. (2.206) we studied how the energy E from eq. (4.1) transforms under an arbitrary sub group $g_{\gamma(t)} \subset \mathbb{G}$

$$\frac{d}{dt}g_{\gamma_t} \circ E\Big|_{t=0} = \int_\Omega \left(\sum_i \left[V_e^\Omega, X_e^{\Omega,i}\right](\phi) \cdot P_i + v^\phi[\mathcal{E}_{tot}]\right) d^2x = 0 \tag{4.42}$$

$$= \int_\Omega \left(\sum_m \alpha_m \widetilde{\boldsymbol{B}}_m \phi + v^\phi[\mathcal{E}_{tot}]\right) d^2x = 0 \tag{4.43}$$

$$\widetilde{\boldsymbol{B}}_m = \sum_{i=1}^q P_i \left[X_e^{m,\Omega}, X_e^{i,\Omega}\right], \quad P_i\left(\boldsymbol{X}_e^\Omega \phi(\boldsymbol{x})\right) = \frac{\delta \mathcal{E}_{prior}}{\delta\left(X_e^{\Omega,i}\phi\right)} \tag{4.44}$$

As we explained in section 2.7 the variation to the field ϕ, v^ϕ and the operator $V_e^\Omega \phi$ which comes from the variation of the spacial coordinates $\boldsymbol{x}$ are taken to be independent from each other. Thus for eq. (4.43) to hold both the Euler-Lagrange

differentials $[\mathcal{E}_{tot}]$ and the product $\widetilde{\boldsymbol{B}}_m \phi$ must vanish

$$[\mathcal{E}_{tot}] (\phi^\star) = 0 \tag{4.45}$$

$$\widetilde{\boldsymbol{B}}_m \phi = 0 \tag{4.46}$$

While eq. (4.45) only holds for the minimizers $\phi^\star$, eq. (4.46) holds according to section 2.7.3 for any image ϕ if the energy E is invariant under pure spacial transformations induced by V_e^Ω. Furthermore eq. (4.46) must hold independently from the coefficient vector $\boldsymbol{\alpha}$, since eq. (4.43) must hold for any $g_{\gamma_t} \in \mathbb{G}$. We will show now that eq. (4.46) defines the level-set operator B. Since the commutator in eq. (4.43) is an element of the Lie algebra $\mathcal{G}$, by eq. (2.169) it can be represented in terms of the basis X_e^i

$$\left[X_e^{m,\Omega}, X_e^{i,\Omega} \right] = \sum_l C_{mi}^l X_e^{\Omega,l} (\phi) \tag{4.47}$$

As the basis elements $X_e^{\Omega,l}$ are represented by the Cartesian gradient operator ∇

$$X_e^{\Omega,l} = \omega_l^\mu (\boldsymbol{x}) \partial_\mu \tag{4.48}$$

the operator $\widetilde{\boldsymbol{B}}_m$ in eq. (4.43) takes the form

$$\widetilde{\boldsymbol{B}}_m = \left\| \widetilde{\boldsymbol{b}}_m \right\| \boldsymbol{B}_m \phi (\boldsymbol{x}), \quad \widetilde{b}_m^\mu (\boldsymbol{x}) = \sum_i P_i (\phi (\boldsymbol{x})) C_{mi}^l \omega_l^\mu (\boldsymbol{x}) \tag{4.49}$$

$$\boldsymbol{B}_m = b_m^\mu (\boldsymbol{x}) \partial_\mu, \quad b_m^\mu = \left\| \widetilde{\boldsymbol{b}}_m \right\|^{-1} \widetilde{b}_m^\mu \tag{4.50}$$

We would like to discuss the normalized operator $\boldsymbol{B}_m$. By eq. (4.49) and eq. (4.46) the operator $\boldsymbol{B}_m$ must satisfy the level-set equation

$$\boldsymbol{B}_m \phi = b_m^\mu \partial_\mu \phi = 0 \tag{4.51}$$

The r operators $\boldsymbol{B}_m$ are the basis of an algebra $\mathcal{A}^B$ and an associated group $\mathbb{A}^B$. For reasons soon to be clear we call the group $\mathbb{A}^B$ the *bending group* and the algebra $\mathcal{A}^B$ the *bending algebra*. The dimension r_B of $\mathcal{A}^B$ is not necessarily equal to the dimension r of the original algebra $\mathcal{G}$, since $\boldsymbol{B}_m \neq 0$ holds only for the non trivial elements of $\mathbb{G}$. For instance in the case $\mathbb{G} = \mathbb{T} \times SO(2)$ only the $SO(2)$ group allows for the construction of a bending group $SO^B(2)$ since E is trivially invariant under $\mathbb{T}$. It is the r_B non trivial operators $\boldsymbol{B}_m \neq 0$ which we use to define the diffusion equation for the coordinate frame Ω from eq. (4.40)

$$\boldsymbol{x} (t) = T_t^B \circ \boldsymbol{x}, \quad \left. \frac{d\boldsymbol{x} (t)}{dt} \right|_{t=0} = \sum_{m=1}^r \beta_m \boldsymbol{B}_m \boldsymbol{x} = \mathbf{B} \cdot \boldsymbol{x} \tag{4.52}$$

The operator $\mathbf{B}$ is a linear combination of the r operators $\boldsymbol{B}_m$ from eq. (4.50) and the coefficient vector $\boldsymbol{\beta}$ is an arbitrary constant vector. Since all $\boldsymbol{B}_m \neq 0$ are level-set operators (eq. (4.51)) the linear combination $\mathbf{B}$ is also a level-set operator. The normalization in eq. (4.50) is needed for the gauge conditions in eq. (4.36) to hold. Thus from the arguments leading to eq. (4.39) the diffusion equation in eq. (4.52) must minimize the total energy E in eq. (4.1). The combination of the flow driven by the bending operator $\mathbf{B}$ in eq. (4.52) and the flow driven by the Euler-Lagrange differentials $[\mathcal{E}_{tot}]$ in eq. (4.11)

$$\dot{\phi}(t, \boldsymbol{x}(t)) = - [\mathcal{E}_{tot}] (\phi(t, \boldsymbol{x}(t))) \tag{4.53}$$
$$\dot{\boldsymbol{x}}(t) = \mathbf{B} \cdot \boldsymbol{x}(t) \tag{4.54}$$

can be interpreted as the defining differential equations of a transformation $T(t) = (T_t^\phi, T_t^B)$ which maps the initial guess ϕ_0 to a minimizer of total energy E in eq. (4.1)

$$\phi(\boldsymbol{x}_0, t) = T_t^\phi \circ \phi_0 (\boldsymbol{x}(t)), \quad \boldsymbol{x}(t) = T_t^B \circ \boldsymbol{x}_0, \quad \boldsymbol{x}_0 \in \Omega_0 \tag{4.55}$$
$$\phi_{tot}^\star(\boldsymbol{x}_0) = \lim_{t \to \infty} \phi(\boldsymbol{x}_0, t) \tag{4.56}$$

The bending flow in eq. (4.54) *cannot* be deduced from eq. (4.53). Hence it provides an *extension* to the principle of least action.

The Curvature Operator K

The diffusion process in eq. (4.52) is a non-linear process since the coefficient vector $\boldsymbol{b}_m(\boldsymbol{x})$ itself (eq. (4.50)) is a function of the coordinates $\boldsymbol{x}(t)$ and the GRF ϕ. It is guided along those operators $\boldsymbol{B}_m$ which do not vanish due to trivial symmetries. We had motivated in figure 4.1 that the diffusion equation in eq. (4.52) bends the coordinate frame so that the level-sets of the image ϕ appear to respect the geometric constraints imposed by the prior energy E^{prior}. The geometric constraints imposed by E^{prior} are constraints on the curvature κ of the level-sets S. For instance the TV prior E_{TV}^{prior} (eq. (2.229)) penalizes level-sets with non-zero mean curvature (eq. (2.242)). The geometric constraints imposed by E^{prior} are fulfilled when the functional derivative of E^{prior}

$$\left[\mathcal{E}^{prior} \right] (\phi) = -\mathrm{Div}\, \boldsymbol{P}(\phi) \tag{4.57}$$

vanishes. In section 4.0.9 we had shown that the energy E is not invariant to the flow in eq. (4.52) and is thus minimized. However if E is minimized, then the diffusion process must also have an effect on the Euler-Lagrange differentials in eq. (4.2), especially on the divergence of the canonical momentum $\boldsymbol{P}$ in eq. (4.57).
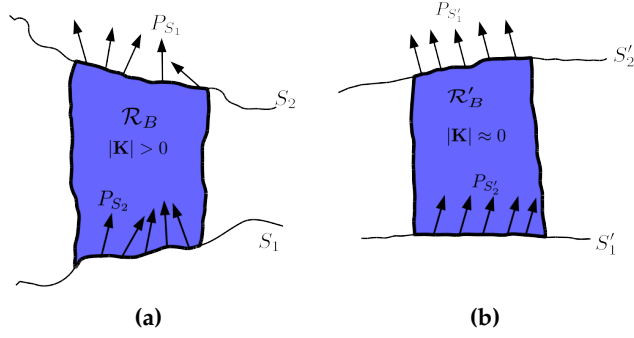
Figure 4.4.: Effect of the diffusion $\boldsymbol{x}' = T_t^B \circ \boldsymbol{x}$ (eq. (4.52)) on the canonical momentum $\boldsymbol{P}$. Figure 4.4a shows a schematic picture of a region $\mathcal{R}_B \subset \Omega$ between two level-sets S_1 and S_2. The canonical momentum (the vectors on the level-sets $S_{1,2}$) is denoted by $\boldsymbol{P}_{S_{1,2}}$. $\boldsymbol{P}$ changes its orientation when shifted along the level-sets S_1 and S_2 since the norm of the curvature operator $\boldsymbol{K}$ (eq. (4.59)) is non zero. In figure 4.4b the level-sets $S_{1,2}$ have been deformed according to $\boldsymbol{x}' = T_t^B \circ \boldsymbol{x}$ such that the canonical momentum $\boldsymbol{P}$ becomes invariant with respect to shifts along $S_{1,2}$. In this case the norm of the curvature operator $\boldsymbol{K}$ vanishes

The effect of T_t^B on $\mathrm{Div}\boldsymbol{P}$ is interesting for the reason that $\boldsymbol{P}$ has a relation to the gradient $\nabla\phi$ and thus to the level-sets S, for instance for the TV prior E_{TV}^{prior} we will show later that $\boldsymbol{P}$ is parallel to $\nabla\phi$. If we know how $\mathrm{Div}\boldsymbol{P}$ changes under the flow in eq. (4.52), we will know how $\boldsymbol{P}$ and thus the level-sets S deform under eq. (4.52).

The rate of change of $\mathrm{Div}\boldsymbol{P}$ can be computed by considering the integral

$$F = \int_\Omega \frac{d}{dt}\, T_t^B \circ \mathrm{Div}\boldsymbol{P}\,(\boldsymbol{x})\Big|_{t=0} \phi\,(\boldsymbol{x})\,\mathcal{N}d^2x \tag{4.58}$$

where $\mathcal{N}d^2x$ is the volume measure preserved under the flow in eq. (4.52). Eq. (4.58) can be transformed into an integral over an operator on ϕ (see appendix C.1)

$$F = \int_\Omega \boldsymbol{K}\phi\,(\boldsymbol{x})\,\mathcal{N}d^2x, \quad \boldsymbol{K} = \left[\mathbf{B}, P_i X_e^i\right] \tag{4.59}$$

We call the operator $\boldsymbol{K}$ the *curvature* operator for two reasons. First it represents the mixed second order derivate of the energy E, since E was first functionally derived by its argument ϕ then by the flow parameter t from eq. (4.52). Second and importantly $\boldsymbol{K}$ relates to a geometrical curvature: Since $\mathbf{B}$ is a level-set operator eq. (4.27), $\boldsymbol{K}$ describes the change the vector $\boldsymbol{P}$ undergoes when being shifted along the level-sets S. In figure 4.4 we have schematically depicted the

Algorithm 3 Basic Newton Algorithm (BNA)

Set $k = 0$
Set Initial guess ϕ^0
Compute residual $r^k = -\left[\mathcal{E}_{tot}\right]\left(\phi^k\right)$
while $\|r\| > \delta$ and $k < N$ **do**
    Compute $\phi^{k+1}\left(\boldsymbol{x}\right) = \phi^k\left(\boldsymbol{x}\right) - \tau^\phi\left[\mathcal{E}_{tot}\right]\left(\phi^k\left(\boldsymbol{x}\right)\right)$
    Recompute $r^{k+1} = -\left[\mathcal{E}_{tot}\right]\left(\phi^{k+1}\right)$
    Store $E^{k+1} = E(\phi^{k+1}, \nabla\phi^{k+1})$ in a vector $\{E^k\}$
    Set $k \rightarrow k + 1$
end while

Algorithm 4 Diffusion Algorithm (DA)

Set $k = 0$
Set Initial guess ϕ^0, $\boldsymbol{x}^0$
while $k < N$ **do**
    Compute $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - \tau^\Omega\boldsymbol{b}\left(\boldsymbol{x}^k\right)$
    Compute $\phi^{k+1}\left(\boldsymbol{x}\right) = \phi^0\left(\boldsymbol{x}^{k+1}\right)$
    Store $E^{k+1} = E(\phi^{k+1}, \nabla\phi^{k+1})$ in a vector $\{E^k\}$
    Set $k \rightarrow k + 1$
end while

Algorithm 5 Extended Least Action Algorithm (ELAA)

Set $k = 0$
Set Initial guess ϕ^0, $\boldsymbol{x}^0$
Compute residual $r^k = -\left[\mathcal{E}_{tot}\right]\left(\phi^k\right)$
while $\|r\| > \delta$ and $k < N$ **do**
    Compute $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - \tau^\Omega\boldsymbol{b}\left(\boldsymbol{x}^k\right)$
    Compute $\phi^{k+1}\left(\boldsymbol{x}^{k+1}\right) = \phi^k\left(\boldsymbol{x}^{k+1}\right) - \tau^\phi\left[\mathcal{E}_{tot}\right]\left(\phi^k\left(\boldsymbol{x}^k\right)\right)$
    Recompute $r^{k+1} = -\left[\mathcal{E}_{tot}\right]\left(\phi^{k+1}\right)$
    Store $E^{k+1} = E(\phi^{k+1}, \nabla\phi^{k+1})$ in a vector $\{E^k\}$
    Set $k \rightarrow k + 1$
end while

---

action of T_t^B on the canonical momentum $\boldsymbol{P}$. Figure 4.4a shows a region $\mathcal{R}_B \subset \Omega$ which is foliated by level-sets ranging from S_1 to S_2. The vector $\boldsymbol{P}$ varies upon shifts along the level-sets $S_{1,2}$. Hence by eq. (4.59) the norm of the curvature operator $\boldsymbol{K}$ is non-zero. In figure 4.4b the region $\mathcal{R}_B$ has been deformed to $\mathcal{R}'_B$
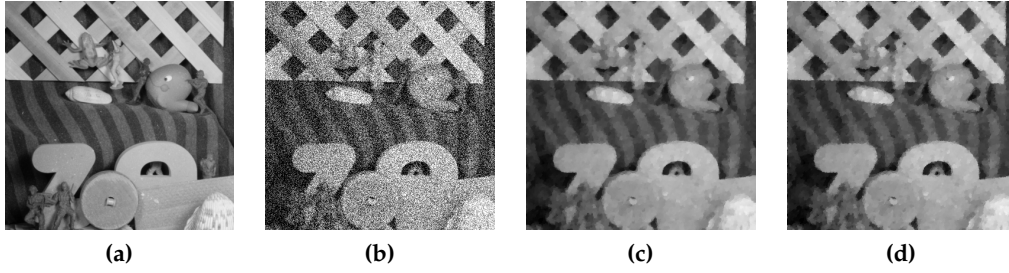
Figure 4.5.: Figure 4.5a shows a picture ϕ^c of a person. ϕ^c is taken to be free of noise. Figure 4.5b is a noise corrupted version of ϕ^c in figure 4.5a, $\phi^d = \phi^c + n$ where n is iid Gaussian noise with a standard deviation $\sigma = 100$. Figure 4.5c shows the result of the ELAA (alg. 5) and figure 4.5d the result of the BNA (alg. 3)

via the diffusion in eq. (4.52). The level-sets $S_{1,2}$ have been deformed such that $\boldsymbol{P}_{S_{1,2}}$ is constant along $S_{1,2}$. Thus the norm of the curvature operator $\boldsymbol{K}$ vanishes. This behavior of the flow in eq. (4.54) justifies the name of the group $\mathbb{A}^B$.

For prior energies E^{prior} for which the canonical momentum $\boldsymbol{P}$ is parallel to the gradient $\nabla\phi$, the vanishing of the norm of the curvature operator $\boldsymbol{K}$, $\|\boldsymbol{K}\| = 0$ means that the gradient $\nabla\phi$ itself is invariant with respect to the diffusion process eq. (4.52), hence the level-sets S of ϕ are lines.

4.1.1. Image De-noising

In section 2.4 we had described the problem of noise contamination of the image $\hat{I}^c$ of an object O recorded by the camera C. The image $\hat{I}^c$ is modeled as the sum of the projection of the object O onto the image plane of C, I_O and additive noise drawn from a distribution p

$$\hat{I}^c_{ij} = I_O\left(\boldsymbol{x}_{i,j}\right) + n \quad n \sim p\left(I^c_{ij}|I_O\left(\boldsymbol{x}_{i,j}\right)\right) \tag{4.60}$$

The problem is that we would like infer the projection I_O given the data $\hat{I}^c$ and knowledge of the distribution p in eq. (4.60). However this inference problem is ill-defined and to make it well-defined we need to consider the geometrical properties of the object O. In eq. (2.121) we reformulated the problem of inference of I_O as a minimization problem

$$I^\star_O = \text{argmin}_{I_O}\left(E_{I^c}\left(I_O\right)\right), \quad E_{I^c}\left(I_O\right) = E^{data}_{I^c}(I_O) + E^{prior}\left(\boldsymbol{X}^\Omega_e I_O\right) \tag{4.61}$$

where the image I_O is considered to be a GRF for which the geometrical neighborhood properties are specified by the prior energy E^{prior} and the noise distribution

Algorithm 6 Image de-noising analysis

    Select ϕ_0 from image database, $i = 0$, $\sigma = 100$
    while i<100 **do**
        Sample noise disturbed image $\phi = \phi_0 + n$, $n \sim \mathcal{N}(0, \sigma)$
        Run BNA, DA or ELAA and obtain vector of energies $\{E^k\}$
        Store in matrix $\mathbf{E}$, $E_{i,k} = E^k$
    end while
    For each iteration k compute the expectation value $\langle E^k \rangle$ and the standard deviation σ_{E^k} from the k-th column vector of $\mathbf{E}$

---

p in eq. (4.60) is connected to the data term $E_{I^c}^{data}$ in eq. (4.61) via the exponential mapping in eq. (2.5).

The goal of this section is to evaluated the Extended Least Action Algorithm (ELAA) (alg. 5) for the inference problem in eq. (4.61). For this we will assume the situation where the noise distribution p in eq. (4.60) is Gaussian

$$E\left(\phi\right) = \frac{1}{2} \int_\Omega |I_O - I^c|^2 d^2 x + E^{prior}\left(\boldsymbol{X}_e^\Omega I_O\right) \tag{4.62}$$

We have run both the Basic Newton (alg. 3) and our Extended Least Action algorithm (alg. 5) to minimize the energy in eq. (4.62) for the total variation prior and our new structure tensor based prior.

Analysis Method

We minimized the image denoising model in eq. (4.62) for both the TV-Prior and the structure tensor prior using the Basic Newton Algorithm (BNA) in alg. (3), the Diffusion Algorithm (DA) in alg. (4) and the Extended Least Action Algorithm (ELAA) in alg. (5). The DA runs the bending flow in eq. (4.54) alone without the Eulerian flow in eq. (4.53) on an initial image ϕ_0 in order to evaluate the effect of eq. (4.54) on the denoising model in eq. (4.61). All three algorithms were analyzed with statistical analysis algorithm (SAA) in alg. (6). The SAA samples Gaussian noise at a standard deviation of $\sigma = 100$ and adds it to the image ϕ_0. Then it runs the BNA, DA, and the ELAA. The energy E^k at each iteration of the BNA, DA and ELAA is stored in a vector. This procedure is repeated 100 times so that for each iteration k of all algorithms we have 100 sample energies E^k. Then the mean energy per iteration k, $\langle E^k \rangle$ and the standard deviation σ_{E^k} are computed. In the same manner we computed the mean $\langle \|\boldsymbol{K}\| \rangle$ and the standard deviation $\sigma_{\|\boldsymbol{K}\|}$ of the norm $\|\boldsymbol{K}\|$. The whole procedure was repeated on a total

of ten images of the middleburry data set [3]. In figure 4.5 we show the Army image of the middleburry data set together with the results of the BNA and the ELAA, and in section C.2 of the appendix we show the results of the other nine images.

Total Variation based Image De-Noising

The TV based image de-noising model is defined by the energy

$$E\left(I_O, \nabla I_O\right) = \frac{1}{2} \int_\Omega |I_O - I^c|^2 d^2x + \int_\Omega \mathcal{E}^{prior}\left(\nabla I_O\left(\boldsymbol{x}\right)\right) d^2x \tag{4.63}$$

$$\mathcal{E}^{prior}\left(\nabla I_O\left(\boldsymbol{x}\right)\right) = \lambda\sqrt{\nabla^T I_O \cdot \nabla I_O}, \quad \boldsymbol{P} = \lambda\frac{\nabla I_O}{\|\nabla I_O\|} \tag{4.64}$$

The prior $\mathcal{E}^{prior}$ in eq. (4.64) is an invariant of the Lie group $\mathbb{G} = \mathbb{T} \times SO(2)$, the group of translation and rotations. However the associated bending operator $\boldsymbol{B}_\mathbb{T}$ vanishes for the translation group $\mathbb{T}$ vanishes since

$$B_\mathbb{T}^x = P^\nu\left[\partial_\nu, \partial_x\right] = 0, \quad B_\mathbb{T}^y = P^\nu\left[\partial_\nu, \partial_y\right] = 0 \tag{4.65}$$

that is $\mathbb{T}$ is a trivial symmetry of $\mathcal{E}^{prior}$ and $E\left(I_O, \nabla I_O\right)$. The bending operator $\boldsymbol{B}_\theta$ associated with the rotation group $SO(2)$ does not vanish, but computes to

$$B_\theta = b_\theta^\mu \partial_\mu, \quad \boldsymbol{b}_\theta = \frac{\nabla^\perp I_O}{\sqrt{\nabla^T I_O \cdot \nabla I_O}} = \boldsymbol{P}^\perp \tag{4.66}$$

We have run the statistical analysis algorithm (SAA) in alg. (6) on the Army image in figure 4.5a and in figure 4.6 we have plotted the results. In figure 4.6a the mean energy $\langle E^k \rangle$ per iteration k is plotted for the BNA (dashed line), the DA (dotted line) and the ELAA (solid line). The DA which only depends on the prior E_{TV}^{prior} converges the slowest. However the ELAA, which is a combination of the DA and the BNA, converges approximately twice as fast as for the BNA and several times faster then the DA. As for the standard deviation σ_{E^k} (figure 4.6b) we see that the ELAA converges more than twice as fast then the BNA. σ_{E^k} is a measure for how robust the solution I_O^k at iteration k is with respect to noise. Thus we conclude that the ELAA is at every iteration robuster to noise then the original BNA. In figure 4.7 we show the results of the SAA for the curvature $\|\boldsymbol{K}\|$. The curvature for the DA follows an exponential behavior. This is expected since by the definition of the mean curvature κ in eq. (2.242) and the definition of the
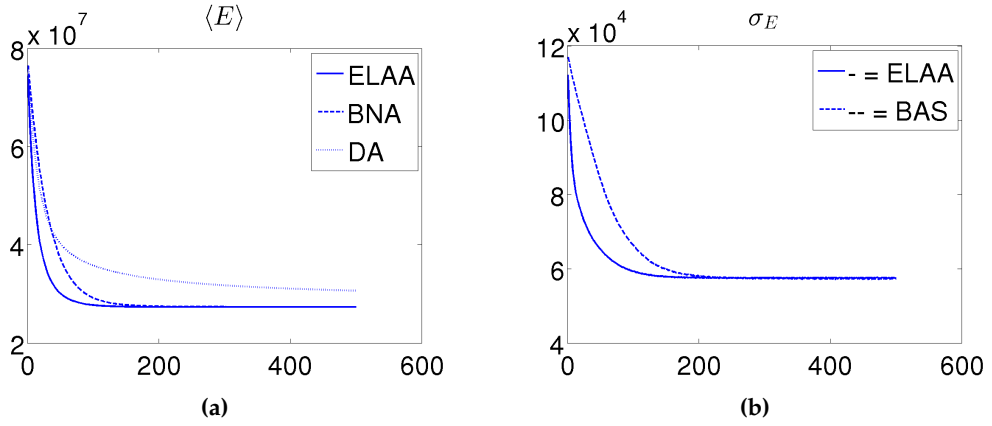
Figure 4.6.: Figure 4.6a shows the mean energy $\langle E^k \rangle$ and figure 4.6b the standard deviation σ_{E^k} per iteration k for the Army image in figure 4.5a. The the ELAA (solid line) converges about twice as fast as the BNA (dashed line) according to figure 4.6a. The standard deviation σ_{E^k} in figure 4.6b converges approximately three times faster for the ELAA then for the BNA indicating that the ELAA is robuster to noise at every iteration k

curvature operator $\boldsymbol{K}$ in eq. (4.59) we have

$$\int_\Omega \frac{d}{dt}\left(g_B^t \circ \kappa\right)\bigg|_{t=0} I_O d^2 x = \int_\Omega \boldsymbol{K} I_O d^2 x \tag{4.67}$$

The left hand side of eq. (4.67) is the rate of change of the curvature κ with respect to the parameter t of the diffusion process in eq. (4.52) and the right hand side is linear in the curvature operator $\boldsymbol{K}$. Thus the curvature κ must have an exponential dependency on t

$$g_B^t \circ \kappa = \alpha \exp\left(-\beta t\right) + \gamma \tag{4.68}$$

We did a least squares fit of the parameters (α, β, γ) in eq. (4.68) to the curvature $\|\boldsymbol{K}\|$ of the DA shown in figure 4.7a. The estimated parameters are $\alpha = 3.46 \cdot 10^5$, $\beta = 0.0044$ and $\gamma = 1.23 \cdot 10^5$. The error of the fit is of the order 10^3, which is two order of magnitude smaller then $\|\boldsymbol{K}\|$. Hence we conclude two things: first the interpretation of the curvature operator $\boldsymbol{K}$ as the curvature of the level-sets is valid. Second the evolution of the curvature of the level-sets under the diffusion process in eq. (4.52) follows an exponential law.
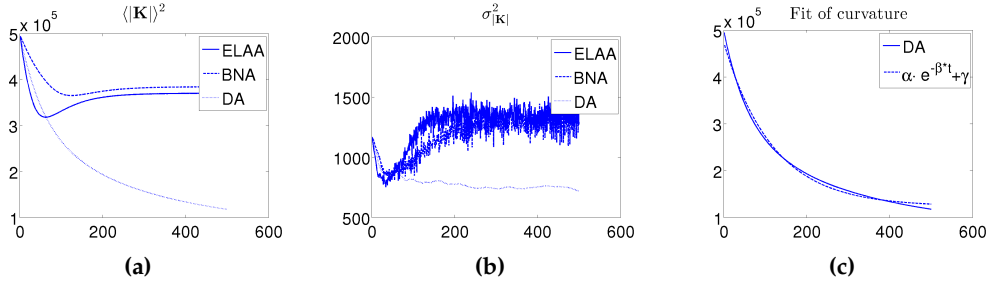
Figure 4.7.: Figure 4.7a shows the mean curvature $\langle \|\boldsymbol{K}\|^k \rangle$ and figure 4.7b the standard deviation $\sigma_{\|\boldsymbol{K}\|^k}$ per iteration k for the Army image in figure 4.5a. For the DA (dotted line), which only depends on the TV prior E_{TV}^{prior}, $\langle \|\boldsymbol{K}\| \rangle$ has an exponential decay. For the ELAA (solid line) $\langle \|\boldsymbol{K}\| \rangle$ drops faster then for the DA, until a point where the data term E^{data} prohibits further smoothing of the level-sets S. Then $\langle \|\boldsymbol{K}\| \rangle$ rises slightly and converges at a higher value. The BNA falls off slower then the ELAA and the DA and converging at a slightly higher value then the ELAA. The standard deviation $\sigma_{\|\boldsymbol{K}\|}$ is for both the ELAA and the BNA comparatively of equal order and small and two orders of magnitude smaller then $\langle \|\boldsymbol{K}\| \rangle$. When comparing the ELAA and the BNA to the DA (dotted line) we can see that the data term E^{data} has an impact on the noise distribution of the curvature $\|\boldsymbol{K}\|$ particularly at later iterations $k > 100$. Figure 4.7c shows a fit of the exponential function in eq. (4.67) to the curvature of the DA algorithm. The difference between the DA (solid line) and the fit (dashed line) is of the order 10^4, an order of magnitude smaller then $\|\boldsymbol{K}\|$

Structure Tensor Prior

In this section we applied our structure tensor prior from eq. (3.14) in section 3.2 to the image de-noising problem

$$E\left(I_O, \nabla I_O\right) = \frac{1}{q} \int_\Omega |I_O - I^c|^q d^2x + \int_\Omega \mathcal{E}_{ST}^{prior}\left(\nabla I_O\left(\boldsymbol{x}\right)\right) d^2x \qquad (4.69)$$

$$\mathcal{E}_{ST}^{prior}\left(\nabla I_O\left(\boldsymbol{x}\right)\right) = \frac{1}{2}\det\left(S\right) \qquad (4.70)$$

In order to apply the ELAA in alg. 5 to the model in eq. (4.70) we need to compute the coefficient vector $\boldsymbol{b}$ of the bending operator $\boldsymbol{B}$. From section 3.2 we know that $\mathcal{E}_{ST}^{prior}$ is invariant to the group $\mathbb{G} = \mathbb{T} \times SO(2)$. Like the TV prior the translation group $\mathbb{T}$ is a trivial symmetry so that it suffices to compute the bending operator $\boldsymbol{B}_\theta$ corresponding to the group $SO(2)$. We remember from eq. (3.17) that the structure tensor prior in eq. (4.70) transforms under the $SO(2)$ in the following way

$$\frac{d}{d\theta} E_{ST}^{prior}\left(S_\theta\right)\bigg|_{\theta=0} = \int_\Omega \text{Tr}\left(\mathbf{B}_{ST} \cdot S\right) d^2x = 0, \quad \mathbf{B}_{ST} = \mathbf{P}^{ST} \cdot \mathbf{M}_\theta \qquad (4.71)$$
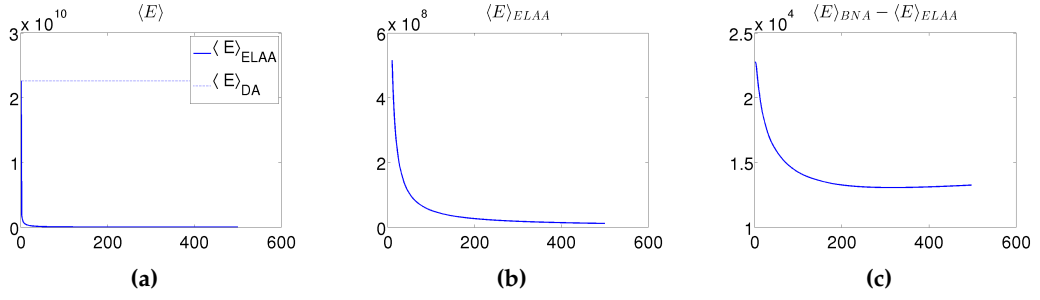
Figure 4.8.: Figure 4.8a shows the mean energy $\langle E \rangle$ as a function of the iteration k for the ELAA (solid line) and the DA (dotted line) for the structure tensor model. Figure 4.8b shows a close up of $\langle E \rangle_{ELAA}$ for $k \geq 10$ and figure 4.8c shows the difference between $\langle E \rangle_{BNA}$ and $\langle E \rangle_{ELAA}$. From figure 4.8b we can see that the mean energy for the ELAA $\langle E \rangle_{ELAA}$ is 2 orders of magnitude smaller then the mean energy for the DA and by figure 4.8c only slightly smaller then $\langle E \rangle_{BNA}$. Thus the effect of the diffusion process in eq. (4.52) on the minimization of the energy E in eq. (4.70) is at most marginal

where $\mathbf{M}_\theta$ is the Pauli matrix (the generator of the algebra $\mathfrak{so}(2)$, eq. (2.176)). The matrix $\mathbf{B}_{ST}$ has a striking similarity to the bending operator $\boldsymbol{B}_m$ in eq. (4.50) since it is a product of the canonical momentum $\mathbf{P}^{ST}$ and the structure constants of the $SO(2)$, the matrix $\mathbf{M}_\theta$. The trace

$$\mathrm{Tr}\left(\mathbf{B}_{ST} \cdot S\right) = B_{ST}^{\mu,\nu} S_{\mu,\nu} = 0 \tag{4.72}$$

is a scalar product which runs over *two* indexes (we used the Einstein summation convention) as opposed to scalar products between two vectors. Hence eq. (4.72) can be seen as the level-set equation for the structure tensor S, much like eq. (4.27) is the level-set equation for the image ϕ.

In order deploy the ELAA in alg. 5 we need to transform the level-set equation of the structure tensor S into a level-set equation like eq. (4.27) with an operator $\boldsymbol{B}_{ST}$, since the diffusion equation in eq. (4.52) necessitates a differential operator of the form in eq. (4.50).

$$\int_\Omega \mathrm{Tr}\left(\mathbf{B}_{ST} \cdot S\right) d^2 x = \int_\Omega \boldsymbol{B}_{ST} I_O d^2 x \tag{4.73}$$

If we insert the definition of the structure tensor from eq. (3.5), $S = \langle \nabla I_O \nabla^T I_O \rangle_{\sigma_{ST}}$ into the left hand side os eq. (4.73) we can shift the convolution operation in S onto the matrix $\mathbf{B}_{ST}$ and use the cyclic periodicity property of the trace to isolate
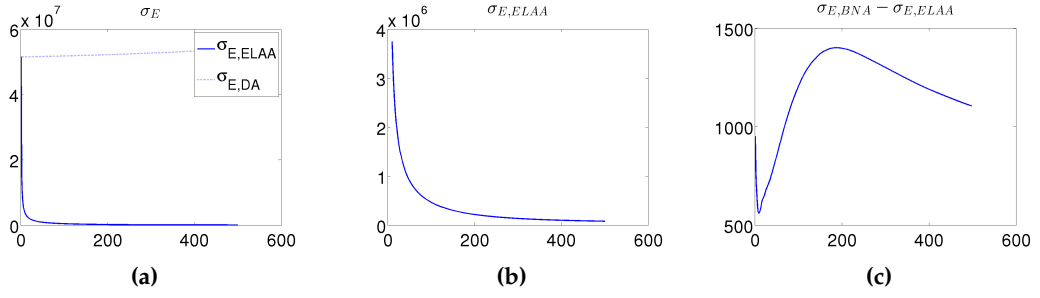
Figure 4.9.: Figure 4.9a shows the standard deviation σ_E as a function of the iteration k for the ELAA (solid line) and the DA (dotted line) for the structure tensor model. Figure 4.9b shows a close up of $\sigma_{E,ELAA}$ for $k \geq 10$ and figure 4.9c shows the difference between $\sigma_{E,BNA}$ and $\sigma_{E,ELAA}$. We essentially see the same behavior for the standard deviation σ_E as for the mean energy in figure 4.8: By figure 4.9b the standard deviation energy for the ELAA $\sigma_{E,ELAA}$ is 1 order of magnitude smaller that of the DA and by figure 4.9c only slightly smaller then $\sigma_{E,BNA}$. Hence the diffusion process eq. (4.52) has a marginal contribution to the statistical robustness of the minimizers of E in eq. (4.70)

the gradient ∇I_O

$$\int_\Omega \text{Tr} \left(\mathbf{B}_{ST} \cdot S \right) d^2 x = \int_\Omega \text{Tr} \left(\mathbf{B}_{ST} \cdot \langle \nabla I_O \nabla^T I_O \rangle_{\sigma_{ST}} \right) d^2 x$$

$$= \int_\Omega \text{Tr} \left(\langle \mathbf{B}_{ST} \rangle_{\sigma_{ST}} \cdot \nabla I_O \nabla^T I_O \right) d^2 x = \int_\Omega \nabla^T I_O \langle \mathbf{B}_{ST} \rangle_{\sigma_{ST}} \nabla I_O d^2 x \quad (4.74)$$

from eq. (4.74) we can read of the form of $\boldsymbol{B}_{ST}$

$$\boldsymbol{B}_{ST} = \frac{1}{\|\boldsymbol{b}_{ST}\|} b_{ST}^\mu \partial_\mu, \quad \boldsymbol{b}_{ST} = \langle \mathbf{B}_{ST} \rangle_{\sigma_{ST}} \nabla I_O \quad (4.75)$$

The matrix $\langle \mathbf{B}_{ST} \rangle_{\sigma_{ST}}$ in eq. (4.75) is the convolution of the elements of $\mathbf{B}_{ST}$ with the weight function $w(\boldsymbol{x})$ from the definition of the structure tensor S in eq. (3.5). Due to eq. (4.72) the operator $\boldsymbol{B}_{ST}$ is also a level-set operator in the sense of eq. (4.27).

In figure 4.8a the energies for the ELAA and the DA algorithm are shown. We can see that the energy $\langle E \rangle_{DA}$ hardly converges at the same rate as the energy $\langle E \rangle_{ELAA}$. $\langle E \rangle_{DA}$ stays within the range of $\langle E \rangle_{DA} \sim 3 \cdot 10^{10}$ while from figure 4.8b ($\langle E \rangle_{ELAA}$ for $k \geq 10$) we can see that $\langle E \rangle_{ELAA}$ drop down by 3 orders of magnitude. Figure 4.8c shows the difference between the mean energies of the BNA and the ELAA. The difference $\langle E \rangle_{BNA} - \langle E \rangle_{ELAA}$ is only in the range of 10^4 which is 4 orders of magnitude smaller than the absolute value of $\langle E \rangle_{ELAA}$ in figure 4.8b. Thus although the diffusion process in eq. (4.52) has a positive impact

(a)

(b)

(c)

(d)
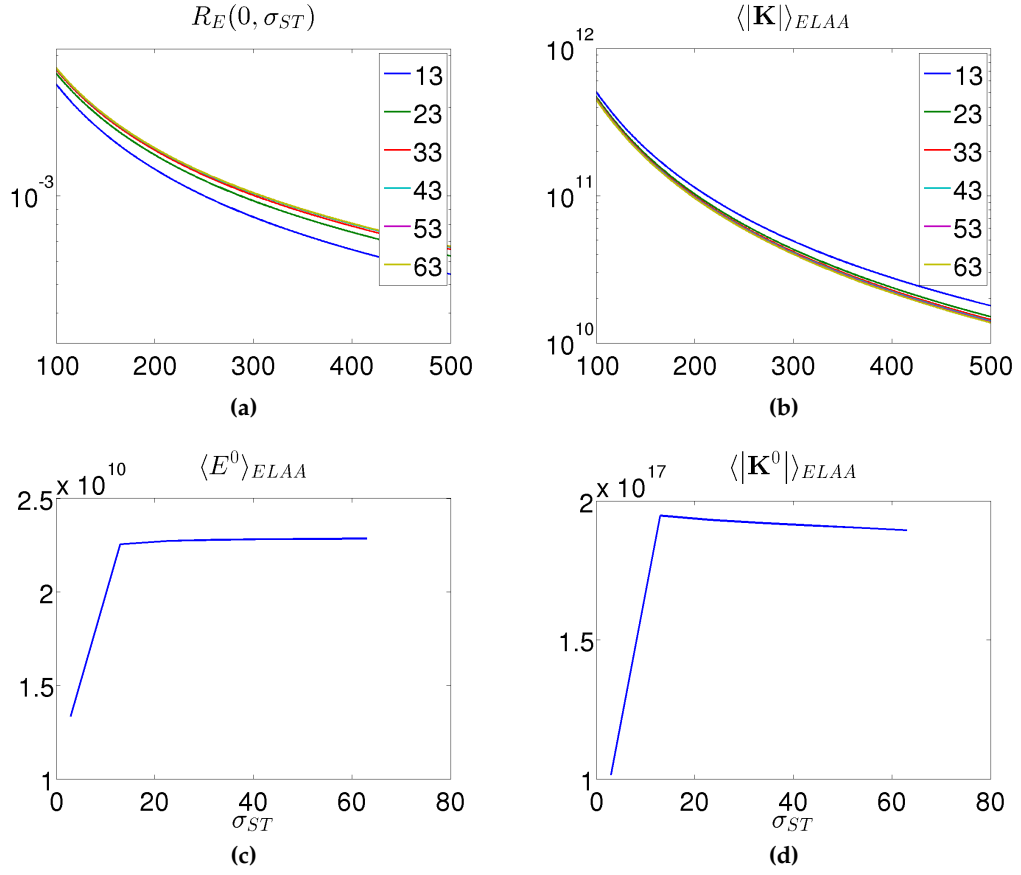
Figure 4.10.: Study of the dependency the mean energy $\langle E^k \rangle_{ELAA}$ and the mean curvature $\langle \|K\| \rangle$ on the window size σ_{ST} of the structure tensor prior E_{ST}^{prior}. Figure 4.10a shows the mean energy $\langle E^k \rangle_{ELAA}$ per iteration $k \geq 100$ for various σ_{ST} and figure 4.10b the mean curvature $\langle \|K\| \rangle$, also for various σ_{ST}. Figures 4.10c and 4.10d show the initial energy $\langle E^k \rangle_{ELAA}$ and the initial curvature $\langle \|K\| \rangle$ for $k = 0$. In figure 4.10a we can see that for smaller σ_{ST} the energy $\langle E^k \rangle_{ELAA}$ converges to lower values. Conversely for larger window sizes σ_{ST} the mean energy profiles $\langle E^k \rangle_{ELAA}$ per σ_{ST} converge. In figure 4.10b we observe a similar behavior for the curvature $\langle \|K\| \rangle$: For small σ_{ST} the curvature $\langle \|K\| \rangle$ is comparatively large. As σ_{ST} rises the profile of $\langle \|K\| \rangle$ per σ_{ST} converge, albeit at lower values. Figures 4.10c and 4.10d show that the initial energy and the initial curvature for $\sigma_{ST} = 3$ have half the values then for the larger window sizes $\sigma_{ST} = 13 \cdots 63$

on the ELAA, this impact is insignificant compared to the impact of eq. (4.52) on the TV-Denoising model (figure 4.6a). The explanation is that the structure tensor prior E_{ST}^{prior} effectively only measures the curvature *up to* the scale determined by the window size σ_{ST}. Loosely speaking, since E_{ST}^{prior} involves the weighted integral of the gradient $\nabla I_O \nabla^T I_O$ over a local neighborhood of size σ_{ST}, level-sets S with higher curvature are integrated out and hence do not contribute to the total energy. To show this we have evaluated the structure tensor model in eq. (4.70) with the ELAA for various window sizes σ_{ST} in figure 4.10. In figure 4.10c the initial energy $\langle E^k \rangle_{ELAA}$ ($k = 0$) and in figure 4.10d the initial curvature $\langle \|K\|^k \rangle$ ($k = 0$) are shown for different window sizes σ_{ST}. For window sizes $\sigma_{ST} \geq 13$ we can see that the energy $\langle E^0 \rangle_{ELAA}$ rises while the curvature $\langle \|K\|^0 \rangle$ falls. Since by eq. (4.59) the curvature operator $\boldsymbol{K}$ is proportional to the decay rate of the Euler-Lagrange differentials $[\mathcal{E}]$ we expect $\langle E^k \rangle_{ELAA}$ to converge at a slower rate for larger window sizes σ_{ST}. In figure 4.10a we have plotted the *relative* energy

$$R_E(k, \sigma_{ST}) = \frac{\langle E^k \rangle}{\langle E^0 \rangle} \tag{4.76}$$

The relative energy $R_E(k, \sigma_{ST})$ tells us how much the energy $\langle E^k \rangle$ has decayed at iteration $k > 0$ relative to the initial energy $\langle E^0 \rangle$ for a specific window size σ_{ST}. Lower values of $R_E(k, \sigma_{ST})$ indicate higher decay rates of $\langle E^k \rangle$ and vice versa. This is supported by figure 4.10b where the curvature $\langle \|K\|^k \rangle$ ($k \geq 100$) is plotted for the same window sizes σ_{ST} as for the energy $\langle E^k \rangle_{ELAA}$ in figure 4.10a. The curvature $\langle \|K\|^k \rangle$ is inverse proportional to the window size σ_{ST} at every iteration k. By eqs. (4.58) and (4.59) the decay rate of Euler-Lagrange differentials $[\mathcal{E}]$ is also inverse proportional to σ_{ST}. Thus we conclude that larger window sizes σ_{ST} have a negative impact on the convergence behavior of the ELAA. On the other hand from figure 4.10c and figure 4.10d we can see that for the smallest window size $\sigma_{ST} = 3$ the mean energy $\langle E^0 \rangle_{ELAA}$ and the curvature $\langle \|K\|^0 \rangle$ both have the smallest values. Thus again by eq. (4.59) the energy $\langle E^k \rangle_{ELAA}$ for $\sigma_{ST} = 3$ has the worst convergence behavior due to the low initial curvature $\langle \|K\|^0 \rangle$.

4.2. summary

We have introduced a new algorithm based on an extension to the principle of least action (PLA). The PLA is the idea that for any problem which can be modeled by a Gibbs random field ϕ it is possible to construct an energy functional

$$E\left(\phi, \boldsymbol{X}_e\phi\right) = E^{data}\left(\phi\right) + \lambda E^{prior}\left(\boldsymbol{X}_e\phi\right) \tag{4.77}$$

The minimizer $\phi^\star$ of the energy E is obtained by solving the Euler-Lagrange equations

$$\left[\mathcal{E}_{tot}\right]\left(\phi(\boldsymbol{x})\right)|_{\phi=\phi^\star} = 0, \quad \left[\mathcal{E}_{tot}\right]\left(\phi(\boldsymbol{x})\right) = \frac{\delta \mathcal{E}}{\delta \phi(\boldsymbol{x})} - \sum_\mu \frac{d}{dx_\mu}\left(\frac{\delta \mathcal{E}}{\partial_\mu \phi(\boldsymbol{x})}\right) \tag{4.78}$$

where $\mathcal{E}_{tot}$ is the integral density function of the energy E. Our extension to the PLA is based around a flow equation for the coordinate frame Ω

$$\boldsymbol{x}\left(t\right) = T_t^B \circ \boldsymbol{x}, \quad \left.\frac{d\boldsymbol{x}\left(t\right)}{dt}\right|_{t=0} = \mathbf{B} \cdot \boldsymbol{x} \tag{4.79}$$

We termed the operator $\mathbf{B}$ the bending operator since it bends the coordinate frame Ω such that the level-sets of the GRF ϕ obtain the geometric form preferred by the prior energy functional $E^{prior}(\boldsymbol{X}_e\phi)$. We showed that the rate of change of the Euler-Lagrange differentials $[\mathcal{E}_{tot}]$ under the flow in eq. (4.79) is equal to $\boldsymbol{K}\phi$, where the curvature operator $\boldsymbol{K}$ describes the curvature of the level-sets of ϕ. The bending operator $\boldsymbol{B}$ obeys the relation $\boldsymbol{B}\phi = 0$ for *any* value of the GRF ϕ at every time instance t of the flow in eq. (4.79) as a consequence of Noethers first theorem. Another consequence is that $\boldsymbol{B}$ and thus $\boldsymbol{K}$ is entirely determined by the prior energy E^{prior} alone.

The flow equation in eq. (4.79) was evaluated in the context of image de-noising. For image de-noising with the total variation (TV) prior from section 2.9 we showed that eq. (4.79) dramatically improves the traditional Newton method for the minimization of the energy functional E, in terms of both speed and robustness of the solution $\phi^\star$ towards noise in the initial guess ϕ^0. The functional derivative of the TV prior $\|\nabla\phi\|$ is equal to $\kappa(\boldsymbol{x}(t))$, the mean curvature of the level-set at the point $\boldsymbol{x}(t)$. We showed that $\kappa(\boldsymbol{x}(t))$ understood as a function of the flow parameter t follows an exponential behavior which is theoretically predicted since its rate of change under the flow in eq. (4.79) is equal to the curvature $\boldsymbol{K}\phi$.

The other model we tested for image de-noising deployed the structure tensor prior E_{ST}^{prior} from section 3. We found that the flow in eq. (4.79) had virtually no effect on the minimization of the energy E. It was shown that the root of the problem is that E_{ST}^{prior} depends on the structure tensor S which integrates the gradient $\nabla\phi$ (more specifically the tensor product $\nabla\phi \cdot \nabla^T\phi$) over a region $\mathcal{A}_{ST} \subset \Omega$ of size σ_{ST} around each pixel $\boldsymbol{x} \in \Omega$. We showed empirically that for

window sizes $\sigma_{ST} > 3$ the curvature $\boldsymbol{K}\phi$ (averaged over Ω) decreases as σ_{ST} increases. Since the curvature $\boldsymbol{K}\phi$ is equal to the rate of change of the Euler-Lagrange differentials, the result is that E_{ST}^{prior} and the total energy E converge at slower rates for larger window sizes σ_{ST} under the flow in eq. (4.79).

5. Conclusions

The focus of this thesis are problems in computer vision which can be modeled by Gibbs random field models (GRF). In section 2.1 we shortly reviewed the theory of GRFs, namely that a GRF is a function $\phi(\boldsymbol{x})$ for which an energy $E_Y(\phi, \boldsymbol{X}_e\phi) = E_Y^{data}(\phi) + E^{prior}(\boldsymbol{X}_e\phi)$ is defined. The GRF ϕ is a hidden variable which cannot be directly observed and the problem of inferring the optimal value of ϕ given the data Y is equal to finding the minimum $\phi_{tot}^\star = \mathrm{argmin}E_Y(\phi, \boldsymbol{X}_e\phi)$. The data term $E_Y^{data}(\phi)$ defines how the GRF ϕ is coupled to the data Y and the prior $E^{prior}(\boldsymbol{X}_e\phi)$ incorporates geometrical constraints on the level-sets of ϕ via the analytical form of the differential operators $\boldsymbol{X}_e = (X_e^1, \cdots, X_e^k)$. For instance it was shown that if the differential operators of the prior are the Cartesian derivatives, $\boldsymbol{X}_e = \nabla$, then $E^{prior}(\nabla\phi)$ penalizes level-sets of ϕ with non-zero curvature κ.

We emphasized that the prior $E^{prior}(\boldsymbol{X}_e\phi)$ must be invariant upon the action of a group of continuous smooth transformations $\mathbb{G}$, $g \circ E^{prior} = E^{prior}$ for all $g \in \mathbb{G}$. The operators $g \in \mathbb{G}$ operate on both the field $\phi(\boldsymbol{x})$ and the coordinates $\boldsymbol{x}$ themselves, $g \circ \phi(\boldsymbol{x}) = g^\phi \circ \phi(g^\Omega \circ \boldsymbol{x})$. The reason for this constraint is that the (local) minimizers $\phi_{tot}^\star \in \mathcal{A}_{tot}$ of the total energy E_Y is a trade-off between the minimizers of the data term E_Y^{data} and the minimizers of the prior energy E_Y^{prior}. The number of minimizers of the data term E_Y^{data} is usually small, since they are highly conditioned on the data Y. However since the energy E_Y has to accommodate many different values of the data Y, the prior energy $E^{prior}(\boldsymbol{X}_e\phi)$ must not penalize ϕ to few distinct minima $\phi^\star$. Instead we argued that the minimizers of E^{prior} must be a large set $\mathcal{A}$ generated by the Lie group $\mathbb{G}$, $\mathcal{A} = \{\phi^\star | \phi^\star = g \circ \phi_0^\star, \quad g \in \mathbb{G}\}$. $\phi_0^\star$ is an arbitrary minimizer of E^{prior}.

Although the prior E^{prior} is invariant under the group $\mathbb{G}$, the total energy E_Y is *generally not* invariant, due to the data term E_Y^{data}, it may be invariant under *local* transformations around the unity, $g \in U_e \subset \mathbb{G}$. In the case that $E_Y(\phi, \boldsymbol{X}_e\phi)$ is symmetric upon the action of the group U_e, $g \circ E_Y = E_Y$ for all $g \in U_e$, Emmy Noether [81, 82] discovered that there exist r vector valued functions $\boldsymbol{W}_m(\boldsymbol{x})$ for which the following relation holds

$$\mathrm{div}\,(\boldsymbol{W}_m) - X_{e,m}^\Omega(\phi)\,[\mathcal{E}] = 0 \tag{5.1}$$

where the generator $X_{e,m}$ is an element of the Lie algebra $\mathcal{G}$ corresponding to the group $\mathbb{G}$. In section 2.6 we reviewed the theory of Lie groups and in section 2.7.2 we transformed eq. (5.1) into an equation involving r differential operators $\boldsymbol{B}_m$ acting on ϕ and derived from the Lie algebra $\mathcal{G}$ of the group $\mathbb{G}$

$$\boldsymbol{B}_m \phi = 0, \quad \boldsymbol{B}_m = \sum_i \frac{\delta \mathcal{E}^{prior}}{\delta X_e^i(\phi)} \cdot \left[X_e^m, X_e^i \right] \tag{5.2}$$

Although the total energy E_Y was taken into account in eqs. (5.1) and (5.2) the $\boldsymbol{B}_m$ only depend on the prior energy density $\mathcal{E}^{prior}$ and stem from the variation of the total energy E_Y under the spatial transformations g^Ω.

In section 2.9 we reviewed the total variation (TV) based prior $E_{TV}^{prior}(\nabla\phi) = \int \|\nabla\phi\| \, d^2x$, considering its transformation properties under the 3 dimensional group $\mathbb{G} = \mathbb{T} \times SO(2)$, the group of translations and rotation on the $\mathbb{R}^2$ plane. E_{TV}^{prior} is a widely used prior since it preserves discontinuities in the GRF ϕ. It was shown that the 2 operators $\boldsymbol{B}_x$ and $\boldsymbol{B}_y$ related to the subgroup $\mathbb{T}$ vanish independently of ϕ, $\boldsymbol{B}_{x/y} = 0$. We called the symmetry under $\mathbb{T}$ a trivial symmetry. The operator $\boldsymbol{B}_\theta$ which is related to $SO(2)$ does not vanish, however it still obeys eq. (5.2), hence $SO(2)$ is a non-trivial symmetry group of E_{TV}^{prior}. It was argued that the level-sets of the minimizers $\phi^\star \in \mathcal{A}$ have zero curvature, thus the level-sets are the lines $\boldsymbol{x}(s) = \boldsymbol{x}_0 + s \cdot \boldsymbol{v}$. By the rotation invariance of E_{TV}^{prior} the orientation of $\boldsymbol{v}$ is arbitrary.

In section 3 we developed a new prior E_{ST}^{prior} based on the structure tensor in [7, 6]. The starting point was that the consideration to build the prior from the differential operator $V = \boldsymbol{v} \cdot \nabla$ with constant vector $\boldsymbol{v}$. We reviewed the structure tensor $S(\nabla\phi)$ which is a weighted sum over the orientations of $\nabla\phi$ with in a window $\mathcal{A}_{\boldsymbol{x}_0,ST}$ of size σ_{ST} around each point $\boldsymbol{x}_0$. Its eigenvector to the strongest eigenvalue is parallel to $\nabla\phi$ and the eigenvector to the weakest eigenvalue is the component vector $\boldsymbol{v}$ from our differential operator V. We proposed to use the determinant of the structure tensor as a prior, $E_{ST}^{prior}(\nabla\phi) = \text{Det}(S)$. The rational behind this proposal is that the minimizers $\phi^\star$ of E_{ST}^{prior} have level-sets which are approximately linear within the regions $\mathcal{A}_{\boldsymbol{x}_0,ST}$. Hence the prior E_{ST}^{prior} smooths the GRF ϕ along the dominant level-set within the region $\mathcal{A}_{\boldsymbol{x}_0,ST}$ at point $\boldsymbol{x}_0$. Like the TV prior E_{TV}^{prior}, our new prior E_{ST}^{prior} is invariant under the group $\mathbb{G} = \mathbb{T} \times SO(2)$ so that it agnostic towards the orientation of the level-sets of ϕ. In section 3 we evaluated the use of both priors in the context of multi-modal optical flow, the results of which are summarized in section 3.7.7.

In section 4 we proposed a new kind of algorithm for the minimization problem $\phi_{tot}^\star = \text{argmin} E_Y(\phi, \boldsymbol{X}_e \phi)$. We focused on describing the minimization problem

as a mapping $T(t)$ from some initial guess ϕ_0 to the minimizer $\phi_{tot}^\star$, $\phi_{tot}^\star = T(t) \circ \phi_0$, $t \to \infty$. Similar to the analysis of the symmetries of E_Y, we proposed splitting the transformation $T(t)$ into a part T_t^ϕ that operates on the GRF ϕ and a part T_t^B which operates only on the coordinate frame Ω

$$\phi(\boldsymbol{x}_0, t) = T_t^\phi \circ \phi_0\left(\boldsymbol{x}(t)\right), \quad \boldsymbol{x}(t) = T_t^B \circ \boldsymbol{x}_0, \quad \boldsymbol{x}_0 \in \Omega_0 \tag{5.3}$$

$$\phi_{tot}^\star(\boldsymbol{x}_0) = \lim_{t \to \infty} \phi(\boldsymbol{x}_0, t) \tag{5.4}$$

where Ω_0 is an initial reference frame. We showed that both operators T_t^ϕ and T_t^B lead to two coupled flow equations for the GRF ϕ and the coordinate system Ω

$$\left.\frac{d\phi(\boldsymbol{x}(t), t)}{dt}\right|_{t=0} = [\mathcal{E}]\left(\phi, \boldsymbol{x}(t)\right) \tag{5.5}$$

$$\left.\frac{d\boldsymbol{x}(t)}{dt}\right|_{t=0} = \mathbf{B} \cdot \boldsymbol{x}(t) \tag{5.6}$$

where the operator $\mathbf{B}$ in eq. (5.6) is a linear combination of the operators $\boldsymbol{B}_m$ in eq. (5.2). The flow equation in eq. (5.5) is the conventional flow equation used in many algorithms for solving the minimization problem $\phi_{tot}^\star = \text{argmin} E_Y(\phi, \boldsymbol{X}_e \phi)$ like steepest descent methods [39]. It is however mostly deployed on the rigid Cartesian coordinate frame Ω_0. This where the flow equation in eq. (5.6) comes into play. The rational behind eq. (5.6) is that it should *bend* the coordinate frame Ω in such a manner that the level-sets of the initial guess ϕ_0 when evaluated on the coordinates $\boldsymbol{x}(t)$, $\phi_0(\boldsymbol{x}(t))$, appear to fulfill the geometric constraints imposed by the prior energy E^{prior}.

We tested the effectiveness of the bending operator $\mathbf{B}$ in the context of image denoising for both the TV prior E_{TV}^{prior} and the structure tensor prior E_{ST}^{prior}. The flow in eq. (5.6) was run on an initial image $\phi_0(\boldsymbol{x}_0)$ which is contaminated with Gaussian noise with the individual priors. Indeed we found that both prior energies get minimized by the flow in eq. (5.6), however E_{ST}^{prior} converges significantly slower then E_{TV}^{prior}. To get a better picture we analytically computed the rate of change of the Euler-Lagrange densities for the priors under the flow in eq. (5.6) and found out that it is equal to an operator equation $\boldsymbol{K}\phi$ which we interpret as the curvature of of the level-sets of ϕ

$$\left.\frac{d\left[\mathcal{E}^{prior}\right](\phi, \boldsymbol{x}(t))}{dt}\right|_{t=0} = \boldsymbol{K}\phi \tag{5.7}$$

The curvature $\boldsymbol{K}$ in eq. (5.7) is functionally dependent on the bending operator $\mathbf{B}$. We showed that due to $\boldsymbol{B}\phi = 0$ (eq. (5.2)) eq. (5.7) holds for the Euler-Lagrange

differentials of the total energy E_Y

$$\left.\frac{d\left[\mathcal{E}\right](\phi, \boldsymbol{x}(t))}{dt}\right|_{t=0} = \boldsymbol{K}\phi \tag{5.8}$$

It was shown that the flow T_t^B in eq. (5.6) alone, that is without the action of T_t^ϕ in eq. (5.5), minimizes the curvature $\boldsymbol{K}$ and hence the total energy E_Y.

In the case of the prior E_{ST}^{prior} we found out through experimental analysis that the norm of the curvature $\boldsymbol{K}$ is inverse proportional to the window size σ_{ST} of the structure tensor. This means that E_{ST}^{prior} decreases *slower* for *larger* window sizes σ_{ST}. We concluded that due to the neighborhood nature of the structure tensor, the window $\mathcal{A}_{ST}$ acts like a drag-net that tampers the speed of the flow in eq. (5.6). Thus the result of minimizing the energy E_Y with the combined flow in eqs. (5.5) and (5.6) is only marginally better then deploying the original flow in eq. (5.5) alone.

On the other hand the image denoising model with the TV based prior E_{TV}^{prior} behaved completely differently. The energy E_Y converged approximately *twice* as fast when running the combined flow in eqs. (5.5) and (5.6) as compared to running the flow in eq. (5.5) alone! Our explanation is that E_{TV}^{prior} does not measure the curvature of the level-sets of ϕ based on local statistics like E_{ST}^{prior} does. Instead it measures the mean curvature $\kappa(\boldsymbol{x}(t))$ of a level-set at the point $\boldsymbol{x}(t)$. More precisely the mean curvature $\kappa(\boldsymbol{x})$ is the (weak) functional derivative of $E_{TV}^{prior}(\nabla\phi)$ with respect to ϕ, thus the Euler-Lagrange differential of E_{TV}^{prior}. By eq. (5.7) $\kappa(\boldsymbol{x})$ is related to the curvature operator $\boldsymbol{K}$

$$\dot{\kappa}(\boldsymbol{x}(t)) = \boldsymbol{K}\phi(\boldsymbol{x}(t)) \tag{5.9}$$

Eq. (5.9) suggests that the mean curvature $\kappa(\boldsymbol{x}(t))$ follows an exponential law, given that only the flow in eq. (5.6) is run. Indeed we were able to run the flow in eq. (5.6) and fit the observed values for κ to the exponential function $\kappa(\boldsymbol{x}(t)) = \alpha\exp(-\beta t) + \gamma$. The relative error of the fit (error divided by the value of κ) is of the order 10^{-1}. Since the mean curvature κ at a single point $\boldsymbol{x}$ can be significantly higher then the averaged curvature over a region $\mathcal{A}_{ST}$ (as is the case for the structure tensor prior E_{ST}^{prior}) the decreasing rate $\dot{\kappa}(\boldsymbol{x}(t))$ can also be high. This explains why the combined flow in eqs. (5.5) and (5.6) is faster then eq. (5.5) alone.

5.1. Outlook

The bending operator $\mathbf{B}$ in eq. (5.6) and the theory behind it was developed for GRF models with total energies $E_Y(\phi, \boldsymbol{X}_e\phi)$ which only contain first order derivatives. However Emmy Noethers first theorem covers models with derivatives of any order. Thus future research should provide the advancement of the proposed theory to higher order derivative models $E_Y(\phi, \boldsymbol{X}_e^k\phi)$, $k > 1$ along the same lines discussed in section 4. Such higher order models allow for constraints on the second or higher order derivative of the GRF ϕ and are thus less restrictive then first order models. Due to the validity Emmy Noethers first for all higher order models, we can expect the same positive results of a higher order bending algebra on the minimization of $E_Y(\phi, \boldsymbol{X}_e^k\phi)$.

Another option for future research is the advancement of the bending algebra theory to Emmy Noethers second theorem which handles the case of *infinite dimensional* Lie groups $\mathbb{G}_\infty$. As briefly explained in section 2.7.1 the Lie algebra $\mathcal{G}_\infty$ of the group $\mathbb{G}_\infty$ is loosely speaking not globally constant but a function on the coordinate frame Ω, the structure 'constants' C_{mn}^l of $\mathcal{G}_\infty$ are actually functions $C_{mn}^l := C_{mn}^l(\boldsymbol{x})$. Models based on the infinite dimensional Lie group $\mathbb{G}_\infty$ are called *gauge* theories. Virtually all models in contemporary physics are gauge theories, from the quantum field theories of the standard model [88] to general relativity [73], Loop Quantum Gravity [98] and String Theory [4]. The development of the extended principle of least action in section 4 to a more general principle of least action encoding $\mathbb{G}_\infty$ might be beneficial not only for the computational algorithms used in the aforementioned gauge theories but may also prove to provide a better understanding the theories themselves.

A. Smooth Manifolds

In section 2.6 we stated that a Lie group is a set G whose elements satisfy certain multiplication rules and which is also a smooth manifold. In this section we want to define what a smooth manifold is. In a nutshell a smooth manifold is a space M whose structure is such that for every point $p \in M$ a tangential space $T_p M$ is defined. The elements $V_p \in T_p M$ at each point $p \in M$ are vectors tangential to M at p. We will show that there exist smooth mappings $V : p \mapsto V_p$ called vector fields which are associated to curves $\theta_p^V(t)$ on M called integral curves, such that V_p is the derivative of the curve $\theta_p^V(t)$ at $t = 0$. The fundamental theorem for flows states that for every smooth vector field V there exist a unique integral curve $\theta_p^V(t)$ at every point $p \in M$. At the end of the section we will introduce the Lie derivative $\mathcal{L}_V W$ which is equal to the rate of change a smooth vector field W undergoes when shifted along the integral curves of V. The Lie derivative $\mathcal{L}_V W$ is central to our modern version of Noethers theorem in section 2.7.2 where we compute the transformation of the energy functional $E(\phi, \mathbf{X}_e^\Omega \phi)$ with respect to an arbitrary Lie group $\mathbb{G}^\Omega$ of spacial transformations since with it we can compute the rate of change of the gradient $\mathbf{X}_e^\Omega \phi$ under the action of $\mathbb{G}^\Omega$.

We will now define what a smooth manifold is by defining what a topological manifold is followed by additional criteria which make it a smooth manifold following [63].

A.0.1. Topological Spaces

A topological manifold has a structure of open subsets called a topological space

Definition 11 (Topological Space). *A set X is a topological space if*

- *X and $\emptyset$ are open*

- *the union of any collection of open subsets $U \subset X$ is open*

- *the intersection of any finite collection of open subsets is open*

The collection of all open subsets of X is called the topology of X.

The following definitions specify the notion of a pre-compact space. A countable collection of Pre-compact spaces serves as a basis for a topological and thus a smooth manifold.

Definition 12 (Compact Space). *A topological space X is compact if every open cover of X has a finite sub-cover*

Definition 13 (Closure). *Let X be a topological space and $S \subset X$ be a sub-space. The closure $\bar{S}$ of S is the intersection of all closed subsets of X containing S.*

Definition 14 (Pre-compact Space). *Let X be a topological space. A subset $S \subset X$ is pre-compact in X if its closure $\bar{S}$ is compact.*

We furthermore need a notion of countability in order to construct the basis of a topological space as a countable set of pre-compact balls.

Definition 15 (Countable Set). *A set A is countable if either*

- *A is finite*

- *A admits a bijection to the natural numbers. In this case A is said to be countable infinite*

An example of a countable infinite set is $\mathbb{Z}$, the set of rational numbers.

Definition 16 (Homeomorphic Map). *Let X and Y be two topological spaces. A homeomorphic map $\psi : X \rightarrow Y$ is a map for which the image $\psi(S) \subset Y$ of an open subset $S \subset X$ is open in Y and the pre-image $\psi^{-1}(V)$ of an open subset $V \subset Y$ is open in X. It follows that both ψ and ψ^{-1} are continuous.*

Definition 17 (Topological Manifold). *A topological manifold M of dimension n is a topological space which satisfies*

- *M is Hausdorff: for all $p, q \in M$ there exist open subsets $U_p, U_q \subset M$ such that $U_p \cap U_q = \emptyset$*

- *M is second countable: There exists a countable collection of open subsets $\mathcal{U} = \{U_i \, | \, U_i \subset M, U_i \text{ open}, 1 \leq i \leq \infty\}$ such that any subset $B \subset M$ is covered by a the union of a sub collection $\mathcal{B} \subset \mathcal{U}$. the collection $\mathcal{U}$ is also called the basis for the topology of M*

- *M is locally Euclidean: for all $p \in M$ there exist a pair (U_p, ψ_p) with the open subset $U_p \subset M$ and the homeomorphic map $\psi_p : U_p \rightarrow \widetilde{U}_p \subset \mathbb{R}^n$*

The pair (U_p, ψ_p) is called a *local chart* of M.

The next lemma states the existence of a countable basis of compact ball which cover every topological space.

Lemma 8 (Countable Basis of a topological Manifold). *Every topological manifold has a countable basis of pre-compact coordinate balls.*

Proof. Let M be a topological manifold. We will first assume the special case that a global chart (M, ψ) exists with $\psi(M) = \widetilde{U} \subset \mathbb{R}^n$. We define a collection of open pre-compact balls $\widetilde{B}_r(\boldsymbol{x}) \subset \widetilde{U}$ of rational radii r located at the rational locations $\boldsymbol{x} \in \mathbb{Z}^n \cap \widetilde{U}$

$$\mathcal{B} = \left\{ \widetilde{B}_r(\boldsymbol{x}) \,\Big|\, r \in \mathbb{Z}, \quad \boldsymbol{x} \in \mathbb{Z}^n \cap \widetilde{U}, \quad \bar{B}_r \subset \widetilde{U} \; open \right\} \tag{A.1}$$

where $\bar{B}_r$ is the closure of $\widetilde{B}_r$. Since $\mathbb{Z}$ and $\mathbb{Z}^n \cap \widetilde{U}$ are both countable sets the collection $\mathcal{B}$ is a countable basis for $\widetilde{U}$. Since ψ is homeomorphic the sets $B_r = \psi^{-1}(\widetilde{B}_r)|_{\widetilde{B}_r \in \mathcal{B}}$ are open and pre-compact in M and thus serve as a countable basis for M.

In the general case a global chart for M does not necessarily exist. However since M is a topological manifold there exist a countable collection of local charts is (U_p, ψ_p) [63]. By the preceding argument the set $U_p \subset M$ must have a countable basis. The collection of countable basis for all subsets $U_p, p \in M$ is itself a countable basis for M of pre-compact balls. $\qquad\square$

A topological space M is basically a space where two points p and q are allowed to be arbitrarily close without meeting each other. In this sense the points in a topological space M lie dense within M.

Even though the points in M may be continuously distributed in M, M itself can be a union of otherwise disjoint topological spaces. In this case M is said to be not connected. In the following we will define different levels of connectivity. The special case of path connectedness plays a role in the context of smooth manifolds in the next subsection.

Definition 18 (Connectivity). *A set X is said to be:*

- *connected: if there do not exist two disjoint open subsets of X whose union is X*

- *path connected: if every pair of points in X can be connected by a path in X*

- *locally path connected: if X has a basis of path connected open sets*

Proposition 3 (Connectivity of a Manifold). *Let M be a topological manifold*

1. *M is locally path connected*

2. *M is connected if and only if it is path connected*

Proof. By Lemma 8 M has a countable basis of pre-compact balls $B_r \subset M$. Each ball B_r is path connected hence M is locally path connected.

To prove 2 we only need to show that M is automatically path connected if it is connected. If M is connected then every pair of subsets U and V satisfying $M = U \cup V$ must also satisfy $N = U \cap V \neq \emptyset$.

We will first assume that both U and V are path connected. If N is not path connected itself, it must be at least locally path connected. Since by lemma 8 M has a basis $\mathcal{B}$ of pre-compact balls B_r a finite sub collection $\mathcal{B}_N$ may be chosen to equal N

$$N = \bigcup_{B_i \in \mathcal{B}_N} B_i \tag{A.2}$$

Since each of the balls $B_i \in \mathcal{B}_N$ are path connected two arbitrary points $p \in U$ and $q \in V$ may be connected with a path going through any of the balls $B_i \in \mathcal{B}_N$. Hence $M = U \cup V$ is path connected. The assumption that U and V are path connected may be liberated to U and V being merely connected since the preceding argument may be iterated until both $U, V \in \mathcal{B}$ for which the case 2 holds for the union $U \cup V$ trivially. $\square$

If M is a path-connected topological space then we can imagine it foliated by integral curves $\theta_p : \mathbb{R} \to M$ (the term *integral* curve will soon be explained) with arbitrary starting points $p \in M$. However on a topological space such curves need not be differentiable and thus need not exist. In the next section we will extend the topological manifold to a smooth manifold. Smooth manifolds admit a structure that always allow the existence of a foliation of M by integral curves.

A.0.2. Smooth Manifolds

Now that we have defined what a topological space M is, we continue to add to it a structure called a smooth atlas thus making it a smooth manifold.

Definition 19 (Smoothly Compatible Charts). *Let M be a topological manifold. Two charts (ψ_i, U_i) and (ψ_j, U_j) of M are said to be smoothly compatible if either $U_i \cap U_j = \emptyset$ or the map $\psi_i \circ \psi_j^{-1}$ is a diffeomorphism. If $\psi_i \circ \psi_j^{-1}$ is a diffeomorphism it follows that both ψ_i and ψ_j are diffeomorphisms.*

Definition 20 (Atlas and Smooth Manifold). *Let M be a topological manifold. An atlas $\mathcal{A}$ is a set of charts (ψ_p, U_p), $p \in M$ which are smoothly compatible with each other. The atlas $\mathcal{A}$ is called maximal if any chart (ρ, V), $V \subset M$ which is smoothly compatible*

with all $(\psi_p, U_p) \in \mathcal{A}$ is also contained in $\mathcal{A}$, $(\rho, V) \in \mathcal{A}$. The pair $(M, \mathcal{A})$ is called a smooth manifold.

When the smooth atlas $\mathcal{A}$ on M is understood, we will drop $\mathcal{A}$ from $(M, \mathcal{A})$ and call M alone a smooth manifold. The following lemma shows that the subsets $U_p \subset M$ of all charts (U_p, ψ_p) are spun by smooth coordinate functions $\xi_p : \mathbb{R}^n \to M$

Lemma 9 (Coordinate functions). *Let M be a smooth manifold and (U_p, ψ_p), $p \in M$ be any smooth chart on M. There exists a smooth map*

$$\xi_p : \psi(U_p) \subset \mathbb{R}^n \to M \quad \xi_p(\boldsymbol{x}) = (\xi_p^1(\boldsymbol{x}), \cdots, \xi_p^n(\boldsymbol{x}))^T \tag{A.3}$$

such that $\xi_p(\psi_p(U_p)) = U_p$. The component functions ξ_p^i of ξ_p are called the local coordinates of U_p.

Proof. Since the map ψ_p of any given chart (U_p, ψ_p) on M is a diffeomorphism, its inverse

$$\psi_p^{-1} : \mathbb{R}^n \to U_p \tag{A.4}$$

is smooth on the subset $\psi_p(U_p) \subset \mathbb{R}^n$. Hence we can conclude the proof by setting $\xi_p = \psi_p^{-1}$. $\qquad\square$

A point $\boldsymbol{x}_q \in \psi_p(U_p)$ is smoothly mapped to mapped to a point $q \in U_p$ by ξ_p, $q = \xi_p(\boldsymbol{x}_q)$. In the following we will express q in terms of the components of $\xi_p(\boldsymbol{x}_q)$, dropping $\boldsymbol{x}_q$ to simplify the notation

$$q = (\xi_p^{1,q}, \cdots, \xi_p^{n,q})^T \tag{A.5}$$

In general the map ψ_p is not unique and there may exist other maps $\widetilde{\psi}_p : U_p \to \mathbb{R}^n$ which are smoothly compatible with ψ_p. In this case we can find a transformation from the coordinates $\xi_p^{i,q}$ of any point $q \in U_p$ to the coordinate functions induced by $\widetilde{\psi}_p$

Definition 21 (Coordinate Transformation). *Let (U_p, ψ_p) be a coordinate chart of a smooth manifold M and $\widetilde{\psi}_p$ be a map with domain U_p and which is smoothly compatible with ψ_p. The map*

$$F = \widetilde{\xi}_p \circ \psi_p : U_p \to U_p, \quad \widetilde{\xi}_p = \psi_p^{-1} \tag{A.6}$$

is called a coordinate transformation. It maps the coordinate representation of a point $q \in U_p$ with respect to ψ_p, $q = (\xi_p^{1,q}, \cdots, \xi_p^{n,q})^T$ to the representation $q = (\widetilde{\xi}_p^{1,q}, \cdots, \widetilde{\xi}_p^{n,q})^T$ with respect to $\widetilde{\psi}_p$.

With the definition of the atlas of a smooth manifold M in definition 20 we have introduced a structure which is differentiable everywhere on M. Thus it is possible to assign a tangential vector X at every point $p \in M$. There are however continuously many ways of assigning such a vector to the point $p \in M$ so that in the next section we will introduce the tangential space T_pM which is the set of all tangential vectors at the point p.

A.1. The Tangent Space T_pM

In this section we would like to introduce the notion of a vector which is tangential to a smooth manifold M at the point $p \in M$. If M is 2-dimensional then we can visualize it by embedding it in $\mathbb{R}^3$. Then a tangential vector X at $p \in M$ is a 3-tuple (X_x, X_y, X_z) pinned at the point p. The set of all tangential vectors at p is the tangential space $T_pM \subset \mathbb{R}^2$ which can be thought of as a plane pinned to $p \in M$. For different $p, q \in M$ the corresponding tangential planes are disjunct, $T_pM \cap T_qM = \emptyset$.

The problem with the above description of the tangential space T_pM of a smooth n-dimensional manifold M is that it relies on embedding M into a higher dimensional euclidean space in which the elements of T_pM may be represented as $(n+1)$-tuples. The fact that T_pM is tangent to M at p reduces the dimensionality of T_pM to $dim(T_pM) = n$. Hence we would like a way of defining T_pM without any reference to an embedding space. The way this goes is by identifying T_pM with the set directional derivatives, or derivations of smooth functions on M as in the following.

Definition 22 (Derivation). *Let M be a smooth manifold, $U_p \subset M$ an open set containing $p \in M$ and $C^\infty(U_p)$ be the set of all smooth functions on U_p. X is called a derivation at $p \in M$ if it satisfies*

$$X(f \cdot g) = f(p)Xg + g(p)Xf \tag{A.7}$$
$$X(c \cdot f + d \cdot g) = cXf + dXg \tag{A.8}$$

for all $f, g \in C^\infty(U_p)$ and $c, d \in \mathbb{R}$.

The derivation X is clearly a derivative since eq. (A.7) is nothing else but the multiplication rule for derivatives. In fact later we will prove the identity $X = X_i \partial_{\xi_i}|_p$ where $\partial_{\xi_i}|_p$ is the i-th component of the gradient in local coordinates of $U_p \subset M$. We are now in a position to define the tangential space T_pU of $U_p \subset M$:

Definition 23 (Tangential Space). *The tangential space T_pU is the set of all derivations at $p \in U_p \subset M$*

Lemma 10 (Linearity)**.** *The tangent space T_pU is a linear vector space*

Proof. Let $X, Y \in T_pU$. We must show that the linear combination $Z = aX + bY$ is also a derivation and hence contained in T_pU. Let $f, g \in C^\infty(U_p)$ we apply the definition in eq. (A.7)

$$Z(f \cdot g) = aX(f \cdot g) + bY(f \cdot g) \tag{A.9}$$
$$= a(f(p)Xg + g(p)Xf) + b(f(p)Yg + g(p)Yf) \tag{A.10}$$
$$= f(p)(aXg + bYg) + g(p)(aXf + bYf) \tag{A.11}$$
$$= f(p)Zg + g(p)Zf \tag{A.12}$$

which show that Z fulfills eq. (A.7). Z also directly fulfills eq. (A.8). $\qquad\square$

Just like an ordinary gradient a derivation vanishes on constant functions as the following lemma shows

Lemma 11 (Properties of Derivations)**.** *Let M be a smooth manifold and T_pU be the tangent space at $p \in U_p \subset M$, then for all $X \in T_pU$*

1. *if $f \in C^\infty(U_p)$ is constant, then $Xf = 0$*

2. *if $f, g \in C^\infty(U_p)$ and $f(p) = g(p) = 0$, then $X(f \cdot g) = 0$*

Proof. Properity 1: By the linearity of X (eq. (A.8)) it suffices to prove 1 for $f = 1$. Since $1 = 1 \cdot 1$ we have

$$X1 = X(1 \cdot 1) = 1X1 + 1X1 = 2X1 \tag{A.13}$$

The second equality in eq. (A.13) comes directly from the definition in eq. (A.7). Eq. (A.13) can only hold if $Xf = 0$.

Property 2: This property is a direct consequence of the definition in eq. (A.7). $\quad\square$

The definition of the tangential space T_pU as the set of all derivations at $p \in U_p$ is a purely local definition: We only considered smooth functions $f \in C^\infty(U_p)$ over the region $U_p \subset M$ for the definition of the derivations in eq. (A.7). However it is not clear whether a tangential space exists for every $p \in M$. On the other side the smooth structure on M defined by its atlas of smoothly compatible charts suggest a smooth structure relating the tangential spaces T_pU and T_qU of two distinct points $p, q \in M$, $p \neq q$ since $\lim_{p \to q} T_pU = T_qU$. One of the technical issues is that the tangential spaces T_pU and T_qU for $p \neq q$ are defined over different function spaces $C^\infty(U_p)$ and $C^\infty(U_q)$. Thus we must first clarify if the definition 23 of

the tangential space T_pU can be transformed to a suitable definition of the set of derivations on the functions $f \in C^\infty(M)$ at point $p \in M$, T_pM.

A.1.1. The Push-Forward

An important object used to solve these issues is the *push-forward.*

Definition 24 (Push-Forward)**.** *Let M and N be smooth manifolds and $F : M \to N$ be a smooth map. Let $q \in N$ be the image of $p \in M$, $q = F(p)$ and $V_q \subset N$ be the image of the neighborhood $U_p \subset M$ located at p, $V_q = F(U_p)$. The map $F_\star$ defined on T_pU by*

$$(F_\star X)f = X(f \circ F) \quad \forall f \in C^\infty(V_{F(p)}),\ X \in T_pU \tag{A.14}$$

maps T_pU to the tangent space $T_qV = T_{F(p)}V$ and is called the push-forward of T_pU.

Lemma 12 (Properties of Push-Forwards)**.** *Let M, N and P be smooth manifolds, $p \in M$, $F : M \to N$, $G : N \to P$ be smooth maps and $U_p \subset M$, $V_{F(p)} \subset N$, $W_{G(F(p))} \subset P$ be local neighborhoods then*

1. *$F_\star : T_pU \to T_{F(p)}V$ is linear*

2. *Chain rule: $(G \circ F)_\star = G_\star \circ F_\star : T_pU \to T_{G(F(p))}W$*

3. *$(Id_{U_p})_\star = Id_{T_pU}$*

4. *F is a diffeomorphism $\Rightarrow F_\star : T_pU \to T_{F(p)}V$ is an isomorphism*

Proof. 1: Let $X, Y \in T_pU$ then for any $a, b \in \mathbb{R}$ the combination $Z = aX + bY$ is also a derivation in T_pU by lemma 10. For any smooth function $f \in C^\infty(V_{F(p)})$ we can apply the definition of $F_\star$ (eq. (A.14)) to Z

$$(F_\star Z)f = Z(f \circ F) = aX(f \circ F) + bY(f \circ F)$$
$$= a(F_\star X)f + b(F_\star Y)f \tag{A.15}$$

Eq. (A.15) shows that $F_\star$ is a linear map from T_pU to $T_{F(p)}V$.

2: For any $X \in T_pU$ and any $f \in C^\infty(W_{G(F(p))})$ we may apply the definition of the push forward in eq. (A.14) twice:

$$(G_\star \circ F_\star)Xf = (G_\star(F_\star X))f = (F_\star X)(f \circ G) = X(f \circ G \circ F) \tag{A.16}$$

On the other hand we can directly apply eq. (A.14) to the smooth map

$$H = G \circ F : U_p \to W_{G(F(p))}$$

which yields the same result $(H_\star X)f = X(f \circ G \circ F)$ thus concluding the proof.

3: This statement follows directly from eq. (A.14) by setting $F = Id_{U_p}$

4: To show that $F_\star$ is an isomorphism we only need to show that its inverse $(F_\star)^{-1}$ exists over all of $T_{F(p)}V$. Since F is a diffeomorphism its inverse $F^{-1} : N \to M$ exists and is smooth. From the chain rule in 2 and the identity lemma in 3 we have

$$(F^{-1})_\star \circ F_\star = (F^{-1} \circ F)_\star = (Id_{U_p})_\star = Id_{T_p U} \tag{A.17}$$

Hence $(F_\star)^{-1} = (F^{-1})_\star$ over $T_{F(p)}V$. $\qquad\square$

In the following we will show that $T_p U$ and $T_p M$ are isomorphic to each other so that every derivation on $C^\infty(U_p)$ can be extended to a derivation on $C^\infty(M)$. For this we show that every $X \in T_p U$ defines an equivalence relation on $C^\infty(M)$.

Proposition 4 (Equivalence Relation). *Let M be a smooth manifold, $p \in M$, $X \in T_p U$ and $f, g \in C^\infty(M)$. Let U_p be a neighborhood of p such that $f|_{U_p} = g|_{U_p}$. Then $X(f|_{U_p}) = X(g|_{U_p})$ and X and f define the equivalence set*

$$[f] = \{g \in C^\infty(M) \,|\, X(g|_{U_p}) = X(f|_{U_p})\} \tag{A.18}$$

Proof. We set $h = f - g$ and show that $X(h|_{U_p}) = 0$ when $h|_{U_p} = 0$. For this we select a subset $B \subset U_p$ such that $p \in B$ and a smooth bump function ψ such that

$$\psi(q) = \begin{cases} \epsilon(q), & \epsilon(q) \in [0,1), \quad q \in U_p, \\ 1, & else \end{cases}$$

$$\epsilon|_{\bar{B}} = 0, \quad \bar{B} \subset U_p$$

The smooth function $\epsilon(q)$ guarantees a smooth transition of ψ from M/U_p to B. Then $h(q) = \psi(q)h(q)$ for all $q \in M$. Since $h|_{U_p} = 0$ and $\psi(p) = 0$ we can use lemma 11 to show that $X(h|_{U_p}) = X(\psi h|_{U_p}) = 0$. $\qquad\square$

The equivalence relation in eq. (A.18) states that the actions of a derivation $X \in T_p U$ on two smooth functions $f, g \in C^\infty(M)$ which agree on an open neighborhood U_p of any $p \in M$ are equal, $X(f|_{U_p}) = X(g|_{U_p})$. However the expression $X(f|_{U_p})$ indicates that X is only defined for smooth functions in $C^\infty(U_p)$. Thus the goal is to show that for every $X \in T_p U$ there exists a $Y \in T_p M$ such that $Yf = X(f|_{U_p})$. In this sense the space $T_p U$ is embedded in the space $T_p M$.

We first need to describe how an open smooth manifold U is embedded into a larger manifold M. This is done with the *inclusion map* ι:

Definition 25 (Inclusion Map). *Let M and U be smooth manifolds. The map*

$$\iota : U \hookrightarrow M \tag{A.19}$$

$$\iota(x) = x \in M, \quad \forall x \in U \tag{A.20}$$

is called the inclusion map. It includes the manifold U into the manifold M by mapping any point $x \in U$ to the same point x now understood as an element in M.

The inclusion map ι is identical with the identity map on M restricted to U, $\iota = Id_M|_U$. However it emphasizes the action of pasting an independent smooth manifold U onto another manifold M such that U becomes a subset of M, $U \subseteq M$. The next proposition shows that the tangential space T_pU within any chart (U_p, ψ_p) is isomorphic to the tangent space T_pM of derivations acting on $C^\infty(M)$

Proposition 5 (Tangential Inclusion Map $\iota_\star$). *Let M be a smooth manifold, $U_p \subset M$ an open subset and $\iota : U_p \hookrightarrow M$ the inclusion map. Then the push-forward*

$$\iota_\star : T_pU \to T_pM \tag{A.21}$$

is an isomorphism for all $p \in M$.

Proof. $\iota_\star$ is injective: Let $X \in T_pU$ and B be a neighborhood of p with $\bar{B} \subset U_p$. For all smooth functions $f \in C^\infty(U_p)$ there exists an extension $\widetilde{f} \in C^\infty(M)$ such that $\widetilde{f}|_{\bar{B}} = f|_{\bar{B}}$ (see Extension Lemma in [63]). By proposition 4 we have

$$Xf = X(\widetilde{f}|_{\bar{B}}) = X(\widetilde{f}|_{U_p}) \tag{A.22}$$

Since $\widetilde{f}|_{U_p} = \widetilde{f} \circ \iota$ by definition 25 we can utilize the definition of the push-forward (definition 24) on ι

$$Xf = X(\widetilde{f}|_{U_p}) = X(\widetilde{f} \circ \iota) = (\iota_\star X)\widetilde{f} \tag{A.23}$$

It follows that if $\iota_\star X = 0$ holds, then $Xf = 0$. Hence $\iota_\star$ is injective.

$\iota_\star$ is surjective: We have to show that for any $Y \in T_pM$ there exists a $X \in T_pU$ such that $Y = \iota_\star X$. Let $\widetilde{f} \in C^\infty(M)$ be any smooth function and f be its restriction to U_p, $f = \widetilde{f}|_{U_p}$ and define an operator $X : C^\infty(U_p) \to \mathbb{R}$ by $Xf = Y\widetilde{f}$. By proposition 4 the derivation $Y\widetilde{f}$ and thus Xf is independent of the choice of $\widetilde{f}$. Hence we can choose $\widetilde{f} = g$ for any $g \in C^\infty(M)$ which satisfies $f = g \circ \iota$. Then

$$Yg = Y\widetilde{f} = Xf = X(g \circ \iota) = (\iota_\star X)g \tag{A.24}$$

Y is a derivation on $C^\infty(M)$ and thus the operator X is a derivation on $C^\infty(U_p)$:

$$Y(g_1 g_2) = g_1(p) Y g_2 + g_2(p) Y g_1 \quad \forall g_1, g_2 \in C^\infty(M) \tag{A.25}$$

If we set $g = g_1 g_2$ in eq. (A.24) we get

$$X(g_1 g_2 \circ \iota) = g_1(p) X(g_2 \circ \iota) + g_2(p) X(g_1 \circ \iota) \quad \forall g_1, g_2 \in C^\infty(M) \tag{A.26}$$

Thus any element $T_p M$ is the image of an element in $T_p U$ and $\iota_\star$ is surjective. $\quad \square$

Proposition 5 shows that every derivation in $T_p U$ is uniquely identified with a derivation in $T_p M$ by means of the inclusion map ι. Hence from now on we will consider the function space $C^(M)$ and identify $T_p U$ with the set of derivations acting on $C^\infty(M)$, $T_p M$.

A.2. The Basis of $T_p M$

In this section we want clarify how the derivations $X \in T_p M$ act on the smooth functions $f \in C^\infty(M)$ on a smooth manifold M. Specifically we will show that for any coordinate chart (U_p, ψ_p) with local coordinates $\xi_p = (\xi_p^1, \cdots, \xi_p^n)$ the tangential space $T_p M$ admits a basis of n differential operators $\partial_{\xi^i}|_q$, $1 \leq i \leq n$. The subscripts q and p indicates that the derivative $\partial_{\xi_p^i}$ acts on the functions $f \in C^\infty(M)$ at the point $q \in U_p \subset M$. We will start with the euclidean manifold $\mathbb{R}^n$. The tangential space $T_a \mathbb{R}$ at any point $a \in \mathbb{R}^n$ is the set of directional derivatives $D_{v|a}$ defined as follows

Definition 26 (Euclidean Directional Derivative). *Let $M = \mathbb{R}^n$, $a \in M$, $v \in \mathbb{R}^n$ and $f \in C^\infty(\mathbb{R}^n)$. The operator $D_{v|a} : C^\infty(\mathbb{R}^n) \to \mathbb{R}$ defined by*

$$D_{v|a} f = \left. \frac{d}{dt} f(a + vt) \right|_{t=0} = v_\mu \partial_\mu f|_a, \quad \partial_\mu \equiv \frac{\partial}{\partial x^\mu} \tag{A.27}$$

is called the euclidean directional derivative of f in direction of the vector v at the point $a \in \mathbb{R}^n$. In eq. (A.27) the index μ appears twice and we adopt the Einstein convention that a sum is carried out over indexes which appear twice.

The directional derivative in eq. (A.27) actually defines a map

$$D_{\cdot|a} : \mathbb{R}^n \to T_a \mathbb{R}^n \tag{A.28}$$

$$v \mapsto D_{v|a} \tag{A.29}$$

With the help of the map in eq. (A.28) we can define the basis of $T_a\mathbb{R}^n$ as the image of the canonical basis e^i of $\mathbb{R}^n$ under the map $D_{\cdot|a}$

Lemma 13 (Basis of $T_a\mathbb{R}^n$)**.** *Let $e_i \in span(\mathbb{R}^n)$ be the canonical basis elements of $\mathbb{R}^n$. The set*

$$span(T_a\mathbb{R}^n) = \left\{ D_{e^i|a} \, |e_i \in span(\mathbb{R}^n) \right\} \tag{A.30}$$

is a basis for $T_a\mathbb{R}^n$.

Proof. First note that $v \mapsto D_{v|a}$ is a linear map. We prove that each $D_{e^i|a}$ is independent from the other $D_{e^j|a}$, $j \neq i$ by contradiction. Suppose

$$D_{e^i|a} = \sum_{j \neq i} \alpha_j D_{e^j|a} \tag{A.31}$$

By linearity of the map $v \mapsto D_{v|a}$ we conclude that $e^i = \sum_{j \neq i} \alpha_j e^j$ which is in contradiction to the assumption that the $e_i \in span(\mathbb{R}^n)$ form a basis in $\mathbb{R}^n$. Hence the $D_{e^i|a}$, $1 \leq i \leq n$ are linearly independent. $\qquad \square$

The extension of the euclidean directional derivative in eq. (A.27) to a directional derivative on a smooth manifold M is straightforward

Definition 27. *Let M be a smooth manifold, (U_p, ψ_p) be a chart of M with local coordinates $\xi_p = \psi_p^{-1}$ and $f \in C^\infty(M)$. The operator $D_{v|p}$ defined by*

$$D_{v|p}f = \frac{d}{dt}f(p+vt)\Big|_{t=0} = v_i\partial_{\xi_p^i}f\Big|_p, \quad \partial_{\xi_p^i} \equiv \frac{\partial}{\partial\xi_p^i} \tag{A.32}$$

is called the directional derivative of f in direction of the vector v at the point $p \in M$.

The operators $\partial_{\xi_p^i}$ in eq. (A.32) form a basis of the tangential space T_pM (the proof of this claim is similar to that of lemma 13). Thus any $X \in T_pM$ can be expressed in terms of the basis $\partial_{\xi_p^i}$ by

$$X = X_i\partial_{\xi_p^i} \tag{A.33}$$

For any smooth map $F : M \to N$ which maps a chart (U_p, ψ_p) of M to another chart (U_q, ψ_q) of N such that $q = F(p)$ for all $p \in M$ and any $f \in C^\infty(N)$, we can express the push-forward $F_\star$ in terms of the local coordinates ξ_p of U_p and ξ_q of U_q:

$$(F_\star X)f = X(f \circ F) = X_i\partial_{\xi_p^i}(f \circ F)\Big|_p = X_i\partial_{\xi_p^i}(F_j)\Big|_p \partial_{\xi_q^j}(f)\Big|_{F(p)} \tag{A.34}$$

Thus the push-forward $F_\star$ maps the basis of T_pM to a basis of $T_{F(p)}N$

$$F_\star : \partial_{\xi_p^i} \mapsto \partial_{\xi_p^i}(F_j)\partial_{\xi_q^j}, \quad q = F(p) \tag{A.35}$$

and $\partial_{\xi_p^i}(F_j)$ is the Jacobian of the map F evaluated at $p \in M$.

In case F is a coordinate transformation of the coordinates ξ_p^i to new coordinates $\widetilde{\xi}_p^i$ (eq. (A.6)) its Jacobian represents a transformation of the basis of T_pM

$$X = X^i \partial_{\xi_p^i} = \widetilde{X}^j \partial_{\widetilde{\xi}_p^j}, \quad \widetilde{X}^j = X^i \frac{\partial \widetilde{\xi}_p^j}{\partial \xi_p^i} \tag{A.36}$$

In eq. (A.36) the coefficients X^i compensate the transformation of the basis of T_pM such that the vector $X \in T_pM$ is invariant with respect to coordinate transformations. Hence the correspondence $p \leftrightarrow T_pM$ is a one-to-one correspondence.

We have shown that the tangential space T_pM at any point $p \in M$ admits a basis of operators $\partial_{\xi_p^i}$. Furthermore given a smooth map between two manifolds, the related push-forward is a transformation of the basis of T_pM. These constructions are important for the next section where we introduce the concept of a vector-field which is basically a smooth map $\widetilde{X} : p \mapsto \widetilde{X}_p \in T_pM$. If there exists a set of smooth vector fields V^i such that the $V^i(p) \in T_pM$ constitute a basis of T_pM then that basis is well defined for all $p \in M$.

A.3. Vector Fields

To understand $\widetilde{X} : p \mapsto X \in T_pM$ as a (possibly smooth) map on M we need to define what its range has to look like. For the time-being we will note the range of $\widetilde{X}$ as TM, $\widetilde{X} : M \to TM$. The map $\widetilde{X}$ maps each point $p \in M$ to an element $X \in T_pM$ in a two-fold manner, first by mapping p to the tangential space T_pM then by selecting the vector $X \in T_pM$. Since the map between a point p and its tangential space T_pM is a one-to-one correspondence and M is an n-dimensional manifold, the set $\{T_pM \,|\, p \in M\}$ of all tangential spaces T_pM is n-dimensional. On the other side each tangential space T_pM is an n-dimensional vector space and the vector X is represented by the coefficients X_i (eq. (A.33)) in the basis of T_pM. Thus the image of the map $\widetilde{X} : M \to TM$ is a $2n$-dimensional structure (n dimensions from the correspondence $p \leftrightarrow T_pM$, n dimensions from each T_pM) and we call TM the *tangential bundle*

Definition 28 (Tangential Bundle). *Let M be a smooth manifold. The set TM defined as the disjoint union of all tangential spaces T_pM*

$$TM = \coprod_{p \in M} T_pM, \quad T_pM \cap T_qM = \emptyset, \; \forall p, q \in M, \, p \neq q \tag{A.37}$$

is called the tangential bundle.

For any $p \in M$ and $X \in T_pM$ the pair $(p, X) \in TM$ is element of the tangential bundle TM and we will make the abbreviation

$$X_p := (p, X) \in TM \tag{A.38}$$

The following lemma shows that TM is itself a smooth manifold

Lemma 14 (Smooth Tangent Bundle). *For any smooth manifold M the tangential bundle TM has a natural topology and a smooth structure making it a $2n$-dimensional smooth manifold*

We will only briefly sketch a proof. The full proof can be found in [63].

Proof. Let (U_p, ψ_p) be a chart of M. Let the tangential bundle $TU_p \subset TM$ be defined as

$$TU_p = \coprod_{q \in U_p} T_qM \tag{A.39}$$

A chart $(TU_p, \widetilde{\psi}_p)$ may be defined such that $\widetilde{\psi}_p : TU_p \to \mathbb{R}^{2n}$

$$\widetilde{\psi}_p(X_q) = \widetilde{\psi}_p \left(X_i \partial_{\xi_q^i} \big|_q \right) = (\psi_p^T(q), X_1, \cdots, X_n)^T \tag{A.40}$$

The image $\widetilde{\psi}_p(TU_p) = \psi_p(U_p) \times \mathbb{R}^n$ is a subset of $\mathbb{R}^{2n}$ and $\widetilde{\psi}_p$ is a smooth map on TU_p. Let $(TU_a, \widetilde{\psi}_a)$, $a \in M$ be another chart on TM with $TU_a \cap TU_p \neq \emptyset$. Then the map $F = \widetilde{\psi}_a \circ \widetilde{\psi}_p^{-1}$ is a diffeomorphism on TM, and thus TM has a structure of smoothly compatible charts $(TU_p, \widetilde{\psi}_p)$. $\qquad\square$

Now that we have established the tangential bundle TM as a smooth manifold we define a vector field Y to be a map between two smooth manifolds, namely M and its tangential bundle TM

Definition 29 (Vector Field). *Let M be a smooth manifold and TM be its tangential bundle. For any $p \in M$ let (U_p, ψ_p) be a chart of M with the local coordinates $\xi_p = (\xi_p^1, \cdots, \xi_p^n)$. The map $Y : M \to TM$ defined by*

$$Y : p \mapsto Y_p = Y^i(p)\partial_{\xi_p^i}\Big|_p \in T_p M \tag{A.41}$$

is called a vector field. Its coefficients Y^i are functions on M.

The vector field Y can be thought of as a map which assigns to each $p \in M$ an arrow Y_p. If the coefficient functions of Y are smooth, $Y^i \in C^\infty(M)$, then Y is called a *smooth vector field* and we define $\mathcal{T}(M)$ to be the set of all smooth vector fields.

Lemma 15 (Smooth Vector Fields). *Let $Y \in \mathcal{T}(M)$ be a smooth vector field and $f \in C^\infty(M)$. Then $Yf \in C^\infty(M)$.*

Proof. For any chart (U_p, ψ_p) of M let g^i be the i-th derivative of f with respect to the local coordinates of U_p, $g^i(p) = \partial_{\xi_p^i} f\Big|_p$. Since f is smooth g^i is also smooth. By definition of $\mathcal{T}(M)$ the coefficients Y^i of Y are also smooth functions. Hence $Yf \in C^\infty(M)$. $\qquad\square$

The vector field $Y \in \mathcal{T}(M)$ maps every point $p \in M$ to a particular tangential vector $Y_p \in T_p M$ which satisfies the multiplication rule for derivations in eq. (A.7). Hence Y itself is a derivation since for any $f, g \in C^\infty(M)$ we have

$$Y(fg) \in C^\infty(M), \quad Y(fg)(p) = f(p)Y_p g + g(p)Y_p f, \quad \forall p \in M \tag{A.42}$$

and $\mathcal{T}(M)$ can also be thought of as the set of all derivations.

A.4. Push-Forwards on $\mathcal{T}(M)$

In definition 24 we introduced the push-forward $F_\star$ of a smooth map $F : M \to N$ which maps a smooth manifold M to another smooth manifold N. We showed that $F_\star$ maps the tangential space $T_p M$ to the tangential space $T_{F(p)} N$ at $F(p) \in N$ for all $p \in M$. The question arises if $F_\star$ also maps the set of smooth vector fields on M, $\mathcal{T}(M)$ to the smooth vector fields on N, $\mathcal{T}(M)$. Naively one would think that for any $X \in \mathcal{T}(M)$ there should exist a $Z \in \mathcal{T}(N)$ such that for all $p \in M$, Z is the push-forward of X

$$Z_{F(p)} f = (F_\star X_p) f = X_p(f \circ F), \quad \forall f \in C^\infty(N), p \in M \tag{A.43}$$

134

however the issue is that the derivation Z must be defined over all N since $Zf \in C^\infty(N)$ is required (see lemma 15) and hence we must be able to calculate $Z_q f$ in terms of X for all $q \in N$, not only for $q \in F(M)$. We will show that $\mathcal{T}(M)$ can only be mapped to $\mathcal{T}(N)$ by F if F is a diffeomorphism and not only merely smooth. In case we already know of the existence of a smooth vector field $Z \in \mathcal{T}(N)$ which is related to $X \in \mathcal{T}(M)$ via eq. (A.43) then we say Z and X are F-related

Definition 30 (*F-Related*). *Let M and N be smooth manifolds, $F : M \to N$, $X \in \mathcal{T}(M)$ and $Z \in \mathcal{T}(N)$. Then X and Z are F-related if*

$$Z_{F(p)} = (F_\star)X_p, \quad \forall p \in M \tag{A.44}$$

The following proposition shows that the vector field Z is smoothly defined over N if F is a diffeomorphism

Proposition 6 (*Push-Forward on $\mathcal{T}(M)$*). *Let M and N be smooth manifolds and $F : M \to N$ be a diffeomorphism. Then for all $X \in \mathcal{T}(M)$ there exists a unique $Z \in \mathcal{T}(N)$ which is F-related to X. Z is called the push-forward of X.*

Proof. Let $Z_{F(p)}$ be the point wise push-forward of X

$$Z_{F(p)} = F_\star X_p, \quad \forall p \in M \tag{A.45}$$

Since F is a diffeomorphism its inverse exists and is surjective, $F^{-1}(N) = M$. Hence we can define Z such that its value Z_q is the push-forward of X_p evaluated at $p = F^{-1}(q)$ for any $q \in N$

$$Z_q = F_\star X_{F^{-1}(q)} \tag{A.46}$$

According to lemma 12 the push-forward $F_\star$ is an isomorphism making Z the unique vector field which is F-related to X. Now we need to show that Z is a smooth vector field. Let (U_p, ψ_p) be a chart of M with local coordinates $\xi_p = \psi_p^{-1}$ and (V_q, ψ_q) be a chart of N with local coordinates $\xi_q = \psi_q^{-1}$ such that $F(U_p) \subset V_q$. According to eq. (A.35) the push-forward $F_\star$ can be written explicitly in local coordinates

$$Z_q = Z^j(q)\partial_{\xi_q^j}, \quad Z^j(q) = \left(\partial_{\xi_p^j} F^i\right)(p) X^i(p)\Big|_{p=F^{-1}(q)} \tag{A.47}$$

Since the coefficient function Z^j in eq. (A.47) is a composite of smooth maps, it is itself smooth, $Z^j \in C^\infty(N)$. Hence $Z \in \mathcal{T}(N)$ $\qquad\square$

A.5. Integral Curves and Flows

Proposition 6 holds for diffeomorphisms which map a smooth manifold M onto itself, $F : M \to M$. A special case of diffeomorphisms on M are the flow maps $\theta_t^V : M \to M$ which are maps that are parameterized by a *time* variable t. Every smooth vector field $V \in \mathcal{T}(M)$ uniquely determines a flow θ_t^V via the relation

$$\frac{d}{dt}\theta_t^V(p) = V_{\theta_t^V(p)}, \quad \forall p \in M \tag{A.48}$$

and we call V the *generator* of the flow θ_t^V, hence the superscript. Eq. (A.48) is at the core of the fundamental theorem for flows, to be stated at the end of the section.

We had shown that the smooth vector field V acts as a derivation, or directional derivative on the set of smooth functions $C^\infty(M)$. Using the fundamental theorem for flows embodied in eq. (A.48) we will introduce the *Lie* derivative $\mathcal{L}_V W$ of a smooth vector field $W \in \mathcal{T}(M)$ which describes how W transforms along the flow θ_t^V governed by V. The crucial result is that $\mathcal{L}_V W$ will be shown to be equal to an anti-symmetric bilinear form $[\cdot, \cdot]$ on $\mathcal{T}(M)$ called the *commutator*

$$\mathcal{L}_V W = [V, W] \in \mathcal{T}(M) \tag{A.49}$$

The commutator $[\cdot, \cdot]$ plays a central role in Lie group theory (see appendix B) and our new approach to the principle of least action in section 4.

To begin with, we will discuss the concept of an *integral curve* of any given smooth vector field $X \in \mathcal{T}(M)$.

Definition 31 (Integral Curve). *Let M be a smooth manifold, $J \subset \mathbb{R}$ be an open interval such that $0 \in J$ and $V \in \mathcal{T}(M)$. The map $\gamma^V : J \to$, $\gamma^V(0) = p \in M$ is called an integral curve if its derivative is a derivation on M*

$$\left.\frac{d}{dt}\gamma^V(t)\right|_{t=t_0} = V_{\gamma(t_0)}, \quad \forall t_0 \in J \tag{A.50}$$

From the definition of γ^V is eq. (A.50) it is not clear whether it exists given the smooth vector field V and the point $p \in M$. Consider any local chart (U_p, ψ_p) in M with local coordinates $\xi_p = \psi_p^{-1}$ and an open subset $D \subset J$ such that $\gamma^V(D) \subseteq U_p$. Then

$$\tilde{\gamma}^V = \psi_p \circ \gamma^V, \quad \tilde{\gamma}^V(t) = (x_1(t), \cdots, x_n(t)) \in \psi_p(U_p), \, t \in D \tag{A.51}$$

is a curve on the open set $\psi_p(U_p) \subset \mathbb{R}^n$ with origin $\widetilde{\gamma}^V(0) = \psi_p(p) = \mathbf{0}$. Hence on D we can express γ^V in terms of the local coordinates of U_p

$$\gamma^V(t) = (\xi_{p,1}^\gamma(t), \cdots, \xi_{p,n}^\gamma(t)) = (\xi_{p,1}(\widetilde{\gamma}^V(t)), \cdots, \xi_{p,n}(\widetilde{\gamma}^V(t))) \tag{A.52}$$

We can now express both sides of eq. (A.50) in terms of the basis of $T_{\gamma^V(t_0)}$

$$\left(\frac{d}{dt} \xi_{p,i}^\gamma \right) \partial_{\xi_{p,i}} = V_i(\gamma^V(t)) \partial_{\xi_{p,i}} \tag{A.53}$$

As the operators $\partial_{\xi_{p,i}}$ are independent from each other eq. (A.53) is satisfied when the curve γ^V satisfies the differential equations

$$\begin{array}{ccc} \frac{d}{dt} \xi_{p,1}^\gamma & = & V_1(\gamma^V(t)) \\ \vdots & = & \vdots \\ \frac{d}{dt} \xi_{p,n}^\gamma & = & V_n(\gamma^V(t)) \end{array} \tag{A.54}$$

The differential equations in eq. (A.54) are first order equations in the time variable t, thus we only need to specify the initial value $\gamma^V(0) = p$ to compute the solution $\gamma_{U_p}^V : D \to U_p$. The curve $\gamma_{U_p}^V$ is only a local section of the unknown curve γ^V. However by selecting a neighboring chart U_q, ψ_q such that $U_q \cap U_p \neq \emptyset$ for $p \neq q$ we can extend $\gamma_{U_p}^V$ to U_q by solving eq. (A.54) in U_q and combining the resulting solution $\gamma_{U_q}^V$ with $\gamma_{U_p}^V$ to a curve with a larger domain.

The integral curve $\gamma^V(t)$ not only depends on the smooth vector field V but also on the starting point $\gamma^V(0) = p \in M$. However an integral curve can be reparameterized to change its starting point since if $a \in \mathbb{R}$ then $\gamma^V(t-a)$ is also an integral curve of V. Thus we would like to characterize the set of integral curves of a smooth vector field V regardless of the starting points. A collection of curves on M is called a *flow*

Definition 32 (Flow). *A Flow domain is an open subset $\mathcal{D} \subset \mathbb{R} \times M$ with the property that for all $p \in M$ the set*

$$\mathcal{D}^p = \{ t \in \mathbb{R} \, | (t, p) \in \mathcal{D} \}, \quad 0 \in \mathcal{D}^p \tag{A.55}$$

is an open interval. A flow is a map $\theta : \mathcal{D} \to M$ which satisfies the group laws:

- *For all $p \in M$*

$$\theta(0, p) = p \tag{A.56}$$

- *For all $s \in \mathcal{D}^p$ and $t \in \mathcal{D}^{\theta(s,p)}$ such that $s + t \in \mathcal{D}^p$ the composition law $\theta(s, \theta(t, p)) = \theta(s + t, p)$ holds*

The map $\theta : (t, p) \mapsto \theta(t, p)$ is called a smooth flow if it is smooth in both t and p. A flow $\bar{\theta} : \bar{\mathcal{D}} \to M$ is called maximal if there exists no other flow $\widetilde{\theta} : \widetilde{\mathcal{D}} \to M$ with larger flow domain $\widetilde{\mathcal{D}} \supset \bar{\mathcal{D}}$ such that $\bar{\theta}$ and $\widetilde{\theta}$ are equal on their common domain. The integral curve $\bar{\theta}_p : \bar{\mathcal{D}}^p \to M$ is called the maximal integral curve.

Each smooth flow $\theta : \mathcal{D} \to M$ defines for fixed $p \in M$ a smooth curve

$$\theta_p = \theta(\cdot, p) : \mathcal{D}^p \to M \tag{A.57}$$

and for fixed $t \in \mathbb{R}$ a diffeomorphism

$$\theta_t = \theta(t, \cdot) : M \to M \tag{A.58}$$

Suppose $V \in \mathcal{T}(M)$ is a smooth vector field. If there exists a flow θ^V as a solution to the differential equation

$$\left. \frac{d}{dt} \theta^V(t, p) \right|_{t=0} = V_p \tag{A.59}$$

then we say that θ^V is generated by V, or V is the *generator* of θ^V. The following fundamental theorem for flows shows that for each smooth vector field $V \in \mathcal{T}(M)$ there exists a unique maximal flow θ^V generated by V

Theorem 3 (Fundamental Theorem on Flows). *Let V be a smooth vector field on a smooth manifold M. There exists a unique maximal flow $\theta^V : \mathcal{D} \to M$ whose generator is V such that*

 a *For all $p \in M$, $\theta_p^V : \mathcal{D}^p \to M$ is the unique maximal integral curve of V with starting point p*

 b *If $s \in \mathcal{D}^p$, then $\mathcal{D}^{\theta(s,p)} = \mathcal{D}^p - s = \{t - s \,|\, t \in \mathcal{D}^p\}$*

 c *For all $s \in \mathbb{R}$ the set $M_s = \{p \in M \,|\, (s, p) \in \mathcal{D}\}$ is open in M and and the map $\theta_s : M_s \to M_{-s}$ is a diffeomorphism with inverse θ_{-s}*

 d *For each $(t, p) \in \mathcal{D}$, $(\theta_t)_\star V_p = V_{\theta_t(p)}$*

We will only make some remarks on theorem 3 and refer the reader for its lengthly proof to [63]. Part b and c of theorem 3 basically state that the defining differential equation of θ^V in eq. (A.59) is *reparameterization invariant* with respect to both the time variable t and the starting point $p \in M$. For in the case of part b if $\theta^V(t, p)$ is a solution to eq. (A.59) with starting point p at $t = 0$ then

$$\widetilde{\theta}^V(t, p) = \theta^V(t + s, p)$$

is also a solution to eq. (A.59) with starting point $\theta^V(s, p)$. In case of part c, the map θ^V_s directly takes the starting point of $\theta^V(t, p)$, p to the starting point of $\widetilde{\theta}^V(t, p)$ which is $\theta^V(s, p)$, the same starting point as for the time reparameterization in part b.

For our needs part d is the most important part of the fundamental theorem for flows. A smooth vector field W is said to be *invariant* under a flow θ if the push-forward of the vector W_p at any point $p \in M$ is equivalent to the evaluation of W at the point $\theta_t(p)$

$$(\theta_t)_\star W_p = W_{\theta_t(p)} \tag{A.60}$$

Part d states that every smooth vector field $V \in \mathcal{T}(M)$ is invariant under its unique maximal flow θ^V. In the next section we will introduce the Lie derivative which is a method to compute the rate of change a smooth vector field undergoes under the flow generated by *another* vector field. This will facilitate the proof of part d in theorem 3.

A.5.1. The Lie Derivative

Let $V, W \in \mathcal{T}(M)$ be two smooth vector fields on the smooth manifold M and $\theta^V : \mathcal{D} \to M$ be the flow corresponding to V. One of the central questions in the theory of smooth manifolds is: how does the vector field W transform under the action of the flow θ^V, or more specifically what is the rate of change of the vector $W_{\theta^V_p(t)} \in T_{\theta^V_p(t)}M$ with respect to changes in $t \in \mathcal{D}^p$? Naively one could directly try to calculate the derivative

$$\lim_{h \to 0} \frac{W_{\theta^V_p(t+h)} - W_{\theta^V_p(t)}}{h} \tag{A.61}$$

The problem with the derivative in eq. (A.61) is that $W_{\theta^V_p(t+h)}$ and $W_{\theta^V_p(t)}$ belong to *different* tangential spaces $T_{\theta^V_p(t+h)}M$ and $T_{\theta^V_p(t)}M$ and thus eq. (A.61) is not defined. However by the fundamental theorem for flows in theorem 3 the map θ^V_t is a diffeomorphism and thus the push-forward $(\theta^V_{-t})_\star$ smoothly maps $T_{\theta^V_p(t)}M$ to T_pM for all $t \in \mathcal{D}^p$. This means the vector $(\theta^V_{-t})_\star W_{\theta^V_p(t)} \in T_pM$ is smooth in t and thus differentiable

Definition 33 (Lie Derivative). *Let $V, W \in \mathcal{T}(M)$ be two smooth vector fields and θ^V be the flow of V. The Lie derivative $\mathcal{L}_V W$ of W along the flow of V is defined as the derivative*

$$(\mathcal{L}_V W)_p = \frac{d}{dt}\left((\theta^V_{-t})_\star W_{\theta^V_p(t)} \right)\Big|_{t=0} \tag{A.62}$$

One of the most important aspects of the Lie derivative $(\mathcal{L}_V W)$ is that it is exceptionally easy to compute *without* reference to the flow of V as the next proposition shows

Proposition 7 (Commutator). *Let $V, W \in \mathcal{T}(M)$ be smooth vector fields. Let (U_p, ψ_p) be a smooth chart at any $p \in M$. The commutator $[V, W]$ is defined by*

$$[V, W]_p = V_p(W^i)\partial_{\xi_p^i} - W_p(V^i)\partial_{\xi_p^i} \tag{A.63}$$

The Lie derivative of W along V is equivalent to the commutator $[V, W]$

$$\mathcal{L}_V W = [V, W] \tag{A.64}$$

Furthermore $\mathcal{L}_V W$ is also a derivation on M, $\mathcal{L}_V W \in \mathcal{T}(M)$.

Proof. We define the open set

$$\mathcal{R}_V = \{p \in M \,|\, V_p \neq 0\} \tag{A.65}$$

We want to show that eq. (A.64) holds on $\mathcal{R}_V$. Let $p \in \mathcal{R}_V$ and $U_p \subset \mathcal{R}_V$ be a neighborhood with local coordinates u_p^i such that the vector field V becomes a directional derivative along u^1, $V = \partial_{u_p^1}$. Then the flow θ^V of V in U_p is linear in the time parameter

$$\theta^V(t, p) = \left(u_p^{1,p} + t, u_p^{2,p}, \cdots, u_p^{n,p}\right) \tag{A.66}$$

where $u_p^{i,p}$ are the coordinates of the point p. We evaluate the Lie derivative as defined in eq. (A.62). The push-forward $(\theta_{-t})_\star$ in the coordinates u_p^i evaluates to the identity matrix for any $t \in \mathcal{D}^p$ such that eq. (A.62) becomes

$$(\mathcal{L}_V W)_p = \frac{d}{dt}\left(W^i(\theta^V(t,p))\partial_{u_p^i}\Big|_p\right)\Big|_{t=0} = \left(\partial_{u_p^1}W^i(p)\right)\partial_{u_p^i}\Big|_p \tag{A.67}$$

On the other hand we can evaluate the commutator in eq. (A.63) in the coordinates u_p^i

$$[V, W]_p = V(W^i)\partial_{u_p^i}|_p - W(V^i)\partial_{u_p^i}|_p = \left(\partial_{u_p^1}W^i(p)\right)\partial_{u_p^i}\Big|_p \tag{A.68}$$

Hence for $p \in \mathcal{R}_V$ we have

$$(\mathcal{L}_V W)_p = [V, W]_p \tag{A.69}$$

For the case $p \in M/\mathcal{R}_V$ where $V_p = 0$ the flow θ^V is the identity for all $t \in \mathbb{R}$,

$\theta(t, p) = p$. It follows that the Lie derivative vanishes on $M/\mathcal{R}_V$, $(\mathcal{L}_V W)|_{M/\mathcal{R}_V} = 0$. On the other side the commutator $[V, W]_p$ also vanishes on $M/\mathcal{R}_V$

$$[V, W]_p = V_p W_p - W_p V_p = 0, \quad \forall p \in M/\mathcal{R}_V \tag{A.70}$$

Hence eq. (A.64) hold on the entire manifold M.

Since V_p and W_p are smooth derivations on M the operator $V_p W_p$ is also a derivation by the chain rule. Hence $\mathcal{L}_V W \in \mathcal{T}(M)$ as was claimed. $\qquad\square$

It hard to overestimate the importance of the equivalence in eq. (A.64). For one, eq. (A.64) allows us to elegantly proof part d of the fundamental theorem of flows (theorem 3). Part d is equivalent to the statement that

$$\mathcal{L}_V V = 0 \tag{A.71}$$

Eq. (A.71) is a direct result of the commutator relationship to the Lie derivative in eq. (A.64). Another important property of the commutator $[V, W]$ is that if $W_p = 0$ at the starting point p of the flow θ^V generated by V then the Lie derivative of W also vanishes, $(\mathcal{L}_V W)_p = 0$. Hence W is invariant along the integral curve θ_p^V.

B. Lie Groups

B.1. The Prolonged Action

The prolonged action of an r-dimensional Lie Group $\mathbb{G}$ on the gradient $\nabla \phi$ is given by

$$D\phi_i^\nu = \frac{d\omega_i^\phi}{dx_\nu} - \sum_\mu \frac{d\omega_\mu^i}{dx_\nu} \partial_\mu \phi \tag{B.1}$$

We want to derive eq. (B.1) following [68, 83] by considering an arbitrary one dimensional subgroup $g_{\gamma(t)} \subset \mathbb{G}$ with $g_{\gamma(0)} = e$. The function $\gamma(t) = (a_1(t) \ldots a_r(t))$ is a continuous path in the parameter space of the group $\mathbb{G}$. The gradient $\nabla\phi$ transforms under the action of $g_{\gamma(t)}$ like

$$F(t, \nabla\phi(\boldsymbol{x})) = g_{\gamma(t)} \circ (\nabla\phi(\boldsymbol{x})) = \widetilde{\nabla}\widetilde{\phi}(\widetilde{\boldsymbol{x}}), \quad \widetilde{\nabla} = J_t^{-T}\nabla \tag{B.2}$$

where J_t is the Jacobian

$$J_{t,\mu\nu} = \frac{d\widetilde{x_\mu}}{dx_\nu}, \quad \widetilde{\boldsymbol{x}} = g_{\gamma(t)} \circ \boldsymbol{x} \tag{B.3}$$

We compute the prolonged action $D\phi_i^\nu$ by taking the derivative of $F(t, \nabla\phi(vx))$ with respect to t at $t = 0$

$$\left.\frac{dF}{dt}\right|_{t=0} = \left.\frac{d}{dt}J_t^{-T}\right|_{t=0}\nabla\phi + J^{-T}\nabla \left.\frac{d}{dt}\widetilde{\phi}(\widetilde{\boldsymbol{x}})\right|_{t=0} \tag{B.4}$$

$$= -J_t^{-T}\left.\frac{d}{dt}J_t\right|_{t=0}J_t^{-T}\nabla\phi + J^{-T}\nabla \left.\frac{d}{dt}\widetilde{\phi}(\widetilde{\boldsymbol{x}})\right|_{t=0} \tag{B.5}$$

The Jacobian evaluated at $t = 0$ yields the identity matrix

$$J_{0,\mu\nu} = \delta_{\mu\nu} \tag{B.6}$$

The derivative of the transformed coordinates $\widetilde{x}$ and the field $\widetilde{\phi}$ with respect to t at $t=0$ yield linear combinations of the infinitesimal variations $\boldsymbol{\omega}^i$ and ω_ϕ^i

$$\left.\frac{d\widetilde{x}_\mu}{dt}\right|_{t=0} = \sum_{i=1}^{r} \omega_\mu^i \alpha_i, \quad \left.\frac{d\widetilde{\phi}(\widetilde{\boldsymbol{x}})}{dt}\right|_{t=0} = \sum_{i=1}^{r} \omega_\phi^i \alpha_i \tag{B.7}$$

$$\omega_\mu^i = \partial_{a_i}\widetilde{x}_\mu|_{g=e}, \quad \omega_\phi^i = \partial_{a_i}\widetilde{\phi}|_{g=e}, \quad \alpha_i = \left.\frac{a_i}{dt}\right|_{t=0} \tag{B.8}$$

The derivative of the Jacobian evaluates to

$$\left.\frac{dJ_{t,\mu\nu}}{dt}\right|_{t=0} = \sum_{i=1}^{r} \frac{d\omega_\mu^i}{dx_\nu}\alpha_i \tag{B.9}$$

So that eq. (B.5) becomes

$$\left.\frac{dF_\nu}{dt}\right|_{t=0} = \sum_{i=1}^{r} \left(\frac{d\omega_i^\phi}{dx_\nu} - \sum_\mu \frac{d\omega_\nu^i}{dx_\mu}\partial_\mu\phi \right) \alpha_i = \sum_{i=1}^{r} D\phi_i^\nu \alpha_i \tag{B.10}$$

The prolonged action for higher order gradients goes along the same lines, expressing the transformed gradient $\widetilde{\nabla}^n$ in terms of the Jacobian J_t

$$\widetilde{\nabla}^n\widetilde{\phi} = (\prod_{l=1}^{n} J^{-T}\nabla)\widetilde{\phi} \tag{B.11}$$

and using the product rule to compute the derivative with respect to t.

B.2. Geometrical Meaning of the Commutator $[\cdot, \cdot]$

In section 2.6 we mentioned that if $X_e \in \mathcal{G}$ is a left invariant vector field evaluated at the identity e of the Lie group $\mathbb{G}$ and $g_{\gamma(t)} \subset \mathbb{G}$ is a one parameter subgroup defined by the vector field V_g

$$\left.\frac{dg_{\gamma(t)}}{dt}\right|_{t=0} = V_e \tag{B.12}$$

then the rate of change of $X_{g_{\gamma(t)}}$ (X_e translated by $g_{\gamma(t)}$) is governed by the commutator between X_e and V_e

$$\left.\frac{dX_{g_{\gamma(t)}}}{dt}\right|_{t=0} = [V_e, X_e] \tag{B.13}$$

Since both X_g and V_g are left invariant the commutator in eq. (B.13) is also left invariant and thus an element of $\mathcal{G}$. We will prove eq. (B.13) for the case when both V_g and X_g are spatial operators

$$X_e = \omega_\Omega^\nu\left(\boldsymbol{x}\right)\partial_\nu, \quad V_e = v_\Omega^\nu\left(\boldsymbol{x}\right)\partial_\nu \tag{B.14}$$

We compute the rate of change of $X_{g_{\gamma(t)}}$

$$\left.\frac{dX_{g_{\gamma(t)}}}{dt}\right|_{t=0} = \left.\frac{d\omega_\Omega^\nu\left(\widetilde{\boldsymbol{x}}\right)}{dt}\right|_{t=0}\partial_\nu + \omega_\Omega^\nu\left(\boldsymbol{x}\right)\left.\frac{d\widetilde{\partial_\nu}}{dt}\right|_{t=0} \tag{B.15}$$

$$= \frac{d\omega_\Omega^\nu}{dx_\mu}\cdot v_\Omega^\mu\cdot\partial_\nu - \omega_\Omega^\nu\cdot\frac{dv_\Omega^\mu}{dx_\nu}\partial_\mu \tag{B.16}$$

$$= \left(\frac{d\omega_\Omega^\mu}{dx_\nu}\cdot v_\Omega^\nu - \omega_\Omega^\nu\cdot\frac{dv_\Omega^\mu}{dx_\nu}\right)\partial_\mu = \left(V_e\left(\omega_\Omega^\mu\right) - X_e\left(v_\Omega^\mu\right)\right)\partial_\mu \tag{B.17}$$

On the other hand we can directly compute the commutator of X_e and V_e

$$[V_e, X_e] = v_\Omega^\mu\frac{d}{dx_\mu}\left(\omega_\Omega^\nu\frac{d}{dx_\nu}\right) - \omega_\Omega^\mu\frac{d}{dx_\mu}\left(v_\Omega^\nu\frac{d}{dx_\nu}\right) \tag{B.18}$$

When carrying out the product rule the terms $v_\Omega^\mu\omega_\Omega^\nu\frac{d}{dx_\mu}\frac{d}{dx_\nu}$ cancels out so that eq. (B.18) equals eq. (B.17) proving eq. (B.13).

B.3. Derivation Of Noethers Theorem

In this section we want to derive Noethers theorem from the variation of the energy

$$E\left(\phi, \boldsymbol{X}_e^\Omega\phi\right) = \int_\Omega \mathcal{E}\left(\phi\left(\boldsymbol{x}\right), \boldsymbol{X}_e^\Omega\phi\left(\boldsymbol{x}\right)\right)\mathcal{N}d^2x \tag{B.19}$$

under an arbitrary one parameter sub group $g_{\gamma(t)} \subset \mathbb{G}$, where the Lie group $\mathbb{G}$ is the group of spatial and intensity transformations $\mathbb{G} = \mathbb{G}_\Omega \times \mathbb{G}_\phi$ of dimension r. The set $\boldsymbol{X}_e^\Omega$ is the commutative sub algebra of the Lie algebra $\mathcal{G}$ of $\mathbb{G}$. The tangential vector V_e at the identity of $g_{\gamma(t)}$ is an element of $\mathcal{G}$ and has the form

$$\left.\frac{dg_{\gamma(t)}}{dt}\right|_{t=0} = V_e = v^\mu(\boldsymbol{x})\partial_\mu + v^\phi(\boldsymbol{x})\frac{\delta}{\delta\phi} \tag{B.20}$$

To reduce clutter we will use $v^\mu := v^\mu(\boldsymbol{x})$ and $v^\phi := v^\phi(\boldsymbol{x})$ in the following, remembering that v^μ and v^ϕ are functions of $\boldsymbol{x}$. Noethers first theorem says that

the energy E in eq. (B.19) is invariant upon the action of the r-dimensional Lie group $\mathbb{G}$

$$\frac{d}{dt}\left(g_{\gamma(t)}\circ E\right)\bigg|_{t=0} = 0 \tag{B.21}$$

if and only if there exists r vector valued functions $\boldsymbol{W}_m(\boldsymbol{x})$ such that the following r relations hold

$$\mathrm{Div}\left(\boldsymbol{W}_m\right) + \left(\omega_m^\phi - \omega_m^{\Omega,\mu}\partial_\mu\phi\right)[\mathcal{E}] = 0, \quad 1 \le m \le r \tag{B.22}$$

where $[\mathcal{E}]$ is the Euler-Lagrange differential

$$[\mathcal{E}] = \frac{\delta\mathcal{E}}{\delta\phi} - \mathrm{Div}\left(\boldsymbol{P}\right), \quad \mathrm{Div}\left(\boldsymbol{P}\right) = \sum_{i=1}^r \frac{d}{dx^\mu}\left(\omega_{\Omega,i}^\mu P^i\right), \quad P^i = \frac{\delta\mathcal{E}}{\delta X_e^{\Omega,i}\phi} \tag{B.23}$$

and $\boldsymbol{P}$ is the canonical momentum of the energy in eq. (B.19). We now want to prove that eq. (B.22) follows from eq. (B.21) thereby calculating the analytical form of the r vector valued functions $\boldsymbol{W}_m(\boldsymbol{x})$.

Proof: The derivative of the one parameter group $g_{\gamma(t)}$ with respect to t in eq. (B.21) yields the operator V_e acting on the energy density $\mathcal{E}$ and the volume form d^2x. Since $\mathcal{E}$ does not directly depend on the coordinates $\boldsymbol{x}$, the partial derivatives vanish, $v^\mu\partial_\mu\mathcal{E} = 0$. Hence we have

$$\frac{d}{dt}\left(g_{\gamma(t)}\circ E\right)\bigg|_{t=0} = \int\left\{v^\phi\frac{\delta\mathcal{E}}{\delta\phi} + \sum_{i=1}^r\left(X_e^{\Omega,i}\left(v^\phi\right)\right)P^i\right. \tag{B.24}$$

$$\left.+ \sum_{i=1}^r\left(\left[V_e^\Omega, X_e^{\Omega,i}\right]\phi\right)P^i + \mathcal{E}\frac{dv^\mu}{dx^\mu}\right\}d^2x \tag{B.25}$$

In eq. (B.24) the coordinates $\boldsymbol{x}$ and the GRF ϕ are treated as independent variables, since the subgroups $\mathbb{G}_\phi$ and $\mathbb{G}_\Omega$ are independent from each other. The summand in eq. (B.25) proportional to $\boldsymbol{P}$ comes from the chain-rule and eq. (B.13) while $\frac{dv^\mu}{dx^\mu}d^2x$ is the rate of change of the volume form d^2x under the action of $g_{\gamma(t)}$. We add zeros in the form

$$v^\mu\partial_\mu\phi\frac{\delta\mathcal{E}}{\delta\phi} - v^\mu\partial_\mu\phi\frac{\delta\mathcal{E}}{\delta\phi} = 0 \tag{B.26}$$

$$\sum_{i=1}^r\left(V_e^\Omega\left(X_e^{\Omega,i}\phi\right)\right)P^i - \sum_{i=1}^r\left(V_e^\Omega\left(X_e^{\Omega,i}\phi\right)\right)P^i = 0 \tag{B.27}$$

to eq. (B.25) and use

$$\left[V_e^\Omega, X_e^{\Omega,i}\right] + X_e^{\Omega,i}\left(v^\phi\right) - V_e^\Omega\left(X_e^{\Omega,i}\phi\right) = X_e^{\Omega,i}\left(v^\phi - V_e^\Omega\phi\right) \tag{B.28}$$

After some reordering of terms in eq. (B.25) and factorizing out $(v^\phi - V_e^\Omega\phi)$ we get

$$\frac{d}{dt}\left(g_{\gamma(t)} \circ E\right)\bigg|_{t=0} = \int \left\{ v^\mu \partial_\mu \phi \frac{\delta\mathcal{E}}{\delta\phi} + \sum_{i=1}^r \left(v^\mu \partial_\mu \left(X_e^{\Omega,i}\phi\right)\right) P^i \right. \tag{B.29}$$

$$\left. + \left(v^\phi - v^\mu\partial_\mu\phi\right)\frac{\delta\mathcal{E}}{\delta\phi} + \sum_{i=1}^r X_e^{\Omega,i}\left(v^\phi - V_e^\Omega\phi\right)P^i + \mathcal{E}\frac{dv^\mu}{dx^\mu}\right\} d^2x \tag{B.30}$$

The summands in eq. (B.29) can be combined to the total derivative acting on the energy density $\mathcal{E}$

$$v^\mu\partial_\mu\phi\frac{\delta\mathcal{E}}{\delta\phi} + \sum_{i=1}^r\left(v^\mu\partial_\mu\left(X_e^{\Omega,i}\phi\right)\right)P^i = v^\mu\frac{d\mathcal{E}}{dx^\mu} \tag{B.31}$$

and we can combine eq. (B.31) with the term $\mathcal{E}\frac{dv^\mu}{dx^\mu}$ from eq. (B.30) to the total divergence

$$v^\mu\frac{d\mathcal{E}}{dx^\mu} + \mathcal{E}\frac{dv^\mu}{dx^\mu} = \frac{d}{dx^\mu}\left(v^\mu\mathcal{E}\right) \tag{B.32}$$

so that eq. (B.30) and eq. (B.31) simplify to

$$\frac{d}{dt}\left(g_{\gamma(t)} \circ E\right)\bigg|_{t=0} = \int \left\{ \frac{d}{dx^\mu}\left(v^\mu\mathcal{E}\right) + \left(v^\phi - v^\mu\partial_\mu\phi\right)\frac{\delta\mathcal{E}}{\delta\phi} \right. \tag{B.33}$$

$$\left. + \sum_{i=1}^r X_e^{\Omega,i}\left(v^\phi - V_e^\Omega\phi\right)P^i \right\} d^2x \tag{B.34}$$

We expand the operator $X_e^{\Omega,i}$ in terms of the Cartesian gradient ∇, $X_e^{\Omega,i} = \omega_{\Omega,i}^\mu \partial_\mu$. Then we can convert the summand in eq. (B.34) into the sum of a total divergence and a summand proportional to the divergence of the canonical momentum $\boldsymbol{P}$

$$\sum_{i=1}^r X_e^{\Omega,i}\left(v^\phi - V_e^\Omega\phi\right)P^i = \sum_{i=1}^r \frac{d}{dx^\mu}\left(\omega_{\Omega,i}^\mu\left(v^\phi - V_e^\Omega\phi\right)P^i\right)$$

$$- \left(v^\phi - V_e^\Omega\phi\right)\mathrm{Div}\left(\boldsymbol{P}\right), \quad \mathrm{Div}\left(\boldsymbol{P}\right) = \sum_{i=1}^r\frac{d}{dx^\mu}\left(\omega_{\Omega,i}^\mu P^i\right) \tag{B.35}$$

We substitute eq. (B.34) with the right hand side of eq. (B.35) and group all the total divergences together as well as the term proportional to $(v^\phi - V_e^\Omega \phi)$

$$\frac{d}{dt}\left(g_{\gamma(t)} \circ E\right)\Big|_{t=0} = \int \left\{ \frac{d}{dx^\mu}\left(v^\mu \mathcal{E} + \sum_{i=1}^r \omega_{\Omega,i}^\mu \left(v^\phi - V_e^\Omega \phi\right) P^i\right) \right.$$
$$\left. + \left(v^\phi - V_e^\Omega \phi\right)[\mathcal{E}]\right\} d^2 x \tag{B.36}$$

$$[\mathcal{E}] = \frac{\delta \mathcal{E}}{\delta \phi} - \text{Div}\,(\boldsymbol{P}) \tag{B.37}$$

If eq. (B.21) holds for the specific path $g_{\gamma(t)}$ with the tangential vector operator V_e then it follows that the integrand of the right hand side of eq. (B.36) must vanish

$$\frac{d}{dx^\mu}\left(v^\mu \mathcal{E} + \sum_{i=1}^r \omega_{\Omega,i}^\mu \left(v^\phi - V_e^\Omega \phi\right) P^i\right) + \left(v^\phi - V_e^\Omega \phi\right)[\mathcal{E}] = 0 \tag{B.38}$$

The statement in eq. (B.21) applies for all one parameter sub groups, therefore we need to expand V_e in terms of the basis $\{X_e^i\}$

$$V_e = \sum_{m=1}^r \alpha_m X_e^m \implies v^\phi = \sum_{m=1}^r \alpha_m \omega_m^\phi, \quad v^{\mu,\Omega} = \sum_{m=1}^r \alpha_m \omega_m^{\mu,\Omega} \tag{B.39}$$

then eq. (B.38) translates to

$$\frac{d}{dt}\left(g_{\gamma(t)} \circ E\right)\Big|_{t=0} = \sum_{m=1}^r \alpha_m \left\{ \frac{d}{dx^\mu}(W_m^\mu) + \left(\omega_m^\phi - X_e^{\Omega,m}\phi\right)[\mathcal{E}]\right\} = 0 \tag{B.40}$$

$$W_m^\mu = \omega_m^{\mu,\Omega}\mathcal{E} + \sum_{i=1}^r \omega_{\Omega,i}^\mu \left(\omega_m^\phi - X_e^{\Omega,m}\phi\right) P^i \tag{B.41}$$

Eq. (B.40) holds for all one parameter sub groups $g_{\gamma(t)} \subset \mathbb{G}$ if the sum vanishes for any configuration of the coefficients α_m. It follows that eq. (B.22) holds for all r vector valued functions $\boldsymbol{W}_m$.

B.3.1. Connection between B_m, W_m and $[\mathcal{E}]$

We want to briefly clarify the relation between the bending operator $B_m = \sum_{i=1}^r P^i \left[X_e^{\Omega,m}, X_e^{\Omega,i}\right]$, the vector valued functions $\boldsymbol{W}_m$ and the Euler-Lagrange differentials $[\mathcal{E}]$. For this we need the rate of change of the energy E (eq. (B.19)) under the one parameter Lie group $g_{\gamma(t)}$ corresponding to the vector V_e (eq. (B.20)).

This rate of change is given by (eq. (B.25))

$$\frac{d}{dt}\left(g_{\gamma(t)} \circ E\right)\Big|_{t=0} = \int \left\{ v^\phi \frac{\delta\mathcal{E}}{\delta\phi} + \sum_{i=1}^{r} \left(X_e^{\Omega,i}\left(v^\phi\right)\right) P^i \right. \tag{B.42}$$

$$\left. + \sum_{i=1}^{r} \left(\left[V_e^\Omega, X_e^{\Omega,i} \right] \phi \right) P^i + \mathcal{E} \frac{dv^\mu}{dx^\mu} \right\} d^2x \tag{B.43}$$

Similar to eq. (B.35) we use the following transformation

$$\sum_{i=1}^{r} X_e^{\Omega,i}\left(v^\phi\right) P^i = \sum_{i=1}^{r} \frac{d}{dx^\mu}\left(\omega_{\Omega,i}^\mu v^\phi P^i \right) - v^\phi \mathrm{Div}\left(\boldsymbol{P}\right) \tag{B.44}$$

to transform eq. (B.42)

$$\frac{d}{dt}\left(g_{\gamma(t)} \circ E\right)\Big|_{t=0} = \int \left\{ v^\phi \left[\mathcal{E}\right] + \sum_{i=1}^{r} \frac{d}{dx^\mu}\left(\omega_{\Omega,i}^\mu v^\phi P^i \right) \right. \tag{B.45}$$

$$\left. + \sum_{i=1}^{r} \left(\left[V_e^\Omega, X_e^{\Omega,i} \right] \phi \right) P^i + \mathcal{E}\frac{dv^\mu}{dx^\mu} \right\} d^2x \tag{B.46}$$

We use eq. (B.39) to express V_e in terms of the basis $\{X_e^i\}$

$$\frac{d}{dt}\left(g_{\gamma(t)} \circ E\right)\Big|_{t=0} = \int \sum_{m=1}^{r} \alpha_m \left\{ \omega_m^\phi \left[\mathcal{E}\right] + \sum_{i=1}^{r} \frac{d}{dx^\mu}\left(\omega_{\Omega,i}^\mu \omega_m^\phi P^i \right) \right.$$

$$\left. + \sum_{i=1}^{r} \left(\left[X_e^{\Omega,m}, X_e^{\Omega,i} \right] \phi \right) P^i + \mathcal{E}\frac{d\omega_{\Omega,m}^\mu}{dx^\mu} \right\} d^2x \tag{B.47}$$

Comparing eq. (B.46) with eq. (B.40) we obtain

$$\boldsymbol{B}_m \phi = \frac{d}{dx^\mu}\left(\widetilde{W}_m^\mu \right) - X_e^{\Omega,m}\phi\left[\mathcal{E}\right] - \mathcal{E}\frac{d\omega_{\Omega,m}^\mu}{dx^\mu} \tag{B.48}$$

$$\boldsymbol{B}_m \phi = \sum_{i=1}^{r} \left(\left[X_e^{\Omega,m}, X_e^{\Omega,i} \right] \phi \right) P^i, \quad \widetilde{W}_m^\mu = W_m^\mu - \sum_{i=1}^{r} \omega_{\Omega,i}^\mu \omega_m^\phi P^i$$

Eq. (B.48) is independent of the intensity variations ω_m^ϕ and hence independent of the entire sub group $\mathbb{G}_\phi$. If the energy E in eq. (B.19) is invariant under the sub group $\mathbb{G}_\Omega$, then it follows that $\boldsymbol{B}_m$ is a level-set operator

$$\boldsymbol{B}_m \phi = 0 \tag{B.49}$$

Eq. (B.49) is valid for all ϕ and poses a constraint on the possible values of the Euler-Lagrange differentials $[\mathcal{E}]$ via eq. (B.48).

148

C. The Bending Algebra

C.1. The curvature operator

Given an energy functional $E(\phi, \boldsymbol{X}_e^\Omega \phi)$ for the GRF ϕ we would like to compute the rate of change of the corresponding Euler-Lagrange differentials $[\mathcal{E}_{tot}]$ under the spatial deformation T_t^B (eq. (4.52))

$$\boldsymbol{x}(t) = T_t^B \circ \boldsymbol{x}, \quad \left.\frac{d\boldsymbol{x}(t)}{dt}\right|_{t=0} = \sum_{m=1}^{r} \beta_m \boldsymbol{B}_m \boldsymbol{x} = \mathbf{B} \cdot \boldsymbol{x} \tag{C.1}$$

We write the rate of change as

$$[\mathcal{E}_{tot}]' = \left.\frac{d}{dt}\left(T_t^B \circ [\mathcal{E}_{tot}]\right)\right|_{t=0} \tag{C.2}$$

where $[\mathcal{E}_{tot}]$ is

$$[\mathcal{E}_{tot}](\phi(\boldsymbol{x})) = \frac{\delta \mathcal{E}^{data}}{\delta \phi(\boldsymbol{x})} - \mathrm{Div}(\boldsymbol{P}) \tag{C.3}$$

$$\mathrm{Div}(\boldsymbol{P}) = \sum_\mu \frac{d}{dx_\mu}\left(\sum_i \omega_{\Omega,i}^\mu P_i\right), \quad P_i = \frac{\delta \mathcal{E}^{prior}}{\delta X_e^{\Omega,i}} \tag{C.4}$$

Since $\mathbf{B}$ in eq. (C.1) is a level-set operator of the GRF ϕ the data energy density $\mathcal{E}^{data}(\phi(\boldsymbol{x}))$ and all its functional derivatives are invariant under the flow in eq. (C.1)

$$\left.\frac{d}{dt}\left(T_t^B \circ \mathcal{E}^{data}\right)\right|_{t=0} = \frac{\delta \mathcal{E}^{data}}{\delta \phi(\boldsymbol{x})}\mathbf{B}\phi = 0 \tag{C.5}$$

Thus $[\mathcal{E}_{tot}]'$ reduces to

$$[\mathcal{E}_{tot}]' = -\left.\frac{d}{dt}\left(T_t^B \circ \mathrm{Div}(\boldsymbol{P})\right)\right|_{t=0} \tag{C.6}$$

In order to compute eq. (C.6) we consider the integral

$$F = \int_\Omega \text{Div}(\boldsymbol{P})\phi(\boldsymbol{x})d^2x \tag{C.7}$$

We compute the rate of change of F in eq. (C.7)

$$\frac{d}{dt}\left(T_t^B \circ F\right)\Big|_{t=0} = \int_\Omega \left[\frac{d}{dt}\left(T_t^B \circ \text{Div}\left(\boldsymbol{P}\right)\right)\Big|_{t=0}\phi + \text{Div}\left(\boldsymbol{P}\right)\mathbf{B}\phi\right]d^2x \tag{C.8}$$

Since $\mathbf{B}\phi = 0$ we have

$$\frac{d}{dt}\left(T_t^B \circ F\right)\Big|_{t=0} = \int_\Omega \frac{d}{dt}\left(T_t^B \circ \text{Div}\left(\boldsymbol{P}\right)\right)\Big|_{t=0}\phi d^2x = \int_\Omega \left[\mathcal{E}_{tot}\right]'\phi d^2x \tag{C.9}$$

The integral in eq. (C.7) is well-defined even in the case when the GRF ϕ contains discontinuities. By the Riesz Representation theorem [11] we can shift the derivatives in eq. (C.7) from $\boldsymbol{P}$ onto the GRF ϕ

$$F = -\int_\Omega \sum_i P_i \omega_{\Omega,i}^\mu \partial_\mu \phi \, d^2x = -\int_\Omega \sum_i P_i X_e^{\Omega,i}\phi d^2x \tag{C.10}$$

where $X_e^{\Omega,i} = \omega_{\Omega,i}^\mu \partial_\mu$ is an element of the commutative Lie algebra $\mathcal{B}$ used to construct the prior $\mathcal{E}^{prior}$. We define the vector field $\mathbf{P}_g$

$$\mathbf{P}_g = \sum_i P_{i,g} \cdot X_g^{\Omega,i}, \quad P_{i,g} = P_i(X_g^{\Omega,i}\phi) \tag{C.11}$$

Then the product $\sum_i P_i X_e^{\Omega,i}$ in eq. (C.10) is the evaluation of the vector field $\mathbf{P}_g$ at the identity $e \in \mathbb{G}$ and eq. (C.10) can be written as

$$F = -\int_\Omega \mathbf{P}_e \phi d^2x \tag{C.12}$$

The rate of change of $\mathbf{P}_g$ under the flow in eq. (C.1) is

$$\frac{d}{dt}\mathbf{P}_{T_t^B}\Big|_{t=0} = [\mathbf{B}, \mathbf{P}_e] \tag{C.13}$$

so that from eq. (C.12) we have

$$\frac{d}{dt}\left(T_t^B \circ F\right)\Big|_{t=0} = -\int_\Omega [\mathbf{B}, \mathbf{P}_e]\phi d^2x \tag{C.14}$$

150

We equate eq. (C.14) with eq. (C.9)

$$\int_\Omega [\mathcal{E}_{tot}]' \, \phi d^2 x = - \int_\Omega \boldsymbol{K} \phi d^2 x, \quad \boldsymbol{K} = [\mathbf{B}, \mathbf{P}_e] \tag{C.15}$$

The operator $\boldsymbol{K}$ is the *curvature operator* and we can see that $[\mathcal{E}_{tot}]'$, the rate of change of the Euler-Lagrange differential $[\mathcal{E}_{tot}]$ under the flow in eq. (C.1), can only be understood as the eigenvalue to the negative curvature operator

$$[\mathcal{E}_{tot}]' \cdot \phi = -\boldsymbol{K}\phi \tag{C.16}$$

The curvature operator $\boldsymbol{K}$ can be expanded in terms of the Cartesian gradient ∇

$$\boldsymbol{K}\phi = \boldsymbol{k}\nabla\phi \tag{C.17}$$

and we define the norm $\|\boldsymbol{K}\|$ as the length of the coefficient vector $\boldsymbol{k}$, $\|\boldsymbol{K}\| = \|\boldsymbol{k}\|$.

C.2. TV Image Denoising, supplementary results

(a) input image   **(b)** ϕ_0 **(c)** $\phi^\star_{ELAA}$ **(d)** $\phi^\star_{BNA}$
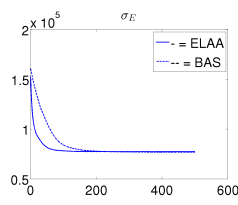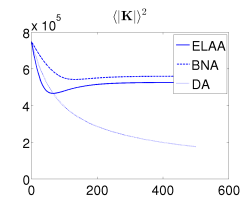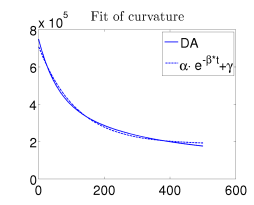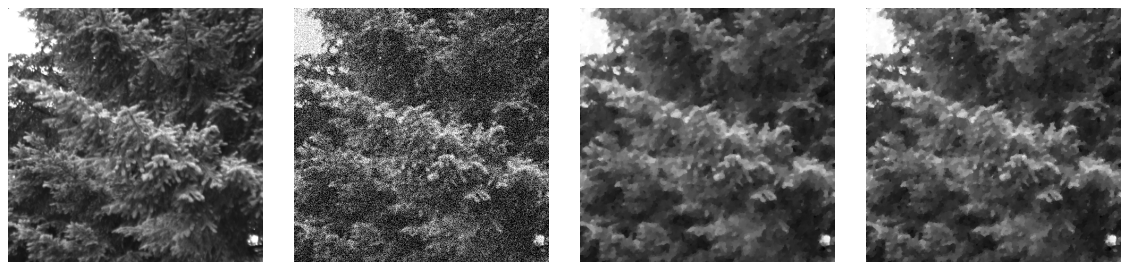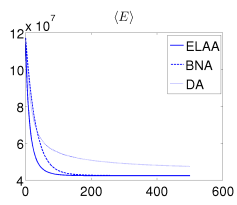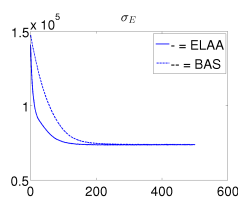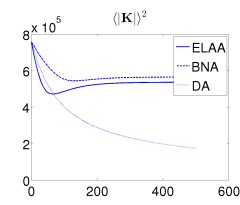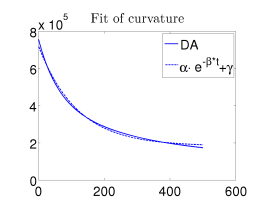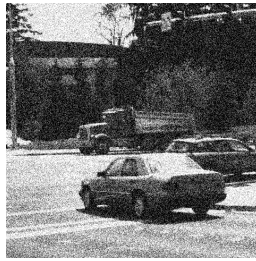
(e)   **(f)**   **(g)**   **(h)**

(i) input image   **(j)** ϕ_0 **(k)** $\phi^\star_{ELAA}$ **(l)** $\phi^\star_{BNA}$

(m)   **(n)**   **(o)**   **(p)**

(a) input image　　**(b)** ϕ_0　　**(c)** $\phi^\star_{ELAA}$　　**(d)** $\phi^\star_{BNA}$

$\langle E \rangle$　　σ_E　　$\langle |\mathbf{K}| \rangle^2$　　Fit of curvature

(e)　　**(f)**　　**(g)**　　**(h)**

(i) input image　　**(j)** ϕ_0　　**(k)** $\phi^\star_{ELAA}$　　**(l)** $\phi^\star_{BNA}$

$\langle E \rangle$　　σ_E　　$\langle |\mathbf{K}| \rangle^2$　　Fit of curvature

(m)　　**(n)**　　**(o)**　　**(p)**

153

(a) input image   **(b)** ϕ_0 **(c)** $\phi^\star_{ELAA}$ **(d)** $\phi^\star_{BNA}$
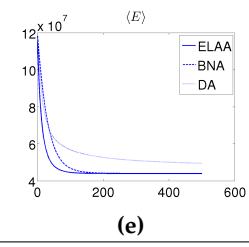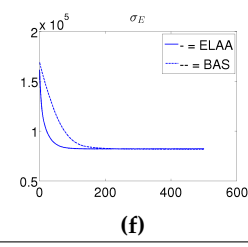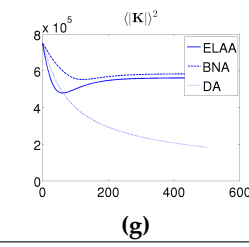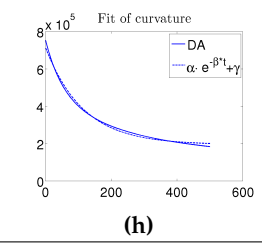
(e)   **(f)**   **(g)**   **(h)**
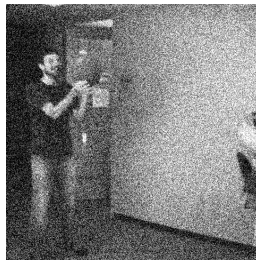
(i) input image   **(j)** ϕ_0 **(k)** $\phi^\star_{ELAA}$ **(l)** $\phi^\star_{BNA}$
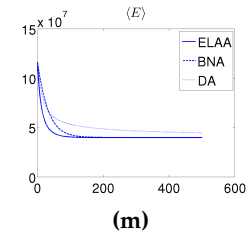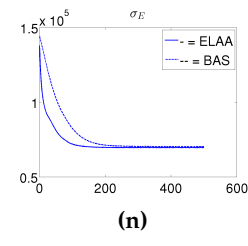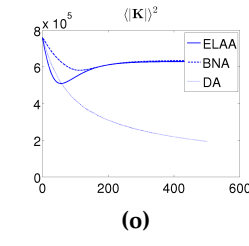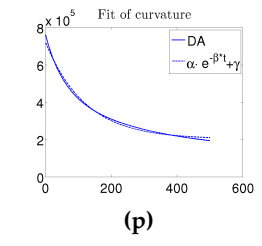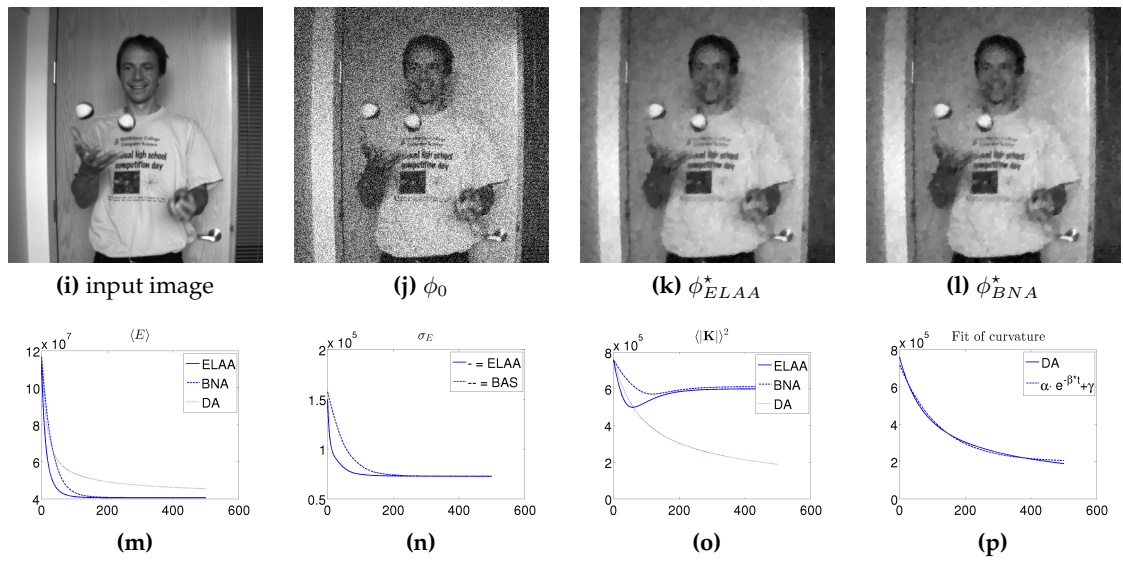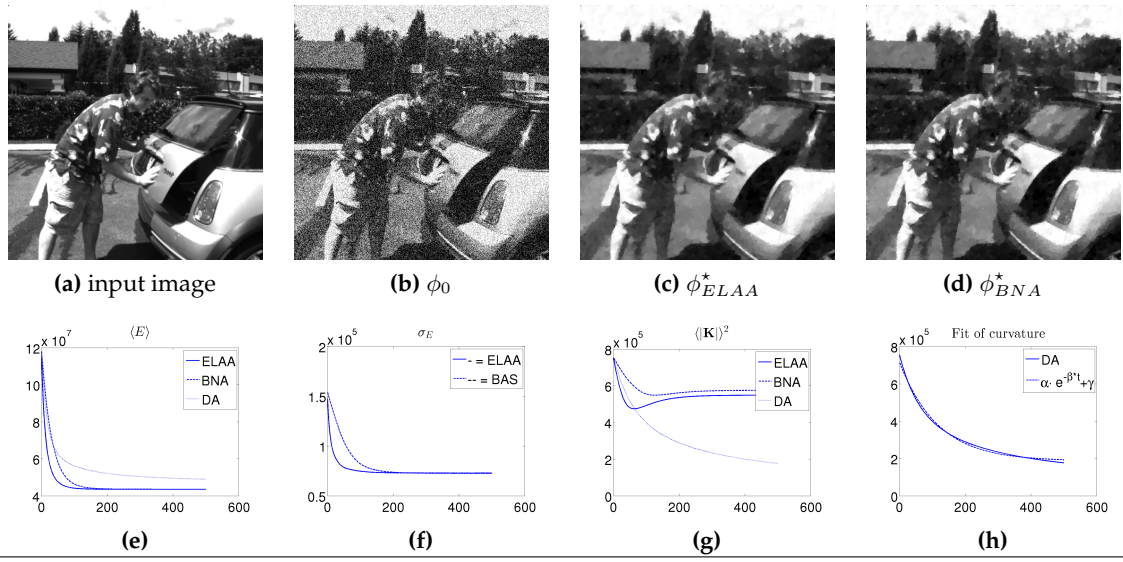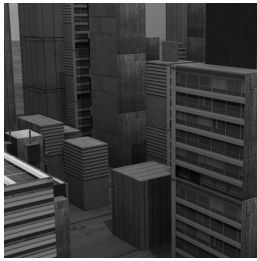
(m)   **(n)**   **(o)**   **(p)**
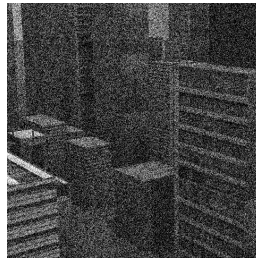
(a) input image   **(b)** ϕ_0 **(c)** $\phi^\star_{ELAA}$ **(d)** $\phi^\star_{BNA}$
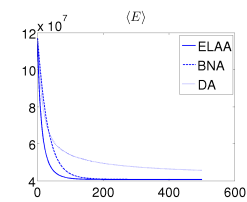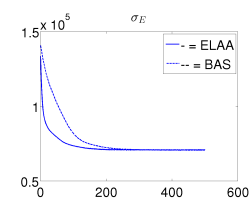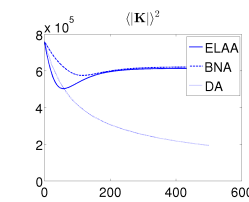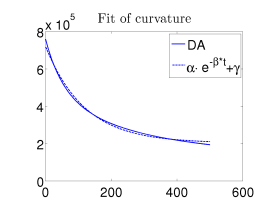
(e)   **(f)**   **(g)**   **(h)**
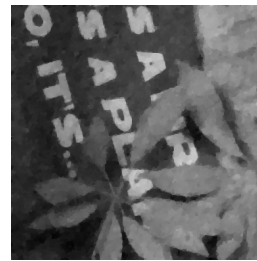
(i) input image   **(j)** ϕ_0 **(k)** $\phi^\star_{ELAA}$ **(l)** $\phi^\star_{BNA}$
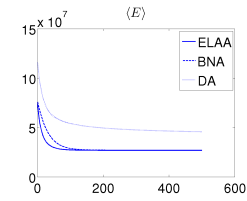
(m)   **(n)**   **(o)**   **(p)**
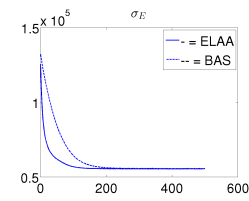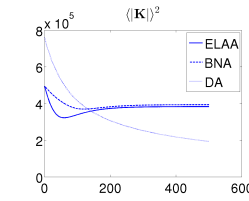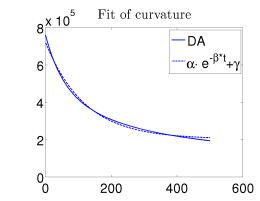
(a) input image   **(b)** ϕ_0 **(c)** $\phi_{ELAA}^{\star}$ **(d)** $\phi_{BNA}^{\star}$
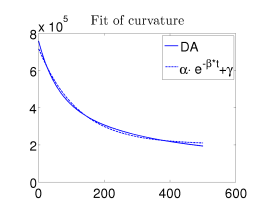

(e)   **(f)**   **(g)**   **(h)**

D. Multimodal Optical Flow

We want to derive the similarity measure $E_{y,I}^{data}$ in a fashion which covers both the global version $E_{y,I}^{data}(\sigma^{sc}, \boldsymbol{d})$ (eq. (3.51)) and the local version $E_{y,I}^{data}(\sigma^{sc}, \boldsymbol{d})$ (eq. (3.56)). We will assume that lower case letters like y stand for low resolution and upper case letters like I stand for high resolution images. The computation deploys the row major lexicographic reordering of images to vectors and filters to (sparse banded) matrices

$$y(\boldsymbol{x}) \to \underline{y}, \quad \int u(\boldsymbol{x})v(\boldsymbol{x})d^2x \to \underline{u}^T \underline{v} \tag{D.1}$$

$$\int_{\mathcal{A}_{\boldsymbol{x}_0}} w(\boldsymbol{x} - \boldsymbol{x}_0) \cdot y(\boldsymbol{x})d^2x \to \mathbf{W}\underline{y} \tag{D.2}$$

To derive the similarity measure $E_{y,I}^{data}$ we considered in section 3.3 a low resolution camera C_y which records the image $\underline{y}$ and a high resolution camera C_I recording the image $\underline{I}$. Since C_y and C_I are physically separated the task was to compute the optical flow $\boldsymbol{d}$ mapping $\underline{I}$ to $\underline{y}$. To handle the difference in optical resolution, $\sigma^{sc,\star}$, between $\underline{y}$ and $\underline{I}$, in section 3.3 the low resolution image $\underline{y}$ is assumed to be result of a sub-sampling process with additive Gaussian noise

$$\underline{y} = \mathbf{W}_{\sigma^{sc}}\underline{Y} + \underline{n}, \quad \underline{n} \sim \mathcal{N}\left(\mathbf{0}, C_n\right) \tag{D.3}$$

where the point-spread function (PSF) $\mathbf{W}_{\sigma^{sc}}$ which models the difference in optical scale is approximately Gaussian. Eq. (D.3) connects the low frequency components of $\underline{Y}$ to the low resolution image $\underline{y}$. From eq. (D.3) the log-likelihood of $\underline{y}$ given $\underline{Y}$ is given by

$$-\ln\left(p\left(\underline{y}|\underline{Y}\right)\right) = \frac{1}{2}\left(\underline{y} - \mathbf{W}_{\sigma^{sc}}\underline{Y}\right)^T \mathbf{C}_n^{-1}\left(\underline{y} - \mathbf{W}_{\sigma^{sc}}\underline{Y}\right) \tag{D.4}$$

To compute the similarity measure $E_{y,I}^{data}$, the correspondence between $\underline{Y}$ and the warped image $\underline{I_d}$ needs to be established. To do this we can compute the conditional distribution $p(\underline{Y}|\underline{I})$: We assume $\boldsymbol{d}$ to be fixed and $\underline{Y}$ and $\underline{I_d}$ to be

jointly Gaussian with the distribution $p(\underline{Y}, \underline{I_d})$

$$-\ln\left(p(\underline{Y}, \underline{I})\right) = \frac{1}{2}\underline{V}^T \mathbf{Q}\underline{V}, \quad \underline{V}^T = \begin{pmatrix} \underline{Y} - \underline{\mu}_Y \\ \underline{I_d} - \underline{\mu}_{I_d} \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} \mathbf{C}_{YY} & \mathbf{C}_{YI_d} \\ \mathbf{C}_{I_dY} & \mathbf{C}_{I_dI_d} \end{pmatrix}^{-1} \quad \text{(D.5)}$$

The covariances $\mathbf{C}_{.,.}$ and the means $\underline{\mu}_{.}$ in eq. (D.5) can be either the global covariance and mean in eq. (2.267) or the local variance and mean in eq. (3.34). In both cases $\mathbf{C}_{.,.}$ is diagonal. The precision matrix $\mathbf{Q}$ can be expressed in terms of the Schur-Complement $\mathbf{S} = \mathbf{C}_{YY} - \mathbf{C}_{YI_d}\mathbf{C}_{I_dI_d}^{-1}\mathbf{C}_{I_dY}$

$$\mathbf{Q} = \begin{pmatrix} \mathbf{S}^{-1} & -\mathbf{S}^{-1}\mathbf{C}_{YI_d}\mathbf{C}_{I_d,I_d}^{-1} \\ \mathbf{C}_{I_d,I_d}^{-1}\mathbf{C}_{I_dY}\mathbf{S}^{-1} & \star \end{pmatrix} \quad \text{(D.6)}$$

then the joint distribution in eq. (D.5) can be split into the product $p(\underline{Y}, \underline{I_d}) = p(\underline{Y}|\underline{I_d})p(\underline{I_d})$ where the conditional distribution $p(\underline{Y}|\underline{I_d})$ can be written in terms of the Schur complement $\mathbf{S}$

$$-\ln\left(p(\underline{Y}|\underline{I_d})\right) = \frac{1}{2}\widetilde{\underline{Y}}^T \mathbf{S}^{-1}\widetilde{\underline{Y}} - \frac{1}{2}\widetilde{\underline{Y}}^T \mathbf{S}^{-1}\mathbf{C}_{YI_d}\mathbf{C}_{I_dI_d}^{-1}\widetilde{\underline{I}}_d - \frac{1}{2}\widetilde{\underline{I}}_d^T \mathbf{C}_{I_dI_d}^{-1}\mathbf{C}_{I_dY}\mathbf{S}^{-1}\widetilde{\underline{Y}}$$
$$\text{(D.7)}$$

$$= \frac{1}{2}\left(\underline{Y} - \underline{\mu}_{Y|I_d}\right)^T \mathbf{C}_{Y|I_d}^{-1}\left(\underline{Y} - \underline{\mu}_{Y|I_d}\right) \quad \text{(D.8)}$$

$$\mathbf{C}_{Y|I_d} = \mathbf{S}, \quad \underline{\mu}_{Y|I_d} = \underline{\mu}_Y + \mathbf{C}_{YI_d}\mathbf{C}_{I_dI_d}^{-1}\widetilde{\underline{I}}_d \quad \text{(D.9)}$$

where $\widetilde{\underline{I}}_d = \underline{I_d} - \underline{\mu}_{I_d}$ and $\widetilde{\underline{Y}} = \underline{Y} - \underline{\mu}_Y$. We combine the conditional $p(\underline{Y}|\underline{I_d})$ from eq. (D.8) together with the likelihood $p\left(\underline{y}|\underline{Y}\right)$ from eq. (D.4) to the posterior $p\left(\underline{Y}|\underline{y}, \underline{I_d}\right) = p\left(\underline{y}|\underline{Y}\right)p(\underline{Y}|\underline{I_d})$ with the energy $E(\underline{Y})$

$$E(\underline{Y}) = \frac{1}{2}\left(\underline{y} - \mathbf{W}_{\sigma^{sc}}\underline{Y}\right)^T \mathbf{C}_n^{-1}\left(\underline{y} - \mathbf{W}_{\sigma^{sc}}\underline{Y}\right)$$
$$+ \frac{1}{2}\left(\underline{Y} - \underline{\mu}_{Y|I_d}\right)^T \mathbf{C}_{Y|I_d}^{-1}\left(\underline{Y} - \underline{\mu}_{Y|I_d}\right) \quad \text{(D.10)}$$

The minimizer $\underline{Y}_d^\star$ of the energy $E(\underline{Y})$ is

$$\underline{Y}_d^\star = \underline{\mu}_{Y|I_d} + \mathbf{C}_{Y|I_d}\mathbf{W}_{\sigma^{sc}}^T \mathbf{H}^{-1}\left(\underline{y} - \mathbf{W}\underline{\mu}_{Y|I_d}\right) \quad \text{(D.11)}$$

$$\mathbf{H} = \mathbf{W}_{\sigma^{sc}}\mathbf{C}_{Y|I_d}\mathbf{W}_{\sigma^{sc}}^T + \mathbf{C}_n \quad \text{(D.12)}$$

We define the similarity measure $E_{y,I}^{data}$ in terms of the conditional distribution $p(\underline{Y}|\underline{I}_{\boldsymbol{d}})$ from eq. (D.8) evaluated at the minimizer $\underline{Y}_{\boldsymbol{d}}^{\star}$

$$E_{y,I}^{data}(\boldsymbol{d}) = -\ln\left(p(\underline{Y}_{\boldsymbol{d}}^{\star}|\underline{I}_{\boldsymbol{d}})\right) \tag{D.13}$$

which after some manipulations becomes

$$E_{y,I}^{data}(\boldsymbol{d}) = \frac{1}{2}\underline{x}^{T}\mathbf{H}^{-1}\mathbf{W}_{\sigma^{sc}}\mathbf{C}_{Y|I_{\boldsymbol{d}}}\mathbf{W}_{\sigma^{sc}}^{T}\mathbf{H}^{-1}\underline{x}, \quad \underline{x} = \underline{y} - \mathbf{W}_{\sigma^{sc}}\underline{\mu}_{Y|I_{\boldsymbol{d}}} \tag{D.14}$$

$E_{y,I}^{data}$ in eq. (D.14) is intractable to compute due to the dense inverse matrix $\mathbf{H}^{-1}$. Similar to [123] we make the simplifications

$$\mathbf{W}_{\sigma^{sc}}\mathbf{C}_{Y|I_{\boldsymbol{d}}}\mathbf{W}_{\sigma^{sc}}^{T} \to \mathbf{C}_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}} \tag{D.15}$$

$$\mathbf{H} \to \widetilde{\mathbf{H}} = \mathbf{C}_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}} + \mathbf{C}_{n} \tag{D.16}$$

and eq. (D.14) reduces to

$$E_{y,I}^{data}(\boldsymbol{d}) = \frac{1}{2}\underline{x}^{T}\widetilde{\mathbf{H}}^{-1}\mathbf{C}_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}\widetilde{\mathbf{H}}^{-1}\underline{x}, \quad \underline{x} = \underline{y} - \mathbf{W}_{\sigma^{sc}}\underline{\mu}_{Y|I_{\boldsymbol{d}}} \tag{D.17}$$

We rewrite eq. (D.17) back in terms of filters and integrals with the inverse lexicographic reordering If we insert the global covariances in eq. (2.267) into the precision matrix $\mathbf{Q}$ in eq. (D.5) we get the global similarity measure from eq. (3.27)

$$E_{y,I}^{data}(\boldsymbol{d}) = \frac{1}{2}\int_{\Omega}\left(y(\boldsymbol{x}) - \mu_{Y} - f\cdot\left(\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}(\boldsymbol{x}) - \mu_{I}\right)\right)^{2}\cdot U^{\sigma^{sc}} \tag{D.18}$$

$$f = C_{y,\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}^{\sigma^{sc}}C_{\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}},\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}^{\sigma^{sc},-1}, \quad U^{\sigma^{sc}} = C_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}^{\sigma^{sc}}\left(C_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}^{\sigma^{sc}} + C_{n}\right)^{-2} \tag{D.19}$$

And if we insert the local covariances in eq. (3.34) into eq. (D.5) we get the local similarity measure $E_{y,I_{\boldsymbol{d}}}^{data,l}$ from eq. (3.38)

$$E_{y,I_{\boldsymbol{d}}}^{data,l}(\sigma^{sc}, a, \boldsymbol{d}) = \frac{1}{2}\int_{\Omega}\left(y(\boldsymbol{x}) - \langle f^{\sigma^{sc},a}I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}(\boldsymbol{x})\right)^{2}\cdot U^{\sigma^{sc},a}(\boldsymbol{x})\,d^{2}x$$

$$U^{\sigma^{sc},a}(\boldsymbol{x}) = C_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}^{\sigma^{sc},a}(\boldsymbol{x})\left(C_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}^{\sigma^{sc},a}(\boldsymbol{x}) + C_{n}\right)^{-2} \tag{D.20}$$

Bibliography

[1] L. Ambrosio and V. M. Tortorelli. Approximation of functional depending on jumps by elliptic functional via t-convergence. *Communications on Pure and Applied Mathematics*, XLIII:999–1036, 1990. 29

[2] K. J. Arrow, L. Hurwicz, H. Uzawa, and H. B. Chenery. Studies in linear and non-linear programming. 1958. 8

[3] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92(1):1–31, 2011. ISSN 1573-1405. doi: 10.1007/s11263-010-0390-2. 71, 87, 106

[4] K. Becker, M. Becker, and J. H. Schwarz. *String theory and M-theory: A modern introduction*. Cambridge University Press, 2006. 119

[5] M. Bhattacharya and A. Das. Registration of multimodality medical imaging of brain using particle swarm optimization. In *Proceedings of the first international conference on intelligent human computer interaction*, pages 131–139. Springer, 2009. 51

[6] J. Bigun. *Vision with Direction A Systematic Introduction to Image Processing and Computer Vision*. Springer, 2006. 39, 86, 116

[7] J. Bigun and G. H. Granlund. Optimal orientation detection of linear symmetry. *Proceedings of the IEEE First International Conference on Computer Vision : London, Great Britain*, pages 433–438, 1987. 61, 86, 116

[8] C. M. Bishop. *Neural networks for pattern recognition*. Oxford university press, 1995. 52

[9] M. J. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer vision and image understanding*, 63(1):75–104, 1996. 51

[10] K. Bredies. A forward–backward splitting algorithm for the minimization of non-smooth convex functionals in banach space. *Inverse Problems*, 25(1): 015005, 2008. 13, 22

[11] K. Bredies and D. Lorenz. Mathematische bildverarbeitung. *Vieweg+ Teubner*, 4(6):12, 2011. 47, 48, 50, 150

[12] T. Bruhn, J. Weickert, and C. Schnörr. Lucas/kanade meets horn/schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, 61,no 3:211–231, 2004. 51

[13] P. Cachier and X. Pennec. 3d non-rigid registration by gradient descent on a gaussian-windowed similarity measure using convolutions. In *Mathematical Methods in Biomedical Image Analysis, 2000. Proceedings. IEEE Workshop on*, pages 182–189. IEEE, 2000. 56

[14] J. B. Campbell and R. H. Wynne. *Introduction to remote sensing*. Guilford Press, 2011. 57

[15] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22 no 1:61–79. 50

[16] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, 20(1-2):89–97, 2004. 7, 8, 29, 47

[17] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*, 40(1):120–145, 2011. 7, 8, 21, 23, 24

[18] C. Chefd'Hotel, G. Hermosillo, and O. Faugeras. A variational approach to multi-modal image matching. In *Variational and Level Set Methods in Computer Vision, 2001. Proceedings. IEEE Workshop on*, pages 21–28. IEEE, 2001. 52, 55

[19] G. H. Chen and R. T. Rockafellar. Convergence rates in forward–backward splitting. *SIAM Journal on Optimization*, 7(2):421–444, 1997. 13

[20] P. L. Combettes. The foundations of set theoretic estimation. *Proceedings of the IEEE*, 81(2):182–208, 1993. 10

[21] P. L. Combettes*. Solving monotone inclusions via compositions of nonexpansive averaged operators. *Optimization*, 53(5-6):475–504, 2004. 13, 22

[22] P. L. Combettes and J.-C. Pesquet. A douglas–rachford splitting approach to nonsmooth convex variational signal recovery. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):564–574, 2007. 13, 22

[23] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In *Fixed-point algorithms for inverse problems in science and engineering*, pages 185–212. Springer, 2011. 10, 22

[24] P. L. M. de Maupertuis. *Loi du repos des corps*. 1740. 24

[25] P. L. M. de Maupertuis. Les loix du mouvement et du repos déduites d'un principe metaphysique. *Hist Acad Roy Sci Belleslett (1746)*, 267:294, 1746. 24

[26] J. Eckstein and D. P. Bertsekas. On the douglas—rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1):293–318, 1992. 13, 22

[27] R. R. Edelman and S. Warach. Magnetic resonance imaging. *New England Journal of Medicine*, 328(10):708–716, 1993. 54

[28] L. Evans and J. Spruck. Motion of level sets by mean curvature. i. In *Fundamental contributions to the continuum theory of evolving phase interfaces in solids*, pages 328–374. Springer, 1999. 50

[29] J. M. Fadili and G. Peyre. Total variation projection with first order schemes. *IEEE Transactions on Image Processing*, 20(3):657–669, Mar. 2011. ISSN 1057-7149. doi: 10.1109/TIP.2010.2072512. 13, 47

[30] Z. Fan, M. H. Santare, and S. G. Advani. Interlaminar shear strength of glass fiber reinforced epoxy composites enhanced with multi-walled carbon nanotubes. *Composites Part A: Applied Science and Manufacturing*, 39 (3):540–554, 2008. 78

[31] O. Faugeras and R. Keriven. *Variational principles, surface evolution, pde's, level set methods and the stereo problem*. IEEE, 2002. 56

[32] M. Fels and P. J. Olver. Moving coframes: I. a practical algorithm. *Acta Applicandae Mathematica*, 51(2):161–213, 1998. 46

[33] M. Fels and P. J. Olver. Moving coframes: Ii. regularization and theoretical foundations. *Acta Applicandae Mathematica*, 55(2):127–208, 1999. 46

[34] M. Ferraro and T. M. Caelli. Relationship between integral transform invariances and lie group theory. *Journal of the Optical Society of America A*, 5 no 5:738–742, 1988. 39

[35] A. D. B. Ferreira, P. R. Nóvoa, and A. T. Marques. Multifunctional material systems: a state-of-the-art review. *Composite Structures*, 151:3–35, 2016. 78

[36] R. P. Feynman. *The principle of least action in quantum mechanics*. PhD thesis, Princeton University Princeton, New Jersey, 1942. 24

[37] R. P. Feynman, R. B. Leighton, and M. Sands. *The Feynman Lectures on Physics*. 1963. ISBN 0201021153, 9780201021158. URL http://www.feynmanlectures.info/. 37

[38] R. P. Feynman, R. B. Leighton, and M. L. Sands. *The Feynman Lectures on Physics: electromagnetism and matter*, volume 2. Addison Wesley Publishing Company, 1964. 25

[39] P. Fieguth. *Statistical image processing and multidimensional modeling*. Springer Science & Business Media, 2010. 6, 13, 93, 117

[40] H. Fischer and H. Kaul. *Mathematik für Physiker*. Springer, 1988. 33

[41] FLIR. URL http://www.flir.com/home/. 58

[42] E. Freitag and R. Busam. *Funktionentheorie 1. 3., neu bearb. und erw. Aufl*. Berlin. 48

[43] O. Güler. On the convergence of the proximal point algorithm for convex minimization. *SIAM Journal on Control and Optimization*, 29(2):403–419, 1991. 12

[44] S.-C. Han and C. I. Podilchuk. Efficient encoding of dense motion fields for motion-compensated video compression. In *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*, volume 1, pages 84–88. IEEE, 1999. 51

[45] S.-C. Han and C. I. Podilchuk. Video compression with dense motion fields. *IEEE Transactions on Image Processing*, 10(11):1605–1612, 2001. 51

[46] R. C. Hardie, K. J. Barnard, J. G. Bognar, E. E. Armstrong, and E. A. Watson. High-resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system. *Optical Engineering*, 37(1):247–260, 1998. 58, 59

[47] R. C. Hardie, M. T. Eismann, and G. L. Wilson. Map estimation for hyperspectral image resolution enhancement using an auxiliary sensor. *IEEE Transactions on Image Processing*, 13(4), 2004. 57, 58, 59, 60, 64, 65

[48] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 3

[49] R. Hentschke. Principle of least action. In *Classical Mechanics*, pages 89–122. Springer, 2017. 24

[50] H. Hertz. *Electric waves: being researches on the propagation of electric action with finite velocity through space*. Dover Publications, 1893. 25

[51] H. Hirschmüller. Stereo processing by semiglobal matching and mutual information. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(2):328–341, 2008. 54

[52] B. K. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17, issues 1-3:185–203, 1981. 51, 52, 53, 57

[53] J. Hsieh. *Computed tomography: principles, design, artifacts, and recent advances*, volume 114. SPIE press, 2003. 54

[54] S. U. Khan and J.-K. Kim. Impact and delamination failure of multiscale carbon nanotube-fiber reinforced polymer composites: a review. *International Journal Aeronautical and Space Sciences*, 12(2):115–133, 2011. 78

[55] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. J. Yezzi. Conformal curvature flows: From phase transitions to active vision. *Archive for rational mechanics and analysis*, 134 issue 3:275–301, 1996. 50

[56] Y. M. Kim, C. Theobalt, J. Diebel, J. Kosecka, B. Miscusik, and S. Thrun. Multi-view image and tof sensor fusion for dense 3d reconstruction. In *Proc. IEEE 12th Int Computer Vision Workshops (ICCV Workshops) Conf*, pages 1542–1549, 2009. doi: 10.1109/ICCVW.2009.5457430. 51

[57] A. A. Kirillov. *An introduction to Lie groups and Lie algebras*, volume 113. Cambridge University Press Cambridge, 2008. 36

[58] C. Kondermann, D. Kondermann, and C. S. Garbe. Postprocessing of optical flows via surface measures and motion inpainting. *Lecture Notes in Computer Science*, 5096:355–364, 2008. 51, 52

[59] T. Kugo. *Eichtheorie*. 1997. doi: 10.1007/978-3-642-59128-0. 37

[60] F. Kuypers. *Klassische Mechanik*. Wiley, 2005. 43

[61] L. Landau and E. Lifshitz. Mechanics, vol. 1. *Course of theoretical physics*, pages 84–93, 1976. 24

[62] L. D. Landau, E. M. Lifshitz, J. B. Sykes, J. S. Bell, and E. Dill. *Electrodynamics of continuous media*. AIP, 1961. 25

[63] J. M. Lee. *Introduction to smooth manifolds*. Shi jie tu shu chu ban gong si Beijing gong si, 2015. 120, 122, 129, 133, 138

[64] P.-L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979. 13, 22

[65] D. G. Lowe. Distictive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004. 52

[66] B. D. Lucas, T. Kanade, et al. An iterative image registration technique with an application to stereo vision. In *IJCAI*, volume 81, pages 674–679, 1981. 51, 52, 53

[67] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE transactions on Medical Imaging*, 16(2):187–198, 1997. 54, 56, 64

[68] E. L. Mansfield. *A practical guide to the invariant calculus*, volume 26. Cambridge University Press, 2010. 34, 35, 43, 46, 142

[69] B. Martinet. Détermination approchée d'un point fixe d'une application pseudo-contractante. *CR Acad. Sci. Paris*, 274(2):163–165, 1972. 12

[70] J. C. Maxwell. On physical lines of force, 1861. *A dynamical theory of the electromagnetic field*, pages 1875–89, 1865. 25

[71] C. Meola, G. M. Carlomagno, A. Squillace, and A. Vitiello. Non-destructive evaluation of aerospace materials with lock-in thermography. *Engineering Failure Analysis*, 13(3):380–388, 2006. 78

[72] B. Mercier. *Topics in finite element solution of elliptic problems*. Springer, 1979. 13

[73] C. W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation*. Macmillan, 1973. 119

[74] J.-J. Moreau. Fonctions convexes duales et points proximaux dans un espace hilbertien. *CR Acad. Sci. Paris Ser. A Math.*, 255:2897–2899, 1962. 7, 10

[75] D. Mumford and J. Shah. Optimal approximation by piecewise smooth functions and associated varia- tional problems. *Comm. Pure and Applied Mathematics*, 42:577–685, 1989. 29, 48

[76] H.-H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (5): 565–593, 1986. 29

[77] R. Nair, F. Lenzen, S. Meister, H. Schäfer, C. Garbe, and D. Kondermann. High accuracy tof and stereo sensor fusion at interactive rates. In *Computer Vision–ECCV 2012. Workshops and Demonstrations*, pages 1–11. Springer Berlin Heidelberg, 2012. 3

[78] T. Netsch, P. Rosch, A. van Muiswinkel, and J. Weese. Towards real-time multi-modality 3-d medical image registration. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1, pages 718–725. IEEE, 2001. 56

[79] I. Newton. *Philosophiae naturalis principia mathematica*, volume 1. excudit G. Brookman; impensis TT et J. Tegg, Londini, 1687. 24

[80] M. Nitzberg and T. Shiota. Nonlinear image filtering with edge and corner enhancement. *IEEE transactions on pattern analysis and machine intelligence*, 14(8):826–833, 1992. 29

[81] E. Noether. Invariante variationsprobleme. *Nachrichten Der Koeniglichen Gesellschaft Der Wissenschaften Zu Goettingen*, Mathematische-Physikalische Klasse:235–257, 1918. 40, 115

[82] E. Noether. Invariant variation problems. *Transport Theory and Statistical Problem*, 1 (3):183–207, 1971. 40, 43, 115

[83] P. J. Olver et al. Symmetry groups and group invariant solutions of partial differential equations. *J. Diff. Geom*, 14:497–542, 1979. 35, 142

[84] S. Osher and L. I. Rudin. Feature-oriented image enhancement using shock filters. *SIAM Journal on Numerical Analysis*, 27(4):919–940, 1990. 47

[85] N. Papenberg, A. Bruhn, T. Brox, S. Didas, and J. Weickert. Highly accurate optic flow computation with theoretically justified warping. *International Journal of Computer Vision*, 67(2):141–158, 2006. 53

[86] N. Parikh, S. Boyd, et al. Proximal algorithms. *Foundations and Trends® in Optimization*, 1(3):127–239, 2014. 13

[87] G. B. Passty. Ergodic convergence to a zero of the sum of monotone operators in hilbert space. *Journal of Mathematical Analysis and Applications*, 72(2):383–390, 1979. 13, 22

[88] M. E. Peskin and D. V. Schroeder. *An Introduction to Quantum Field Theory (Frontiers in Physics)*. Perseus Books, 1995. 24, 37, 46, 119

[89] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the mumford-shah functional. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 1133–1140. IEEE, 2009. 8, 12, 13, 21, 22, 23

[90] M. Pollefeys and L. Koch, R.and Gool. Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters. *International Journal of Computer Vision*, 32(1):7–25, 1999. ISSN 0920-5691. doi: 10.1023/A:1008109111715. 51

[91] L. D. Popov. A modification of the arrow-hurwicz method for search of saddle points. *Mathematical notes of the Academy of Sciences of the USSR*, 28 (5):845–848, 1980. 8

[92] J. Repko and M. Pollefeys. 3d models from extended uncalibrated video sequences: Addressing key-frame selection and projective drift. In *3-D Digital Imaging and Modeling, 2005. 3DIM 2005. Fifth International Conference on*, pages 150–157. IEEE, 2005. 51

[93] A. Roche, G. Malandain, N. Ayache, and X. Pennec. Multimodal image registration by maximization of the correlation ratio. 1998. 55, 56

[94] A. Roche, G. Malandain, X. Pennec, and N. Ayache. The correlation ratio as a new similarity measure for multimodal image registration. In *Medical Image Computing and Computer-Assisted Intervention*, pages 1115–1124. Springer, 1998. 55, 64

[95] A. Roche, G. Malandain, and N. Ayache. *Unifying maximum likelihood approaches in medical image registration*. PhD thesis, Inria, 1999. 54, 55, 64

[96] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM journal on control and optimization*, 14(5):877–898, 1976. 7

[97] R. T. Rockafellar. *Convex analysis*. Princeton university press, 2015. 7, 15, 16

[98] C. Rovelli. *Quantum gravity*. Cambridge university press, 2007. 119

[99] L. I. Rudin. Images, numerical analysis of singularities and shock filters. 1987. 47

[100] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992. 8, 29, 47

[101] H. Rue and L. Held. *Gaussian Markov random fields: theory and applications*. CRC Press, 2005. 59

[102] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, volume 1, 2003. doi: 10.1109/CVPR.2003.1211354. 3, 51

[103] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. Variational methods in imaging, volume 167 of applied mathematical sciences, 2009. 7, 47, 50

[104] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: exploring photo collections in 3d. In *ACM transactions on graphics (TOG)*, volume 25, pages 835–846. ACM, 2006. 51

[105] N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the world from internet photo collections. *International Journal of Computer Vision*, 80(2):189–210, 2008. 51

[106] C. Spiessberger, A. Gleiter, and B. G. Data fusion of lockin-thermography phase images. *9th International Conference on Quantitative InfraRed Thermography*, 2008. 78

[107] D. Sun, S. Roth, and M. J. Black. Secrets of optical flow estimation and their principles. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2432–2439. IEEE, 2010. 5, 51

[108] B. F. Svaiter. On weak convergence of the douglas–rachford method. *SIAM Journal on Control and Optimization*, 49(1):280–287, 2011. 13, 22

[109] P. Tanskanen, K. Kolev, L. Meier, F. Camposeco, O. Saurer, and M. Pollefeys. Live metric 3d reconstruction on mobile phones. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 65–72. IEEE, 2013. 51

[110] M. Tehrani, A. Boroujeni, T. Hartman, T. Haugh, S. Case, and M. Al-Haik. Mechanical characterization and impact damage assessment of a woven carbon fiber reinforced carbon nanotube–epoxy composite. *Composites Science and Technology*, 75:42–48, 2013. 78

[111] G. H. Valadez. *Variational Methods for Multimodal Image Matching*. PhD thesis, École Doctorale Sciences et Technologies de l'Information et de la Communication, 2002. 52, 54, 55, 56

[112] Y. Vardi, L. Shepp, and L. Kaufman. A statistical model for positron emission tomography. *Journal of the American Statistical Association*, 80(389): 8–20, 1985. 54

[113] L. A. Vese and T. F. Chan. A multiphase level set framework for image segmentation using the mumford and shah model. *International Journal of Computer Vision*, 50 no 3:271–293, 2002. 29, 48

[114] L. Wasserman. *All of Statistics, A Concise Course inStatistical Inference*. Springer, 2004. 81

[115] A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers. An improved algorithm for tv-l 1 optical flow. In *Statistical and Geometrical Approaches to Visual Motion Analysis*, pages 23–45. Springer, 2009. 47, 51

[116] J. Weickert. Theoretical foundations of anisotropic diffusion in image processing. *Computing, Suppl.*, 11:221–236, 1996. 29

[117] J. Weickert. *Anisotropic diffusion in image processing*, volume 1. Teubner Stuttgart, 1998. 29

[118] D. Wu and G. Busse. Lock-in thermography for nondestructive evaluation of materials. *Revue générale de thermique*, 37(8):693–703, 1998. 78

[119] A. Yezzi, L. Zöllei, and T. Kapur. A variational framework for joint segmentation and registration. *Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, page 44, 2001. 51

[120] D. C. Youla and H. Webb. Image restoration by the method of convex projections: Part 1theory. *IEEE transactions on medical imaging*, 1(2):81–94, 1982. 10

[121] C. Zach, T. Pock, and H. Bischof. A duality based approach for realtime tv-l 1 optical flow. In *Joint Pattern Recognition Symposium*, pages 214–223. Springer, 2007. 53

[122] C. Zetzsche and E. Barth. Fundamental limits of linear filters in the visual processing of two-dimensional signals. *Vision research*, 30(7):1111–1117, 1990. 52

[123] Y. Zhang. Spatial resolution enhancement for hyperspectral image based on wavelet baysian fusion. *4th International Congress on Image and Signal Processing*, 2011. 60, 159

[124] J. Zhu, L. Wang, R. Yang, and J. Davis. Fusion of time-of-flight depth and stereo for high accuracy depth maps. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition CVPR 2008*, pages 1–8, 2008. doi: 4587761. 3