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:

A Modern Version Of Emmy Noether's First Theorem
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 Optimierung 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.
Zur Behandlung solcher Probleme geben wir eine kleine Einführung in die Theorie der Dualen Optimierung. Innerhalb dieser Theorie spielen das Fenchel Dualitäts Theorem und die Kuhn-Tucker bedingen die zentrale Rolle bei der lösung eines dualen optimierungs Problems.
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 Theorie der dualen Optimierung, 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.
For the treatment of such problems we give a small introduction to the theory of dual optimization. Within this theory, Fenchel's duality theorem and the Kuhn-Tucker conditions play the central role in solving a dual optimization problem.
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 theory of dual optimization 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 5 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 $\boldsymbol{B}$abette, my two daughters $\mathbb{A}$ranka and Emy (thanks for the welcome distraction), my mother Marga and my sisters Sandra and Ingrid, my parents in law $\mathcal{G}$erhard and Elisabeth.

Contents

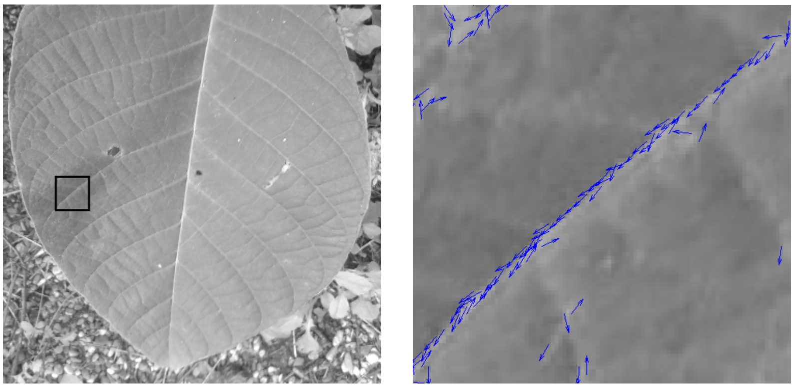
1. Introduction

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)**

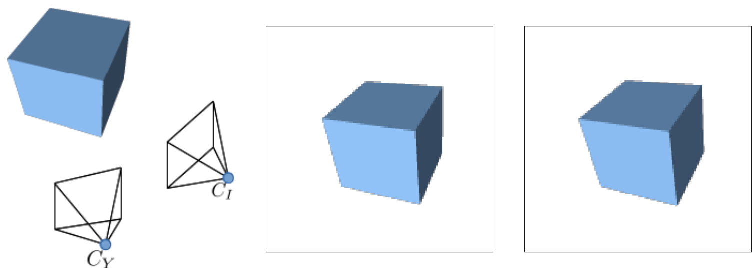
Figure 1.1.: Figure 1.1a shows an image of a leaf. The leaf clearly has no global symmetry. Figure 1.1b shows a close-up of the region around a vein of the leaf, indicated by the box in figure 1.1a. The vectors in figure 1.1b along the vein indicate local translations which leave the vein invariant.

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 valued function $\boldsymbol{b}(\boldsymbol{x})$ 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 ([45, 118, 69, 95]), 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

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

In section 2.2 we will give a brief overview of the topic of *convex optimization* by primal dual splitting [18, 19]. Thereby we introduce the Kuhn-Tucker conditions [88] for the optimum $\phi^\star$ in eq. (1.8) to exist. Sections 2.3 to 2.5 prepare the stage for the introduction of Emmy Noethers celebrated first theorem [73, 74, 11, 71] in section 2.6. 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 $\boldsymbol{W}_m$. 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! In section 3 we show that Noethers Theorem leads to a generalized optimization principle, which not only takes into account the variation of the energy E with respect to the GRF ϕ but also deformations of the domain Ω. We will introduce a criterion for the optimal deformation of Ω based on the curvature of the level-sets of ϕ. Building on section 3 we demonstrate in section 4 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 4.3 we will use the prior developed in section 4 in the context of optical flow [101]. In section 5 we will introduce a generalization of the primal dual algorithm in section 2.2 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

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

called a Gibbs-Random-Field (GRF) [37]. In this thesis we will generally assume that the GRF ϕ is an element of the Hilbert space $\Phi(\Omega)$.

Definition 1 (Hilbert Space)**.** *A Hilbert space $\Phi(\Omega)$ [46, 66, 60] is a set equipped with an scalar product $\langle\cdot,\cdot\rangle$ such that*

- $\langle\phi,\phi\rangle \geq 0$ *holds for all* $\phi \in \Phi(\Omega)$
- $\langle\phi,\phi\rangle = 0 \Leftrightarrow \phi = 0$
- $\langle\cdot,\cdot\rangle$ *is bilinear*

$$\langle h, \alpha\phi + \beta p\rangle = \alpha\left\langle h, f\phi\right\rangle + \beta\left\langle h, p\right\rangle \tag{2.3}$$

$$\langle\alpha\phi + \beta p, h\rangle = \bar{\alpha}\left\langle\phi, h\right\rangle + \bar{\beta}\left\langle p, h\right\rangle, \quad \forall\alpha,\beta \in \mathbb{C}, \phi, p, h \in \Phi(\Omega) \tag{2.4}$$

- $\langle\cdot,\cdot\rangle$ *is Hermitian,* $\langle p,\phi\rangle = \overline{\langle\phi,p\rangle}$

We will assume the functions $\phi \in \Phi(\Omega)$ to be real valued and the scalar product $\langle\cdot,\cdot\rangle$ to be the integral

$$\langle p,\phi\rangle = \int_\Omega p(\boldsymbol{x})\phi(\boldsymbol{x})d^2x \tag{2.5}$$

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. $E^{prior}(\phi, \nabla\phi)$ describes how the elements of C interact with each other. Together both energy functionals form the total energy of the system C

$$E_Y(\phi, \nabla\phi) = E_Y^{data}(\phi) + E^{prior}(\phi, \nabla\phi) \tag{2.6}$$

which is related to the probability distribution

$$p(\phi, \nabla\phi|Y) = p(Y|\phi) \cdot p(\phi, \nabla\phi) \sim \exp\left(-E_Y(\phi, \nabla\phi)\right) \tag{2.7}$$

$$p(Y|\phi) = \exp\left(-E_Y^{data}(\phi)\right) \tag{2.8}$$

$$p(\phi, \nabla\phi) = \exp\left(-E^{prior}(\phi, \nabla\phi)\right) \tag{2.9}$$

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)

$$\phi^\star = \operatorname{argmin}_\phi\left(E_Y(\phi, \nabla\phi)\right) \tag{2.10}$$

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. (2.7) and the solution to the inverse problem

$$\phi^\star = C^{-1}(Y) \tag{2.11}$$

In this thesis we will not be concerned with the probability interpretation of GRF models via eq. (2.7) but rather directly with the formulation of a GRF model via the energy functional $E_Y(\phi, \nabla\phi)$ in eq. (2.6). 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. (2.10) generally fall into two categories: convex and non-convex problems. Convex energy functionals [90, 18, 19] 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 [90, 96]. 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 $\text{prox}\,(\phi|\tau f)$ [66, 89]. 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

$$f(\phi) \geq f(\phi^\star), \quad \forall \phi \in X \tag{2.12}$$

Our general goal is to introduce the method of Pock et. al. [80, 19] for minimizing a class of convex functions of the form $E(\phi, \boldsymbol{A}\phi) = g(\boldsymbol{A}\phi) + f(\phi)$ where $\boldsymbol{A}$ is a linear operator, since the energy functional $E_Y(\phi, \nabla\phi)$ in eq. (2.7) is in this class.

Central to the method of [80] is the concept of *duality*.

Definition 2 (Dual Space)**.** *Let X be a Hilbert space. The dual space $X^\dagger$ is the set of linear maps*

$$L_p : X \to \mathbb{R} \tag{2.13}$$

The Riesz representation theorem [84] states that there exists an element $p \in X$ such that

$$L_p(\phi) = \langle p, \phi \rangle\,, \quad \forall \phi \in X \tag{2.14}$$

Hence $X^\dagger$ is isomorphic to X and we can identify L_p with p.

Definition 3 (Dual Operator)**.** *Let X and Y be two Hilbert spaces and $\boldsymbol{A} : X \to Y$ be a linear operator. The dual operator $\boldsymbol{A}^\dagger : Y^\dagger \to X^\dagger$ maps $Y^\dagger$ to $X^\dagger$ such that*

$$\left\langle \boldsymbol{A}^\dagger \boldsymbol{p}, \phi \right\rangle = \langle \boldsymbol{p}, \boldsymbol{A}\phi \rangle \tag{2.15}$$

holds.

It turns out that every convex function $E(\phi, \boldsymbol{A}\phi)$ on X has a unique *concave* dual conjugate $E^\star(\boldsymbol{p}, \boldsymbol{A}^\dagger\boldsymbol{p})$ on $Y^\dagger$ such that the minimum of E equals the maximum of the conjugate $E^\star$

$$E(\phi^\star, \boldsymbol{A}\phi^\star) = \min_{\phi \in X} E(\phi, \boldsymbol{A}\phi) = \max_{\boldsymbol{p} \in Y^\dagger} E^\star(\boldsymbol{p}, \boldsymbol{A}^\dagger \boldsymbol{p}) = E^\star(\boldsymbol{p}^\star, \boldsymbol{A}^\dagger \boldsymbol{p}^\star) \tag{2.16}$$

Eq. (2.16) is called a saddle-point problem with the pair $(\phi^\star, \boldsymbol{p}^\star)$ being the saddle-point. Thus to obtain the minimizer $\phi^\star$ of E the scheme devised in [80], based on the work of Arrow and Hurrwicz in [2] and Popov in [82], starts from an initial guess $(\phi^0, \boldsymbol{p}^0)$ and progresses to the saddle-point $(\phi^\star, \boldsymbol{p}^\star)$ through alternating steps of *descent* in the primal variable ϕ and *ascent* in the dual variable $\boldsymbol{p}$. The benefits of such a primal-dual approach are that the evaluation of the function $E(\phi, \boldsymbol{A}\phi)$ 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 [93] it was shown in [18] 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 4 (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*

$$dom(f) = \{\phi \in X |\, f(\phi) < \infty\} \tag{2.17}$$

is not empty.

Definition 5 (Lower Semi-Continuous Function)**.** *Let $f : X \to \mathbb{R} \cup \infty$. f is lower semi-continuous if the level-sets*

$$S_\alpha^f = \{\phi \in X |\, f(\phi) \leq \alpha\} \tag{2.18}$$

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

Definition 6 (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*

$$\lambda f(\phi_0) + (1 - \lambda) f(\phi_1) \geq f(\lambda \phi_0 + (1 - \lambda)\phi_1) \tag{2.19}$$

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 $\mathrm{dom}(f)$ of any convex function f is convex itself, since for any $\phi_0, \phi_1 \in \mathrm{dom}(f)$ the image $f(\phi_\lambda)$ of the line segment $\phi_\lambda = \lambda\phi_0 + (1 - \lambda)\phi_1$ is bounded by virtue of eq. (2.19). Hence ϕ_λ is contained in $\mathrm{dom}(f)$ for all $\lambda \in [0, 1]$ and $\mathrm{dom}(f)$ is convex.

Definition 7 (Epigraph). *Let $f : X \to \mathbb{R} \cup \infty$. The epigraph of f is defined as the set*

$$epi(f) = \{(\phi, \mu) \in X \times \mathbb{R} | \, f(\phi) \leq \mu\} \tag{2.20}$$

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 $\phi \in X$

$$f(\phi) \geq f(\phi^\star), \quad \forall \phi \in X \Leftrightarrow \quad \left.\frac{\delta f(\phi)}{\delta \phi}\right|_{\phi = \phi^\star} = 0 \tag{2.21}$$

However eq. (2.21) 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 $\frac{\delta}{\delta \phi} f(\phi) = \frac{\phi}{\|\phi\|}$ does not exist at $\phi = 0$ and we need a broader definition of differentiability which also encompasses the case of non-smooth functions.

Definition 8 (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*

$$f(v) - f(\phi) \geq \langle p, v - \phi \rangle, \quad \forall v \in X \tag{2.22}$$

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*

$$\mathbf{0} \in \partial f(\phi^\star) \tag{2.23}$$

Proof. If eq. (2.23) holds then by setting $p = \mathbf{0}$ in the definition of the subdifferential ∂f in eq. (2.22) we get $f(v) - f(\phi^\star) \geq 0$ for all $v \in X$ hence $\phi^\star$ minimizes f. The reversed argument also holds. □

2.2.1. The Proximal Operator

One of the basic algorithms of non-smooth convex optimization is the Projection Onto Convex Sets (POCS) algorithm ([114, 25, 22]). The problem is stated as follows:

Definition 9 (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*

$$z = P_C(\phi) = \underset{y \in C}{\mathit{argmin}} \left\{ \frac{\|\phi - y\|^2}{2} \right\} \tag{2.24}$$

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

Eq. (2.24) can be rewritten in an unconstrained form

$$z = P_C(\phi) = \underset{y \in X}{\operatorname{argmin}} \left\{ D_C(y, \phi) \right\}, \quad D_C(y, \phi) = \delta_C(y) + \frac{\|\phi - y\|^2}{2} \tag{2.25}$$

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

$$\delta_C(\phi) = \begin{cases} 0 & \phi \in C \\ \infty & otherwise \end{cases} \tag{2.26}$$

The minimization in eq. (2.25) 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 [66] a generalization of the projection operator in eq. (2.25) for the whole class of convex lower semi-continuous functions $f \in \Gamma_0$ was proposed called the *proximal operator*

Definition 10 (Proximal Operator)**.** *Let $f \in \Gamma_0$. The proximal operator is defined as*

$$\mathit{prox}(\phi | f) = \underset{y \in X}{\mathit{argmin}} \left\{ f(y) + \frac{\|\phi - y\|^2}{2} \right\} \tag{2.27}$$

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

Lemma 2 (Uniqueness)**.** *For any $\phi \in X$ the mapping $z = \mathit{prox}(\phi | f)$ 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 $\text{prox}(\phi|f)$ is the unique minimizer. $\square$

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

$$\phi^\star = prox(\phi^\star|f) \tag{2.28}$$

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

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

$$\mathbf{0} \in \partial f(z) + z - \phi \tag{2.29}$$

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. (2.28) 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*

$$\|prox(\phi|f) - prox(y|f)\| \leq \|\phi - y\| \tag{2.30}$$

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*

$$\langle y_1 - y_2, \phi_1 - \phi_2\rangle \geq 0 \tag{2.31}$$

holds.

Proof. By the definition of the subdifferential in eq. (2.22) we have

$$\begin{aligned} y_1 \in \partial f(\phi_1) &\Rightarrow f(v) - f(\phi_1) \geq \langle y_1, v - \phi_1\rangle \,\forall v \in X \\ y_2 \in \partial f(\phi_2) &\Rightarrow f(v') - f(\phi_2) \geq \langle y_2, v' - \phi_2\rangle \,\forall v' \in X \end{aligned} \tag{2.32}$$

Since f is convex the inequalities in eq. (2.32) hold also for all $\phi_1, \phi_2 \in X$. After some manipulations we get

$$\langle y_1, \phi_1 - v\rangle \geq f(\phi_1) - f(v) \tag{2.33}$$

$$\langle -y_2, v' - \phi_2\rangle \geq f(\phi_2) - f(v') \tag{2.34}$$

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

$$\langle y_1 - y_2, \phi_1 - \phi_2 \rangle \geq 0 \tag{2.35}$$

□

proof of proposition 1. Let $\phi_1 = \text{prox}(y_1|f)$ and $\phi_2 = \text{prox}(y_2|f)$. From eq. (2.29) it follows that

$$y_1 - \phi_1 \in \partial f(\phi_1), \quad y_2 - \phi_2 \in \partial f(\phi_2) \tag{2.36}$$

By lemma 4 we have the inequality

$$\langle y_1 - \phi_1 - (y_2 - \phi_2), \phi_1 - \phi_2 \rangle \geq 0 \quad \Rightarrow \quad \langle y_1 - y_2, \phi_1 - \phi_2 \rangle \geq \|\phi_1 - \phi_2\|^2 \tag{2.37}$$

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

$$\|y_1 - y_2\| \geq \|\phi_1 - \phi_2\| \tag{2.38}$$

which proves the proposition. □

The proximal operator is at the center of the proximal point algorithm which was proposed in [62]. Since the proximal point algorithm predates the splitting algorithm in [80] 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

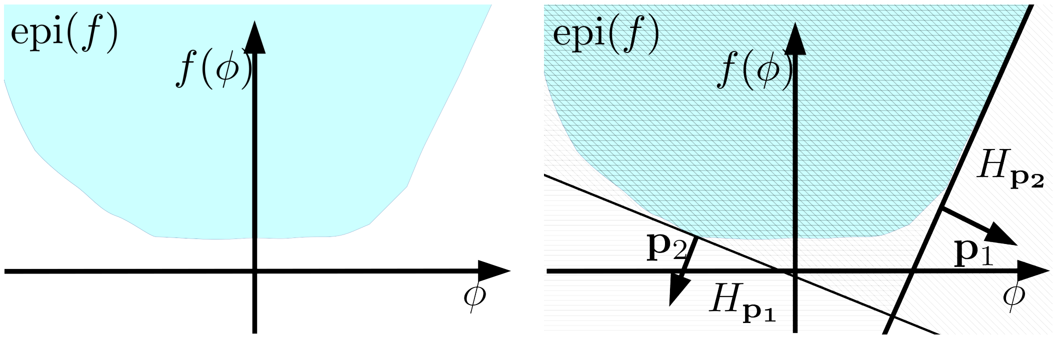
$$\phi_{k+1} = \text{prox}(\phi_k|\tau f) \tag{2.39}$$

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 [40] it was shown that the function values $f(\phi_n)$ satisfy

$$f(\phi_n) - f(\phi^\star) \leq \frac{\|\phi_0 - \phi^\star\|^2}{\sigma_n}, \quad \lim_{n \to \infty} \sigma_n = \infty \tag{2.40}$$

If f is smooth such that $\partial f = \left\{ \frac{\delta}{\delta \phi} f \right\}$ the proximal point iteration in eq. (2.39) reduces to

$$\phi_{k+1} = \phi_k - \tau_k \frac{\delta}{\delta \phi} f(\phi_{k+1}) \tag{2.41}$$



(a) **(b)**

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 $p \in Y^\dagger$.

by setting $z = \phi_{k+1}$ and $\phi = \phi_k$ in eq. (2.29). Eq. (2.41) is called the *backward gradient step* (see [77]) and is similar to the *forward gradient step* defined as

$$\widetilde{\phi}_{k+1} = \phi_k - \tau_k \frac{\delta}{\delta\phi} f(\phi_k) \tag{2.42}$$

also known as steepest descent [37, 77]. The backward gradient step in eq. (2.41) 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. (2.42) 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 [77] 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 [12, 102, 21, 24, 23, 26, 29, 56, 64, 78] 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. [80] 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 $f : X \to \mathbb{R}$ is that of a mapping of *points* $\phi \in X$ to values $f(\phi) \in \mathbb{R}$. In many cases the proximal operator $\operatorname{prox}(\cdot|\tau f)$ can evaluated in an elegant manner by transforming f to a description in which the *slopes* of lines tangent to the epigraph $\operatorname{epi}(f)$ are mapped to the image of f. In figure 2.1a a convex function f and its epigraph $\operatorname{epi}(f)$ are depicted. Since $f \in \Gamma_0$ the epigraph is convex itself. In figure 2.1b the half-planes $H_{p_{1,2}}$ 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}, $\operatorname{epi}(f) \subset H_{p_1} \cap H_{p_2}$. 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 $\operatorname{epi}(f)$ then $\operatorname{epi}(f)$ is identical to the intersection of all such half-planes

$$\operatorname{epi}(f) = \bigcap_{p \in X^\dagger} H_p, \quad \operatorname{epi}(f) \subset H_p \tag{2.43}$$

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

$$h_p(\phi) = \langle p, \phi \rangle - b, \quad p \in X^\dagger, b \in \mathbb{R} \tag{2.44}$$

In eq. (2.44) the constant b is the offset of the line $h_p(\phi)$ at $\phi = 0$. Since $\operatorname{epi}(f)$ is contained in each H_p the function f majorizes the corresponding linear function h_p

$$f(\phi) \geq \langle p, \phi \rangle - b \quad \forall \phi \in X \tag{2.45}$$

Eq. (2.45) imposes a bound on the possible values for the offset b since

$$b \geq \langle p, \phi \rangle - f(\phi) \tag{2.46}$$

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

Definition 11 (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*

$$f^\star(p) = \max_{\phi \in X} \{\langle p, \phi \rangle - f(\phi)\} \tag{2.47}$$

is called the convex conjugate function of f.

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

$$g_\star(p) = \min_{\phi} \{\langle p, \phi \rangle - g(\phi)\} \tag{2.48}$$

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. The function $h = -g$ is convex and its conjugate $h^\star$ is related to the concave conjugate $g_\star$ by*

$$g_\star(p) = -h^\star(-p) \tag{2.49}$$

Proof. By the definition of the conjugate for concave function in eq. (2.48) we have

$$\begin{aligned} g_\star(p) &= \min_x \{\langle p, \phi\rangle - g(\phi)\} = \min_x \{\langle p, \phi\rangle + h(\phi)\} && (2.50)\\ &= \min_x \{-(\langle -p, \phi\rangle - h(\phi))\} = -\max_x \{\langle -p, \phi\rangle - h(\phi)\} && (2.51)\\ &= -h^\star(-p) && (2.52) \end{aligned}$$

□

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. (2.47). From the definition of the convex conjugate $f^\star$ in eq. (2.47) it directly follows that

$$f(\phi) + f^\star(p) \geq \langle p, \phi\rangle \tag{2.53}$$

Eq. (2.53) is called *Fenchel's inequality* [90]. For a concave function g a similar inequality holds:

$$\langle p, \phi\rangle \geq g(\phi) + g_\star(p) \tag{2.54}$$

The following lemma shows an important consequence of the special case of equality in eq. (2.53)

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. (2.47) we have

$$f^\star(p_0) \geq \langle p_0, \phi\rangle - f(\phi), \quad \forall \phi \in X \tag{2.55}$$

Let $f(\phi_0) + f^\star(p_0) = \langle p_0, \phi_0\rangle$ hold. Then from eq. (2.55)

$$\begin{aligned} \langle p_0, \phi_0\rangle = f(\phi_0) + f^\star(p_0) &\geq f(\phi_0) - f(\phi) + \langle p_0, \phi\rangle\,, \quad \forall \phi \in X \\ f(\phi) - f(\phi_0) &\geq \langle p_0, \phi - \phi_0\rangle \quad \forall \phi \in X && (2.56) \end{aligned}$$

By the definition of the subdifferential in eq. (2.22) 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 [90]) and thus

$$f(\phi_0) \geq \langle p, \phi_0 \rangle - f^\star(p), \quad \forall p \in X^\star \tag{2.57}$$

Similar to eq. (2.56) we obtain

$$\begin{aligned} \langle p_0, \phi_0 \rangle = f(\phi_0) + f^\star(p_0) \geq f^\star(p_0) - f^\star(p) + \langle p, \phi_0 \rangle\,, \quad \forall p \in X^\star \\ f^\star(p) - f^\star(p_0) \geq \langle p - p_0, \phi_0 \rangle \quad \forall p \in X^\star \end{aligned} \tag{2.58}$$

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

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

$$f(\phi) - f(\phi_0) \geq \langle p_0, \phi - \phi_0 \rangle \quad \forall \phi \in X \tag{2.59}$$

$$\Rightarrow \langle p_0, \phi_0 \rangle \geq \langle p_0, \phi \rangle - f(\phi) + f(\phi_0) \quad \forall \phi \in X \tag{2.60}$$

We apply the definition of the conjugate function $f^\star$ in eq. (2.47) to eq. (2.60)

$$\langle p_0, \phi_0 \rangle \geq f^\star(p_0) + f(\phi_0) \tag{2.61}$$

Combining eq. (2.61) with Fenchel's inequality in eq. (2.53) 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

$$\mathbf{0} \in \partial f(\phi^\star) \tag{2.62}$$

By lemma 6 the condition in eq. (2.62) 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

$$f(\phi^\star) = -f^\star(\mathbf{0}) \tag{2.63}$$

must hold. Since the conjugate function $f^\star$ is convex its negative is concave. In the following we want to show that eq. (2.63) 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 [88, 90, 97], to be introduced soon.

Similar to the definition of the indicator function in eq. (2.26) we can define a conjugate indicator function

$$\delta^{\alpha}_{C^\star}(p) = \begin{cases} -\alpha & p \in C^\star \\ \infty & otherwise \end{cases} \quad C^\star = \{\mathbf{0}\} \tag{2.64}$$

with finite α and reformulate the right hand side of eq. (2.63) as a maximization over $p \in X^\star$

$$-f^\star(\mathbf{0}) + \alpha = \max_{p\in X^\star} \{-\delta^{\alpha}_{C^\star}(p) - f^\star(p)\} \tag{2.65}$$

such that eq. (2.63) becomes

$$\min_{\phi\in X} \{f(\phi) + \alpha\} = \max_{p\in X^\star} \{-\delta^{\alpha}_{C^\star}(p) - f^\star(p)\} \tag{2.66}$$

It is easy to see that $g_\star(p) = \delta^{\alpha}_{C^\star}(p)$ is the conjugate of the constant function $g(\phi) = \alpha$ and thus eq. (2.66) relates the minimizers of $f(\phi) + \alpha$ to the maximizers of the concave dual function $-\delta^{\alpha}_{C^\star}(p) - f^\star(p)$. In fact as the next theorem due to Rockafellar [88] will show this duality between the minimizers of the convex function $f - g$, 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 $\boldsymbol{A} : X \to Y$ a linear map. Furthermore let*

$$E(\phi, \boldsymbol{A}\phi) = f(\phi) - g(\boldsymbol{A}\phi) \tag{2.67}$$

$$E^\star(\boldsymbol{p}, \boldsymbol{A}^\dagger\boldsymbol{p}) = g_\star(\boldsymbol{p}) - f^\star(\boldsymbol{A}^\dagger\boldsymbol{p}) \tag{2.68}$$

Then

$$\min_{\phi\in X} E(\phi, \boldsymbol{A}\phi) = \max_{\boldsymbol{p}\in Y^\star} E^\star(\boldsymbol{p}, \boldsymbol{A}^\dagger\boldsymbol{p}) \tag{2.69}$$

must hold.

Eq. (2.69) shows that the convex function $E(\phi, \boldsymbol{A}\phi)$ is bounded from below by the maximum of the concave conjugate function $E^\star(\boldsymbol{p}, \boldsymbol{A}^\dagger\boldsymbol{p})$. 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. (2.69) a saddle-point equation and the point $(\phi^\star, \boldsymbol{p}^\star) \in X \times Y^\star$ for which eq. (2.69) is fulfilled is a saddle-point.

Proof. To begin with, we first prove eq. (2.69) for the case $A = \mathbb{1}$. Thus $Y = X$

and eq. (2.69) reduces to

$$\min_{\phi \in X} \{f(\phi) - g(\phi)\} = \max_{p \in X^\star} \{g_\star(p) - f^\star(p)\} \tag{2.70}$$

By Fenchel's inequality in eq. (2.53) we have

$$f(\phi) + f^\star(p) \geq \langle p, \phi \rangle \geq g(\phi) + g_\star(p) \quad \Rightarrow f(\phi) - g(\phi) \geq g_\star(p) - f^\star(p) \tag{2.71}$$

and thus

$$\alpha = \min_x \{f(\phi) - g(\phi)\} \geq \max_p \{g_\star(p) - f^\star(p)\} \tag{2.72}$$

We have to prove that

$$\max_p \{g_\star(p) - f^\star(p)\} \geq \alpha \tag{2.73}$$

Let the sets F and G be the epigraphs

$$F = \{(\phi, \mu) | \, \phi \in X, \mu \geq f(\phi)\} \tag{2.74}$$
$$G = \{(\phi, \mu) | \, \phi \in X, \mu \leq g(\phi) + \alpha\} \tag{2.75}$$

Since F and G are convex sets and $f(\phi) \geq g(\phi) + \alpha$ by the definition of α in eq. (2.72) 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 $h_p(\phi) = \langle p, \phi \rangle - \beta$

$$H = \{(\phi, \mu) | \, \phi \in X, \mu \geq h_p(\phi)\} \tag{2.76}$$

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

$$f(\phi) \geq h_p(\phi) \geq g(\phi) + \alpha \tag{2.77}$$

The left inequality yields

$$\beta \geq \langle p, \phi \rangle - f(\phi) \quad \Rightarrow \beta \geq f^\star(p) \tag{2.78}$$

where eq. (2.78) follows from the definition of the conjugate $f^\star$ in eq. (2.47). The right inequality in eq. (2.77) yields in combination with eq. (2.78)

$$\langle p, \phi \rangle - g(\phi) \geq \beta + \alpha \tag{2.79}$$
$$\Rightarrow g_\star(p) \geq \beta + \alpha \geq f^\star(p) + \alpha \tag{2.80}$$

From eq. (2.80) we conclude

$$\max_{p} \{g_\star(p) - f^\star(p)\} \geq \alpha \tag{2.81}$$

which in combination with eq. (2.72) proves eq. (2.70).

For the general case in eq. (2.69) let $\widetilde{g}(\phi) = g(\boldsymbol{A}\phi)$. Then we can compute the conjugate $\widetilde{g}_\star(p)$ by finding an alternative representation of $-g(\boldsymbol{A}\phi)$

$$-g(\boldsymbol{A}\phi) = -\max_{y\in Y} \left\{ g(y) + \min_{\boldsymbol{p}_y \in Y^\star} \langle \boldsymbol{p}_y, \boldsymbol{A}\phi - y\rangle \right\} \tag{2.82}$$

$$= \min_{y\in Y} \left\{ -g(y) - \min_{\boldsymbol{p}_y \in Y^\star} \langle \boldsymbol{p}_y, \boldsymbol{A}\phi - y\rangle \right\} \tag{2.83}$$

$$= \min_{y\in Y} \max_{\boldsymbol{p}_y \in Y^\star} \{ -g(y) - \langle \boldsymbol{p}_y, \boldsymbol{A}\phi - y\rangle \} \tag{2.84}$$

then

$$\widetilde{g}_\star(p) = \min_{\phi'} \{\langle p, \phi'\rangle - \widetilde{g}(\phi')\} = \min_{\phi'} \{\langle p, \phi'\rangle - g(\boldsymbol{A}\phi')\} \tag{2.85}$$

$$= \min_{y\in Y} \max_{\boldsymbol{p}_y \in Y^\star} \min_{\phi' \in X} \{\langle p, \phi'\rangle - g(y) - \langle \boldsymbol{p}_y, \boldsymbol{A}\phi' - y\rangle\} \tag{2.86}$$

$$= \min_{y\in Y} \max_{\boldsymbol{p}_y \in Y^\star} \min_{\phi' \in X} \left\{ \left\langle p - \boldsymbol{A}^\dagger \boldsymbol{p}_y, \phi' \right\rangle - g(y) + \langle \boldsymbol{p}_y, y\rangle \right\} \tag{2.87}$$

$$= \max_{\boldsymbol{p}_y \in Y^\star} \min_{\phi' \in X} \left\{ \left\langle p - \boldsymbol{A}^\dagger \boldsymbol{p}_y, \phi' \right\rangle + g_\star(\boldsymbol{p}_y) \right\} \tag{2.88}$$

In eq. (2.87) we used definition 3 of the dual operator $\boldsymbol{A}^\dagger$. We insert $\widetilde{g}_\star(p)$ into eq. (2.70)

$$\min_{\phi\in X} \{f(\phi) - \widetilde{g}(\phi)\} = \max_{p\in X^\star} \max_{\boldsymbol{p}_y \in Y^\star} \min_{\phi' \in X} \left\{ \left\langle p - \boldsymbol{A}^\dagger \boldsymbol{p}_y, \phi' \right\rangle + g_\star(\boldsymbol{p}_y) - f^\star(p) \right\} \tag{2.89}$$

and carry out the maximization over p and minimization over ϕ'. These two operations fix the constraint $p = \boldsymbol{A}^\dagger \boldsymbol{p}_y$ thus we get

$$\min_{\phi\in X} \{f(\phi) - g(\boldsymbol{A}\phi)\} = \max_{\boldsymbol{p}_y \in Y^\star} \left\{ g_\star(\boldsymbol{p}_y) - f(\boldsymbol{A}^\dagger \boldsymbol{p}_y) \right\} \tag{2.90}$$

□

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

$$\min_{\phi\in X} \{f(\phi) + g(\boldsymbol{A}\phi)\} = \max_{\boldsymbol{p}\in Y^\star} \left\{ -g^\star(-\boldsymbol{p}) - f^\star(\boldsymbol{A}^\dagger \boldsymbol{p}) \right\} \tag{2.91}$$

The next proposition shows that the saddle-points $(\phi^\star, \boldsymbol{p}^\star)$ of the saddle-point equation in eq. (2.69) relate the subdifferentials of f and $g_\star$ to one another.

Proposition 2 (Kuhn-Tucker conditions)**.** *Let eq. (2.69) be satisfied for a saddle-point $(\phi^\star, \boldsymbol{p}^\star)$. Then the subdifferential equations*

$$\boldsymbol{A}^\dagger \boldsymbol{p}^\star \in \partial f(\phi^\star), \quad \boldsymbol{A}\phi^\star \in \partial g_\star(\boldsymbol{p}^\star) \tag{2.92}$$

$$\phi^\star \in \partial f^\star(\boldsymbol{A}^\dagger \boldsymbol{p}^\star), \quad \boldsymbol{p}^\star \in \partial g(\boldsymbol{A}\phi^\star) \tag{2.93}$$

hold.

Proof. Let $(\phi^\star, \boldsymbol{p}^\star)$ be a saddle-point of Fenchel's duality theorem in eq. (2.69). It follows that

$$f(\phi^\star) + f^\star(\boldsymbol{A}^\dagger \boldsymbol{p}^\star) = g_\star(\boldsymbol{p}^\star) + g(\boldsymbol{A}\phi^\star) \tag{2.94}$$

Eq. (2.94) simultaneously imposes equality in eqns. (2.53) and (2.54)

$$f(\phi^\star) + f^\star(\boldsymbol{A}^\dagger \boldsymbol{p}^\star) = \langle \boldsymbol{p}^\star, \boldsymbol{A}\phi^\star \rangle \tag{2.95}$$

$$g(\boldsymbol{A}\phi^\star) + g_\star(\boldsymbol{p}^\star) = \langle \boldsymbol{p}^\star, \boldsymbol{A}\phi^\star \rangle \tag{2.96}$$

and the conditions in eq. (2.92) follow by applying lemma 6 to eqns. (2.95) and (2.96). The conditions in eq. (2.93) follow directly from lemma 6. □

Corollary 1 (Convex Kuhn-Tucker conditions)**.** *For functionals $E(\phi, \boldsymbol{A}\phi) = f(\phi) + g(\boldsymbol{A}\phi)$ where both f and g are convex the Kuhn-Tucker conditions are obtained from eqns. (2.92) and (2.93) by the substitutions $\boldsymbol{p}^\star \to -\boldsymbol{p}^\star$, $g \to -g$ and $g_\star \to -g^\star$*

$$\begin{aligned} -\boldsymbol{A}^\dagger \boldsymbol{p}^\star \in \partial f(\phi^\star), &\quad -\boldsymbol{A}\phi^\star \in \partial g^\star(-\boldsymbol{p}^\star) \\ \phi^\star \in \partial f^\star(-\boldsymbol{A}^\dagger \boldsymbol{p}^\star), &\quad \boldsymbol{p}^\star \in \partial g(\boldsymbol{A}\phi^\star) \end{aligned} \tag{2.97}$$

With the help of the Kuhn-Tucker conditions in eq. (2.92) we can formulate a condition for the minimizer $\phi^\star$ alone which resembles the condition in $0 \in \partial f(\phi^\star)$ (eq. (2.23)) but for the function $E(\phi, \boldsymbol{A}\phi)$: Since the Kuhn-Tucker conditions in eq. (2.92) hold for the saddle-point $(\phi^\star, \boldsymbol{p}^\star)$ and $\boldsymbol{p}^\star \in \partial g(\boldsymbol{A}\phi^\star)$ (by lemma 6) we have

$$0 = \boldsymbol{A}^\dagger \boldsymbol{p}^\star - \boldsymbol{A}^\dagger \boldsymbol{p}^\star \in \partial f(\phi^\star) - \boldsymbol{A}^\dagger \partial g(\boldsymbol{A}\phi^\star) = \partial E(\phi^\star, \boldsymbol{A}\phi^\star) \tag{2.98}$$

where $\boldsymbol{A}^\dagger \partial g(\boldsymbol{A}\phi^\star)$ implies that $\boldsymbol{A}^\dagger$ is applied to every element of the set $\partial g(\boldsymbol{A}\phi^\star)$. Eq. (2.98) states that the minimizer $\phi^\star$ of $E(\phi, \boldsymbol{A}\phi)$ 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*

$$\min_{\phi \in X} \{f(\phi) + w(z - \phi)\} + \min_{p \in X^\star} \{f^\star(p) + w(z - p)\} = w(z) \tag{2.99}$$

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

$$g_\star(p) = -w(z - p) + w(z) \tag{2.100}$$

We insert f, g and eq. (2.100) into eq. (2.69) of Fenchel's Duality theorem and directly verify eq. (2.99)

$$\min_{\phi \in X} \{f(\phi) + w(z - \phi)\} = \max_{p \in X^\star} \{-w(z - p) - f^\star(p)\} + w(z) \tag{2.101}$$

$$= -\min_{p \in X^\star} \{w(z - p) + f^\star(p)\} + w(z) \tag{2.102}$$

$$\min_{\phi \in X} \{f(\phi) + w(z - \phi)\} + \min_{p \in X^\star} \{w(z - p) + f^\star(p)\} = w(z) \tag{2.103}$$

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

$$\bar{p} \in \partial f(\bar{\phi}), \quad \bar{\phi} \in \partial g_\star(\bar{p}) \tag{2.104}$$

However since $g_\star$ is smooth we have

$$\bar{\phi} = \nabla g_\star(p)|_{p=\bar{p}} = z - \bar{p} \tag{2.105}$$

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

From from the decomposition eq. (2.105) and the definition of the proximal operator in eq. (2.27) it follows that

$$z = \text{prox}\,(z|f) + \text{prox}\,(z|f^\star) \tag{2.106}$$

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

Algorithm 1 Chambolle Pock Primal-Dual Splitting

Choose parameters τ, $\sigma > 0$, $\theta \in (0,1)$ and initial guesses $\bar{\phi}^0$, $\boldsymbol{p}^0$ and $\phi^0 = \bar{\phi}^0$
while $G(\phi^n, \boldsymbol{p}^n) > \epsilon$ or $n < M$ **do**
  Ascending step: $\boldsymbol{p}^{n+1} = \text{prox}(\boldsymbol{p}^n + \sigma \boldsymbol{A}\phi^n | \sigma g^\star)$
  Descending step: $\bar{\phi}^{n+1} = \text{prox}(\bar{\phi}^n - \tau \boldsymbol{A}^\dagger \boldsymbol{p}^{n+1} | \tau f)$
  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 $\boldsymbol{p}$ in the direction which maximizes the dual function $E^\star(\boldsymbol{p}, \boldsymbol{A}^\dagger \boldsymbol{p})$ and the descending step moves the primal variable ϕ in the direction which minimizes the primal function $E(\phi, \boldsymbol{A}\phi)$. Since $\bar{\phi}^{n+1}$ is computed from $\boldsymbol{p}^{n+1}$ 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.

2.2.3. Primal Dual Splitting

In this section we will introduce the splitting method first described by Pock et. al in [80] and refined in [19] which is a primal-dual method for solving the saddle-point problem in eq. (2.69) for a given convex function $E(\phi, \boldsymbol{A}\phi) = g(\boldsymbol{A}\phi) + f(\phi)$. Splitting methods generally include an iterative strategy in which 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 [56] and further developed in [12, 23, 78]. The forward-backward algorithm combines a forward-gradient step involving g (eq. (2.42)) with a proximal step (eq. (2.39)) involving f

$$\phi^{n+1} = \text{prox}\left(\phi^n - \tau \frac{\delta}{\delta\phi} g(\boldsymbol{A}\phi^n) \middle| \sigma f\right) \tag{2.107}$$

The iteration in eq. (2.107) however is based upon the assumption that the function g is smooth while the function f is possibly non-smooth. The Douglas-Rachford algorithm [56, 26, 24, 102] relaxes the smoothness criterion on g by introducing an auxiliary variable v and the update rules

$$\lambda^n \in [\epsilon, 2 - \epsilon] \tag{2.108}$$
$$v^{n+1} = v^n + \lambda^n \left(\text{prox}\left(2\phi^n - v^n | \tau g\right) - \phi^n\right)$$
$$\phi^{n+1} = \text{prox}\left(v^{n+1} | \tau f\right) \tag{2.109}$$

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 [25] for an overview of both algorithms and their applications.

The splitting algorithm which was proposed by Pock et. al. in [80] 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. (2.47), $g = g^{\star\star}$ and thus

$$g(\boldsymbol{A}\phi) = \max_{\boldsymbol{p}\in Y^\star} \{\langle \boldsymbol{p}, \boldsymbol{A}\phi\rangle - g^\star(\boldsymbol{p})\} \tag{2.110}$$

Thus the minimization problem

$$\min_{\phi\in X} E(\phi, \boldsymbol{A}\phi) = \min_{\phi\in X} \{g(\boldsymbol{A}\phi) + f(\phi)\} \tag{2.111}$$

is equivalent to the saddle-point point problem

$$\min_{\phi\in X} E(\phi, \boldsymbol{A}\phi) = \min_{\phi\in X} \max_{\boldsymbol{p}\in Y^\star} \{\langle \boldsymbol{p}, \boldsymbol{A}\phi\rangle - g^\star(\boldsymbol{p}) + f(\phi)\} \tag{2.112}$$

Hence the authors in [80] propose an algorithm (see alg. 1) which beginning with a pair of initial guesses $(\phi^0, \boldsymbol{p}^0)$ consists of an *ascending* iteration in the dual variable $\boldsymbol{p}$ and a *descending* iteration and subsequent interpolation step in the primal variable ϕ. After computing the dual update $\boldsymbol{p}^{n+1}$ from $\boldsymbol{p}^n$ and ϕ^n the update $\bar{\phi}^{n+1}$ is computed from $\boldsymbol{p}^{n+1}$. Thus $\bar{\phi}^{n+1}$ is always *one step ahead* of $\boldsymbol{p}^{n+1}$. 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 $\boldsymbol{p}$. In the case of interpolation with $\theta = 1$ and $\tau\sigma \left\|\boldsymbol{A}\right\|^2 \leq 1$ convergence of the algorithm is guaranteed [80, 19].

To access the convergence of the algorithm the authors in [80] introduced the *partial primal-dual Gap*

$$G_{X\times Y^\star}(\phi, \boldsymbol{p}) = E(\phi, \boldsymbol{A}\phi) - E^\star(\boldsymbol{A}^\dagger\boldsymbol{p}, \boldsymbol{p}) \tag{2.113}$$

$$E(\phi, \boldsymbol{A}\phi) = \max_{\boldsymbol{p}'\in Y^\star} \{\langle \boldsymbol{p}', \boldsymbol{A}\phi\rangle - g^\star(\boldsymbol{p}') + f(\phi)\} \tag{2.114}$$

$$E^\star(\boldsymbol{p}, \boldsymbol{A}^\dagger\boldsymbol{p}) = \min_{\phi'\in X} \left\{\left\langle \boldsymbol{A}^\dagger\boldsymbol{p}, \phi'\right\rangle - g^\star(\boldsymbol{p}) + f(\phi')\right\} \tag{2.115}$$

From Fenchel's theorem in eq. (2.91) we have

$$\min_{\phi\in X} E(\phi, \boldsymbol{A}\phi) = \max_{\boldsymbol{p}\in Y^\star} E^\star(\boldsymbol{p}, \boldsymbol{A}^\dagger\boldsymbol{p}) \tag{2.116}$$

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

$$G_{X\times Y^\star}(\phi, \boldsymbol{p}) \geq 0 \tag{2.117}$$

The points $(\phi^\star, \boldsymbol{p}^\star)$ for which the primal-dual gap vanishes, $G_{X\times Y^\star}(\phi^\star, \boldsymbol{p}^\star) = 0$ are saddle-points. Given the initial guesses ϕ^0 and $\boldsymbol{p}^0$ the authors in [19] 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

$$G_{X\times Y^\star}(\phi^N, \boldsymbol{p}^N) \leq \frac{D(\phi^0, \boldsymbol{p}^0)}{N}, \quad \phi^N = \frac{1}{N}\sum_{i=1}^{N}\phi^i, \ \boldsymbol{p}^N = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{p}^i \tag{2.118}$$

And in the limit $N \to \infty$ the cluster point $(\phi^N, \boldsymbol{p}^N)$ converges to a saddle-point $(\phi^\star, \boldsymbol{p}^\star)$ of the saddle-point problem in eq. (2.69). Furthermore in the case when X and Y are finite dimensional, $X \times Y \subset \mathbb{R}^q \times \mathbb{R}^r$ the sequence $(\phi^n, \boldsymbol{p}^n)$ itself converges to the same saddle-point $(\phi^\star, \boldsymbol{p}^\star)$. Hence the iterations $(\phi^n, \boldsymbol{p}^n)$ of any implementation of alg. 1 on a computer necessarily converge to an approximate solution of the saddle-point problem in eq. (2.69). An improved version of alg. 1 was additionally proposed in [19] 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 5 we will propose an extension which yields a convergence rate faster then $O(\frac{1}{N})$.

2.3. Image De-Noising

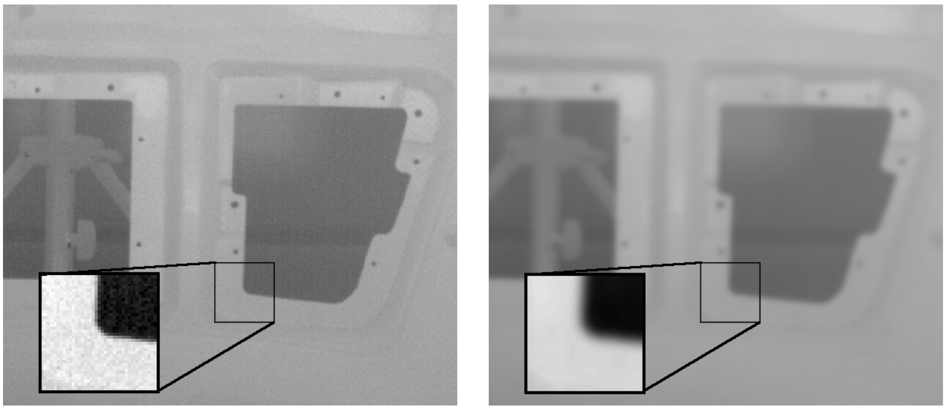
An example of a physical system containing a GRF is a camera C recording an object O. The domain $\Omega \subset \mathbb{R}^2$ is the focal plane of the camera C and the object O is naturally projected onto the focal plane Ω producing the projection I_O. In theory the projection I_O is a continuous function 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 CCD collecting the photons of the light at discrete positions $\boldsymbol{x}_{i,j}$ called pixels

$$I^c_{ij} = I^c(\boldsymbol{x}_{i,j}) \in \mathbb{R}, \quad \boldsymbol{x}_{i,j} \in \Omega \quad 1 < i < n, 1 < j < m \tag{2.119}$$

The observables Y are the recorded intensities I^c_{ij} 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^c_{ij}

$$I^c_{ij} = C_{ij}\left(I_O\right) \tag{2.120}$$

The intensity I^c_{ij} 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



(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. (2.124) with the prior in eq. (2.125). The noise is removed but the boundaries of O are over smoothed

rather the result of a stochastic process described as independently identically distributed (iid) noise

$$\hat{I}_{ij}^c = I_O\left(\boldsymbol{x}_{i,j}\right) + n \quad n \sim p\left(I_{ij}^c | I_O\left(\boldsymbol{x}_{i,j}\right)\right) \tag{2.121}$$

$p(I_{ij}^c|I_O\left(\boldsymbol{x}_{i,j}\right))$ 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. (2.8) 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. (2.121). Such regularity can be achieved by correlating the intensities $I_O(\boldsymbol{x})$ at all pixels with each other in the prior

$$p\left(I_O\right) = \exp\left(-E^{prior}\left(I_O\right)\right) \tag{2.122}$$

$$E^{prior}\left(I_O\right) = \int_\Omega \mathcal{E}\left(I_O\left(\boldsymbol{x}\right), I_O\left(\Omega / \left\{\boldsymbol{x}\right\}\right)\right) dx \tag{2.123}$$

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

$$I_O^\star = \text{argmin}_{I_O}\left(E_{I^c}\left(I_O\right)\right), \quad E_{I^c}\left(I_O\right) = E_{I^c}^{data}(I_O) + E^{prior}\left(\nabla I_O\right) \tag{2.124}$$

However in practice for a $n \times n$ dimensional image I^c the minimization in eq. (2.124) 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. (2.124) are practically infeasible.

To reduce the complexity we want the integrand $\mathcal{E}$ in eq. (2.123) 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 $\nabla I_O(\boldsymbol{x})$

$$E_{L_2}^{prior}\left(\nabla I_O\right) = \int_\Omega \left\|\nabla\, I_O\left(\boldsymbol{x}\right)\right\|^2 dx \tag{2.125}$$

where the gradient operation ∇ can be realized by finite differences. While the prior in eq. (2.125) 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. (2.121). Figure 2.2b shows the result of the minimization in eq. (2.124) with the L_2 prior in eq. (2.123). $E_{L_2}^{prior}$ reduces the noise in I_O but due to its isotropic nature it over-smooths the boundaries of O. In section 2.4 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.

2.4. Lie Groups and the Noether Theorem

2.4.1. Motivation 1, the problem

In section 2.3 we had claimed that the problem with the L_2 prior

$$E_{L_2}\left(\phi\right) = \int_\Omega \left\|\nabla\phi\right\|^2 \tag{2.126}$$

is that it 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 $\phi = const$

$$A_c = \left\{\phi_c^\star \middle|\, \phi_c^\star = \text{argmin}_\phi\left(E_{L_2}\left(\nabla\phi\right)\right) = c, \quad c \in \mathbb{R}\right\} \tag{2.127}$$

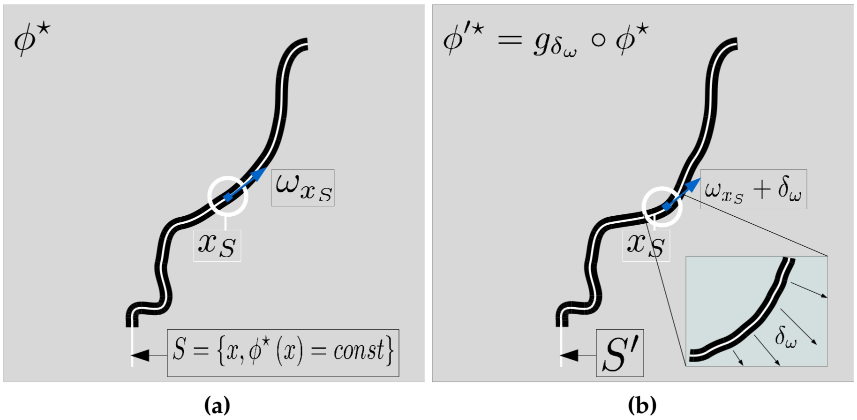


Figure 2.3.: Local transformation of an image ϕ with a level-set S. Figure 2.3a shows an image $\phi(\boldsymbol{x})$ with a line S along which the intensity values are constant. At each point $\boldsymbol{x}_S$ the vector $\boldsymbol{\omega}_S$ is the tangential vector on S. Figure 2.3b shows the result 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 $\boldsymbol{\delta}_\omega(\boldsymbol{x})$

In the context of the minimization problem in eq. (2.124) the minimizer set A_c in eq. (2.127) emphasizes that the prior E_{L_2} does not allow for the solution $I_O^\star$ (eq. (2.124)) 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

$$\boldsymbol{x}' = \mathbf{R}_\theta \boldsymbol{x}, \quad \mathbf{R}_\theta = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \in SO(2) \tag{2.128}$$

The gradient $\nabla\phi$ transforms under the rotation in eq. (2.128) 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. (2.128)

$$E'_{L_2} = \int_\Omega \nabla^T \phi \mathbf{R}_\theta^T \mathbf{R}_\theta \nabla \phi d^2x = \int_\Omega \|\nabla\phi\|^2 d^2x \tag{2.129}$$

In general the invariance of the prior energy $E^{prior}(\nabla\phi)$ of a GRF ϕ with respect to the rotations in eq. (2.128) 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. (2.124) 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 [93, 18] which will be introduced in section 2.7 , anisotropic difusion guided by directional operators like the structure tensor [72, 110, 111] and level set methods of the Mumford-Shah type [67, 1, 107]. Among the earliest attempts for anisotropic regularization was the work of Nagel et. al. [68]. They introduced a quadratic prior

$$E_D^{prior}(\nabla\phi) = \int (\nabla\phi(\boldsymbol{x}))^T D(\boldsymbol{x}) (\nabla\phi(\boldsymbol{x})) d^2x \tag{2.130}$$

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 $\nabla\phi$. For instance in eq. (2.124) we can insert eq. (2.130) 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

$$D(\boldsymbol{x}) = \lambda \boldsymbol{b}(\boldsymbol{x}) \boldsymbol{b}(\boldsymbol{x})^T, \quad \boldsymbol{b}(\boldsymbol{x}_0) \perp \langle \nabla I^c(\boldsymbol{x}) \rangle (\boldsymbol{x}_0) \tag{2.131}$$

$$\langle \nabla I^c(\boldsymbol{x}) \rangle (\boldsymbol{x}_0) = \int_A w(\|\boldsymbol{x} - \boldsymbol{x}_0\|) \nabla I^c(\boldsymbol{x}) d^2x \tag{2.132}$$

the prior E_D^{prior} penalizes the tangential component of ∇I_O along $\boldsymbol{b}$ in the minimization in eq. (2.124). 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. (2.128). Thus the minimization in eq. (2.124) can produce a solution $I_O^\star$ in eq. (2.124) 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. (2.124).

On the other side a potential anisotropic prior E^{prior} which *is* rotation invariant would lead to a solution $I_O^\star$ in eq. (2.124) 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 $\phi^\star = \text{argmin}_\phi E^{prior}(\nabla\phi)$ while remaining invariant to a specified but more general set of spacial transformations $\mathbb{G}^\Omega$.

2.4.2. 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. (2.127)) 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. (2.127) 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

$$g_\cdot : \mathbb{R} \to \mathbb{G}_{const}, \quad g_d = \cdot + d, \quad g_d \in \mathbb{G}_{const} \tag{2.133}$$

With the help of the construction in eq. (2.133) we can restate the invariance of E_{L_2} in the following way

$$g_d \circ E_{L_2}(\nabla\phi) = E_{L_2}(\nabla(\phi + d)) = E_{L_2}(\nabla\phi) \tag{2.134}$$

and A_c in eq. (2.127) can be viewed as being spun by *one* constant function $\phi_0^\phi(\boldsymbol{x}) = c$ and the set $\mathbb{G}_{const}$

$$A_c = \{\phi^\star |\, \phi^\star = g_d \circ \phi_0^\star, \quad g_d \in \mathbb{G}_{const}\} \tag{2.135}$$

With the constructions in eq. (2.133) and eq. (2.135) 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 2.4.1. Hence E^{prior} is assumed to be invariant to the set $\mathbb{G}^{\Omega\phi} = \mathbb{G}^\phi \times \mathbb{G}^\Omega$ with the actions

$$g_{\omega^\phi} \circ \phi(\boldsymbol{x}) = \phi(\boldsymbol{x}) + \omega^\phi(\boldsymbol{x}), \quad g_{\omega^\phi} \in \mathbb{G}^\phi \tag{2.136}$$

$$g_{\omega^\Omega} \circ \boldsymbol{x} = \boldsymbol{x} + \boldsymbol{\omega}^\Omega(\boldsymbol{x}), \quad g_{\omega^\Omega} \in \mathbb{G}^\Omega \tag{2.137}$$

The transformations in eq. (2.136) and eq. (2.137) formally capture all the possible transformations the prior energy E^{prior} is invariant to. In this sense $\mathbb{G}^{\Omega\phi}$ is maximal and E^{prior} is invariant upon the action the entire set $\mathbb{G}^{\Omega\phi}$

$$g \circ E^{prior} = E^{prior}, \quad \forall g \in \mathbb{G} \tag{2.138}$$

For instance the prior energy E_{L_2} is invariant to the set $\mathbb{G}^{\Omega\phi} = \mathbb{G}_{const} \times SO(2)$, the set of addition of the variable ϕ with constants (eq. (2.133)) and the set of rotations in Ω (see eq. (2.128)).

Similar to the definition of A_c in eq. (2.135) we can describe the minimizers of E^{prior} as being related to each other by the elements of $\mathbb{G}^{\Omega\phi}$

$$A = \left\{\phi^\star \middle| \, \phi^\star = g \circ \phi_0^\star \quad g \in \mathbb{G}^{\Omega\phi}\right\} \tag{2.139}$$

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

$$A_\Omega\left(\phi_c^\star\right) = \left\{\phi^\star \middle| \, \phi^\star\left(\boldsymbol{x}\right) = \phi_c^\star\left(g_{\omega^\Omega} \circ \boldsymbol{x}\right), \quad g_{\omega^\Omega} \in \mathbb{G}^\Omega\right\} \tag{2.140}$$

$$A = \left\{A_\Omega\left(\phi_c^\star\right) \middle| \, \phi_c^\star = g_{\omega^\phi} \circ \phi_0^\star, \quad g_{\omega^\phi} \in \mathbb{G}^\phi\right\} \tag{2.141}$$

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 $\mathbb{G}^\Omega$

$$A_{red} = \left\{\phi_c^\star \middle| \, \phi_c^\star = g_{\omega^\phi} \circ \phi_0^\star, \quad g_{\omega^\phi} \in \mathbb{G}^\phi\right\} \tag{2.142}$$

$$\phi_d^\star\left(\boldsymbol{x}\right) \neq \phi_c^\star\left(g_{\omega^\Omega} \circ \boldsymbol{x}\right) \quad \forall g_{\omega^\Omega} \in \mathbb{G}^\Omega, \phi_{c,d}^\star \in A_{red} \tag{2.143}$$

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

Definition 12 (Level-sets). *Let $c \in \phi(\Omega)$, where $\phi(\Omega)$ is the image of the Euclidean space Ω under the map ϕ. The set S_c defined by*

$$S_c = \left\{\boldsymbol{x} \middle| \, \phi\left(\boldsymbol{x}\right) = c\right\} \tag{2.144}$$

is called a level-set of ϕ. The union S of level-sets is

$$S = \cup_{c \in \phi(\Omega)} S_c \tag{2.145}$$

and we will refer to S as the level-sets of ϕ.

By the definition of the action of g_{ω^Ω} in eq. (2.137) we see that g_{ω^Ω} is a geometrical transformation that deforms the level-sets S_c (see figure 2.3). We are free to define g_{ω^ϕ} so that it is orthogonal to g_{ω^Ω} in the sense that the level-sets S are invariant under g_{ω^ϕ}

$$S' = g_{\omega^\phi} \circ S = S \tag{2.146}$$

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

$$S = \{\boldsymbol{x} | \, \boldsymbol{\omega}_S(\boldsymbol{x}) \cdot \nabla \phi_0^\star(\boldsymbol{x}) = 0\} \tag{2.147}$$

In figure 2.3b we show an example of a level-set S which is distorted by the operator $g_\delta \in \mathbb{G}^\Omega$. The resulting level-set S' has the VVF $\boldsymbol{\omega}_S'(\boldsymbol{x}) = \boldsymbol{\omega}_S(\boldsymbol{x}) + t\boldsymbol{\delta}(\boldsymbol{x})$ as tangent vectors.

$$S' = \{\boldsymbol{x} | \, (\boldsymbol{\omega}_S(\boldsymbol{x}) + t\boldsymbol{\delta}(\boldsymbol{x})) \cdot \nabla \phi_0^\star(\boldsymbol{x}) = 0\} \tag{2.148}$$

The parameter t controls the extent to which S_c is deformed by $\boldsymbol{\delta}(\boldsymbol{x})$. However it also possible to represent S_c' with the help of a deformation of the gradient operator ∇ itself

$$S' = \{\boldsymbol{x}' | \, \boldsymbol{\omega}_S(\boldsymbol{x}') \cdot \nabla_{t\boldsymbol{\delta}} \phi_0^\star(\boldsymbol{x}') = 0\} \tag{2.149}$$

The operator $\nabla_{t\boldsymbol{\delta}}$ loosely speaking encodes a reversal of the action of g_{ω^Ω} on $\boldsymbol{x}$ so that S' can be represented with the same tangential VVF 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 ∇. A formal introduction of the push-forward is given in appendix A in the context of smooth manifolds. 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 VVF $\boldsymbol{\omega}(\boldsymbol{x})$

$$\frac{d}{dt} E^{prior}(\nabla_{t\boldsymbol{\delta}}\phi)\bigg|_{t=0} = 0 \tag{2.150}$$

Given a specific form of the operators in $\mathbb{G}^\Omega$, eq. (2.150) poses constraints on the form of the differential operators in the prior $E^{prior}(\nabla_{t\boldsymbol{\delta}}\phi)$. Eq. (2.150) 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. (2.149)).

2.5. 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. (2.136) and eq. (2.137). Following this section we will make the invariance property of E^{prior} in eq. (2.150) formally more explicit in sections 2.6 and 3. In appendix A we have laid out the theory of smooth manifolds. According to definition 32 a smooth manifold M is a structure such that each point $p \in M$ is equipped with a chart $\{U_p, \psi_p\}$, where $U_p \subset M$ is a neighborhood at the point p and $\psi_p : A \subset \mathbb{R}^n \rightarrow U_p$ is a smooth diffeomorphism that maps a subset $A \subset \mathbb{R}^n$ of the Euclidean space $\mathbb{R}^n$ to M such that $\psi_p(\mathbf{0}) = p$. A Lie group $\mathbb{G}$ is such a smooth manifold with the addition that the points $g \in \mathbb{G}$ are actually *operators* that operator on some base vector space $\mathcal{M}$

Definition 13 (Lie Group)**.** *Let $\mathbb{G}$ be an r-dimensional smooth manifold as defined in definition 32. $\mathbb{G}$ is called a Lie Group if there exists a smooth multiplication operation $'\cdot : \mathbb{G} \times \mathbb{G} \rightarrow \mathbb{G}'$ and a smooth inverse operator $i : \mathbb{G} \rightarrow \mathbb{G}$ such that the following elements exist*

- *Neutral element e: $e \cdot g = g$ for all $g \in \mathbb{G}$*
- *Inverse element: $\mathfrak{i}(g) = g^{-1} \in \mathbb{G}$ for all $g \in \mathbb{G}$*
- *Product: $p \cdot q \in \mathbb{G}$ for all $p, q \in \mathbb{G}$*

Examples of Lie groups are

- The *general linear group* $GL(m, \mathbb{R})$ of $m \times m$ invertible matrices. $GL(m, \mathbb{R})$ is a group under matrix multiplication. The multiplication AB is smooth in the entries of $A \in GL(m, \mathbb{R})$ and $B \in GL(m, \mathbb{R})$. The inversion operator $\mathfrak{i}(A) = A^{-1}$ is also smooth since by Cramer's rule the entries of A^{-1} are smooth polynomials in the entries of A of finite order and the unit element e is the unit matrix $\mathbb{1}_{m \times m}$.
- The Euclidean space $\mathbb{R}^n$ is a Lie group group with the multiplication operation $x + y$ and the inversion $\mathfrak{i}(x) = -x$ which are both smooth.

Due to the smooth manifold structure of the Lie group $\mathbb{G}$ each element $p \in \mathbb{G}$ has at least one chart $\{U_p, \psi_p\}$ assigned to it. The elements $g \in U_p$ may be expressed in local coordinates $\boldsymbol{\xi}_p^g$ (see eq. (A.4))

$$g = \left\{ \xi_p^{1,g}, \cdots, \xi_p^{n,g} \right\} \tag{2.151}$$

For instance the coordinates of the general linear group $GL(n, \mathbb{R})$ are the entries a_{ij} of the matrices $A \in GL(n, \mathbb{R})$. In the following we will drop the subscript p from the local coordinates $\boldsymbol{\xi}_p^g$ of g thereby leaving the exact chart $\{U_p, \psi_p\}$ to which g belongs implicit. Hence the local coordinates of any $g \in \mathbb{G}$ are denoted by $\boldsymbol{\xi}^g$.

The group operation $'\cdot'$ gives rise to two group homeomorphisms on $\mathbb{G}$ called the *left action* l_g and the *right action* r_g

Definition 14 (Left and Right Action)**.** *The left action l_g and the right action r_g on a Lie group $\mathbb{G}$ are defined as the mappings*

$$l_g : \mathbb{G} \to \mathbb{G} \quad l_g(h) = g \cdot h \quad g, h \in \mathbb{G} \tag{2.152}$$
$$r_g : \mathbb{G} \to \mathbb{G} \quad r_g(h) = h \cdot g \quad g, h \in \mathbb{G} \tag{2.153}$$

Since $\mathbb{G}$ is by definition a smooth manifold the following lemma holds

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

The proof is simple: since $g \cdot h$ and $\mathfrak{i}(g) = g^{-1}$ exist and are smooth for all $g, h \in \mathbb{G}$, the inverse of l_g, $l_{g^{-1}}(h) = g^{-1} \cdot h$ exists and is smooth. Hence l_g is a diffeomorphism. A similar argument holds for the right action r_g.

According to appendix A there exists a tangential bundle $T\mathbb{G}$ of vector fields $V \in T\mathbb{G}$ and associated flows $\theta^V(t, g) : \mathbb{R} \times \mathbb{G} \to \mathbb{G}$ such that the differential relation

$$\frac{d}{dt}\theta^V(t, g)\bigg|_{t=0} = V_g \in T_g\mathbb{G}, \quad V_g = v^i(\boldsymbol{\xi}^g)\partial_{\xi^{i},g} \tag{2.154}$$

holds, by the fundamental theorem for flows (theorem 4). In eq. (2.154) we used the Einstein convention for summation over the index i.

Definition 15 (Einstein Convention)**.** *If a product contains an index exactly twice then summation over said index is implied*

$$a_i b_i c_j := \sum_{i=1}^{max(i)} (a_i b_i) c_j \tag{2.155}$$

$max(i)$ is the maximum value the index i can have.

The vector field V in eq. (2.154) is a function on $\mathbb{G}$ mapping its argument $g \in \mathbb{G}$ to the image V_g. The subscript of V_g indicates that the coefficient functions v^i in eq. (2.154) depend on, and the differential operators $\partial_{\xi^{i},g}$ operate on the local coordinates of g. The set $T_g\mathbb{G}$ is the tangential space of the smooth manifold $\mathbb{G}$ at the point $g \in \mathbb{G}$ and the differential operators $\partial_{\xi^{i},g}$ form a basis of the particular tangential space $T_g\mathbb{G}$.

Let $W \in T\mathbb{G}$ be another smooth vector field on the Lie group $\mathbb{G}$. A central question is how does W transform under the flow θ^V associated with the vector field V by eq. (2.154). The Lie derivative (eq. (A.62)) $\mathcal{L}_V W$ answers this question

Definition 16 (Lie Derivative). *Let $V, W \in \mathcal{T}(M)$ be two smooth vector fields and θ^{V^Ω} be the flow of V. The Lie derivative $\mathcal{L}_V W$ of W along the flow of V is defined as the derivative*

$$(\mathcal{L}_V W)_h = \frac{d}{dt}\left((\theta^V_{-t})_\star W_{\theta^{V^\Omega}_h(t)}\right)\Big|_{t=0} \tag{2.156}$$

where $(\theta^V_{-t})_\star W$ is the push-forward (eq. (A.47)) of W along the flow θ^V.

According to the proposition 9 in appendix A the Lie derivative $\mathcal{L}_V W$ is equivalent to the commutator between V and W

$$(\mathcal{L}_V W)_h = [V_h, W_h] \tag{2.157}$$

Thus it is possible to compute $\mathcal{L}_V W$ without any reference to the flow θ^{V^Ω}.

In the case of Lie groups there exists a special subset of the tangential bundle $T\mathbb{G}$ which is invariant under the left action l_g for any $g \in \mathbb{G}$ called the *Lie Algebra*

Definition 17 (Lie Algebra). *Let $\mathbb{G}$ be a Lie group and $T\mathbb{G}$ be its tangential bundle. The Lie algebra $\mathcal{G} \subset T\mathbb{G}$ is the set of vector fields V_h which are invariant under the left action l_g*

$$(l_g)_\star V_h = V_{gh} \tag{2.158}$$

for all $g \in \mathbb{G}$

One of the most important aspects of the Lie algebra $\mathcal{G}$ is that it is *closed* under the Lie derivative $\mathcal{L}_V$ from eq. (2.156)

Proposition 3 (Left invariant commutator). *Let V and W be smooth vector fields in $\mathcal{G}$ and l_g be the left action with respect to any element $g \in \mathbb{G}$. The commutator $Z = [V, W]$ is invariant with respect to l_g*

$$(l_g)_\star [V_h, W_h] = [V_{gh}, W_{gh}] \tag{2.159}$$

It follows that $Z \in \mathcal{G}$

Eq. (2.159) tells us that if we can compute the rate of change of W along the flow of V at the location $h \in \mathbb{G}$, which is equivalent to $\mathcal{L}_V W|_h$, then we can compute it at *any* other location in $\mathbb{G}$.

Due to the left-invariance of the Lie algebra $\mathcal{G}$ and the commutator in eq. (2.159) it follows that the values $V_g \in T_g\mathbb{G}$ of any vector field $V \in \mathcal{G}$ are isomorphic to the evaluation of V at the identity $e \in \mathbb{G}$, $V_e \in T_e\mathbb{G}$ since by eq. (2.158) we have

$$V_g = (l_g)_\star V_e \tag{2.160}$$

It follows that the Lie algebra $\mathcal{G}$ can be *reduced* to the tangential space at the identity $e \in \mathbb{G}$, $T_e\mathbb{G}$.

2.5.1. 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 $\mathbb{R}^2$. The subgroup $SO(2)$ generates the unit circle S^1 by rotating the point $\boldsymbol{x}_0 \in \Omega$ by the angle $\alpha \in \mathbb{R}$

$$\boldsymbol{x}(\alpha) = g_\alpha \circ \boldsymbol{x}_0 = R_\alpha \boldsymbol{x}_0, \quad R_\alpha = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{pmatrix} \tag{2.161}$$

The subgroup $\mathbb{T}$ generates lines with orientation vector $\boldsymbol{t} \in \mathbb{R}^2$

$$g_t \circ \boldsymbol{x}_0 = \boldsymbol{x}_0 + \boldsymbol{t} \tag{2.162}$$

Thus the local coordinate vector of the group $\mathbb{T} \times SO(2)$ is the vector $\boldsymbol{\xi}_g = (t_1, t_2, \alpha)$.

The algebra $\mathcal{G} = \mathfrak{t} \times \mathfrak{so}(2)$ of $\mathbb{T} \times SO(2)$ has the basis $\{X_e^{\Omega,x}, X_e^{\Omega,y}, X_e^{\Omega,\alpha}\}$. The subset $\mathfrak{t} = \{X_e^{\Omega,x}, X_e^{\Omega,y}\}$ is the set generators of infinitesimal translations

$$X_e^{\Omega,x} = \partial_x, \quad X_e^{\Omega,y} = \partial_y \tag{2.163}$$

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

$$X_e^{\Omega,\alpha} = -y\partial_x + x\partial_y \tag{2.164}$$

From eq. (2.164) we can see that $X_e^{\Omega,\alpha}$ does not commute with $\mathfrak{t}$ and the commutators for the basis $\left\{X_e^{\Omega,x}, X_e^{\Omega,y}, X_e^{\Omega,\alpha}\right\}$ are easily computed

$$\left[X_e^{\Omega,\alpha}, X_e^{\Omega,x}\right] = -X_e^{\Omega,y} \quad \left[X_e^{\Omega,\alpha}, X_e^{\Omega,y}\right] = X_e^{\Omega,x} \quad \left[X_e^{\Omega,x}, X_e^{\Omega,y}\right] = 0 \tag{2.165}$$

In eq. (2.165) we see that the translation algebra $\mathfrak{t}$ is closed under commutation with the basis of $\mathfrak{so}(2)$. It is an example of what is called the *ideal* of the Lie algebra $\mathcal{G} = \mathfrak{t} \times \mathfrak{so}(2)$

Definition 18 (Ideal of a Lie Algebra). *A q-dimensional sub-algebra $\mathcal{X}$ of a Lie algebra $\mathcal{G}$ is called an ideal if it is closed under the commutator* $[\cdot, \cdot]$

$$[V, X] \in \mathcal{X}, \quad \forall V \in \mathcal{G}, X \in \mathcal{X} \subset \mathcal{G} \tag{2.166}$$

Let X^i be a basis element of $\mathcal{X}$ and ∂_{ξ^j} be a basis element of $\mathcal{G}$ (eq. (2.154)). Then the commutator $\left[\partial_{\xi^j}, X^i\right]$ can be expressed in the basis of $\mathcal{X}$

$$\left[\partial_{\xi^j_g}, X^i_g\right] = \sum_{l=1}^{q} C^l_{j,i} X^l_g, \quad \forall g \in \mathbb{G} \tag{2.167}$$

where $C^l_{j,i}$ are the structure constants from definition 46.

The meaning of the first two commutators in eq. (2.165) is that the gradient operator ∇ is rotated by 90° counter clockwise under the action of $X_e^{\Omega,\alpha}$

$$\frac{d}{d\alpha} \nabla_{\boldsymbol{x}(\alpha)} \bigg|_{\alpha=0} = \left[X_e^{\Omega,\alpha}, \nabla\right] = \mathbf{M}_\alpha \cdot \nabla, \quad \mathbf{M}_\alpha = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{2.168}$$

The structure constant matrix $\mathbf{M}_\alpha$ is one of the Pauli matrices [54]. 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 [54, 79, 36].

2.6. Noether's First Theorem

In her original paper [73, 74] Emmy Noether handles the question: Given a model of a physical system, encoded in an action

$$E = \int_\Omega \left(\mathcal{E}\left(\boldsymbol{x}, \{\phi_\rho\}, \{\nabla_K \phi_\rho\}\right)\right) d^n x \tag{2.169}$$

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

$$E' = g_\gamma \circ E = E \quad \forall g_\gamma \in \mathbb{G} \tag{2.170}$$

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 2.5. 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 $\alpha_1, \dots, \alpha_r$ 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

$$\phi'_\rho\left(\boldsymbol{x}\right) = \phi_\rho\left(\boldsymbol{x}\right) + \sum_{m=1}^{r} \alpha_m \omega_m^{\phi_\rho}(\boldsymbol{x}) \quad \boldsymbol{x}' = \boldsymbol{x} + \sum_{m=1}^{r} \alpha_m \boldsymbol{\omega}_m^{\Omega}(\boldsymbol{x}) \tag{2.171}$$

She proved that if the action E is invariant under g_γ in eq. (2.170), then there exists r vectors $\boldsymbol{W}_m$ such the integral relationship

$$E - E' = \int_\Omega \sum_{m=1}^{r} \alpha_m \left[\sum_\rho \bar{\omega}_m^{\phi_\rho}\left[\mathcal{E}\right]_\rho + \mathrm{div}\left(\boldsymbol{W}_m\right)\right] = 0 \tag{2.172}$$

$$\bar{\omega}_m^{\phi_\rho} = \left(\omega_m^{\phi_\rho} - \omega_m^{\mu\Omega}\partial_\mu\phi_\rho\right), \quad \left[\mathcal{E}\right]_\rho = \frac{\delta\mathcal{E}}{\delta\phi_\rho} - \frac{d}{dx^\mu}\left(\frac{\delta\mathcal{E}}{\delta\phi_{\rho,\mu}}\right) \tag{2.173}$$

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

$$\int f \cdot \partial_\mu g d^n x = \int \partial_\mu\left(f \cdot g\right) d^n x - \int \partial_\mu f \cdot g d^n x \tag{2.174}$$

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

$$\sum_\rho \bar{\omega}_m^{\phi_\rho}\left[\mathcal{E}\right]_\rho + \mathrm{div}\left(\boldsymbol{W}_m\right) = 0 \quad m = 1, \dots, r \tag{2.175}$$

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 $\rho \leq r$ the system of equations in eq. (2.175) is overdetermined, eq. (2.172) can only hold if *all* the divergences and *all* the

Euler-Lagrange equations vanish

$$[\mathcal{E}]_\rho\left(\phi_1^\star,\ldots,\phi_\rho^\star\right)=0,\quad \operatorname{div}\left(\boldsymbol{W}_m\right)\left(\phi_1^\star,\ldots,\phi_\rho^\star\right)=0,\quad \rho\leq r \tag{2.176}$$

Eq. (2.176) 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 $\mathbb{R}^3$

$$V\left(\mathbf{r}_e(t)\right)=-\frac{m_e\cdot m_s}{r}\quad r=\|\mathbf{r_e}-\mathbf{r_s}\| \tag{2.177}$$

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

$$L\left(\mathbf{r}_e(t)\right)=\frac{1}{2}m_e\dot{r}_e^2+\frac{1}{2}m_e\dot{r}_s^2-V\left(\mathbf{r}_e(t)\right) \tag{2.178}$$

The Euler-Lagrange equations are easily computed

$$\ddot{r}-\frac{m_s+m_e}{r^2}=0 \tag{2.179}$$

The parameter t is the time parameter of the two body system and eq. (2.179) describes the motion of the total mass around the center of mass of the two body system. The Kepler Lagrangian in eq. (2.178) 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 $SO(3)\times\mathbb{R}$)

$$t'=t+\delta t \tag{2.180}$$

$$\mathbf{r}'=\mathbf{r}+\partial_{\theta_i}\mathbf{r}'\delta\theta_i\quad i=x,y\,\text{or}\,z \tag{2.181}$$

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

$$W_t = \mathcal{H} = \frac{1}{2} m_e \dot{r}^2 + V\left(\mathbf{r}_e\left(t\right)\right) \quad \textit{time shift} \tag{2.182}$$

$$W_x = z\dot{y} - y\dot{z} \quad \textit{Rotation around x-axis} \tag{2.183}$$

$$W_y = z\dot{x} - x\dot{z} \quad \textit{Rotation around y-axis} \tag{2.184}$$

$$W_z = x\dot{y} - y\dot{x} \quad \textit{Rotation around z-axis} \tag{2.185}$$

The conserved quantity $\mathcal{H}$ in eq. (2.182) 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. (2.183) to eq. (2.185)) 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. (2.179) 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 $SO\left(3\right)$ comes the fact that $S_{\boldsymbol{W}}$ is actually a flat Euclidean plane embedded in a 3-dimensional Euclidean space.

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 [79] directly implement electrodynamical 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 [61, 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

$$\widetilde{E}_C(\kappa_1, \ldots, \kappa_r) = E_C(\phi_i(\kappa_1, \ldots, \kappa_r), \nabla\phi_i(\kappa_1, \ldots, \kappa_r)) \tag{2.186}$$

The mathematical expression of the re-parameterized energy $\widetilde{E}_C(\kappa_1, \dots, \kappa_r)$ 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.

2.7. Total Variation

In this section we will introduce a widely used method for anisotropic regularization of the GRF ϕ called Total Variation (TV) [93, 18, 109, 29, 13]. In the context of shock-filtering [93, 75, 92] it was shown that the functional

$$E_{L_1}(\phi) = \int |\nabla\phi| dx \tag{2.187}$$

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, [93] chose the approximative prior

$$E_{L_1 approx}(\phi) = \int \sqrt{|\nabla\phi|^2 + \epsilon} dx \tag{2.188}$$

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 [13, 96]. To do this we need to explore the function-space the minimizers of eq. (2.187) might belong to. Smooth functions ϕ_{smooth} are functions for which $\nabla\phi$ exists everywhere, thus they may be minimizers of eq. (2.187). 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. (2.187) is possible if one assumes $\nabla\phi$ to be a distribution, more precisely a Radon measure [13, 96] 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

$$\int_\Omega \phi \cdot \mathrm{Div}\boldsymbol{p} d^2x = -\int \boldsymbol{p}^T d\boldsymbol{\mu} < \infty \tag{2.189}$$

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

$$BV = \{\phi | \int_\Omega \phi \cdot \mathrm{Div}\boldsymbol{p} d^2x < \infty, \|\boldsymbol{p}\| \leq 1\Big\} \tag{2.190}$$

Now it is possible to define a norm on BV. By virtue of the Hölder inequality [39] there exists a scalar C for which we can determine the upper bound of eq. (2.189)

$$\int_{\Omega} \phi \cdot \mathrm{Div}\boldsymbol{p} dx \leq C\|\phi\|_{\infty} \tag{2.191}$$

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

$$TV\left(\phi\right) = \sup\left\{\int_{\Omega} \phi \cdot \mathrm{Div}\boldsymbol{p} d^2x \,|\quad \|\boldsymbol{p}\| \leq 1\right\} \tag{2.192}$$

As was discussed in [13] 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 $TV\left(\phi\right)$ can be written in the form

$$TV\left(\phi\right) = \sum_k \|\nabla\phi_{\Omega_k}\|_{L_1} + \sum_{k<m} L_{lm} \tag{2.193}$$

$$L_{lm} = \int_{\Omega_l \cap \Omega_m} |\phi_l - \phi_m|\, dl_{lm} \tag{2.194}$$

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. (2.193) penalizes the smooth parts of ϕ (the gradients $(\nabla\phi)_{\Omega_k}$). Similar to eq. (2.134) $\|\nabla\phi_{\Omega_k}\|_{L_1}$ is invariant to shifts of ϕ_{Ω_k} by constants c_k'

$$\|\nabla\left(\phi_{\Omega_k} + c_k'\right)\|_{L_1} = \|\nabla\phi_{\Omega_k}\|_{L_1} \tag{2.195}$$

Thus due to eq. (2.195) we can view the smooth functions ϕ_{Ω_k} as being centered around constants c_k

$$\phi_{\Omega_k}\left(\boldsymbol{x}\right) = c_k + \widetilde{\phi}_{\Omega_k}\left(\boldsymbol{x}\right) \tag{2.196}$$

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

$$E\left(\phi, \nabla\phi\right) = \sum_k \int_{\Omega_k} (\phi_d - \phi_{\Omega_k})^2 dx + \lambda TV\left(\phi\right) \tag{2.197}$$

then in [67, 107] it was shown that the c_k can be computed to be the *mean* of the data ϕ_d within the area Ω_k, $c_k = \int_{\Omega_k} \phi_d d^2x$ given that the deviations $\widetilde{\phi}_{\Omega_k}$ are penalized by the first term in eq. (2.193).

The second term in eq. (2.193) 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. (2.197) then we obtain

$$\widetilde{E}\left(\phi, \nabla\phi\right) = \sum_k \int_{\Omega_k} (\phi_d - c_k)^2 dx + \lambda TV\left(\phi\right) \tag{2.198}$$

which is of course only an approximation to eq. (2.197). The data term in eq. (2.198) is a measure for the variance of ϕ_d in Ω_k. The two terms in eq. (2.193) together with the data term $\sum_k \int_{\Omega_k} (\phi_d - c_k)^2 dx$ in eq. (2.198) 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. (2.197) and (2.198) marks the trade-off between the TV term and the data term in eqs. (2.197) and (2.198) and thus it determines the size of the partitions Ω_k.

2.7.1. The Mean Curvature of Total Variation

In eq. (2.193) 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. (2.193), 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. (2.194) can be rewritten essentially as a measure for the length of the level-set S_{lm} interfacing Ω_l and Ω_m

$$L_{lm} = |\phi_k - \phi_m| \, \|S\|_{lm}\,, \quad \|S\|_{lm} = \int_0^T \left\| \frac{d}{dt}\left(\boldsymbol{x}(t)\right) \right\| dt \tag{2.199}$$

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

$$\boldsymbol{x}(t) = g_t^{V^\Omega} \circ \boldsymbol{x}_0, \quad \boldsymbol{x}_0 \in \Omega_l \cap \Omega_m \tag{2.200}$$

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

$$\|S\|_{lm} = s\left(T\right), \quad s\left(t\right) = \int_0^t \left\|\boldsymbol{v}\left(\boldsymbol{x}(t')\right)\right\| dt' \tag{2.201}$$

The function $s(t)$ in eq. (2.201) is called the *arc length* of the curve $\boldsymbol{x}(t)$. By virtue of the definition of the arc length $s(t)$ in eq. (2.201) we can express derivatives with respect to s by

$$\frac{d}{ds} = \frac{1}{\|\boldsymbol{v}\|}\frac{d}{dt} \tag{2.202}$$

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. (2.202)

$$\kappa(\boldsymbol{x}_0) = \left\| \frac{d^2}{ds^2}\boldsymbol{x}(s) \right\|_{s=0} \tag{2.203}$$

$$= \frac{1}{\|\boldsymbol{v}\|^3}\left(v_x \cdot \frac{dv_y}{dt} - v_y \cdot \frac{dv_x}{dt} \right)\Bigg|_{t=0} \tag{2.204}$$

In [13] it is shown that the expression for the curvature $\kappa(\boldsymbol{x}_0)$ at the point $\boldsymbol{x}_0$ in eq. (2.204) is equivalent to the *mean curvature* [28, 17, 51]

$$\kappa = \mathrm{Div}\left(\frac{\nabla\,\phi}{|\nabla\phi|} \right) \tag{2.205}$$

which is the functional derivative of the TV norm in eq. (2.187) with respect to ϕ

$$\kappa = -\partial TV(\phi) \tag{2.206}$$

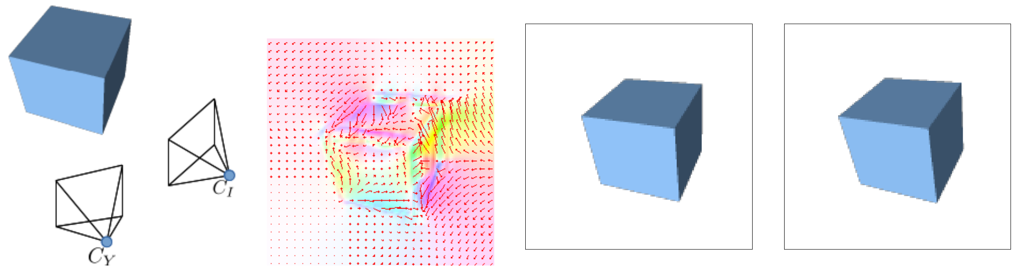
For a thorough derivation of the mean curvature κ in terms of weak derivatives in BV spaces see [96]. The Euler-Lagrange equations of any energy function $E = \int \mathcal{E} d^2x$ including the TV functional in eq. (2.187) as a prior

$$[\mathcal{E}](\boldsymbol{x}) = \frac{\delta\mathcal{E}}{\delta\phi}(\boldsymbol{x}) + \lambda\kappa(\boldsymbol{x}) = 0 \tag{2.207}$$

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 $SE(2)$, the group of rotations and translations, TV is also an invariant of that group.

2.8. 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



(a) **(b)** $\boldsymbol{d}$ **(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 $\boldsymbol{d}$. The vectors in figure 2.4b indicate which pixels $\boldsymbol{x}'$ in I and $\boldsymbol{x}$ in Y are mapped to each other.

mapped to the data Y by the model C, see eq. (2.11). A prime example of an inverse problem in computer vision is optical flow [48, 58, 101, 10, 14, 53, 109]. 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 [42, 41] 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 [58, 81, 83, 103, 99, 98, 52, 95]. 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 [113].

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

$$\boldsymbol{x}' = \boldsymbol{x} + \boldsymbol{d}\left(\boldsymbol{x}\right) \tag{2.208}$$

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 [48, 58, 105, 20] for the estimation of the optical flow $\boldsymbol{d}$ is to model $\boldsymbol{d}$ as a GRF with a given data term $E^{data}_{Y,I}\left(\boldsymbol{d}\right)$. Without further information of the mapping between Y and I from another source (e.g. sparse feature mapping with SIFT features [57]), the data term $E^{data}_{Y,I}(\boldsymbol{d})$ 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 $I_{\boldsymbol{d}}(\boldsymbol{x}) = I(\boldsymbol{x} + \boldsymbol{d}(\boldsymbol{x}))$ [105, 20]

$$E^{data}_{Y,I}(\boldsymbol{d}) = F(Y, I_{\boldsymbol{d}}), \quad I_{\boldsymbol{d}}\left(\boldsymbol{x}\right) = I\left(\boldsymbol{x} + \boldsymbol{d}(\boldsymbol{x})\right) \tag{2.209}$$

In general the mapping of $Y(\boldsymbol{x})$ and $I_{\boldsymbol{d}}(\boldsymbol{x})$ via $\boldsymbol{d}(\boldsymbol{x})$ is ill-determined: According to [53] the regions $\mathcal{A} \subset \Omega$ of an image ϕ can be given an *intrinsic dimension* [9, 116] 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 $E^{data}_{Y,I}(\boldsymbol{d})$ 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 $E^{data}_{Y,I}(\boldsymbol{d})$. 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 $E^{data}_{Y,I}(\boldsymbol{d})$ 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 $\boldsymbol{d}(\boldsymbol{x})$ is well defined over all Ω. The total energy

$$E_{Y,I}\left(\boldsymbol{d}\right) = E^{data}_{Y,I}\left(\boldsymbol{d}\right) + E^{prior}\left(\nabla \boldsymbol{d}\right) \tag{2.210}$$

is then a trade-off between the similarity of Y and $I_{\boldsymbol{d}}$ and the smoothness of the optical flow $\boldsymbol{d}(\boldsymbol{x})$.

2.8.1. Uni-Modal Optical Flow

Among the earliest methods for optical flow estimation are the methods described in the seminal papers of Horn and Schunck [48] and Lukas and Kanade [58]. In [48] the following model for computing the flow between two frames of a video was proposed

$$E_{Y,I}(\boldsymbol{d}) = E_{Y,I}^{data}(\boldsymbol{d}) + E^{prior}(\boldsymbol{d}) \tag{2.211}$$

$$E_{Y,I}^{data}(\boldsymbol{d}) = \int_{\Omega} \Big(Y(\boldsymbol{x}) - I_{\boldsymbol{d}}(\boldsymbol{x})\Big)^2 dx, \quad I_{\boldsymbol{d}}(\boldsymbol{x}) = I(\boldsymbol{x} + \boldsymbol{d}(\boldsymbol{x})) \tag{2.212}$$

$$E^{prior}(\boldsymbol{d}) = \lambda \int_{\Omega} \sum_i \|\nabla d_i\|^2 \, dx \tag{2.213}$$

In eq. (2.212) the frame I is warped back to the frame Y by the field $\boldsymbol{d}(\boldsymbol{x})$. The prior energy $E^{prior}(\boldsymbol{d})$ in eq. (2.213) imposes an isotropic smoothness constraint on the flow field $\boldsymbol{d}$. As we discussed in section 2.4.1 the main limitation of the L_2 prior in eq. (2.211) is that it does not preserve edges in the flow field $\boldsymbol{d}(\boldsymbol{x})$. To overcome this limitation [76] and [115] used the TV prior in eq. (2.187) as a smoothing term for each of the components of $\boldsymbol{d}$

$$E_{Y,I}(\boldsymbol{d}) = E_{Y,I}^{data}(\boldsymbol{d}) + \lambda \int_{\Omega} \sum_i \|\nabla d_i\| \, dx \tag{2.214}$$

According to section 2.7 the level-sets of each component $d_i(\boldsymbol{x})$ are smoothed while the discontinuities are preserved.

2.8.2. Multi-Modal Optical Flow

The next issue with the model in eq. (2.211) is that the likelihood $E_{Y,I}^{data}(\boldsymbol{d})$ in eq. (2.212) 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*

$$Y \overset{d}{\approx} I \tag{2.215}$$

Thus it is possible to find an optical flow field $\boldsymbol{d}^\star$ such that for each pixel $\boldsymbol{x} \in \Omega$ the warped image $I_{\boldsymbol{d}}$ approximates the image Y, $Y(\boldsymbol{x}) \approx I_{\boldsymbol{d}}(\boldsymbol{x})$. However there exists multi-modal setups where the cameras are *not* sensitive to the same spectra. In section 4.7.5 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 $I_{vsc} \overset{d}{\approx} y_{tc}$.

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

$$\mathbb{E}(X) = \int X \cdot p(X) \tag{2.216}$$

$$\mathbb{V}ar(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) \tag{2.217}$$

The three similarity measures all avoid the brightness constancy assumption implied by the data term in eq. (2.212) 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) [59, 87, 105, 47] is a popular similarity measure used mainly in medical imaging where images from different modalities including Magnetic Resonance Imaging (MRI) [27], Computed Tomography (CT) [49] and Positron Emission Tomography (PET) [106] 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

$$MI(Y, I) = \int p\left(\hat{Y}, \hat{I}\right) \ln \frac{p\left(\hat{Y}, \hat{I}\right)}{p\left(\hat{Y}\right) \cdot p\left(\hat{I}\right)} d\hat{Y} d\hat{I} \tag{2.218}$$

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. (2.218) 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 $I_{\boldsymbol{d}}$

$$E^{data}_{Y,I}(\boldsymbol{d}) = -MI(Y, I_{\boldsymbol{d}}) \tag{2.219}$$

However, as [85] 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, [85] 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 $I(\boldsymbol{x})$ and $Y(\boldsymbol{x})$ 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

$$p\left(\hat{Y}\middle|\, \hat{I}=I\right) = \frac{p\left(\hat{Y}, \hat{I}=I\right)}{p\left(\hat{I}=I\right)} \tag{2.220}$$

encodes the spatial functional relationship between Y and I. They introduced the Correlation Ratio (CR) [86, 87, 105, 20]

$$\eta_{CR}\left(I|Y\right) = \frac{\mathbb{V}ar\left(\phi^{\star}\left(Y\right)\right)}{\mathbb{V}ar\left(I\right)}, \quad E^{data}_{Y,I}\left(\boldsymbol{d}\right) = -\eta_{CR}\left(I_{\boldsymbol{d}}|Y\right) \tag{2.221}$$

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

$$\phi^{\star}\left(Y\right) = \mathbb{E}\left(\hat{I}\middle|\, \hat{Y}=Y\right) = \int I p\left(I|\, Y\right) dI \tag{2.222}$$

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. (2.221) 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 continuous, is measurable in the L_2-sense. Thus CR is a much stronger constraint then MI and has fewer, but more meaningful minima [85].

Cross Correlation

Cross Correlation [31, 70, 15, 105] is the strongest constrained similarity measure. It is basically an additional constraint to CR, namely that the functional

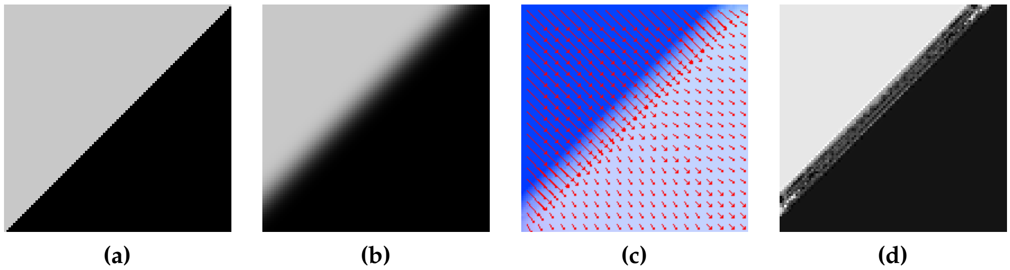


Figure 2.5.: Figure 2.5a shows a synthetic high resolution image I^{syn}. In figure 2.5b we show a low resolution image Y^{syn}. Y^{syn} is computed by down sampling I^{syn} by a factor $\sigma^{sc} = 5$, creating y^{syn} followed by cubic interpolation and translated by 10 pixels relative to I^{syn}. An optical flow model which incorporates knowledge of the scale difference between Y^{syn} and I^{syn} should produce a flow $\widetilde{\boldsymbol{d}}$, such that the warped image $I^{syn}_{\widetilde{\boldsymbol{d}}}$ matches the image Y^{syn} *up to scale* σ^{sc}, thereby preserving the features of I^{syn}. Figure 2.5c shows the flow $\boldsymbol{d}$ computed with the model in eq. (2.211). Since the model in eq. (2.211) does not incorporate knowledge of the scale difference σ^{sc}, the features of the warped image $I^{syn}_{\boldsymbol{d}}$ (figure 2.5d) are heavily distorted

relationship in eq. 2.221 must be linear. Then η reduces to

$$\eta_{CC}\left(I|Y\right) = \frac{Cov\left(Y, I\right)}{\mathbb{V}ar\left(I\right)\cdot \mathbb{V}ar\left(Y\right)} \qquad Y = f \cdot I + \beta \tag{2.223}$$

As we will see in section 4.4 a measure similar to eq. (2.223) 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. (2.219) [59] 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

$$i(\boldsymbol{x}) = (G \star I)(\boldsymbol{x}), \quad \boldsymbol{x} \in \Omega \tag{2.224}$$

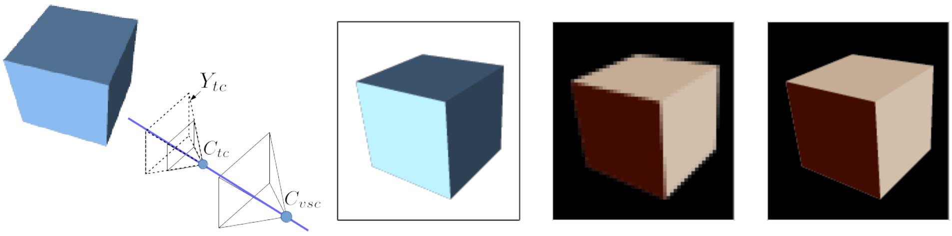
and evaluate the similarity measures on the image pair y and i. The negative impact is that we could only estimate an optical flow $\boldsymbol{d}$ with the same low resolution as the image y. Conversely we could up-sample the image y with some interpolation scheme to produce a high resolution image 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 [48]

$$E(\boldsymbol{d}^{syn}) = \frac{1}{2}\int_{\Omega}\Big(Y^{syn}(\boldsymbol{x}) - I^{syn}_{\boldsymbol{d}^{syn}}(\boldsymbol{x})\Big)^2 dx + \frac{\lambda}{2}\sum_i \int_{\Omega} \|\nabla d_i^{syn}(\boldsymbol{x})\|^2 \, dx \tag{2.225}$$

(see eq. (2.211)) to compute the optical flow $\boldsymbol{d}^{syn}$ mapping I^{syn} to Y^{syn} (see figure 2.5c). Figure 2.5d shows the image $I^{syn}_{\boldsymbol{d}^{syn}}(\boldsymbol{x}) = I^{syn}\big(\boldsymbol{x} + \boldsymbol{d}^{syn}(\boldsymbol{x})\big)$. 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. (2.225) 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}.

2.9. Image Fusion

In this section we will introduce the image fusion method of Hardie et. al. [44]. 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 [16] where the TC and the VSC are built in a co-aligned fashion within the body of a satellite. In section 4.3 the goal is to extend this method to the case where 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}.



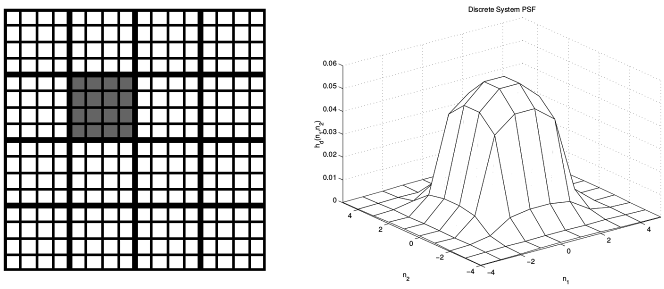
(a) **(b)** I_{vsc} **(c)** y_{tc} **(d)** Y_{tc}

Figure 2.6.: Figure 2.6a shows a schematic setup of the camera configuration considered by Hardie et. al. [44]. 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 [44] takes the data y_{tc} and I_{vsc} to produce a higher resolution thermographic image Y_{tc}, shown in figure 2.6d

The method of [44] 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 [38]. The model, obtained by Hardie et al. [43], 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 $W_{\sigma^{sc}}$

$$y_{tc} = \langle Y_{tc} \rangle_{\sigma^{sc}} + n \quad n \sim \mathcal{N}(0|C_n), \quad \langle Y_{tc} \rangle_{\sigma^{sc}} = W_{\sigma^{sc}} Y_{tc} \tag{2.226}$$

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. (2.226) since the high spacial frequency components of Y_{tc} are filtered out by $W_{\sigma^{sc}}$. In [44] 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



(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. [43]. It shows that each pixel in the TC image has a non uniform response over its surface to incoming photons.

avoiding limitations such as the noise model of y_{tc}.

The key ingredient in the model of [44] 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 [94]

$$p(Y_{tc}|I_{vsc}) = \frac{p(Y_{tc}, I_{vsc})}{p(I_{vsc})} \sim \mathcal{N}\left(\mu_{Y_{tc}|I_{vsc}} | C_{Y_{tc}|I_{vsc}}\right) \tag{2.227}$$

$$C_{Y_{tc}|I_{vsc}} = C_{Y_{tc},Y_{tc}} - C^2_{Y_{tc},I_{vsc}} \cdot C^{-1}_{I_{vsc},I_{vsc}} \tag{2.228}$$

$$\mu_{Y_{tc}|I_{vsc}}(\mathbf{x}) = \mu_{Y_{tc}} + C_{Y_{tc},I_{vsc}} \cdot C^{-1}_{I_{vsc},I_{vsc}} \left(I_{vsc}(\boldsymbol{x}) - \mu_{I_{vsc}}\right) \tag{2.229}$$

where the variances and means are computed globally

$$C_{u,v} = \int_\Omega (u(\boldsymbol{x}) - \mu_u) \cdot (v(\boldsymbol{x}) - \mu_v) d^2x \tag{2.230}$$
$$\mu_u = \int_\Omega u(\boldsymbol{x})\, d^2x$$

We see that the mean of Y_{tc} conditioned on I_{vsc}, $\mu_{Y_{tc}|I_{vsc}}$ (eq. (2.229)) 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. (2.227) with the Gaussian likelihood in eq. (2.226) to the posterior

$$p(Y_{tc}|y_{tc}, I_{vsc}) \sim p(y_{tc}|Y_{tc}) \cdot p(Y_{tc}|I_{vsc}) = \exp(-E_{y_{tc},I_{vsc}}(Y_{tc})) \tag{2.231}$$

with the associated energy

$$
\begin{aligned}
E_{y_{tc},I_{vsc}}\left(Y_{tc}\right) =& \frac{1}{2}\int_{\Omega}\Big(y_{tc}\left(\boldsymbol{x}\right) - W_{\sigma^{sc}}Y_{tc}\left(\boldsymbol{x}\right)\Big)^{2}\cdot C_{n}^{-1}d^{2}x \\
&+\frac{1}{2}\int_{\Omega}\Big(Y_{tc}\left(\boldsymbol{x}\right) - \mu_{Y_{tc}|I_{vsc}}\left(\boldsymbol{x}\right)\Big)^{2}\cdot C_{Y_{tc}|I_{vsc}}^{-1}d^{2}x
\end{aligned}
\tag{2.232}
$$

The minimization of eq. (2.232) and thus maximization of (2.231) with respect to Y_{tc} gives the analytical solution [44] in lexicographic reordering (appendix C)

$$
\begin{aligned}
&Y_{tc}^{\star} = \mu_{Y_{tc}|I_{vsc}} + C_{Y_{tc}|I_{vsc}}\cdot W_{\sigma^{sc}}^{T}\mathbf{H}^{-1}\cdot\Big(y_{tc} - \Big\langle\mu_{Y_{tc}|I_{vsc}}\Big\rangle_{\sigma^{sc}}\Big) && (2.233)\\
&\Big\langle\mu_{Y_{tc}|I_{vsc}}\Big\rangle_{\sigma^{sc}} = W_{\sigma^{sc}}\mu_{Y_{tc}|I_{vsc}} \\
&\mathbf{H} = \Big(W_{\sigma^{sc}}\cdot C_{Y_{tc}|I_{vsc}}\cdot W_{\sigma^{sc}}^{T} + C_{n}\Big)
\end{aligned}
$$

Eq. (2.233) 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 [117] a computationally tractable approximation was introduced

$$
\hat{Y}_{tc} = \mu_{Y_{tc}|I_{vsc}} + C_{\langle Y_{tc}\rangle_{\sigma^{sc}}|\langle I_{vsc}\rangle_{\sigma^{sc}}}\cdot\widetilde{\mathbf{H}}^{-1}\Big(y_{tc} - \Big\langle\mu_{Y_{tc}|I_{vsc}}\Big\rangle_{\sigma^{sc}}\Big) \tag{2.234}
$$

$$
\langle I_{vsc}\rangle_{\sigma^{sc}} = W_{\sigma^{sc}}I_{vsc},\quad \langle Y_{tc}\rangle_{\sigma^{sc}} = W_{\sigma^{sc}}Y_{tc}\approx y_{tc} \tag{2.235}
$$

$$
\widetilde{\mathbf{H}} = \Big(C_{\langle Y_{tc}\rangle_{\sigma^{sc}}|\langle I_{vsc}\rangle_{\sigma^{sc}}} + C_{n}\Big) \tag{2.236}
$$

$$
\mu_{\langle Y_{tc}\rangle_{\sigma^{sc}}|I_{vsc}}\left(\mathbf{x}\right) = \mu_{\langle Y_{tc}\rangle_{\sigma^{sc}}} + C_{\langle Y_{tc}\rangle_{\sigma^{sc}},I_{vsc}}\cdot C_{I_{vsc},I_{vsc}}^{-1}\left(I_{vsc}\left(\boldsymbol{x}\right) - \mu_{I_{vsc}}\right) \tag{2.237}
$$

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. (2.234) 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 4.3. 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 4.3.

3. Noether's First Theorem: A Modern Version

Noether's Theorem in section 2.6 introduces the concept of invariance of an energy functional $E(\phi, \nabla\phi)$ with respect to the action of an n-dimensional Lie group $\mathbb{G}^{\Omega\phi}$ on the functions $\phi(\boldsymbol{x})$ and the Euclidean space Ω (see eqs. 2.136 and 2.137). The main conclusion of the theorem is that if $E(\phi, \nabla\phi)$ is invariant under the action of $\mathbb{G}^{\Omega\phi}$ then there exists n vector valued functions $\boldsymbol{W}_m(\phi, \boldsymbol{x})$, $1 \leq m \leq n$ which are divergence free as soon as $E(\phi, \nabla\phi)$ is at its minimum.

However in its current form Noether's theorem has two flaws. First the energy functional $E(\phi, \nabla\phi)$ is assumed to be differentiable in the function ϕ. Second it does not explain under what conditions a minimizer $\phi^\star$ for $E(\phi, \nabla\phi)$ exists and thus $\mathrm{div}\boldsymbol{W}_m(\phi^\star, \boldsymbol{x}) = 0$ for $1 \leq m \leq n$. Within the context of convex analysis in section 2.2 Fenchel's duality theorem was introduced. An important result of this theorem is that the energy functional $E(\phi, \boldsymbol{A}\phi)$ is at its minimum if the Kuhn-Tucker conditions in eq. (2.92) are fulfilled. The Kuhn-Tucker conditions also handle the case when $E(\phi, \boldsymbol{A}\phi)$ is non-smooth in ϕ.

Hence the goal in this section is an attempt to reconcile Noether's theorem with the theory of convex analysis, especially the Kuhn-Tucker conditions in eq. (2.92). Our approach is facilitated by considering the Lie group $\mathbb{G}^{\Omega\phi}$ and convex energy functionals E which are functions on $\mathbb{G}^{\Omega\phi}$

$$E\left(\phi_g, \boldsymbol{X}_g^\Omega \phi_g\right) = E^{data}(\phi) + E^{prior}(\boldsymbol{X}_g^\Omega \phi), \quad \phi_g = g \circ \phi, \quad \phi \in \Phi(\Omega) \tag{3.1}$$

$$E^{data}(\phi_g) = \int_\Omega \mathcal{E}^{data}\left(\phi_g\right) d^2x, \quad E^{prior}(\boldsymbol{X}_g^\Omega \phi_g) = \int_\Omega \mathcal{E}^{prior}\left(\boldsymbol{X}_g^\Omega \phi_g\right) d^2x \tag{3.2}$$

The operators $\boldsymbol{X}^\Omega = \left\{X^{\Omega,1}, X^{\Omega,2}\right\}$ are the basis of a 2-dimensional sub-algebra $\mathcal{X}^\Omega \subset \mathcal{G}^{\Omega\phi}$, where $\mathcal{G}^{\Omega\phi}$ is the algebra of $\mathbb{G}^{\Omega\phi}$. We consider $\mathcal{X}^\Omega$ to be the ideal of $\mathcal{G}^{\Omega\phi}$ (see definition 18). Furthermore $\Phi(\Omega)$ is a Hilbert space which depends on the prior E^{prior}. For instance if E^{prior} is the total variation prior in section 2.7 then $\Phi(\Omega)$ is the set of functions of bounded variation BV (eq. (2.190)).

From the Kuhn-Tucker conditions in lemma 6 it follows that there exists a minimizer $\phi_g^\star \in \Phi(\Omega)$ such that

$$-\mathrm{Div}(\boldsymbol{p}_g^{\star,pr}) = \boldsymbol{X}_g^{\dagger,\Omega}\boldsymbol{p}_g^{\star,pr} \in \partial E^{data}(\phi_g^\star), \quad \boldsymbol{p}_g^{\star,pr} \in \partial E^{prior}(\boldsymbol{X}_g^\Omega \phi_g^\star) \subset \Phi^\star(\Omega) \tag{3.3}$$

The main result in this section is that if E is invariant under the Lie group $\mathbb{G}^{\Omega\phi}$ then the Kuhn-Tucker conditions in eq. (3.3) are satisfied if the algebra $\mathcal{X}^\Omega$ is commutative.

3.1. The action of $\mathbb{G}^{\Omega\phi}$ on Functionals

The action of $\mathbb{G}^{\Omega\phi}$ on $\Phi(\Omega)$

We will assume $\mathbb{G}^{\Omega\phi}$ to be a n-dimensional Lie group such that its elements $g \in \mathbb{G}^{\Omega\phi}$ act on both the function space $\Phi(\Omega)$ *as well as on the Euclidean frame* Ω

$$\phi_g(\boldsymbol{x}) := g \circ (\phi(\boldsymbol{x})) = \phi_{g^\phi}(\boldsymbol{x}_{g^\Omega}), \quad \phi \in \Phi(\Omega) \tag{3.4}$$

$$g^\phi := g|_{\Phi(\Omega)}, \quad \phi_{g^\phi}(\boldsymbol{x}) := \left(g^\phi \circ \phi\right)(\boldsymbol{x}), \quad g^\phi \circ \boldsymbol{x} = \boldsymbol{x}, \quad \forall \boldsymbol{x} \in \Omega \tag{3.5}$$

$$g^\Omega := g|_\Omega, \quad \boldsymbol{x}_{g^\Omega} := g^\Omega \circ \boldsymbol{x}, \quad \boldsymbol{x} \in \Omega \tag{3.6}$$

The operator g^ϕ in eq. (3.5) is the restriction of g on to the function values $\phi(\boldsymbol{x})$. It only transforms the function values at each location $\boldsymbol{x}$ *and not the locations* $\boldsymbol{x}$ *themselves*. On the other side the operator g^Ω is the restriction of g on the Euclidean space Ω. It describes how g deforms Ω. Unlike the independence of Ω with respect to g^ϕ in eq. (3.5) g^Ω naturally has an effect on the functions $\phi \in \Phi(\Omega)$, since they are functions on Ω.

Example 1. *An example for $\mathbb{G}^{\Omega\phi}$ is the set $\mathbb{G}^\phi \times SO(2)$ of variations of the functions $\phi \in \Phi(\Omega)$ and rotations on Ω (see section 2.5.1). The variations $g^\phi \in \mathbb{G}^\phi$ are parameterized by one parameter ξ^ϕ and a function $\boldsymbol{\omega}$ which vanishes on the boundary of Ω*

$$\phi_{g^\phi(\xi^\phi)}(\boldsymbol{x}) = \phi(\boldsymbol{x}) + \xi^\phi \boldsymbol{\omega}(\boldsymbol{x}), \quad \xi^\phi \in \mathbb{R}, \, \boldsymbol{\omega} \in \Phi(\Omega), \, \boldsymbol{\omega}|_{\partial\Omega} = 0 \tag{3.7}$$

An element $g \in \mathbb{G}^\phi \times SO(2)$ combines the action of g^ϕ in eq. (3.7) with that of the $SO(2)$

$$\phi_g(\boldsymbol{x}) = \phi(\boldsymbol{x}_\alpha) + \xi^\phi \boldsymbol{\omega}(\boldsymbol{x}_\alpha), \quad \boldsymbol{x}_\alpha = R_\alpha \boldsymbol{x}, \quad R_\alpha \in SO(2) \tag{3.8}$$

The local coordinates of g in eq. (3.8) are the variational parameter ξ^ϕ and the rotation angle α.

In example 1 the representation operators g^ϕ and g^Ω of g are mutually independent since g^ϕ is a function of ξ^ϕ and g^Ω is only a function of the angle α.

However for an arbitrary Lie group $\mathbb{G}^{\Omega\phi}$ acting on $\Phi(\Omega)$ as in eq. (3.4) we will consider both representations g^ϕ and g^Ω to depend on all n local coordinates (eq. (2.151)) $\xi^{i,g}$ of g.

Let $V \in \mathcal{G}^{\Omega\phi}$ be a smooth vector field on $\mathbb{G}^{\Omega\phi}$ and θ^V the corresponding flow (see eq. (2.154)) expressed in the basis of $\mathcal{G}^{\Omega\phi}$

$$V_g = v^i\left(\boldsymbol{\xi}^g\right)\partial_{\xi^{i,g}} \tag{3.9}$$

Like g in eq. (3.4) the flow θ^V acts simultaneously on $\Phi(\Omega)$ and Ω

$$\phi_{\theta^V(t,g)}(\boldsymbol{x}) := \phi_{\theta^{V^\phi}(t,g)}\left(\boldsymbol{x}_{\theta^{V^\Omega}(t,g)}\right) \tag{3.10}$$

and the action of V in ϕ_g decomposes in the following fashion

$$\frac{d}{dt}\phi_{\theta^V(t,g)}(\boldsymbol{x})\bigg|_{t=0} = V_g\phi_g(\boldsymbol{x}) = V_g^\phi\phi_g(\boldsymbol{x}) + V_g^\Omega\phi_g(\boldsymbol{x}) \tag{3.11}$$

$$V_g^\phi = v^i\left(\boldsymbol{\xi}^g\right)\partial^\phi_{\xi^{i,g}} \quad V_g^\Omega = v^i\left(\boldsymbol{\xi}^g\right)\partial^\Omega_{\xi^{i,g}} \tag{3.12}$$

The basis operators $\partial^\phi_{\xi^{i,g}}$ are restrictions of the basis $\partial_{\xi^{i,g}}$ onto the local coordinate vector $\boldsymbol{\xi}^g$ of the operator g^ϕ (eq. (3.5)). Similarly the basis operators $\partial^\Omega_{\xi^{i,g}}$ are restrictions of the basis $\partial_{\xi^{i,g}}$ onto the operator g^Ω (eq. (3.6)). However the coefficient functions $v^i\left(\boldsymbol{\xi}^g\right)$ are identical for both V^ϕ and V^Ω.

Let X^Ω be a smooth vector field in the restricted basis $\partial^\Omega_{\xi^{i,g}}$. Since X^Ω and V^ϕ act on different subspaces, they commute. Hence the rate of change of $X_g^\Omega\phi_g$ under the flow θ^V results in

$$\frac{d}{dt}X^\Omega_{\theta^V(t,g)}\phi_{\theta^V(t,g)}\bigg|_{t=0} = \left[V_g^\Omega, X_g^\Omega\right]\phi_g + X_g^\Omega(V_g\phi_g) \tag{3.13}$$

The first summand on the right hand side of eq. (3.13) is the Lie derivative (eq. (2.156)) of the vector field X_g^Ω along the flow θ^V of V. It only depends on the spatial component V^Ω of V. On the other side the second summand in eq. (3.13) depends on the entire vector field V.

In general the action of $\mathbb{G}^{\Omega\phi} \times \Omega$ on Ω in eq. (3.6) also induces a rescaling of the spatial integral measure d^2x by a divergence

$$\frac{d}{dt}\left(\theta_t^V \circ d^2x\right)\bigg|_{t=0} = \frac{dv^\mu}{dx^\mu}d^2x, \quad v^\mu = V_g^\Omega x^\mu_{g^\Omega} \tag{3.14}$$

However in the following sections we will only consider *volume-preserving* transformations

$$\frac{dv^{\mu}}{dx^{\mu}} d^2x = 0 \tag{3.15}$$

for the following reason: The transformation in eq. (3.14) will have no effect on the considerations to be made below and the new optimization scheme to be introduced in section 5. Therefore by assuming eq. (3.15) the following calculations are kept uncluttered.

The action of $\mathbb{G}^{\Omega\phi}$ on functionals of $\Phi(\Omega)$

The transformed functions $\phi_g(\boldsymbol{x})$ in eq. (3.4) depend on both the locations $\boldsymbol{x} \in \Omega$ and (the local coordinates $\boldsymbol{\xi}^g$ of) the Lie group elements $g \in \mathbb{G}^{\Omega\phi}$ (see eq. (3.6)). In the following we want to shift from viewing the elements $\phi \in \Phi(\Omega)$ as functions on Ω to viewing them as functions on $\mathbb{G}^{\Omega\phi}$. The reason is that we want to examine the action of the Lie group $\mathbb{G}^{\Omega\phi}$ on the functional $E(\phi_g, \boldsymbol{X}_g^{\Omega}\phi_g)$ in eq. (3.1). In eq. (3.1) the locations $\boldsymbol{x}$ are integrated away and $E(\phi_g, \boldsymbol{X}_g^{\Omega}\phi_g)$ only depends on $\phi \in \Phi(\Omega)$ as a whole and $g \in \mathbb{G}^{\Omega\phi}$. We view the Euclidean space Ω as *fixed* and consider the local coordinates of $g \in \mathbb{G}^{\Omega\phi}$ (see eq. (2.151)) as the variables of ϕ_g.

We begin with establishing a coordinate expression for the dual vector operator $\boldsymbol{X}^{\dagger,\Omega}$ on the dual space $\Phi^{\star}(\Omega)$ in eq. (3.3). Consider the vector field $X \in \mathcal{G}^{\Omega\phi}$. By lemma 14 X^{Ω}, the restriction of X on to Ω is also a derivation on $\Phi(\Omega)$. With the help of the scalar product in eq. (2.5) we want to show that there exists a *dual vector field* $X^{\dagger,\Omega}$ which acts on the dual space $\Phi^{\star}(\Omega)$

$$\left\langle X_g^{\dagger,\Omega} p, \phi \right\rangle = \left\langle p, X_g^{\Omega}\phi \right\rangle \tag{3.16}$$

The next lemma shows that there exists a dual basis $\partial^{\dagger,\Omega}_{\xi^{i,g}}$ in which $X^{\dagger,\Omega}$ can be expressed

Lemma 8 (Dual Basis)**.** *Let $\mathcal{G}$ be a Lie algebra such that for every $X \in \mathcal{G}$ the restricted vector field X^{Ω} (eq. (3.42)) is expressed in the restricted basis of $\mathcal{G}$*

$$X_g^{\Omega} = X^i(\boldsymbol{\xi}^g)\partial^{\Omega}_{\xi^{i,g}} \tag{3.17}$$

Then there exists a dual algebra $\mathcal{G}^{\star}$ with elements $X^{\star}$ which are linear combinations of the basis of $\mathcal{G}^{\star}$

$$X_g^{\dagger,\Omega} = X^i(\boldsymbol{\xi}^g)\partial^{\dagger,\Omega}_{\xi^{i,g}} \tag{3.18}$$

The basis elements $\partial^{\star}_{\xi^{i,g}}$ have the form

$$\partial^{\dagger,\Omega}_{\xi^{i,g}} p(\boldsymbol{x}') = -\partial'_{\mu}\left(\frac{\partial x'^{\mu}}{\partial \xi^{i,g}} p(\boldsymbol{x}')\right), \quad \boldsymbol{x}' = g^{\Omega} \circ \boldsymbol{x},\, p \in \Phi^{\star}(\Omega) \tag{3.19}$$

Proof. We begin by proving the existence of the dual basis $\partial^{\dagger,\Omega}_{\xi^{i,g}}$ of the dual algebra $\mathcal{G}^{\star}$. Let $\boldsymbol{x}' = g^{\Omega} \circ \boldsymbol{x}$. Then by the chain-rule we can express the operators $\partial^{\Omega}_{\xi^{i,g}}$ in terms of the Cartesian derivatives ∂_{μ}

$$\partial^{\Omega}_{\xi^{i,g}} = \frac{\partial x'^{\mu}}{\partial \xi^{i,g}} \partial'_{\mu} \tag{3.20}$$

Let $\phi \in \Phi(\Omega)$ and $p \in \Phi^{\star}(\Omega)$ such that the condition

$$p(\boldsymbol{x}) = 0, \quad \forall \boldsymbol{x} \in \partial\Omega \tag{3.21}$$

holds. Then it follows that

$$\left\langle p, \partial^{\Omega}_{\xi^{i,g}} \phi \right\rangle = \int p(\boldsymbol{x}') \partial^{\Omega}_{\xi^{i,g}} \phi(\boldsymbol{x}') d^2x' = \int p(\boldsymbol{x}') \frac{\partial x'^{\mu}}{\partial \xi^{i,g}} \partial'_{\mu} \phi(\boldsymbol{x}') d^2x' \tag{3.22}$$

$$= -\int \partial'_{\mu}\left(\frac{\partial x'^{\mu}}{\partial \xi^{i,g}} p(\boldsymbol{x}')\right) \phi(\boldsymbol{x}') d^2x' = \left\langle \partial^{\dagger,\Omega}_{\xi^{i,g}} p, \phi \right\rangle \tag{3.23}$$

where we used condition eq. (3.21) and integration by parts in eq. (3.23). Eq. (3.23) proves the existence of the dual basis. We consider the vector field $X \in \mathcal{G}$ in eq. (3.17). Note that the coefficient functions $X^i(\boldsymbol{\xi}^g)$ only depend on the local coordinates of g and not on the coordinates $\boldsymbol{x} \in \Omega$. Thus we have

$$\left\langle p, X^{\Omega}_{g} \phi \right\rangle = \sum_i X^i(\boldsymbol{\xi}^g) \left\langle p, \partial^{\Omega}_{\xi^{i,g}} \phi \right\rangle \tag{3.24}$$

using eq. (3.23) we obtain

$$\sum_i X^i(\boldsymbol{\xi}^g) \left\langle p, \partial^{\Omega}_{\xi^{i,g}} \phi \right\rangle = \sum_i X^i(\boldsymbol{\xi}^g) \left\langle \partial^{\dagger,\Omega}_{\xi^{i,g}} p, \phi \right\rangle = \left\langle X^{\dagger,\Omega}_{g} p, \phi \right\rangle \tag{3.25}$$

which proves eq. (3.18) □

Now we can compute the rate of change of the energy functional $E(\phi_g, \boldsymbol{X}^{\Omega}_{g} \phi_g)$ under the flow θ^V corresponding to the vector field V in eq. (3.11)

$$\frac{d}{dt} E\left(\phi_{\theta^V(t,g)}, \boldsymbol{X}^{\Omega}_{\theta^V(t,g)} \phi\right)\Big|_{t=0} \tag{3.26}$$

However since E is not necessarily smooth in ϕ_g the derivative in eq. (3.26) is not defined. Nevertheless it is possible to compute the subdifferential of E with respect to the flow parameter t. We first consider how the function $E^{data}(\phi)$ transforms under $\theta^V(t,g)$ and define the subdifferential $\partial_t^V E^{data}$ in the following fashion

Definition 19 (Flow Subdifferential)**.** *Let $\partial E^{data}(\phi_g)$ be the subdifferential of E^{data} at ϕ_g. The flow subdifferential $\partial_t^V E^{data}$ of $E^{data}(\phi_g)$ along the flow θ^V is defined by*

$$\partial_t^V E^{data}(\phi)|_{t=0} = \left\{ \left\langle p^d, V_g\phi \right\rangle \middle| \, p^d \in \partial E^{data}(\phi_g) \right\} \tag{3.27}$$

$\partial_t^V E^{data}(\phi)|_{t=0}$ in eq. (3.27) is a subset of $\mathbb{R}$.

In the case when $E^{data}(\phi)$ is smooth with respect to ϕ the subdifferential ∂E^{data} contains only one element, namely the functional derivative of E^{data} with respect to ϕ, $\partial E^{data} = \left\{ \frac{\delta E^{data}}{\delta \phi} \right\}$ and eq. (3.27) reduces to the chain-rule

$$\partial_t^V E^{data}(\phi) = \left\{ \frac{d}{dt} E^{data}\left(\phi_{\theta(t,g)}\right)\Big|_{t=0} \right\} = \left\{ \left\langle \frac{\delta E^{data}}{\delta \phi}, V_g\phi \right\rangle \right\} \tag{3.28}$$

Next we consider the rate of change of $E^{prior}(\boldsymbol{X}_g^{\Omega}\phi)$ under the flow θ^V. The flow subdifferential $\partial_t^V E^{prior}$ is similar to eq. (3.27) when we take eq. (3.13) into account

$$\partial_t^V E^{prior} = \left\{ \sum_{j=1}^{2} \left\langle p_j^{pr}, \left[V_g^{\Omega}, X_g^{\Omega,j}\right] \phi_g + X_g^{\Omega,j}\left(V_g\phi_g\right) \right\rangle \middle| \, \boldsymbol{p}^{pr} \in \partial E^{prior} \right\} \tag{3.29}$$

$$= \left\{ \sum_{j=1}^{2} \left[\left\langle p_j^{pr}, \left[V_g^{\Omega}, X_g^{\Omega,j}\right] \phi_g \right\rangle - \left\langle \mathrm{Div}(\boldsymbol{p}_g^{pr}), V_g\phi_g \right\rangle \right] \middle| \, \boldsymbol{p}^{pr} \in \partial E^{prior} \right\} \tag{3.30}$$

where Div is the divergence with respect to the dual operator $\boldsymbol{X}_g^{\dagger,\Omega}$ which is defined in eq. (3.19)

$$\mathrm{Div}(\boldsymbol{p}_g^{pr}) = -\boldsymbol{X}_g^{\dagger,\Omega}\boldsymbol{p}_g^{pr} \tag{3.31}$$

We combine the expressions for $\partial_t^V E^{data}$ and $\partial_t^V E^{prior}$ in eqs. (3.27) and (3.30) to an expression for the total flow subdifferential of the functional $E(\phi, \boldsymbol{X}\phi)$

$$\partial_t^V E(\phi_g, \boldsymbol{X}_g^\Omega \phi_g) = \partial_t^V E^{data}(\phi) + \partial_t^V E^{prior}(\boldsymbol{X}_g^\Omega \phi) \tag{3.32}$$

$$= \left\{ A(\boldsymbol{p}^{pr}, p^d) + B(\boldsymbol{p}^{pr}) \middle| p^d \in \partial E^{data}(\phi_g), p_i^{pr} \in \partial^i E^{prior}(\boldsymbol{X}_g^\Omega \phi_g) \right\} \tag{3.33}$$

$$A = \left\langle [\mathcal{E}]_{p^d, \boldsymbol{p}^{pr}}, V_g \phi_g \right\rangle, \quad B = \sum_{j=1}^{2} \left\langle p_j^{pr}, \left[V_g^\Omega, X_g^{\Omega,j} \right] \phi_g \right\rangle \tag{3.34}$$

$$[\mathcal{E}]_{p^d, \boldsymbol{p}^{pr}}(\boldsymbol{x}) = p^d(\boldsymbol{x}_{g^\Omega}) - \mathrm{Div}(\boldsymbol{p}_g^{pr})(\boldsymbol{x}_{g^\Omega}) \tag{3.35}$$

We now make some remarks on eq. (3.33). The subdifferential $\partial_t^V E$ has two constituents: The first constituent $A(\boldsymbol{p}^{pr}, p^d)$ is the scalar product of the density $[\mathcal{E}]$ with the variation of ϕ, $V_g\phi$. $[\mathcal{E}]$ is a combination of elements of the subdifferentials of the data term E^{data} and the prior E^{prior}. It is a generalization of the Euler-Lagrange differentials in eq. (2.172) to the case of non-smooth functionals. On the other side the second constituent of $\partial_t^V E$, $B(\boldsymbol{p}^{pr})$ is the scalar product of the subdifferential of E^{prior} alone with the smooth vector field $\left[V^\Omega, \boldsymbol{X}^\Omega\right]$ acting on ϕ. While A in eq. (3.35) encodes the *total* contribution to $\partial_t^V E$ from the variation of the function $\phi \in \Phi(\Omega)$, $V_g\phi$, the component B in eq. (3.34) encodes the contribution of the vector field $\boldsymbol{X}^\Omega$ alone, meaning *independently of* ϕ, to $\partial_t^V E$. In the case when V^Ω and $\boldsymbol{X}^\Omega$ commute

$$\left[V, \boldsymbol{X}^\Omega\right] = 0 \tag{3.36}$$

B in eq. (3.34) vanishes and $\partial_t^V E$ is dominated by $V_g\phi$. Eq. (3.36) is what we call a *trivial symmetry* of the energy $E(\phi_g, \boldsymbol{X}_g^\Omega \phi)$. In section 3.2 we will show that it is possible for B to vanish with non zero commutator $\left[V^\Omega, \boldsymbol{X}^\Omega\right]$ yielding a *non-trivial* symmetry. The non-trivial symmetries are at the core of our modernization of Noether's Theorem.

3.2. Noether's First Theorem: A Modern Version

The energy functional $E(\phi, \nabla\phi)$ considered in Emmy Noethers original paper [73] (section 2.6) was taken to be smooth in the function ϕ. In section 3.1 we derived the action of the Lie group $\mathbb{G}^{\Omega\phi}$ on the functional $E(\phi_g, \boldsymbol{X}_g^\Omega \phi_g)$ in eq. (3.1) with the following result for the rate of change of E along the flow θ^V

$$
\begin{aligned}
\partial_t^V E(\phi_g, \boldsymbol{X}_g^\Omega \phi_g) &= \partial_t^V E^{data}(\phi) + \partial_t^V E^{prior}(\boldsymbol{X}_{g^\Omega}^\Omega \phi) \\
&= \left\{ A(\boldsymbol{p}^{pr}, p^d) + B(\boldsymbol{p}^{pr}) \middle| p^d \in \partial E^{data}(\phi), \boldsymbol{p}_g^{pr} \in \partial E^{prior} \right\} \qquad (3.37)
\end{aligned}
$$

$$
A = \left\langle [\mathcal{E}]_{p^d, \boldsymbol{p}^{pr}}, V_g \phi_g \right\rangle, \quad [\mathcal{E}]_{p^d, \boldsymbol{p}^{pr}}(\boldsymbol{x}) = p^d(\boldsymbol{x}) - \operatorname{Div}(\boldsymbol{p}_g^{pr})(\boldsymbol{x}) \qquad (3.38)
$$

$$
B = \sum_{j=1}^{2} \left\langle p_{g,j}^{pr}, \left[V_g^\Omega, X_g^{\Omega,j} \right] \phi_g \right\rangle \qquad (3.39)
$$

Our goal is to recover the classical identity in eq. (2.172) from eq. (3.37) by considering the functional E in eq. (3.37) to be smooth in ϕ_g and $\boldsymbol{X}_g^\Omega \phi_g$. In this case the dual variables p^d and $\boldsymbol{p}^{pr}$ in eq. (3.37) are unique and eq. (3.37) simplifies to

$$
\partial_t^V E(\phi_g, \boldsymbol{X}_g^\Omega \phi_g) = \left\{ \int_\Omega \left(\sum_{i=1}^{2} p_i^{pr} \left[V_g^\Omega, X_g^{i,\Omega} \right] \phi_g + [\mathcal{E}] \, V_g \phi_g \right) d^2 x \right\} \qquad (3.40)
$$

$$
[\mathcal{E}] = \frac{\delta \mathcal{E}}{\delta \phi_g} - \operatorname{Div}(\boldsymbol{p}_g^{pr}), \quad \boldsymbol{p}_g^{pr} = \frac{\delta \mathcal{E}^{prior}}{\delta \boldsymbol{X}_g^\Omega \phi_g} \qquad (3.41)
$$

We use the expansion V and V^Ω in the basis of $\mathcal{G}^{\Omega\phi}$ in eq. (3.11) and define the variations ω_i^ϕ and $\boldsymbol{\omega}_i^\Omega$ as the actions of the basis of $\mathcal{G}^{\Omega\phi}$ on ϕ_g and $\boldsymbol{x}_{g^\Omega}$

$$
\begin{aligned}
V_g \phi_g(\boldsymbol{x}) &= v^i(\boldsymbol{\xi}^g) \omega_i^\phi(\boldsymbol{x}), \quad \omega_i^\phi(\boldsymbol{x}) := \partial_{\xi^{i,g}} \phi_g(\boldsymbol{x}) \\
V_g^\Omega \boldsymbol{x}_{g^\Omega} &= v^i(\boldsymbol{\xi}^g) \boldsymbol{\omega}_i^\Omega(\boldsymbol{x}), \quad \boldsymbol{\omega}_i^\Omega(\boldsymbol{x}) := \partial_{\xi^{i,g}}^\Omega \boldsymbol{x}_{g^\Omega} \qquad (3.42)
\end{aligned}
$$

such that eq. (3.40) translates to

$$
\partial_t^V E(\phi_g, \boldsymbol{X}_g^\Omega \phi_g) = \left\{ \int_\Omega \sum_m v^i(\boldsymbol{\xi}^g) \left(\boldsymbol{B}_g^{\Omega,m} \phi_g + [\mathcal{E}] \, \omega_m^\phi \right) d^2 x \right\} \qquad (3.43)
$$

$$
\boldsymbol{B}_g^{\Omega,m} = \sum_{i=1}^{2} p_i^{pr} \left[\partial_{\xi^{m,g}}^\Omega, X_g^{i,\Omega} \right] \qquad (3.44)
$$

The operator $\boldsymbol{B}^{\Omega,m}$ can be expressed in the basis of $\mathcal{X}^\Omega$

$$
\boldsymbol{B}_g^{\Omega,m} = \boldsymbol{b}_g^{\Omega,m,T} \boldsymbol{X}_g^\Omega, \quad b_{g,j}^{\Omega,m} = C_{m,i}^j p_i^{pr} \qquad (3.45)
$$

The constants $C_{m,i}^j$ are the structure constants (definition 46) from the commutator in eq. (3.44). With the definitions of ω_i^ϕ and $\boldsymbol{\omega}_i^\Omega$ we prove in appendix B the following identity

$$\operatorname{div}(\boldsymbol{W}_m) - [\mathcal{E}]\,\omega_m^{\Omega,\mu}\partial_\mu\phi_g = \boldsymbol{B}_g^{\Omega,m}\phi_g \tag{3.46}$$

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

$$W_m^\mu = \omega_m^{\Omega,\mu}\mathcal{E} + \widetilde{\omega}_m^\phi \sum_{i=1}^{2} \omega_i^{X,\mu} p_i^{pr}, \quad \boldsymbol{\omega}_i^X = X_g^{\Omega,i}\boldsymbol{x}_{g^\Omega}, \quad \widetilde{\omega}_m^\phi = \omega_m^\phi - \omega_m^{\Omega,\nu}\partial_\nu\phi_g \tag{3.47}$$

The functions $(\partial_\nu\phi)_g$ are the components of the Cartesian gradient $\nabla\phi$ under the action of g (see eq. (3.4)). With eq. (3.46) the differential $\partial_t^V E$ in eq. (3.37) transforms to

$$\partial_t^V E(\phi_g, \boldsymbol{X}_g^\Omega\phi_g) = \left\{ \int_\Omega \sum_m v^m(\boldsymbol{\xi}^g)\left(\operatorname{div}(\boldsymbol{W}_m) + \widetilde{\omega}_m^\phi[\mathcal{E}]\right) d^2x \right\} \tag{3.48}$$

We assume $E(\phi_g, \boldsymbol{X}_g^\Omega\phi_g)$ to be invariant under the flow θ^V of any vector field $V \in \mathcal{G}^{\Omega\phi}$. Thus the integrand in eq. (3.48) must vanish for any coefficient functions $v^m(\boldsymbol{\xi}^g)$

$$\partial_t^V E(\phi_g, \boldsymbol{X}_g^\Omega\phi_g) = \{0\} \Leftrightarrow \left(\operatorname{div}(\boldsymbol{W}_m) + \widetilde{\omega}_m^\phi[\mathcal{E}]\right) = 0 \tag{3.49}$$

Eq. (3.49) is equivalent to eq. (2.175). In section 2.6 the variations ω_i^ϕ to the function ϕ and the variations $\boldsymbol{\omega}_i^\Omega$ of the Euclidean space Ω from eq. (3.42) are assumed to be mutually independent. It was explained that $[\mathcal{E}] = 0$ and $\operatorname{div}(\boldsymbol{W}_m) = 0$ must hold individually so that eq. (3.49) can only hold if ϕ_g minimizes $E(\phi_g, \boldsymbol{X}_g^\Omega\phi_g)$.

In our framework the mutual independence of the variations ω_i^ϕ and $\boldsymbol{\omega}_i^\Omega$ is identical to the mutual independence of the action $V_g\phi_g$ and the restricted action $V_g^\Omega\phi_g$ for any $V \in \mathcal{G}^{\Omega\phi}$ since by eq. (3.13)

$$V_g^\phi\phi_g = V_g\phi_g - V_g^\Omega\phi_g, \quad V_g^\Omega\phi_g \perp V_g\phi_g \tag{3.50}$$

and thus $V_g^\phi\phi_g$ makes up for any differences between $V_g\phi_g$ and $V_g^\Omega\phi_g$. The invariance of $E(\phi_g, \boldsymbol{X}_g^\Omega\phi_g)$ under the flow θ^V results by eq. (3.43) in

$$\partial_t^V E = \{0\} \Leftrightarrow \boldsymbol{B}_g^{\Omega,m}\phi_g = 0, \quad [\mathcal{E}] = 0 \tag{3.51}$$

While $[\mathcal{E}] = 0$ only holds for the minimizers $\phi^\star$ of E as in eq. (3.49), the condition $\boldsymbol{B}_g^{\Omega,m}\phi_g = 0$ holds for any ϕ_g since it is a property of the prior E^{prior} of the model as the following section shows.

3.2.1. Pure spatial Symmetries

We shall now study the special subset $\mathcal{G}^{\Omega} \subset \mathcal{G}^{\Omega\phi}$ of smooth vector fields $Q \in \mathcal{G}^{\Omega}$ which act in such a way as to leave the function space $\Phi(\Omega)$ invariant

Definition 20 (Pure spatial Algebra $\mathcal{G}^{\Omega}$)**.** *The pure spatial algebra $\mathcal{G}^{\Omega}$ is defined as the set of smooth vector fields $Q \in \mathcal{G}^{\Omega}$ for which there exist a $\phi \in \Phi(\Omega)$ such that*

$$\frac{d}{dt}\phi_{\theta^Q(t,g)}\bigg|_{t=0} = Q_g\phi_g = 0, \quad \forall g \in \mathbb{G}^{\Omega\phi} \tag{3.52}$$

holds.

The condition in Eq. (3.52) does *not* confine $\phi_g(\boldsymbol{x})$ to be constant in Ω. It is rather a condition on the restriction Q_g^{ϕ} (eq. (3.11)) that its action on ϕ_g cancels that of the spatial action Q_g^{Ω}

$$Q_g^{\phi}\phi_g = -Q_g^{\Omega}\phi_g \tag{3.53}$$

The intuitive explanation of eqns. (3.52) and (3.53) goes as follows. The flow θ^Q traces out a path $\boldsymbol{x}(t) = \boldsymbol{x}_{\theta^Q(t,g^{\Omega})}$. Due to eq. (3.52) the function values $\phi_{\theta^Q(t,g)}(\boldsymbol{x})$ are *dragged* along the path $\boldsymbol{x}(t)$.

Now let the energy functional $E(\phi_g, \boldsymbol{X}_g^{\Omega}\phi_g)$ be invariant along the flow θ^Q of a smooth vector field $Q \in \mathcal{G}^{\Omega}$ with $Q_g = q^i(\boldsymbol{\xi}^g)\partial_{\xi^i,g}$. From eqns. (3.40) and (3.52) it follows that

$$\partial_t^Q E = \{0\} \Longleftrightarrow \int_{\Omega}\sum_{m=1}^{n} q^m(\boldsymbol{\xi}^g)\boldsymbol{B}_g^{\Omega,m}\phi_g d^2x = 0, \quad \forall\phi \in \Phi(\Omega), g \in \mathbb{G}^{\Omega\phi} \tag{3.54}$$

If eq. (3.54) holds for all $Q \in \mathcal{G}^{\Omega}$ it then follows that

$$\boldsymbol{B}_g^{\Omega,m}\phi_g = \sum_{i=1}^{2} p_i^{pr}\left[\partial^{\Omega}_{\xi^m,g}, X_g^{i,\Omega}\right]\phi_g = 0 \tag{3.55}$$

Eq. (3.55) is specifically a constraint on the prior energy E^{prior} since the dual variable $\boldsymbol{p}^{pr}$ (eq. (3.41)) only depends on the prior energy density $\mathcal{E}^{prior}$ from eq. (3.2). There are three cases to consider such that eq. (3.55) can hold:

- Case a: The Lie algebra $\mathcal{G}^{\Omega}$ commutes with that of $\mathcal{X}^{\Omega}$, $[\partial^{\Omega}_{\xi^m,g}, X_g^{\Omega,i}] = 0$ for $i = \{1, 2\}$
- Case b: $p_i^{pr} = 0$ for $i = \{1, 2\}$

- Case c: If we have $[\partial^{\Omega}_{\xi^{i,g}}, X^{\Omega,m}_g] \neq 0$ for some i and m the dual variable $\boldsymbol{p}^{pr}$ if non-vanishing must be orthogonal to the vector $\mathbf{M}_m$, which is a vector valued function over Ω for fixed m defined as $(\mathbf{M}_m)_i = [\partial^{\Omega}_{\xi^{i,g}}, X^{\Omega,m}_g]\phi_g$.

We call cases a and b trivial symmetries and case c a non-trivial symmetry.

Example 2 (Symmetries of the prior $E^{prior}_{L_2}(\nabla\phi)$)**.** *The prior $E^{prior}_{L_2}(\nabla\phi)$ in section 2.4.1 is constructed with the basis $\{\partial_x, \partial_y\}$ of the algebra $\mathfrak{t}$ of the translation group $\mathbb{T}^{\Omega}$ (section 2.5.1). It is invariant under the group of rotations and translation $\mathbb{G}^{\Omega} = \mathbb{T}^{\Omega} \times SO(2)$. The invariance under translations is a trivial symmetry since $[\partial_x, \partial_y] = 0$. However the subgroup $SO(2)$ is different: The gradient $\nabla\phi_g$ is turned by $90°$ counter-clockwise by the basis of $\mathfrak{so}(2)$ (eq. (2.165))*

$$\nabla^{\perp}\phi_g = [\partial_\theta, \nabla]\,\phi_g, \quad g \in \mathbb{T}^{\Omega} \times SO(2) \tag{3.56}$$

In addition the sub differential of $E^{prior}_{L_2}$ has one element, $\partial E^{prior}_{L_2}(\nabla\phi_g) = \{\boldsymbol{p}^{pr} = \nabla\phi_g\}$ and we have from the definition of the operator $\boldsymbol{B}$ in eq. (3.44)

$$\boldsymbol{B}_\theta\phi_g = \left(\boldsymbol{p}^{pr,T} \cdot \nabla^{\perp}\right)\phi_g = 0 \tag{3.57}$$

Hence $SO(2)$ is a non-trivial symmetry group of the prior $E^{prior}_{L_2}$.

In chapter 4 we will introduce a prior $E^{prior}_{ST}(\nabla\phi_g)$ which is invariant to the group $\mathbb{G}^{\Omega} = \mathbb{T} \times SO(2)$ based on the structure tensor [7, 6, 8] and preserves edges in ϕ (in contrast to $E^{prior}_{L_2}$).

3.3. Embedding Geometrical Constraints into Prior Energies

In section 2.4.1 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

$$A = \{\phi^\star |\, \phi^\star = \mathrm{argmin}_\phi \left(-E^{prior}\left(\nabla\phi\right)\right)\} \tag{3.58}$$

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. (3.58) be generated by the group $\mathbb{G}$

$$A = \{\phi^\star |\, \phi^\star = g_\omega \circ \phi^\star_0 \quad g_\omega \in \mathbb{G}\} \tag{3.59}$$

In eq. (2.141) 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^Ω on the coordinate frame Ω.

We consider the $E(\phi_g, \boldsymbol{X}_g^\Omega \phi_g)$ in eq. (3.1) in which we set $E^{data} = 0$.

$$E(\phi_g, \boldsymbol{X}_g^\Omega \phi_g) = E^{prior}(\boldsymbol{X}_g^\Omega \phi_g) \tag{3.60}$$

We would like to make an assessment on what the precise conditions are for the minimizer set $A^{\boldsymbol{X}_g^\Omega}$

$$A^{\boldsymbol{X}_g^\Omega} = \left\{ \phi_g^\star \middle| \, \phi_g^\star = \operatorname{argmin}_{\phi_g} \left(-E^{prior} \left(\boldsymbol{X}_g^\Omega \phi_g \right) \right) \right\} \tag{3.61}$$

to exist. By the Kuhn-Tucker conditions in eq. (3.3) there exists a vector $\boldsymbol{p}_g^{\star,pr}$ in the subdifferential of E^{prior} at every $\phi_g^{\star,pr} \in A^{\boldsymbol{X}_g^\Omega}$ for which

$$\mathrm{Div}(\boldsymbol{p}_g^{\star,pr}) = 0, \quad \boldsymbol{p}_g^{\star,pr} \in \partial E^{prior}(\boldsymbol{X}_g^\Omega \phi_g^\star), \quad \forall \phi_g^\star \in A^{\boldsymbol{X}_g^\Omega} \tag{3.62}$$

holds. We will call eq. (3.62) the *free Kuhn-Tucker conditions*. The following proposition shows that $\boldsymbol{p}_g^{\star,pr}$ is *constant* with respect to the derivations $\left\{ X_g^{\Omega,1}, X_g^{\Omega,2} \right\}$ if $E^{prior}(\boldsymbol{X}_g^\Omega \phi_g)$ is invariant under the pure spatial group $\mathbb{G}^\Omega$.

Proposition 4 (Constance of the Dual Space). *Let $\mathbb{G}^\Omega$ be a symmetry group of $E^{prior}(\boldsymbol{X}_g^\Omega \phi_g)$ such that eq. (3.55) holds. The local coordinates $\boldsymbol{\xi}^g$ of any $g \in \mathbb{G}^\Omega$ are taken to have the form $\boldsymbol{\xi}^g = \left\{ \xi_X^{1,g}, \xi_X^{2,g}, \xi^{3,g}, \cdots, \xi^{n,g} \right\}$ such that the differential operators $\boldsymbol{X}_g^\Omega$ have the canonical form*

$$X_g^{\Omega,1} = \partial_{\xi_X^{1,g}}, \quad X_g^{\Omega,2} = \partial_{\xi_X^{2,g}} \tag{3.63}$$

Furthermore let $\phi_g^{\star,pr} \in A^{\boldsymbol{X}_g^\Omega}$ be a minimizer of E^{prior} such that there exists a vector $\boldsymbol{p}_g^{\star,pr}$ in the subdifferential of E^{prior} at $\phi_g^{\star,pr}$ for which

$$Div(\boldsymbol{p}_g^{\star,pr}) = 0, \quad p_{g,j}^{\star,pr} \in \partial^j E^{prior}(\boldsymbol{X}_g^\Omega \phi_g^\star), \quad \forall \phi_g^\star \in A^{\boldsymbol{X}_g^\Omega} \tag{3.64}$$

holds by the Kuhn-Tucker conditions. The dual variable $\boldsymbol{p}_g^{\star,pr}$ is then constant with respect to the sub coordinates $\left\{ \xi_X^{1,g}, \xi_X^{2,g} \right\}$.

We have to postpone the proof of proposition 4 to section 5.2 where we will introduce the concept of a curvature operator $\boldsymbol{K}$ which operates on the function space $\Phi(\Omega)$.

Due to the invariance of E^{prior} under the pure spatial group $\mathbb{G}^\Omega$ eq. (3.55) must hold. We expand the commutator in eq. (3.55) with respect to the basis operators $\boldsymbol{X}_g^{\Omega} = \left\{X_g^{\Omega,1}, X_g^{\Omega,2}\right\}$

$$\boldsymbol{B}_g^{\star,\Omega,m}\phi_g^{\star} = \boldsymbol{b}_g^{\star,\Omega,m,T}\boldsymbol{X}_g^{\Omega}\phi_g^{\star} = 0 \tag{3.65}$$

The coefficient vector $\boldsymbol{b}^{\star,\Omega,m,T}$ is the optimum of the vector $\boldsymbol{b}^{\Omega,m,T}$ in eq. (3.45). Eq. (3.65) is a condition for $\phi_g^{\star}$, called a *level-set* equation, that holds for all $g \in \mathbb{G}^{\Omega\phi}$. In the following we shall examine the geometry of the minimizers $\phi_g^{\star}$ which satisfy eq. (3.65) and argue that the operators $\boldsymbol{X}_g^{\Omega}$ commute with each other.

Definition 21 (Level-set)**.** *Let $\mathbb{X}^{\Omega}$ be a two dimensional group with the algebra $\mathcal{X}^{\Omega}$. Each element $h^{\Omega} \in \mathbb{X}^{\Omega}$ has two local coordinates $h^{\Omega} = \left(\xi_X^{1,h}, \xi_X^{2,h}\right)$. Furthermore let $Z^{\Omega} \in \mathcal{X}^{\Omega}$ such that there exists a function $\phi^Z \in \Phi(\Omega)$ which is constant under the flow $\theta^{Z^{\Omega}}$*

$$\frac{d}{dt}\phi^Z\left(\boldsymbol{x}_{\theta^{Z^{\Omega}}(t,h^{\Omega})}\right)\bigg|_{t=0} = Z_{h^{\Omega}}^{\Omega}\phi^Z(\boldsymbol{x}_{h^{\Omega}}) = 0, \quad \boldsymbol{x}_{h^{\Omega}} = h^{\Omega}\circ\boldsymbol{x}, \forall h^{\Omega} \in \mathbb{X}^{\Omega} \tag{3.66}$$

for any location $\boldsymbol{x} \in \Omega$. The level-set $S_Z(\boldsymbol{x}) \subset \Omega$ through the point $\boldsymbol{x}$ is defined as the set of locations $\boldsymbol{x}_{h^{\Omega}} \in \Omega$ which satisfy eq. (3.66)

$$S_Z(\boldsymbol{x}) = \left\{\boldsymbol{x}_{h^{\Omega}} = h^{\Omega}\circ\boldsymbol{x}\,\middle|\, Z_{h^{\Omega}}^{\Omega}\phi^Z(\boldsymbol{x}_{h^{\Omega}}) = 0, \quad h^{\Omega} \in \mathbb{X}^{\Omega}\right\} \tag{3.67}$$

Since $\mathbb{X}^{\Omega}$ is a Lie group it contains the identity $e \in \mathbb{X}^{\Omega}$. Hence the point $\boldsymbol{x}$ is an element of $S_Z(\boldsymbol{x})$, $\boldsymbol{x} \in S_Z(\boldsymbol{x})$.

The definition of the level-set S_Z in eq. (3.67) is based on the assumption that the function ϕ^Z exists for which the condition in eq. (3.66) holds. In [6, 34] it was shown that a function $\phi^Z \in \Phi(\Omega)$ may exist with non empty level-set S_Z (eq. (3.67)) if and only if the following lemma holds

Lemma 9 (Existence of ϕ^Z)**.** *The function ϕ^Z in eq. (3.67) exists for $S_Z \neq \emptyset$ if and only if $\mathcal{X}^{\Omega}$ is a commutative algebra*

$$S_Z \neq \emptyset \Leftrightarrow \left[Z^{\Omega}, W^{\Omega}\right] = 0, \quad \forall Z^{\Omega}, W^{\Omega} \in \mathcal{X}^{\Omega} \tag{3.68}$$

Thus the coefficient functions of the vector fields $Z^{\Omega} \in \mathcal{X}^{\Omega}$ are constants with respect to the local coordinates $\left\{\xi_{\Omega}^{1,b}, \xi_{\Omega}^{2,b}\right\}$ of h^{Ω}

$$Z_{h^{\Omega}}^{\Omega} = z^i\partial_{\xi_{\Omega}^{i,h}}, \quad z^i = const, \quad i \in \{1,2\}, \quad \forall h^{\Omega} \in \mathbb{X}^{\Omega} \tag{3.69}$$

and the flow θ^{Z^Ω} of Z^Ω expressed in local coordinates $\left\{\xi^{1,h}, \xi^{2,h}\right\}$ is a line

$$\theta^{Z^\Omega}(t, h^\Omega) = \boldsymbol{\xi}^h + t\boldsymbol{z} \tag{3.70}$$

An illustrative example is the translation group $\mathbb{T}$

$$T^\Omega \in \mathbb{T},\ \boldsymbol{x}_{T^\Omega} = \boldsymbol{x} + \boldsymbol{t} \tag{3.71}$$

The algebra $\mathfrak{t}$ of $\mathbb{T}^\Omega$ is spun by the Cartesian operators $\{\partial_x, \partial_y\}$ (see section 2.5.1) and the vector fields $V^\Omega \in \mathfrak{t}$ have the form

$$V^\Omega_{T^\Omega} = v^1\partial_{t^1} + v^2\partial_{t^2} \in \mathfrak{t} \tag{3.72}$$

and the corresponding level-sets $S^{lin}_{\boldsymbol{v}}$ are the straight lines with tangential vector $\boldsymbol{v} = (v^1, v^2)$

$$S^{lin}_{\boldsymbol{v}}(\boldsymbol{x}) = \left\{\boldsymbol{x}_{T^\Omega} \middle|\, V^\Omega_{T^\Omega}\phi\left(\boldsymbol{x}_{T^\Omega}\right) = 0, T^\Omega \in \mathbb{T}\right\} \tag{3.73}$$

The tangential vector $\boldsymbol{v}$ is a constant function in Ω so that the level-set $S^{lin}_{\boldsymbol{v}}(\boldsymbol{x}) \subset \Omega$ is a collection of lines $\boldsymbol{x}_{\boldsymbol{t}}(t)$, each with the offset $\boldsymbol{x}_{T^\Omega}$ and the orientation $\boldsymbol{v}$

$$\boldsymbol{x}_{T^\Omega}(t) = \boldsymbol{x}_{T^\Omega} + t\boldsymbol{v} \tag{3.74}$$

From lemma 9 and proposition 4 it follows that the basis operators $\boldsymbol{X}^\Omega_g = \left\{X^{\Omega,1}_g, X^{\Omega,2}_g\right\}$ which are used to defined the prior $E^{prior}(\boldsymbol{X}^\Omega_g\phi_g)$ must commute. Furthermore the level-sets S_B corresponding to $\boldsymbol{B}^{\star,\Omega,m}$ in eq. (3.65)

$$S_B(\boldsymbol{x}) = \left\{\boldsymbol{x}_{g^\Omega} = g^\Omega \circ \boldsymbol{x} \middle|\, \boldsymbol{b}^{\star,\Omega,m,T}_g \boldsymbol{X}^\Omega_g \phi^\star_g(\boldsymbol{x}_{h^\Omega}) = 0, \quad g \in \mathbb{G}^{\Omega\phi}\right\} \tag{3.75}$$

are lines in the direction of the tangential vector $\boldsymbol{b}^{\star\Omega,m}_g$ and thus $\boldsymbol{B}^{\star,\Omega,m}$ commutes with $\boldsymbol{X}^\Omega$, $\boldsymbol{B}^{\star,\Omega,m} \in \mathcal{X}^\Omega$. Since the ideal $\mathcal{X}^\Omega$ is closed under the commutator by definition 18 the transformation of $\boldsymbol{B}^{\star,\Omega,m}$ under the flow of *any* vector field $V^\Omega \in \mathcal{G}^\Omega$ resides in $\mathcal{X}^\Omega$

$$\widetilde{\boldsymbol{B}}^{\star,\Omega,m} = \left[V^\Omega, \boldsymbol{B}^{\star,\Omega,m}\right] \in \mathcal{X}^\Omega \tag{3.76}$$

From eq. (3.76) it follows that a function $\widetilde{\phi}^\star_g$ which satisfies $\widetilde{\boldsymbol{B}}^{\star,\Omega,m}_g\widetilde{\phi}^\star_g = 0$ is also a minimizer of $E^{prior}(\boldsymbol{X}^\Omega_g\phi_g)$, $\widetilde{\phi}^\star_g \in A^{\boldsymbol{X}^\Omega}$.

In section 2.7 we introduced the total variation prior E^{prior}_{TV}. The symmetry group $\mathbb{G}^\Omega$ is the group $SO(2) \times \mathbb{T}^\Omega$ with the algebra $\mathcal{G}^\Omega = \mathfrak{so}(2) \times \mathfrak{t}$. The ideal $\mathcal{X}^\Omega$ is the algebra $\mathfrak{t}$. Thus E^{prior}_{TV} penalizes level-sets with non-zero curvature so that its minimizers $\phi^\star$ only contain linear level-sets $S^{lin}_{\boldsymbol{v}}$. However the orientation vector $\boldsymbol{v}$ is not constrained. In section 4 we will introduce the prior E^{prior}_{ST} which is based on the structure tensor [7, 8, 6] which has the same property as E^{prior}_{TV} in that it c penalizes level-sets with non-zero curvature.

From now on we shall consider priors $E^{prior}(\boldsymbol{X}^{\Omega}_{g}\phi)$ which are constructed from the basis elements $X^{1,\Omega}_{g^\Omega} \in \mathcal{X}^\Omega$ and $X^{2,\Omega}_{g^\Omega} \in \mathcal{X}^\Omega$ of the commutative algebra $\mathcal{X}^\Omega$. We will consider the prior $E^{prior}(\boldsymbol{X}^{\Omega}_{g}\phi)$ to be invariant under the pure spatial group $\mathbb{G}^\Omega$.

4. Linearized Priors

4.1. The Linear Structure Tensor

We shall now proceed to introduce a prior based on the considerations made in chapter 2.4. 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

$$V_e^\Omega = v_x(\boldsymbol{x})\,\partial_x + v_y(\boldsymbol{x})\,\partial_y \in \mathfrak{t} \quad \boldsymbol{v}(\boldsymbol{x})|_A = const \tag{4.1}$$

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 $\{\partial_x, \partial_y\}$

$$\left[V_e^\Omega, \partial_{x,y}\right] = 0 \tag{4.2}$$

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

$$S_X = \left\{\boldsymbol{x} |\, \boldsymbol{v}^T \cdot \nabla\phi(\boldsymbol{x}) = 0\right\} \tag{4.3}$$

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

$$J(\boldsymbol{v}) = \frac{1}{2}\int_A w(\|\boldsymbol{x}\|)\,\boldsymbol{v}^T \cdot \left(\nabla\phi(\boldsymbol{x})\,\nabla^T\phi(\boldsymbol{x})\right)\boldsymbol{v} = \frac{1}{2}\boldsymbol{v}^T S \boldsymbol{v} \tag{4.4}$$

$$S = \int_A w(\|\boldsymbol{x}\|)\left(\nabla\phi(\boldsymbol{x})\,\nabla^T\phi(\boldsymbol{x})\right) d^2x = \left\langle \nabla\phi\nabla^T\phi \right\rangle \tag{4.5}$$

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. (4.5).

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

$$S = V^T D V \quad D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \quad V = (\boldsymbol{V}_1, \boldsymbol{V}_2) \tag{4.6}$$

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_α be the structure tensor where the local coordinate frame A is rotated by the $SO(2)$ (see eq. (2.161))

$$S_\alpha = \int_A w\left(\|\boldsymbol{x}(\alpha)\|\right) \left(\nabla_{\boldsymbol{x}(\alpha)} \phi\left(\boldsymbol{x}(\alpha)\right) \nabla^T_{\boldsymbol{x}(\alpha)} \phi\left(\boldsymbol{x}(\alpha)\right)\right) d^2 x_\alpha \tag{4.7}$$

The $SO(2)$ only rotates the domain A and does not deform it otherwise, thus the integral measure d^2x_α is independent of α, $d^2x_\alpha = d^2x$. Since the weighting function w only depends on the norm $\|\boldsymbol{x}(\alpha)\|$ which is preserved by the $SO(2)$, it is also invariant. The only component which changes is the gradient $\nabla_{\boldsymbol{x}(\alpha)}$. Using eq. (2.168) and the product rule we can compute the derivative of S_α at $\alpha = 0$

$$\begin{aligned} \frac{d}{d\alpha} S_\alpha \bigg|_{\alpha=0} &= \int_A w\left(\|\boldsymbol{x}\|\right) \left(\mathbf{M}_\alpha \nabla\phi \nabla^T \phi + \nabla\phi\nabla^T\phi \mathbf{M}_\alpha^T\right) d^2x && (4.8) \\ &= \mathbf{M}_\alpha \cdot S - S \cdot \mathbf{M}_\alpha = [\mathbf{M}_\alpha, S] && (4.9) \end{aligned}$$

In eq. (4.9) we used $\mathbf{M}_\alpha^T = -\mathbf{M}_\alpha$. We can get some information on the magnitude of the rate of change $\frac{d}{d\alpha} S_\alpha \Big|_{\alpha=0}$ by multiplying the commutator in eq. (4.9) with the eigenvectors $\boldsymbol{v}_{1,2}$

$$\widetilde{\boldsymbol{v}}_{1,2} = [\mathbf{M}_\alpha, S]\, \boldsymbol{v}_{1,2} \tag{4.10}$$

It is easy to show that both projections $\widetilde{\boldsymbol{v}}_{1,2}$ in eq. (4.10) have the same norm

$$\|\widetilde{\boldsymbol{v}}_{1,2}\| = |\lambda_1 - \lambda_2| \tag{4.11}$$

With the help of eq. (4.11) 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 $\mathbb{R}^2$

4.2. Structure Tensor Based Prior

Since the vector field V_e^Ω in eq. (4.1) 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 $R_\alpha \in SO\,(2)$

$$S' = R_\alpha^T S R_\alpha \tag{4.12}$$

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

$$E_{ST}^{prior} = \int_\Omega \mathcal{E}_{ST}\,(S)\, d^2x \tag{4.13}$$

$$\mathcal{E}_{ST}\,(S) = \frac{\lambda}{2}\det\,(S) \tag{4.14}$$

We want to show that E_{ST}^{prior} is invariant under the $SO(2)$. We insert S_α from eq. (4.7) into the determinant in eq. (4.14) and evaluate the derivative of E_{ST}^{prior} with respect to α

$$\left.\frac{d}{d\alpha} E_{ST}^{prior}(S_\alpha)\right|_{\alpha=0} = \int_\Omega \mathrm{Tr}\left(\mathbf{P}^{ST} \cdot [\mathbf{M}_\alpha, S]\right) d^2x, \quad P_{ij}^{ST} = \frac{\delta \mathcal{E}_{ST}}{\delta S_{ij}} \tag{4.15}$$

The matrix $\mathbf{P}^{ST}$ is the dual variable 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. (4.15) can be further transformed

$$\mathrm{Tr}\left(\mathbf{P}^{ST} \cdot [\mathbf{M}_\alpha, S]\right) = 2 \cdot \mathrm{Tr}\left(\mathbf{P}^{ST} \cdot \mathbf{M}_\alpha \cdot S\right) \tag{4.16}$$

The matrix under the trace on the right hand side of eq. (4.16) 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 $SO(2)$

$$\frac{d}{d\alpha} E_{ST}^{prior}(S_\alpha)\Big|_{\alpha=0} = 2\int_\Omega \mathrm{Tr}\left(\mathbf{P}^{ST}\cdot \mathbf{M}_\alpha \cdot S\right) d^2x = 0 \tag{4.17}$$

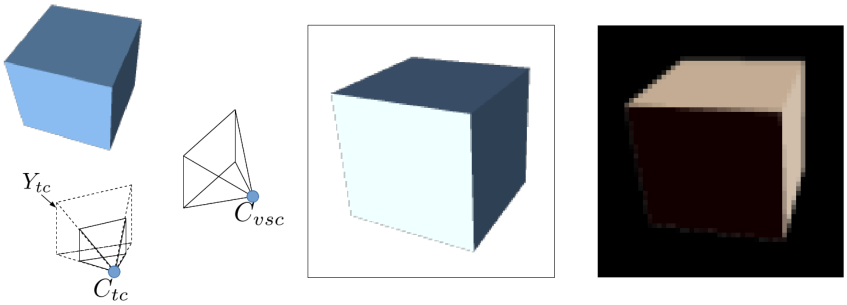
We note that the symmetry expressed by eq. (4.17) is a non-trivial symmetry, since only the trace as a whole vanishes.

4.3. Geometrical Optical Flow Model

In section 2.8 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. (2.210)) 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 $I_{\hat{\boldsymbol{d}}}(\boldsymbol{x})$ (defined in eq. (2.209)) are given an estimate $\hat{\boldsymbol{d}}$ of the optical flow field $\boldsymbol{d}$.

In the uni-modal case in section 2.8.1 the similarity measure $E_{Y,I}^{data}$ was defined as the pixel difference between the $Y(\boldsymbol{x})$ and $I_{\boldsymbol{d}}(\boldsymbol{x})$ (see eq. (2.212)). In section 2.8.2 it was explained that the measure $E_{Y,I}^{data}$ defined in eq. (2.212) 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. (2.218), [59]), Correlation Ratio (CR) (eq. (2.221), [86]) and Cross Correlation (CC) (eq. (2.223), [87]). The similarity measure based on MI, CR and CC pose constraints on the similarity $Y(\boldsymbol{x})$ and $I_{\hat{\boldsymbol{d}}}(\boldsymbol{x})$ in ascending order with CC posing the strongest constraint $Y(\boldsymbol{x}) = f \cdot I(\boldsymbol{x}) + \beta$ (see eq. (2.223)). However all three measures assume that the images $Y(\boldsymbol{x})$ and $I_{\boldsymbol{d}}(\boldsymbol{x})$ 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 2.9 we outlined the method of Hardie et. al. [44] 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. (2.226). However in their paper [44] 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 4.1.: Figure 4.1a shows the setup of a thermographic camera (TC), C_{tc}, and a visual spectrum camera (VSC), C_{vsc}, recording an object O. Figure 4.1b shows the image I which is recorded by C_{vsc} and figure 4.1c the lower resolution image y recorded by C_{tc}. The solid line cone of C_{tc} in figure 4.1a 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 $\boldsymbol{d}$ (the mapping between I and y) by the model in eq. (4.25)

Y). The result, a model capable of jointly estimating $\boldsymbol{d}(\boldsymbol{x})$ 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.

4.4. Multi-Modal Optical Flow with Differing Resolutions

We consider the camera setup in figure 4.1a 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. [44] (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 2.9 to include the unknown optical flow field $\boldsymbol{d}(\boldsymbol{x})$ 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 $I_{\boldsymbol{d}}$

$$I_{\boldsymbol{d}}\left(\boldsymbol{x}\right) = I\left(\boldsymbol{x} + \boldsymbol{d}(\boldsymbol{x})\right) \tag{4.18}$$

for a given optical flow field $\boldsymbol{d}(\boldsymbol{x})$. From eq. (2.233) we can compute the super resolved image $\hat{Y}_{\boldsymbol{d}}(\boldsymbol{x})$ which is a function of $\boldsymbol{d}$. In principle the energy in eq. (2.232) with the image $\hat{Y}_{\boldsymbol{d}}(\boldsymbol{x})$ then serves as a similarity measure between y and $I_{\boldsymbol{d}}$

$$E^{data}_{y,I}\left(\boldsymbol{d}\right) = E_{y,I_{\boldsymbol{d}}}\left(\hat{Y}_{\boldsymbol{d}}\right) \tag{4.19}$$

$\hat{Y}_{\boldsymbol{d}}$ is an implicit function which does not need to be estimated directly since by virtue of eq. (2.233) it is easily computed when needed. However since we are only interested in $\boldsymbol{d}$ we do not need to explicitly evaluate $\hat{Y}_{\boldsymbol{d}}$.

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

We will now briefly compute the exact form of $E_{y,I}^{data}(\boldsymbol{d})$ in eq. (4.19). A detailed description can be found in appendix C. In section 2.9 the model for computing the super- resolved image Y is given by the energy $E(Y)$ (see eq. (2.232))

$$E_{y,I}(Y) = \frac{1}{2}\int_{\Omega}\Big(y(\boldsymbol{x}) - W_{\sigma^{sc}}Y(\boldsymbol{x})\Big)^2 \cdot C_n^{-1} d^2x \tag{4.20}$$

$$+\frac{1}{2}\int_{\Omega}\Big(Y(\boldsymbol{x}) - \mu_{Y|I}(\boldsymbol{x})\Big)^2 \cdot C_{Y|I}^{-1} d^2x \tag{4.21}$$

with the conditional variance and mean

$$C_{Y|I} = C_{Y,Y}^{\sigma^{sc}} - C_{Y,I}^{\sigma^{sc},2} \cdot C_{I,I}^{\sigma^{sc},-1} \tag{4.22}$$

$$\mu_{Y|I}(\boldsymbol{x}) = \mu_Y + f \cdot (I(\boldsymbol{x}) - \mu_I), \quad f = C_{Y,I}^{\sigma^{sc}} \cdot C_{I,I}^{\sigma^{sc},-1} \tag{4.23}$$

The first integrand of $E_{y,I}(Y)$ in eq. (4.20) 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. (4.21) models the relationship between the *intensities* of Y and I. This done by transforming the spectrum of I via the factor f (eq. (4.23)) to match that of Y. Since this is done on a pixel-by-pixel basis, eq. (4.21) pins down the high resolution component of Y.

At this point we incorporate the optical flow $\boldsymbol{d}(\boldsymbol{x})$ which separates the cameras C_{tc} and C_{vsc} by assuming Y to be co-aligned the warped image $I_{\boldsymbol{d}}$ (eq. (4.18)). Thus we substitute I for $I_{\boldsymbol{d}}$ in the integrand in eq. (4.21)

$$E(Y,\boldsymbol{d}) = \frac{1}{2}\int_{\Omega}\Big(y(\boldsymbol{x}) - \langle Y\rangle_{\sigma^{sc}}(\boldsymbol{x})\Big)^2 \cdot C_n^{-1} d^2x \tag{4.24}$$

$$+\frac{1}{2}\int_{\Omega}\Big(Y(\boldsymbol{x}) - \mu_{Y|I_{\boldsymbol{d}}}(\boldsymbol{x})\Big)^2 \cdot C_{Y|I_{\boldsymbol{d}}}^{-1} d^2x \tag{4.25}$$

with expressions for $C_{Y|I_{\boldsymbol{d}}}$ and $\mu_{Y|I_{\boldsymbol{d}}}$ similar to those in eqs. 4.22 and 4.23.

While keeping $\boldsymbol{d}$ fixed we minimize $E(Y,\boldsymbol{d})$ with respect to Y and obtain a

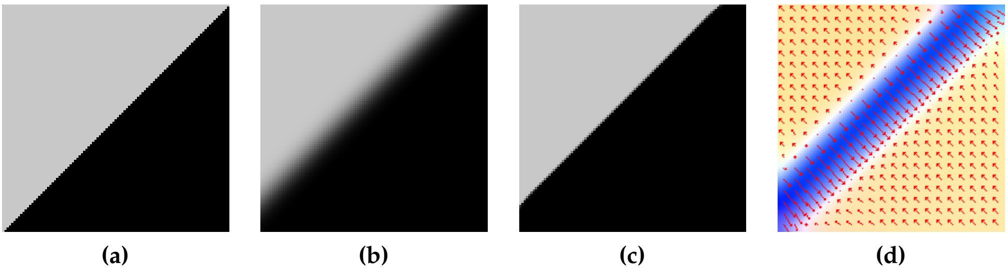


Figure 4.2.: Figure 4.2a shows a synthetic high resolution image I^{syn}. In figure 4.2b we show a low resolution image y^{syn}. y^{syn} is computed by convolution of I^{syn} with Gaussian $G_{\sigma^{sc}}$ with standard deviation $\sigma^{sc} = 5$ and translated by 10 pixels relative to I^{syn}. Figure 4.2d shows the flow $\boldsymbol{d}$ computed with the model in eq. (4.27), which incorporates knowledge of the scale difference between y^{syn} and I^{syn} and figure 4.2c show the warped image $I_{\boldsymbol{d}}$

simplified closed form solution for Y similar to eq. (2.234)

$$\hat{Y}_{\boldsymbol{d}} = \mu_{Y|I_{\boldsymbol{d}}} + C^{\sigma^{sc}}_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}} \cdot \left(C^{\sigma^{sc}}_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}} + C_n\right)^{-1} \left(y - \widetilde{\mu}_{Y|I_{\boldsymbol{d}}}\right) \tag{4.26}$$

We insert the closed form expression for $\hat{Y}$ from eq. (4.26) into $E\left(Y, \boldsymbol{d}\right)$ and obtain an energy measuring the similarity between y and $\langle I\rangle_{\sigma^{sc},\boldsymbol{d}} = W_{\sigma^{sc}} I_{\boldsymbol{d}}$

$$E^{data}_{y,I}\left(\boldsymbol{d}\right) = E_{y,I_{\boldsymbol{d}}}\left(\hat{Y}_{\boldsymbol{d}}\right)$$

$$= \frac{1}{2}\int_{\Omega}\left(y\left(\boldsymbol{x}\right) - \mu_Y - f \cdot \left(\langle I\rangle_{\sigma^{sc},\boldsymbol{d}}\left(\boldsymbol{x}\right) - \mu_I\right)\right)^2 \cdot U^{\sigma^{sc}} \tag{4.27}$$

$$f = C^{\sigma^{sc}}_{y,\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}} C^{\sigma^{sc},-1}_{\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}},\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}} \tag{4.28}$$

$$U^{\sigma^{sc}} = C^{\sigma^{sc}}_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}} \left(C^{\sigma^{sc}}_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}} + C_n\right)^{-2} \tag{4.29}$$

$E^{data}_{y,I}\left(\boldsymbol{d}^{\star}\right)$ in eq. (4.27) is minimal when the specific optical flow $\boldsymbol{d}^{\star}$ satisfies

$$y\left(\boldsymbol{x}\right) \approx f \cdot \langle I\rangle_{\sigma^{sc},\boldsymbol{d}^{\star}} + \left(\mu_Y - f \cdot \mu_I\right) \tag{4.30}$$

Comparing eq. (4.30) to the CC eq. (2.223) we can see that with $\beta = \left(\mu_Y - f \cdot \mu_I\right)$ eq. (4.30) is also a CC-type equation, however it correlates the low resolution images y and $\langle I\rangle_{\sigma^{sc},\boldsymbol{d}^{\star}}$.

It is easy to show that

$$C^{\sigma^{sc}}_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}} = C^{\sigma^{sc}}_{Y,Y}\left(1 - \eta_{CC}(Y|I_{\boldsymbol{d}})\right) \tag{4.31}$$

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. (4.27) 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 4.2d. Notice the blurred boundary $\boldsymbol{d}^{syn}$ around the linear feature in I^{syn} (figure 4.2a). This is the result of $E^{data}_{y,I}$ in eq. (4.27) measuring the difference between y^{syn} and the blurred image $\widetilde{I}^{syn}_{\boldsymbol{d}}=W_{\sigma^{sc}}I_{\boldsymbol{d}}$. In figure 4.2c we see $I_{\boldsymbol{d}^{syn}}$. The linear boundary has been warped by $\boldsymbol{d}^{syn}$ without being corrupted like in figure 2.5d.

4.5. Localization

The similarity measure $E^{data}_{y,I}$ in eq. (4.27) basically compares the image y with the transformed image $f\cdot\langle I\rangle_{\sigma^{sc},\boldsymbol{d}}$, where f is defined in eq. (4.28). 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

$$y\left(\boldsymbol{x}\right)\approx f\cdot\widetilde{I}_{\boldsymbol{d}^\star}\left(\boldsymbol{x}\right) \tag{4.32}$$

Eq. (4.32) 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. (4.32) is not valid. For instance in figure 4.9 in section 4.7.5 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. (4.32)

$$y\left(\boldsymbol{x}\right)|_{\mathcal{A}}\approx f\cdot\langle I\rangle_{\sigma^{sc},\boldsymbol{d}^\star}\left(\boldsymbol{x}\right)|_{\mathcal{A}} \tag{4.33}$$

might hold provided that the region $\mathcal{A}$ is occupied by the same object in both images which will be the case in section 4.7.5. Therefore we propose a local version of the variance and the mean in eq. (2.230)

$$C_{u,v}\left(\boldsymbol{x}_0\right) = \int_{\Omega} \omega\left(\boldsymbol{x}-\boldsymbol{x}_0\right)\left(u\left(\boldsymbol{x}\right)-\mu_u\left(\boldsymbol{x}_0\right)\right)\cdot\left(v\left(\boldsymbol{x}\right)-\mu_v\left(\boldsymbol{x}_0\right)\right)d^2x \tag{4.34}$$
$$\mu_u\left(\boldsymbol{x}_0\right) = \int_{\Omega} \omega\left(\boldsymbol{x}-\boldsymbol{x}_0\right)u\left(\boldsymbol{x}\right)d^2x$$

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

$$\omega\left(\boldsymbol{x}\right) = \begin{cases} \frac{1}{\left|U^a_{\boldsymbol{x}_0}\right|-1} & \boldsymbol{x} \in U^a_{\boldsymbol{x}_0}, \quad width\left(U^a_{\boldsymbol{x}_0}\right) = height\left(U^a_{\boldsymbol{x}_0}\right) = a \\ 0 & else \end{cases} \tag{4.35}$$

With this definition the conditional variance $C_{Y|I_{\boldsymbol{d}}}$ and the factor f (eq. (4.28)) become functions of the coordinates $\boldsymbol{x} \in \Omega$ and the parameter a

$$C^{\sigma^{sc},a}_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}\left(\boldsymbol{x}\right) = C^{\sigma^{sc},a}_{Y,Y}\left(\boldsymbol{x}\right) - C^{\sigma^{sc},a,2}_{y,\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}\left(\boldsymbol{x}\right)\cdot C^{\sigma^{sc},a,-1}_{\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}},\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}\left(\boldsymbol{x}\right) \tag{4.36}$$
$$f^{\sigma^{sc},a}\left(\boldsymbol{x}\right) = C^{\sigma^{sc},a}_{y,\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}\left(\boldsymbol{x}\right)\cdot C^{\sigma^{sc},a,-1}_{\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}},\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}\left(\boldsymbol{x}\right) \tag{4.37}$$

We substitute $C^{\sigma^{sc},a}_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}$ and $f^{\sigma^{sc},a}$ in eq. (4.27) with the local versions from eqs. 4.36 and 4.37 and obtain

$$E^{data,l}_{y,I_{\boldsymbol{d}}}\left(\sigma^{sc},a,\boldsymbol{d}\right) = \frac{1}{2}\int_{\Omega}\Big(y\left(\boldsymbol{x}\right) - \Big\langle f^{\sigma^{sc},a}I_{\boldsymbol{d}}\Big\rangle_{\sigma^{sc}}\left(\boldsymbol{x}\right)\Big)^2\cdot U^{\sigma^{sc},a}\left(\boldsymbol{x}\right)d^2x$$
$$U^{\sigma^{sc},a}\left(\boldsymbol{x}\right) = C^{\sigma^{sc},a}_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}(\boldsymbol{x})\left(C^{\sigma^{sc},a}_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}(\boldsymbol{x}) + C_n\right)^{-2} \tag{4.38}$$

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.

4.6. The Multigrid Newton algorithm

We combine now our global similarity measures in eq. (4.27) and together with the structure tensor prior E^{prior}_{ST} (eq. (4.14)) and the TV prior E^{prior}_{TV} (eq. (2.192)) to the following two models

$$E^g_{ST}\left(\boldsymbol{d}\right) = E^{data}_{Y,I}(\sigma^{scale},\boldsymbol{d}) + \lambda_{ST}\sum_{i=1}^{2}E^{prior}_{ST}\left(\nabla d_i\right) \tag{4.39}$$

$$E^g_{TV}\left(\boldsymbol{d}\right) = E^{data}_{Y,I}(\sigma^{scale},\boldsymbol{d}) + \lambda_{TV}\sum_{i=1}^{2}E^{prior}_{TV}\left(\nabla d_i\right) \tag{4.40}$$

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 $\|\boldsymbol{r}\| > \epsilon$ 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 $\boldsymbol{d}_{k+1} = \boldsymbol{d}_k + \alpha \boldsymbol{\delta}$, $k \rightarrow k+1$
  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. 4.39 to 4.40 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. 4.39 to 4.40). The linearized model is solved by a conjugate gradients algorithm with block Jacobi preconditioning.

When minimizing the structure tensor based model E^g_{ST} (eq. (4.39)) the following major numeric problem occurs: The problem arises in step 9 of the MOF algorithm. The second functional derivative $Q^g_{k,ST}$ of the energy model E^g_{ST} consists of one part comming from the likelihood and one part coming from the prior, $Q^g_k = Q^{data,g}_k + \lambda_{ST} Q^{prior}_{k,ST}$. The matrix $Q^{prior}_{k,ST}$ is the second derivative of the prior E^{prior}_{ST} with respect to the individual components of $\boldsymbol{d}$

$$Q^{prior,ii}_{k,ST} = \frac{\delta^2}{\delta d_i^2} E^{prior}_{ST} (\nabla d_i) \quad Q^{prior,ij} = 0 \tag{4.41}$$

Since $E^{prior}_{ST}(\nabla d_i)$ is purely quartic in d_i by definition (eq. (4.14)) the matrix $Q^{prior}_{k,ST}$ in eq. (4.41) 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^{data}_k is the second derivative of the data term $E^{data}_{Y,I}$ in eq. (4.39) and eq. (4.40). In regions where there is no motion the eigen values of Q^{data}_k 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

$$\widetilde{E}_{ST}^{prior}(\nabla d_i) = E_{ST}^{prior}(\nabla d_i) + \lambda_2 \left\|\nabla d_i\right\|^2 \tag{4.42}$$

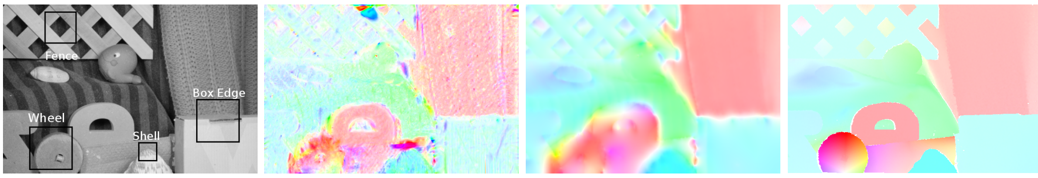
With the L_2 prior in 4.42 the linearized solution in step 9 becomes numerically stable. Thus we substitute E_{ST}^{prior} for $\widetilde{E}_{ST}^{prior}$ (eq. (4.42)) in eq. (4.39)

$$E_{ST}^{g}(\boldsymbol{d}) = E_{Y,I}^{data}(\sigma^{scale}, \boldsymbol{d}) + \lambda_{ST} \sum_{i=1}^{2} \widetilde{E}_{ST}^{prior}(\nabla d_i) \tag{4.43}$$

The challenge is to find the appropriate setting for λ_2 in eq. (4.42) 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. (4.42) does not over weigh the original anisotropic prior $\widetilde{E}_{ST}^{prior}$. In section 4.7.6 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. (4.42).

4.7. Results

In this section we want to evaluate and compare the two optical flow models in eqs. 4.39 and 4.40. The methodology for this section is as follows: In section 4.7.1 we want to study the effect of the size of the structure tensor S in eq. (4.39) on the estimation of the optical flow $\boldsymbol{d}$. We evaluate the models in eqs. 4.39 and 4.40 on a sequence of two images obtained from *one* camera from the Middleburry data-set [3] (figure 4.3). 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. 4.39 and 4.40. Since the data in figure 4.3 is collected from one camera we use the global similarity measure in eq. (4.27) thereby setting $\sigma = 0$. Thus eq. (4.27) effectively reduces to the brightness constancy constraint in eq. (2.212). We compare the results of the models 4.39 and 4.40 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 4.7.1 to include a synthesized multi-modal image set in section 4.7.4. 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 4.3a, filtering it with a Gaussian filter of standard deviation ${\sigma^{sc}}_{test}$, thus creating an artificial scale difference and inverting the result. We will show that $E_{y_{test}^{scale},I}^{data}(\sigma^{sc}, \mathbf{0})$ is minimal at $\sigma^{sc} = \sigma^{sc,\star}$.



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

Figure 4.3.: Rubberwhale Sequence: Figure 4.3a shows one frame of the sequence. figure 4.3b shows the estimated optical flow $\boldsymbol{d}^{\star}_{ST}$, figure 4.3c the flow $\boldsymbol{d}^{\star}_{TV}$ and figure 4.3d 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 4.7.5. As we discussed in section 4.5 the similarity measure $E^{data}_{Y,I}(\sigma^{sc}, \boldsymbol{d})$ 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 $E^{data,l}_{Y,I}(\sigma^{sc}, \boldsymbol{d})$ introduced in eq. (4.38) in the localized optical flow models

$$E_{ST}(\boldsymbol{d}) = E^{data,l}_{y,I}(\sigma^{sc}, a, \boldsymbol{d}) + \lambda_{ST} E^{prior}_{ST}(\nabla \boldsymbol{d}) \tag{4.44}$$

$$E_{TV}(\boldsymbol{d}) = E^{data,l}_{y,I}(\sigma^{sc}, a, \boldsymbol{d}) + \lambda_{TV} E^{prior}_{TV}(\nabla \boldsymbol{d}) \tag{4.45}$$

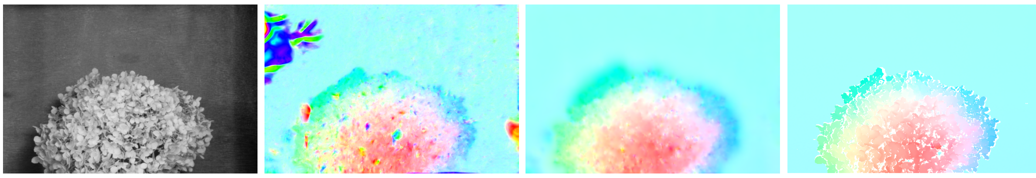
We will show that our strategy in section 4.7.4 for finding the true scale difference $\sigma^{sc,\star}$ between the images y and I also yields a strategy for finding the correct window size a for $E^{data,l}_{y,I}(\sigma^{sc}, a, \boldsymbol{d})$ such that $y = \alpha I + \beta$ *locally* holds.

4.7.1. 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)

$$EPE(\boldsymbol{x}) = \|\boldsymbol{d} - \boldsymbol{d}_{gt}\|(\boldsymbol{x}) \tag{4.46}$$

The EPE meassures how well the computed optical flow $\boldsymbol{d}$ fits the true optical flow $\boldsymbol{d}_{gt}$.



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

Figure 4.4.: Hydrangea Sequence: Figure 4.4a shows one frame of the sequence. figure 4.4b shows the estimated optical flow $\boldsymbol{d}^{\star}_{ST}$, figure 4.4c the flow $\boldsymbol{d}^{\star}_{TV}$ and figure 4.4d shows the provided ground truth

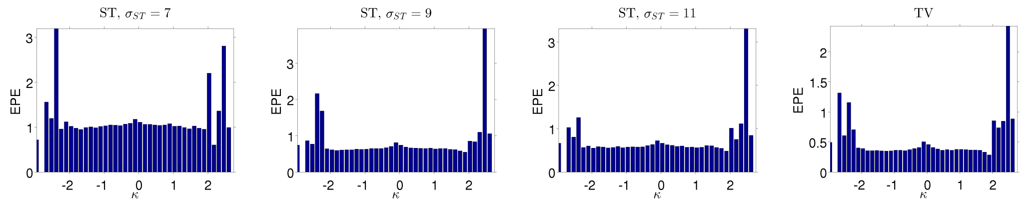
We will further need the definition of the mean curvature (eq. (2.205))

$$\kappa = \text{Div}\left(\frac{\nabla I}{|\nabla I|}\right) \tag{4.47}$$

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

4.7.2. 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. (4.39)) on the quality of the optical flow mapping $\boldsymbol{d}$ 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 $\boldsymbol{d}^{\star}_{ST}$ the optical flow obtained by E_{ST} and $\boldsymbol{d}^{\star}_{TV}$ the flow obtained by E_{TV}. In figure 4.3 the rubber whale sequence of the middleburry data set is shown, and in figure 4.3d the corresponding ground truth $\boldsymbol{d}_{gt}$. 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 4.3a 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 4.3b the computed flow-field $\boldsymbol{d}^{\star}_{ST}$ for the energy $E_{ST}(\boldsymbol{d})$ (eq. (4.39)) is shown for a filter size of $\sigma_{ST} = 11$, while in figure 4.3c the resulting flow for the TV model is shown. We can observe from the comparison between figures 4.3b and 4.3c that the TV model produces smoother results which are closer to the ground truth (figure 4.3d). In figure 4.5 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. (4.39) (varying σ_{ST} in figures 4.5a to 4.5c) and for the TV model in eq. (4.40) (figure 4.5d). In all plots in figure 4.5 the



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

Figure 4.5.: EPE to level-set curvature: Figures 4.5a to 4.5d show plots of the EPE (eq. (4.46)) against the curvature κ (eq. (4.47)) for the rubber whale sequence (figure 4.3). Figures 4.5a to 4.5c show the results for the structure tensor model E_{ST} and figure 4.5d 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
Wheel	7	2.36	Fence	7	0.46
	9	1.32		9	0.39
	11	1.15		11	0.35
	TV	1.38		TV	0.18
Shell	7	0.86	Box Edge	7	0.44
	9	0.62		9	0.34
	11	0.50		11	0.30
	TV	0.17		TV	0.09

Table 4.1.: EPE for different filter-sizes σ_{ST} for the model E^g_{ST} (eq. (4.39)) and for the TV model E^g_{TV} (eq. (4.40)). 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^g_{ST} 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*).

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 2.8) 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^{prior}_{ST} (eq. (4.39)) and E^{prior}_{TV} (eq. (4.40)) 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^{prior}_{TV} this is more evident since according to eqs. 2.205 and 2.206 in section 2.7 the curvature κ is the functional derivative of E^{prior}_{TV} and thus explicitly forced to vanish. Now comparing the plots in figures 4.5a to 4.5c we can see larger window sizes σ_{ST}

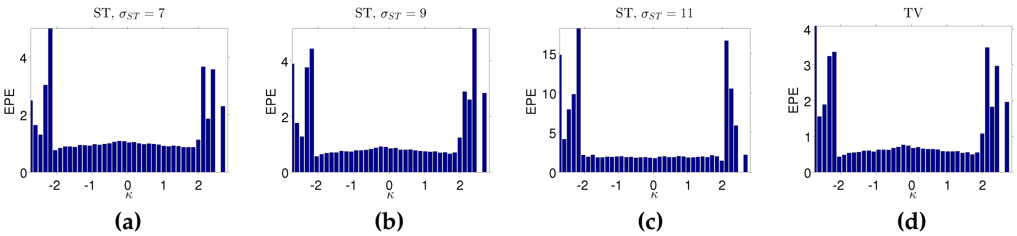


Figure 4.6.: EPE to level-set curvature: Figures 4.6a to 4.6d show plots of the EPE (eq. (4.46)) against the curvature κ (eq. (4.47)) for the hydrangea sequence (figure 4.4). Figures 4.6a to 4.6c show the results for the structure tensor model E_{ST} and figure 4.5d the result for the TV model E_{TV}. The curvature κ was split into 40 bins and the height of the bars is the average EPE per curvature bin.

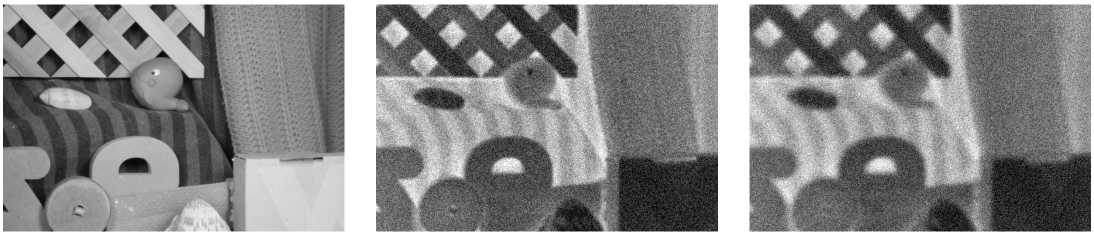
of the structure tensor in eq. (4.39) lead to small values for the EPE and thus more accurate optical flow estimates. In comparison with the TV model in eq. (4.40) (figure 4.5d) the EPE values structure tensor model E_{ST} with $\sigma_{ST} = 11$ in figure 4.5c are closest to those of the TV model in figure 4.5d with an average discrepancy of 0.3 between figure 4.5c and figure 4.5d. For this reason we will set $\sigma_{ST} = 11$ for the rest of this section.

Table 4.1 shows the median EPE values for different ROI's in figure 4.3a. 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.

4.7.3. Hydrangea Sequence

In figure 4.4a we show the hydrangea sequence of the middleburry dataset. In contrast to the rubber whale sequence in figure 4.3a figure 4.4a 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 $\boldsymbol{d}^{\star}_{ST}$ (figure 4.4b) and $\boldsymbol{d}^{\star}_{TV}$ (figure 4.4c) compared to the ground truth $\boldsymbol{d}_{gt}$ in figure 4.4d. In figure 4.6 we have again plotted the EPE against the curvature κ in a fashion similar to figure 4.5. Other than in figure 4.5 we can see that for increasing window sizes $\sigma_{ST} = 7$ (figure 4.6a) to $\sigma_{ST} = 11$ (figure 4.6c) for the model E_{ST} the EPE values increase, especially at higher curvatures $\|\kappa\| > 2$. Since the background in figure 4.4a is largely constant the condition defining the level-sets S_V (see eq. (3.67))

$$S_V = \{\boldsymbol{x} |\, V_e I\,(\boldsymbol{x}) = 0\} \quad V_e = \boldsymbol{v}\nabla \tag{4.48}$$



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

Figure 4.7.: Synthesized multi-modal data. This data simulates the camera arrangement in figure 2.6. The image I in figure 4.7a is from the rubberwhale data set in figure 4.3. Figures 4.7b and 4.7c show the image $y_{\sigma^{sc}{}_{test}}$ (eq. (4.50)) 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 4.4a 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 4.6c. Comparing the E_{ST} model to the TV based model E_{TV} we see that the EPE values in figure 4.6d are comparatively smaller then those for the E_{ST} model in figures 4.6a to 4.6c . The EPE values in figure 4.6d however are larger then the corresponding values in figure 4.5d for the rubber whale sequence (figure 4.3a).

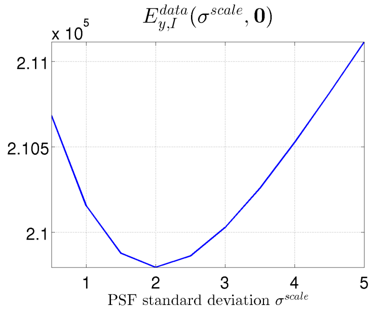
4.7.4. Estimation of the Scale Difference σ^{sc}

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 4.7): I is taken from the rubberwhale data set in figure 4.3. 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

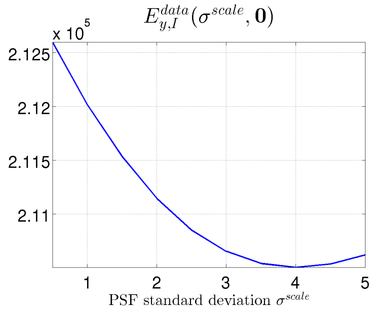
$$Y(\boldsymbol{x}) = -I(\boldsymbol{x}) + I_{min} + I_{max} \tag{4.49}$$

$$y^{test}_{\sigma}(\boldsymbol{x}) = \langle Y \rangle_{\sigma^{sc}{}_{test}}(\boldsymbol{x}) + n(\boldsymbol{x}), \quad n \sim \mathcal{N}(0, \chi_{std}) \tag{4.50}$$

I_{min} and I_{max} in eq. (4.49) are the minimum respectively maximum intensity of the image I and the standard deviation χ_{std} in eq. (4.50) was set to $\chi_{std} = 50$. In figure 4.7 the image I is shown along with the synthesized images y_2 and y_4 (y^{test}_{σ} from eq. (4.50) with $\sigma^{sc}{}_{test} = 2$ rsp. $\sigma^{sc}{}_{test} = 4$). The goal is now to show that $E^{data}_{y,I}(\sigma^{sc}, \mathbf{0})$ 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 4.8.: Figures 4.8a and 4.8b show plots the similarity measure $E^{data}_{y,I}(\sigma^{sc}, \boldsymbol{d})$ for the cases $y = y_2$, ${\sigma^{sc}}_{test} = 2$, and $y = y_4$, ${\sigma^{sc}}_{test} = 4$. We can observe that $E^{data}_{y,I}(\sigma^{sc}, \boldsymbol{d})$ is minimal with respect to σ^{sc} at the correct scales ${\sigma^{sc}}_{test}$

This might seem obvious since $E^{data}_{y,I}$ which was computed in eq. (4.27)

$$E^{data}_{y,I}\left(\sigma^{sc}, \boldsymbol{d}\right) = \frac{1}{2}\int_{\Omega}\Big(y\left(\boldsymbol{x}\right) - f \cdot \langle I \rangle_{\sigma^{sc}}\left(\boldsymbol{x}\right)\Big)^2 \cdot F \tag{4.51}$$

$$f = C^{\sigma^{sc}}_{y,\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}} C^{\sigma^{sc},-1}_{\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}},\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}} \tag{4.52}$$

$$F = C^{\sigma^{sc}}_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}\left(C^{\sigma^{sc}}_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}} + \lambda C_n\right)^{-2} \tag{4.53}$$

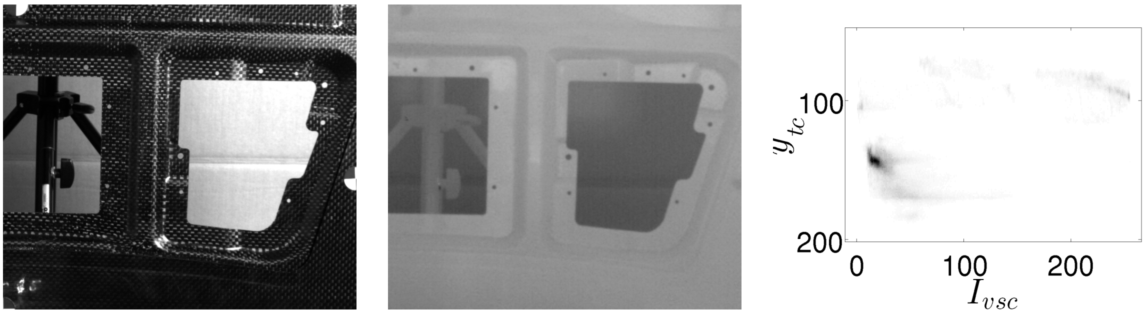
was derived from the basic assumption that y is the result of the convolution of the PSF $W_{\sigma^{sc}}$ with the high resolution image Y along with additive noise (see eq. (2.226))

$$y = \langle Y\rangle_{\sigma^{sc}} + n \quad n \sim \mathcal{N}\left(0|C_n\right) \tag{4.54}$$

which is similar to how we generated the test images y^{test}_{σ} in eq. (4.50). However the factor F in eq. (4.51) is highly non-linear in σ^{sc}, so we want to show that despite this non-linearity $E^{data}_{y,I}(\sigma^{sc}, \mathbf{0})$ has a global minimum at $\sigma^{sc,\star} = {\sigma^{sc}}_{test}$. In figure 4.8 we have plotted $E^{data}_{y,I}(\sigma^{sc}, \mathbf{0})$ over σ^{sc} for the cases ${\sigma^{sc}}_{test} = 2$ (figure 4.8a) and $\sigma^{sc} = 4$ (figure 4.8b). For both cases $E^{data}_{y,I}(\sigma^{sc}, \mathbf{0})$ is minimal at the correct scale ${\sigma^{sc}}_{test}$. From this we learn that $E^{data}_{y,I}(\sigma^{sc}, \boldsymbol{d})$ is sensitive to the scale difference σ^{sc} between y and I and we can use it to determine the true scale difference ${\sigma^{sc}}_{test}$.

4.7.5. Real Multimodal Optical Flow Data

We would now like to evaluate our optical flow model in eq. (4.39) on the data in figure 4.9. The image y_{tc} (figure 4.9b) was recorded with a thermographic

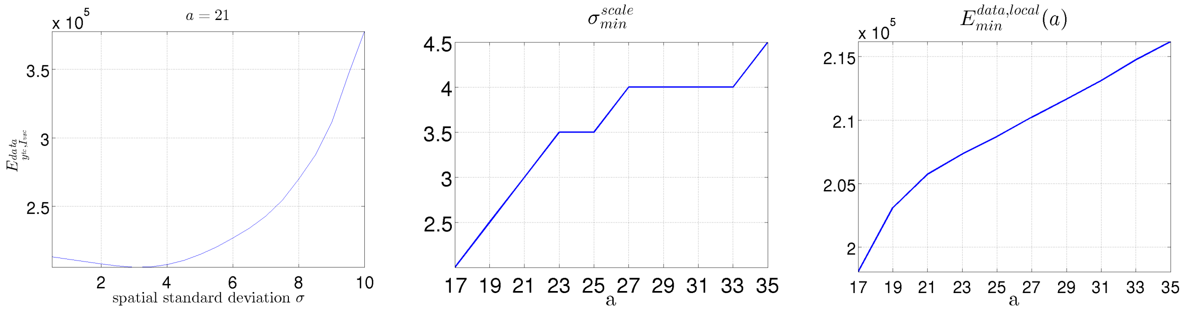


(a) **(b)** **(c)**

Figure 4.9.: figure 4.9a shows an image from a visual spectrum camera (VSC). The object recorded is a carbon-fiber reinforced polymer (CFRP). Figure 4.9b 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 4.9b correspond to warmer objects (the CFRP) and lower intensities to colder objects (the background). As in figure 4.1a 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 $\boldsymbol{d}(\boldsymbol{x})$ (see eq. (2.208)) which maps every pixel in the TC image to the corresponding pixel in the VSC image. Figure 4.9c shows the joint histogram of the VSC and TC image. It shows a complex mapping of the intensities of figure 4.9a to those of figure 4.9b 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, 50, 104, 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 [112, 100, 63] 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



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

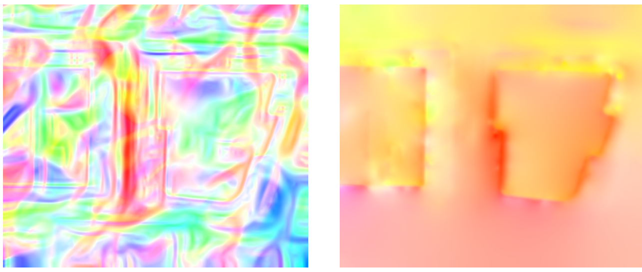
Figure 4.10.: Figure 4.10a: Plot $E^{data,l}_{y_{tc},I_{vsc}}(\sigma^{sc}, a, \mathbf{0})$ over the PSF scale difference σ^{sc} for the images y_{tc} and I_{vsc} in figure 4.9 for the window size $a = 25$. Figure 4.10b shows the minimum scale ${\sigma^{sc}}_{min}$ defined in eq. (4.59) as a function over the window size a and figure 4.10c the similarity measure $E^{data,l}_{min}$ (eq. (4.60)) 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 $E^{data,l}_{min}(21)$ 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. (4.39) 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 $E^{data}_{y,I}(\sigma^{sc}, \boldsymbol{d})$ in eq. (4.39) 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 4.9 lack this linearity relationship. This can be seen in the joint histogram in figure 4.9c where there is no unique correspondence between the intensities of y_{tc} to those of I_{vsc}. In section 4.5 we therefore introduced the local similarity measure $E^{data,l}_{y,I}(\sigma^{sc}, a, \boldsymbol{d})$ which constrains the similarity assessment of the images y and I to the local square regions $\mathcal{A}^a_{\boldsymbol{x}_0}$ of size a centered around $\boldsymbol{x}_0 \in \Omega$

$$E^{data,l}_{y_{tc},I_{vsc,\boldsymbol{d}}}(\sigma^{sc}, a, \boldsymbol{d}) = \frac{1}{2}\int_{\Omega}\left(y_{tc} - \left\langle f^{\sigma^{sc},a} I_{vsc,\boldsymbol{d}}\right\rangle_{\sigma^{sc}}\right)^2 \cdot U^{\sigma^{sc},a} d^2x \tag{4.55}$$

$$U^{\sigma^{sc},a}(\boldsymbol{x}) = C^{\sigma^{sc},a}_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}(\boldsymbol{x})\left(C^{\sigma^{sc},a}_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}(\boldsymbol{x}) + C_n\right)^{-2} \tag{4.56}$$

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 4.11.: Resulting optical flows of the local models E^{l}_{ST} ($\boldsymbol{d}^{\star}_{ST}$, eq. (4.44)) and E^{l}_{TV} ($\boldsymbol{d}^{\star}_{TV}$, eq. (4.45)). 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. (2.193).

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 4.7.4 we showed that the global similarity measure in eq. (4.27) is convex in the optical scale parameter σ^{sc}. For fixed widow sizes a the local similarity measure in eq. (4.56) 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 4.9 were used to set the zoom of the lenses such that the aperture angles of C_{tc} and C_{vsc} are approximately equal

$$\frac{w_{tc}}{f_{tc}} \approx \frac{w_{vsc}}{f_{vsc}} \tag{4.57}$$

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

$$\sigma^{sc,\star} = \frac{f_{vsc}}{f_{tc}} \approx \frac{w_{vsc}}{w_{tc}} \tag{4.58}$$

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

$$\sigma^{sc}{}_{min}(a) = \underset{\sigma}{\operatorname{argmin}} \left(E^{data,l}_{y_{tc},I_{vsc}} \left(\sigma^{sc}, a, \mathbf{0}\right)\right) \tag{4.59}$$

$$E^{data,l}_{min}(a) = E^{data,l}_{y_{tc},I_{vsc}} \left(\sigma^{sc}{}_{min}(a), a, \mathbf{0}\right) \tag{4.60}$$

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 4.10c shows the plot of $E^{data,l}_{min}$ (eq. (4.60)) 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. (4.56) 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 4.10b 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 $y_{tc}|_{\mathcal{A}^{a^\star}_{\boldsymbol{x}_0}}$ and $I_{vsc}|_{\mathcal{A}^{a^\star}_{\boldsymbol{x}_0}}$ are expected to be linearly dependent

$$y_{tc}|_{\mathcal{A}^{a^\star}_{\boldsymbol{x}_0}} \approx f \cdot \left\langle I_{\boldsymbol{d}^\star,vsc}\right\rangle_{\sigma^{sc}} |_{\mathcal{A}^{a^\star}_{\boldsymbol{x}_0}} \tag{4.61}$$

With the values $a^\star = 21$ and $\sigma^{sc,\star} = 3$ we compute the optical flow $\boldsymbol{d}$ for the models E^l_{ST} and E^l_{TV}

$$E^l_{ST}(\boldsymbol{d}) = E^{data,l}_{Y,I}(\sigma^{sc,\star}, a^\star, \boldsymbol{d}) + \lambda_{ST} \sum_{i=1}^{2} \widetilde{E}^{prior}_{ST}(\nabla d_i) \tag{4.62}$$

$$E^l_{TV}(\boldsymbol{d}) = E^{data,l}_{Y,I}(\sigma^{sc,\star}, a^\star, \boldsymbol{d}) + \lambda_{TV} \sum_{i=1}^{2} E^{prior}_{TV}(\nabla d_i) \tag{4.63}$$

In figure 4.11 we show the resulting optical flow $\boldsymbol{d}^\star_{ST}$ and $\boldsymbol{d}^\star_{TV}$. The model E_{TV} produces a piece-wise smooth optical flow $\boldsymbol{d}^\star_{TV}$ due to the piece-wise smoothing term of TV in eq. (2.193). 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. (4.42). The value $\lambda_2 = 10^{-6}$ was chosen purely for stabilizing the structure tensor models in eq. (4.39) and eq. (4.62) 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. (4.61) we deployed Pearson's χ^2 statistic [108].

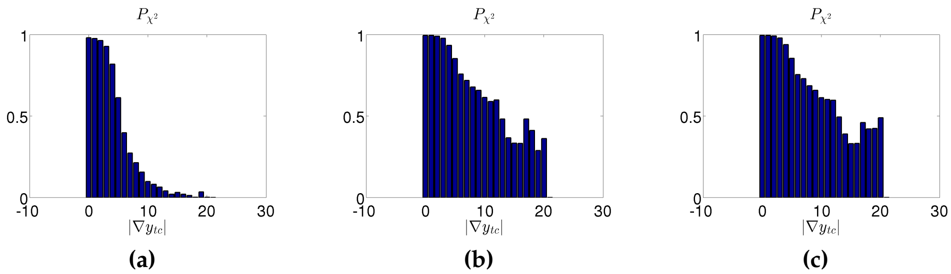


Figure 4.12.: Comparison of the p-values (eq. (4.66)) for the hypotheses (eq. (4.69)) $H_{\hat{\boldsymbol{d}}=\boldsymbol{0}}$ (figure 4.12a), $H_{\hat{\boldsymbol{d}}=\boldsymbol{d}^\star_{ST}}$ (figure 4.12b) and $H_{\hat{\boldsymbol{d}}=\boldsymbol{d}^\star_{TV}}$ (figure 4.12c). The p-values where computed for windows $\mathcal{A}_{\boldsymbol{x}_0}$ around each pixel $\boldsymbol{x}_0 \in \Omega$ 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 4.9 obey the linear relation in eq. (4.69) regardless of the optical flow $\hat{\boldsymbol{d}}$. For higher values of the gradient ∇y the hypothesis $H_{\hat{\boldsymbol{d}}=\boldsymbol{0}}$ in figure 4.12a fails as expected since the p-values tend to zero. The p-values at higher gradients for the hypotheses $H_{\hat{\boldsymbol{d}}=\boldsymbol{d}^\star_{ST}}$ (figure 4.12b) and $H_{\hat{\boldsymbol{d}}=\boldsymbol{d}^\star_{TV}}$ (figure 4.12c) are significantly higher then for $H_{\hat{\boldsymbol{d}}=\boldsymbol{0}}$ with $H_{\hat{\boldsymbol{d}}=\boldsymbol{d}^\star_{TV}}$ having the highest p-values meaning that the total variation model E^l_{TV} in eq. (4.45) best fulfills the linearity hypothesis in eq. (4.69).

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 $\hat{X}_i$, $1 \leq i \leq k$. . 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

$$V = \sum_{i=1}^{k} \frac{(X_i - E_i)^2}{E_i} \tag{4.64}$$

is a random variable that follows the χ^2 distribution with cumulative distribution function P

$$P(\bar{V} < v) = \int_0^v f(v')dv', \quad f(v) = \frac{v^{\frac{k}{2}-1} e^{-\frac{v}{2}}}{2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right)} \tag{4.65}$$

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

$$p(v) = 1 - P(\bar{V} < v) \tag{4.66}$$

$p(v)$ 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

$$\hat{V} = \sum_{i=1}^{k} \frac{\left(\hat{X}_i - E_i\right)^2}{E_i} \tag{4.67}$$

If the p-value of the *particular* realization $\hat{V}$ satisfies

$$p(\hat{V}) > \alpha \tag{4.68}$$

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 $\boldsymbol{d}^\star_{ST}$ and $\boldsymbol{d}^\star_{TV}$. The hypotheses to test are generated by the assumption that the local data term $E^{data,\tilde{l}}_{Y,I_{\boldsymbol{d}}}$ in eq. (4.27) is minimal for a particular optical flow $\hat{\boldsymbol{d}}$

$$H_{\hat{\boldsymbol{d}}}: \quad y\left(\boldsymbol{x}\right) \approx \mu_{\widetilde{Y}|\widetilde{I}_{\hat{\boldsymbol{d}}}}\left(\boldsymbol{x}\right) \quad \mu_{\widetilde{Y}|\widetilde{I}_{\hat{\boldsymbol{d}}}}\left(\boldsymbol{x}\right) = \left\langle f^{\sigma^{sc},\star,a^\star}_{\hat{\boldsymbol{d}}} \cdot I_{\hat{\boldsymbol{d}}} \right\rangle_{\sigma^{sc}}\left(\boldsymbol{x}\right) \tag{4.69}$$

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 $\mu_{\widetilde{Y}|\widetilde{I}_{\hat{\boldsymbol{d}}}}(\boldsymbol{x})$ are the expectations E_i of the X_i. We calculate the local observations $\hat{V}(\boldsymbol{x}_0)$

$$\hat{V}_{\hat{\boldsymbol{d}}}(\boldsymbol{x}_0) = \int_{\mathcal{A}_{\boldsymbol{x}_0}} \left(y(\boldsymbol{x}) - \mu_{\widetilde{Y}|\widetilde{I}_{\hat{\boldsymbol{d}}}}\left(\boldsymbol{x}\right)\right)^2 \cdot \mu^{-1}_{\widetilde{Y}|\widetilde{I}_{\hat{\boldsymbol{d}}}}\left(\boldsymbol{x}\right) d^2x \tag{4.70}$$

and compute the p-values $p(\hat{V}_{\hat{\boldsymbol{d}}}(\boldsymbol{x}_0))$ from eq. (4.65) and eq. (4.66) which we abbreviate by $p_{\hat{\boldsymbol{d}}}(\boldsymbol{x}_0)$. In figure 4.12 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 $\hat{\boldsymbol{d}} = \boldsymbol{0}$, $\hat{\boldsymbol{d}} = \boldsymbol{d}^\star_{TV}$ and $\hat{\boldsymbol{d}} = \boldsymbol{d}^\star_{ST}$. 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. (4.69) 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 $H_{\boldsymbol{d}^\star_{TV}}$ and $H_{\boldsymbol{d}^\star_{ST}}$ (eq. (4.69)) are valid for regions

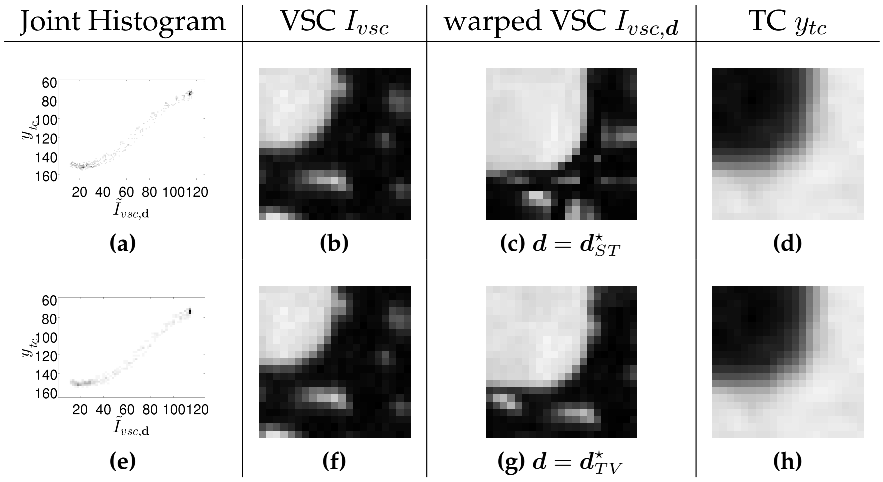
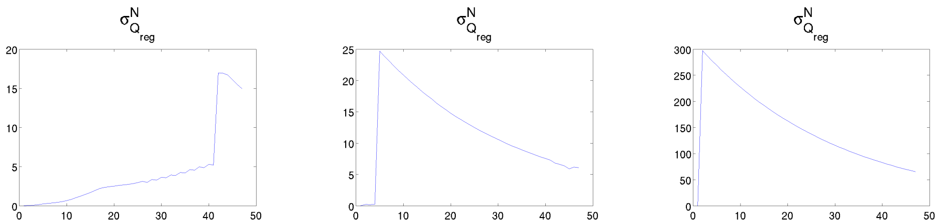


Figure 4.13.: Comparison of region of interests (ROI) of size $a^{\star} = 21$. Figures 4.13b and 4.13f show a ROI of I_{vsc} and 4.13d and 4.13h the corresponding ROI of the image y_{tc}. Figures 4.13c and 4.13g show figure 4.13b warped by the flows $\boldsymbol{d}^{\star}_{ST}$ and $\boldsymbol{d}^{\star}_{TV}$. 4.13a and 4.13e show the histograms between 4.13d and the filtered roi's $\widetilde{I}_{vsc,\boldsymbol{d}} = W_{\sigma^{sc,\star}} \star I_{vsc,\boldsymbol{d}}$

with gradients $\|\nabla y_{tc}\| < 10$ since the corresponding p-values in figure 4.12c and figure 4.12b are exceed α. For higher valued gradients $\|\nabla y_{tc}\| > 10$ the p-value $p_{\boldsymbol{d}^{\star}_{TV}}$ (figure 4.12c) is slightly higher than $p_{\boldsymbol{d}^{\star}_{ST}}$ (figure 4.12b) hence the model total variation model E^{l}_{TV} in eq. (4.45) better fulfills the hypothesis of linearity in eq. (4.69) then the structure tensor based model E^{l}_{ST} in eq. (4.44) and is thus better suited for multi-modal optical flow.

Figure 4.13 shows a comparison between an ROI in the image y_{tc} (figure 4.13h) and the corresponding ROI from the image I_{vsc} (figure 4.13b) warped by $\boldsymbol{d}^{\star}_{ST}$ (figure 4.13c) and by $\boldsymbol{d}^{\star}_{TV}$ (figure 4.13g). The gradients within this ROI are of the order $\|\nabla y_{tc}\| \approx 10$ and hence the corresponding p-values $p_{\boldsymbol{d}^{\star}_{TV}}$ and $p_{\boldsymbol{d}^{\star}_{ST}}$ are in the accepted range, $p_{\boldsymbol{d}^{\star}_{TV}} > \alpha$ and $p_{\boldsymbol{d}^{\star}_{ST}} > \alpha$. Hence the linearity hypothesis in eq. (4.69) holds for both $\hat{\boldsymbol{d}} = \boldsymbol{d}^{\star}_{TV}$ and $\hat{\boldsymbol{d}} = \boldsymbol{d}^{\star}_{ST}$ and the histograms in figure 4.13e (hypothesis $\hat{\boldsymbol{d}} = \boldsymbol{d}^{\star}_{TV}$) and figure 4.13a (hypothesis $\hat{\boldsymbol{d}} = \boldsymbol{d}^{\star}_{ST}$) visually reflect the linear dependence of the ROI in the image y_{tc} (figure 4.13h) and the corresponding ROI from the image I_{vsc} (figure 4.13b).



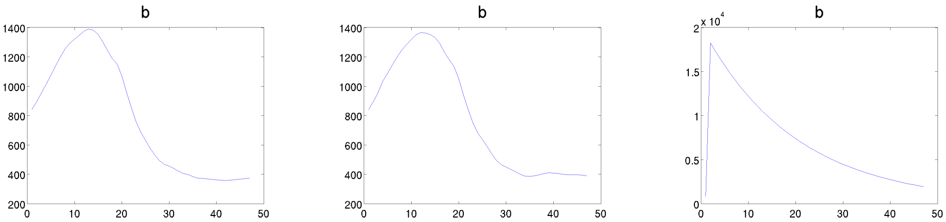
(a) $\lambda_2 = 10^{-3}$ **(b)** $\lambda_2 = 10^{-6}$ **(c)** $\lambda_2 = 10^{-9}$

Figure 4.14.: The largest eigenvalue σ_Q^k of Q^{reg} plotted over the iterations k for three values of λ_2 in eq. (4.42). Initially we have $\sigma_Q^k \approx 8\lambda_2$ which is the eigenvalue of the L_2 term in eq. (4.42). 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 4.14b and 4.14c) the jump occurs sooner indicating quicker an-isotropic behavior of E_{ST}^{prior}.

4.7.6. Eigenvalue analysis and the stabilization parameter λ_2

In chapter 4.3 we stated that the L_2 term in eq. (4.42) is needed to support the numerical stability of the model. We will back this statement now. Figures 4.14a, 4.14b and 4.14c show the largest eigenvalue of Q^i_{reg}, 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 4.3 but we found similar results for the hydrangea sequence in figure 4.4 and the CFRP data in figure 4.9. 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. (4.42) 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 4.14c) 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 4.14c is an indication that E_{ST}^{prior} is smoothing an-isotropically. On the other side Figures 4.15a, 4.15b and 4.15c show the residual vector $\boldsymbol{b}$ for different values of λ_2. Comparing the magnitude of the residual vector $\boldsymbol{b}$ in Figures 4.15a, 4.15b and 4.15c 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



(a) $\lambda_2 = 10^{-3}$ **(b)** $\lambda_2 = 10^{-6}$ **(c)** $\lambda_2 = 10^{-9}$

Figure 4.15.: The residual vector $\boldsymbol{b}$ plotted over the iterations k for three values of λ_2 in eq. (4.42). 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 4.14b the structure tensor only needs 4 iterations until its eigenvalues over-weigh the eigenvalues of the L_2 term.

4.7.7. 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 2.4, 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

$$y(\boldsymbol{x}) \approx f \, \langle I_{\boldsymbol{d}^\star} \rangle_{\sigma^{sc}} (\boldsymbol{x}) + \beta \tag{4.71}$$

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 2.7 to the optical flow models

$$E_{ST}^{g}(\boldsymbol{d}) = E_{Y,I}^{data}(\sigma^{scale}, \boldsymbol{d}) + \lambda_{ST} \sum_{i=1}^{2} E_{ST}^{prior}(\nabla d_i) \tag{4.72}$$

$$E_{TV}^{g}(\boldsymbol{d}) = E_{Y,I}^{data}(\sigma^{scale}, \boldsymbol{d}) + \lambda_{TV} \sum_{i=1}^{2} E_{TV}^{prior}(\nabla d_i) \tag{4.73}$$

To compare E_{ST}^{prior} with the total variation prior E_{TV}^{prior} we deployed both models in eqs. (4.72) and (4.73) on the rubber whale sequence (figure 4.3) and the hydrangea sequence (figure 4.4) from the middleburry 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. (4.72) 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. (4.73) 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. (4.71). The goal was to find out if the similarity measure $E_{y_{\sigma}^{test},I^{test}}^{data}(\sigma^{sc}, \mathbf{0})$ 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 $E_{y_{\sigma}^{test},I^{test}}^{data}(\sigma^{sc}, \mathbf{0})$ 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. (4.71)), the similarity measure $E_{y,I}^{data}(\sigma^{sc}, \mathbf{0})$ 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. (4.71), 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

$$H^{\star}_{\boldsymbol{d}} : y_{tc}|_{\mathcal{A}^{a^\star}_{\boldsymbol{x}_0}} \approx f \cdot \langle I_{\boldsymbol{d}^\star,vsc}\rangle_{\sigma^{sc}} |_{\mathcal{A}^{a^\star}_{\boldsymbol{x}_0}} \tag{4.74}$$

where $\mathcal{A}^{a^\star}_{\boldsymbol{x}_0}$ 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 $E^{data,l}_{y_{tc},I_{vsc}}(\sigma^{sc}, a, \boldsymbol{0})$ 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

$$E_{ST}\left(\boldsymbol{d}\right) = E^{data,l}_{y,I}(\sigma^{sc,\star}, a^\star, \boldsymbol{d}) + \lambda_{ST} E^{prior}_{ST}\left(\nabla \boldsymbol{d}\right) \tag{4.75}$$

$$E_{TV}\left(\boldsymbol{d}\right) = E^{data,l}_{y,I}(\sigma^{sc,\star}, a^\star, \boldsymbol{d}) + \lambda_{TV} E^{prior}_{TV}\left(\nabla \boldsymbol{d}\right) \tag{4.76}$$

To analyze the validity of the local linearity hypothesis $H^\star_{\boldsymbol{d}}$ in eq. (4.74) for the computed flows $\boldsymbol{d}^\star = \boldsymbol{d}^\star_{ST}$ (from eq. (4.75)) and $\boldsymbol{d}^\star = \boldsymbol{d}^\star_{TV}$ (eq. (4.76)) we deployed Pearson's χ^2 test. The p-value $p^\star_{\boldsymbol{d}}(\boldsymbol{x})$ in Pearson's χ^2 test is an indicator for the validity of the hypothesis $H^\star_{\boldsymbol{d}}$ such that $p^\star_{\boldsymbol{d}}(\boldsymbol{x}) = 1$ means that $H^\star_{\boldsymbol{d}}$ is definitely valid at the point $\boldsymbol{x}$ and $p^\star_{\boldsymbol{d}}(\boldsymbol{x}) = 0$ means that $H^\star_{\boldsymbol{d}}$ must be rejected.

The p-values $p_{\boldsymbol{d}^\star_{ST}}(\boldsymbol{x})$ and $p_{\boldsymbol{d}^\star_{TV}}(\boldsymbol{x})$ in regions with small gradients $\|y_{tc}\| < 10$ were sufficient, $p_{\boldsymbol{d}^\star_{ST}/\boldsymbol{d}^\star_{TV}} \approx 1$, to accept the linearity hypotheses $H_{\boldsymbol{d}^\star_{ST}}$ and $H_{\boldsymbol{d}^\star_{TV}}$. For higher gradients $\|y_{tc}\| > 10$ both p-values dropped off below the 50% quantile, however with $p_{\boldsymbol{d}^\star_{TV}}$ being slightly larger then $p_{\boldsymbol{d}^\star_{ST}}$. Thus although both models in eq. (4.76) and eq. (4.75) are not fully consistent with the linearity hypothesis in eq. (4.74) at the boundaries of the CFRP, the TV model in eq. (4.76) produced an optical flow $\boldsymbol{d}^\star_{TV}$ which is more consistent with the data y_{tc} and I_{vsc} at the boundaries then the structure tensor based model in eq. (4.75).

The good p-values in the regions $\|y_{tc}\| < 10$ do not give any information about how consistent the computed flows $\boldsymbol{d}^\star_{ST}$ (from eq. (4.75)) and $\boldsymbol{d}^\star_{TV}$ (eq. (4.76)) are themselves. This is why we visually compared $\boldsymbol{d}^\star_{ST}$ and $\boldsymbol{d}^\star_{TV}$ to each other: The flow $\boldsymbol{d}^\star_{ST}$ had a lot of artificial linear boundaries in regions where y_{tc} and I_{vsc} are largely constant. In contrast the flow $\boldsymbol{d}^\star_{TV}$ was smooth everywhere except at the physical boundary locations of the CFRP. This is due to the behavior of $E^{prior}_{ST}(\nabla \boldsymbol{d})$, namely that it only penalizes the curvature of the level-sets of $\boldsymbol{d}$. On the other hand the TV prior E^{prior}_{TV} not only penalizes the curvature of the level-sets but also enforces smoothness of the solution $\boldsymbol{d}^\star_{TV}$ in regions where the images y_{tc} and I_{vsc} are not discontinuous. Thus we conclude two things: First the TV prior E^{prior}_{TV} is superior to our structure tensor based prior E^{prior}_{ST} for multi-modal optical flow since the TV model in eq. (4.76) produces visually more consistent optical flows. Secondly the local linearity hypothesis in eq. (4.74) which emerges from our proposed similarity measure $E^{data,l}_{y_{tc},I_{vsc}}(\sigma^{sc}, a, \boldsymbol{d})$ 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 $E^{data,l}_{y_{tc},I_{vsc}}(\sigma^{sc}, a, \boldsymbol{d})$ as we had only encoded the basic assumption that TC produces Gaussian noise (see eq. (2.226) in section 2.9). Including a more realistic noise distribution into $E^{data,l}_{y_{tc},I_{vsc}}$ 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.

5. The Extended Least Action Algorithm

In section 2.2 we reviewed Fenchel's duality theorem [88] which is a central theorem in the analysis of convex functionals $E(\phi, \boldsymbol{A}\phi)$ of the form

$$E(\phi, \boldsymbol{A}\phi) = E^{data}(\phi) + E^{prior}(\boldsymbol{A}\phi), \quad \phi \in \Phi(\Omega) \tag{5.1}$$

where $\boldsymbol{A}$ is a linear operator on the Hilbert space $\Phi(\Omega)$. The main consequence of Fenchel's duality theorem is that there exists a pair $(\boldsymbol{p}^{\star,pr}, \phi^\star)$ where $\phi^\star$ is a minimizer of the functional E in eq. (5.1). The dual variable $\boldsymbol{p}^{\star,pr}$ is in the subdifferential of the prior $E^{prior}(\boldsymbol{A}\phi^\star)$ and fulfills the Kuhn-Tucker conditions (eq. (2.97))

$$-\boldsymbol{A}^\dagger \boldsymbol{p}^{\star,pr} \in \partial E^{data}(\phi^\star), \quad \boldsymbol{p}^{\star,pr} \in \partial E^{prior}(\boldsymbol{A}\phi^\star) \tag{5.2}$$

Section 3.2 was dedicated to the analysis of the energy E under the action of an n-dimensional finite Lie Group $\mathbb{G}^{\Omega\phi}$ which acts both on the GRF ϕ and on the coordinate frame Ω. We therefore considered E to depend on the transformed GRF ϕ_g and the sub algebra $\mathcal{X}^\Omega \subset \mathcal{G}^{\Omega\phi}$

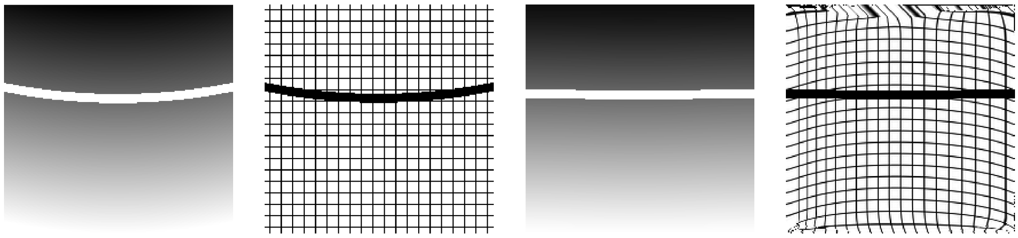
$$E\left(\phi_g, \boldsymbol{X}_g^\Omega \phi_g\right) = E^{data}(\phi_g) + E^{prior}(\boldsymbol{X}_g^\Omega \phi_g), \quad \phi_g = g \circ \phi, \quad \phi \in \Phi(\Omega) \tag{5.3}$$

The Kuhn-Tucker conditions for the functional in eq. (5.3) then translate to

$$\mathrm{Div}(\boldsymbol{p}_g^{\star,pr}) = -\boldsymbol{X}_g^{\dagger,\Omega} \boldsymbol{p}_g^{\star,pr} \in \partial E^{data}(\phi_g^\star), \quad \boldsymbol{p}_g^{\star,pr} \in \partial E^{prior}(\boldsymbol{X}_g^\Omega \phi_g^\star) \tag{5.4}$$

We had shown if $E\left(\phi_g, \boldsymbol{X}_g^\Omega \phi_g\right)$ is *invariant* under the action of $\mathbb{G}^{\Omega\phi}$ then it must be also be invariant under the subgroup of pure spatial transformations $\mathbb{G}^\Omega \subset \mathbb{G}^{\Omega\phi}$ with the algebra $\mathcal{G}^\Omega$ (eq. (3.52))

$$\left.\frac{d}{dt}\phi_{\theta^Q(t,g)}\right|_{t=0} = Q_g \phi_g = 0, \quad \forall g \in \mathbb{G}^{\Omega\phi} \tag{5.5}$$



(a) $\phi_0(\boldsymbol{x})$ **(b)** Ω_0 **(c)** $\widetilde{\phi}(\boldsymbol{x})$ **(d)** $\Omega^\epsilon = \theta^{B^\Omega} \circ \Omega_0$

Figure 5.1.: Figure 5.1a shows an image ϕ_0 with parabolic level-sets according to eq. (5.16). The white line indicates the level-sets $S_{\phi_0,c}$ with $39 < c < 43$. In figure 5.1b the coordinate frame Ω_0 is shown together with the level-sets $S_{\phi_0,c}$. Figure 5.1c shows the warped image $\widetilde{\phi}(\boldsymbol{x}) = \phi_0(\theta^{B^\Omega} \circ \boldsymbol{x})$ and figure 5.1d the transformed coordinate frame $\widetilde{\Omega} = \theta^{B^\Omega} \circ \Omega_0$. $\widetilde{\Omega}$ has been deformed by the algorithm in eq. (5.17) 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 $E_{TV}^{prior}(\nabla\phi)$.

From eq. (5.5) it followed that the identity

$$\boldsymbol{B}_g^{\Omega,m}\phi_g = 0, \qquad \boldsymbol{p}_g^{pr} \in \partial E^{prior}(\boldsymbol{X}_g^\Omega \phi_g) \tag{5.6}$$

holds *for any* $\phi \in \Phi(\Omega)$ and $g \in \mathbb{G}^{\Omega\phi}$, where $\boldsymbol{B}^{\Omega,m}$ and $\boldsymbol{p}_g^{pr}$ are related by

$$\boldsymbol{B}_g^{\Omega,m} = \boldsymbol{b}_g^{\Omega,m,T}\boldsymbol{X}_g^\Omega, \quad b_{g,j}^{\Omega,m} = p_{g,i}^{pr} C_{m,i}^j \tag{5.7}$$

In the case of smooth functionals $E\left(\phi_g, \boldsymbol{X}_g^\Omega \phi_g\right)$ we showed that eq. (5.6) reduces to the Noether identity (eq. (3.46))

$$\operatorname{div}(\boldsymbol{W}_m) - [\mathcal{E}]\,\omega_m^{\Omega,\mu}(\partial_\mu\phi)_g = \boldsymbol{B}_g^{\Omega,m}\phi_g = 0 \tag{5.8}$$

where $[\mathcal{E}]$ is the Euler-Lagrange differential in eq. (3.41). The coupling of the Euler-Lagrange differentials $[\mathcal{E}]$ and the product $\boldsymbol{B}_g^{\Omega,m}\phi_g$ in eq. (5.8) suggest that the symmetry identity in eq. (5.6) and thus $\mathbb{G}^\Omega$ impact the solutions space of the Kuhn-Tucker conditions in eq. (5.2). Indeed we will show that there exists an optimal deformation of the coordinate frame Ω which serves as an aid to solving the Kuhn-Tucker conditions in eq. (5.2). The deformation algorithm will exploit the non-trivial symmetries of the energy in eq. (5.3) in order to narrow down the space of possible solutions to the Kuhn-Tucker conditions in eq. (5.2). We will evaluate its effectiveness for both the TV prior E_{TV}^{prior} which we reviewed in section 2.7 and the structure tensor prior E_{ST}^{prior} which we introduced in section 4.

5.1. The Bending Flow

We will start the motivation of the new algorithm in the following way: In section 3.3 we argued that from the free Kuhn-Tucker conditions in eq. (3.62) and the invariance identity $\boldsymbol{B}_g^{\Omega,m}\phi_g = 0$ in eq. (5.6) it follows by proposition 4 that the optimal dual variable $\boldsymbol{p}^{\star,pr}$ which solves the free Kuhn-Tucker conditions must be constant in the local coordinates $\left\{\xi_X^{1,g}, \xi_X^{2,g}\right\}$ corresponding to the algebra $\mathcal{X}^\Omega$ (see eq. (3.63)). From lemma 9 it follows that the algebra $\mathcal{X}^\Omega$ is a commutative algebra and that the level-sets

$$S_B(\boldsymbol{x}) = \left\{\boldsymbol{x}_{g^\Omega} = g^\Omega \circ \boldsymbol{x} \middle| \boldsymbol{b}_g^{\star,\Omega,m,T}\boldsymbol{X}_g^\Omega \phi_g^\star(\boldsymbol{x}_{h^\Omega}) = 0, \quad g \in \mathbb{G}^{\Omega\phi}\right\} \tag{5.9}$$

are lines in the coordinates $\left\{\xi_X^{1,g}, \xi_X^{2,g}\right\}$.

Two examples of prior energies previously introduced, The structure tensor based prior energy in eq. (4.14) in section 4.3 and the total variation (TV) prior in eq. (2.192), 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. (2.204))) in the Cartesian coordinate frame Ω.

Our idea is to find a method for obtaining a transformation flow *of the coordinate frame* Ω, θ^{B^Ω} such that given *an arbitrary* image $\phi_0(\boldsymbol{x})$ the level-sets of the *warped* image $\phi_0(\theta^{B^\Omega} \circ \boldsymbol{x})$ satisfies the geometrical constraints imposed by E^{prior}, specifically such that the level-sets

$$S = \left\{\boldsymbol{x}' \middle| \sum_i \alpha_i X_e^{\Omega,i}\phi\left(\boldsymbol{x}'\right) = 0\right\}, \quad \boldsymbol{x}' = \theta^{B^\Omega} \circ \boldsymbol{x} \tag{5.10}$$

are linear in the image $\theta^{B^\Omega} \circ \Omega$ (see figure 5.1). In general it is not possible to derive θ^{B^Ω} analytically since an analytical expression for the level-sets of an arbitrary image ϕ_0 is not available. We propose therefore the following flow equation

Definition 22 (Bending Flow). *The Bending flow θ^{B^Ω} is defined by the differential equation*

$$\left.\frac{d}{dt}\theta^{B^\Omega}(t,g)\right|_{t=0} = \widetilde{\mathbf{B}}_g^\Omega = \sum_{m=1}^n \beta_m \widetilde{\boldsymbol{B}}_g^{\Omega,m} \tag{5.11}$$

The operators $\widetilde{\boldsymbol{B}}^{\Omega,m}$ are obtained from eq. (5.6) by normalizing the coefficients $\boldsymbol{b}^{\Omega,m}$

$$\widetilde{\boldsymbol{B}}_g^{\Omega,m} = \widetilde{\boldsymbol{b}}_g^{\Omega,m,T} \boldsymbol{X}_g^{\Omega}, \quad \widetilde{\boldsymbol{b}}_g^{\Omega,m} = \frac{\boldsymbol{b}_g^{\Omega,m}}{\max(1, \left\| \boldsymbol{b}_g^{\Omega,m} \right\|)} \tag{5.12}$$

The normalization in eq. (5.12) makes sure that the flow in eq. (5.11) is numerically stable in the cases $\|\boldsymbol{p}^{pr}\| \gg 1$ and $\|\boldsymbol{p}^{pr}\| \approx 0$. The operators $\widetilde{\boldsymbol{B}}^{\Omega,m}$ in eq. (5.12) are called the bending operators. They are purely spatial operators since they are spun by the spatial basis operators $\boldsymbol{X}^{\Omega} = \{X_g^{\Omega,1}, X_g^{\Omega,2}\}$. Hence the flow $\theta^{B^{\Omega}}$ only operates on Ω

$$\phi_{\theta^{B^{\Omega}}(t,g)}(\boldsymbol{x}) = \phi_{g^{\phi}}(\boldsymbol{x}_{\theta^{B^{\Omega}}(t,g^{\Omega})}), \quad \frac{d}{dt}\phi_{\theta^{B^{\Omega}}(t,g)}\bigg|_{t=0} = \widetilde{\mathbf{B}}_g^{\Omega}\phi_g = 0 \tag{5.13}$$

The flow $\theta^{B^{\Omega}}(t, g^{\Omega})$ is a map from the point $\boldsymbol{x}_{g^{\Omega}}$ to the point

$$\boldsymbol{x}_{\theta^{B^{\Omega}}(t,g^{\Omega})} = \int_0^t \mathbf{B}^{\Omega}_{\theta^{B^{\Omega}}(t',g^{\Omega})} \boldsymbol{x}_{\theta^{B^{\Omega}}(t',g^{\Omega})} dt' \tag{5.14}$$

The vector field $\mathbf{B}^{\Omega}$ in eq. (5.12) only depends on the prior energy E^{prior} and *not* on the data term E^{data}. Therefore we consider the total energy $E(\phi_g, \boldsymbol{X}_g^{\Omega}\phi_g)$ in eq. (5.3) only to consist of the TV prior E_{TV}^{prior}

$$E\left(\phi_g, \boldsymbol{X}_g^{\Omega}\phi_g\right) = E_{TV}^{prior}\left(\boldsymbol{X}_g^{\Omega}\phi_g\right) \tag{5.15}$$

What follows is a display how the flow equation in eq. (5.11) helps to (partially) minimize E_{TV}^{prior}. Consider the image $\phi_0(\boldsymbol{x})$ in figure 5.1a. It's level-sets have the shape of a parabola

$$S_{\phi_0,y_0} = \{\boldsymbol{x} |\, y = (x - x_0)^2 + y_0, \quad (x,y)^T = \boldsymbol{x}, y_0 \in \mathbb{R}\Big\} \tag{5.16}$$

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 5.1a 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 (alg. 4) that solves the integration problem in eq. (5.14) by splitting the time domain $\{0, T\}$ into N time steps t_n, $0 \leq n \leq N$ and iteratively computing a new coordinate frame Ω_{n+1} given an old estimate Ω_n

$$\Omega_{n+1}: \quad \boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \tau_{\Omega}\mathbf{B}_e^{\Omega}(\boldsymbol{x}_n), \quad \boldsymbol{x}_n \in \Omega_n \tag{5.17}$$

where Ω_n is the transformed coordinate frame at time step t_n and $e \in \mathbb{G}^{\Omega\phi}$ is the unit element. In the following all algorithms shall operate at $e \in \mathbb{G}^{\Omega\phi}$ since by the left-invariance of the symmetry algebra $\mathcal{G}^{\Omega\phi}$ (eq. (2.159)) the presented approach is also left-invariant.

The initial coordinate frame Ω_0 (figure 5.1b) is the Cartesian coordinate frame and the black line in figure 5.1b indicates the level-set S_{ϕ_0,y_0}. The simple procedure in eq. (5.17) deforms the coordinate frame Ω_0 under the influence of the prior E_{TV}^{prior} such that it assumes the form $\widetilde{\Omega} = \theta^{B^\Omega} \circ \Omega_0$ in figure 5.1d after $N = 2500$ iterations. In figure 5.1c the image $\phi_0(\theta^{B^\Omega} \circ \boldsymbol{x})$ is shown which is the result of transforming the original image $\phi_0(\boldsymbol{x})$ (figure 5.1a) to the frame $\widetilde{\Omega}$ in figure 5.1d. 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 5.1d 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 5.1a is not *cancelled* but merely inverted and deferred to the coordinate frame Ω.

5.1.1. Newtonian Minimization

One of the basic algorithms for solving the Kuhn-Tucker conditions in eq. (5.2) is the method of steepest descent (see [37] 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 $[\mathcal{E}]$

$$\dot{\phi}(t, \boldsymbol{x}_0) = -\left[\mathcal{E}\right]\left(\phi(t, \boldsymbol{x}_0)\right) \tag{5.18}$$

such that for $t \to t^\star$ where the limit $t^\star$ might by infinite, $\phi(t, \boldsymbol{x}_0)$ converges to $\phi^\star(\boldsymbol{x}_0)$

$$\dot{\phi}(t, \boldsymbol{x}_0)\Big|_{t \to t^\star} = 0 \Longrightarrow \phi(t, \boldsymbol{x}_0)\big|_{t \to t^\star} = \phi^\star(\boldsymbol{x}_0) \tag{5.19}$$

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 $[\mathcal{E}]$

$$\phi^{n+1}(\boldsymbol{x}_0) = \phi^n(\boldsymbol{x}_0) - \tau^\phi \left[\mathcal{E}\right]\left(\phi^n(\boldsymbol{x}_0)\right) \tag{5.20}$$

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. (5.11) with the concept of steepest descent for the image ϕ (in eq. (5.18)). The combination of the discretized eqs. (5.17) and (5.20) is straightforward

$$\phi^{n+1}(\boldsymbol{x}_n) = \phi^n(\boldsymbol{x}_n) - \tau^{\phi}\left[\mathcal{E}\right]\left(\phi^n(\boldsymbol{x}_n)\right) \tag{5.21}$$

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \tau_{\Omega}\mathbf{B}_e^{\Omega}\boldsymbol{x}_n \tag{5.22}$$

We will discuss how both the flow for the image ϕ in eq. (5.18) and the bending flow $\theta^{B^{\Omega}}$ in eq. (5.11) effect the evolution of the level-sets of ϕ for the generic energy functional $E(\phi_g, \boldsymbol{X}_g^{\Omega}\phi_g)$ in eq. (5.3).

5.1.2. The dynamics of the level-sets S

Noethers Theorem states that if the energy functional E is invariant under a Lie group $\mathbb{G}^{\Omega\phi}$ of dimension n, then there exists n identities relating the Euler-Lagrange differentials $[\mathcal{E}]$ and the divergences of r vector valued functions $\boldsymbol{W}_m$ via eq. (3.46). Thus the integral identity

$$\int_{\Omega}\left(\mathrm{div}\left(\boldsymbol{W}_m\right)\right)d^2x = \int_{\Omega}\left(X_e^{m,\Omega}\left(\phi\right)\left[\mathcal{E}\right]\right)d^2x \tag{5.23}$$

holds for any $\phi \in \Phi(\Omega)$.

Dynamics of the normal vector $\boldsymbol{n}_S$

Eq. (5.23) must hold for any integration domain Ω which means that the integrands themselves must be equal

$$\mathrm{div}\left(\boldsymbol{W}_m\right) = X_e^{m,\Omega}\left(\phi\right)\left[\mathcal{E}\right] \tag{5.24}$$

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 $\partial\mathcal{A}$

$$\int_{\mathcal{A}}\mathrm{div}\left(\boldsymbol{W}_m\right)d^2x = \int_{\partial\mathcal{A}}\boldsymbol{W}_m d\boldsymbol{n}_S \tag{5.25}$$

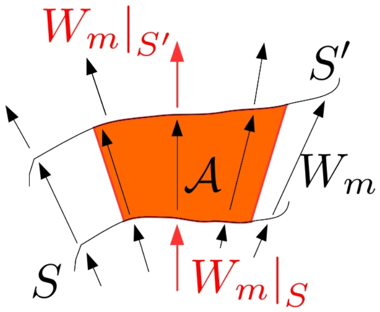


Figure 5.2.: This figure shows a transformation of the level-set S to S' along the vector $\boldsymbol{W}_m(\boldsymbol{x})$. The region $\mathcal{A} \subset \Omega$ is the region a section of S traverses as it is shifted along $\boldsymbol{W}_m$ to the end position S'. If the divergence of $\boldsymbol{W}_m$ vanishes, this means that the incoming flux of $\boldsymbol{W}_m$ equals the outgoing flux (both indicated by the red arrows), $\boldsymbol{W}_m|_S = \boldsymbol{W}_m|_{S'}$

where $\boldsymbol{n}$ is the normal vector on the surface $\partial\mathcal{A}$. Thus from eq. (5.24) we have

$$\int_{\partial\mathcal{A}} \boldsymbol{W}_m d\boldsymbol{n}_S = \int_{\mathcal{A}} \left(X_e^{m,\Omega}(\phi)[\mathcal{E}] \right) d^2x \tag{5.26}$$

In figure 5.2 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 $S \to S'$. 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

$$\int_S \boldsymbol{W}_m d\boldsymbol{n}_S - \int_{S'} \boldsymbol{W}_m d\boldsymbol{n}_{S'} = \int_{\mathcal{A}} X_e^{m,\Omega}(\phi)[\mathcal{E}]\, dx^2 \tag{5.27}$$

From Eq. (5.27) we see that the Euler-Lagrange differentials $[\mathcal{E}]$ 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. (5.27) vanishes. More precisely eq. (5.27) can interpreted as an equation of motion for the *normal vector* on S, $\boldsymbol{n}_S$.

5.2. The Curvature Operator K

In section 3.3 we had examined the *free* Kuhn-Tucker conditions in eq. (3.62)

$$\mathrm{Div}(\boldsymbol{p}_e^{\star,pr}) = 0, \quad \boldsymbol{p}_g^{\star,pr} \in \partial E^{prior}(\boldsymbol{X}_g^{\Omega}\phi_g^{\star}), \quad \forall \phi_g^{\star} \in A^{\boldsymbol{X}_g^{\Omega}} \tag{5.28}$$

hold for the minimizer set $A^{\boldsymbol{X}^{\Omega}}$ (eq. (3.61)) of the prior $E(\boldsymbol{X}_g^{\Omega}\phi_g)$ in eq. (3.2) which is invariant under the pure spatial group $\mathbb{G}^{\Omega}$ with the algebra $\mathcal{G}^{\Omega}$ (eq. (3.52)).

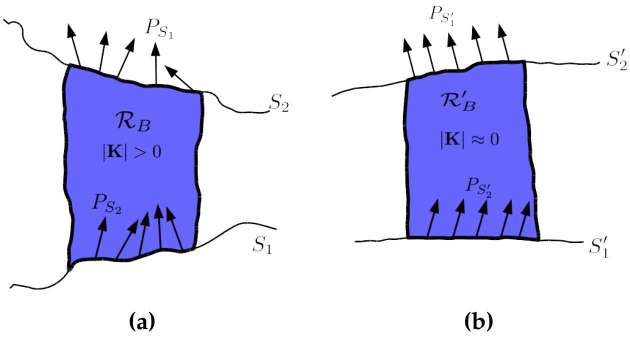


Figure 5.3.: Effect of the diffusion $\boldsymbol{x}' = \theta^{B^\Omega} \circ \boldsymbol{x}$ (eq. (5.11)) on the canonical momentum $\boldsymbol{p}^{pr}$. Figure 5.3a shows a schematic picture of a region $\mathcal{R}_B \subset \Omega$ between two level-sets S_1 and S_2. The canonical momentum (the vectors on the level-sets $S_{1,2}$) is denoted by $\boldsymbol{p}^{pr}_{S_{1,2}}$. $\boldsymbol{p}^{pr}$ changes its orientation when shifted along the level-sets S_1 and S_2 since the norm of the curvature operator $\mathbf{K}$ (eq. (5.31)) is non zero. In figure 5.3b the level-sets $S_{1,2}$ have been deformed according to $\boldsymbol{x}' = \theta^{B^\Omega} \circ \boldsymbol{x}$ such that the canonical momentum $\boldsymbol{p}^{pr}$ becomes invariant with respect to shifts along $S_{1,2}$. In this case the norm of the curvature operator $\mathbf{K}$ vanishes

Proposition 4 of that section states that the dual variable $\boldsymbol{p}_e^{\star,pr}$ in eq. (3.62) is constant with respect to the local variables $\{\xi_X^{1,g}, \xi_X^{2,g}\}$. By lemma 9 the minimizer $\phi_g^\star$ must have level-sets S_B (eq. (3.75)) which are *linear* with respect to the coordinates $\{\xi_X^{1,g}, \xi_X^{2,g}\}$ and thus have vanishing curvature. To prove proposition 4 we will introduce a new notion of the curvature of a level-set.

Lemma 10 (Curvature Operator)**.** *Let the energy functional $E(\phi_g, \boldsymbol{X}_g^\Omega \phi_g)$ in eq. (5.3) be invariant under the npure spatial group $\mathbb{G}^\Omega$ in eq. (5.5) such that by Noether's theorem the identities in eq. (5.8)*

$$\boldsymbol{B}_g^{\Omega,m} \phi_g = 0, \quad 1 \le m \le n \Rightarrow \mathbf{B}\phi_g = 0 \tag{5.29}$$

hold. Then the rate of change of the divergence Div($\boldsymbol{p}^{pr}$) in eq. (5.2) under the flow θ^{B^Ω} in eq. (5.11) satisfies

$$\int_\Omega \frac{d}{dt} \theta^{B^\Omega} \circ Div \boldsymbol{p}_g^{pr}(\boldsymbol{x})\Big|_{t=0} \phi_g(\boldsymbol{x})\, d^2x = -\int_\Omega \mathbf{K}\phi_g(\boldsymbol{x})\, d^2x \tag{5.30}$$

with the operator

$$\mathbf{K} = \left[\mathbf{B}, \mathbf{P}_g^{pr}\right] = k^l X_g^{\Omega,l}, \quad k^l = \mathbf{B}(p_g^{pr,l}) - \mathbf{P}_g^{pr}(b^l), \quad \mathbf{P}_g^{pr} = p_g^{pr,i} X_g^{\Omega,i} \tag{5.31}$$

called the curvature operator.

Proof. Let F_g be defined as

$$F_g = \int_\Omega \mathrm{Div}\boldsymbol{p}_g^{pr}(\boldsymbol{x})\phi_g(\boldsymbol{x})d^x \tag{5.32}$$

Then we can use the definition of the dual basis operators in eq. (3.19) to transform F_g

$$F_g = \int_\Omega \mathrm{Div}\boldsymbol{p}_g^{pr}(\boldsymbol{x})\phi_g(\boldsymbol{x})d^x = -\int_\Omega \mathbf{P}_g^{pr}\phi_g(\boldsymbol{x})d^2x \tag{5.33}$$

The operator $\mathbf{P}_g^{pr} = p_g^{pr,i}X_g^{\Omega,i}$ is a derivation on the Hilbert space $\Phi(\Omega)$. As such its transformation under the flow θ^{B^Ω} in eq. (5.11) is equivalent to the Lie derivative $\mathcal{L}_\mathbf{B}\mathbf{P}_g$

$$\frac{d}{dt}\theta^{B^\Omega} \circ \mathbf{P}_g\Big|_{t=0} = \mathcal{L}_\mathbf{B}\mathbf{P}_g = \mathbf{K} \tag{5.34}$$

Thus the rate of change of F_g under the flow θ^{B^Ω} is equivalent to

$$\frac{d}{dt}\theta^{B^\Omega} \circ F_g\Big|_{t=0} = -\int_\Omega \{\mathbf{K}\phi_g + \mathbf{P}_g(\mathbf{B}\phi_g)\}\, d^2x = -\int_\Omega \mathbf{K}\phi_g d^2x \tag{5.35}$$

where we used $\mathbf{B}\phi_g = 0$ in the second identity in eq. (5.35). On the other side

$$\frac{d}{dt}\theta^{B^\Omega} \circ F_g\Big|_{t=0} = \int_\Omega \frac{d}{dt}\,\theta^{B^\Omega} \circ \mathrm{Div}\boldsymbol{p}_g^{pr}\,(\boldsymbol{x})\Big|_{t=0}\phi_g\,(\boldsymbol{x})\,d^2x \tag{5.36}$$

Eqns. (5.35) and (5.36) are equivalent and thus eq. (5.30) is proven. □

We call the operator $\mathbf{K}$ 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. (5.11). Second and importantly $\mathbf{K}$ relates to a geometrical curvature: Since $\mathbf{B}$ is a level-set operator eq. (5.6), $\mathbf{K}$ describes the change the vector $\boldsymbol{p}_e^{pr}$ undergoes when being shifted along the level-sets of the GRF ϕ. In figure 5.3 we have schematically depicted the action of θ^{B^Ω} on the canonical momentum $\boldsymbol{p}^{pr}$. Figure 5.3a shows a region $\mathcal{R}_B \subset \Omega$ which is foliated by level-sets ranging from S_1 to S_2. The vector $\boldsymbol{p}^{pr}$ varies upon shifts along the level-sets $S_{1,2}$. Hence by eq. (5.31) the norm of the curvature operator $\mathbf{K}$ is non-zero. In figure 5.3b the region $\mathcal{R}_B$ has been deformed to $\mathcal{R}'_B$ via the diffusion in eq. (5.11). The level-sets $S_{1,2}$ have been deformed such that $\boldsymbol{p}_{S_{1,2}}^{pr}$ is constant along $S_{1,2}$. Thus the norm of the curvature operator $\mathbf{K}$ vanishes. We are now ready to prove proposition 4 in section 3.3

Proof of proposition 4. Let lemma 10 and the free Kuhn-Tucker conditions in eq. (5.28) hold for a pair $(\boldsymbol{p}_g^{\star,pr}, \phi_g^\star)$. From eq. (5.30) it follows that the curvature associated to $\boldsymbol{p}^{\star,pr}$ vanishes

$$\mathbf{K}^\star = \left[\mathbf{B}^{\star,\Omega}, \mathbf{P}_g^\star\right] = 0 \Leftrightarrow k^{\star,l} = 0 \tag{5.37}$$

For any $\psi \in \Phi(\Omega)$ the scalar product with the curvature coefficient functions vanish

$$\left\langle k^{\star,l}, \psi \right\rangle = \left\langle \mathbf{B}^{\star,\Omega}(p_g^{\star,pr,l}) - \boldsymbol{p}_g^{\star,pr,T} \boldsymbol{X}_g^\Omega(b^{\star,\Omega,l}), \psi \right\rangle = 0, \quad \boldsymbol{b}^{\star,\Omega} = \sum_{m=1}^{n} \beta_m \boldsymbol{b}^{\star\Omega,m}$$
$$\left\langle \boldsymbol{b}^{\star,\Omega,T} \boldsymbol{X}_g^\Omega(p_g^{\star,pr,l}), \psi \right\rangle = \left\langle \boldsymbol{p}_g^{\star,pr,T} \boldsymbol{X}_g^\Omega(b^{\star,\Omega,l}), \psi \right\rangle \tag{5.38}$$

We use Dirichlet conditions $\psi|_{\partial\Omega} = 0$ and the definition of the dual operator $\boldsymbol{X}_g^{\dagger,\Omega}$ in eq. (3.18) on the right hand side of eq. (5.38)

$$\left\langle \boldsymbol{p}_g^{\star,pr,T} \boldsymbol{X}_g^\Omega(b^{\star,\Omega,l}), \psi \right\rangle = \left\langle \boldsymbol{X}_g^{\dagger,\Omega} \boldsymbol{p}_g^{\star,pr} \cdot b^{\star,\Omega,l}, \psi \right\rangle \tag{5.39}$$

Since the Kuhn-Tucker conditions in eq. (5.28) hold the right hand side of eq. (5.38) vanishes and we have

$$\left\langle \boldsymbol{b}^{\star,\Omega,T} \boldsymbol{X}_g^\Omega(p_g^{\star,pr,l}), \psi \right\rangle = 0 \tag{5.40}$$

The function $\psi \in \Phi(\Omega)$ has arbitrary values in the interior of Ω. Hence the first factor in the scalar product in eq. (5.40) must vanish point-wise in Ω

$$\boldsymbol{b}^{\star,\Omega,T} \boldsymbol{X}_g^\Omega(p_g^{\star,pr,l})(\boldsymbol{x}) = 0, \quad \boldsymbol{x} \in \Omega \tag{5.41}$$

Eq. (5.41) is a level set equation for each component of $\boldsymbol{p}^{\star,pr}$. From lemma 9 it follows that the coefficient vector $\boldsymbol{b}^{\star,\Omega}$ must be constant function in the local coordinates $\{\xi_X^{1,g}, \xi_X^{2,g}\}$ corresponding to the commuting basis $\boldsymbol{X}^\Omega = \{X^{\Omega,1}, X^{\Omega,2}\}$. By the fixed relationship between $\boldsymbol{b}^\Omega$ and $\boldsymbol{p}_g^{pr}$ in eq. (5.6) we conclude that $\boldsymbol{p}_g^{\star,pr}$ is also constant with respect ot the coordinates $\{\xi_X^{1,g}, \xi_X^{2,g}\}$. $\square$

5.2.1. Image De-noising

In section 2.3 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

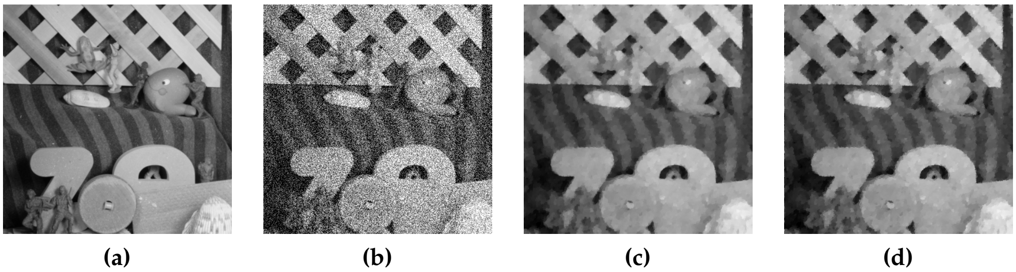


Figure 5.4.: Figure 5.4a shows a picture I^c of a person. I^c is taken to be free of noise. Figure 5.4b is a noise corrupted version of I^c in figure 5.4a, $\phi^d = I^c + n$ where n is iid Gaussian noise with a standard deviation $\sigma = 100$. Figure 5.4c shows the result of the ELAA (alg. 5) and figure 5.4d the result of the BNA (alg. 3)

drawn from a distribution p

$$\hat{I}^c_{ij} = I_O\left(\boldsymbol{x}_{i,j}\right) + n \quad n \sim p\left(I^c_{ij}|I_O\left(\boldsymbol{x}_{i,j}\right)\right) \tag{5.42}$$

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. (5.42). 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. (2.124) we reformulated the problem of inference of I_O as a minimization problem

$$I_O^\star = \text{argmin}_{I_O}\left(E_{I^c}\left(I_O\right)\right), \quad E_{I^c}\left(I_O\right) = E^{data}_{I^c}(I_O) + E^{prior}\left(\boldsymbol{X}^\Omega_e I_O\right) \tag{5.43}$$

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 p in eq. (5.42) is connected to the data term $E^{data}_{I^c}$ in eq. (5.43) via the exponential mapping in eq. (2.8).

The goal of this section is to evaluated the Extended Least Action Algorithm (ELAA) (alg. 5) for the inference problem in eq. (5.43). For this we will assume the situation where the noise distribution p in eq. (5.42) is Gaussian

$$E\left(\phi\right) = \frac{1}{2}\int_\Omega |I_O - I^c|^2 d^2x + E^{prior}\left(\boldsymbol{X}^\Omega_e I_O\right) \tag{5.44}$$

We have run both the Basic Newton (alg. 3) and our Extended Least Action algorithm (alg. 5) to minimize the energy in eq. (5.44) for the total variation prior and our new structure tensor based prior.

Algorithm 3 Basic Newton Algorithm (BNA)

Set $k = 0$
Set Initial guess ϕ^0
Compute residual $r^k = -[\mathcal{E}]\left(\phi^k\right)$
while $\|r\| > \delta$ and $k < N$ **do**
  Compute $\phi^{k+1}(\boldsymbol{x}) = \phi^k(\boldsymbol{x}) - \tau^\phi [\mathcal{E}]\left(\phi^k(\boldsymbol{x})\right)$
  Recompute $r^{k+1} = -[\mathcal{E}]\left(\phi^{k+1}\right)$
  Store $E^{k+1} = E(\phi^{k+1}, \nabla\phi^{k+1})$ in a vector $\{E^k\}$
  Set $k \to k+1$
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 $\phi^{k+1}(\boldsymbol{x}) = \phi^0\left(\boldsymbol{x}^{k+1}\right)$
  Store $E^{k+1} = E(\phi^{k+1}, \nabla\phi^{k+1})$ in a vector $\{E^k\}$
  Set $k \to k+1$
end while

Algorithm 5 Extended Least Action Algorithm (ELAA)

Set $k = 0$
Set Initial guess ϕ^0, $\boldsymbol{x}^0$
Compute residual $r^k = -[\mathcal{E}]\left(\phi^k\right)$
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 $\phi^{k+1}\left(\boldsymbol{x}^{k+1}\right) = \phi^k\left(\boldsymbol{x}^{k+1}\right) - \tau^\phi [\mathcal{E}]\left(\phi^k\left(\boldsymbol{x}^k\right)\right)$
  Recompute $r^{k+1} = -[\mathcal{E}]\left(\phi^{k+1}\right)$
  Store $E^{k+1} = E(\phi^{k+1}, \nabla\phi^{k+1})$ in a vector $\{E^k\}$
  Set $k \to k+1$
end while

Analysis Method

We minimized the image denoising model in eq. (5.44) for both the TV-Prior and the structure tensor prior using the Basic Newton Algorithm (BNA) in alg. (3), the

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 $\mathbf{E}$

Diffusion Algorithm (DA) in alg. (4) and the Extended Least Action Algorithm (ELAA) in alg. (5). The DA runs the bending flow in eq. (5.22) alone without the Eulerian flow in eq. (5.21) on an initial image ϕ_0 in order to evaluate the effect of eq. (5.22) on the denoising model in eq. (5.43). 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 $\langle \|\mathbf{K}\| \rangle$ and the standard deviation $\sigma_{\|\mathbf{K}\|}$ of the norm $\|\mathbf{K}\|$. The whole procedure was repeated on a total of ten images of the middleburry data set [3]. In figure 5.4 we show the Army image of the middleburry data set together with the results of the BNA and the ELAA, and in section D of the appendix we show the results of the other nine images.

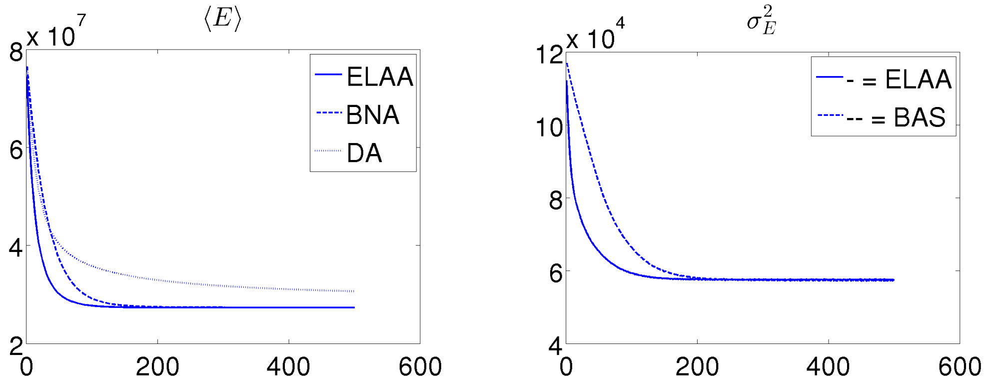
Total Variation based Image De-Noising

The TV based image de-noising model is defined by the energy

$$E\left(I_O, \nabla I_O\right) = \frac{1}{2}\int_\Omega |I_O - I^c|^2 d^2x + \int_\Omega \mathcal{E}^{prior}\left(\nabla I_O\left(\boldsymbol{x}\right)\right) d^2x \tag{5.45}$$

$$\mathcal{E}^{prior}\left(\nabla I_O\left(\boldsymbol{x}\right)\right) = \lambda\sqrt{\nabla^T I_O \cdot \nabla I_O}, \quad \boldsymbol{p}^{pr} = \lambda\frac{\nabla I_O}{\|\nabla I_O\|} \tag{5.46}$$

The prior $\mathcal{E}^{prior}$ in eq. (5.46) 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 $\boldsymbol{B}_{\mathbb{T}}$



(a) **(b)**

Figure 5.5.: Figure 5.5a shows the mean energy $\langle E^k \rangle$ and figure 5.5b the standard deviation σ_{E^k} per iteration k for the Army image in figure 5.4a. The the ELAA (solid line) converges about twice as fast as the BNA (dashed line) according to figure 5.5a. The standard deviation σ_{E^k} in figure 5.5b converges approximately three times faster for the ELAA then for the BNA indicating that the ELAA is robuster to noise at every iteration k

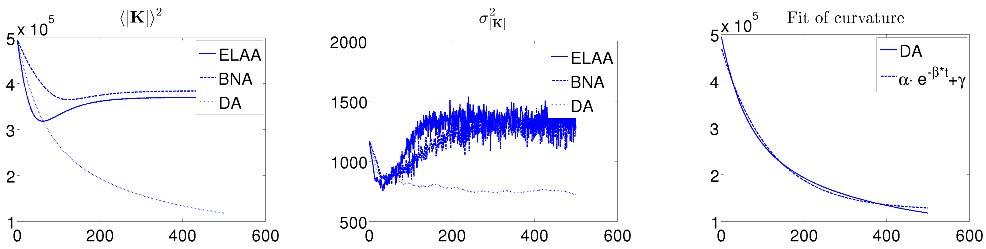
vanishes for the translation group $\mathbb{T}$ vanishes since

$$B^x_{\mathbb{T}} = P^\nu \left[\partial_\nu, \partial_x\right] = 0, \quad B^y_{\mathbb{T}} = P^\nu \left[\partial_\nu, \partial_y\right] = 0 \tag{5.47}$$

that is $\mathbb{T}$ is a trivial symmetry of $\mathcal{E}^{prior}$ and $E\left(I_O, \nabla I_O\right)$. The bending operator $\boldsymbol{B}^{\Omega,\alpha}$ associated with the rotation group $SO(2)$ does not vanish, but computes to

$$\boldsymbol{B}^{\Omega,\alpha} = b^{\Omega,\alpha}_{e,\mu}\partial_\mu, \quad \boldsymbol{b}^{\Omega,\alpha} = \frac{\nabla^\perp I_O}{\sqrt{\nabla^T I_O \cdot \nabla I_O}} = \boldsymbol{p}^{pr,\perp} \tag{5.48}$$

We have run the statistical analysis algorithm (SAA) in alg. (6) on the Army image in figure 5.4a and in figure 5.5 we have plotted the results. In figure 5.5a 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^{prior}_{TV} 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 5.5b) 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^k_O 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 5.6 we show the results of the SAA for the curvature $\|\mathbf{K}\|$. The curvature for the DA follows an exponential behavior. This is expected since by the definition of the mean curvature κ in eq. (2.205) and the definition of the



(a) (b) (c)

Figure 5.6.: Figure 5.6a shows the mean curvature $\langle \|\mathbf{K}\|^k \rangle$ and figure 5.6b the standard deviation $\sigma_{\|\mathbf{K}\|^k}$ per iteration k for the Army image in figure 5.4a. For the DA (dotted line), which only depends on the TV prior E^{prior}_{TV}, $\langle \|\mathbf{K}\| \rangle$ has an exponential decay. For the ELAA (solid line) $\langle \|\mathbf{K}\| \rangle$ drops faster then for the DA, until a point where the data term E^{data} prohibits further smoothing of the level-sets S. Then $\langle \|\mathbf{K}\| \rangle$ 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 $\sigma_{\|\mathbf{K}\|}$ is for both the ELAA and the BNA comparatively of equal order and small and two orders of magnitude smaller then $\langle \|\mathbf{K}\| \rangle$. 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 $\|\mathbf{K}\|$ particularly at later iterations $k > 100$. Figure 5.6c shows a fit of the exponential function in eq. (5.49) 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 $\|\mathbf{K}\|$

curvature operator $\mathbf{K}$ in eq. (5.31) we have

$$\int_\Omega \frac{d}{dt} \left(\theta^{B^\Omega}(t,e) \circ \kappa \right) \Big|_{t=0} I_O d^2x = \int_\Omega \mathbf{K} I_O d^2x \tag{5.49}$$

The left hand side of eq. (5.49) is the rate of change of the curvature κ with respect to the parameter t of the diffusion process in eq. (5.11) and the right hand side is linear in the curvature operator $\mathbf{K}$. Thus the curvature κ must have an exponential dependency on t

$$\theta^{B^\Omega}(t,e) \circ \kappa = \alpha \exp\left(-\beta t\right) + \gamma \tag{5.50}$$

We did a least squares fit of the parameters (α, β, γ) in eq. (5.50) to the curvature $\|\mathbf{K}\|$ of the DA shown in figure 5.6a. 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 $\|\mathbf{K}\|$. Hence we conclude two things: first the interpretation of the curvature operator $\mathbf{K}$ 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. (5.11) follows an exponential law.

Structure Tensor Prior

In this section we applied our structure tensor prior from eq. (4.14) in section 4.2 to the image de-noising problem

$$E\left(I_O, \nabla I_O\right) = \frac{1}{q}\int_\Omega \left|I_O - I^c\right|^q d^2x + \int_\Omega \mathcal{E}_{ST}^{prior}\left(\nabla I_O\left(\boldsymbol{x}\right)\right) d^2x \tag{5.51}$$

$$\mathcal{E}_{ST}^{prior}\left(\nabla I_O\left(\boldsymbol{x}\right)\right) = \frac{1}{2}\det\left(S\right) \tag{5.52}$$

In order to apply the ELAA in alg. 5 to the model in eq. (5.52) we need to compute the coefficient vector $\boldsymbol{b}$ of the bending operator $\boldsymbol{B}$. From section 4.2 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 $\boldsymbol{B}^{\Omega,\alpha}$ corresponding to the group $SO(2)$. We remember from eq. (4.17) that the structure tensor prior in eq. (5.52) transforms under the $SO(2)$ in the following way

$$\left.\frac{d}{d\theta} E_{ST}^{prior}(S_\theta)\right|_{\theta=0} = \int_\Omega \mathrm{Tr}\left(\mathbf{B}_{ST} \cdot S\right) d^2x = 0, \quad \mathbf{B}_{ST} = \mathbf{P}^{ST} \cdot \mathbf{M}_\theta \tag{5.53}$$

where $\mathbf{M}_\theta$ is the Pauli matrix (the generator of the algebra $\mathfrak{so}(2)$, eq. (2.168)). The matrix $\mathbf{B}_{ST}$ has a striking similarity to the bending operator $\boldsymbol{B}_m$ in eq. (5.12) 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

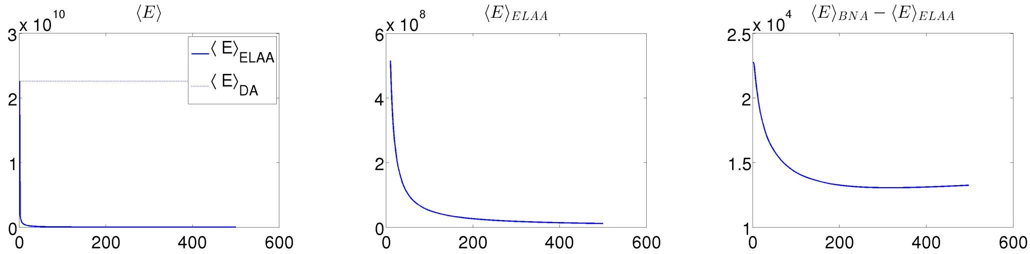
$$\mathrm{Tr}\left(\mathbf{B}_{ST} \cdot S\right) = B_{ST}^{\mu,\nu} S_{\mu,\nu} = 0 \tag{5.54}$$

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. (5.54) can be seen as the level-set equation for the structure tensor S, much like eq. (3.65) 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. (3.65) with an operator $\boldsymbol{B}_{ST}$, since the diffusion equation in eq. (5.11) necessitates a differential operator of the form in eq. (5.12).

$$\int_\Omega \mathrm{Tr}\left(\mathbf{B}_{ST} \cdot S\right) d^2x = \int_\Omega \boldsymbol{B}_{ST} I_O d^2x \tag{5.55}$$

If we insert the definition of the structure tensor from eq. (4.5), $S = \left\langle \nabla I_O \nabla^T I_O \right\rangle_{\sigma_{ST}}$ into the left hand side os eq. (5.55) 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



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

Figure 5.7.: Figure 5.7a 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 5.7b shows a close up of $\langle E \rangle_{ELAA}$ for $k \geq 10$ and figure 5.7c shows the difference between $\langle E \rangle_{BNA}$ and $\langle E \rangle_{ELAA}$. From figure 5.7b 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 5.7c only slightly smaller then $\langle E \rangle_{BNA}$. Thus the effect of the diffusion process in eq. (5.11) on the minimization of the energy E in eq. (5.52) is at most marginal

the gradient ∇I_O

$$\begin{aligned}\int_\Omega \operatorname{Tr}\left(\mathbf{B}_{ST} \cdot S\right) d^2x &= \int_\Omega \operatorname{Tr}\left(\mathbf{B}_{ST} \cdot \left\langle \nabla I_O \nabla^T I_O \right\rangle_{\sigma_{ST}}\right) d^2x \\ &= \int_\Omega \operatorname{Tr}\left(\langle \mathbf{B}_{ST} \rangle_{\sigma_{ST}} \cdot \nabla I_O \nabla^T I_O\right) d^2x = \int_\Omega \nabla^T I_O \left\langle \mathbf{B}_{ST} \right\rangle_{\sigma_{ST}} \nabla I_O d^2x \end{aligned} \tag{5.56}$$

from eq. (5.56) we can read of the form of $\boldsymbol{B}_{ST}$

$$\boldsymbol{B}_{ST} = \frac{1}{\|\boldsymbol{b}_{ST}\|} b^\mu_{ST} \partial_\mu, \quad \boldsymbol{b}_{ST} = \langle \mathbf{B}_{ST} \rangle_{\sigma_{ST}} \nabla I_O \tag{5.57}$$

The matrix $\langle \mathbf{B}_{ST} \rangle_{\sigma_{ST}}$ in eq. (5.57) 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. (4.5). Due to eq. (5.54) the operator $\boldsymbol{B}_{ST}$ is also a level-set operator in the sense of eq. (3.65).

In figure 5.7a 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 5.7b ($\langle E \rangle_{ELAA}$ for $k \geq 10$) we can see that $\langle E \rangle_{ELAA}$ drop down by 3 orders of magnitude. Figure 5.7c 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 5.7b. Thus although the diffusion process in eq. (5.11) has a positive impact on the ELAA, this impact is insignificant compared to the impact of eq. (5.11) on the TV-Denoising model (figure 5.5a). The explanation is that the structure tensor

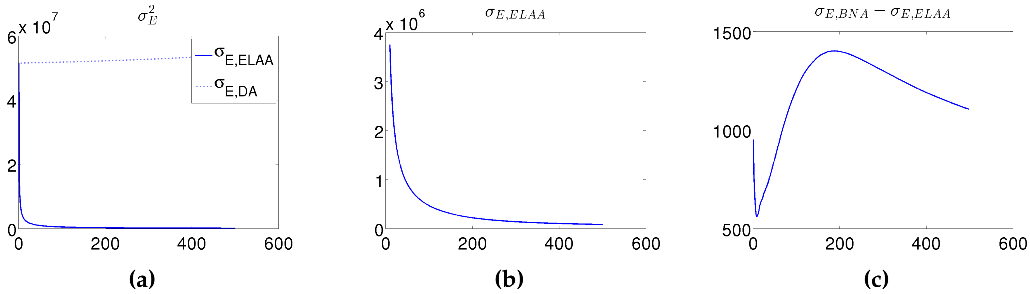


Figure 5.8.: Figure 5.8a shows the standard deviation σ_E as a function of the iteration k for the ELAA (solid line) and the DA (dotted line) for the structure tensor model. Figure 5.8b shows a close up of $\sigma_{E,ELAA}$ for $k \geq 10$ and figure 5.8c shows the difference between $\sigma_{E,BNA}$ and $\sigma_{E,ELAA}$. We essentially see the same behavior for the standard deviation σ_E as for the mean energy in figure 5.7: By figure 5.8b the standard deviation energy for the ELAA $\sigma_{E,ELAA}$ is 1 order of magnitude smaller that of the DA and by figure 5.8c only slightly smaller then $\sigma_{E,BNA}$. Hence the diffusion process eq. (5.11) has a marginal contribution to the statistical robustness of the minimizers of E in eq. (5.52)

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. (5.52) with the ELAA for various window sizes σ_{ST} in figure 5.9. In figure 5.9c the initial energy $\langle E^k \rangle_{ELAA}$ ($k = 0$) and in figure 5.9d 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. (5.31) the curvature operator $\mathbf{K}$ 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 5.9a we have plotted the *relative* energy

$$R_E(k, \sigma_{ST}) = \frac{\langle E^k \rangle}{\langle E^0 \rangle} \tag{5.58}$$

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 5.9b 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 5.9a. The curvature $\langle \|K\|^k \rangle$ is inverse proportional to the window size σ_{ST} at every iteration k. By eqs. (5.30) and (5.31) the decay rate of Euler-Lagrange differentials

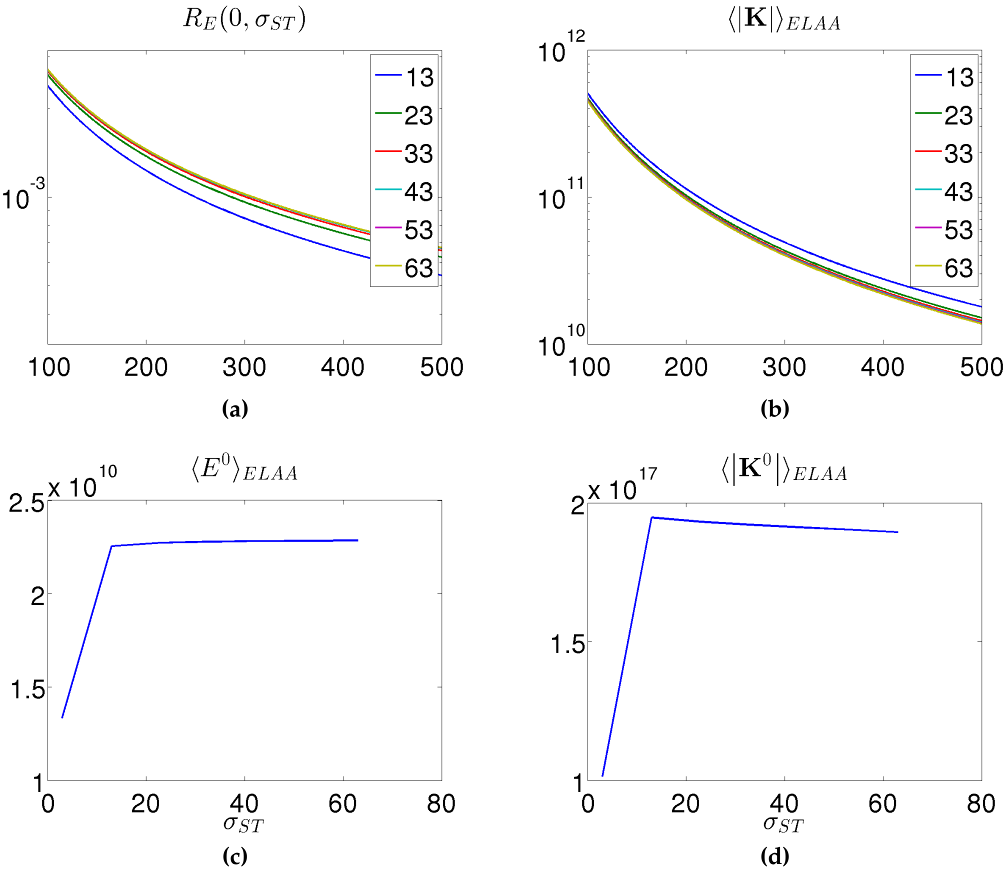


Figure 5.9.: Study of the dependency the mean energy $\langle E^k \rangle_{ELAA}$ and the mean curvature $\langle \|K\| \rangle$ on the window size σ_{ST} of the structure tensor prior E_{ST}^{prior}. Figure 5.9a shows the mean energy $\langle E^k \rangle_{ELAA}$ per iteration $k \geq 100$ for various σ_{ST} and figure 5.9b the mean curvature $\langle \|K\| \rangle$, also for various σ_{ST}. Figures 5.9c and 5.9d show the initial energy $\langle E^k \rangle_{ELAA}$ and the initial curvature $\langle \|K\| \rangle$ for $k = 0$. In figure 5.9a 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 5.9b 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 5.9c and 5.9d show that the initial energy and the initial curvature for $\sigma_{ST} = 3$ have half the values then for the larger window sizes $\sigma_{ST} = 13 \cdots 63$

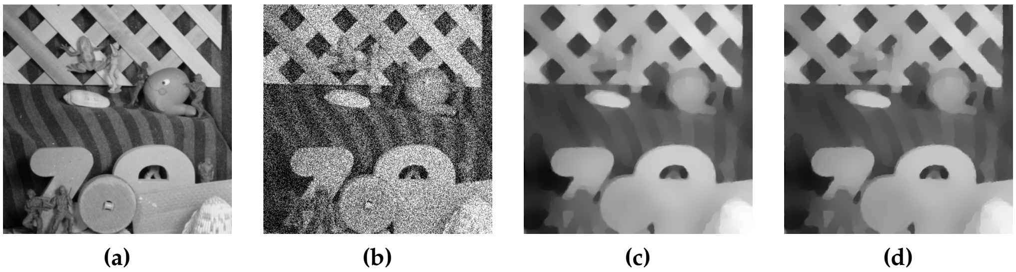


Figure 5.10.: Figure 5.10a shows a picture I^c of a scene. I^c is taken to be free of noise. Figure 5.4b is a noise corrupted version of I^c in figure 5.10a, $\phi^d = I^c + n$ where n is iid Gaussian noise with a standard deviation $\sigma = 100$. Figure 5.10c shows the result of the EPDS (alg. 8) and figure 5.10d the result of the CPPDS (alg. 7)

$[\mathcal{E}]$ 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 5.9c and figure 5.9d 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. (5.31) the energy $\langle E^k \rangle_{ELAA}$ for $\sigma_{ST} = 3$ has the worst convergence behavior due to the low initial curvature $\langle \|K\|^0 \rangle$.

5.3. The Extended Primal Dual Algorithm

In this section we shall review the primal dual splitting algorithm of Pock et. al. [80, 19] (alg. 1) for the total variation based denoising functional $E(I_O, \nabla I_O)$ in eq. (5.46). Alg. 7 is the adaptation of alg. 1 for the functional $E(I_O, \nabla I_O)$. The proximal operators $\text{prox}\left(\boldsymbol{p}|\sigma E_{TV}^{\star,prior}\right)$ and $\text{prox}\left(\widetilde{I}|\sigma E^{data}\right)$ in alg. 7 for the denoising model $E(I_O, \nabla I_O)$ in eq. (5.46) are according to [19]

$$\text{prox}\left(\widetilde{\boldsymbol{p}}|\sigma E_{TV}^{\star,prior}\right) = \frac{\widetilde{\boldsymbol{p}}}{\max(1, \|\widetilde{\boldsymbol{p}}\|)} \tag{5.59}$$

$$\text{prox}\left(\widetilde{I}|\tau E^{data}\right) = \frac{\widetilde{I} + \frac{\tau}{\lambda} I^c}{1 + \frac{\tau}{\lambda} I^c} \tag{5.60}$$

In [80] it was shown that the optimal values for the step parameters σ and τ and the interpolation parameter θ are

$$\tau = \sigma = \frac{1}{\sqrt{12}}, \quad \theta = 1 \tag{5.61}$$

Algorithm 7 Chambolle Pock Primal-Dual Splitting (CPPDS)

Normalize data $I^c \in \{-1, 1\}$
Choose parameters τ, $\sigma > 0$, $\theta \in (0,1)$ and initial guesses $\bar{I_O}^0$, p^0 and $I_O^0 = \bar{I_O}^0$
while $G(I_O^n, \boldsymbol{p}^n) > \epsilon$ or $n < M$ **do**
  Ascending step: $\boldsymbol{p}^{n+1} = \text{prox}(\boldsymbol{p}^n + \sigma \nabla I_O^n | \sigma E_{TV}^{\star, prior})$
  Descending step: $\bar{I_O}^{n+1} = \text{prox}(\bar{I_O}^n - \tau \nabla^\star p^{n+1} | \tau E^{data})$
  Interpolation step: $I_O^{n+1} = \bar{I_O}^{n+1} + \theta(\bar{I_O}^{n+1} - \bar{I_O}^n)$
  Store primal-dual gap $G(I_O^n, \boldsymbol{p}^n)$
end while

The normalizing step ensures that the smoothing parameter λ can be set to a value of order $O(1)$ independently of the data I^c. The ascending step moves the dual variable p in the direction which maximizes the dual function $E^\star(p, \nabla^\star p)$ and the descending step moves the primal variable I_O in the direction which minimizes the primal function $E(I_O, \nabla I_O)$. Since $\bar{I_O}^{n+1}$ 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.

since with these values and the normalization step in alg. 7 the algorithm is by theory guaranteed to converge.

In the Extended Primal-Dual splitting (EPDS) algorithm (alg. 8) we have extended alg. 7 to include the bending flow in eq. (5.11) where due to the invariance of the total variation prior E_{TV} under the group $SO(2)$ the coefficient vector $\boldsymbol{b}^{\Omega,\alpha}$ is perpendicular to the dual variable $\boldsymbol{p}$

$$\boldsymbol{B}^{\Omega,\alpha} = b_{e,\mu}^{\Omega,\alpha} \partial_\mu, \quad \boldsymbol{b}^{\Omega,\alpha} = \boldsymbol{p}^{\perp,n} \tag{5.62}$$

The step parameter $\tau^\Omega = 3$ was chosen since with this value all experiments were stable for various values of the smoothing parameter λ in eq. (5.46). We applied the SAA algorithm (alg. 6) to both the EPDS and the CPPDS algorithms. Figures 5.10c and 5.10d show the result of the EPDS (alg. 8) and the CPPDS (alg. 7) algorithms for $\lambda = 0.625$ and figure 5.11 shows the mean primal-dual gap $\langle G_{BV \times Y^\star} \rangle$ and its variance σ_G^2 plotted over the time steps $1 \leq k \leq 500$ for both algorithms. BV is the Hilbert space of bounded variation defined in eq. (2.190) and Y is the image of ∇ acting on BV. Figure 5.11 also shows plots of the square curvature $\langle \|K\| \rangle^2$ and its variance $\sigma_{\|K\|}^2$.

Both the EPDS and the CPPDS algorithms converge significantly faster then the Newtonian algorithms (BNA in alg. 3 and GNA in alg. 5) and for $k > 100$ the primal dual gaps of both algorithms converge to approximately the same value (the difference is of the order ~ 100). Figure 5.11 shows the result of the EPDS

Algorithm 8 Extended Primal-Dual Splitting (EPDS)

Normalize data $I^c \in \{-1, 1\}$

Choose parameters τ, τ^Ω, $\sigma > 0$, $\theta \in (0, 1)$ and initial guesses $\bar{I_O}^0$, p^0 and $I_O^0 = \bar{I_O}^0$

while $G(I_O^n, \boldsymbol{p}^n) > \epsilon$ or $n < M$ **do**

  Ascending step: $\boldsymbol{p}^{n+1} = \text{prox}(\boldsymbol{p}^n + \sigma \nabla I_O^n | \sigma E_{TV}^{\star,prior})$

  Warping step: $I_O^{n,warp}(\boldsymbol{x}) = I_O^n(\boldsymbol{x}^{n+1}), \quad \boldsymbol{x}^{n+1} = \boldsymbol{x}^n + \tau^\Omega \boldsymbol{p}^{\perp,n+1}$

  Descending step: $\bar{I_O}^{n+1} = \text{prox}(I_O^{n,warp} - \tau \nabla^\star \boldsymbol{p}^{n+1} | \tau E^{data})$

  Interpolation step: $I_O^{n+1} = \bar{I_O}^{n+1} + \theta(\bar{I_O}^{n+1} - \bar{I_O}^n)$

  Store primal-dual gap $G(I_O^n, \boldsymbol{p}^n)$

end while

As in alg. 1 the ascending step moves the dual variable p in the direction which maximizes the dual function $E^\star(p, \nabla^\star p)$ and the descending step moves the primal variable I_O in the direction which minimizes the primal function $E(I_O, \nabla I_O)$. Additionally the warping step warps the primal variable I_O in the direction perpendicular to ∇I_O, thereby minimizing the curvature of its level-sets.

and CPPDS for different values of the smoothing parameter λ. We can see that in the initial steps, at $k = 6$, the influence of the bending flow in eq. (5.11) is proportional to the smoothing parameter λ. The higher λ is, the faster EPDS initially converges compared to CPPDS. The EPDS also appears to be more stable regarding the influence of noise in the data I^c, since both variances σ_G^2 and $\sigma_{\|K\|}^2$ are favorable for the EPDS. We explain this observation by the fact that the bending operator $\mathbf{B}^\Omega$ in the flow in eq. (5.11) does not depend on the data term in eq. (5.46) but only on the current state of the dual variable $\boldsymbol{p}^n$. The warping step in the EPDS algorithm (alg. 8) smooths the level-sets of I_O such that the descent and ascent steps are less affected by the noise in the data I^c. This is supported by the plots of the mean curvature $\langle \|K\| \rangle$ in figure 5.11 which shows that the curvature $\|\mathbf{K}\|$ initially sinks for $k \leq 15$ and then slightly rises to a stable value for $k > 100$.

Figure 5.12 shows the evolution of a region of interest (ROI) in I_0^k for the EPDS and the CPPDS algorithm. In the region $k > 100$ the ROI $I_{O,EPDS}^k$ is oversmoothed at corner-like features compared to the ROI $I_{O,CPPDS}^k$. Indeed in figure 5.11 the curvature $\|\mathbf{K}\|$ of the EPDS algorithm is slightly lower then that of the CPPDS. This observation is due again due to the independence of the bending flow in eq. (5.11) and the data I^c which is this case is not a desirable feature, since it leads to a higher value primal-dual gap $G_{BV \times Y^\star}$ for the EPDS. To mitigate this one could extend the EPDS algorithm with a variable warping step size τ_k^Ω which gets smaller with increasing iteration k.

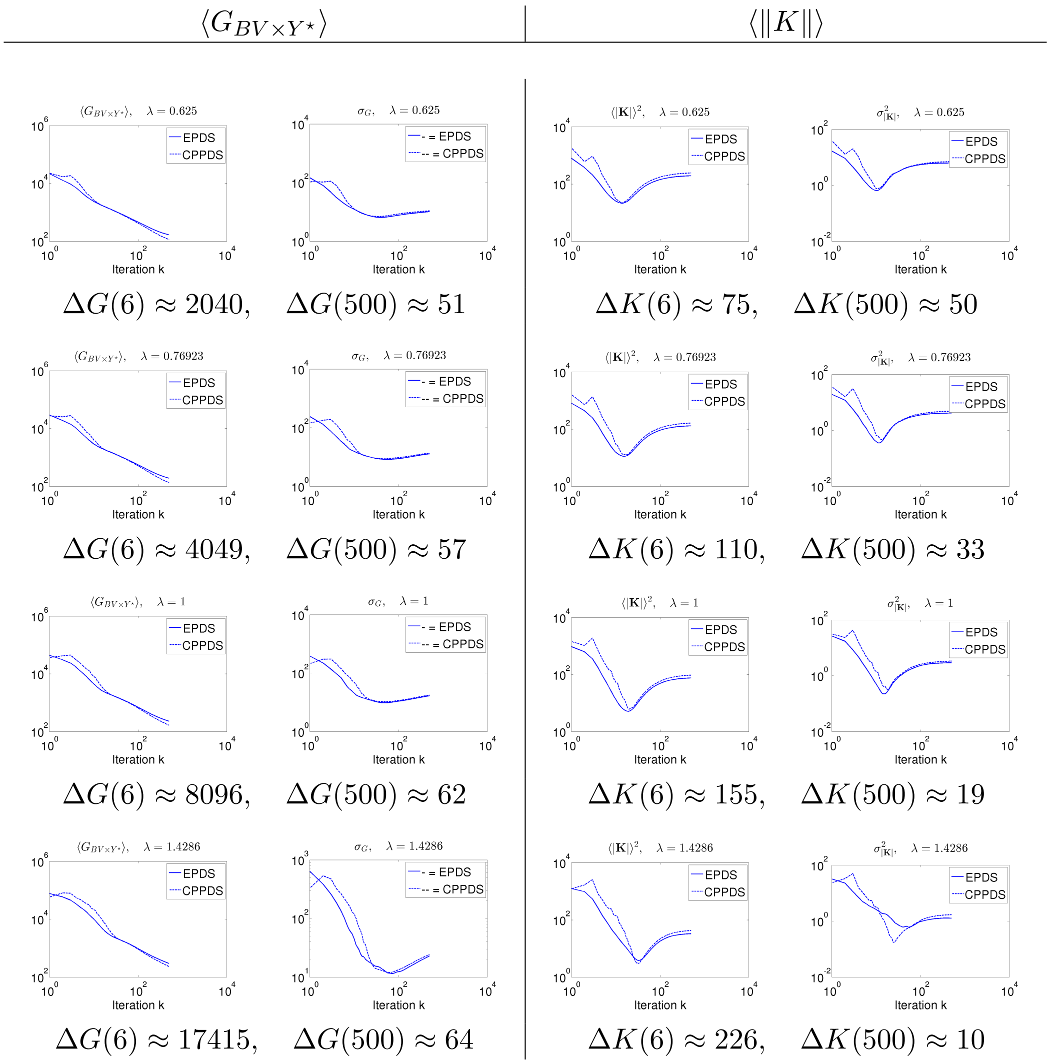


Figure 5.11.: This table shows the results of the EPDS (alg. 8) and the CPPDS algorithm (alg. 7) on the data in figure 5.10d. The rows correspond to four different values of the smoothing parameter λ in ascending order. The left column shows the mean and the variance of the primal dual gap $G_{BV\times Y^\star}$ in a log-plot over the iteration counter k for the EPDS and the CPPDS algorithm and The right column shows the same plots for the curvature $\|K\|$. $\Delta G(k)$ ($\Delta K(k)$) is the difference of the primal-dual gap (the curvature) of the EPDS and the CPPDS algorithms at iteration k.

5.4. summary

We have introduced a new algorithm based on the principles developed in section 3. In that section we had shown that for any convex energy functional of the

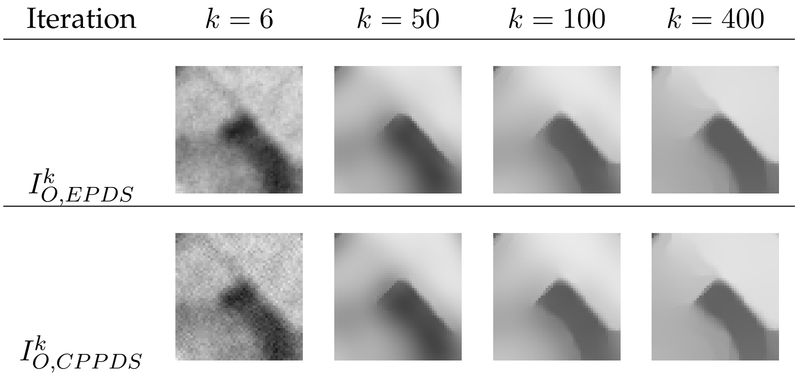


Figure 5.12.: This figure shows a region of interest (ROI) of the image I_O^k for different iterations k for the EPDS (upper row, alg. 8) and the CPPDS (lower row, alg. 7). The ROI $I^{10}_{O,EPDS}$ is smoother then $I^{10}_{O,CPPDS}$. However for higher iterations k EPDS over-smooths $I^k_{O,EPDS}$ at corner-like features in contrast to CPPDS

form

$$E\left(\phi, \boldsymbol{X_e}\phi\right) = E^{data}\left(\phi\right) + \lambda E^{prior}\left(\boldsymbol{X_e}\phi\right) \tag{5.63}$$

which is invariant under the action of an n-dimensional Lie group $\mathbb{G}^{\Omega\phi}$ there exists up to n differential operators satisfying

$$\widetilde{\mathbf{B}}^{\Omega}_g \phi_g = 0, \qquad \boldsymbol{p}^{pr}_g \in \partial E^{prior}(\boldsymbol{X}^{\Omega}_g \phi_g) \tag{5.64}$$

We introduced an optimization scheme based around a flow equation for the coordinate frame Ω

$$\boldsymbol{x}\left(t\right) = \theta^{B^{\Omega}} \circ \boldsymbol{x}, \qquad \left.\frac{d\boldsymbol{x}\left(t\right)}{dt}\right|_{t=0} = \mathbf{B}^{\Omega} \cdot \boldsymbol{x} \tag{5.65}$$

and termed the operator $\mathbf{B}^{\Omega}$ 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 divergence Div$(\boldsymbol{p}^{pr}_g)$, $\boldsymbol{p}^{pr}_g \in \partial E^{prior}$, under the flow in eq. (5.65) is equal to $\mathbf{K}\phi$, where the curvature operator $\mathbf{K}$ describes the curvature of the level-sets of ϕ. The bending operator $\mathbf{B}^{\Omega}$ obeys the relation $\mathbf{B}^{\Omega}\phi = 0$ for *any* value of the GRF ϕ at every time instance t of the flow in eq. (5.65) as a consequence of Noethers first theorem. Another consequence is that $\mathbf{B}^{\Omega}$ and thus $\mathbf{K}$ is entirely determined by the prior energy E^{prior} alone. The flow in eq. (5.65) minimizes the curvature $\mathbf{K}\phi_g$ and thus narrows down the solution space to the Kuhn-Tucker conditions in eq. (5.2).

For image de-noising with the total variation (TV) prior from section 2.7 we showed that eq. (5.65) dramatically improves the traditional Newton method for the minimization of the denoising energy functional $E(I_O, \nabla I_O)$, in terms of both speed and robustness of the solution $I_O^\star$ towards noise in the initial guess I_O^0. The functional derivative of the TV prior $\|\nabla\phi\|$ is equal to $\kappa(\boldsymbol{x}_{\theta B^\Omega(t,g^\Omega)})$, the mean curvature of the level-set at the point $\boldsymbol{x}_{\theta B^\Omega(t,g^\Omega)}$. We showed that $\kappa(\boldsymbol{x}_{\theta B^\Omega(t,g^\Omega)})$ 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. (5.65) is equal to the curvature $\mathbf{K}I_O$.

The other model we tested for image de-noising deployed the structure tensor prior E_{ST}^{prior} from section 4. We found that the flow in eq. (5.65) 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 $\nabla I_O \cdot \nabla^T I_O$) 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 $\mathbf{K}\phi$ (averaged over Ω) decreases as σ_{ST} increases. Since the curvature $\mathbf{K}\phi$ 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. (5.65).

In section 5.3 we incorporated the bending flow in eq. (5.65) into the primal-dual splitting algorithm of Chambolle and Pock [19] (CPPDS, alg. 7). The resulting algorithm dubbed the extended primal-dual splitting algorithm (EPDS, alg. 8) showed marginal improvements over CPPDS. While EPDS converged faster than CPPDS in the initial iterations it tends to over-smooth the image $I_O^\star$ compared to CPPDS. We argued that this observation is due to the fixed warping step τ^Ω in alg. 8 and that a future implementation of EPDS should adopt a dynamical warping step τ_k^Ω which decreases at higher iterations k of the algorithm.

6. 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 a convex energy functional

$$E(\phi, \boldsymbol{A}\phi) = E_Y^{data}(\phi) + E^{prior}(\boldsymbol{A}\phi) = \int_\Omega \mathcal{E}\left\{\phi(\boldsymbol{x}), \boldsymbol{A}\phi(\boldsymbol{x})\right\} d^2x, \quad \phi \in \Phi(\Omega) \tag{6.1}$$

is defined. The data term $E_Y^{data}(\phi)$ defines how the GRF ϕ is coupled to the data Y and the prior $E^{prior}(\boldsymbol{A}\phi)$ incorporates geometrical constraints on the level-sets of ϕ. Furthermore the function space $\Phi(\Omega)$ is a Hilbert space and the vector operator $\boldsymbol{A}$ is a linear operator on $\Phi(\Omega)$.

We outlined the theory of convex analysis in section 2.2 for convex functionals. The fundamental theorem in convex analysis is Fenchel's duality theorem from which it follows that the minimizers $\phi^\star$ of $E(\phi, \boldsymbol{A}\phi)$ in eq. (6.1) satisfies the Kuhn-Tucker conditions

$$\boldsymbol{A}^\dagger \boldsymbol{p}^{\star,pr} \in \partial E_Y^{data}(\phi^\star), \quad \boldsymbol{p}^{\star,pr} \in \partial E^{prior}(\boldsymbol{A}\phi^\star) \tag{6.2}$$

We emphasized that in order for the Kuhn-Tucker conditions of the energy $E(\phi, \boldsymbol{A}\phi)$ in eq. (6.1) to be satisfied for a minimizer $\phi^\star = \text{argmin} E(\phi, \boldsymbol{A}\phi)$ for a wide variety of data Y the prior $E^{prior}(\boldsymbol{A}\phi)$ must be invariant upon the action of a group of continuous smooth transformations $\mathbb{G}$, $g \circ E^{prior} = E^{prior}$ for all $g \in \mathbb{G}$. For instance in section 2.7 we reviewed the total variation prior

$$E_{TV}^{prior}(\nabla\phi) = \sup_{\boldsymbol{p}} \left\{ \int_\Omega \boldsymbol{p} \nabla\phi d^2x \,\middle|\, \|\boldsymbol{p}\| \leq 1 \right\} \tag{6.3}$$

The Kuhn-Tucker conditions for a GRF model containing E_{TV}^{prior} in eq. (6.3) are

$$\kappa = \text{div}\boldsymbol{p}^{\star,pr} \in \partial E_Y^{data}(\phi^\star), \quad \boldsymbol{p}^{\star,pr} \in \partial E_{TV}^{prior}(\nabla\phi^\star) \tag{6.4}$$

where $\kappa(\boldsymbol{x})$ is the curvature a level-set of $\phi^\star$ at the point $\boldsymbol{x} \in \Omega$. The total variation prior E_{TV}^{prior} penalizes level-sets with non-zero curvature κ. However E_{TV}^{prior} and thus κ are invariant under rotations of the domain Ω and hence the orientation of the level-sets of $\phi^\star$ is solely fixed by the data term E_Y^{data}.

To broaden this idea of invariant priors we reviewed the theory of Lie groups and their algebras in section 2.5. An n-dimensional Lie group $\mathbb{G}$ is a smooth manifold such that its elements $g \in \mathbb{G}$ are parameterized by the n-dimensional local coordinate vector $\boldsymbol{\xi}^g$. In addition $\mathbb{G}$ has the structure of a group such that for every $g \in \mathbb{G}$ there exists an inverse $g^{-1} \in \mathbb{G}$ and an identity element $e \cdot g = e$. The Lie algebra $\mathcal{G}$ is the n-dimensional vector space of smooth vector fields $V \in \mathcal{G}$ which are differential operators with respect to the local coordinates $\boldsymbol{\xi}^g$. V is itself a function of (the local coordinates of) $\mathbb{G}$ and the image V_g for $g \in \mathbb{G}$ can be understood as a tangential vector at g. Every vector field $V \in \mathcal{G}$ has a one-to-one relationship with a map $\theta^V : \mathbb{R} \times \mathbb{G} \to \mathbb{G}$ called a flow

$$\frac{d}{dt}\theta^V(t,g)\Big|_{t=0} = V_g \tag{6.5}$$

One of the core aspects of the Lie algebra $\mathcal{G}$ is that it is closed under the commutator $[\cdot,\cdot]$. The commutator itself has a geometric meaning: it describes how one vector field W transforms as it is transported along the flow θ^V of another vector field

$$\frac{d}{dt}\theta^V(t,\cdot) \circ W\Big|_{t=0} = [V,W] \in \mathcal{G}, \quad \forall V,W \in \mathcal{G} \tag{6.6}$$

In her famous paper [73, 74] Emmy Noether considered a Lie group $\mathbb{G}^{\Omega\phi}$ which operates on the functions $\phi \in \Phi(\Omega)$ by operating on the function values $\phi(\boldsymbol{x})$ at each $\boldsymbol{x} \in \Omega$ as well as on the locations $\boldsymbol{x}$ themselves. In the case that $E_Y(\phi, \nabla\phi)$ is smoothly differentiable and symmetric upon the action of the group $\mathbb{G}^{\Omega\phi}$, $g \circ E_Y = E_Y$ for all $g \in \mathbb{G}^{\Omega\phi}$, she discovered that there exist n vector valued functions $\boldsymbol{W}_m(\boldsymbol{x})$ for which the following relation holds

$$\operatorname{div}(\boldsymbol{W}_m) + (\omega_m^\phi - \omega_m^{\Omega,\mu}\partial_\mu\phi)\,[\mathcal{E}] = 0 \tag{6.7}$$

where ω_m^ϕ comes from the action of $\mathbb{G}^{\Omega\phi}$ on ϕ, the vector $\boldsymbol{\omega}_m^\Omega$ from the action of $\mathbb{G}^{\Omega\phi}$ on Ω and $\mathcal{E}$ are the Euler-Lagrange differentials

$$[\mathcal{E}] = \frac{\delta}{\delta\phi}\mathcal{E} - \operatorname{Div}(\boldsymbol{p}^{pr}), \quad \boldsymbol{p}^{pr} \in \partial E^{prior}(\boldsymbol{A}\phi) \tag{6.8}$$

Eq. (6.7) also holds for pure spatial group actions which leave the GRF ϕ invariant itself ($\omega_m^\phi = 0$)

$$\operatorname{div}(\boldsymbol{W}_m) - \omega_m^{\Omega,\mu}\partial_\mu\phi\,[\mathcal{E}] = 0 \tag{6.9}$$

In section 3 we combined the insight presented by Emmy Noether that led to eq. (6.7) with Fenchel's duality theorem in section 2.2 into a single variational principle. We considered the Lie group $\mathbb{G}^{\Omega\phi}$ to act in a two-fold manner on the function space $\Phi(\Omega)$

$$\phi_g(\boldsymbol{x}) = g \circ \phi(\boldsymbol{x}) = (g^{\phi} \circ \phi)(\boldsymbol{x}_{g^{\Omega}}), \quad \boldsymbol{x}_{g^{\Omega}} = g^{\Omega} \circ \boldsymbol{x} \tag{6.10}$$

and a general energy functional

$$E(\phi_g, \boldsymbol{X}_g^{\Omega}\phi_g) = E_Y^{data}(\phi_g) + E^{prior}(\boldsymbol{X}_g^{\Omega}\phi_g), \quad \phi \in \Phi(\Omega) \tag{6.11}$$

where the operators $\boldsymbol{X}^{\Omega} = \{X^{\Omega,1}, X^{\Omega,2}\}$ are the basis of a 2-dimensional subalgebra $\mathcal{X}^{\Omega} \subset \mathcal{G}^{\Omega\phi}$. It was shown that if the energy functional $E(\phi_g, \boldsymbol{X}_g^{\Omega}\phi_g)$ is invariant under the action in eq. (6.10) then there exists n spatial differential operators $\boldsymbol{B}^{\Omega,m}$ which act on the local coordinates of g^{Ω} in eq. (6.10). The operators $\boldsymbol{B}^{\Omega,m}$ satisfy

$$\text{div}\,(\boldsymbol{W}_m) - \omega_m^{\Omega,\mu}\partial_\mu\phi_g\,[\mathcal{E}] = \boldsymbol{B}_g^{\Omega,m}\phi_g = 0, \quad \boldsymbol{B}^{\Omega,m} = \boldsymbol{b}_g^{\Omega,m,T}\boldsymbol{X}_g^{\Omega} \tag{6.12}$$

for any function $\phi \in \Phi(\Omega)$ and any $g \in \mathbb{G}^{\Omega\phi}$. The coefficient function vector $\boldsymbol{b}_g^{\Omega,m,T}$ has a fixed relationship with the subdifferential of the prior $E^{prior}(\boldsymbol{X}_g^{\Omega}\phi_g)$

$$b_{g,j}^{\Omega,m} = \sum_i^n C_{m,i}^j p_i^{pr}, \quad \boldsymbol{p}^{pr} \in \partial E^{prior}(\boldsymbol{X}_g^{\Omega}\phi_g) \tag{6.13}$$

We showed that due to the invariance of the energy functional in eq. (6.11), encoded in eq. (6.12) the coefficient vector $\boldsymbol{b}_g^{\Omega}$ is constant in g when the Kuhn-Tucker conditions in eq. (6.2) are satisfied and from

$$\boldsymbol{b}_g^{\star,\Omega,T}\boldsymbol{X}_g^{\Omega}\phi_g^{\star} = 0 \tag{6.14}$$

it follows that the minimizer $\phi_g^{\star}$ has level-sets with the fixed tangential vector $\boldsymbol{b}_g^{\star,\Omega}$.

In section 4 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 4 we evaluated the use of both priors in the context of multi-modal optical flow, the results of which are summarized in section 4.7.7.

In section 5 we proposed a new kind of algorithm for the minimization problem $\phi_{tot}^\star = \text{argmin} E_Y(\phi, \boldsymbol{X}_e\phi)$. The algorithm is a steepest descent algorithm with two coupled flow equations: one for the GRF ϕ and one for the coordinate system Ω

$$\left.\frac{d\phi(\boldsymbol{x}(t),t)}{dt}\right|_{t=0} = [\mathcal{E}]\,(\phi, \boldsymbol{x}(t)) \tag{6.15}$$

$$\left.\frac{d\boldsymbol{x}(t)}{dt}\right|_{t=0} = \mathbf{B}\cdot\boldsymbol{x}(t) \tag{6.16}$$

where the operator $\mathbf{B}$ in eq. (6.16) is a linear combination of the operators $\boldsymbol{B}_m$ in eq. (6.12). The flow equation in eq. (6.15) 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 [37]. It is however mostly deployed on the rigid Cartesian coordinate frame Ω_0. This where the flow equation in eq. (6.16) comes into play. The rational behind eq. (6.16) 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 called $\mathbf{B}$ *bending* operator and tested the effectiveness of it 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. (6.16) 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. (6.16), 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. (6.16) 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

$$\left.\frac{d}{dt}\text{Div}(\boldsymbol{p}^{pr})\right|_{t=0} = \boldsymbol{K}\phi, \quad \boldsymbol{p}^{pr} \in \partial E^{prior}(\boldsymbol{X}_e^\Omega\phi) \tag{6.17}$$

The curvature $\boldsymbol{K}$ in eq. (6.17) is functionally dependent on the bending operator $\boldsymbol{B}$. We showed that due to $\boldsymbol{B}\phi = 0$ (eq. (6.12)) eq. (6.17) holds for the Euler-

Lagrange differentials of the total energy E_Y

$$\left.\frac{d\left[\mathcal{E}\right](\phi, \boldsymbol{x}(t))}{dt}\right|_{t=0} = \boldsymbol{K}\phi \tag{6.18}$$

It was shown that the flow θ^{B^Ω} in eq. (6.16) alone, that is without the Eulerian flow in eq. (6.15), 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. (6.16). Thus the result of minimizing the energy E_Y with the combined flow in eqs. (6.15) and (6.16) is only marginally better then deploying the original flow in eq. (6.15) 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. (6.15) and (6.16) as compared to running the flow in eq. (6.15) 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. (6.17) $\kappa(\boldsymbol{x})$ is related to the curvature operator $\boldsymbol{K}$

$$\dot{\kappa}(\boldsymbol{x}(t)) = \boldsymbol{K}\phi(\boldsymbol{x}(t)) \tag{6.19}$$

Eq. (6.19) suggests that the mean curvature $\kappa(\boldsymbol{x}(t))$ follows an exponential law, given that only the flow in eq. (6.16) is run. Indeed we were able to run the flow in eq. (6.16) 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. (6.15) and (6.16) is faster then eq. (6.15) alone.

In section 5.3 we extended the primal-dual splitting algorithm of Chambolle and Pock [19] (CPPDS) to include the bending flow in eq. (6.16). Our algorithm, dubbed the extended primal-dual splitting (EPDS) algorithm is, like CPPDS an iterative algorithm. We applied both the EPDS and the CPPDS to the image denoising problem with the total variation prior E_{TV}^{prior}, and found that while EPDS converged faster than CPPDS in the initial iterations it tends to oversmooth at higher iterations. We attributed this observation to the fixed step size in the discrete version of the flow $\theta^{B^{\Omega}}$ in eq. (6.16).

6.1. Outlook

The bending operator $\mathbf{B}$ in eq. (6.16) and the theory behind it was developed for GRF models with total energies $E(\phi_g, \boldsymbol{X}_g^{\Omega}\phi_g)$ 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 $E_Y(\phi_g, \boldsymbol{X}_g^k\phi_g)$, $k > 1$ along the same lines discussed in section 5. 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 theorem for all higher order models, we can expect the same positive results of a higher order bending algebra on the minimization of $E(\phi_g, \boldsymbol{X}_g^k\phi_g)$.

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 2.6 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 [79] to general relativity [65], Loop Quantum Gravity [91] and String Theory [4]. The development of the extended principle of least action in section 5 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 2.5 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_pM is defined. The elements $V_p \in T_pM$ 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 3.2 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 [55].

A.0.1. Topological Spaces

A topological manifold has a structure of open subsets called a topological space

Definition 23 (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 24 (Compact Space)**.** *A topological space X is compact if every open cover of X has a finite sub-cover*

Definition 25 (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 26 (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 27 (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 28 (Homeomorphic Map)**.** *Let X and Y be two topological spaces. A homeomorphic map $\psi : X \to Y$ 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 29 (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 $\psi_p : U_p \to \widetilde{U}_p \subset \mathbb{R}^n$*

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 11 (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 $\boldsymbol{x} \in \mathbb{Z}^n \cap \widetilde{U}$

$$\mathcal{B} = \left\{ \widetilde{B}_r(\boldsymbol{x}) \middle| r \in \mathbb{Z}, \quad \boldsymbol{x} \in \mathbb{Z}^n \cap \widetilde{U}, \quad \bar{B}_r \subset \widetilde{U} \; open \right\} \tag{A.1}$$

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) [55]. 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. □

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 30 (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 5 (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 11 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 11 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. □

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 31 (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 32 (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 12 (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 $\xi_p = \psi_p^{-1}$. □

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 33 (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 32 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 34 (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 35 (Tangential Space)**.** *The tangential space T_pU is the set of all derivations at $p \in U_p \subset M$*

Lemma 13 (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). □

Just like an ordinary gradient a derivation vanishes on constant functions as the following lemma shows

Lemma 14 (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). □

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 35 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 36 (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 15 (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 13. For any smooth function $f \in C^\infty(V_{F(p)})$ we can apply the definition of $F_\star$ (eq. (A.14)) to Z

$$\begin{aligned}(F_\star Z)f = Z(f \circ F) &= aX(f \circ F) + bY(f \circ F) \\ &= a(F_\star X)f + b(F_\star Y)f\end{aligned} \tag{A.15}$$

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

$$(F^{-1})_\star \circ F_\star = (F^{-1} \circ F)_\star = (Id_{U_p})_\star = Id_{T_pU} \tag{A.17}$$

Hence $(F_\star)^{-1} = (F^{-1})_\star$ over $T_{F(p)}V$. □

In the following we will show that T_pU and T_pM 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 $X \in T_pU$ defines an equivalence relation on $C^\infty(M)$.

Proposition 6 (Equivalence Relation). *Let M be a smooth manifold, $p \in M$, $X \in T_pU$ 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 14 to show that $X(h|_{U_p}) = X(\psi h|_{U_p}) = 0$. □

The equivalence relation in eq. (A.18) states that the actions of a derivation $X \in T_pU$ 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 $X \in T_pU$ there exists a $Y \in T_pM$ such that $Yf = X(f|_{U_p})$. In this sense the space T_pU is embedded in the space T_pM.

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 37 (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 7 (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 [55]). By proposition 6 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 37 we can utilize the definition of the push-forward (definition 36) 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 6 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_pM is the image of an element in T_pU and $\iota_\star$ is surjective. $\square$

Proposition 7 shows that every derivation in T_pU is uniquely identified with a derivation in T_pM by means of the inclusion map ι. Hence from now on we will consider the function space $C^(M)$ and identify T_pU with the set of derivations acting on $C^\infty(M)$, T_pM.

A.2. The Basis of T_pM

In this section we want clarify how the derivations $X \in T_pM$ 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_pM admits a basis of n differential operators $\partial_{\xi_p^i}|_q$, $1 \leq i \leq n$. 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 38 (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*

$$D_{v|a} f = \frac{d}{dt} f(a + vt)\bigg|_{t=0} = v_\mu \partial_\mu f\big|_a, \quad \partial_\mu \equiv \frac{\partial}{\partial x^\mu} \tag{A.27}$$

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 16 (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*

$$span(T_a\mathbb{R}^n) = \left\{ D_{e^i|a} \middle| \, e_i \in span(\mathbb{R}^n) \right\} \tag{A.30}$$

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. □

The extension of the euclidean directional derivative in eq. (A.27) to a directional derivative on a smooth manifold M is straightforward

Definition 39. *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 16). 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 vectorfield 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 40 (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 17 (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 [55].

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 $(TU_p, \widetilde{\psi}_p)$. □

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 41 (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 18 (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 $Yf \in C^\infty(M)$. □

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 36 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

$$Z_{F(p)} f = (F_\star X_p) f = X_p(f \circ F), \quad \forall f \in C^\infty(N), \, p \in M \tag{A.43}$$

however the issue is that the derivation Z must be defined over all N since $Zf \in C^\infty(N)$ is required (see lemma 18) 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 42 (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 8 (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 15 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

$$Z_q = Z^j(q)\partial_{\xi_q^j}, \quad Z^j(q) = (F_\star)^i_j\, X^i(p)\Big|_{p=F^{-1}(q)}, \quad (F_\star)^i_j = \partial_{\xi_p^j} F^i \tag{A.47}$$

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 $Z \in \mathcal{T}(N)$ ☐

A.5. Integral Curves and Flows

Proposition 8 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 and our new approach to the principle of least action in section 5.

To begin with, we will discuss the concept of an *integral curve* of any given smooth vector field $X \in \mathcal{T}(M)$.

Definition 43 (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

$$\gamma^V(t) = (\xi^\gamma_{p,1}(t), \cdots, \xi^\gamma_{p,n}(t)) = (\xi_{p,1}(\widetilde{\gamma}^V(t)), \cdots, \xi_{p,n}(\widetilde{\gamma}^V(t))) \tag{A.52}$$

We can now express both sides of eq. (A.50) in terms of the basis of $T_{\gamma^V(t_0)}$

$$\left(\frac{d}{dt}\xi^\gamma_{p,i}\right)\partial_{\xi_{p,i}} = V_i(\gamma^V(t))\partial_{\xi_{p,i}} \tag{A.53}$$

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

$$\begin{array}{rcl} \frac{d}{dt}\xi^\gamma_{p,1} & = & V_1(\gamma^V(t)) \\ \vdots & = & \vdots \\ \frac{d}{dt}\xi^\gamma_{p,n} & = & V_n(\gamma^V(t)) \end{array} \tag{A.54}$$

The existence of the eq. (A.54) is guaranteed by the ODE theorem

Theorem 3 (ODE Existence, Uniqueness and Smoothness)**.** *Let M be a smooth manifold and $V \in TM$ a smooth vector field on M. Then the solution $\boldsymbol{\xi}^\gamma_p$ of the differential equations in eq. (A.54) exists and is both smooth and unique*

The proof for theorem 3 is rather lengthly and we refer the interested reader to [55]. However we will give a short motivation of why theorem 3 holds. 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 $\gamma^V_{U_p} : D \to U_p$. Since M is a topological space and thus locally euclidean (see 29) the subset $U_p \subset M$ can be assumed to be euclidean. Hence $\gamma^V_{U_p}(t)$ exists uniquely and is approximately a straight line with the tangent V_p. The curve $\gamma^V_{U_p}$ 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 $\gamma^V_{U_p}$ to U_q by solving eq. (A.54) in U_q and combining the resulting solution $\gamma^V_{U_q}$ with $\gamma^V_{U_p}$ to a curve with a larger domain. Since the charts (ψ_p, U_p) and (ψ_q, U_q) are smoothly compatible and the vector field V is smooth, there exists a smooth transition from the solution $\gamma^V_{U_p}$ to the solution $\gamma^V_{U_q}$. Thus $\gamma^V(t)$ is smooth across its entire 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 44 (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

$$\frac{d}{dt}\theta^V(t,p)\bigg|_{t=0} = V_p \tag{A.59}$$

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 4 (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 $p \in M$, $\theta^V_p : \mathcal{D}^p \to M$ 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 4 and refer the reader for its lengthly proof to [55]. Part b and c of theorem 4 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 4.

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

$$\lim_{h\to 0} \frac{W_{\theta_p^V(t+h)} - W_{\theta_p^V(t)}}{h} \tag{A.61}$$

The problem with the derivative in eq. (A.61) is that $W_{\theta_p^V(t+h)}$ and $W_{\theta_p^V(t)}$ belong to *different* tangential spaces $T_{\theta_p^V(t+h)}M$ and $T_{\theta_p^V(t)}M$ and thus eq. (A.61) is not defined. However by the fundamental theorem for flows in theorem 4 the map θ_t^V is a diffeomorphism and thus the push-forward $(\theta_{-t}^V)_\star$ smoothly maps $T_{\theta_p^V(t)}M$ to T_pM for all $t \in \mathcal{D}^p$. This means the vector $(\theta_{-t}^V)_\star W_{\theta_p^V(t)} \in T_pM$ is smooth in t and thus differentiable

Definition 45 (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*

$$(\mathcal{L}_V W)_p = \frac{d}{dt}\left((\theta_{-t}^V)_\star W_{\theta_p^V(t)}\right)\Big|_{t=0} \tag{A.62}$$

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 9 (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 ineq. (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. □

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 4). 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.

Definition 46 (Structure Constants). *Let $\partial_{\xi_p^i}$ and $\partial_{\xi_p^j}$ be two basis elements of the n-dimensional tangential space T_pM at the point $p \in M$. Since T_pM is a linear vector* *space, the commutator* $\left[\partial_{\xi_p^i}, \partial_{\xi_p^j}\right]$ *can be expressed in the basis of* T_pM

$$\left[\partial_{\xi_p^i}, \partial_{\xi_p^j}\right] = \sum_{l=1}^{n} C_{i,j}^{l} \partial_{\xi_p^l} \tag{A.72}$$

The coefficients $C_{i,j}^l$ *are called the structure constants. They are antisymmetric with respect to* i *and* j, $C_{i,j}^l = -C_{j,i}^l$.

B. Derivation Of Noethers Theorem

In this section we want to derive Noethers theorem from the variation of under the action of the flow θ^V of any smooth vector field $V \in \mathcal{G}$ (eq. (2.154)).

Theorem 5 (Noethers Theorem). *Let $\mathbb{G}$ be a finite n-dimensional Lie group and $\mathcal{G}$ its Lie algebra. The energy*

$$E\left(\phi_g, \boldsymbol{X}_g^{\Omega}\phi_g\right) = \int_{\Omega} \mathcal{E}\left(\phi_g\left(\boldsymbol{x}\right), \boldsymbol{X}_g^{\Omega}\phi_g\left(\boldsymbol{x}\right)\right) d^2x \tag{B.1}$$

is invariant upon the action of the flow θ_t^V of any smooth vector field $V \in \mathcal{G}$

$$\partial_t^V E|_{t=0} = 0 \tag{B.2}$$

if and only if there exists n vector valued functions $\boldsymbol{W}_m(\boldsymbol{x})$ such that the following n relations hold

$$Div\left(\boldsymbol{W}_m\right) + \left(\omega_m^{\phi} - \omega_m^{\Omega,\mu}\partial_\mu\phi_g\right)[\mathcal{E}] = 0, \quad 1 \le m \le r \tag{B.3}$$

where $[\mathcal{E}]$ is the Euler-Lagrange differential

$$[\mathcal{E}] = \frac{\delta\mathcal{E}}{\delta\phi_g} - Div(\boldsymbol{p}_g^{pr}), \quad \boldsymbol{p}_g^{pr} = \frac{\delta\mathcal{E}^{prior}}{\delta\boldsymbol{X}_g^{\Omega}\phi_g} \tag{B.4}$$

and the variations ω_i^{ϕ} and $\boldsymbol{\omega}_i^{\Omega}$ are defined as (see eq. (3.42))

$$\omega_i^{\phi}(\boldsymbol{x}) := \partial_{\xi^{i,g}}\phi_g(\boldsymbol{x}) \quad \boldsymbol{\omega}_i^{\Omega}(\boldsymbol{x}) := \partial_{\xi^{i,g}}^{\Omega}\boldsymbol{x}_{g^{\Omega}} \tag{B.5}$$

Proof. As a starting point we take the expression of the flow subdifferential $\partial_t^V E$ in eq. (3.40)

$$\partial_t^V E(\phi_g, \boldsymbol{X}_g^{\Omega}\phi_g) = \left\{\int_{\Omega}\left(\sum_{i=1}^{2} p_i^{pr}\left[V_g^{\Omega}, X_g^{i,\Omega}\right]\phi_g + [\mathcal{E}]\, V_g\phi_g\right) d^2x\right\} \tag{B.6}$$

$$[\mathcal{E}] = \frac{\delta\mathcal{E}}{\delta\phi_g} - \mathrm{Div}(\boldsymbol{p}_g^{pr}), \quad \boldsymbol{p}_g^{pr} = \frac{\delta\mathcal{E}^{prior}}{\delta\boldsymbol{X}_g^{\Omega}\phi_g} \tag{B.7}$$

We use the identity

$$-\int_{\Omega}\left(\mathrm{Div}(\boldsymbol{p}_g^{pr})V_g\phi_g\right)d^2x=\int_{\Omega}\left(\boldsymbol{p}_g^{pr,T}\boldsymbol{X}_g^{\Omega}V_g\phi_g\right)d^2x \tag{B.8}$$

Hence we have

$$\partial_t^V E=\int\left\{V_g\phi_g\frac{\delta\mathcal{E}}{\delta\phi}+\sum_{i=1}^{2}p_g^{i,pr}\left(X_g^{\Omega,i}\left(V_g\phi_g\right)+\left[V_g^{\Omega},X_g^{\Omega,i}\right]\phi_g\right)\right\}d^2x \tag{B.9}$$

We add zeros in the form

$$V_g^{\Omega}\phi_g\frac{\delta\mathcal{E}}{\delta\phi}-V_g^{\Omega}\phi_g\frac{\delta\mathcal{E}}{\delta\phi}=0 \tag{B.10}$$

to eq. (B.9) and use

$$X_g^{\Omega,i}\left(V_g\phi_g\right)+\left[V_g^{\Omega},X_g^{\Omega,i}\right]\phi_g=X_g^{\Omega,i}\left(V_g\phi_g-V_g^{\Omega}\phi\right)+V_g^{\Omega}\left(X_g^{\Omega,i}\phi\right) \tag{B.11}$$

After some reordering of terms in eq. (B.9) and factorizing out $(V_g\phi_g-V_g^{\Omega}\phi)$ we get

$$\partial_t^V E=\int\left\{V_g^{\Omega}\phi\frac{\delta\mathcal{E}}{\delta\phi}+\sum_{i=1}^{2}\left(V_g^{\Omega}\left(X_g^{\Omega,i}\phi\right)\right)p_g^{i,pr}\right. \tag{B.12}$$

$$\left.+\left(V_g\phi_g-V_g^{\Omega}\phi\right)\frac{\delta\mathcal{E}}{\delta\phi}+\sum_{i=1}^{2}p_g^{i,pr}X_g^{\Omega,i}\left(V_g\phi_g-V_g^{\Omega}\phi\right)\right\}d^2x \tag{B.13}$$

The summands in eq. (B.12) can be combined to the total derivative acting on the energy density $\mathcal{E}$

$$\left(V_g^{\Omega}\phi\right)\frac{\delta\mathcal{E}}{\delta\phi}+\sum_{i=1}^{2}\left(V_g^{\Omega}\left(X_g^{\Omega,i}\phi\right)\right)p_g^{i,pr}=v^{\mu}\frac{d\mathcal{E}}{dx^{\mu}},\quad v^{\mu}=V_g^{\Omega}x_{g^{\Omega}}^{\mu} \tag{B.14}$$

Since we assumed the vector field V to be volume-preserving (eq. (3.15)) v^{μ} in eq. (B.14) is divergence free and thus

$$v^{\mu}\frac{d\mathcal{E}}{dx^{\mu}}=\frac{d}{dx^{\mu}}\left(v^{\mu}\mathcal{E}\right) \tag{B.15}$$

so that eq. (B.13) and eq. (B.14) simplify to

$$\partial_t^V E = \int \left\{ \frac{d}{dx^\mu} \left(v^\mu \mathcal{E} \right) + \left(V_g \phi_g - V_g^\Omega \phi \right) \frac{\delta \mathcal{E}}{\delta \phi} \right. \tag{B.16}$$

$$\left. + \sum_{i=1}^{2} X_g^{\Omega,i} \left(V_g \phi_g - V_g^\Omega \phi \right) p_g^{i,pr} \right\} d^2 x \tag{B.17}$$

We expand the operator $X_g^{\Omega,i}$ in terms of the Cartesian gradient ∇, $X_g^{\Omega,i} = \omega_{X,i}^{\Omega,\mu} \partial_\mu$. Then we can convert the summand in eq. (B.17) into the sum of a total divergence and a summand proportional to the divergence of the dual variable $\boldsymbol{p}_g^{pr}$

$$\sum_{i=1}^{2} X_g^{\Omega,i} \left(V_g \phi_g - V_g^\Omega \phi \right) p_g^{i,pr} = \sum_{i=1}^{2} \frac{d}{dx^\mu} \left(\omega_{X,i}^{\Omega,\mu} \left(V_g \phi_g - V_g^\Omega \phi \right) p_g^{i,pr} \right)$$
$$- \left(V_g \phi_g - V_g^\Omega \phi \right) \mathrm{Div} \left(\boldsymbol{p}_g^{pr} \right), \quad \mathrm{Div} \left(\boldsymbol{p}_g^{pr} \right) = \sum_{i=1}^{2} \frac{d}{dx^\mu} \left(\omega_{X,i}^{\Omega,\mu} p_g^{i,pr} \right) \tag{B.18}$$

We substitute eq. (B.17) with the right hand side of eq. (B.18) and group all the total divergences together as well as the term proportional to $(V_g \phi_g - V_g^\Omega \phi)$

$$\partial_t^V E = \int \left\{ \frac{d}{dx^\mu} \left(R_g^\mu \right) + \left(V_g \phi_g - V_g^\Omega \phi \right) [\mathcal{E}] \right\} d^2 x \tag{B.19}$$

$$R_g^\mu = v^\mu \mathcal{E} + \left(V_g \phi_g - V_g^\Omega \phi \right) \sum_{i=1}^{2} \omega_{X,i}^{\Omega,\mu} p_g^{i,pr}, \quad [\mathcal{E}] = \frac{\delta \mathcal{E}}{\delta \phi} - \mathrm{Div} \left(\boldsymbol{p}^{pr} \right) \tag{B.20}$$

If eq. (B.2) holds for the specific flow θ_t^V of the vector field V_g then it follows that the integrand of the right hand side of eq. (B.19) must vanish

$$\frac{d}{dx^\mu} \left(R^\mu \right) + \left(V_g \phi_g - V_g^\Omega \phi \right) [\mathcal{E}] = 0 \tag{B.21}$$

The vector field $\boldsymbol{R}_g$ in eq. (B.19) measures the flux of energy through the boundary $\partial\Omega$. Eq. (B.21) means that if E is invariant with respect to the flow θ_t^V then the density $[\mathcal{E}]$ in the interior of Ω is balanced by the flux $\boldsymbol{R}_g$ on the boundary $\partial\Omega$. The statement in eq. (B.2) applies for all vector fields in the algebra $\mathcal{G}$, therefore we need to expand V_g in terms of the basis of $\mathcal{G}$, $\partial_{\xi^{m,g}}$ (see eq. (2.154))

$$V_g = v^m(\boldsymbol{\xi}^g) \partial_{\xi^{m,g}} \Longrightarrow V_g \phi_g = v^m(\boldsymbol{\xi}^g) \omega_m^\phi, \quad v^{\mu,\Omega} = v^m(\boldsymbol{\xi}^g) \omega_m^{\mu,\Omega}, \quad 1 \le m \le n \tag{B.22}$$

then eq. (B.21) translates to

$$\partial_t^V E = \sum_{m=1}^{n} v^m(\boldsymbol{\xi}^g)\left\{\frac{d}{dx^\mu}\left(W_m^\mu\right) + \left(\omega_m^\phi - \omega_m^{\Omega,\mu}\partial_\mu\phi_g\right)[\mathcal{E}]\right\} = 0 \tag{B.23}$$

$$W_m^\mu = \omega_m^{\Omega,\mu}\mathcal{E} + \sum_{i=1}^{2}\omega_{X,i}^{\Omega,\mu}\left(\omega_m^\phi - \omega_m^{\Omega,\mu}\partial_\mu\phi_g\right)p_g^{i,pr} \tag{B.24}$$

Eq. (B.23) holds for all smooth vector fields $V \in \mathcal{G}$ if the sum vanishes for any configuration of the coefficients $v^i(\boldsymbol{\xi}^g)$. It follows that eq. (B.3) holds for all n vector valued functions $\boldsymbol{W}_m$. □

C. Multimodal Optical Flow

We want to derive the similarity measure $E^{data}_{y,I}$ in a fashion which covers both the global version $E^{data}_{y,I}(\sigma^{sc}, \boldsymbol{d})$ (eq. (4.51)) and the local version $E^{data}_{y,I}(\sigma^{sc}, \boldsymbol{d})$ (eq. (4.56)). 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

$$y(\boldsymbol{x}) \rightarrow \underline{y}, \quad \int u(\boldsymbol{x})v(\boldsymbol{x})d^2x \rightarrow \underline{u}^T\underline{v} \tag{C.1}$$

$$\int_{\mathcal{A}_{\boldsymbol{x}_0}} w(\boldsymbol{x}-\boldsymbol{x}_0)\cdot y(\boldsymbol{x})d^2x \rightarrow \mathbf{W}\underline{y} \tag{C.2}$$

To derive the similarity measure $E^{data}_{y,I}$ we considered in section 4.3 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 4.3 the low resolution image $\underline{y}$ is assumed to be result of a sub-sampling process with additive Gaussian noise

$$\underline{y} = \mathbf{W}_{\sigma^{sc}}\underline{Y} + \underline{n}, \quad \underline{n} \sim \mathcal{N}(\mathbf{0}, C_n) \tag{C.3}$$

where the point-spread function (PSF) $\mathbf{W}_{\sigma^{sc}}$ which models the difference in optical scale is approximately Gaussian. Eq. (C.3) connects the low frequency components of $\underline{Y}$ to the low resolution image $\underline{y}$. From eq. (C.3) the log-likelihood of $\underline{y}$ given $\underline{Y}$ is given by

$$-\ln\left(p\left(\underline{y}|\underline{Y}\right)\right) = \frac{1}{2}\left(\underline{y} - \mathbf{W}_{\sigma^{sc}}\underline{Y}\right)^T \mathbf{C}_n^{-1}\left(\underline{y} - \mathbf{W}_{\sigma^{sc}}\underline{Y}\right) \tag{C.4}$$

To compute the similarity measure $E^{data}_{y,I}$, the correspondence between $\underline{Y}$ and the warped image $\underline{I}_{\boldsymbol{d}}$ 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 $\underline{I}_{\boldsymbol{d}}$ to be jointly Gaussian with the distribution $p(\underline{Y}, \underline{I}_{\boldsymbol{d}})$

$$-\ln\left(p(\underline{Y}, \underline{I})\right) = \frac{1}{2}\underline{V}^T \mathbf{Q} \underline{V}, \quad \underline{V}^T = \begin{pmatrix} \underline{Y} - \underline{\mu}_Y \\ \underline{I}_{\boldsymbol{d}} - \underline{\mu}_{I_{\boldsymbol{d}}} \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} \mathbf{C}_{YY} & \mathbf{C}_{YI_{\boldsymbol{d}}} \\ \mathbf{C}_{I_{\boldsymbol{d}}Y} & \mathbf{C}_{I_{\boldsymbol{d}}I_{\boldsymbol{d}}} \end{pmatrix}^{-1} \tag{C.5}$$

The covariances $\mathbf{C}_{\cdot,\cdot}$ and the means $\underline{\mu}_{\cdot}$ in eq. (C.5) can be either the global covariance and mean in eq. (2.230) or the local variance and mean in eq. (4.34). In both cases $\mathbf{C}_{\cdot,\cdot}$ is diagonal. The precision matrix $\mathbf{Q}$ can be expressed in terms of the Schur-Complement $\mathbf{S} = \mathbf{C}_{YY} - \mathbf{C}_{YI_{\boldsymbol{d}}}\mathbf{C}^{-1}_{I_{\boldsymbol{d}}I_{\boldsymbol{d}}}\mathbf{C}_{I_{\boldsymbol{d}}Y}$

$$\mathbf{Q} = \begin{pmatrix} \mathbf{S}^{-1} & -\mathbf{S}^{-1}\mathbf{C}_{YI_{\boldsymbol{d}}}\mathbf{C}^{-1}_{I_{\boldsymbol{d}},I_{\boldsymbol{d}}} \\ \mathbf{C}^{-1}_{I_{\boldsymbol{d}},I_{\boldsymbol{d}}}\mathbf{C}_{I_{\boldsymbol{d}}Y}\mathbf{S}^{-1} & \star \end{pmatrix} \tag{C.6}$$

then the joint distribution in eq. (C.5) can be split into the product $p(\underline{Y}, \underline{I}_{\boldsymbol{d}}) = p(\underline{Y}|\underline{I}_{\boldsymbol{d}})p(\underline{I}_{\boldsymbol{d}})$ where the conditional distribution $p(\underline{Y}|\underline{I}_{\boldsymbol{d}})$ can be written in terms of the Schur complement $\mathbf{S}$

$$-\ln\left(p(\underline{Y}|\underline{I}_{\boldsymbol{d}})\right) = \frac{1}{2}\underline{\widetilde{Y}}^T\mathbf{S}^{-1}\underline{\widetilde{Y}} - \frac{1}{2}\underline{\widetilde{Y}}^T\mathbf{S}^{-1}\mathbf{C}_{YI_{\boldsymbol{d}}}\mathbf{C}^{-1}_{I_{\boldsymbol{d}}I_{\boldsymbol{d}}}\underline{\widetilde{I}}_{\boldsymbol{d}} - \frac{1}{2}\underline{\widetilde{I}}_{\boldsymbol{d}}^T\mathbf{C}^{-1}_{I_{\boldsymbol{d}}I_{\boldsymbol{d}}}\mathbf{C}_{I_{\boldsymbol{d}}Y}\mathbf{S}^{-1}\underline{\widetilde{Y}} \tag{C.7}$$

$$= \frac{1}{2}\left(\underline{Y} - \underline{\mu}_{Y|I_{\boldsymbol{d}}}\right)^T \mathbf{C}^{-1}_{Y|I_{\boldsymbol{d}}}\left(\underline{Y} - \underline{\mu}_{Y|I_{\boldsymbol{d}}}\right) \tag{C.8}$$

$$\mathbf{C}_{Y|I_{\boldsymbol{d}}} = \mathbf{S}, \quad \underline{\mu}_{Y|I_{\boldsymbol{d}}} = \underline{\mu}_Y + \mathbf{C}_{YI_{\boldsymbol{d}}}\mathbf{C}^{-1}_{I_{\boldsymbol{d}}I_{\boldsymbol{d}}}\underline{\widetilde{I}}_{\boldsymbol{d}} \tag{C.9}$$

where $\underline{\widetilde{I}}_{\boldsymbol{d}} = \underline{I}_{\boldsymbol{d}} - \underline{\mu}_{I_{\boldsymbol{d}}}$ and $\underline{\widetilde{Y}} = \underline{Y} - \underline{\mu}_Y$. We combine the conditional $p(\underline{Y}|\underline{I}_{\boldsymbol{d}})$ from eq. (C.8) together with the likelihood $p\left(\underline{y}|\underline{Y}\right)$ from eq. (C.4) to the posterior $p\left(\underline{Y}|\underline{y}, \underline{I}_{\boldsymbol{d}}\right) = p\left(\underline{y}|\underline{Y}\right)p(\underline{Y}|\underline{I}_{\boldsymbol{d}})$ with the energy $E(\underline{Y})$

$$\begin{aligned} E(\underline{Y}) = &\frac{1}{2}\left(\underline{y} - \mathbf{W}_{\sigma^{sc}}\underline{Y}\right)^T \mathbf{C}_n^{-1}\left(\underline{y} - \mathbf{W}_{\sigma^{sc}}\underline{Y}\right) \\ &+ \frac{1}{2}\left(\underline{Y} - \underline{\mu}_{Y|I_{\boldsymbol{d}}}\right)^T \mathbf{C}^{-1}_{Y|I_{\boldsymbol{d}}}\left(\underline{Y} - \underline{\mu}_{Y|I_{\boldsymbol{d}}}\right) \end{aligned} \tag{C.10}$$

The minimizer $\underline{Y}^\star_{\boldsymbol{d}}$ of the energy $E(\underline{Y})$ is

$$\underline{Y}^\star_{\boldsymbol{d}} = \underline{\mu}_{Y|I_{\boldsymbol{d}}} + \mathbf{C}_{Y|I_{\boldsymbol{d}}}\mathbf{W}^T_{\sigma^{sc}}\mathbf{H}^{-1}\left(\underline{y} - \mathbf{W}\underline{\mu}_{Y|I_{\boldsymbol{d}}}\right) \tag{C.11}$$

$$\mathbf{H} = \mathbf{W}_{\sigma^{sc}}\mathbf{C}_{Y|I_{\boldsymbol{d}}}\mathbf{W}^T_{\sigma^{sc}} + \mathbf{C}_n \tag{C.12}$$

We define the similarity measure $E_{y,I}^{data}$ in terms of the conditional distribution $p(\underline{Y}|\underline{I}_{\boldsymbol{d}})$ from eq. (C.8) evaluated at the minimizer $\underline{Y}_{\boldsymbol{d}}^{\star}$

$$E_{y,I}^{data}\left(\boldsymbol{d}\right) = -\ln\left(p(\underline{Y}_{\boldsymbol{d}}^{\star}|\underline{I}_{\boldsymbol{d}})\right) \tag{C.13}$$

which after some manipulations becomes

$$E_{y,I}^{data}\left(\boldsymbol{d}\right) = \frac{1}{2}\underline{x}^T\mathbf{H}^{-1}\mathbf{W}_{\sigma^{sc}}\mathbf{C}_{Y|I_{\boldsymbol{d}}}\mathbf{W}_{\sigma^{sc}}^T\mathbf{H}^{-1}\underline{x}, \quad \underline{x} = \underline{y} - \mathbf{W}_{\sigma^{sc}}\underline{\mu}_{Y|I_{\boldsymbol{d}}} \tag{C.14}$$

$E_{y,I}^{data}$ in eq. (C.14) is intractable to compute due to the dense inverse matrix $\mathbf{H}^{-1}$. Similar to [117] we make the simplifications

$$\mathbf{W}_{\sigma^{sc}}\mathbf{C}_{Y|I_{\boldsymbol{d}}}\mathbf{W}_{\sigma^{sc}}^T \to \mathbf{C}_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}} \tag{C.15}$$

$$\mathbf{H} \to \widetilde{\mathbf{H}} = \mathbf{C}_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}} + \mathbf{C}_n \tag{C.16}$$

and eq. (C.14) reduces to

$$E_{y,I}^{data}\left(\boldsymbol{d}\right) = \frac{1}{2}\underline{x}^T\widetilde{\mathbf{H}}^{-1}\mathbf{C}_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}\widetilde{\mathbf{H}}^{-1}\underline{x}, \quad \underline{x} = \underline{y} - \mathbf{W}_{\sigma^{sc}}\underline{\mu}_{Y|I_{\boldsymbol{d}}} \tag{C.17}$$

We rewrite eq. (C.17) back in terms of filters and integrals with the inverse lexicographic reordering If we insert the global covariances in eq. (2.230) into the precision matrix $\mathbf{Q}$ in eq. (C.5) we get the global similarity measure from eq. (4.27)

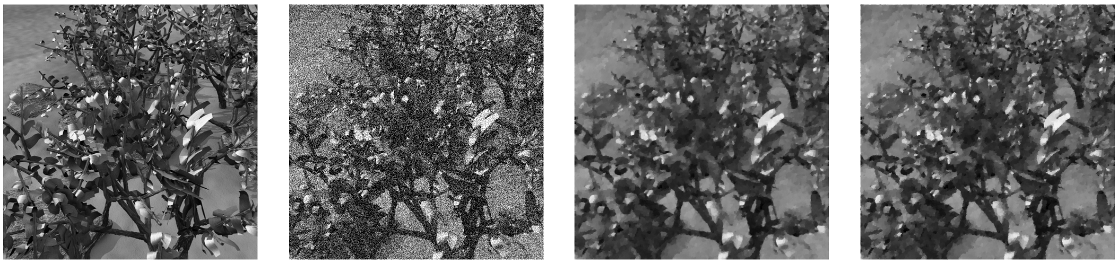
$$E_{y,I}^{data}\left(\boldsymbol{d}\right) = \frac{1}{2}\int_{\Omega}\Big(y\left(\boldsymbol{x}\right) - \mu_Y - f\cdot\left(\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}\left(\boldsymbol{x}\right) - \mu_I\right)\Big)^2\cdot U^{\sigma^{sc}} \tag{C.18}$$

$$f = C_{y,\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}^{\sigma^{sc}} C_{\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}},\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}^{\sigma^{sc},-1}, \quad U^{\sigma^{sc}} = C_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}^{\sigma^{sc}}\Big(C_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}^{\sigma^{sc}} + C_n\Big)^{-2} \tag{C.19}$$

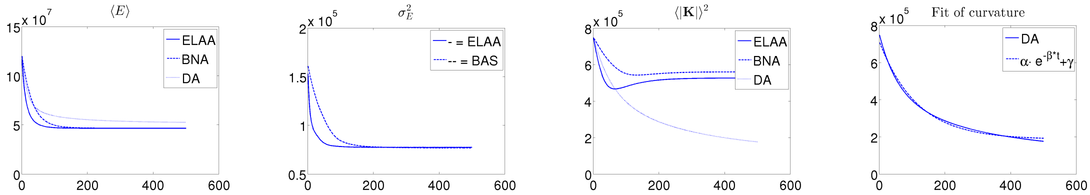
And if we insert the local covariances in eq. (4.34) into eq. (C.5) we get the local similarity measure $E_{y,I_{\boldsymbol{d}}}^{data,l}$ from eq. (4.38)

$$\begin{aligned} E_{y,I_{\boldsymbol{d}}}^{data,l}\left(\sigma^{sc},a,\boldsymbol{d}\right) &= \frac{1}{2}\int_{\Omega}\Big(y\left(\boldsymbol{x}\right) - \Big\langle f^{\sigma^{sc},a}I_{\boldsymbol{d}}\Big\rangle_{\sigma^{sc}}\left(\boldsymbol{x}\right)\Big)^2\cdot U^{\sigma^{sc},a}\left(\boldsymbol{x}\right)d^2x \\ U^{\sigma^{sc},a}\left(\boldsymbol{x}\right) &= C_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}^{\sigma^{sc},a}(\boldsymbol{x})\Big(C_{\langle Y\rangle_{\sigma^{sc}}|\langle I_{\boldsymbol{d}}\rangle_{\sigma^{sc}}}^{\sigma^{sc},a}(\boldsymbol{x}) + C_n\Big)^{-2} \end{aligned} \tag{C.20}$$

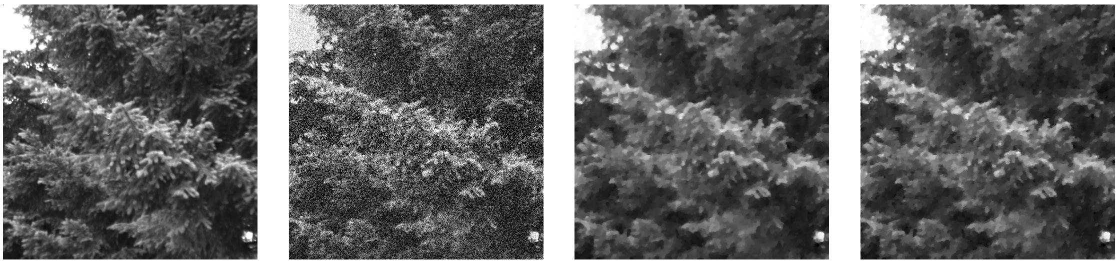
D. De-noising, supplementary results



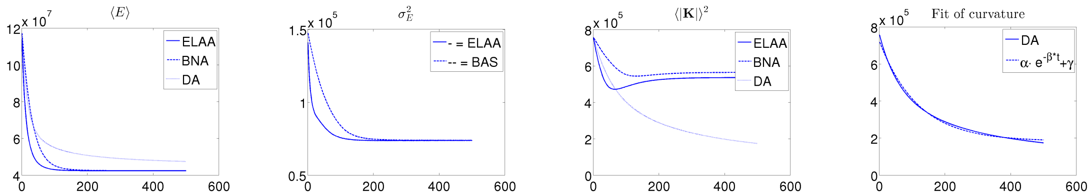
(a) input image **(b)** ϕ_0 **(c)** $\phi^\star_{ELAA}$ **(d)** $\phi^\star_{BNA}$



(e) **(f)** **(g)** **(h)**



(i) input image **(j)** ϕ_0 **(k)** $\phi^\star_{ELAA}$ **(l)** $\phi^\star_{BNA}$



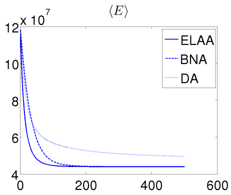
(m) **(n)** **(o)** **(p)**

(a) input image

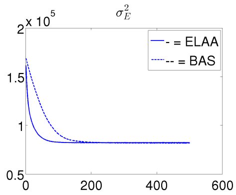
(b) ϕ_0

(c) $\phi^{\star}_{ELAA}$

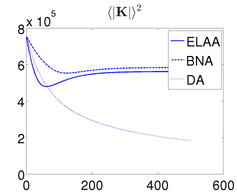
(d) $\phi^{\star}_{BNA}$



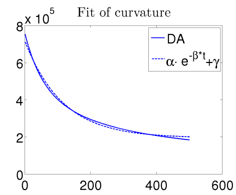
(e)



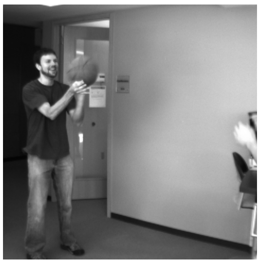
(f)



(g)

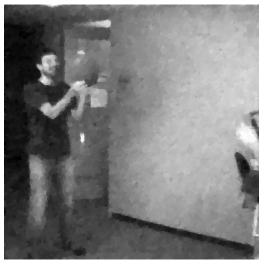


(h)

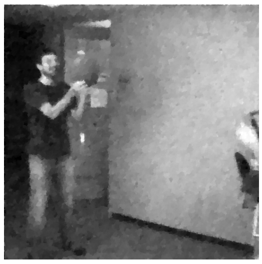


(i) input image

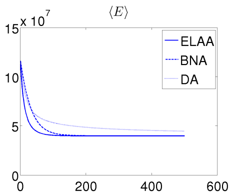
(j) ϕ_0



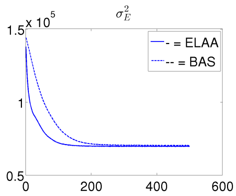
(k) $\phi^{\star}_{ELAA}$



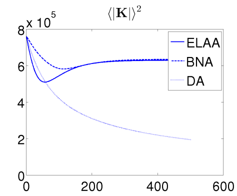
(l) $\phi^{\star}_{BNA}$



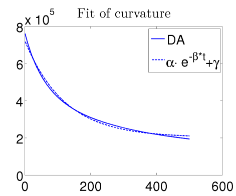
(m)



(n)



(o)



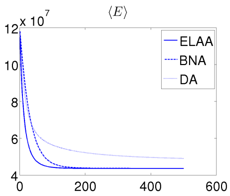
(p)

(a) input image

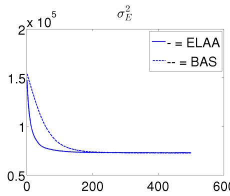
(b) ϕ_0

(c) $\phi^{\star}_{ELAA}$

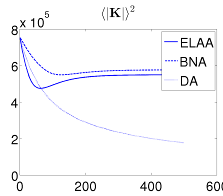
(d) $\phi^{\star}_{BNA}$



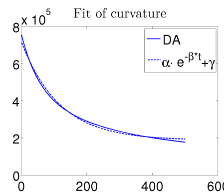
(e)



(f)



(g)



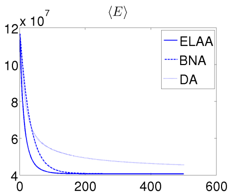
(h)

(i) input image

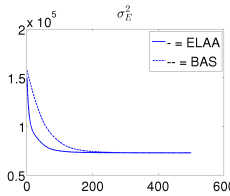
(j) ϕ_0

(k) $\phi^{\star}_{ELAA}$

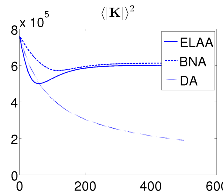
(l) $\phi^{\star}_{BNA}$



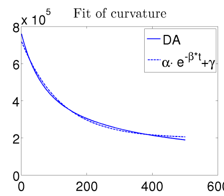
(m)



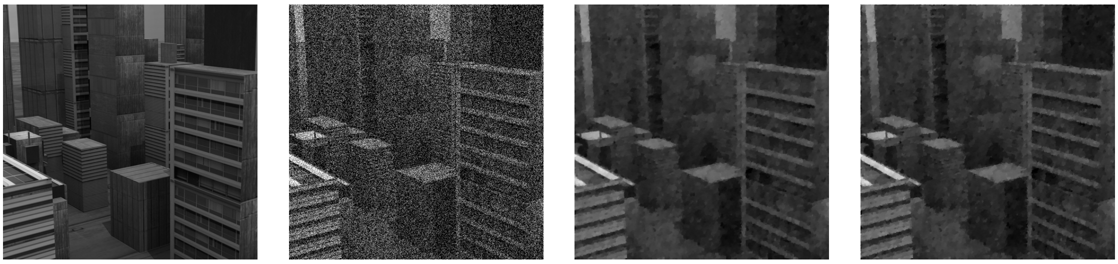
(n)



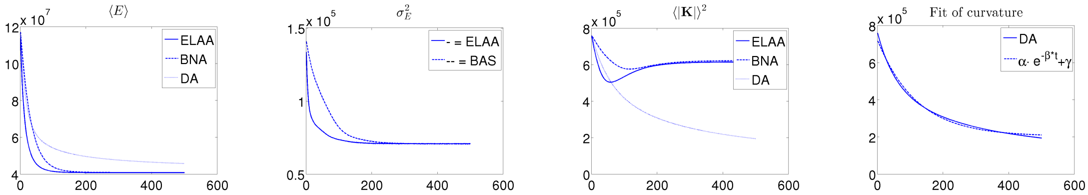
(o)



(p)

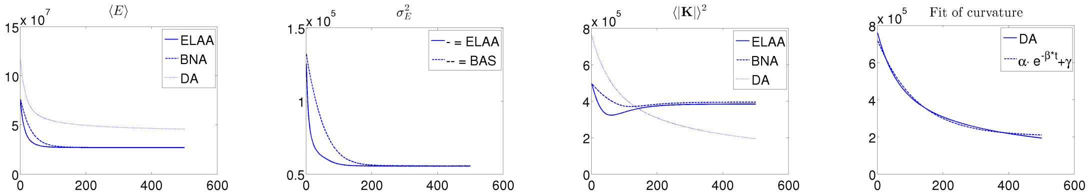


(a) input image **(b)** ϕ_0 **(c)** $\phi^\star_{ELAA}$ **(d)** $\phi^\star_{BNA}$



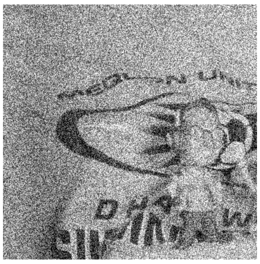
(e) **(f)** **(g)** **(h)**

(i) input image **(j)** ϕ_0 **(k)** $\phi^\star_{ELAA}$ **(l)** $\phi^\star_{BNA}$



(m) **(n)** **(o)** **(p)**

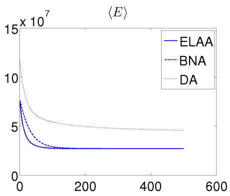
(a) input image



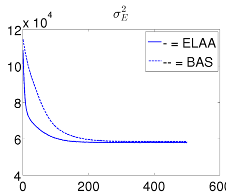
(b) ϕ_0

(c) $\phi^{\star}_{ELAA}$

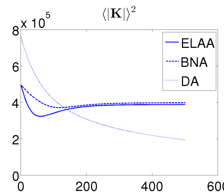
(d) $\phi^{\star}_{BNA}$



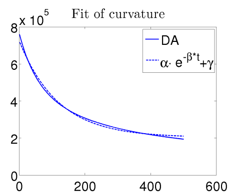
(e)



(f)



(g)



(h)

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. 9

[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. 80, 96, 111

[4] K. Becker, M. Becker, and J. H. Schwarz. *String theory and M-theory: A modern introduction*. Cambridge University Press, 2006. 129

[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. 45

[6] J. Bigun. *Vision with Direction A Systematic Introduction to Image Processing and Computer Vision*. Springer, 2006. 65, 67, 69, 95, 126

[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. 65, 69, 70, 95, 126

[8] J. Bigun, T. Bigun, and K. Nilsson. Recognition by symmetry derivatives and the generalized structure tensor. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 26 no. 12:1590–1605, 2004. 65, 69

[9] C. M. Bishop. *Neural networks for pattern recognition*. Oxford university press, 1995. 46

[10] 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. 45

[11] K. Brading and H. R. Brown. Symmetries and noether's theorems. *Symmetries in physics: Philosophical reflections*, pages 89–109, 2003. 5

[12] K. Bredies. A forward–backward splitting algorithm for the minimization of non-smooth convex functionals in banach space. *Inverse Problems*, 25(1): 015005, 2008. 14, 23

[13] K. Bredies and D. Lorenz. Mathematische bildverarbeitung. *Vieweg+ Teubner*, 4(6):12, 2011. 41, 42, 44

[14] 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. 45

[15] 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. 49

[16] J. B. Campbell and R. H. Wynne. *Introduction to remote sensing*. Guilford Press, 2011. 51

[17] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22 no 1:61–79. 44

[18] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, 20(1-2):89–97, 2004. 5, 7, 9, 28, 41

[19] 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. 5, 7, 8, 23, 24, 25, 118, 123, 128

[20] 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. 46, 49

[21] G. H. Chen and R. T. Rockafellar. Convergence rates in forward–backward splitting. *SIAM Journal on Optimization*, 7(2):421–444, 1997. 14

[22] P. L. Combettes. The foundations of set theoretic estimation. *Proceedings of the IEEE*, 81(2):182–208, 1993. 11

[23] P. L. Combettes*. Solving monotone inclusions via compositions of nonexpansive averaged operators. *Optimization*, 53(5-6):475–504, 2004. 14, 23

[24] 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. 14, 23

[25] 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. 11, 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. 14, 23

[27] R. R. Edelman and S. Warach. Magnetic resonance imaging. *New England Journal of Medicine*, 328(10):708–716, 1993. 48

[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. 44

[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. 14, 41

[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. 87

[31] O. Faugeras and R. Keriven. *Variational principles, surface evolution, pde's, level set methods and the stereo problem*. IEEE, 2002. 49

[32] M. Fels and P. J. Olver. Moving coframes: I. a practical algorithm. *Acta Applicandae Mathematica*, 51(2):161–213, 1998. 40

[33] M. Fels and P. J. Olver. Moving coframes: Ii. regularization and theoretical foundations. *Acta Applicandae Mathematica*, 55(2):127–208, 1999. 40

[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. 67

[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. 87

[36] R. P. Feynman, R. B. Leighton, and M. Sands. *The Feynman Lectures on Physics*. 1963. ISBN 0201021153, 9780201021158. URL http://www.feynmanlectures.info/. 37

[37] P. Fieguth. *Statistical image processing and multidimensional modeling*. Springer Science & Business Media, 2010. 6, 14, 103, 127

[38] FLIR. URL http://www.flir.com/home/. 52

[39] E. Freitag and R. Busam. *Funktionentheorie 1. 3., neu bearb. und erw. Aufl*. Berlin. 42

[40] O. Güler. On the convergence of the proximal point algorithm for convex minimization. *SIAM Journal on Control and Optimization*, 29(2):403–419, 1991. 13

[41] 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. 45

[42] S.-C. Han and C. I. Podilchuk. Video compression with dense motion fields. *IEEE Transactions on Image Processing*, 10(11):1605–1612, 2001. 45

[43] 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. 52, 53

[44] 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. 51, 52, 53, 54, 73, 74

[45] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 3

[46] D. Hilbert, J. v. Neumann, and L. Nordheim. Über die grundlagen der quantenmechanik. *Mathematische Annalen*, 98(1):1–30, 1928. 6

[47] H. Hirschmüller. Stereo processing by semiglobal matching and mutual information. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(2):328–341, 2008. 48

[48] B. K. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17, issues 1-3:185–203, 1981. 45, 46, 47, 51

[49] J. Hsieh. *Computed tomography: principles, design, artifacts, and recent advances*, volume 114. SPIE press, 2003. 48

[50] 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. 87

[51] 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. 44

[52] 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. 45

[53] 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. 45, 46

[54] T. Kugo. *Eichtheorie*. 1997. doi: 10.1007/978-3-642-59128-0. 37

[55] J. M. Lee. *Introduction to smooth manifolds*. Shi jie tu shu chu ban gong si Beijing gong si, 2015. 130, 132, 139, 143, 147, 149

[56] 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. 14, 23

[57] D. G. Lowe. Distictive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004. 46

[58] 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. 45, 46, 47

[59] 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. 48, 50, 73

[60] S. G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. 11(7):674–693, 1989. doi: 10.1109/34.192463. 6

[61] E. L. Mansfield. *A practical guide to the invariant calculus*, volume 26. Cambridge University Press, 2010. 40

[62] B. Martinet. Détermination approchée d'un point fixe d'une application pseudo-contractante. *CR Acad. Sci. Paris*, 274(2):163–165, 1972. 13

[63] 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. 87

[64] B. Mercier. *Topics in finite element solution of elliptic problems*. Springer, 1979. 14

[65] C. W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation*. Macmillan, 1973. 129

[66] J.-J. Moreau. Fonctions convexes duales et points proximaux dans un espace hilbertien. *CR Acad. Sci. Paris Ser. A Math.*, 255:2897–2899, 1962. 6, 8, 11

[67] 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, 42

[68] 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

[69] 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

[70] 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. 49

[71] D. E. Neuenschwander. *Emmy Noether's wonderful theorem*. JHU Press, 2017. 5

[72] 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

[73] E. Noether. Invariante variationsprobleme. *Nachrichten Der Koeniglichen Gesellschaft Der Wissenschaften Zu Goettingen*, Mathematische-Physikalische Klasse:235–257, 1918. 5, 37, 61, 125

[74] E. Noether. Invariant variation problems. *Transport Theory and Statistical Problem*, 1 (3):183–207, 1971. 5, 37, 125

[75] S. Osher and L. I. Rudin. Feature-oriented image enhancement using shock filters. *SIAM Journal on Numerical Analysis*, 27(4):919–940, 1990. 41

[76] 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. 47

[77] N. Parikh, S. Boyd, et al. Proximal algorithms. *Foundations and Trends® in Optimization*, 1(3):127–239, 2014. 14

[78] 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. 14, 23

[79] M. E. Peskin and D. V. Schroeder. *An Introduction to Quantum Field Theory (Frontiers in Physics)*. Perseus Books, 1995. 37, 40, 129

[80] 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, 9, 13, 14, 23, 24, 118

[81] 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. 45

[82] 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. 9

[83] 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. 45

[84] F. Riesz. *Sur une espèce de géométrie analytique des systèmes de fonctions sommables*. Gauthier-Villars, 1907. 8

[85] A. Roche, G. Malandain, N. Ayache, and X. Pennec. Multimodal image registration by maximization of the correlation ratio. 1998. 48, 49

[86] 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 Interventation*, pages 1115–1124. Springer, 1998. 49, 73

[87] A. Roche, G. Malandain, and N. Ayache. *Unifying maximum likelihood approaches in medical image registration*. PhD thesis, Inria, 1999. 48, 49, 73

[88] R. T. Rockafellar. *Convex Analysis*. 1970. 5, 17, 18, 99

[89] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM journal on control and optimization*, 14(5):877–898, 1976. 8

[90] R. T. Rockafellar. *Convex analysis*. Princeton university press, 2015. 7, 8, 16, 17

[91] C. Rovelli. *Quantum gravity*. Cambridge university press, 2007. 129

[92] L. I. Rudin. Images, numerical analysis of singularities and shock filters. 1987. 41

[93] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992. 9, 28, 41

[94] H. Rue and L. Held. *Gaussian Markov random fields: theory and applications*. CRC Press, 2005. 53

[95] 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, 45

[96] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. Variational methods in imaging, volume 167 of applied mathematical sciences, 2009. 8, 41, 44

[97] I. Singer. A fenchel-rockafellar type duality theorem for maximization. *Bulletin of the Australian Mathematical Society*, 20(2):193–198, 1979. 17

[98] 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. 45

[99] 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. 45

[100] C. Spiessberger, A. Gleiter, and B. G. Data fusion of lockin-thermography phase images. *9th International Conference on Quantitative InfraRed Thermography*, 2008. 87

[101] 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, 45

[102] B. F. Svaiter. On weak convergence of the douglas–rachford method. *SIAM Journal on Control and Optimization*, 49(1):280–287, 2011. 14, 23

[103] 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. 45

[104] 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. 87

[105] G. H. Valadez. *Variational Methods for Multimodal Image Matching*. PhD thesis, École Doctorale Sciences et Technologies de l'Information et de la Communication, 2002. 46, 48, 49

[106] 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. 48

[107] 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, 42

[108] L. Wasserman. *All of Statistics, A Concise Course inStatistical Inference*. Springer, 2004. 90

[109] 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. 41, 45

[110] J. Weickert. Theoretical foundations of anisotropic diffusion in image processing. *Computing, Suppl.*, 11:221–236, 1996. 29

[111] J. Weickert. *Anisotropic diffusion in image processing*, volume 1. Teubner Stuttgart, 1998. 29

[112] D. Wu and G. Busse. Lock-in thermography for nondestructive evaluation of materials. *Revue générale de thermique*, 37(8):693–703, 1998. 87

[113] 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. 45

[114] 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. 11

[115] 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. 47

[116] 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. 46

[117] Y. Zhang. Spatial resolution enhancement for hyperspectral image based on wavelet baysian fusion. *4th International Congress on Image and Signal Processing*, 2011. 54, 159

[118] 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