UNIVERSITÄT ZU LÜBECK
INSTITUTE OF MATHEMATICS AND
IMAGE COMPUTING

Iterative Bregman Regularization For Lifted Variational Problems

Iterative Bregman-Regularisierung für geliftete Variationsprobleme

Masterarbeit

im Rahmen des Studiengangs
Mathematik in Medizin und Lebenswissenschaften
der Universität zu Lübeck

Vorgelegt von
Danielle Bednarski

Ausgegeben und betreut von
Prof. Dr. rer. nat. Jan Lellmann
Institute of Mathematics and Image Computing

2. Oktober 2019

Eidesstattliche Erklärung

Ich versichere an Eides statt, die vorliegende Arbeit selbstständig und nur unter Benutzung der angegebenen Quellen und Hilfsmittel angefertigt zu haben.

Lübeck,
October 2, 2019

Autor

Kurzfassung

Variationsmodelle sind ein wichtiges Werkzeug in der modernen Bildverarbeitung. Sie finden in einer Vielzahl von Anwendungsgebieten Verwendung, wie zum Beispiel dem Entrauschen, der Segmentierung und der optischen Tiefenschätzung. Um bessere – idealerweise global optimale – Minimierer nichtkonvexer Probleme zu finden, wurden Liftingmethoden entwickelt. Diese approximieren das ursprüngliche Problem mit einem höherdimensionalen, dafür aber konvexen Modell. In dieser Masterarbeit wird vorgeschlagen, die iterative Bregman-Regularisierung auf geliftete Varia-tionsprobleme anzuwenden. Insbesondere wird der Zusammenhang der iterativen Bregman-Regularisierung auf dem ursprünglichen und dem gelifteten Problem bei einer sublabelgenauen Diskretisierung untersucht. Am Ende werden erste experimentelle Ergebnisse bezüglich zweier Variationsmodelle vorgestellt: einem konvexen Modell zum Entrauschen und einem nicht konvexen Modell zur optischen Tiefenschätzung.

Abstract

Variational models are an important tool in modern image processing. They can be used for a variety of applications such as denoising, segmentation, and depth estimation. In order to find better – ideally globally optimal – minimizers of non-convex models, functional lifting strategies have been developed. They approximate the original problem using a larger but convex model. In this thesis, we propose to combine the lifting approach with the iterative Bregman regularization. In particular, we investigate the relation between the Bregman iteration on the original and on the lifted problem using a sublabel-accurate discretization. Finally, we show experimental results for the convex denoising and non-convex stereo matching model.

Contents

Notation

Abbreviations

ISS	inverse scale space
ROF	Rudin-Osher-Fatemi (Def. 2.1)
TV	total variation (Def. 3.1)

Mathematical Symbols

$\overline{\mathbb{R}}$	$\mathbb{R} \cup \{-\infty, \infty\}$
$\operatorname{con} C$	convex hull of set C (Def. 3.10)
$\operatorname{con} f$	convex hull of function f (Def. 3.10)
f^*	convex conjugate of function f (Def. 3.11)
f^{**}	biconjugate of function f (Def. 3.11)
$\hat{\partial} f(x)$	regular subdifferential of function f (Def. 3.12)
$\partial f(x)$	(general) subdifferential of function f (Def. 3.12)

Spaces and Norms (Ch. 3.1)

$BV(\Omega, \Gamma)$	functions of bounded variation (L^1-integrable and finite TV norm)
$\mathcal{H}$	Hausdorff metric
L^p	Lebesgue spaces (functions with finite $\|\| \cdot \|\|_{L^2}$ norm)
W_p^k	Sobolev spaces (functions in L^p whose weak partial derivates up to order k exist and are in L^p)

Variational Problems

f	input image
u	output image
Ω	image domain
Γ	image range
$G(f, u)$	data term
$R(u)$	regularization term

Sublabel-Accurate Lifting

$k+1$	number of labels
$u(\cdot) = \gamma_i^\alpha$	one-dimensional image value (Eq. (66))
$\mathbf{u}(\cdot) = \mathbf{1}_i^\alpha$	k-dimensional lifted image value (Eq. (67))
$L : \Gamma \to Q'$	lifting function for image value, $L(\gamma_i^\alpha) = \mathbf{1}_i^\alpha$ (Eq. (86))
$\mathbf{G}(f, \mathbf{u})$	lifted data term
$g(x, u(x))$	integrand of data term
$\mathbf{g}^{**}(x, \mathbf{u}(x))$	integrand of lifted data term (Eq. (72))
$\mathbf{TV}(\mathbf{u})$	lifted TV regularization term
$\mathbf{\Phi}^{**}(x, \mathbf{u}(x))$	integrand of lifted TV regularization term (Eq. (77))

Introduction

Motivation

In modern image processing tasks, *variational problems* play an important role. They are a certain type of optimization problem, where a functional is to be minimized over a solution space consisting of functions. They are well-suited to impose constraints such as smoothness on the solution. A typical form for a variational minimization problem in image processing is

$$u^* \in \arg\min_{u \in U} \{G(f,u) + \lambda R(u)\}, \tag{1}$$

where U is a suitable solution space for the special task given, e.g. a suitable space of images. In a mathematical context, images can be described as functions $u : \Omega \to \Gamma$, where $\Omega \subset \mathbb{R}^d$ is the *image domain*, and $\Gamma \subset \mathbb{R}^k$ the *image range* or *label space*. Typically, the image domain is two-dimensional and the image range one-dimensional for gray-scale images or three-dimensional for RGB images. $G(f,u)$ is called the *data term* and fits the output u to the given input data f. $R(u)$ is called the *regularization term* and inflicts certain constraints on the solution u. The weighting parameter $\lambda \in \mathbb{R}^+$ is used in order to give more weight to one term or the other. Together, $G(f,u) + \lambda R(u)$ is called an *energy term* or *objective function*.

The scope for variational problems in image processing is diverse (AK06, SGG$^+$09). Depending on the choice of the data and regularization term, variational problems can be used for tasks such as denoising (ROF92, BCM05), segmentation (CV01) and depth estimation (SCD$^+$06, ZGFN08). Here, we consider the *Rudin-Osher-Fatemi* model (ROF) as an example.

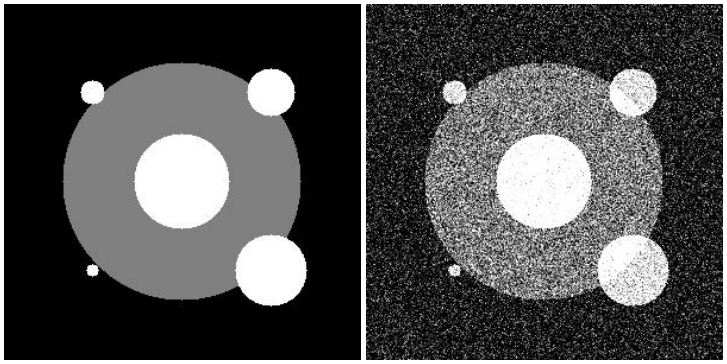
Definition 2.1 (*Rudin-Osher-Fatemi (ROF92)*)

Given an image $f \in L^2(\Omega, \mathbb{R}^k)$, we assume that it can be decomposed into a noise-free image u and normally distributed noise v, such that $f = u + v$ holds. With weighting parameter $\lambda \in \mathbb{R}^+$, the *Rudin-Osher-Fatemi* (ROF) model is given by

$$\inf_{u \in \mathrm{BV}(\Omega,\mathbb{R}^k)} \frac{1}{2} \int_\Omega ||u(x) - f(x)||_2^2 \mathrm{d}x + \lambda \,\mathrm{TV}(u). \tag{2}$$

Here, $\mathrm{BV}(\Omega, \mathbb{R}^k)$ denotes the space of function with bounded variation (Def. 3.3) and $\mathrm{TV}(u)$ denotes the total variation of u (Def. 3.1).

The ROF model plays an important role in image denoising. Its popularity is due to the convexity of the energy term and the models ability to remove noise while still preserving edges. Slight drawbacks of the model are its inclination to produce cartoonish images, staircasing effects, and possible loss of contrast (see Ex. 4.1). An exemplary result for the ROF model with different weighting parameters λ can be seen in Fig. 1. Since the ROF model is an established model, it is often used as a prototypical problem when developing new methods or algorithms (e.g. (MLM$^+$15, Bre67)).



(a) Input image (b) Input image with noise

(c) Solution of ROF, $\lambda = 0.5$ (d) Solution of ROF, $\lambda = 5$ (e) Solution of ROF, $\lambda = 20$

Figure 1: **Rudin-Osher-Fatemi.** Top row: Input image and input image with artificially added noise. Bottom row: Solution of the ROF model with different weighting parameters λ. The choice of the weighting parameter λ strongly affects the solution. If λ is chosen too small, not all of the noise is removed. If λ is chosen too large, structural detail of the original image is lost. The results were computed using CVX (GB14, GB08).

Bregman Iteration. As we have seen in Fig. 1, the weighting parameter λ plays an important role in the ROF model and has a large influence on the solution. Instead of having to solve the ROF model for different weighting parameters and selecting the best one based on the solution, Osher et al. proposed an iterative regularization method (OBG$^+$05), which is based on the Bregman distance (Bre67) with respect to the regularization term:

$$u_l^* \in \arg\min_{u \in \mathrm{BV}(\Omega,\mathbb{R}^k)} \left\{ \frac{1}{2} \int_\Omega ||u(x) - f(x)||_2^2 \mathrm{d}x + \lambda \,\mathrm{TV}(u) - \langle p_{l-1}, u \rangle \right\}, \tag{3}$$

where $p_0 = 0$ and $p_l \in \partial \lambda \,\mathrm{TV}(u_{l-1})$ for $l \geq 1$. We refer to Def. 3.13 and Sec. 4.1 for mathematical details on the Bregman distance and iteration.

The Bregman iteration effectively steps through a range of regularization parameters. Using a suitable stopping criterion such as the discrepancy principle (OBG$^+$05), this provides an intuitive way for selecting the regularization strength. The continuous limit of the Bregman iteration leads to the *inverse scale space flow* (BGM$^+$16). The basic idea behind scale spaces is that an input image f is put together by details of different scale with respect to the regularizer R. The inverse scale space flow starts with the mean of the input image and later on it recovers more and more detail of the input image. Structures appearing early in the iteration correspond to nonlinear eigenfunctions of the regularizer R with small eigenvalues and are called large-scale details. Details appearing later correspond to eigenfunctions of the regularizer R with large eigenvalues and are called small-scale details (Sec. 4.1). In Fig. 2 we show some exemplary results of the Bregman iteration on an artificial image, and in Fig. 3 a real life example on the application of scale spaces.

Functional Lifting. Another potential difficulty when working with variational problems is the possibility of non-convex energy terms. Many solvers work with gradient-descent methods and, therefore, they might get stuck in a local optimum. Let us consider *stereo matching* as an example.

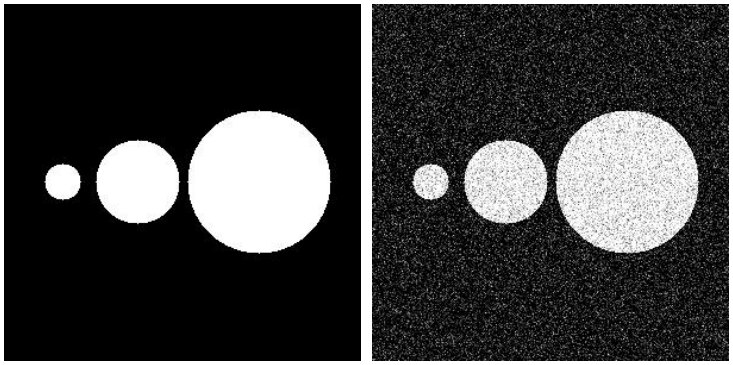
Stereo matching is an established approach used to estimate depth. The model expects two or more input images, showing the same scene from slightly different viewpoints. The goal is to estimate a depth map $u : \Omega \to \mathbb{R}$ of the scene (see Fig. 4). Usually the images are rectified in a preprocessing step, so that the epipolar lines in the images align. We focus on formulating the variational problem and refer to (SHK$^+$14) for more detail on the preprocessing step.

There are multiple formulations for the stereo matching problem. The basic concept is as follows:

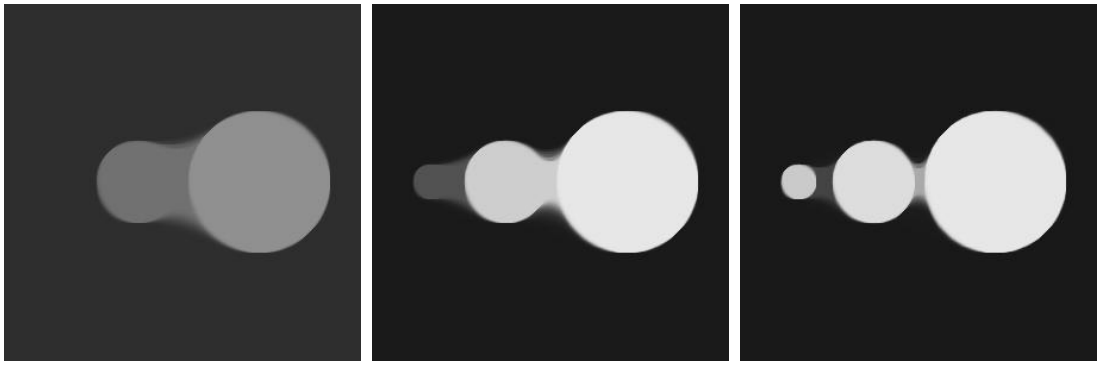
Definition 2.2 (*Stereo Matching*)

Given two rectified input images $f_1, f_2 \in L^2(\Omega, \mathbb{R}^k)$ and a parameter $\lambda \in \mathbb{R}^+$, the *stereo matching* model is given as

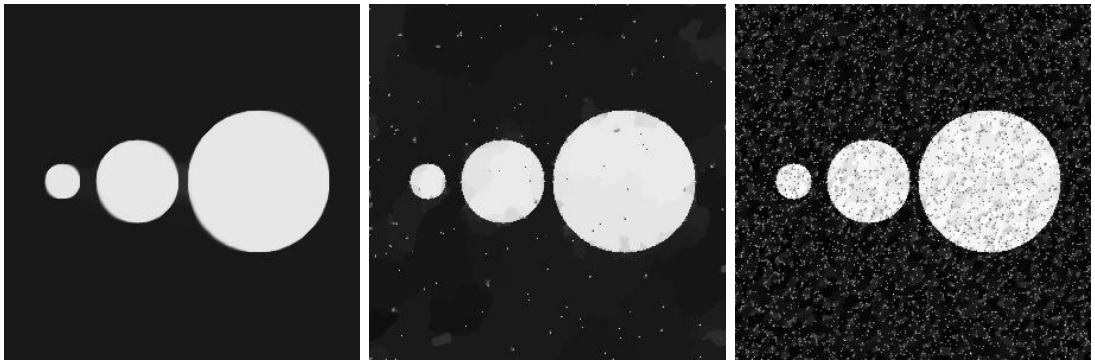
$$\inf_{u \in BV(\Omega,\mathbb{R}^k)} \frac{1}{2} \int_\Omega |f_1(x) - f_2(x + u(x))| \mathrm{d}x + \lambda TV(u(x)). \tag{4}$$



(a) Input image (b) Input image with noise



(c) Bregman iteration, l=1 (d) Bregman iteration, l=2 (e) Bregman iteration, l=3



(f) Bregman iteration, l=4 (g) Bregman iteration, l=50 (h) Bregman iteration, l=100

Figure 2: **Bregman iteration on ROF model.** First row: Input image and input image with artificially added noise. Second and third row: A selection of solutions of the Bregman iteration on the ROF model. Each iteration was solved using CVX (GB14, GB08). We did not specify a stopping criterion for the iteration in order to make the connection to the inverse scale space flow clear. In the beginning, only the two biggest white circles appear. As the iteration proceeds, the third smaller circle and finally the noise is recovered. Applying the usual discrepancy principle as a stopping criterion would have terminated the iteration after the fourth step.

Figure 3: **Scale Spaces.** Real-life example on the application of scale spaces. Left: input image showing a woman. Right, output image after some scales with respect to the total variation have been removed. In particular, scales associated with wrinkles have been removed in order to reverse the effect of aging. Images from (BGM+16).

In the data term of the above definition, pixels of the second input image are moved over the first input image and a penalty term is calculated depending on the offset/depth $u(x)$. The aim of the data term is to find an offset $u(x)$ for each pixel, such that the value of the shifted pixels in the second image matches the value in the first image. The total variation is used as regularizer in order to ensure that the shifting is sufficiently regular.

The data term of the stereo matching problem is in general non-convex. This raises the question whether the problem can be reformulated as to obtain an equivalent convex formulation. This would allow to find a global solution. Another advantage would be the possibility to use standard convex solvers, such as the PDHG (CP11) or ADMM (Hes69, EB92) algorithm. These solvers are efficient on large-scale problems.

Consider a variational problem of the form (1), where R is convex but G is not. In (PCBC10) a functional lifting method was introduced, which uses the indicator function of the subgraph of a function $u \in U$ in order to represent u in a higher-dimensional space. Using this reformulation, a new energy term can be formulated on the lifted function space, which is altogether convex. The authors show that projecting the minimizer of the lifted problem onto the original function space results in a minimizer of the unlifted problem. Further details on this method can be found in Ch. 4.2.2.

In this thesis we want to consider the possibility of combining the two powerful methods, namely the iterative Bregman regularization and the functional lifting. We will primarily concentrate on convex variational problems and examine the question, whether the solution of the Bregman iteration on an unlifted convex problem is equivalent to the one on the according to Ch. 4.2.3 lifted problem.

Figure 4: **Stereo Matching.** The two images on the left are the input images. They were taken from slightly different viewpoints and have already been rectified. Right image: Depth map as found by minimizing the stereo matching problem. Warm colors indicate regions that are close to the camera position, cold colors indicate regions that are further away. Source of input images: (SHK+14). The picture on the right was calculated for this thesis using Matlab and the libraries `prost` and `sublabel_relax` (MLM+15).

Related Work

Bregman Iteration. The Bregman iteration was first introduced by Osher et al. in (OBG+05) as an extension for the ROF model. Instead of the total variation, the Bregman iteration uses the Bregman distance (Bre67) associated with the total variation as regularizer and iteratively approaches the denoised image. The primary goal was to improve the result of ROF image denoising. In (BGO+06) and (BGM+16) a connection to the inverse scale space flow was pointed out. In (YOGD08) the Bregman iteration was extended to the l_1 regularizer. A linearized extension was investigated in (OMDY11) and (COS09). Based on the linearized Bregman iteration the very effective split Bregman algorithm was derived in (GO09).

Functional Lifting. Functional lifting strategies originated in discrete labeling problems. In a discrete setting, the labeling problem can be described by a graph using the theory of Markov Random Fields (KL80). The minimization problem is then given by a penalty term for the edges and a penalty term for the nodes. Ishikawa and Geiger proposed a lifting approach (Ish03a, IG98), which allows – under certain conditions – to reformulate the multi-label problem as a higher dimensional graph and solve it globally.

Variational models can be considered as labeling problems in a spatially continuous framework. Here, too, functional lifting methods can be applied in order to find a higher-dimensional, convex representation of a non-convex problem. This allows to postpone the discretization until implementation. First lifting approaches were made in binary image segmentation for the well-known Chan-Vese model (CV01):

$$\arg \min_{C \subset \Omega, c_1 \in \Gamma, c_2 \in \Gamma} \left\{ \int_C (f(x) - c_1)^2 \mathrm{d}x + \int_{\Omega \setminus C} (f(x) - c_2)^2 \mathrm{d}x + \lambda \mathcal{H}^{d-1}(C) \right\}. \tag{5}$$

Here, the input image f is to be separated into two regions C and $\Omega \backslash C$. Variable c_1 describes the mean value of f on C, c_2 the mean value of f on $\Omega \backslash C$ and $\mathcal{H}^{d-1}$ the $(d-1)$–dimensional Hausdorff measure, or – more precisely – the perimeter of set C. Let us assume that c_1 and c_2 are known and fixed. The authors suggested representing set C by the zero level set of a Lipschitz function. Application of the Heaviside function then leads to a representation of set C by its characteristic function:

$$\arg \min_{u:\Omega \to \{0,1\}} \int_\Omega (f(x) - c_1)^2 * u(x) + (f(x) - c_2)^2 * u(x) \mathrm{d}x + \lambda \, \mathrm{TV}(u). \tag{6}$$

In (CEN06) a relaxation method for this problem was proposed. The solution space was extended to its convex hull:

$$\arg \min_{u:\Omega \to [0,1]} \int_\Omega (f(x) - c_1)^2 * u(x) + (f(x) - c_2)^2 * u(x) \mathrm{d}x + \lambda \, \mathrm{TV}(u). \tag{7}$$

The authors showed that for any solution u^* of Eq.(7), thresholding $\tilde{u}^* = 1_{u^* > \mu}$ results in a minimizer $\tilde{u}^*$ of Eq. (6) for almost any $\mu \in [0, 1]$.

A couple of years ago, several lifting approaches for continuous multi-label segmentation problems emerged at about the same time (CCP12, ZGFN08, LKY$^+$09, LLWS13). In this thesis we concentrate on a continuous lifting method, which is based on the discrete lifting method by Ishikawa (Ish03a, IG98) and the continuous lifting method for the Chan-Vese model (CEN06). For an introduction to this method with respect to the total variation regularizer see (PSG$^+$08). An extension to a more general problem class can be found in (PCBC10). A formulation that allows for sublabel-accurate discretization is given in (MLM$^+$15).

Primal-Dual Hybrid Gradient Algorithm. We will use the primal-dual hybrid gradient (PDHG) algorithm to solve primal-dual problems of the form

$$\min_{u \in \mathbb{R}^n} \max_{v \in \mathbb{R}^m} \Phi(u, v), \quad \Phi(u, v) := g(u) + \langle Ku, v \rangle - f^*(v), \tag{8}$$

where K is a linear operator, $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $f^* : \mathbb{R}^m \to \overline{\mathbb{R}}$ are convex functionals.

The algorithm was first introduced by Zhu and Chan (ZC08) in order to solve total variation minimization problems in image processing (see Def.2.1). In (EZC10), the algorithm was generalized to a broader class of convex optimization problems, and in (PCBC09) it was applied to the Mumford-Shah functional. After choosing some starting values for the primal variable u and the dual variable v, the algorithm alternately performs a gradient-ascent step on the maximization problem $\max_{v \in \mathbb{R}^m} \Phi(u, v)$ and a gradient-descent step on the minimization problem $\min_{u \in \mathbb{R}^n} \Phi(u, v)$. In order to improve the performance of the algorithm and to guarantee convergence, Pock and Chambolle incorporated diagonal preconditioning into the algorithm (PC11). Since the choice of the step size for the gradient-ascent and -descent step proved to be crucial for the speed of convergence, the use of adaptive step schemes was suggested in (GLY$^+$13).

Overview

This thesis is structured as follows: In **Chapter 3** we introduce the required mathematical concepts. In particular, we discuss possible solution spaces for variational problems, and important properties used in convex analysis to describe objective functions. In **Chapter 4** we introduce existing methods for solving variational problems. The first method is the iterative Bregman regularization, which improves the result of convex variational problems and leads (in the continuous limit) to the inverse scale space flow. The second method is functional lifting, which is used on non-convex, spatially continuous objective functions in order to find a higher-dimensional convex formulation for the problem. We discuss two related approaches, one formulated on a continuous and the other on a discretized label space. The latter is also known as sublabel-accurate. In **Chapter 5** we discuss the combination of the iterative Bregman regularization and the sublabel-accurate functional lifting. In particular, we examine whether the solution of the Bregman iteration on an unlifted convex problem is equivalent to the one on the sublabel-accurate lifted problem. In **Chapter 6** we present details on the implementation and first numerical results achieved with this new combined method. **Chapter 7** concludes this thesis with a summary, outlook to future research and open questions concerning this topic.

Mathematical Preliminaries

Variational problems play an important role in modern image processing and can be used for many different tasks, e.g. image denoising or depth estimation. The general aim is to solve a problem of the form

$$u^* \in \arg\min_{u \in U} \{F(u, f)\}, \tag{9}$$

where F is the objective function, f the given input data and U an appropriately chosen solution space. The former strongly depend on the given task. Considering image denoising as an example, f would be a noisy input image and u^* the denoised output image. As for a depth estimation problem we would consider two input images f_1 and f_2 of the same scene taken from slightly different viewpoints and u^* would be a depth map of the scene. In this chapter, we introduce some basic concepts and vocabulary used in the context of variational problems. First, we start by considering the solution spaces used in this thesis. Then, we introduce some properties which help classify objective functions.

Function Spaces

By choosing a function space U in Eq. (9), we can postulate properties for the solution. Considering image denoising, the aim is to find a smooth representation of the input image, which preserves sharp edges at the same time. This means that color jumps in the output image cannot be punished stronger than smooth color gradients of the same magnitude. In this scenario the *total variation* (TV) semi-norm is of particular interest.

Definition 3.1 (*Total Variation, (AFP00, Def. 3.4) and (Lel11, Def. A.1)*)

Consider a function $u : \Omega \to \mathbb{R}^k$ with $\Omega \subset \mathbb{R}^d$ open and bounded. Furthermore, let u be $||\cdot||_{L^1}$-integrable, thus $||u||_{L^1} := \int_\Omega |u(x)| \mathrm{d}x < \infty$. The *total variation* of u is defined as

$$TV(u) := \sup_{\phi \in C_c^1(\Omega,\Gamma)} \left\{ -\int_\Omega \langle u(x), \mathrm{Div}\ \phi(x) \rangle \mathrm{d}x \,\middle|\, ||\phi(x)||_2 \leq 1 \ \ \forall x \in \Omega \right\}, \tag{10}$$

where $C_c^1(\Omega, \Gamma)$ is the space of once continuously differentiable functions $\phi : \Omega \to \Gamma$, that have a compact support on Ω. Furthermore, Div is defined as

$$\mathrm{Div}\phi(x) := (\mathrm{div}\phi_1(x), ..., \mathrm{div}\phi_k(x))^\top. \tag{11}$$

If u is differentiable in Ω, the total variation of u can be calculated as $\mathrm{TV}(u) = \int_\Omega |\nabla u| \mathrm{d}x$. Furthermore, the total variation is related to the distributional derivative. Here, we will only state the final proposition as presented in (Lel11, Prop. A.2), for more information the reader is referred to (AFP00, Prop. 3.6) and (Lel11, Ch. A.1.1).

Proposition 3.2 (*Summarized from (AFP00, Prop. 3.6) and (Lel11, Prop. A.2)*)

Consider a function $u : \Omega \to \mathbb{R}^k$ with $\Omega \subset \mathbb{R}^d$ open and bounded. Then the condition $u \in \mathrm{BV}(\Omega, \mathbb{R}^k)$ is equivalent to $\int_\Omega |u(x)|\mathrm{d}x < \infty$ and its distributional derivative corresponding to a finite Radon measure. The latter means, that $\int_\Omega |u_j(x)|\mathrm{d}x < \infty$ holds for all partial derivates u_j and there exist measures $\mathrm{D}u_j = (\mathrm{D}_1 u_j, ..., \mathrm{D}_d u_j)$ for $j = 1, ..., k$ on Borel subsets $B(\Omega) \subset \Omega$ such that

$$-\sum_{j=1}^{k} \int_\Omega u_j \mathrm{div} v^j \mathrm{d}x = \sum_{j=1}^{k} \sum_{i=1}^{d} \int_\Omega v_i^j \mathrm{d}\mathrm{D}_i u_j, \tag{12}$$

for all $v \in C_c^\infty(\Omega, \mathbb{R}^{d\times k})$.

With the notation introduced in Prop. 3.2, the following equality holds:

$$\mathrm{TV}(u) = |\mathrm{D}u|(\Omega) = \int_\Omega \mathrm{d}|\mathrm{D}u|(\Omega). \tag{13}$$

The distributional derivative $\mathrm{D}u$ can be decomposed into mutually singular measures: $\mathrm{D}^a u$ for absolutely continuous parts, $\mathrm{D}^j u$ for jump parts and the so-called Cantor part $\mathrm{D}^c u$ for the remaining parts - see (AFP00, Cor. 3.89) for more details. This theoretical result together with the polar decomposition for measures introduced in (AFP00, Cor. 1.29) is used later on for the method discussed in Ch. 4.2.2.

We now return to the main topic of this section. Let us define three function spaces, which we will use later on:

Definition 3.3 (*Function Spaces (AFP00, Def. 3.1, Def. 2.4)*)

Consider a function $u : \Omega \to \mathbb{R}^k$, with $\Omega \subset \mathbb{R}^d$ open and bounded.

- The function u is in the $L^p(\Omega, \mathbb{R}^k)$ space (sometimes called *Lebesgue space*) if it is measurable with respect to the d-dimensional Lebesgue measure and

$$\int_\Omega |u(x)|^p \mathrm{d}x < \infty. \tag{14}$$

- The function u is of *bounded variation* and an element of the space $BV(\Omega, \mathbb{R}^k)$ if it is $||\cdot||_{L^1}$- integrable and if its total variation $TV(u)$ is finite.

- If u lies in the Lebesgue space $L^p(\Omega, \mathbb{R}^k)$ and if its weak partial derivates v for the multi-index $\alpha \in \mathbb{N}_0^s$, with

$$\int_\Omega v(x)\phi(x)\mathrm{d}x = (-1)^\alpha \int_\Omega u(x) \left(\frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}}, ..., \frac{\partial^{\alpha_s}}{\partial x_s^{\alpha_s}} \right) \phi(x)\mathrm{d}x \tag{15}$$

for all test functions $\phi \in C_c^\infty(\Omega)$ exist up to order k and also lie in $L^p(\Omega, \mathbb{R}^k)$, we write $u \in W_p^k$ and say that u lies in the *Sobolev space*.

There is a weak version of the *coarea formula* for functions in $\mathrm{BV}(\Omega, \mathbb{R}^k)$, which was first proven in (FR60). Here, we will only state the final theorem, without going into detail. For further information we refer the reader to (FR60) and (AFP00).

Theorem 3.4 (*Coarea formula in* BV *(FR60), (AFP00, Thm. 3.40)*)

Consider $u \in \mathrm{BV}(\Omega, \mathbb{R}^k)$ with $\Omega \subset \mathbb{R}^d$. The set $\{u > t\}$ has for L^1-a.e. $t \in \mathbb{R}$ finite perimeter in Ω and for any Borel set $B \subset \Omega$ we have

$$|\mathrm{D}u|(B) = \int_{-\infty}^{+\infty} |\mathrm{D}1_{u>t}|(B)\mathrm{d}t, \tag{16}$$

for the characteristic function of the subgraph of u defined as

$$1_{u>t}(x) := \begin{cases} 1, & \text{if } u(x) > t, \\ 0, & \text{else.} \end{cases} \tag{17}$$

Convex Analysis

In this section we consider the objective function of variational problems. Other than in Eq. (9) we denote them by lowercase f. We begin by discussing some properties which guarantee existence of a minimizer. Then, we introduce the concepts of *convexity*, *convex hulls* and *convex conjugates*. Afterwards, we define *subdifferentials* and the *Bregman distance*. In the following, let the extended real line be defined as $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$.

Existence of Minimizers

Objective functions that become negative infinity for at least one value in the image domain, or that become positive infinity on the whole image domain result in trivial optimization problems. Therefore, we define:

Definition 3.5 (*Properness (RW09)*)

A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is called *proper*, if $f(x) > -\infty$ for all $x \in \mathbb{R}^n$ and if there is at least one $y \in \mathbb{R}^n$ for which f is finite.

Let us, furthermore, consider this weak version of continuity:

Definition 3.6 (*Lower Semi-continuity (RW09, Def. 1.5)*)

A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is called *lower semi-continuous*, if

$$\liminf_{y \to x} f(y) = \lim_{\varepsilon \to 0^+} \left(\inf_{y \in B(x,\varepsilon)} f(y) \right) = f(x) \tag{18}$$

holds for all $x \in \mathbb{R}^n$. Here, $B(x, \varepsilon)$ denotes a ball around x with radius ε.

Using the above definitions, we can formulate requirements for the existence of a minimizer for the energy term f.

Corollary 3.7 (*Existence of a Minimum (RW09, Cor. 1.10)*)

Let $f : \Omega \to \overline{\mathbb{R}}$ be proper and lower semi-continuous for $\Omega \subset \mathbb{R}^n$ compact and bounded. The function f then attains its minimum on Ω.

Cor. 3.7 not only guarantees the existence of a minimum, but implicitly also the existence of at least one minimizer $x \in \Omega$.

Convexity

Another important property used to describe energy terms is convexity. Solvers cannot always distinguish between local and global minimizers. As we see in Thm. (3.9), this is not a problem when working with convex functions.

Definition 3.8 (*Convexity (RW09, Def. 2.1)*)

A *set* $\Omega \in \mathbb{R}^n$ is called *convex*, if

$$(1-\lambda)x_1 + \lambda x_2 \in \Omega \tag{19}$$

holds for any choice of $x_1, x_2 \in \Omega$ and $\lambda \in (0,1)$. A *function* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ on a convex set $\mathbb{R}^n$ is called *convex*, if the inequality

$$f((1-\lambda)x_1 + \lambda x_2) \leq (1-\lambda)f(x_1) + \lambda f(x_2) \tag{20}$$

is fulfilled for any choice of $x_1, x_2 \in \Omega$ and $\lambda \in (0,1)$.

An immediate result of the second part of Def. 3.8 is Jensen's inequality (RW09, Thm. 2.2). It extends the inequality to convex combinations $x = \sum_{i=1}^{n} \lambda_i x_i$ for nonnegative $\lambda_1, ..., \lambda_n$ summing up to 1 and $x_1, ..., x_n \in \Omega$ in the following way:

$$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leq \sum_{i=1}^{n} \lambda_i f(x_i). \tag{21}$$

For convex functions, the following theorem on minimizers holds:

Theorem 3.9 (*Minimizers of convex functions (RW09, Thm. 2.6)*)

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex. Then every local minimizer is also a global minimizer. Furthermore, the set of all minimizers $\{x \in \mathbb{R}^n | f(x) \leq f(y) \text{ for all } y \in \mathbb{R}^n\}$ is also convex.

Working with a proper, lower semi-continuous and convex energy term on a compact and bounded image domain, the existence of a minimizer is guaranteed according to Cor. 3.7, which is according to Thm. 3.9 not only locally but also globally optimal.

Later on, we will work with proper and lower semi-continuous, yet possibly non-convex energy terms. In order to be able to use certain solvers we can approximate a non-convex function by a convex one. For the approximation we need the concept of convex hulls.

Definition 3.10 (*Convex hull (RW09, p. 53ff)*)

The convex hull of a set $\Omega \subset \mathbb{R}^n$ is defined as the smallest convex set including Ω. The convex hull g of a function f on a convex set Ω is the largest convex function for which $g(x) \leq f(x)$ holds for every $x \in \Omega$.

The convex hull g of a function f still contains the information on the global minimum of f, although information on local minima and maxima is lost. Calculating the minimum of g provides the minimum of f, but with the addition that we can use convex solvers. Attention should be paid though, since we are not guaranteed to find the actual global minimizer of f. Consider a two-dimensional function with the shape of the letter Λ. The convex hull of this function does not contain the information of the local maximum, in fact the convex hull is a line with slope zero and every point of the domain could be returned as a global minimizer.

Convex hulls can be obtained by different means. In (RW09, Thm. 2.27 and Prop. 2.31), convex combinations are used to determine the convex hull of a set or function respectively. For a non-convex set $C \in \mathbb{R}^n$ the convex hull consists of all points satisfying

$$\Omega = \text{con } C := \left\{ \sum_{i=1}^{p} \lambda_i x_i \,\middle|\, \lambda_1, ..., \lambda_p \geq 0, \sum_{i=1}^{p} \lambda_i = 1, x_i \in C, p \geq 0 \right\}. \tag{22}$$

Equivalently, the convex hull g of a function f on a convex set $\Omega \subset \mathbb{R}^n$ can be determined as

$$g(x) = (\text{con} f)(x) := \inf \left\{ \sum_{i=1}^{n} \lambda_i f(x_i) \,\middle|\, \lambda_1, ..., \lambda_n \geq 0, \sum_{i=1}^{n} \lambda_i = 1, x_i \in \Omega, \sum_{i=1}^{n} \lambda_i x_i = x \right\}. \tag{23}$$

Alternatively, the convex hull of a function can be determined with the *Legendre-Fenchel transform*, also called convex conjugate.

Definition 3.11 (*Convex Conjugate (RW09, p. 473)*)

Given a function $f : \Omega \to \overline{\mathbb{R}}$ on a real topological vector space Ω with dual space Ω^*, the *conjugate* $f^* : \Omega^* \to \overline{\mathbb{R}}$ is defined as

$$f^*(x^*) := \sup_{x \in \Omega} \{\langle x^*, x \rangle - f(x)\}. \tag{24}$$

The conjugate of f^* is called the *biconjugate* f^{**} and is equal to the convex hull of f. We can get an intuition for this by considering a hyperplane lying below a graph f. The following explanation is visualized in Fig. 5. Such a hyperplane is defined by slope m and intercept b, and needs to satisfy

$$f(x) \geq \langle m, x \rangle + b \tag{25}$$

for all $x \in \Omega$. For a fixed slope m, the intercept b can be computed as

$$-b \geq \sup_{x \in \Omega} \{\langle m, x \rangle - f(x)\} = f^*(m). \tag{26}$$

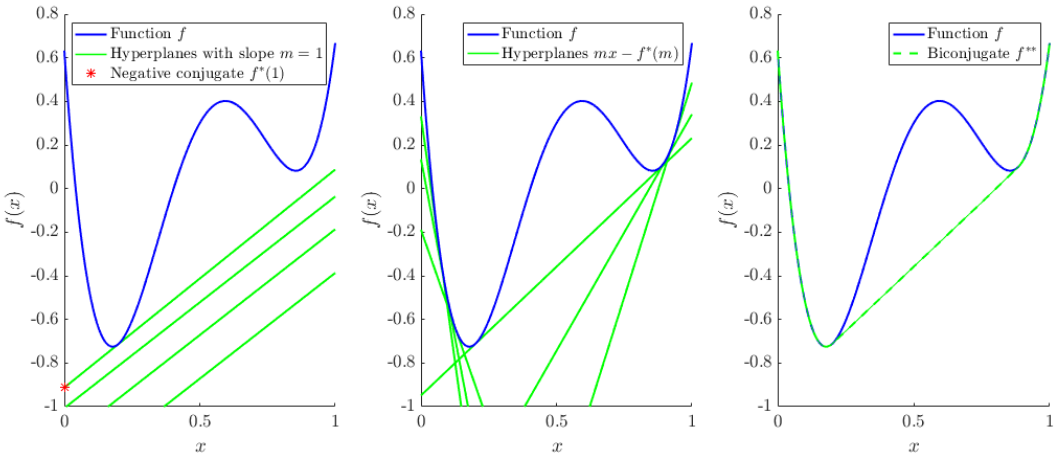


Figure 5: **Convex Conjugates.** Dark blue: Non-convex function f. Left: Selection of hyperplanes with slope $m = 1$ which lie below the function f. The red star marks the greatest intercept for which the slope still lies beneath f and which is the negative conjugate evaluated at $m = 1$. Middle: Selection of hyperplanes $y_i = \langle m_i, x\rangle - f^*(m_i)$. Right: Supremum over all hyperplanes $y_i = \langle m_i, x\rangle - f^*(m_i)$, which also is the biconjugate f^{**}.

The closest approximation of Eq. (26) is given by $b = -f^*(m)$. Therefore, the convex conjugate $f^*(m)$ is the negative intercept of the closest hyperplane with slope m which lies below the graph. The biconjugate is the pointwise supremum over all hyperplanes which lie below f and, therefore, it provides the convex envelope of f by

$$f^{**}(x) = \sup_{m\in\Omega^*} \left\{\langle m, x\rangle - f^*(m)\right\}. \tag{27}$$

According to (RW09, Thm. 11.1), both the conjugate and biconjugate of a function with a proper convex hull are also proper, lower semi-continuous and convex. Furthermore, the biconjugate is equivalent to

$$f^{**}(x) = \mathrm{cl}(\ \mathrm{con}\ f) = \mathrm{cl}\left(\sup_{\substack{g \text{ convex},\\ g\leq f}} g\right), \tag{28}$$

with cl denoting the closure. The closure of a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is its lower semi-continuous hull: $\mathrm{cl}\ f(x) = \liminf_{y\to x} f(y)$.

Subdifferential

Later on, we will work with objective functions that are convex, usually continuous, but generally non-smooth. Therefore, we use a more general definition of differentiability, which allows us to determine *subgradients* at points in the domain of the function where differentiability is not given.

Definition 3.12 (*Subdifferential (RW09, Def. 8.8)*)

The *regular subdifferential* $\hat{\partial} f(x)$ of a convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ at a finite point $x \in \mathbb{R}^n$ is defined as

$$\hat{\partial} f(x) := \{p \in \mathbb{R}^n |\ \ f(y) \geq f(x) + \langle p, y - x \rangle,\ \ \forall y \in \mathbb{R}^n\} \tag{29}$$

and p is called a *regular subgradient*. The *(general) subdifferential* $\partial f(x)$ at a finite point $x \in \mathbb{R}^n$ is defined as

$$\partial f(x) := \left\{ q \in \mathbb{R}^n \middle|\ \exists \left\{x^k\right\} \subset \mathbb{R}^n, \left\{p^k\right\} \subset \hat{\partial} f(x^k) : x^k \to x, f(x^k) \to f(x) \text{ and } p^k \to q \right\} \tag{30}$$

and q is called a *(general) subgradient*.

Note the equivalence of the regular and general subdifferential for any proper and convex function (RW09, Prop. 8.12). According to (RW09, Cor. 10.9), the subdifferential of the sum of proper and lower semi-continuous functionals $f_1, ..., f_m$ with $f_i : \mathbb{R}^n \to \mathbb{R}$ includes the sum of subdifferentials of the functionals:

$$\partial f(x) \supset \partial f_1(x) + ... + \partial f_m(x). \tag{31}$$

If the functionals $f_1, ..., f_m$ are convex and their domains cannot be separated,

$$\partial f(x) \subset \partial f_1(x) + ... + \partial f_m(x) \tag{32}$$

holds. Furthermore, if each f_i is regular at x, their sum is also regular and equality in Eq. (31) holds.

The *Bregman distance* was first introduced in (Bre67), where it was used for solving problems in convex programming. In (OBG+05) it was reinterpreted as an iterative regularization term and in that context we will use it in chapter 4.1.

Definition 3.13 (*Bregman Distance*)

Let $f : \Gamma \to \mathbb{R}$ be a convex and nonnegative functional and let p be a subgradient of f evaluated at v. The *Bregman distance* is defined as

$$B_f^p(u, v) := f(u) - f(v) - \langle p, u - v \rangle. \tag{33}$$

Note that the Bregman distance is only unique as long as f is differentiable at v, otherwise the choice of an arbitrary subgradient $p \in \partial f(v)$ can affect the distance. The Bregman distance is no metric in the actual sense, since symmetry and the triangle inequality usually are not fulfilled. However, the Bregman distance is nonnegative for any v and zero for $v = u$. If f is strictly convex, this property extends to $B_f^p(u, v)$ with respect to u.

Methods

Using variational models in image processing poses some challenges. First of all, one has to choose a good model and suitable solution space for the given task. The choice of the weighting parameter can heavily influence the quality of the solution, as we have seen in Fig. 1. Not all variational models, i.e., stereo matching (Def. 2.2), are convex and, therefore, one might not be able to find a global optimal solution, even if it exists. Also, the choice of the solver influences the quality of the result and the duration of the computation time needed.

In this chapter, we introduce two methods used in the context of variational models. First, we consider the iterative Bregman regularization, which obviates the search for a good weighting parameter and also has been shown to lead to the inverse scale space flow in its continuous limit. In order to understand the meaning of the term scale, we consider the spectral frequency representation with respect to the regularizer. Then, we take a look at two closely related functional lifting methods. These methods can be used in order to obtain a convex formulation for an originally non-convex problem.

Iterative Bregman Regularization

The iterative Bregman regularization was first introduced by Osher et al. in (OBG$^+$05) and builds on the Bregman distance as introduced in (Bre67). The primary goal of the new method was to improve ROF image denoising by reducing the loss of contrast induced by the TV regularizer, when the regularization parameter λ is chosen too small. An account on this problem can be found in (Mey01):

Example 4.1 (*ROF result on disc ((Mey01), p.36)*)

Let the input image be $f(x,y) = \alpha\mathcal{X}_r(x,y)$ for any $r \in \mathbb{R}^+$ and

$$\mathcal{X}_r(x,y) = \begin{cases} 1, & \text{if } \sqrt{x^2+y^2} \leq r, \\ 0, & \text{else.} \end{cases} \tag{34}$$

We assume $f = u + v$, where u is the denoised image and v the noise. Formulating the ROF problem as

$$u^* \in \arg\min_{u \in \mathrm{BV}(\Omega,\mathbb{R}^k)} \left\{ \lambda \int_\Omega ||u(x) - f(x)||_2^2 \mathrm{d}x + TV(u(x)) \right\}, \tag{35}$$

we get the decomposition

$$\begin{aligned} u^* = (\alpha - \frac{1}{\lambda r})\mathcal{X}_r \text{ and } v = \frac{1}{\lambda r}\mathcal{X}_r, &\quad \text{for } \alpha\lambda r \geq 1 \\ u^* = 0 \text{ and } v = f, &\quad \text{for } \alpha\lambda r \leq 1. \end{aligned} \tag{36}$$

We can see, that the choice of λ compromises the decomposition, especially when chosen too small. In particular, there will be a loss of contrast for $\alpha = 1$ and $\lambda > \frac{1}{r}$.

Instead of solving the ROF model for a variety of regularization parameters λ in search for the best one, the authors of (OBG$^+$05) proposed an iterative regularization method. The new method iteratively approaches the best solution by taking the Bregman distance associated with the actual regularization functional into account.

Let the class of considered variational problems be

$$u^* \in \arg\min_u \{G(u, f) + \lambda R(u)\}, \tag{37}$$

where $G(u, f)$ is a nonnegative convex data term with respect to u for fixed f, and $R(u)$ is a nonnegative convex regularization term. The proposed algorithm is:

Algorithm 4.2 (*Bregman Iteration*)

1. Initialize $p_0 = 0$ and $u_0 = 0$.
2. For $l = 0, 1, ...$ and while the stopping criterion is not fulfilled:

   Solve

   $$u_l := \arg\min_u \{G(u, f) + \lambda(R(u) - R(u_{l-1}) - \langle p_{l-1}, u - u_{l-1}\rangle)\} \tag{38}$$
   $$= \arg\min_u \{G(u, f) + \lambda(R(u) - \langle p_{l-1}, u\rangle)\}. \tag{39}$$

   Choose $p_l \in \partial R(u_l)$.

Instead of using the original regularization term $R(u)$, we use the Bregman distance $B_R^{p_{l_1}}(u, u_l) = R(u) - R(u_{l-1}) - \langle p_{l-1}, u - u_{l-1}\rangle$ associated with it. We can drop the constant terms $-R(u_{l-1}) - \langle p_{l-1}, -u_{l-1}\rangle$ in Eq. (38). Refer to Def. 3.13 for more details on the Bregman distance.

Let us now consider the special case of the ROF problem. The authors of (OBG$^+$05) prove the well-definedness of the iterates p_l and u_l. They also prove that u_l monotonically converges to f in L^2 for $l \to \infty$. Furthermore, they show that the Bregman distance between u_l and the noise-free image u decreases, as long as the data term evaluated at the iterate u_l is smaller than at the noise-free image u. In the ROF scenario, the latter is the noise level of the image and can often be estimated. If λ is chosen sufficiently large and the iteration is stopped as soon as the data term evaluated at u_l exceeds the noise level, the sequence $\{u_l\}_{l\in\mathbb{N}}$ converges towards the noise free image u. This stopping criterion is called *discrepancy principle*. In its continuous limit, the Bregman iteration leads to the *inverse scale space flow* (ISS), which is further elaborated in (BGO$^+$06) and (BGM$^+$16). The ISS is given as

$$\partial_s p(s) = f - u(s), \quad p(s) \in \partial R(u(s)), \quad p(0) = 0, \tag{40}$$

where f is the input image and $p(s)$ a subgradient of R evaluated at the solution $u(s)$ (BGM$^+$16). Basically, the flow u starts as the mean of the input image f, then it slowly recovers details of the input image according to their scale with respect to the

regularizer, until it finally converges to f. Large scale features are recovered prior to small scale features. This is of interest in image denoising, since small scale features are usually associated with noise. Stopping the flow at a suitable time renders the noise-free image.

In order to understand the term scale and the significance of the ISS flow, we introduce the concept of spectral representation.

Definition 4.3 (*Eigenvalue and eigenfunction, (BGM$^+$16)*)

Let $R : \xi \to \mathbb{R}^+$ be a convex functional. We refer to $v \in \xi$ with $||v||_2 = 1$ as *eigenfunction* with *eigenvalue* $\mu \in \mathbb{R}$, if

$$\mu v \in \partial(R(v)) \tag{41}$$

is satisfied.

Let us follow (BGM$^+$16) and consider the ISS (40) for an eigenfunction v with respect to R. Let μ be the corresponding eigenvalue and assume v has zero mean. The solution of the ISS (40) is then calculated as

$$u(s) = \begin{cases} 0, & \text{if } s \leq \mu \\ v, & \text{otherwise,} \end{cases} \tag{42}$$

and has a piecewise constant behavior in time. The greater the eigenvalue μ, the later in time the according eigenfunction v is recovered. The spectral frequency representation with respect to R is defined as

$$\phi_s := \partial_s u(s) = v\delta_\mu(s). \tag{43}$$

For more general input images f it can be shown that ϕ_s meets

$$f = \int_0^\infty \mathrm{d}\phi_s. \tag{44}$$

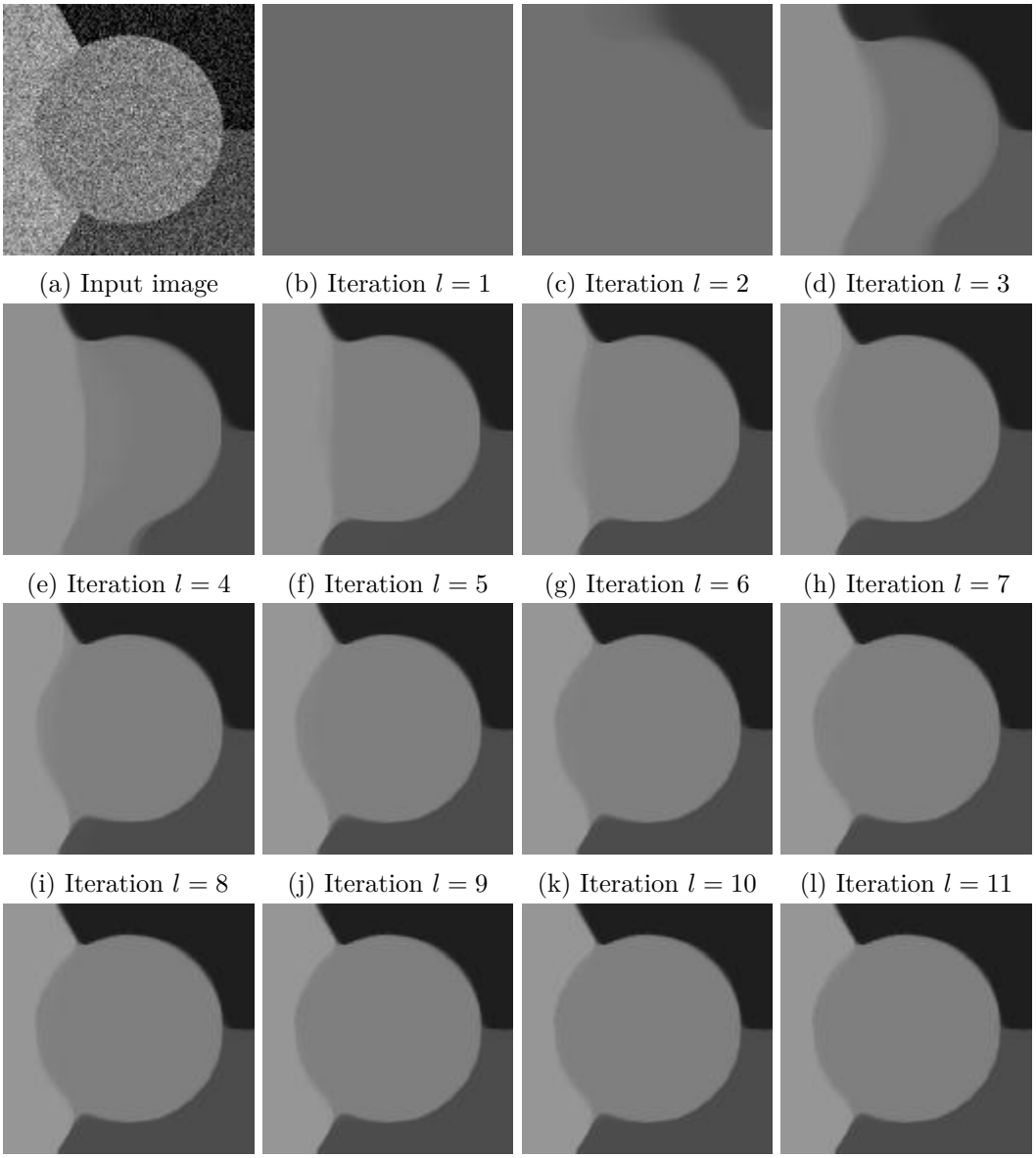
Every input image f can be decomposed into details of different scale. These details correspond to eigenfunctions associated with the regularizer R. Details that appear soon correspond to eigenfunctions with small eigenvalues and we call them large-scale details. Details that appear later on correspond to eigenfunctions with large eigenvalues and we call them small-scale details. We can see in Eq. (43) that the solution of the ISS gives us a spectral frequency representation with respect to R. For a real-life example see Fig. 6.

Fig. 7 shows the results of the iterative Bregman regularization on the ROF model, where the ISS flow qualities can be observed. Since λ is chosen large, we get an over-smoothed first solution u_1. As the iteration proceeds, more and more detail is recovered, starting with cartoon-like structures and contours. Fine structures such as the salt-and-pepper noise would have been regained if we had continued the iteration.

Figure 6: **Scale spaces.** Left: Input image. Middle: Large scale details with respect to the total variation regularizer of the input image. Right: Small scale details with respect to the total variation regularizer of the input image. Images from (Gil17).

In (OBG$^+$05) the possibility of applying the iterative Bregman regularization to further variational problems is discussed. The authors stress that a generalization with respect to different data terms is quite complex, since the compactness of the level sets of $G(u, f) + R(u) - \langle p_{l-1}, u\rangle$ is needed for the well-definedness of the iterations. They show that the before-mentioned properties hold for data terms of the form $\frac{\lambda}{2}||f - Ku||^2$, where $K \in L^2(\Omega)$ is a linear operator, which is bounded and whose kernel does not include the space of continuous functions.

Fixing the data term $G(u, f) = \frac{\lambda}{2}||f - Ku||^2$, the authors consider different regularization terms. For the well-definedness of the iterates the regularizer needs to be locally bounded, convex and nonnegative on a Banach space $U \in L^2(\Omega)$. Furthermore the level sets $\{u \in U|\ R(u) \leq \gamma \in \mathbb{R}\}$ need to be compact in $L^2(\Omega)$, nonempty for $\gamma > 0$, and R needs to be extendable to a weakly lower semi-continuous functional. If in addition $G(u, f) + R(u)$ is strictly convex, then there is always a unique minimizer u_l and $p_l = p_{l-1} + q_l$ for some $p_l \in \partial R(u_l)$, and $q_l \in \partial G(u_l, f)$ holds.



(a) Input image (b) Iteration $l = 1$ (c) Iteration $l = 2$ (d) Iteration $l = 3$

(e) Iteration $l = 4$ (f) Iteration $l = 5$ (g) Iteration $l = 6$ (h) Iteration $l = 7$

(i) Iteration $l = 8$ (j) Iteration $l = 9$ (k) Iteration $l = 10$ (l) Iteration $l = 11$

(m) Iteration $l = 12$ (n) Iteration $l = 13$ (o) Iteration $l = 14$ (p) Iteration $l = 15$

Figure 7: **Iterative Bregman regularization for ROF problem.** Top left: Input image from *prost* library (MLM$^+$15) with artificially added Gaussian white noise. The other images have been generated for this thesis with CVX (GB14, GB08). Left to right, top to bottom: Solutions of the iterative Bregman regularizer at different steps. The solution of the first iteration step is over-smoothed (in fact it is the mean of the input image). As the iteration proceeds, more and more detail is being recovered.

Functional Lifting

Not all variational models are convex (e.g. stereo matching Def. 2.2) and, therefore, it might not be numerically feasible to find a global optimal solution, even if it exists. Many algorithms used for solving variational models incorporate a form of gradient descent and can easily get stuck in local minima. Motivated by techniques for approximating combinatorial labeling problems (Ish03a) (see also (LLWS13) for an overview in the segmentation context), the authors of (PSG+08) present a framework which allows to approximate an originally non-convex variational model by a convex problem in a higher-dimensional space. They consider problems of the form

$$\arg \min_{u \in \mathrm{BV}(\Omega,\Gamma)} F(u), \;\; F(u) := \int_\Omega g(x, u(x)) \mathrm{d}x + TV(u), \tag{45}$$

where $\Omega \subset \mathbb{R}^2$, $\Gamma = [\gamma_{\min}, \gamma_{\max}] \subset \mathbb{R}^+$ and $g(x, u(x))$ continuous in x and $u(x)$ but not necessarily convex in $u(x)$. An extension to the more general problem class with arbitrary regularization term and infinite label space can be found in (PCBC10). The general class presents as

$$\arg \min_{u \in W^{1,1}(\Omega,\mathbb{R})} F(u), \;\; F(u) := \int_\Omega f(x, u(x), \nabla u(x)) \mathrm{d}x, \tag{46}$$

where f is continuous in the first two variables and convex in the last one.

The problem given by Eq. (45) was further considered in (MLM+15), where a refinement of the original method is presented. This refined approach begins with a discretization of the label space and allows for a tighter approximation of the data term after discretization (see Fig. 10).

Problem Formulation

In the following, we combine the notation of the above-mentioned papers and consider problems of the form

$$\arg \min_{u \in \mathrm{BV}(\Omega,\Gamma)} F(u), \;\; F(u) := \int_\Omega f(x, u(x), \mathrm{D}u(x)) \mathrm{d}x \tag{47}$$

$$:= \int_\Omega g(x, u(x)) \mathrm{d}x + TV(u), \tag{48}$$

for $\Omega \subset \mathbb{R}^2$, $\Gamma = \mathbb{R}$ and $\mathrm{D}u(x)$ the derivative of u in a distributional sense (see Prop. 3.2). The cost function $g(x, u(x))$ for assigning label $u(x)$ to x is continuous in x and u but not necessarily convex in u. Therefore, f is continuous in the first two variables, convex in the last one, and – depending on the data term – possibly non-convex in the second one.

In the following sections, we take a look at the method introduced in (PSG+08) and describe the proposed lifting method for variational models with TV regularizer using the notation of (PCBC10). Afterwards, we present the alternative approach proposed in (MLM+15) and establish a relation between both approaches. The following section heavily builds on the three papers (PSG+08, PCBC10, MLM+15).

Functional Lifting

The main goal of functional lifting is to find a higher-dimensional representation of a given non-convex variational problem, such that the new formulation is convex. The method presented in (PSG$^+$08, PCBC10) comprises four steps: First, a lifted representation $\phi \in \mathrm{BV}(\Omega \times \Gamma, \{0,1\})$ for $u \in \mathrm{BV}(\Omega, \Gamma)$, as well as the search space over these representations is defined. Then, the energy term is reformulated as a convex variational problem with respect to the lifted search space. In the third step, the lifted search space is extended to its convex envelope. One can then argue that any solution of the lifted problem over the convex envelope of the lifted search space translates to a solution of the original problem.

1st Step. We begin by expressing functions u by their γ-superlevel-sets. Given a function $u \in \mathrm{BV}(\Omega, \Gamma)$, we define the binary function $\phi : \Omega \times \Gamma \to \{0,1\}$ as

$$\phi(x,\gamma) := 1_{\{u>\gamma\}}(x) := \begin{cases} 1, & \text{if } u(x) > \gamma, \\ 0, & \text{otherwise.} \end{cases} \tag{49}$$

This function is equivalent to the indicator function of the γ-superlevel-sets of u (see Fig. 8) or, to put it in another way, the characteristic function of the subgraph of $u(x)$ (see Fig. 9). In particular, ϕ is a representation of u on a higher-dimensional domain and it is easy to see, that we can retrieve u by applying what is in (PSG$^+$08) called the layer-cake-formula

$$u(x) = \int_{-\infty}^{0} (\phi(x,\gamma) - 1)\,\mathrm{d}\gamma + \int_{0}^{\infty} \phi(x,\gamma)\mathrm{d}\gamma. \tag{50}$$

The new search space is defined as

$$C' := \left\{ \phi \in \mathrm{BV}(\Omega \times \mathbb{R}, \{0,1\}) \,\middle|\, \lim_{\gamma\to-\infty} \phi(x,\gamma) = 1, \lim_{\gamma\to\infty} \phi(x,\gamma) = 0 \right\}. \tag{51}$$

Note that feasible ϕ have to decline with respect to γ, such that $\phi_\gamma(x,\gamma) \leq 0$ holds for the the partial derivative. This constraint is later on enforced by the lifted energy term (53).

This lifting approach is similar to the one introduced for the Chan-Vese model in (CEN06), where indicator functions are used in order to describe sets. For references to alternative lifting approaches, a short overview and comparison thereof we refer the reader to (CCP12).

2nd Step. Let Γ_u be the jump set of ϕ and let $\nu_{\Gamma_u} := (p^x, p^\gamma) := (\mathrm{D}_x\phi(x,\gamma), \partial_\gamma\phi(x,\gamma))$ denote the inner unit normal of ϕ at Γ_u. The initial problem (48) can be expressed in terms of the *interfacial energy* of ϕ evaluated at Γ_u. In (PCBC10) the authors denote this interfacial energy as

$$\int_{\Gamma_u} h(x,\gamma,\nu_{\Gamma_u})\mathrm{d}\mathcal{H}^2(x), \tag{52}$$

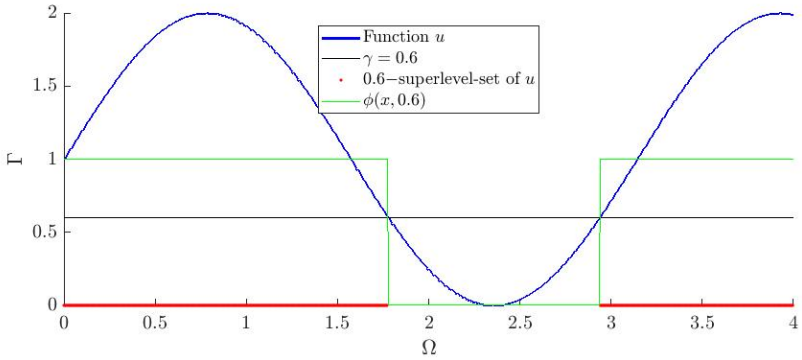


Figure 8: **Example of γ-superlevel-set of u.** Visualizing the lifting procedure. The blue line represents the function u. The black line marks $\gamma = 0.6$. The red area is the 0.6-superlevel-set of u, which is a subset of Γ. The green line is the lifted representation of u for fixed $\gamma = 0.6$ and equivalent to the indicator function of the 0.6-superlevel-set of u.

where $\mathcal{H}^2$ is the 2-dimensional Hausdorff-measure and h is defined as

$$h(x, \gamma, (p^x, p^\gamma)) := \begin{cases} |p^\gamma| f\left(x, \gamma, \frac{p^x}{|p^\gamma|}\right), & \text{if } p^\gamma < 0, \\ \lim_{\lambda \to +\infty} \frac{1}{\lambda} f(x, \gamma, \lambda p^x), & \text{if } p^\gamma = 0, \\ +\infty, & \text{if } p^\gamma > 0. \end{cases} \tag{53}$$

The authors point out, that h is lower semi-continuous, convex and one-homogeneous with respect to (p^x, p^γ).

In order to develop an intuition for h, we will specify some characteristics shown in (PCBC10). For a visual reference we refer the reader to Fig. 9. As we have mentioned earlier, feasible inner normals ν_{Γ_u} have to be non-positive in p^γ, otherwise they would imply that ϕ increases at some point with respect to γ. The third case in the definition of h (53) ensures that this request is met. For $u \in C^1(\Omega, \Gamma)$ we have $p^\gamma = -1$ and h satisfies

$$\int_{\Gamma_u} h(x, \gamma, \nu_{\Gamma_u}(x)) \mathrm{d}\mathcal{H}^d(x) = \int_\Omega h(x, u(x), (\nabla u(x), -1)) \mathrm{d}x \tag{54}$$

$$\overset{Eq.\ (53)}{=} \int_\Omega |-1| f\left(x, u(x), \frac{\nabla u(x)}{|-1|}\right) \mathrm{d}x \tag{55}$$

$$= \int_\Omega f(x, u(x), \nabla u(x)) \mathrm{d}x. \tag{56}$$

In this scenario the lifted and unlifted problem are equivalent. In case of u not being sufficiently smooth, it has a jump at some $x \in \Omega$, which means $p^\gamma = 0$ and $p^x \neq 0$. In

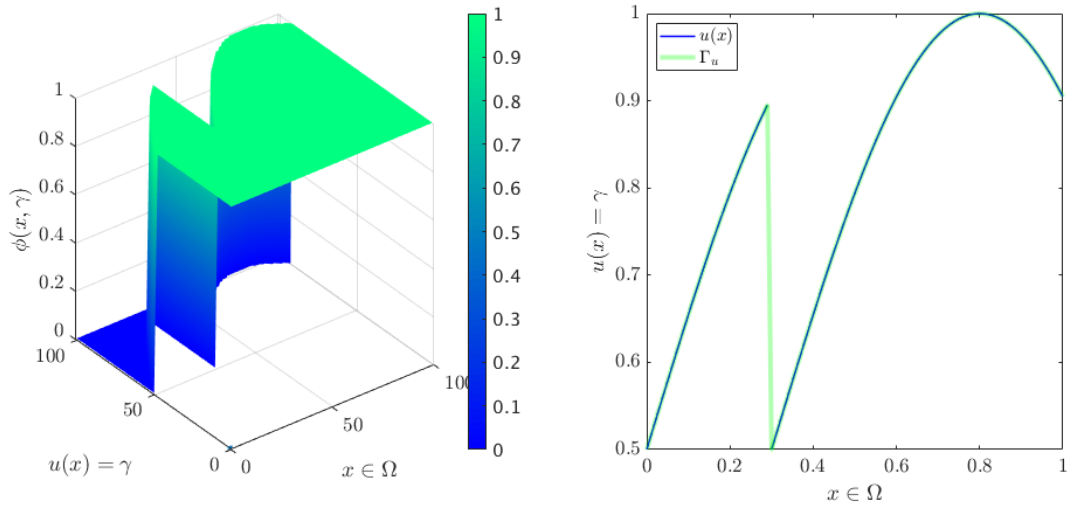


Figure 9: **Relaxation using the characteristic function of the subgraph.** Right plot: Functional $u(x)$ as blue line. Note the jump of $u(x)$ at $x = 0.3$. The light green line indicates the jump set of $\phi(x, y)$. Left plot: Characteristic function of the subgraph of $u(x)$ as the two horizontal light green and dark blue areas. The vertical part with color gradient represents the jump set of $\phi(x, \gamma)$, which we will now call Γ_u. Consider the normal of Γ_u at $x = 0.8$. We have $\nu_{\Gamma_u} = (\nabla u(0.8), -1)$. At the jump $x = 0.3$ on the other hand the functional u is non-differentiable. For $x = 0.3$ and $\gamma \leq 0.9$ we get $p^\gamma = 0$.

this case, h becomes

$$\int_{\Gamma_u} h(x, \gamma, \nu_{\Gamma_u}(x)) \mathrm{d}\mathcal{H}^d(x) = \int_\Omega h(x, u(x), (\nabla u(x), 0)) \mathrm{d}x \tag{57}$$

$$\overset{Eq.\ (53)}{=} \int_\Omega \lim_{\lambda \to +\infty} \frac{1}{\lambda} f(x, \gamma, \lambda p^x) \mathrm{d}x \tag{58}$$

$$= \int_\Omega |p^x| \mathrm{d}x. \tag{59}$$

Again, the lifted and unlifted formulation are equivalent.

We can now formulate a convex and one-homogeneous lifted version of the minimization problem (48). Due to the fact that h is convex and one-homogeneous with respect to (p^x, p^γ), we can furthermore show the equivalence to the original formulation (48):

$$F(\phi(x, \gamma)) := \int_{\Omega \times \mathbb{R}} h(x, \gamma, \mathrm{D}\phi(x, \gamma)) \mathrm{d}(x, \gamma) \tag{60}$$

$$= \int_{\Omega \times \mathbb{R}} h(x, \gamma, \nu_{\Gamma_u}(x, \gamma)) \mathrm{d}\mathcal{H}^d \llcorner \Gamma_u(x, \gamma) \tag{61}$$

$$= \int_{\Gamma_u} h(x, \gamma, \nu_{\Gamma_u}(x)) \mathrm{d}\mathcal{H}^d(x) = \int_\Omega f(x, u(x), \mathrm{D}u(x)) \mathrm{d}x. \tag{62}$$

3rd Step. Minimizing F over C' is still not a convex problem due to the non-convexity of the search space. Therefore, we replace C' with its convex envelope

$$C := \left\{ \phi \in \mathrm{BV}(\Omega \times \mathbb{R}, [0,1]) \middle| \lim_{\gamma \to -\infty} \phi(x,\gamma) = 1, \lim_{\gamma \to \infty} \phi(x,\gamma) = 0 \right\}. \tag{63}$$

Again, feasible ϕ have to decrease with respect to γ. The definition of h ensures that this condition is met.

4th Step. In (PCBC10) and (PSG+08), the authors show that thresholding of a global minimizer $\phi^* \in \arg\min_{\phi \in C} F(\phi(x,\gamma))$ of the lifted problem (60) results in a global minimizer $1_{\phi^* > s} = \arg\min_{\phi \in C'} F(\phi(x,\gamma))$ of the lifted problem on the non-convex set C' for almost any $s \in [0,1]$. Depending on the choice of s we reach different functionals in C'. Applying the layer-cake-formula (50) to $1_{\phi^* > s}$ provides a minimizer $u^* = \int_{-\infty}^{0} (1_{\phi^* > s}(x,\gamma) - 1)\, \mathrm{d}\gamma + \int_{0}^{\infty} 1_{\phi^* > s}(x,\gamma) \mathrm{d}\gamma$ of the initial non-convex problem (46). In case of total variation regularization, we explicitly get

$$\arg\min_{\phi \in C} F(\phi), \quad F(\phi) = \int_{\Omega \times \Gamma} g(x, u(x)) |\partial_\gamma \phi(x,\gamma)| \mathrm{d}(x,\gamma) + TV(\phi), \tag{64}$$

$$C = \left\{ \phi \in \mathrm{BV}(\Omega \times \mathbb{R}, [0,1]) \middle| \lim_{\gamma \to -\infty} \phi(x,\gamma) = 1, \lim_{\gamma \to \infty} \phi(x,\gamma) = 0 \right\}. \tag{65}$$

Here, we only considered the problem class (45). We started with a variational problem (48) that was possibly non-convex in one variable. Now, we have a lifted representation (65) of the initial problem. This lifted representation is due to the properties of the function h (53) convex and lower semi-continuous. If the initial data term g is proper, the lifted variational problem is also proper. In this case we are guaranteed the existence of a minimizer (Cor. 3.7), which is a global minimizer (Thm. 3.9).

In (PCBC10) it is shown that this method can also be applied to problems of the form (46) with slight changes on the lifted search spaces C' and C, and a dual formulation of the lifted problem is presented. Furthermore, details on the discretization and implementation can be found. After discretization, it turns out that the data term is approximated in a linear fashion in between the discretized labels (MLM+15). In the following section, we introduce a slightly different approach which allows for an even tighter approximation (see Fig. 10).

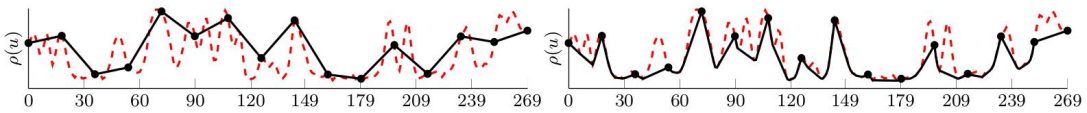


Figure 10: **Approximation of the data term after discretization.** Red dotted line: non-convex integrand of the data term $g(\cdot, u(\cdot))$. Left: Piecewise linear approximation of the data term in between the labels, using the lifting method introduced in Sec. 4.2.2. Right: Piecewise convex approximation of the data term in between the labels, using the lifting method proposed in Sec. 4.2.3. Images from (MLM+15).

Sublabel-Accurate Functional Lifting

A slightly different approach for the problem class (48) was suggested in (MLM$^+$15). Unlike the method introduced in the last section, this approach assumes a discretized label space $\Gamma = [\gamma_{\min}, \gamma_{\max}]$. Although the overall notion for both methods is quite similar, they use a different notation. We now present this new notation and explain the connection to the one in the previous section.

The method can be divided into two steps. First, we partition the label space Γ into k intervals $\Gamma_i = [\gamma_i, \gamma_{i+1}]$, such that $\Gamma = \bigcup_{i=1}^k \Gamma_i$ holds for $\gamma_1 < ... < \gamma_{k+1} \in \Gamma$. Using this partition, a lifted representation of $u(\cdot)$ and the search space are defined. Then, we reformulate the data and regularization term separately by lifting them into the higher-dimensional space and determining the respective biconjugates.

1st Step. After choosing some partition $\bigcup_{i=1}^k [\gamma_i, \gamma_{i+1}]$ of the label space Γ, every $u(\cdot) \in \Gamma_i$ can be rewritten with the help of the chosen labels and a variable $\alpha \in [0, 1]$:

$$u(\cdot) = \gamma_i^\alpha := \gamma_i + \alpha(\gamma_{i+1} - \gamma_i). \tag{66}$$

Using these variables i and α, we can find a lifted representation $\mathbf{u}(\cdot) \in \mathbb{R}^k$ for $u(\cdot)$:

$$\mathbf{u}(\cdot) = \mathbf{1}_i^\alpha := \mathbf{1}_{i-1} + \alpha(\mathbf{1}_i - \mathbf{1}_{i-1}), \tag{67}$$

where $\mathbf{1}_i$ represents a k-dimensional vector of i ones followed by zeros. Even though we have discretized the label space and and are using a vectorial representation, we still allow for all sublabels due to the continuity of the variable $\alpha \in [0, 1]$. This vectorial representation is furthermore unique, in the sense that any $\mathbf{u}$ can uniquely be mapped to a $u \in \Gamma$ by

$$u(x) = \gamma_1 + \sum_{i=1}^k \mathbf{u}_i(x)(\gamma_{i+1} - \gamma_i). \tag{68}$$

Note however, that any $u \in \{\gamma_2, ..., \gamma_k\}$ has two lifted representations, since the equality $\mathbf{1}_i^1 = \mathbf{1}_{i+1}^0$ holds for any $i \in \{1, 2, ..., k\}$.

Let us denote the set of all feasible $\mathbf{u}$ by

$$Q' = \left\{ \mathbf{u} = \mathbf{1}_i^\alpha \in \mathbb{R}^k \,\middle|\, \mathbf{u}_1 = ... = \mathbf{u}_{i-1} = 1, \mathbf{u}_i = \alpha, \mathbf{u}_{i+1} = ... = \mathbf{u}_{L-1} = 0 \right\}. \tag{69}$$

This Q' corresponds to C' in section 4.2.2: each element $\mathbf{u} \in Q'$ is a vectorial representation of the characteristic function of the subgraph $\phi(\cdot, \gamma)$ of some $u(\cdot)$. However, Q' is non-convex and later on we consider its convex envelope $[0, 1]^k$. In fact, we even consider $\mathbb{R}^k$, yet due to the definition of the lifted integrand only solutions in $[0, 1]^k$ are possible.

2nd Step for Data Term. Next, we are going to introduce the lifted representation of the energy term. For this step, data and regularization term are considered separately. For fixed $x \in \Omega$, the integrand of the data term $g(u) := g(\cdot, u(\cdot)) : \Gamma \to \mathbb{R}$ is to be minimized over Γ. In (MLM$^+$15) the authors introduce the integrand of the lifted data term $\mathbf{g} : \mathbb{R}^k \to \mathbb{R} \cup +\infty$ as

$$\mathbf{g}(\mathbf{u}) = \min_{1 \le i \le k} \mathbf{g}_i(\mathbf{u}), \tag{70}$$

$$\mathbf{g}_i(\mathbf{u}) = \begin{cases} g(\gamma_i^\alpha), & \text{if } \mathbf{u} = \mathbf{1}_i^\alpha, \ \ \alpha \in [0,1], \\ \infty, & \text{else.} \end{cases} \tag{71}$$

They stress that minimizing $\mathbf{g}(\mathbf{u})$ over $\mathbb{R}^k$ gives a solution in Q'. Inserting this solution into the layer-cake-formula (68) yields a minimizer of g over Γ. They show that the convex envelope of the integrand (70) of the lifted data term is given as

$$\mathbf{g}^{**}(\mathbf{u}) = \sup_{\mathbf{v}\in\mathbb{R}^k} \left\{ \langle \mathbf{u}, \mathbf{v} \rangle - \max_{1\le i\le k} \mathbf{g}_i^*(\mathbf{v}) \right\} \tag{72}$$

with

$$\begin{aligned} \mathbf{g}_i^*(\mathbf{v}) &= c_i(\mathbf{v}) + g_i^* \left(\frac{\mathbf{v}_i}{\gamma_{i+1} - \gamma_i} \right), \\ c_i(\mathbf{v}) &= \langle \mathbf{1}_{i-1}, \mathbf{v} \rangle - \frac{\gamma_i}{\gamma_{i+1} - \gamma_i}, \\ g_i &= g + \delta_{\Gamma_i}, \\ \delta_C(x) &= \begin{cases} 0, & x \in C, \\ \infty, & x \notin C. \end{cases} \end{aligned} \tag{73}$$

We now have an integrand $\mathbf{g}^{**}(\mathbf{u})$ which is convex in $\mathbf{u}$. In particular, the originally non-convex integrand $g(u)$ is approximated piecewise convex on every interval Γ_i. See Fig. 11 for a simple demonstration of the lifting process and Fig. 10 for a comparison with the method introduced in the previous section.

2nd Step for Regularization Term: For the lifted representation of the regularization term, the authors of (MLM$^+$15) proceed analogously. Using the notion introduced in Prop. 3.2, they formulate the total variation of the lifted representation $\mathbf{u}$ with a finite $\mathbb{R}^{k\times d}$-valued Radon measure $\mathrm{D}\mathbf{u}$ as

$$TV(\mathbf{u}) = \int_\Omega \mathrm{d}\mathbf{\Phi}(x, \mathrm{D}\mathbf{u}), \tag{74}$$

For fixed x, they introduce the lifted regularizer as

$$\mathbf{\Phi}(\mathbf{z}) = \min_{1\le i\le j\le k} \mathbf{\Phi}_{i,j}(\mathbf{z}). \tag{75}$$

Since they want to penalize jumps from some value γ_i^α to γ_j^β (with $\alpha, \beta \in [0,1]$) in the direction of $v \in \mathbb{R}^d$, they choose $\mathbf{\Phi}_{i,j} : \mathbb{R}^{k\times d} \to \mathbb{R} \cup \{\infty\}$ as

$$\mathbf{\Phi}_{i,j}(\mathbf{z}) = \begin{cases} |\gamma_i^\alpha - \gamma_j^\beta| \cdot |v|_2, & \text{if } \mathbf{z} = (\mathbf{1}_i^\alpha - \mathbf{1}_j^\beta) v^T, \\ \infty, & \text{else.} \end{cases} \tag{76}$$

This lifted representation of the total variation is non-convex, therefore, the authors determine the convex envelope. Crucially, they show that for ordered labels $\gamma_1 < \gamma_2 < ...$, the convex envelope is given by

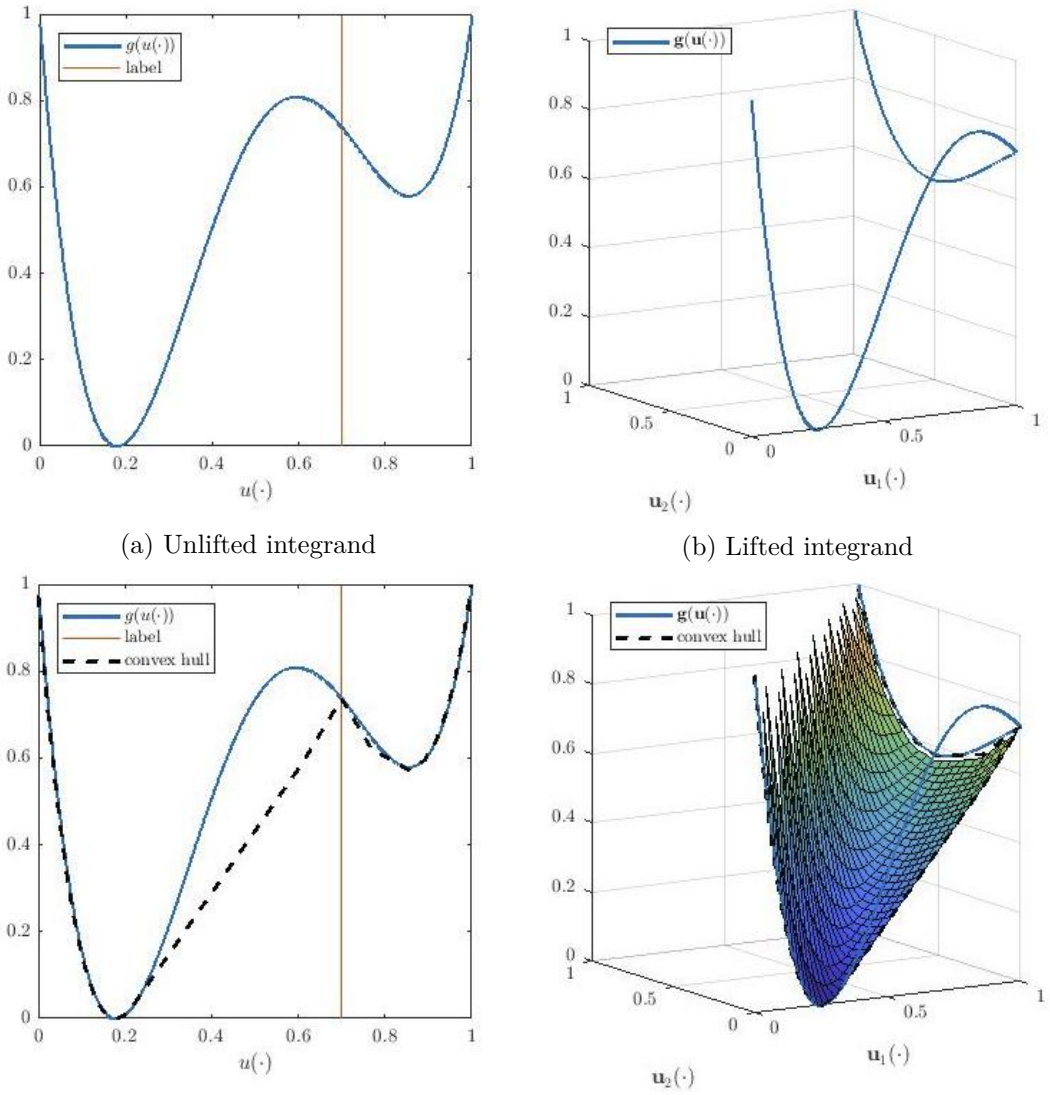
$$\mathbf{\Phi}^{**}(\mathbf{z}) = \sup_{\mathbf{q}\in K} \langle \mathbf{q}, \mathbf{z} \rangle, \tag{77}$$

with the constraint set

$$K = \left\{ \mathbf{q} \in \mathbb{R}^{k \times d} \middle| \ |\mathbf{q}_i|_2 \leq \gamma_{i+1} - \gamma_i, \ \ \forall i \right\}. \tag{78}$$

This constraint set can be efficiently enforced due to the finite and – with respect to the number of labels – linear number of constraints.

How does the lifting and convexification of the data and regularization term connect to the approach in section 4.2.2? In the first lifting approach, we formulated the convex envelope of the indicator functions, by replacing them with functions that vary between 0 and 1, and decrease with respect to γ. The latter was being enforced by the lifted formulation of the problem. In this sublabel-accurate lifting, we formally work on the solution space $\mathbb{R}^k$ but restrict possible solutions to the convex hull $[0,1]^k$ of Q', by the choice of the lifted data term $\mathbf{g}$ and use of its biconjugate $\mathbf{g}^{**}$. Here, the expression "decreasing with respect to γ" translates to $\mathbf{u}_1 \geq \mathbf{u}_2 \geq ... \geq \mathbf{u}_k$. However, the sublabel-accurate lifting does not seem to enforce this restriction, as we will get back to in Ch. 6



(a) Unlifted integrand

(b) Lifted integrand

(c) Unlifted convex hull

(d) Lifted convex hull

Figure 11: **Relaxation by subsampling of label space.** Top left: Integrand $g(u)$ of non-convex data term for fixed $x \in \Omega$ and $u \in \Gamma = [0, 1]$. The red line marks the second of the chosen labels $[0, 0.7, 1]$. Top right: Lifted integrand of data term (z-axis) for the lifted $\mathbf{u}$ (x- and y-axis). The lifting process creases the integrand at the second label and spreads each interval to $[0, 1]$. Bottom right: Convex hull of the lifted integrand. Bottom left: Projection of the intersection of the convex hull of the lifted integrand and the x-, y-axis into the original space (black dotted line). Overall we have convex integrands in the data term and, therefore, a convex data term. Furthermore, the original integrand of the data term is approximated by the convex hull in between the labels.

Iterative Bregman Regularization for Sublabel-Accurate Lifting

In the last chapter, we presented two successful methods used in variational image processing: the iterative Bregman regularization and the sublabel-accurate lifting. We showed how the sublabel-accurate lifting can be used to obtain a convex formulation of an originally non-convex variational problem. Furthermore, we discussed some advantages offered by the Bregman iteration and its connection to the inverse scale space flow. Until now, the application of the Bregman iteration has been restricted to convex variational problems. In this chapter we want to discuss the possibility of applying the Bregman iteration to variational problems with total variation regularizer and a possibly non-convex data term by using the sublabel-accurate lifting.

Our contribution in this thesis is the formulation of a lifted Bregman iteration. We address some arising questions and conjecture that the solution of the Bregman iteration on an unlifted convex problem is equivalent to the one on the sublabel-accurate lifted problem. We make first steps towards proving this hypothesis. Details on the implementation of the proposed method and first experimental results for the ROF and stereo matching problem can be found in the next chapter.

Formulation of the Problem

Let us assume we have a problem of the form

$$\arg\min_{u\in BV(\Omega,\Gamma)} F(u), \quad F(u) := G(f,u) + TV(u), \tag{79}$$

$$G(f,u) := \int_\Omega g(x,u(x))\mathrm{d}x, \tag{80}$$

$$TV(u) := \int_\Omega \Phi(x,\mathrm{D}u)\mathrm{d}x, \tag{81}$$

for some input image f, closed domain $\Omega \subset \mathbb{R}^2$, closed range $\Gamma \subset \mathbb{R}^+$, distributional derivate $\mathrm{D}u$, and a data term $G(f,u)$ which is nonnegative, lower semi-continuous and possibly non-convex in u. We refer to this problem as unlifted or original problem. For simplicity reasons, we drop the dependency on the input image f from now on. According to chapter 4.2.3 we choose some labels $\gamma_1, ..., \gamma_{k+1}$ and perform the sublabel-accurate lifting. This results in the lifted problem

$$\arg\min_{\mathbf{u}\in BV(\Omega,\mathbb{R}^k)} \mathbf{F}(\mathbf{u}), \quad \mathbf{F}(\mathbf{u}) := \mathbf{G}(\mathbf{u}) + \mathbf{TV}(\mathbf{u}), \tag{82}$$

$$\mathbf{G}(\mathbf{u}) := \int_\Omega \mathbf{g}^{**}(x,\mathbf{u}(x))\mathrm{d}x, \tag{83}$$

$$\mathbf{TV}(\mathbf{u}) := \int_\Omega \mathbf{\Phi}^{**}(x,\mathrm{D}\mathbf{u}(x))\mathrm{d}x, \tag{84}$$

with $\mathbf{g}^{**}(x,\mathbf{u}(x))$ as in equation (72) and $\mathbf{\Phi}^{**}(x,\mathrm{D}\mathbf{u}(x))$ as in equation (77).

The lifted problem has a nonnegative convex data term $\mathbf{G}(\mathbf{u})$ and a nonnegative convex regularizer $\mathbf{TV}(\mathbf{u})$. Therefore, it fulfills two basic requirements for the Bregman iteration. We can formulate the Bregman iteration on the lifted problem as follows:

Algorithm 5.1 (*Lifted Bregman Iteration*)

1. Initialize $\mathbf{p}_0 = 0$ and $\mathbf{u}_0 = 0$,
2. For $l = 0, 1, ...,$ and while the stopping criterion is not fulfilled:

   Solve

   $$\mathbf{u}_l := \arg\min_{\mathbf{u} \in BV} \left\{ \mathbf{G}(\mathbf{u}) + \mathbf{TV}(\mathbf{u}) - \langle \mathbf{p}_{l-1}, \mathbf{u} \rangle \right\}. \tag{85}$$

   Choose $\mathbf{p}_l \in \partial \mathbf{TV}(\mathbf{u}_l)$.

Some questions arise immediately. Does the Bregman iteration on an unlifted convex problem provide the same results as the lifted Bregman iteration? Under which requirements on the data term is the lifted Bregman iteration well-posed? Does the lifted Bregman iteration result in scale spaces with respect to the regularizer for variational problems with a non-convex data term?

Lifted Bregman Iteration on the ROF

This section is dedicated to the question whether the Bregman iteration on a convex problem, namely the ROF, provides the same solution as the lifted Bregman iteration. We conjecture that this is the case and take some first steps towards proving our assumption. In order to distinguish between the lifted and unlifted setting, we use the notation presented in the previous section. In particular, bold letters indicate the lifted setting.

The following questions need to be considered:

1. How does the subdifferential of the lifted problem relate to the subdifferential of the original problem? For the Bregman iteration on the ROF model the authors of (OBG$^+$05) show in particular that $p_l := p_{l-1} - (u_l - f)$ lies in the subdifferential of $TV(u_l)$ and use this subgradient for their proofs concerning the well-definedness of the iterates and convergence results. Can we find a subgradient in the lifted total variation, which is equivalent to this subgradient of the unlifted total variation?

2. Assuming the existence of a lifted subgradient in $\partial \mathbf{TV}(\mathbf{u}_l)$ which is equivalent to the unlifted subgradient $p_l := p_{l-1} - (u_l - f) \in \partial TV(u_l)$, are the solutions of the next unlifted and lifted Bregman iteration equivalent?

3. The implementation of the sublabel-accurate lifting presents a straightforward alternative way for choosing a subgradient of the lifted regularizer. Can we also use

this subgradient for the lifted Bregman iteration and how does the choice of the subgradient influence the results?

The first and second question are closely related. In this section, we empirically consider the first question and leave the second question to future work. The experimental results presented in Ch. 6 give some first hints to the answer of the third question.

Since the Bregman distance does not play a role in the first iteration, we assume that the solutions of the first unlifted and lifted Bregman iteration are equivalent. More precisely, we assume that the solution of the first lifted Bregman iteration is given by a $\mathbf{u}_1 \in BV(\Omega, Q')$, where Q' is given by Eq. (69). Furthermore, we assume that this lifted solution is uniquely linked to the solution $u_1 \in \mathrm{BV}(\Omega, \mathbb{R})$ of the first unlifted Bregman iteration via the layer-cake formula (68). Let us denote this mapping for fixed $x \in \Omega$ via the functional $L : \Gamma \to Q'$,

$$L(\gamma_i^\alpha) = \mathbf{1}_i^\alpha \text{ and } L^{-1}(\mathbf{1}_i^\alpha) = \gamma_i^\alpha, \tag{86}$$

for $u_1(\cdot) = \gamma_i^\alpha$ as in Eq. (66) and $\mathbf{u}_1(\cdot) = \mathbf{1}_i^\alpha$ as in Eq. (67). Furthermore, we denote by $p_1 \in \partial TV(u_1)$ the subgradient chosen after the first unlifted Bregman iteration and by $u_2 \in BV(\Omega, \Gamma)$ the solution of the second unlifted Bregman iteration. It follows, that $0 \in \partial G(u_2) + \partial TV(u_2) - p_1$ holds.

In the following, we will suggest a lifting $\mathbf{p}_l = L_{\delta_i}(p_l)$ for subgradients $p_l \in TV(u_l)$ and $u_l = \gamma_i^\alpha$. We leave it to future work to verify whether

$$L(u_{l+1}) = \arg\min_{\mathbf{u} \in \mathrm{BV}} \{\mathbf{G}(\mathbf{u}) + \mathbf{TV}(\mathbf{u}) - \langle \mathbf{p}_l, \mathbf{u} \rangle\} \tag{87}$$

holds, which would answer the second question.

Lifted Integrand of the Data Term

Before considering the question whether we can find a lifted representation for subgradients of the regularizer, we take a closer look at the lifted integrand of the data term. The following considerations will help us understand the lifted data term and to define a lifting for subgradients later on. For simplicity reasons we drop the dependency on x from now on. For data terms with convex integrands $g(u)$,

$$\mathbf{g}^{**}(\mathbf{u}) = \begin{cases} g(L^{-1}(\mathbf{u})), & \text{if } \mathbf{u} \in Q', \\ \infty, & \text{if } \mathbf{u} \in \mathbb{R}^k \backslash [0,1]^k, \end{cases} \tag{88}$$

holds. In other words, the lifted and unlifted integrand coincide on the respective part of Γ and Q'. Outside of the convex hull of Q' the lifted integrand becomes infinity. In the following we formulate a representation with convex combinations for the lifted integrand evaluated at a point within the convex hull of Q': Essentially, the value of $\mathbf{g}^{**}$ at any point $u \in (\operatorname{con} Q') \backslash Q'$ can be expressed as a convex combination of its values at boundary points, i.e., at points in Q'.

Lemma 5.2 (*Convex Combination for Lifted Integrand of the Data Term*)

Let $g(u(\cdot)) : \Gamma \to \overline{\mathbb{R}}$ be lower semi-continuous, proper and convex. Furthermore, let $\mathbf{g}^{**}$ be defined as in Eq. (72). Every $\mathbf{u} \in [0,1]^k$ can be represented by a convex combination $\mathbf{u} = \sum_{i=1}^k \lambda_i \mathbf{1}_i^{\alpha_i}$ with $\lambda_1, ..., \lambda_k \in [0,1]$, $\sum_{i=1}^k \lambda_i = 1$ and $\mathbf{1}_1^{\alpha_1}, ..., \mathbf{1}_k^{\alpha_k} \in Q'$ such that

$$\mathbf{g}^{**}\left(\sum_{i=1}^k \lambda_i \mathbf{1}_i^{\alpha_i}\right) = \sum_{i=1}^k \lambda_i \mathbf{g}^{**}\left(\mathbf{1}_i^{\alpha_i}\right) \tag{89}$$

holds.

Proof: Consider a family of functionals $\{f_i\}_{i=1}^k$ with

$$f_i(\mathbf{u}) := \begin{cases} g(L^{-1}(\mathbf{u})), & \text{if } \mathbf{u} = \mathbf{1}_i^{\alpha} \in Q', \\ +\infty, & \text{else.} \end{cases} \tag{90}$$

Following relationship between $\{f_i\}_{i=1}^k$ and $\mathbf{g}^{**}$ holds, since g is convex:

$$f_i(\mathbf{u}) = \mathbf{g}^{**}(\mathbf{u}), \quad \text{if } \mathbf{u} = \mathbf{1}_i^{\alpha} \in Q'. \tag{91}$$

Due to the definition of the integrand g, all functionals of the family $\{f_i\}_{i=1}^k$ are proper, lower semi-continuous and convex. According to (Roc70, Cor. 17.1.3), the convex hull f of the family $\{f_i\}_{i=1}^k$ evaluated at any $\mathbf{u}(\cdot)$ is given as

$$f(\mathbf{u}) = \inf_{\substack{\lambda_i \geq 0, \\ \mathbf{u}_i \in \mathbb{R}^k}} \left\{ \sum_{i=1}^k \lambda_i f_i(\mathbf{u}_i) \,\middle|\, \mathbf{u} = \sum_{i=1}^k \lambda_i \mathbf{u}_i \right\}. \tag{92}$$

For now, we focus on $\mathbf{u} \in [0,1]^k$. We only need to consider linear combinations with $\mathbf{u}_i$ which have a representation $\mathbf{1}_i^{\alpha_i} \in Q'$, since by definition all functionals of the family $\{f_i\}_{i=1}^k$ become infinity outside of Q'. Therefore,

$$f(\mathbf{u}) = \inf_{\substack{\lambda_i \in [0,1], \\ \mathbf{1}_i^{\alpha_i} \in Q'}} \left\{ \sum_{i=1}^k \lambda_i f_i(\mathbf{1}_i^{\alpha_i}) \,\middle|\, \sum_{i=1}^k \lambda_i = 1, \mathbf{u} = \sum_{i=1}^k \lambda_i \mathbf{1}_i^{\alpha_i} \right\}. \tag{93}$$

All functionals $f_1, ..., f_k$ are lower semi-continuous and their (positive weighted) sum is also lower semi-continuous. Therefore, the search space is a bounded and closed subspace of $\mathbb{R}^{2k}$ and consequently compact. Every lower semi-continuous function has a minimizer in any compact set. Therefore, we can replace the infimum by a minimum:

$$f(\mathbf{u}) = \min_{\substack{\lambda_i \in [0,1], \\ \mathbf{1}_i^{\alpha_i} \in Q'}} \left\{ \sum_{i=1}^k \lambda_i f_i(\mathbf{1}_i^{\alpha_i}) \,\middle|\, \sum_{i=1}^k \lambda_i = 1, \mathbf{u} = \sum_{i=1}^k \lambda_i \mathbf{1}_i^{\alpha_i} \right\}. \tag{94}$$

Due to the relationship (91) between f_i and $\mathbf{g}^{**}$, this is equivalent to

$$f(\mathbf{u}) = \min_{\substack{\lambda_i \in [0,1], \\ \mathbf{1}_i^{\alpha_i} \in Q'}} \left\{ \sum_{i=1}^k \lambda_i \mathbf{g}^{**}(\mathbf{1}_i^{\alpha_i}) \,\middle|\, \sum_{i=1}^k \lambda_i = 1, \mathbf{u} = \sum_{i=1}^k \lambda_i \mathbf{1}_i^{\alpha_i} \right\}. \tag{95}$$

We proceed with the proof of (89). Since $\mathbf{g}^{**}$ is convex, Jensen's inequality

$$\mathbf{g}^{**}\left(\sum_{i=1}^{k}\lambda_i\mathbf{1}_i^{\alpha_i}\right) \leq \sum_{i=1}^{k}\lambda_i\mathbf{g}^{**}(\mathbf{1}_i^{\alpha_i}) \tag{96}$$

holds for arbitrary $\mathbf{1}_i^{\alpha_i} \in Q'$ and $\lambda_i \in [0,1]$ adding up to one. Note that the linear combination used to describe $\mathbf{u}$ is not unique, yet the equality $\mathbf{g}^{**}\left(\sum_{i=1}^{k}\lambda_i\mathbf{1}_i^{\alpha_i}\right) = \mathbf{g}^{**}\left(\sum_{j=1}^{k}\lambda_j\mathbf{1}_j^{\alpha_j}\right)$ holds for all possible linear combinations. As a consequence, we can approximate the left-hand side of (96) by the minimum of the right-hand side:

$$\mathbf{g}^{**}(\mathbf{u}) \leq \min_{\substack{\lambda_i\in[0,1],\\ \mathbf{1}_i^{\alpha_i}\in Q'}} \left\{ \sum_{i=1}^{k}\lambda_i\mathbf{g}^{**}(\mathbf{1}_i^{\alpha_i}) \,\middle|\, \sum_{i=1}^{k}\lambda_i = 1, \mathbf{u} = \sum_{i=1}^{k}\lambda_i\mathbf{1}_i^{\alpha_i} \right\}. \tag{97}$$

Looking at equation (95), we get

$$\mathbf{g}^{**}(\mathbf{u}) \leq f(\mathbf{u}), \tag{98}$$

which concludes the first part of the proof. Since the convex hull of $\mathbf{g}$ on $[0,1]^k$ is proper, the following equality holds according to equation (28)

$$\mathbf{g}^{**}(\mathbf{u}) = \mathrm{cl}\ (\mathrm{con}\ \mathbf{g}(\mathbf{u})) = \mathrm{cl}\left(\sup_{\substack{\phi\leq\mathbf{g}\\ \phi \text{ convex}}} \phi(\mathbf{u})\right). \tag{99}$$

The convex hull f, which is defined in equation (95), is convex and bounds $\mathbf{g}$ from below for any $\mathbf{u}$. This means

$$\mathbf{g}^{**}(\mathbf{u}) = \mathrm{cl}\left(\sup_{\substack{\phi\leq\mathbf{g}\\ \phi \text{ convex}}} \phi(\mathbf{u})\right) \geq f(\mathbf{u}). \tag{100}$$

The claimed equality (89) follows with the two inequalities (98) and (100). ■

In this section we examined the integrand of a sublabel-accurate lifted data term, where the unlifted data term is convex. We now have the following understanding of the integrand: For any $\mathbf{u} \in \mathbb{R}^k \backslash [0,1]^k$ the integrand becomes infinity, for any $\mathbf{u} \in Q'$ the integrand is equal to the unlifted integrand evaluated at $L^{-1}(\mathbf{u})$. For any other point $\mathbf{u} \in [0,1]^k \backslash Q'$ we can find in accordance to Lemma 5.2 a convex combination $\mathbf{u} = \sum_{i=1}^{k}\lambda_i\mathbf{1}_i^{\alpha}$, such that $\mathbf{g}^{**}(\mathbf{u}) = \sum_{i=1}^{k}\lambda_i\mathbf{g}^{**}(\mathbf{1}_i^{\alpha})$ holds for the lifted integrand.

Subgradient of the Lifted Integrand of the Data Term

Let us now consider the subgradient used in Alg. 5.1. For the l-th lifted Bregman iteration we need a subgradient $\mathbf{p}_{l-1}$ of the lifted regularizer evaluated at the solution of the previous step, or more explicitly $\mathbf{p}_{l-1} \in \mathbf{TV}(\mathbf{u}_{l-1})$. Since $\mathbf{u}_l$ is the minimizer of the l-th iteration,

$$0 \in \partial \mathbf{G}(\mathbf{u}_l) + \partial \mathbf{TV}(\mathbf{u}_l) - \mathbf{p}_{l-1} \tag{101}$$

holds. Therefore, we can express the subdifferential of the lifted total variation via the subdifferential of the lifted data term:

$$\partial \mathbf{TV}(\mathbf{u}_l) = \mathbf{p}_{l-1} - \partial \mathbf{G}(\mathbf{u}_l). \tag{102}$$

In this section we study the subdifferential of the lifted data term. Our aim is to define a lifting for a subgradient q_l of the unlifted data term $G(u_l)$, such that Eq. (87) holds for $\mathbf{p}_l = \mathbf{p}_{l-1} - L(q_l)$. Here, we will only define the lifting and leave the proof of Eq. (87) to future work.

Considering Eq. (31), we expect

$$\partial G(u) \supset \int_\Omega \partial g(x, u(x)) \mathrm{d}x, \tag{103}$$

$$\partial \mathbf{G}(\mathbf{u}) \supset \int_\Omega \partial \mathbf{g}^{**}(x, \mathbf{u}(x)) \mathrm{d}x \tag{104}$$

to hold, since g and $\mathbf{g}^{**}$ are both proper, convex, and lower semi-continuous in u. However, a rigorous proof would be required for this integral version of Eq.(31). We point the reader to (CHPA18).

From now on we drop the dependency on x and proceed with a couple of assumptions. Having found the solution $u_l = \gamma_i^\alpha$ of the l-th unlifted Bregman iteration, we assume that this solution has a lifted counterpart of the form $L(u_l) = \mathbf{1}_i^\alpha \in Q'$, which is the solution of the l-th lifted Bregman iteration. Furthermore, we assume that any subgradient $p_l \in \partial\, G(u_l)$ can be decomposed according to equation (104) and consider the integrand only. Given $p \in \partial g(u_1)$, we define a lifting functional L_{∂_i} which provides a lifted counterpart $\mathbf{p} \in \partial \mathbf{g}^{**}(L(u_1))$.

Theorem 5.3 (*Lifting of a Subgradient of the Integrand of the Data Term*)

Let $u = \gamma_i^\alpha \in \Gamma$ be the minimizer of the unlifted problem (79) evaluated at a fixed point $x \in \Omega$. If the unlifted integrand g of the data term in Eq. (79) is convex and if p is a subgradient of g evaluated at u, define

$$\mathbf{p} := L_{\partial i}(p) := \begin{cases} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ p(\gamma_{i+1} - \gamma_i) \\ \vdots \\ p(\gamma_{k+1} - \gamma_k) \end{pmatrix}, & \text{if } p \leq 0, \\ \begin{pmatrix} p(\gamma_2 - \gamma_1) \\ \vdots \\ p(\gamma_{i+1} - \gamma_i) \\ 0 \\ \vdots \\ 0 \end{pmatrix}, & \text{if } p \geq 0, \end{cases} \tag{105}$$

with $p(\gamma_{i+1} - \gamma_i)$ in the i-th row. Then $\mathbf{p} = L_{\partial i}(p)$ satisfies $\mathbf{p} \in (\partial(\mathbf{g}^{**}))(L(\gamma_i^\alpha))$ with $\mathbf{g}^{**}$ defined in Eq. (72) and L given in Eq. (86). In other words: for any subgradient of g evaluated at u we can find a lifted counterpart $L_{\partial i}(p)$ in the subdifferential of the lifted integrand $\mathbf{g}^{**}$ evaluated at $L(u)$.

Proof: In order to show that Thm. 5.3 holds, we need to prove that $\mathbf{p} \in (\partial(\mathbf{g}^{**}))(L(u))$ holds, which is equivalent to the following inequation:

$$\mathbf{g}^{**}(\mathbf{v}) \geq \mathbf{g}^{**}(L(u)) + \langle \mathbf{p}, \mathbf{v} - L(u) \rangle, \quad \forall \mathbf{v} \in [0,1]^k. \tag{106}$$

Please note that the case $\mathbf{v} \in \mathbb{R}^k \backslash [0,1]^k$ is trivial, since $\mathbf{g}^{**}$ becomes infinity outside of $[0,1]^k$. We are going to consider the two cases $\mathbf{v} \in Q'$ and $\mathbf{v} \in [0,1]^k \backslash Q'$ separately.

First Case Let $\mathbf{v} \in Q'$ and define $\mathbf{p}$ according to (105). Let

$$L(u) = \mathbf{u} = \mathbf{1}_i^\alpha \in Q', \tag{107}$$

$$L(v) = \mathbf{v} = \mathbf{1}_j^\beta \in Q', \tag{108}$$

$$u = \gamma_i + \alpha(\gamma_{i+1} - \gamma_i), \tag{109}$$

$$v = \gamma_j + \beta(\gamma_{j+1} - \gamma_j). \tag{110}$$

Since g is convex, $\mathbf{g}^{**}(L(v)) = g(v)$ and $\mathbf{g}^{**}(L(u)) = g(u)$ hold for any $u, v \in \Gamma$. Due to the assumption $p \in \partial g(u)$,

$$g(v) \geq g(u) + p(v - u) \tag{111}$$

$$\overset{(109)(110)}{\Leftrightarrow} g(v) \geq g(u) + p(\gamma_j + \beta(\gamma_{j+1} - \gamma_j) - \gamma_i - \alpha(\gamma_{i+1} - \gamma_i)) \tag{112}$$

holds for any v. In case of $i = j$ this reduces to

$$g(v) \geq g(u) + p(\gamma_{i+1} - \gamma_i)(\beta - \alpha). \tag{113}$$

In the following, we consider all possible cases for p, i, j. For each case we proceed by substituting $\mathbf{p}$, $L(u) = \mathbf{1}_i^\alpha$, $L(v) = \mathbf{1}_i^\beta$, $\mathbf{g}^{**}(L(v)) = g(v)$ and $\mathbf{g}^{**}(L(u)) = g(u)$ on the right-hand side of equation (106) and consequently we show that the inequality holds by referring to the equations (112) and (113).

Let $i = j$ and $p \in \partial g(u)$ arbitrary. After the substitution we get

$$g(v) \geq g(u) + p(\gamma_{i+1} - \gamma_i)(\beta - \alpha), \quad \forall \mathbf{v} \in Q', \tag{114}$$

which holds according to (113).

Let us next consider the case of a non-negative subgradient $p \geq 0$. This means, that its lifted representation $\mathbf{p}$ is a vector which contains zeros at the positions $(i+1)$ to k. For $j > i$ the vector $L(v) - L(u)$ has $(i-1)$ zeros followed by $(1-\alpha)$ and some other terms, which are of no interest to us. We can estimate

$$\begin{aligned}
&\mathbf{g}^{**}(L(u)) + \langle \mathbf{p}, L(v) - L(u) \rangle && (115)\\
&= g(u) + p(\gamma_{i+1} - \gamma_i)(1 - \alpha) && (116)\\
&\leq g(u) + p((\gamma_j - \gamma_i) - \alpha(\gamma_{i+1} - \gamma_i)) && (117)\\
&\leq g(u) + p((\gamma_j - \gamma_i) - \alpha(\gamma_{i+1} - \gamma_i)) + \beta(\gamma_{j+1} - \gamma_j)) && (118)\\
&\overset{(112)}{\leq} g(v) = \mathbf{g}^{**}(\mathbf{v}). && (119)
\end{aligned}$$

For $j < i$ the vector $L(v) - L(u)$ is non-zero at the positions $j, ..., i$. After the substitution and with equation (112) we get

$$\begin{aligned}
&\mathbf{g}^{**}(L(u)) + \langle \mathbf{p}, L(v) - L(u) \rangle && (120)\\
&= g(u) + p(\beta - 1)(\gamma_{j+1} - \gamma_j) - \sum_{k=j+1}^{i-1} p(\gamma_{k+1} - \gamma_k) - \alpha p(\gamma_{i+1} - \gamma_i) && (121)\\
&= g(u) + \beta p(\gamma_{j+1} - \gamma_j) - \sum_{k=j}^{i-1} p(\gamma_{k+1} - \gamma_k) - \alpha p(\gamma_{i+1} - \gamma_i) && (122)\\
&= g(u) + \beta p(\gamma_{j+1} - \gamma_j) + p\gamma_j - p\gamma_i - \alpha p(\gamma_{i+1} - \gamma_i) && (123)\\
&= g(u) + p((\gamma_j - \gamma_i) - \alpha(\gamma_{i+1} - \gamma_i)) + \beta(\gamma_{j+1} - \gamma_j)) && (124)\\
&\overset{(112)}{\leq} g(v) = \mathbf{g}^{**}(\mathbf{v}). && (125)
\end{aligned}$$

Let us now consider the case of a non-positive subgradient $p \leq 0$. In this case, the proposed lifting results in a vector whose first i entries are zero. For $j < i$, the vector $L(v) - L(u)$ is $(-\alpha)$ at the i-th position followed by zeros. This leads to the estimation

$$\begin{aligned}
&\mathbf{g}^{**}(L(u)) + \langle \mathbf{p}, L(v) - L(u) \rangle && (126)\\
&= g(u) + p(\gamma_{i+1} - \gamma_i)(-\alpha) && (127)\\
&\leq g(u) + p((\gamma_{i+1} - \gamma_i)(-\alpha) + (\beta - 1)(\gamma_{j+1} - \gamma_j)) && (128)\\
&\leq g(u) + p((\gamma_{i+1} - \gamma_i)(-\alpha) + \beta(\gamma_{j+1} - \gamma_j) - (\gamma_i - \gamma_j)) && (129)\\
&\overset{(112)}{\leq} g(v) = \mathbf{g}^{**}(L(v)). && (130)
\end{aligned}$$

For $j > i$ the vector $L(v) - L(u)$ has non-zero entries at the positions $i, ..., j$. One last time we apply the substitution and get

$$\mathbf{g}^{**}(L(u)) + \langle \mathbf{p}, L(v) - L(u) \rangle \tag{131}$$

$$= g(u) + (1-\alpha)p(\gamma_{i+1} - \gamma_i) + \sum_{k=i+1}^{j-1} p(\gamma_{k+1} - \gamma_k) + \beta p(\gamma_{j+1} - \gamma_j) \tag{132}$$

$$= g(u) - \alpha p(\gamma_{i+1} - \gamma_i) + \sum_{k=i}^{j-1} p(\gamma_{k+1} - \gamma_k) + \beta p(\gamma_{j+1} - \gamma_j) \tag{133}$$

$$= g(u) - \alpha p(\gamma_{i+1} - \gamma_i) + p\gamma_j - p\gamma_i + \beta p(\gamma_{j+1} - \gamma_j) \tag{134}$$

$$= g(u) - p\alpha(\gamma_{i+1} - \gamma_i) + p(\gamma_j + \beta(\gamma_{j+1} - \gamma_j) - \gamma_i) \tag{135}$$

$$\overset{(112)}{\leq} g(v) = \mathbf{g}^{**}(L(v)). \tag{136}$$

This concludes the proof of the first case.

Second Case It remains to consider $\mathbf{v} \in [0,1]^k \backslash Q'$. According to Lemma (5.2), there exists a convex combination $\mathbf{v} = \sum_{j=1}^k \lambda_j \mathbf{1}_j^{\beta_j}$ with $\lambda_1, ..., \lambda_k \in [0,1]$, $\sum_{j=1}^k \lambda_j = 1$, and $\beta_1, ..., \beta_k \in [0,1]$, for which

$$\mathbf{g}^{**}\left(\sum_{j=1}^k \lambda_j \mathbf{1}_j^{\beta_j}\right) = \sum_{j=1}^k \lambda_j \mathbf{g}^{**}\left(\mathbf{1}_j^{\beta_j}\right) \tag{137}$$

holds. For arbitrary $\mathbf{v} = \sum_{j=1}^k \lambda_j \mathbf{1}_j^{\beta_j} \in [0,1]^k \backslash Q'$ with $\mathbf{v}_j := \mathbf{1}_j^{\beta_j} \in Q'$ and $\mathbf{p}$ as in (105), the following approximation holds:

$$\mathbf{g}^{**}(\mathbf{v}) = \mathbf{g}^{**}\left(\sum_{j=1}^k \lambda_j \mathbf{1}_j^{\beta_j}\right) = \sum_{j=1}^k \lambda_j \mathbf{g}^{**}\left(\mathbf{1}_j^{\beta_j}\right) = \sum_{j=1}^k \lambda_j g(\mathbf{v}_j) \tag{138}$$

$$\geq \sum_{j=1}^k \lambda_j (g(L(u)) + \langle \mathbf{p}, \mathbf{v}_j - L(u) \rangle) \tag{139}$$

$$= \sum_{j=1}^k \lambda_j g(L(u)) + \sum_{j=1}^k \lambda_j \langle \mathbf{p}, \mathbf{v}_j - L(u) \rangle \tag{140}$$

$$= \sum_{j=1}^k \lambda_j g(L(u)) + \langle \mathbf{p}, \sum_{j=1}^k \lambda_j (\mathbf{v}_j - L(u)) \rangle \tag{141}$$

$$= \sum_{j=1}^k \lambda_j g(L(u)) + \langle \mathbf{p}, \sum_{j=1}^k \lambda_j \mathbf{v}_j - \sum_{j=1}^k \lambda_j L(u) \rangle \tag{142}$$

$$= g(L(u)) + \langle \mathbf{p}, \sum_{j=1}^k \lambda_j \mathbf{v}_j - L(u) \rangle \tag{143}$$

$$= g(L(u)) + \langle \mathbf{p}, \mathbf{v} - L(u) \rangle. \tag{144}$$

This concludes the proof of (105). We have indeed found a lifting, which gives us for any subgradient of the integrand of the unlifted data term a subgradient of the integrand of the lifted data term. ■

In this chapter, we introduced the lifted Bregman iteration Alg. 5.1. We made the conjecture that the Bregman iteration and its lifted counterpart lead to equivalent results on convex variational problems with total variation regularizer. We took a first step towards proving this assumption by defining a lifting for subgradients of the integrands of the unlifted data term. For a comprehensive proof it remains to show that the inclusion in Eq. (103) holds, and that using the subgradient and its lifted counterpart in the original and lifted Bregman iteration respectively results in equivalent minimizers. As a useful supplementary insight, we proved that any integrand of the sublabel-accurate lifted data term evaluated at some point in $[0,1]^k$ can be rewritten as a convex combination of evaluations at points in Q'.

Implementation and Numerical Results

We aim to experimentally validate some of our conjectures concerning the lifted Bregman iteration, which we introduced in the last chapter. First, we discuss details on the discretization and implementation of the algorithm. Then, consider the ROF model in order to compare the results and behavior of the original and lifted Bregman iteration on convex problems. Finally, we show first results achieved with the novel algorithm on the non-convex stereo matching problem.

Discretization and Implementation

Discretization an Numerical Optimization. We begin by providing the formulation for the discretization of the sublabel-accurate lifted problem and recall some steps deduced for numerical optimization in (MLM$^+$15). As we will see later, this formulation of the discretized lifted problem is easily extended to the Bregman iteration and offers a straightforward way of choosing subgradients. For more details and interim steps on the numerical optimization, we refer the reader to the original paper.

Let Ω^h denote the 2-dimensional Cartesian grid with grid spacing h representing the discretized image domain $\Omega \subset \mathbb{R}^2$. According to (MLM$^+$15), the sublabel-accurate lifted problem (82) is discretized to

$$\arg \min_{\mathbf{u}:\Omega^h \to \mathbb{R}^k} \sum_{x \in \Omega^h} \left\{ \mathbf{g}^{**}(x, \mathbf{u}(x)) + \mathbf{\Phi}^{**}(x, \nabla \mathbf{u}(x)) \right\}, \tag{145}$$

where $\mathbf{g}^{**}(x, \mathbf{u}(x))$ is given by Eq. (72) and $\mathbf{\Phi}^{**}(x, \nabla \mathbf{u}(x))$ by Eq. (77). Following (MLM$^+$15), we drop the dependency on x when possible for simplicity reasons. The data term requires to calculate the maximum over the conjugates $\mathbf{g}_i^*$ (73) with respect to the labels $i = 1, ..., k$. This is simplified by introducing the real variable t (147). We get the formulation:

$$\arg \min_{\mathbf{u}:\Omega^h \to \mathbb{R}^k} \max_{\substack{t:\Omega^h \to \mathbb{R} \\ \mathbf{v}:\Omega^h \to \mathbb{R}^k \\ \mathbf{q}:\Omega^h \to \mathbb{R}^{k \times d}}} \left\{ \langle \mathbf{u}, \mathbf{v} \rangle - \sum_{x \in \Omega^h} t(x) + \langle \mathbf{q}, \nabla \mathbf{u} \rangle \right\}, \tag{146}$$

$$\text{s.t. } t(x) \geq \mathbf{g}_i^*(\mathbf{v}(x)), \tag{147}$$

$$|\mathbf{q}_i|_2 \leq \gamma_{i+1} - \gamma_i, \tag{148}$$

with $\mathbf{q}_i$ denoting the i-th row of $\mathbf{q}$. Solving this problem using proximal methods such as PDHG, requires to implement explicit projections on the constraint sets. In order to simplify the projections onto (147),

the authors introduce another auxiliary variable $\mathbf{z} : \Omega^h \to \mathbb{R}^k$ (150) with rows $\mathbf{z}_i$, which results in the problem

$$\arg \min_{\mathbf{u}:\Omega^h \to \mathbb{R}^k} \max_{\substack{t:\Omega^h \to \mathbb{R} \\ \mathbf{v}:\Omega^h \to \mathbb{R}^k \\ \mathbf{q}:\Omega^h \to \mathbb{R}^{k \times d}}} \langle \mathbf{u}, \mathbf{v} \rangle - \sum_{x \in \Omega^h} t(x) + \langle \mathbf{q}, \nabla \mathbf{u} \rangle, \tag{149}$$

$$\text{s.t. } \mathbf{z}_i(x) = (t(x) - c_i(\mathbf{v}(x)))(\gamma_{i+1} - \gamma_i), \tag{150}$$

$$\mathbf{z}_i(x) \geq \mathbf{g}_i^* \left(\frac{\mathbf{v}_i(x)}{\gamma_{i+1} - \gamma_i} \right) (\gamma_{i+1} - \gamma_i), \tag{151}$$

$$|\mathbf{q}_i|_2 \leq \gamma_{i+1} - \gamma_i. \tag{152}$$

The first equality constraint (150) is implemented with the help of a Lagrange multiplier $\mathbf{s} : \Omega^h \to \mathbb{R}^k$. For the second constraint (151), the authors provide two options for the orthogonal projections onto the epigraphs of the conjugates $\mathbf{g}_i^*$ for quadratic and piecewise linear convex pieces $\mathbf{g}_i$.

Most importantly, the discretized total variation is given by the term

$$F^h(\nabla \mathbf{u}) := \sum_{x \in \Omega^h} \mathbf{\Phi}^{**}(\nabla \mathbf{u}(x)) = \max_{\mathbf{q}:\Omega^h \to \mathbb{R}^{k \times d}} \left\{ \sum_{x \in \Omega^h} \langle \mathbf{q}(x), \nabla \mathbf{u}(x) \rangle - \delta_K(\mathbf{q}(x)) \right\}. \tag{153}$$

with set K given by Eq. (78). Here, the F^h – which denotes the sum over $\mathbf{\Phi}^{**}$ – is proper and convex, ∇ is a linear transform, and the range of ∇ contains a point of the relative interior of the domain of F^h. Therefore, we can apply (Roc70, Thm. 23.9) and get

$$\partial(F^h \circ \nabla)(\mathbf{u}(x)) = \nabla^\top (\partial F^h)(\nabla \mathbf{u}(x)). \tag{154}$$

Since $\mathbf{\Phi}^{**}$ is proper and lower semi-continuous, we get with Eq. (31)

$$\partial F^h(\nabla \mathbf{u}(x)) \supset \sum_{x \in \Omega^h} \partial \mathbf{\Phi}^{**}(\nabla \mathbf{u}(x)). \tag{155}$$

In addition, $\mathbf{\Phi}^{**}$ is convex and for fixed $x \in \Omega^h$ a subgradient of $\mathbf{\Phi}^{**}$ fulfills (RW09, Prop. 11.3)

$$\mathbf{p} \in \partial \mathbf{\Phi}^{**}(z) \Leftrightarrow \mathbf{p} \in \arg \max_{\mathbf{p} \in \mathbb{R}^{k \times d}} \left\{ \langle \mathbf{p}, z \rangle - \delta_K(\mathbf{p}) \right\}. \tag{156}$$

This means that $\nabla^\top \mathbf{q}$, with $\mathbf{q}$ being the miximizer of Eq. (153) is a subgradient of $F^h(\nabla \mathbf{u}_l)$. We can use this for the discretized lifted Bregman iteration by incorporating the maximizer $\mathbf{q}$ ascertained in the preceding iteration into the subsequent iteration. This results in the following formulation for the discretization of the lifted Bregman iteration:

Algorithm 6.1 (*Discretized Lifted Bregman Iteration*)

1. Initialize $\mathbf{p}_0 = 0$.
2. For $l = 0, 1, ...$ and while the stopping criterion is not fulfilled, solve

$$\mathbf{u}_l = \arg \min_{\mathbf{u}:\Omega^h \to \mathbb{R}^k} \max_{\substack{t:\Omega^h \to \mathbb{R} \\ \mathbf{v}:\Omega^h \to \mathbb{R}^k \\ \mathbf{q}:\Omega^h \to \mathbb{R}^{k \times d}}} \left\{ \langle \mathbf{u}, \mathbf{v} \rangle - \sum_{x \in \Omega^h} t(x) + \langle \mathbf{q}, \nabla \mathbf{u} \rangle - \langle \mathbf{p}_{l-1}, \nabla \mathbf{u} \rangle \right\}, \tag{157}$$

$$\text{s.t. } \mathbf{z}_i(x) = (t(x) - c_i(\mathbf{v}(x)))(\gamma_{i+1} - \gamma_i), \tag{158}$$

$$\mathbf{z}_i(x) \geq \mathbf{g}_i^* \left(\frac{\mathbf{v}_i(x)}{\gamma_{i+1} - \gamma_i} \right) (\gamma_{i+1} - \gamma_i), \tag{159}$$

$$|\mathbf{q}_i|_2 \leq \gamma_{i+1} - \gamma_i. \tag{160}$$

Set $\mathbf{p}_l = \mathbf{q}$.

Implementation. The implementation in this thesis is based on the libraries `prost` and `sublabel_relax` (MLM$^+$15). `Prost` allows to solve large-scale problems of the form

$$\min_{u \in \mathbb{R}^n} \max_{v \in \mathbb{R}^m} \{ g(u) + \langle Ku, v \rangle - f^*(v) \}. \tag{161}$$

Here, K is a linear operator, $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $f^* : \mathbb{R}^m \to \overline{\mathbb{R}}$ are convex functionals. The library provides the ADMM (Hes69, EB92) and the PDHG (EZC10, PCBC09, PC11, GLY$^+$13) solver. Both solvers are implemented in CUDA. `Sublabel_relax` offers an extension of `prost` to the functional-lifting setting. In this thesis, we extend the libraries to the lifted Bregman iteration, in particular we reuse and extend some of the examples presented in the libraries. Since the libraries are not well documented, we provide a short introduction and some helpful information in the appendix.

Numerical Results

We present the results of the lifted Bregman iteration on two well-known problems. First, we consider the convex ROF model. Since the original Bregman iteration is also applicable to the ROF model, we can compare the results obtained in the lifted and unlifted setting. The second considered example is the non-convex stereo matching model. Unlike the first example, there is no empirical way to verify the results, since the combination with the sublabel-accurate lifting allows the use of the iterative Bregman regularization on this specific problem for the first time.

The experiments were run on a computer with the following specifications:

- Ubuntu 16.04.3 LTS
- CPU Intel(R) Core(TM) i7-2600 CPU @ 3.40GHz
- RAM 15 GB
- NVIDIA GeForce GTX 480, 1536 MB
- NVIDIA CUDA, release 8.0, V8.0.61
- Matlab 2017a

Iterative Bregman Regularization on Lifted ROF

Consider the ROF energy term

$$\int_\Omega ||u(x) - f(x)||_2^2 \mathrm{d}x + \lambda TV(u(x)), \tag{162}$$

where f is the input picture. We want to compare the results of the iterative Bregman regularizer on the unlifted and the sublabel-accurate lifted problem. We use an image w from the `prost` library (MLM$^+$15) and add artificial Gaussian white noise v with standard deviation 0.1, such that our input image is $f = w + v$.

First, we solve the unlifted problem using CVX, a package for specifying and solving convex programs (GB14, GB08). We choose $\lambda = 20$ and use the forward Neumann difference operator for the discretized TV in order to match the implemented example in `sublabel_relax`. We stop the iteration according to the discrepancy principle as soon as we have a solution u_l such that $||f - u_l||_{L^2} < ||f - v||_{L^2}$.

For the lifted Bregman iteration we consider the case of two- and five-label lifting. The first case is equivalent to the unlifted problem and we can use its solution as an indicator of the error introduced by projecting the lifted solution onto the original label space.

We choose the PDHG solver. Before running the lifted Bregman iteration, we need to adjust the parameters of the solver. Our aim is to avoid deviations of the solution of the lifted from the ones of the unlifted Bregman iteration induced by the choice of the solver parameters. Therefore, we begin by solving the first iteration step of the two

	two labels	five labels
step-size	alg2	alg2
residual_iter	16	12
alg2_gamma	0.001 λ	0.00001 λ
tau0	15	13
sigma0	0.09	0.11

Table 1: Parameters of PDHG solver for lifted Bregman iteration on ROF

lifted problems. We project both solutions onto the original label space and calculate both objective function values without the Bregman term according to Eq. (162). We compare these values to the optimality value of the solution of the first unlifted iteration step. We then fine-tune the parameters of the solver, until the difference between the lifted and the unlifted optimality value becomes minimal. This results in the parameters listed in Tab. 1.

After specifying the parameters, we continue by solving the remaining iteration steps. Since the ROF is a convex problem, we expect the results of the lifted and unlifted Bregman iteration to be equal up to machine accuracy. Both the unlifted Bregman iteration and the two-label lifted Bregman iteration take 15 iteration steps before the discrepancy principle is fulfilled. The five-label lifted Bregman iteration requires 18 iteration steps and takes almost twice as long as the two-label lifted iteration. For details on the runtime, see Fig. 13. We calculate the optimality values of all solutions by inserting the solutions into Eq. (162). The result is shown in Fig. 12. The optimality values of the unlifted and the two-label lifted problem are equal up to machine accuracy. The optimality values of the unlifted and the five-label lifted problem show some larger deviations, especially around iteration three to six.

Next, we compare the actual output images. In Fig. 14 we show a selection of interesting interim steps. The solutions of the unlifted and the two-label lifted problem are visually identical. The five-label lifted iteration shows a similar qualitative behavior. However, we observe that in the case of the five-label lifting, the algorithm needs more iteration steps to recover the basic outline of the shape. This is particularly noticeable in iteration three to six, as is shown in Fig. 15.

The different convergence behavior of the unlifted and five-label lifted problem might be due to the fact, that we were not able to determine solver parameters (see Tab. 1) for which the solution of the first five-label lifted iteration came as close to the solution of the first unlifted iteration, as the solution of the first two-label lifted iteration. Another possible explanation is that the choice of the subgradient affected the results. For the unlifted Bregman iteration we followed (OBG$^+$05) and chose subgradients by solving $p_l := p_{l-1} - u_l + f$. For the lifted Bregman iteration, we chose the subgradient according to Alg. 6.1.

Although we were not able to reproduce the exact results of the Bregman iteration with the five-label lifted version, the results look very promising. It appears that the lifted Bregman iteration shows overall the same qualitative properties as the original

Bregman iteration.

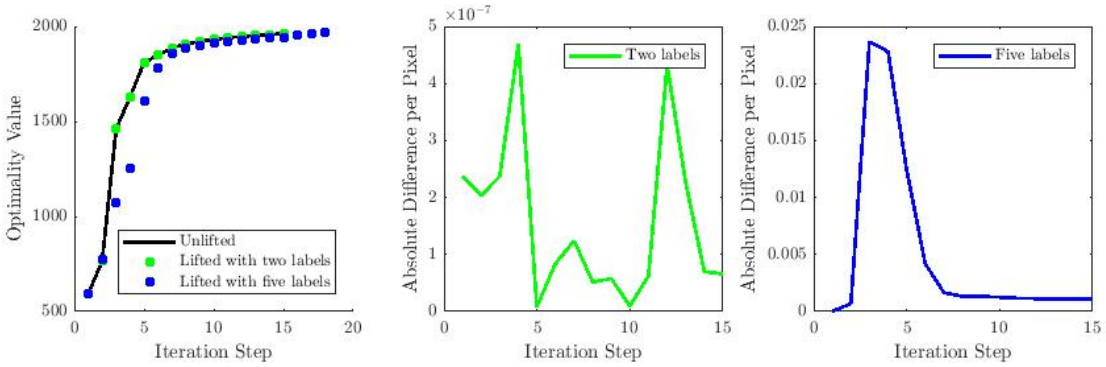


Figure 12: **Objective function values.** Left: Values obtained by plugging the solutions of the Bregman iteration into the objective function of the unlifted ROF problem (162). Middle: Per-pixel averaged absolute difference between the optimality of the two-label lifted and unlifted iteration. Right: Respective difference between the solutions of the five-label lifted and unlifted iteration. Note the different scales of the y-axes of the middle and right plot. The unlifted and two-label lifted iteration render the same optimality up to machine accuracy. The optimality values of the five-label lifted iteration deviate especially for the first iterations. When we fitted the parameters of the PDHG solver so as to obtain preferably equivalent solutions in the first unlifted and lifted iteration, the difference in optimality was larger for the five-label lifting than in the two-label lifting scenario (10^{-4} as opposed to 10^{-7}). This difference might have propagated through the choice of subgradients. For solutions of the iterations see Fig. 15.

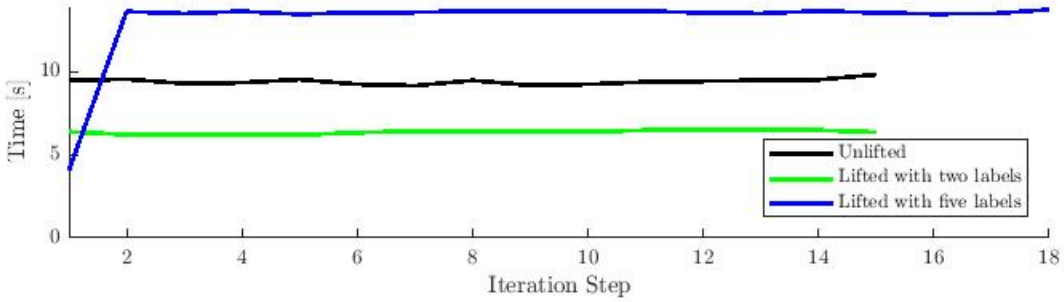
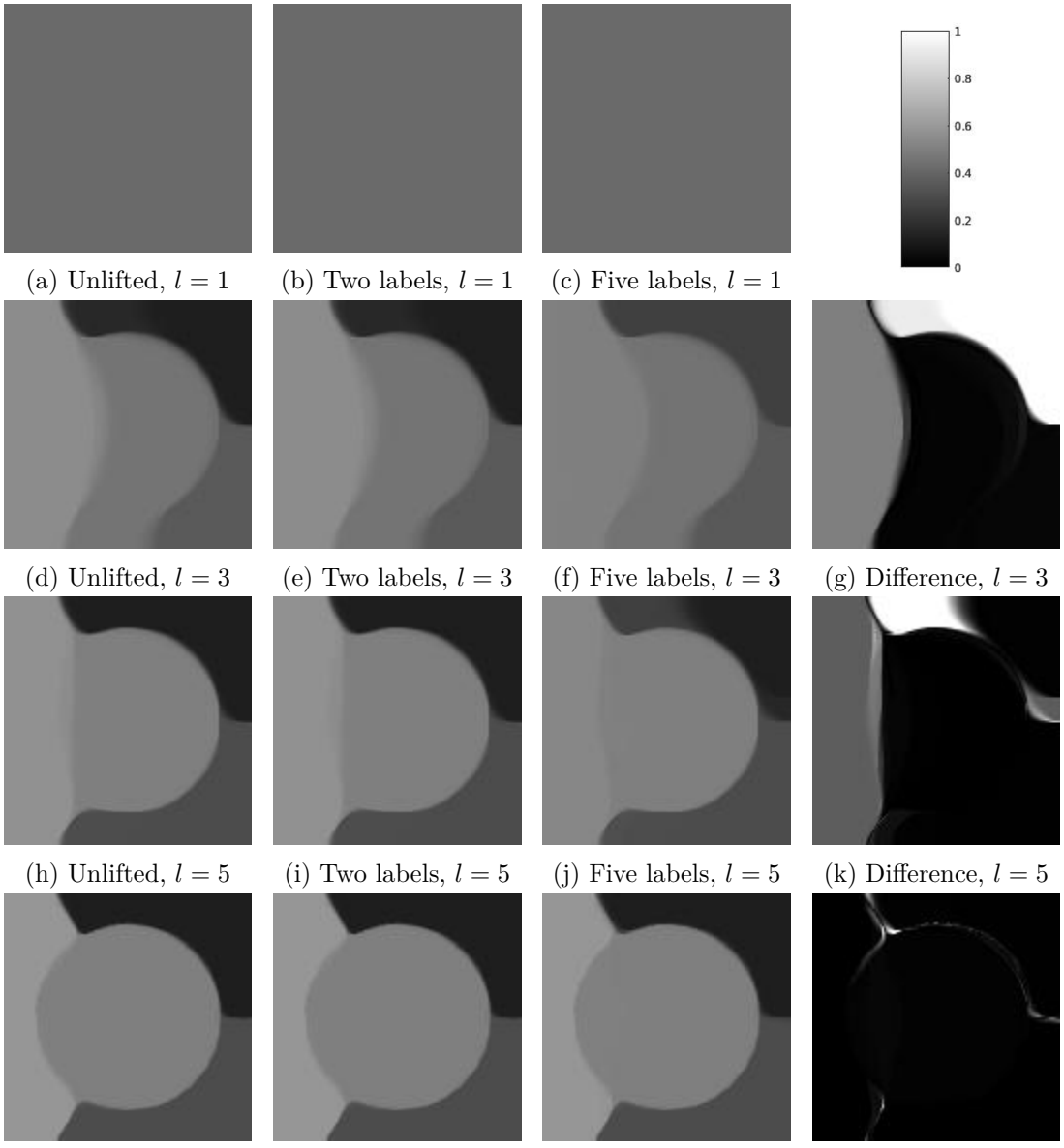


Figure 13: **Duration of iteration steps.** The plot depicts the runtime needed to solve the single Bregman iterations on the ROF problem. The two-label lifted Bregman iteration takes half as long as the five-label lifted one. Overall, the runtime does not change much over the iterations. The only exception is the runtime of the first iteration in the five-label lifting scenario. We do not have an explanation for this outlier.



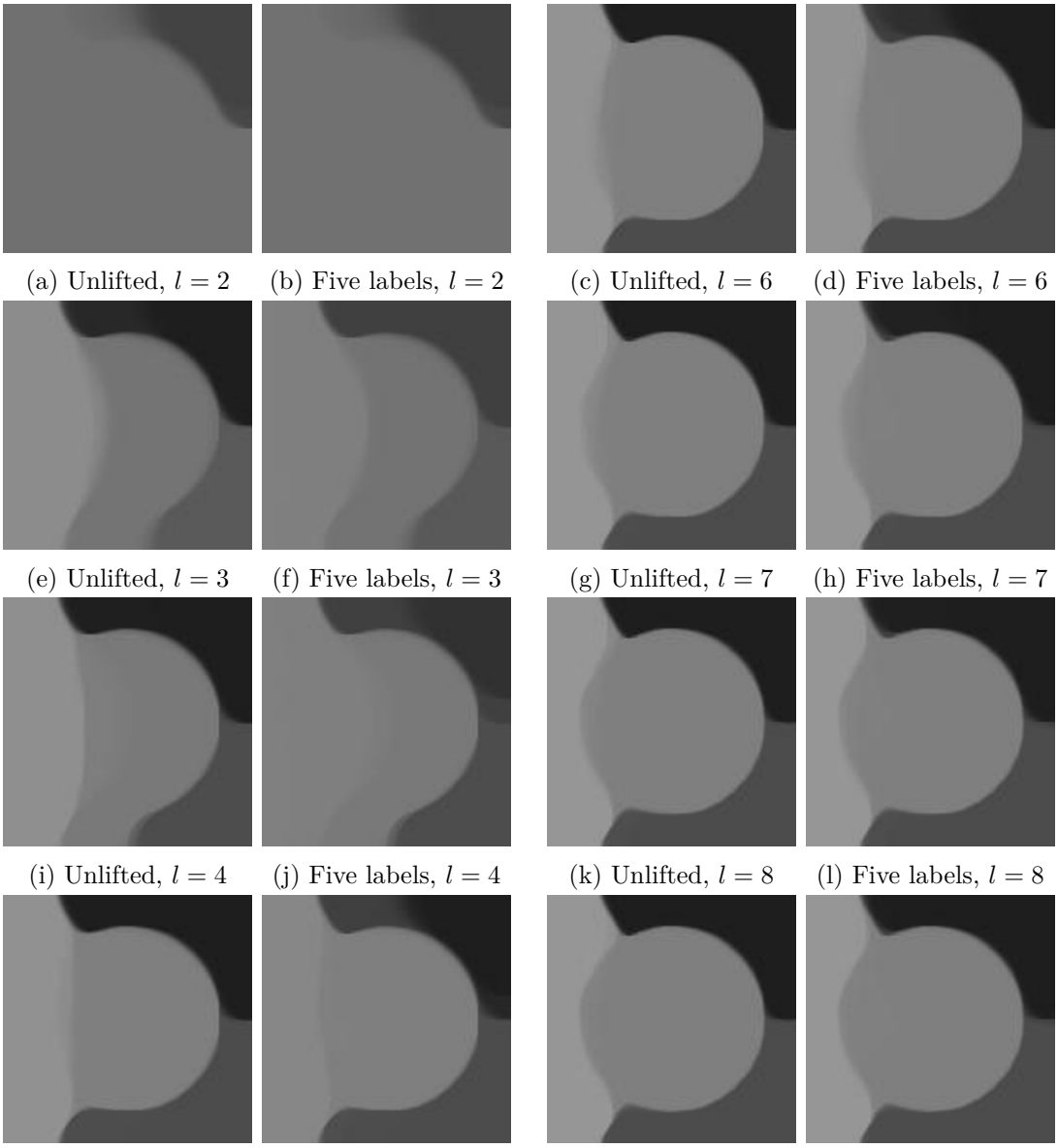
(a) Unlifted, $l = 1$ (b) Two labels, $l = 1$ (c) Five labels, $l = 1$

(d) Unlifted, $l = 3$ (e) Two labels, $l = 3$ (f) Five labels, $l = 3$ (g) Difference, $l = 3$

(h) Unlifted, $l = 5$ (i) Two labels, $l = 5$ (j) Five labels, $l = 5$ (k) Difference, $l = 5$

(l) Unlifted, $l = 15$ (m) Two labels, $l = 15$ (n) Five labels, $l = 15$ (o) Difference, $l = 15$

Figure 14: **Results of the Bregman iteration on ROF (1).** Solutions of the first, third, fifth, and last Bregman iteration on the unlifted (left column), two-label lifted (second column) and five-label lifted (third column) ROF problem. The right column shows the absolute difference of the two-label and five-label lifted iteration multiplied by factor ten. The solutions of the unlifted and two-label lifted problem show no visible difference. The Bregman iteration on the five-label lifted ROF problem needs three more steps until the iteration is terminated according to the discrepancy principle and some differences are visible in the third and fifth iteration. The plots have been generated using CVX (GB14, GB08), and the `prost` and `sublabel_relax` library (MLM$^+$15).



(a) Unlifted, $l = 2$ (b) Five labels, $l = 2$ (c) Unlifted, $l = 6$ (d) Five labels, $l = 6$

(e) Unlifted, $l = 3$ (f) Five labels, $l = 3$ (g) Unlifted, $l = 7$ (h) Five labels, $l = 7$

(i) Unlifted, $l = 4$ (j) Five labels, $l = 4$ (k) Unlifted, $l = 8$ (l) Five labels, $l = 8$

(m) Unlifted, $l = 5$ (n) Five labels, $l = 5$ (o) Unlifted, $l = 9$ (p) Five labels, $l = 9$

Figure 15: **Results of the Bregman iteration on ROF (2).** Solutions of the second to ninth Bregman iteration on the unlifted (first and third column) and five-label lifted (second and fourth column) ROF problem. In between the third and sixth iteration some basic contours are recovered faster in the unlifted than in the five-label lifted scenario. As of the seventh iteration, the solutions of the unlifted and five-label lifted problem show no more visible differences. The plots have been generated using CVX (GB14, GB08), and the `prost` and `sublabel_relax` library (MLM$^+$15).

Iterative Bregman Regularization on Lifted Stereo Matching

For our second example we consider the stereo matching model as a non-convex problem. Following (MLM$^+$15), we use a truncated sum of absolute gradient differences calculated on 4×4 patches as penalty term. We use two preprocessed images from (SHK$^+$14) as input. We solve the lifted Bregman iteration Alg. 6.1 with $\lambda = 20$ for four and eight labels respectively. Again, we use the forward Neumann difference operator and the PDHG solver.

In this scenario, we cannot use the discrepancy principle as a stopping criterion for the Bregman iteration, since we do not have a ground truth. Therefore, we stop the iteration after fifty steps. The runtime of the iteration steps is depicted in Fig. 16. The eight-label lifted Bregman iteration takes more than twice as long as the four-label lifted one. Whereas the ROF Bregman iteration steps were solved within seconds, the stereo matching iteration steps take around 4-9 minutes.

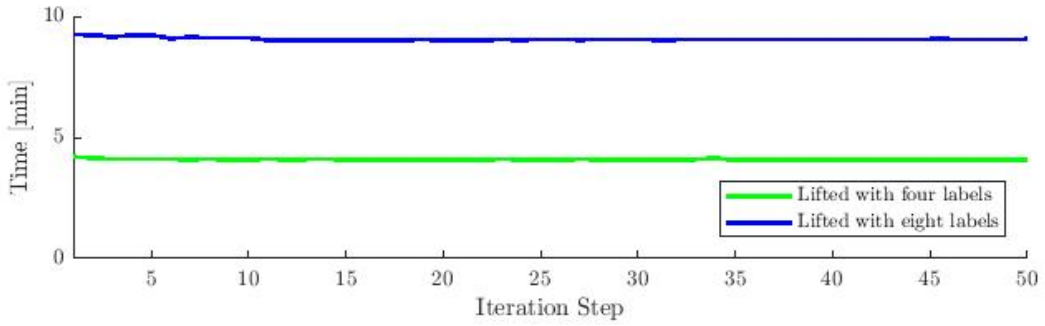
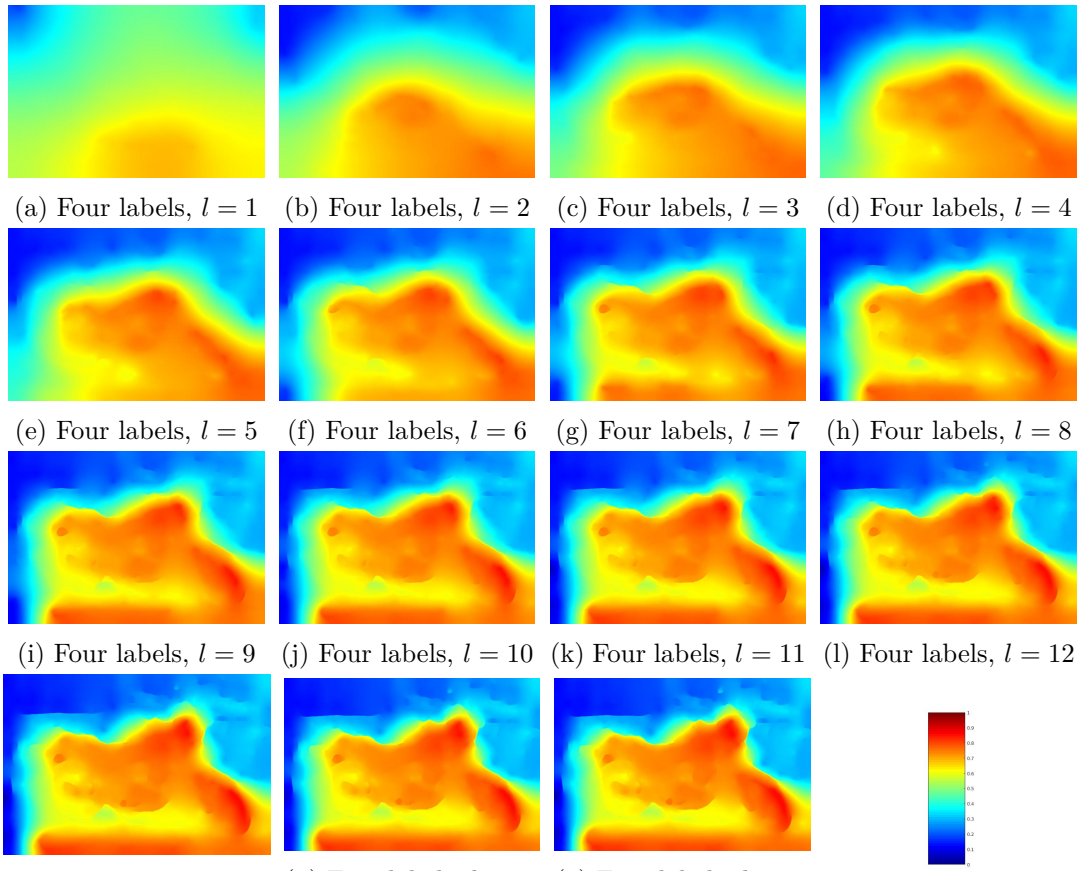


Figure 16: **Duration of iteration steps.** Runtime needed to solve single iterations on the stereo matching problem. In the eight-label lifting scenario, the runtime is approximately twice as long as in the four-label lifting scenario. The runtime almost does not change over the iterations.

The results of the four- and eight-label lifted Bregman iteration are shown in Fig. 17 and Fig. 18. Although we have not yet analyzed the lifted Bregman iteration on non-convex problems theoretically, the first experimental results look promising: The lifted Bregman iteration appears to present the same inverse scale-space properties on the stereo matching problem as on the ROF problem. The solution of the first iteration shows an oversimplified approximation of the actual depth and as the iteration proceeds more and more detail is added to this approximation. In the eight-label lifting case, details appear slightly sooner than in the four-label lifting case.



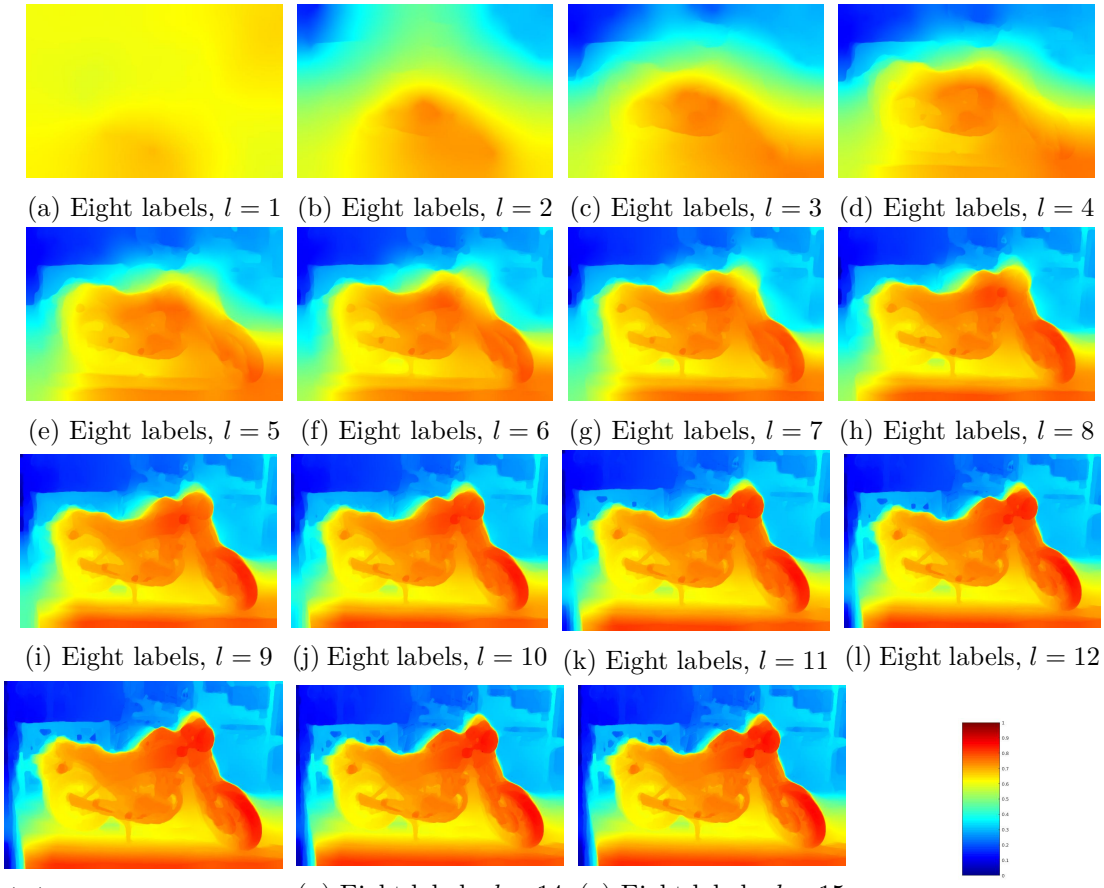
(a) Four labels, $l = 1$ (b) Four labels, $l = 2$ (c) Four labels, $l = 3$ (d) Four labels, $l = 4$

(e) Four labels, $l = 5$ (f) Four labels, $l = 6$ (g) Four labels, $l = 7$ (h) Four labels, $l = 8$

(i) Four labels, $l = 9$ (j) Four labels, $l = 10$ (k) Four labels, $l = 11$ (l) Four labels, $l = 12$

(m) Four labels, $l = 13$ (n) Four labels, $l = 14$ (o) Four labels, $l = 15$

Figure 17: **Results of the Bregman iteration on stereo matching (1).** Left to right, top to bottom: Solutions of the four-label lifted Bregman iteration on the stereo matching model. The first solution gives very vague depth information and as the iteration continues more and more details become visible. This reminds of inverse scale spaces. The iteration was solved using the `prost` and `sublabel_relax` libraries (MLM+15).



(a) Eight labels, $l = 1$ (b) Eight labels, $l = 2$ (c) Eight labels, $l = 3$ (d) Eight labels, $l = 4$

(e) Eight labels, $l = 5$ (f) Eight labels, $l = 6$ (g) Eight labels, $l = 7$ (h) Eight labels, $l = 8$

(i) Eight labels, $l = 9$ (j) Eight labels, $l = 10$ (k) Eight labels, $l = 11$ (l) Eight labels, $l = 12$

(m) Eight labels, $l = 13$ (n) Eight labels, $l = 14$ (o) Eight labels, $l = 15$

Figure 18: **Results of the Bregman iteration on stereo matching (2).** Left to right, top to bottom: Solutions of the first to fifteenth Bregman iteration on the eight-label lifted stereo matching model, solved with the libraries `prost` and `sublabel_relax` (MLM$^+$15). Consistent to the four-label lifted scenario, the first solution gives very vague information on the depth and as the iteration continues more and more details is gained. However, the details appear slightly sooner than in the four-label lifted scenario. Maybe the eight-label lifting gives a better approximation of the originally non-convex energy term, or the choice of different subgradients lead to better results.

Conclusion and Discussion

In this thesis, we proposed a novel approach that combines two powerful methods used in imaging. The iterative Bregman regularization (OBG$^+$05), which reduces the loss of contrast and results in inverse scale spaces, and functional lifting, which is used in order to find better – ideally globally optimal – minimizers of non-convex model, by approximating the original problem with a larger although convex model. In particular, we considered a functional lifting method, which allows for sublabel-accurate discretization (MLM$^+$15). By combining both methods, we obtained the lifted Bregman iteration, which allows to apply the iterative Bregman regularization to non-convex variational problems. We anticipated that the proposed method results in equivalent minimizers as the original Bregman iteration, when applied to convex problems.

We investigated the latter assumption both theoretically and experimentally. In Ch. 5, we took some first steps towards proving that both iterations lead to equivalent results on convex problems with total variation regularizer. Since the first iteration in both the old and the new method do not include the Bregman term, we took equivalency in this step for granted (PSG$^+$08, MLM$^+$15). We then argued, that for any subgradient used in the unlifted iteration, we should be able to find an equivalent lifted subgradient. In Thm. 5.3 we deduced a lifting for subgradients of the unlifted data-term integrands, which results in subgradients of the lifted data-term integrands. In order to complete the proof a few more steps are required, which we leave to future work. It still remains to be shown, that we can find a decomposition of the subgradient used in the unlifted iteration with respect to the integrands of the data term. Furthermore, it needs to be shown, that the use of this lifted subgradient in the lifted Bregman iteration results in a minimizer, which is equivalent to the minimizer of the unlifted Bregman iteration.

In the experimental section we examined two problems numerically. The convex ROF model on artificial input data and the non-convex stereo matching model on a real-life example. In order to solve the lifted Bregman iteration, we used the two libraries `prost` and `sublabel_relax` (MLM$^+$15).

We performed both the unlifted and lifted Bregman iteration on the ROF problem. For the lifted iteration we considered the case of two- and five-label lifting. The two-label lifting does in fact not imply a lifting, since the solution space remains of the same dimension. We used this case to estimate how much difference in the results should be attributed to machine inaccuracy and the use of different solvers – `cvx` (GB14, GB08) in the unlifted and PDHG (ZC08) in the lifted case. The results on the lifted and unlifted Bregman iteration were only for the two-label lifting equivalent up to machine accuracy. The five-label lifting resulted in a different convergence velocity, yet overall the five-label lifted Bregman iteration showed the same behavior as the unlifted one. We conjecture that the difference in convergence velocity is caused by the fact that the solutions of the first lifted and unlifted iteration were already not equivalent up to machine accuracy. For the five-label lifting we were not able to adjust the parameters of the PDHG solver as well as in the two-label lifting case. Due to the different first solutions it is standing to reason, that subgradients which are not equivalent were chosen

and used in the following iterations. In order to further investigate this issue, future experiments could include using lifted versions (Thm. 5.3) of the unlifted subgradients in the lifted Bregman iteration.

In our second experiment, we considered the stereo matching model as a non-convex variational problem. Although we have not yet analyzed this scenario theoretically, the first experimental results look promising. The lifted Bregman iteration appears to show the same inverse scale space flow properties as in the convex scenario. The first result is a vague approximation of the actual depth map and as the iteration continues the output acquires more and more detail.

We made some observations concerning the library `sublabel_accurate`, which are not directly related to the topic of this thesis, but might be of interest for future research. Due to the connection of the implemented lifting approach to the one introduced in Ch. 4.2.2, we have some expectations concerning the solution of a sublabel-accurate lifted problem: for the lifted formulation of an originally convex problem, we expect to find a minimizer in Q' defined by Eq. (69). Concerning originally non-convex problems, we expect to find a minimizer $\mathbf{u} = (\mathbf{u}_1, ..., \mathbf{u}_k)$ for which $\mathbf{u}_1 \geq \mathbf{u}_2 \geq ... \geq \mathbf{u}_k$ holds. This was only the case, when we solved minimization problems without regularization term. As soon as we added the total variation regularizer to the equation, the results did not take the expected forms anymore. This behavior is another possible explanation for the deviations in the five-label lifted ROF scenario. The possibility of adding another constraint to the sublabel-accurate lifting should be examined in the future.

In summary, the lifted Bregman iteration is an interesting new approach, which might extend the application field of the iterative Bregman regularization to non-convex problems. There are many open mathematical and numerical questions; under which conditions is the iteration well defined? How does the choice of the subgradient affect the result of the Bregman iteration? How can our proof in Ch. 5.2 be completed? Does the lifted Bregman iteration – for convex and non-convex problems alike – result in an inverse scale space flow in its continuous limit? Although there remain many open theoretical questions, the first experimental results look promising and entice to future research.

Appendix

Source Code Overview

Since the libraries `sublabel_relax` and `prost` are not well documented, we provide a short introduction for convenience reasons. As mentioned in chapter 6.1, `prost` is designed to solve variational problems of the form

$$\min_{u\in\mathbb{R}^n} \max_{v\in\mathbb{R}^m} g(u) + \langle Ku, v\rangle - f^*(v). \tag{163}$$

Here, K is a linear operator, $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $f^* : \mathbb{R}^m \to \overline{\mathbb{R}}$ are convex functionals. The `prost` library allows the user to describe a given variational problem quite naturally. This is done by following these steps:

1. Declaration of variables $u_i \in \mathbb{R}^{n_i}$

   $\to$ u_1 = prost.variable(n_1);

   $\to$ u_2 = prost.variable(n_2);

2. Declaration of sub-variables if necessary $vz = (v, z) \in \mathbb{R}^{n_1+n_2}$

   $\to$ vz = prost.variable($n_1 + n_2$);

   v = prost.sub_variable(vz, n_1);

   z = prost.sub_variable(vz, n_2);

3. Allocation of primal or dual status to the variables

   $\to$ problem = prost.min_max_problem($\{u, s\}$,$\{vz, q, p\}$);

   Minimizes over u and s

   Maximizes over vz, q and p

4. Description of variational problem

   $\to$ problem.add_function(q, prost.function.(...));

   Describes $g(u)$ and $f^*(v)$ in equation 163

   $\to$ problem.add_dual_pair(u, v, prost.block.(...));

   Describes $\langle Ku, v\rangle$ in equation 163, where K = prost.block.(...)

5. Choice of backend

   $\to$ backend = prost.backend.pdhg(...);

   $\to$ backend = prost.backend.admm(...);

6. Specification of solver options

   → opts = prost.options(...);

7. Solution of problem

   → solution = prost.solve(problem, backend, opts);

In the third step it is also possible to describe a primal or dual problem. In the fourth step we usually need to add multiple functions to describe the problem, since each added function can only depend on one variable. Both libraries provide a couple of implemented functions, such as the Euclidean or absolute difference. In order to add a dual pair we always need one primal and one dual variable. The variables are connected as a semi inner product, where the matrix given in the third argument is a prefactor of the primal variable. Both libraries provide some implemented matrices, such as sparse or the forward Neumann gradient operator.

The fourth step can be quite restrictive, since there is only a limited amount of implemented functions and matrices. In Alg. 6.1 we can replace the Bregman term with the equivalent expression $\langle \mathbf{q}_{l-1}, \nabla \mathbf{u} \rangle$, were $\mathbf{q}_{l-1}$ is a real-valued vector and $\mathbf{u}$ a primal variable. Since the difference operator ∇ is only implemented for primal dual pairs, we need to declare an auxiliary dual variable a with the size of $\mathbf{q}_{l-1}$. The restriction $a = \mathbf{q}_{l-1}$ can be specified as an auxiliary function. Only then the artificial dual variable a and the primal variable $\mathbf{u}$ can be combined. The formulation of $\langle \mathbf{q}_{l-1}, \nabla \mathbf{u} \rangle$ then presents as $\langle a, \nabla \mathbf{u} \rangle + \delta_{\mathbf{q}_{l-1}}(a)$, where $\delta_{\mathbf{q}_{l-1}}(a)$ is only zero for $a = \mathbf{q}_{l-1}$ and infinity otherwise.

References

[AFP00] Ambrosio, L. ; Fusco, N. ; Pallara, D.: *Functions of bounded variation and free discontinuity problems*. Oxford mathematical monographs, 2000

[AK06] Aubert, G. ; Kornprobst, P.: *Mathematical problems in image processing: partial differential equations and the calculus of variations*. Bd. 147. Springer Science & Business Media, 2006

[BCM05] Buades, A. ; Coll, B. ; Morel, J.: A review of image denoising algorithms, with a new one. In: *Multiscale Modeling & Simulation* 4 (2005), Nr. 2, S. 490–530

[BGM+16] Burger, M. ; Gilboa, G. ; Moeller, M. ; Eckardt, L. ; Cremers, D.: Spectral decompositions using one-homogeneous functionals. In: *SIAM Journal on Imaging Sciences* 9 (2016), Nr. 3, S. 1374–1408

[BGO+06] Burger, M. ; Gilboa, G. ; Osher, S. ; Xu, J. u. a.: Nonlinear inverse scale space methods. In: *Communications in Mathematical Sciences* 4 (2006), Nr. 1, S. 179–212

[Bre67] Bregman, L. M.: The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. In: *USSR computational mathematics and mathematical physics* 7 (1967), Nr. 3, S. 200–217

[CCP12] Chambolle, A. ; Cremers, D. ; Pock, T.: A convex approach to minimal partitions. In: *J. Imaging Sci.* 5 (2012), Nr. 4, S. 1113–1158

[CEN06] Chan, T. F. ; Esedoglu, S. ; Nikolova, M.: Algorithms for finding global minimizers of image segmentation and denoising models. In: *SIAM journal on applied mathematics* 66 (2006), Nr. 5, S. 1632–1648

[CHPA18] Correa, Rafael ; Hantoute, Abderrahim ; Pérez-Aros, Pedro: Complete characterizations of the subdifferential of convex integral functions II: Qualification conditions, conjugate and sequential formulae. In: *arXiv preprint arXiv:1804.10705* (2018)

[COS09] Cai, J. ; Osher, S. ; Shen, Z.: Linearized Bregman iterations for compressed sensing. In: *Mathematics of Computation* 78 (2009), Nr. 267, S. 1515–1536

[CP11] Chambolle, A. ; Pock, T.: A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. In: *Journal of Mathematical Imaging and Vision* 40 (2011), May, Nr. 1, S. 120–145

[CV01] Chan, T. F. ; Vese, L. A.: Active Contours Without Edges. In: *IEEE Trans. Image Proc.* 10 (2001), Nr. 2, S. 266–277

[EB92] Eckstein, J. ; Bertsekas, D. P.: On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. In: *Math. Prog.* 55 (1992), S. 293–318

[EZC10] Esser, E. ; Zhang, X. ; Chan, T. F.: A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. In: *SIAM Journal on Imaging Sciences* 3 (2010), Nr. 4, S. 1015–1046

[FR60] Fleming, W. H. ; Rishel, R.: An integral formula for total gradient variation. In: *Archiv der Mathematik* 11 (1960), Nr. 1, S. 218–222

[GB08] G., Michael ; B., Stephen: Graph implementations for nonsmooth convex programs. In: Blondel, V. (Hrsg.) ; Boyd, S. (Hrsg.) ; Kimura, H. (Hrsg.): *Recent Advances in Learning and Control*. Springer-Verlag Limited, 2008 (Lecture Notes in Control and Information Sciences), S. 95–110. – `http://stanford.edu/~boyd/graph_dcp.html`

[GB14] G., Michael ; B., Stephen: *CVX: Matlab Software for Disciplined Convex Programming, version 2.1*. `http://cvxr.com/cvx`, März 2014

[Gil17] Gilboa, G.: Semi-inner-products for convex functionals and their use in image decomposition. In: *Journal of Mathematical Imaging and Vision* 57 (2017), Nr. 1, S. 26–42

[GLY+13] Goldstein, T. ; Li, M. ; Yuan, X. ; Esser, E. ; Baraniuk, R.: Adaptive primal-dual hybrid gradient methods for saddle-point problems. In: *arXiv preprint arXiv:1305.0546* (2013)

[GO09] Goldstein, T. ; Osher, S.: The split Bregman method for L1-regularized problems. In: *SIAM journal on imaging sciences* 2 (2009), Nr. 2, S. 323–343

[Hes69] Hestenes, M.: Multiplier and gradient methods. In: *Journal of optimization theory and applications* 4 (1969), Nr. 5, S. 303–320

[IG98] Ishikawa, Hiroshi ; Geiger, Davi: Segmentation by grouping junctions. In: *cvpr* Bd. 98 Citeseer, 1998, S. 125

[Ish03a] Ishikawa, H.: Exact optimization for Markov random fields with convex priors. In: *Patt. Anal. Mach. Intell.* 25 (2003), Nr. 10, S. 1333–1336

[Ish03b] Ishikawa, H.: Exact optimization for Markov random fields with convex priors. In: *IEEE transactions on pattern analysis and machine intelligence* 25 (2003), Nr. 10, S. 1333–1336

[KL80] Kindermann, R. ; L., Snell J.: *Markov Random Fields and Their Applications*. American Mathematical Society, 1980

[Lel11] Lellmann, J.: *Nonsmooth convex variational approaches to image analysis*, University of Heidelberg, Diss., 2011

[LKY+09] Lellmann, J. ; Kappes, J. ; Yuan, J. ; Becker, F. ; Schnörr, C.: Convex multi-class image labeling by simplex-constrained total variation. In: *Scale Space Var. Meth.* Bd. 5567, 2009 (Springer LNCS), S. 150–162

[LLWS13] Lellmann, J. ; Lellmann, B. ; Widmann, F. ; Schnörr, C.: Discrete and Continuous Models for Partitioning Problems. In: *Int. J. Comp. Vis* 104 (2013), Nr. 3, S. 241–269

[Mey01] Meyer, Y.: *Oscillating patterns in image processing and nonlinear evolution equations: the fifteenth Dean Jacqueline B. Lewis memorial lectures*. Bd. 22. American Mathematical Soc., 2001

[MLM+15] Möllenhoff, T. ; Laude, E. ; Möller, M. ; Lellmann, J. ; Cremers, D.: Sublabel-Accurate Relaxation of Nonconvex Energies. In: *CoRR* abs/1512.01383 (2015)

[OBG+05] Osher, S. ; Burger, M. ; Goldfarb, D. ; Xu, J. ; Yin, W.: An iterative regularization method for total variation-based image restoration. In: *Multiscale Modeling & Simulation* 4 (2005), Nr. 22, S. 460–489

[OMDY11] Osher, S. ; Mao, Y. ; Dong, B. ; Yin, W.: Fast linearized Bregman iteration for compressive sensing and sparse denoising. In: *arXiv preprint arXiv:1104.0262* (2011)

[PC11] Pock, T. ; Chambolle, A.: Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In: *2011 International Conference on Computer Vision* IEEE, 2011, S. 1762–1769

[PCBC09] Pock, T. ; Cremers, D. ; Bischof, H. ; Chambolle, A.: An algorithm for minimizing the Mumford-Shah functional. In: *2009 IEEE 12th International Conference on Computer Vision* IEEE, 2009, S. 1133–1140

[PCBC10] Pock, T. ; Cremers, D. ; Bischof, H. ; Chambolle, A.: Global solutions of variational models with convex regularization. In: *SIAM Journal on Imaging Sciences* 3 (2010), Nr. 4, S. 1122–1145

[PSG+08] Pock, T. ; Schoenemann, T. ; Graber, G. ; Bischof, H. ; Cremers, D.: A convex formulation of continuous multi-label problems. (2008), S. 792–805

[Roc70] Rockafellar, R. T.: *Convex analysis*. Bd. 28. Princeton university press, 1970

[ROF92] Rudin, L. ; Osher, S. ; Fatemi, E.: Nonlinear total variation based noise removal algorithms. In: *Physica D: Nonlinear Phenomena* 60 (1992), November, Nr. 1-4, S. 259–268

[RW09] Rockafellar, R. T. ; Wets, R. J.: *Variational analysis*. Bd. 317. Springer Science & Business Media, 2009

[SCD+06] Seitz, S. M. ; Curless, B. ; Diebel, J. ; Scharstein, D. ; Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* Bd. 1 IEEE, 2006, S. 519–528

[SGG+09] Scherzer, O. ; Grasmair, M. ; Grossauer, H. ; Haltmeier, M. ; Lenzen, F.: *Variational methods in imaging*. Springer, 2009

[SHK+14] Scharstein, D. ; Hirschmüller, H. ; Kitajima, Y. ; Krathwohl, G. ; Nešić, N. ; Wang, X. ; Westling, P.: High-resolution stereo datasets with subpixel-accurate ground truth. In: *German conference on pattern recognition* Springer, 2014, S. 31–42

[YOGD08] Yin, W. ; Osher, S. ; Goldfarb, D. ; Darbon, J.: Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing. In: *SIAM Journal on Imaging sciences* 1 (2008), Nr. 1, S. 143–168

[ZC08] Zhu, M. ; Chan, T.: An efficient primal-dual hybrid gradient algorithm for total variation image restoration. In: *UCLA CAM Report* 34 (2008)

[ZGFN08] Zach, C. ; Gallup, D. ; Frahm, J.-M. ; Niethammer, M.: Fast global labeling for real-time stereo using multiple plane sweeps. In: *Vis. Mod. Vis*, 2008, S. 243–252