Master's Thesis

Spatially-Adapted Regularization
for
Inverse Problems

Lucas Plagwitz

Matr. 418502

Supervisor: Prof. Dr. Benedikt Wirth
Assisting Supervisor: Dr. Frank Wübbeling

Faculty of Mathematics and Computer Science, University of Münster, Germany

April 1st, 2021

Abstract

The present work deals with enhanced regularization techniques in the context of imaging problems. Besides image denoising, two inverse problems of this domain are the focus of this work: the deconvolution and the reconstruction of computerized tomography measurements. A direct inversion is not feasible due to the ill-posedness of the underlying forward operator. The construction of an optimization problem, which balances between faithfulness to the data and regularizing property, is the main venture of this work. A first step is to establish the mathematical foundation for investigations of the underlying operators, developing a suitable regularization property, and implement a solving algorithm. The basis for this is the total variation regularization, which is widespread in the context of imaging.

Furthermore, this work focuses on a spatially-adapted weighting between data fidelity and regularization term. As images typically consist of various objects at different scales, this also motivates the locally-dependent weighting. I.e., a smooth surface needs less data fidelity than a fine pixel-structure. In addition to the derivation of this approach, comparisons to another enhancement, i.e. the Bregman-based TV-regularization, is made. It will turn out that the comparability of both is limited, since they improve objecting problems, respectively. A new combination of both ways is described to obtain the best possible reconstruction. Finally, simulations of this novel and competing methods over the three imaging problems, which are based both on real and simulated data, are presented.

Contents

CHAPTER 1

Introduction

The reconstruction of an unknown cause from a known effect is often denoted *inverse problem*. These problems play a fundamental role in various and interdisciplinary fields, including geophysics, computer science, and medicine. Inverse problems are typically present when a quantity cannot be measured directly but only indirectly. One prominent example may be the calculation of a mass distribution from measured gravitational forces. Further examples include the decrease of X-ray intensities in absorbing materials, which is known as Beer's law and takes an important role throughout this work. The inverse problem in this regard constitutes the determination of density and further concentrations from tomography data.

The main difficulty of inverse problems is their fundamental non-linear nature: Two distinct causes can produce similar effects. In addition, the underlying measurements are in most cases not perfect and include noise and inaccuracies, hence creating a highly non-trivial problem. Therefore, an effect z is represented via an operator $A : X \to Z$

$$z = Au + \eta, \tag{1.1}$$

where X, Z represent space of cause and effect respectively and η denotes the noise related to the underlying forward problem. The described sensitivity of the cause u, which is based on the measurement z, is formally described as an ill-posed problem.

The fundamental definition of such ill-posed problems reaches back to the French mathematician Jacques Hadamard [1]. According to his definition, a problem is called ill-posed if one of the conditions

1. existence of a solution u for every $z \in Z$
2. uniqueness of those solutions
3. continuity of the inverse map $A^{-1} : Z \to X$

is not fulfilled.

Therefore, direct reconstruction of the cause via an inversion formula A^{-1} is not advisable.

In this work, an alternative solution of inverse problems is investigated. This alternative reformulates the inverse problem by using an optimization scheme

$$\operatorname*{arg\,min}_{u \in X} \; F(u) + G(Ku). \tag{1.2}$$

Here, $F : X \to \mathbb{R} \cup \{\infty\}$ is called fidelity term as it describes the proximity to the measurement z.[1] The regularization $G : Y \to \mathbb{R} \cup \{\infty\}$ only depends on the solution sought and, in addition, enforces certain wanted properties. The latter are expressed by a linear operator $K : X \to Y$ and the associated space Y. For special problems, this regulator requires a more versatile approach, e.g., spatially-dependent fidelity to regularization weights and this work focuses heavily on such.

This thesis is structured as followed: In chapter 2 theoretical foundations are laid, followed by the explicit formulation of the spatially-adaptive regularization for inverse problems in chapter 3. The penultimate section 4 is devoted to the numerical implementation and their results. Finally, we conclude in section 5.

1.1. Computerized Tomography

In 1901, the physicist *Wilhelm Conrad Röntgen* was awarded the Nobel prize for the discovery of a new type of radiation, the so-called X-rays. When fast electrons are decelerated in an electromagnetic field, highly energetic photons are released [2]. The wavelength of these photons reach from 10^{-8} m to 10^{-13} m, depending on the deceleration force, i.e., the voltage between anode and cathode. This type of radiation can easily penetrate matter and, depending on its type and certain physical quantities like density, they are weakened differently as described by Beer's law

$$\bar{y} \triangleq \mathbb{E}(y) = I_0 e^{-\int_l u(x) \mathrm{d}x}. \tag{1.3}$$

Here, I_0 denotes the intensity of the incident X-ray beam, y the number of detected X-ray photons, and u the function measured along the line integral defined by the ray l.

This process is, however, imperfect. Two main error sources can be identified. First, quantum noise arises that can be, ignoring the energy variability of different photons, modeled via a Poisson distribution. Secondly, the electronic noise is given by a Gaussian distribution. As offsets can be directly measured by a so-called dark scan and then substracted from the original data, it is possible to assume zero-mean. As consequence, the measured

---

[1] In case of the presented form of an inverse problem, F is usually expressed as $\|Au - z\|$ for a suitable norm $\|\cdot\|$.

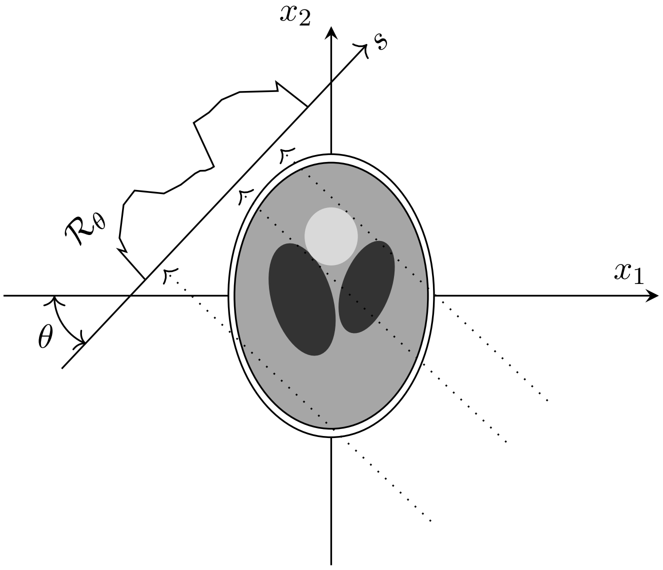


Figure 1.1.: Sketch of Radon transform for $\theta = 0.26\pi$. The dashed lines symbolize a portion of the X-ray beams directed toward the detectors. The brighter the areas are, the fewer X-rays are detected.

photon intensity can be written as

$$y \sim \text{Poisson}(\bar{y}) + \mathcal{N}(0, \sigma^2).$$

Since the resulting function $u(x)$ is scalar, multiple X-ray measurements have to be connected (typically this involves a circular arrangement of various detectors or circular moving detectors) to finally deduct a two-dimensional X-ray image. Assuming $u \in L^2(\mathbb{R}^2)$ describes the function to be measured, the Radon transform $\mathcal{R}$ (see Fig. 1.1) describes the integration over all directions $\theta \in [0, \pi]$

$$(\mathcal{R}u)(\theta, s) = \int_{x \cdot \theta = s} u(x) \mathrm{d}\sigma(x), \tag{1.4}$$

where the integral is taken with respect to the measure $\mathrm{d}\sigma$.

In the next chapter the Radon operator and its associated reconstruction are discussed in detail. As it turns out, the Radon transform is, according to Hadamard, ill-posed and therefore leads to an amplification of errors. In the following, we restrict ourselves to additive noise, neglecting multiplicative contributions, hence accounting only for electronic noise.

1.2. Image Processing

In addition to computerized tomography, there are other image-based problems that can be solved with Eq. (1.2). By having the same target structure, i.e. an image, many concepts can be adopted between these problems. In general, an image u is described as an element in the space $L^2(\mathbb{R}^2)$. In the following chapter, we construct a decomposition of u into one part that contains the basic structures and another containing high oscillations. This procedure is necessary to explicitly formulate a mathematical framework. Furthermore, it allows to separate the high noise oscillations more easily.

In reality, an image exists only in its discrete form: an image is merely the arrangement of different pixel values. We consider a Cartesian grid of size $N_1 \times N_2$

$$\{(i,j) : 1 \leq i \leq N_1, 1 \leq j \leq N_2\},$$

where (i,j) denote the indices and pixel locations. A discrete space for images can then be defined via the finite dimensional vector space $X = \mathbb{R}^{N_1 N_2}$ together with the standard scalar product.

In order to describe the oscillation properties, a gradient on X can be defined. In this discrete setting the gradient is a vector in $Z = X \times X$

$$\nabla : X \to Z \qquad (\nabla u)_{i,j} = \begin{pmatrix} (\nabla u)^1_{i,j} \\ (\nabla u)^2_{i,j} \end{pmatrix},$$

$$(\nabla u)^1_{i,j} = \begin{cases} \frac{u_{i+1,j} - u_{i,j}}{h} & i < N_1 \\ 0 & i = N_1 \end{cases}, \qquad (\nabla u)^2_{i,j} = \begin{cases} \frac{u_{i,j+1} - u_{i,j}}{h} & j < N_2 \\ 0 & j = N_2 \end{cases}.$$

The next step is to set up a minimization problem that seeks the optimal solution between data fidelity and small gradient norm. The gradient norm is defined as

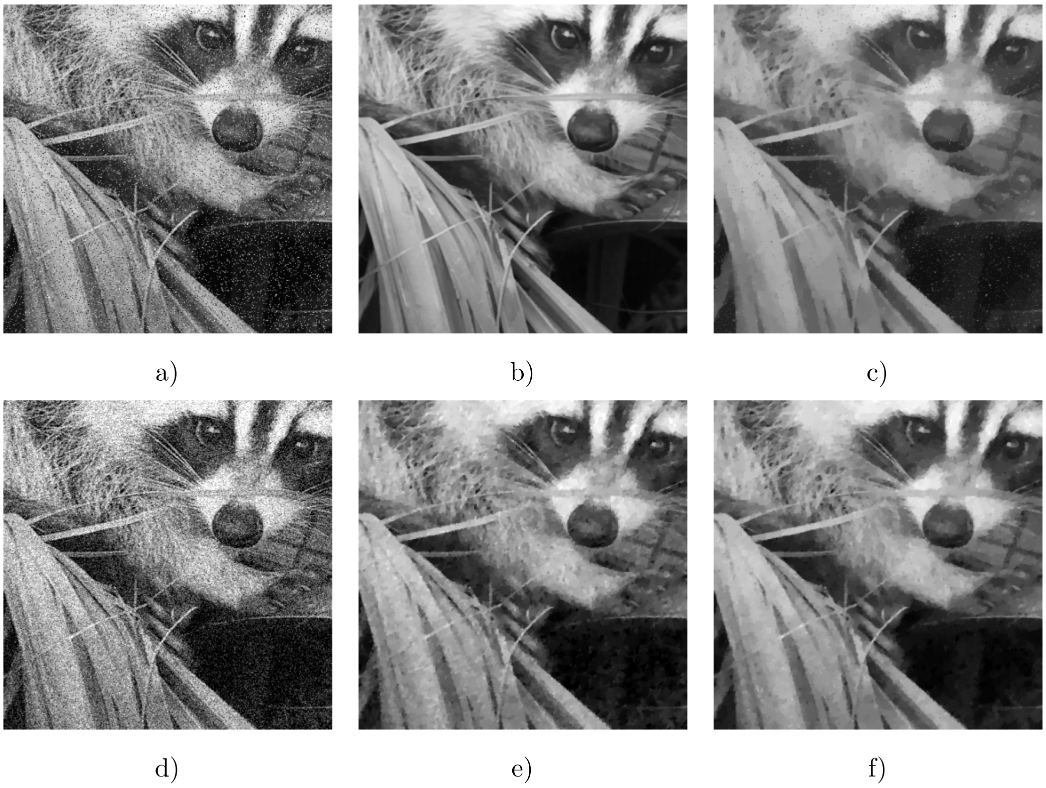
$$\|\nabla u\|_1 = \sum_{i,j} |(\nabla u)_{i,j}|, \qquad |(\nabla u)_{i,j}| = \sqrt{((\nabla u)^1_{i,j})^2 + (\nabla u)^2_{i,j})^2}.$$

This constitutes the discrete version of the so-called *isotropic total variation norm.* The main benefit of said norm when compared to, e.g. the 2-norm[2], lies in the preservation of the image's edges, since it penalizes large and small fluctuations of gradients equally. Altogether, this defines the problem

$$\operatorname*{arg\,min}_{u \in X} \; F(u) + \|\nabla u\|_1 . \tag{1.5}$$

---

[2] The terms 1-norm and 2-norm are used here for the discrete version of the norms in $L^1(\Omega)$ and $L^2(\Omega)$, i.e. $\|\cdot\|_1 := \sum_{i,j} |\cdot|$, $\|\cdot\|_1 := \sqrt{\sum_{i,j} (\cdot)^2}$.



	PSNR	MAE
a) Salt-pepper noise	15.27	0.047
b) L^1-fidelity	28.47	0.018
c) L^2-fidelity	22.69	0.053

	PSNR	MAE
d) Gaussian noise	13.98	0.159
e) L^1-fidelity	23.93	0.046
f) L^2-fidelity	24.66	0.041

Figure 1.2.: Comparison between different fidelity terms on salt-pepper and Gaussian distributed noise. Salt-pepper noise with a balanced amount of 0.1, see subfigures a)-c), and pixel by pixel Gaussian noise $\mathcal{N}(0, \frac{1}{25})$, see subfigures d)-f), are shown. Image taken from Ref. [3].

One prominent example in image processing is the denoising of images. This denoising can also be formulated via the initial equation of inverse problems (1.1), where the forward operator can be neglected

$$z = u + \eta.$$

In contrast to the regularization, the norm of the fidelity term depends more strictly on the underlying noise η. Here, one makes use of the 1-norm $F(u) = \lambda \|u - w\|_1$ or the 2-norm $F(u) = \frac{\lambda}{2} \|u - w\|_2^2$ with some weight $\lambda > 0$ as shown in Fig. 1.2. The figure summarizes the behavior of both norms on Gaussian and salt-pepper noise. The results show, that large deviations in pixels values between u and z are, in case of salt-pepper noise, penalized more heavily in the 2-norm than in the 1-norm. Gaussian noise is, on the other hand, penalized

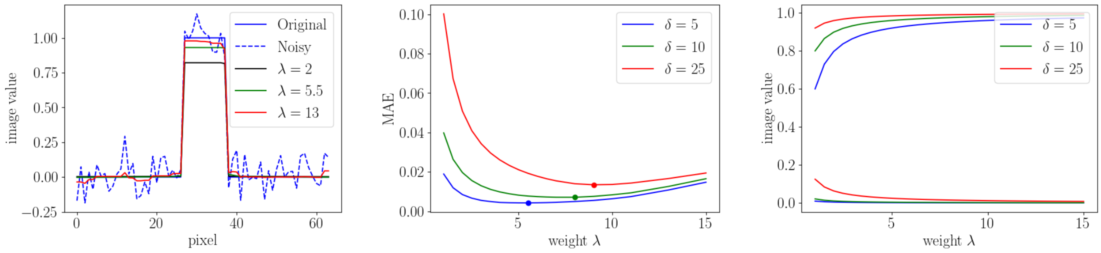


Figure 1.3.: Interaction between width and weighting of the optimization example 1.1. The subfigure on the left shows a cross section of the two-dimensional object with $\delta = 5$. Different weights are applied to the noisy data with $\eta_{i,j} \sim \mathcal{N}(0, \frac{1}{100})$. The underlying images are illustrated in Fig. C.1. In the middle subfigure the behavior of the mean absolute error (MAE) in relation to three different widths is shown. The optimal λ is marked by a dot. The subfigure on the right shows the behavior of the contrast depending on width and weighting. The upper half corresponds to h_{in}, whereas the lower half corresponds to h_{out}.

more in the 1-norm than in the 2-norm. As Gaussian noise will be the main focus of this work, we focus on L^2-data-fidelity in the following.

A main question of this work is to find the appropriate weighting λ. Even though higher noise leads to lower data fidelity, the following example shows basic problems of denoising via total variation. We simplify the problem by considering a single and simple image object.

Example 1.1. Let $(N_1, N_2) = (64, 64)$. Furthermore, $z \in \mathbb{R}^{64\times 64}$ defines the image with

$$z = \chi_{\Omega_\delta} \qquad \Omega_\delta := \{(32-\delta, 32+\delta) \times (32-\delta, 32+\delta)\} \subset \mathbb{R}^{64\times 64} \quad 32 > \delta \in \mathbb{N}.$$

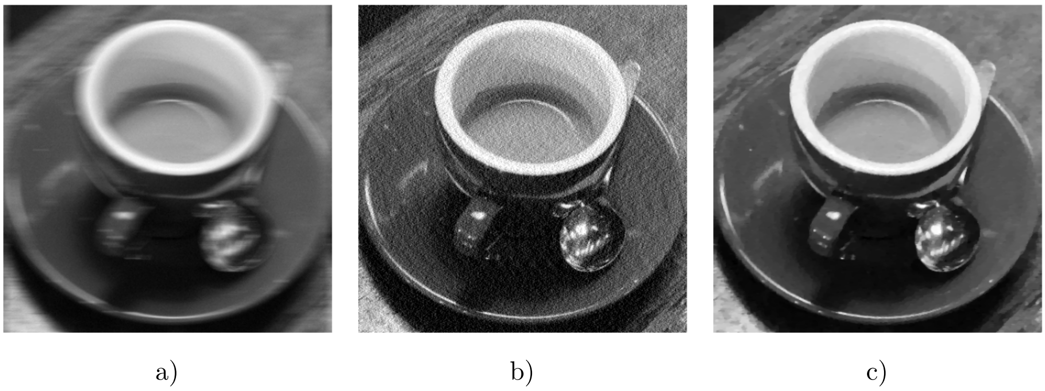
We now consider the regularization with L^2-fidelity of the unperturbed image. It can be assumed that the solution u consists of two parts, h_{out} outside of Ω_δ and h_{in} inside of Ω_δ, with $h_{\text{out}} \leq h_{\text{in}}$. Thus, the equation (1.5) results in[3]

$$\begin{gathered}
\arg\min_u \frac{\lambda}{2} \|u - z\|_2^2 + \|\nabla u\|_1 \\
\Leftrightarrow \arg\min_u \frac{\lambda}{2} \cdot (4\delta^2 \cdot (1-h_{in})^2 + (64^2 - 4\delta^2) \cdot h_{\text{out}}^2) + 8\delta \cdot (h_{\text{in}} - h_{\text{out}}) \\
\frac{\partial}{\partial h_{\text{in}}} = 4\lambda\delta^2 \cdot (h_{\text{in}} - 1) + 8\delta \qquad \frac{\partial}{\partial h_{\text{out}}} = \lambda(64^2 - 4\delta^2) \cdot h_{\text{out}} - 8\delta \\
h_{\text{in}} = 1 - \frac{2}{\lambda\delta} \qquad h_{\text{out}} = \frac{2\delta}{\lambda(32^2 - \delta^2)}.
\end{gathered}$$

---

[3] To keep the model as simple as possible, the isotropic total variation term has been slightly simplified here.

The levels $h_{\text{out}}, h_{\text{in}}$ are thus dependent on the width δ and on the weighting λ between regularization and faithfulness. The smaller the width and the faithful term, the smaller h_{in} becomes. This effect leads to a general loss of contrast. It also hints that objects of different scales should be weighted differently. Fig. 1.3 illustrates this behavior.


a) b) c)

	PSNR	MAE
a) Degraded	22.87	0.039
b) Minimum-Norm	18.84	0.091
c) TV	29.75	0.018

Figure 1.4.: Motion-blurred reconstruction. Subfigure a) contains Gaussian noise with $\eta_{i,j} \sim \mathcal{N}(0, \frac{1}{100^2})$. While subfigure b) displays a minimum-norm solution, the image in subfigure c) is regularized with the total variation term. Image taken from Ref. [4].

Another field of application for total-variation-based regularization is the deconvolution of images [5]. Images can be degraded by, for example, blur or motion. These effects can be represented by problem (1.1) with a convolution operator A and associated kernel k_A, i.e. $Au = k_A * u$. For an efficient image reconstruction one needs to determine the convolution kernel k_A. It is conceivable to obtain this information from, e.g., internal sensors like the embedded gyroscope of a smartphone [6], or directly on image-based estimations via sharp edges [7].

Even if the kernel is known, the reconstruction of the original image is not guaranteed. To illustrate this, consider the convolution theorem

$$\mathcal{F}(k_A * u) = 2\pi \cdot \mathcal{F}(u) \cdot \mathcal{F}(k_A).$$

A further description and proof can be found in theorem 2.15. Here, $\mathcal{F}$ denotes the Fourier transform. A straightforward reconstruction reads

$$u = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(z)}{\mathcal{F}(k_A)}\right).$$

At first glance, the reconstruction becomes unsafe for $\mathcal{F}(k_A) \to 0$. In subsection 2.2.1 we prove the ill-posedness of the underlying convolution operator. This is the reason why we will use a regularization approach to solve

$$\arg\min_{u \in X} \frac{\lambda}{2} \left\| Au - z \right\|^2 + \left\| \nabla u \right\|_1 ,$$

where A is the convolution operator. Fig. 1.4 shows a total-variation-based reconstruction of a deblurred image compared to a minimum-norm solution.

As we will show in this work, the here posed regularization scheme builds the foundation for spatially-adapted regularization, which we will then generalize for the Radon operator.

CHAPTER 2

Mathematical Foundations

In this chapter the mathematical basics are introduced. First, we focus on the optimization equation (1.2). In this context, the necessary theory of convex analysis is discussed. Secondly, the underlying operators are examined, with a focus on the ill-posedness according to Hadamard. As a third point, a suitable space for images and a resulting regularization functional is introduced. Lastly, an extended regularization technique with the iterative Bregman divergence is presented.

2.1. Convex Analysis

Solving the optimization equation (1.2) depends significantly on the properties of F and G. Differentiability is a far too restricting requirement with respect to the 1-norm. Therefore, we cannot make use of basic constructs like the gradient. As it turns out however, convexity is always given in our formulation. Hence, we need to investigate the theory of optimization over convex functions. The goal is to obtain an algorithm that converges to the optimum of the sum of two convex functions.

All basic terms are formulated for a general function $f : \mathcal{H} \to \mathbb{R} \cup \{\infty\}$ over a Hilbert space $\mathcal{H}$. At the end of this section, the switch back to the functions F and G is then straightforwardly possible.

Definition 2.1.

1. f is called *convex*, if

   $$f\left(\lambda x + (1-\lambda)y\right) \leq \lambda f(x) + (1-\lambda) f(y)$$

   for $x, y \in \mathcal{H}$ and $\lambda \in [0,1]$.
2. f is called *proper*, if the domain of f $\operatorname{dom}(f) \coloneqq \{x \in H \mid f(x) < \infty\}$ is not empty.
3. f is called *lower semi-continuous* at $x \in \mathcal{H}$, if $\liminf_{n\to\infty} f(x_n) \geq f(x)$ for every sequence $(x_n)_n$ with $x_n \to x$.

To obtain a condition of optimality, we define the subdifferential on the basis of reference [8].

Definition 2.2 (Subdifferential)**.** Let $f : \mathcal{H} \to \mathbb{R} \cup \{\infty\}$ be a proper function. The *subdifferential* is the set-valued operator

$$\partial f : \mathcal{H} \to 2^{\mathcal{H}} \qquad x \mapsto \{u \in \mathcal{H} \mid f(y) \geq f(x) + \langle y - x, u \rangle \quad \forall y \in \mathcal{H}\}.$$

Elements of $\partial f(x_0)$ are called *subgradients* of f at x_0.

In case of convexity $\partial f(x)$ is non-empty for all $x \in \operatorname{dom}(f)$. In general, the subdifferential determines all possible slopes of an affine lower bound that is touching the respective point. The connection to a derivation in the differentiable case becomes clear. If f is proper, convex, and Gâteaux-differentiable[1] in x_0 then

$$\partial f(x_0) = \{\nabla f(x_0)\}.$$

Via the subdifferential, we receive an adjustment of the first order optimality condition for convex non-smooth functions. This is known as *Fermat's rule*

$$\begin{aligned} x \in \operatorname*{arg\,min}_{u \in \mathcal{H}} f(u) &\Leftrightarrow f(y) \geq f(x) + \langle y - x, 0 \rangle \text{ for all } y \in \mathcal{H} \\ &\Leftrightarrow 0 \in \partial f(x). \end{aligned}$$

Furthermore, this lead to a characterization for the following operator.

Definition 2.3 (Proximal operator)**.** Let $f : \mathcal{H} \to \mathbb{R} \cup \{\infty\}$ be convex, lower semi-continuous, and proper. The *proximal operator* $\operatorname{Prox}_f : \mathcal{H} \to \mathcal{H}$ is given by

$$x \mapsto \operatorname*{arg\,min}_{u \in \mathcal{H}} \left(\frac{1}{2} \|x - u\|^2 + f(u) \right).$$

Intuitively, the proximal operator takes a step towards the optimum, while large distances to x are penalized. This principle is similar to the well-known gradient descent. However, Prox_f is also defined for non-smooth functions. At first glance, this seems to be of little help to the problem, since we exchange one optimizing problem with another. On the other hand, by taking a closer look at the following subdifferential characterization, it is possible to calculate the proximal operator.

Proposition 2.4. Let $f : \mathcal{H} \to \mathbb{R} \cup \{\infty\}$ be convex, lower semi-continuous, and proper with $\operatorname{arg\,min} f \neq \emptyset$. Then,

---

[1] A function f is called Gâteaux-differentiable in $x_0 \in \operatorname{dom}(f)$, if there exists a unique $\nabla f(x) \in \mathcal{H}$ such that for any $y \in \mathcal{H}$ the directional derivative is given by $\lim_{\alpha \searrow 0} \frac{1}{\alpha}(f(x + \alpha y) - f(x)) = \langle y, \nabla f(x) \rangle$.

1. for $x \in \mathcal{H}$ holds
$$p = \operatorname{Prox}_f(x) \;\Leftrightarrow\; x - p \in \partial f(p).$$
2. $x \in \arg\min f \;\Leftrightarrow\; x = \operatorname{Prox}_f(x)$
3. $x \notin \arg\min f \;\Rightarrow\; f(\operatorname{Prox}_f(x)) < f(x)$
4. the sequence $(x^{(k)})_k$
$$x^{(k+1)} = \operatorname{Prox}_{\gamma f}(x^{(k)})$$
is a minimizing sequence of f and converges weakly to some point in $\arg\min f$, where $\gamma > 0$ and $x^{(0)} \in \mathcal{H}$.

Proof.

1. See reference [9].
2. $x = \operatorname{Prox}_f(x) \Leftrightarrow x \in x + \partial f(x) \Leftrightarrow 0 \in \partial f(x) \overset{\text{Fermat's rule}}{\Longleftrightarrow} x \in \arg\min f$
3. By assumption $x \notin \arg\min f$. Let $p = \operatorname{Prox}_f(x)$. With 2. $p \neq x$
$$f(x) = f(x) + \frac{1}{2}\|x - x\|^2 > \frac{1}{2}\|x - p\|^2 + f(p),$$
which implies $f(x) - f(p) > \frac{1}{2}\|x - p\|^2 > 0$.
4. See references [9, 10].

□

The iteration rule from Prop. 2.4.4. leads to the so-called *proximal point algorithm*. With respect to Eq. (1.2), a calculation of $\operatorname{Prox}_{F(u)+G(Ku)}$ is difficult to perform directly. Further, terms of F and G with respect to K have to be considered separately.

Definition 2.5 (Fenchel-Legendre conjugate)**.** Let $f : \mathcal{H} \to [-\infty, \infty]$. The *Fenchel-Legendre conjugate* of f is

$$f^* : \mathcal{H} \to [-\infty, \infty], \qquad u \mapsto \sup_{x \in \mathcal{H}} \langle x, u\rangle - f(x).$$

The bi-conjugate of f is $(f^*)^* = f^{**}$.

It can be shown that $f^{**} \leq f$. In addition, f^{**} is the pointwise supremum over all continuous affine lower bounds on f. Figure 2.1 shows a sketch of the introduced basic definitions.

Proposition 2.6 (Fenchel-Young inequality)**.** Let $f : \mathcal{H} \to \mathbb{R} \cup \{\infty\}$ be proper. Then for all $x, u \in \mathcal{H}$:
$$f(x) + f^*(u) \geq \langle x, u\rangle.$$

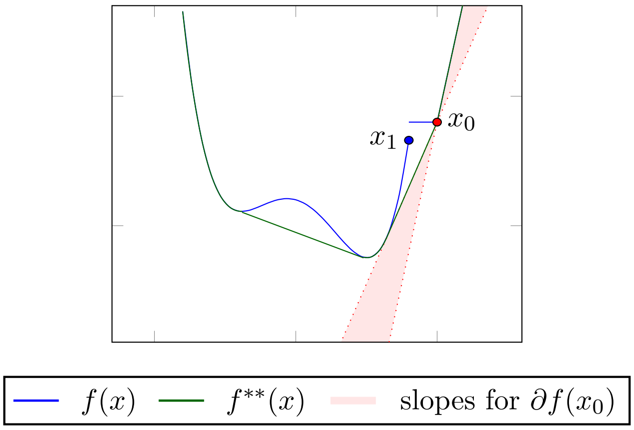


Figure 2.1.: Basic definitions in convex analysis. Since f is a non-convex function, there are regions where the subdifferential is empty. At x_1 a lower semi-continuous point is present. The red area marks all possible elements in $\partial f(x_0)$. The bi-conjugate f^{**} generates the convex lower bound.

Proof. Let $x, u \in \mathcal{H}$. Since f is proper, $f^* > -\infty$. If $f(x) = \infty$, the inequality holds obviously. If, on the other hand, $f(x) < \infty$, then

$$f^*(u) = \sup_{y \in \mathcal{H}} \langle u, y \rangle - f(y) \geq \langle u, x \rangle - f(x).$$

□

The following theorem states a connection between ∂f and ∂f^*.

Proposition 2.7. Let $f : \mathcal{H} \to \mathbb{R} \cup \{\infty\}$ be proper, lower semi-continuous, and convex. Let $x, u \in \mathcal{H}$. Then

$$u \in \partial f(x) \quad \Leftrightarrow \quad f(x) + f^*(u) - \langle x, u \rangle = 0 \quad \Leftrightarrow \quad x \in \partial f^*(u).$$

Having introduced these concepts of convex analysis, equation (1.2) can be solved. For this purpose, we switch back from the general function f to the (explicit) functions F, G and redefine the problem.

Definition 2.8 (Primal-Dual problem)**.** Let X, Y be Hilbert spaces. For convex functions $F : X \to \mathbb{R} \cup \{\infty\}$, $G : Y \to \mathbb{R} \cup \{\infty\}$, and a linear operator $K : X \to Y$ we define

$$P(x) := F(x) + G(Kx) \quad \text{and} \quad D(y) := -F^*(-K^*y) - G^*(y).$$

The problem $\inf_{x \in X} P(x)$ is called *primal problem*, whereas $\sup_{y \in Y} D(y)$ is called *dual* *problem.* With the Fenchel-Young inequality 2.6, one can conclude

$$P(x) \geq \inf_{x' \in X} P(x') \geq \sup_{y' \in Y} D(y') \geq D(y).$$

Furthermore, $\Delta(x, y) = P(x) - D(y)$ is called *duality gap.* If $\Delta(x, y) = 0$, then x and y must be optimizers of the primal and dual problem, respectively.

Proposition 2.9. The pair $(x, y) \in X \times Y$ is a solution to the primal and dual problem if and only if $Kx \in \partial G^*(y)$ and $-K^*y \in \partial F(x)$.

Proof.

$$\begin{aligned}
(x, y) \text{ is a solution } &\Leftrightarrow P(x) = D(y) \\
&\Leftrightarrow F(x) + G(Kx) = -F^*(-K^*y) - G^*(y) \\
&\Leftrightarrow (F(x) + F^*(-y) - \langle x, -y \rangle) + (G(x) + G^*(y) - \langle x, y \rangle = 0) \\
&\Leftrightarrow (F(x) + F^*(-K^*y) - \langle x, -K^*y \rangle) = 0 \\
&\quad\ \wedge\ (G(Kx) + G^*(y) - \langle Kx, y \rangle = 0
\end{aligned}$$

The third equivalence holds with the help of the Fenchel-Young inequality. Through proposition 2.7 follows

$$Kx \in \partial G^*(y) \ \wedge\ -K^*y \in \partial F(x).$$

□

With this knowledge, an algorithm can be constructed, which needs only the proximal operator of F and G^*, separately.

Algorithm 1: Primal-Dual algorithm

```
Data: $F, G$ proper, lower semi-continuous, and convex; $K : X \to Y$
Result: $\arg\min_x F(x) + G(Kx)$
Initialization: $\tau, \gamma > 0$ with $\tau\gamma \|K\|^2 \leq 1$,
                $(x^{(0)}, y^{(0)}) \in X \times Y$ and set $\tilde{x}^{(0)} = x^{(0)}$.
while $\left\| x^{(k)} - x^{(k-1)} \right\| / \left\| x^{(k)} \right\| > tol$ do
    $y^{(k+1)} = \mathrm{Prox}_{\gamma G^*}(y^{(k)} + \gamma K \tilde{x}^{(k)})$
    $x^{(k+1)} = \mathrm{Prox}_{\tau F}(x^{(k)} - \tau K^* y^{(k+1)})$
    $\tilde{x}^{(k+1)} = x^{(k+1)} + (x^{(k+1)} - x^{(k)})$
    $k = k + 1$
```

Theorem 2.10 (Primal-Dual convergence)**.** The sequences $(x^{(k)})_k$, $(y^{(k)})_k$ created by the Primal-Dual algorithm converge weakly to solutions of the primal and dual problem.

For the underlying proof and other convergence properties, see reference [11].

2.2. Ill-Posedness of Linear Operators

The description of inverse problems in equation (1.1) is introduced via a linear operator A. A more profound analysis of the underlying linear operator in the Radon and convolution problem is presented in the following. The ill-posedness according to Hadamard's criteria is examined for both.

Definition 2.11 (Linear operator)**.** Let X, Z be normed spaces then $A : X \to Z$ is called a *linear operator* when for all $u_0, u_1 \in X$ and $\lambda \in \mathbb{R}$ the following conditions apply:

1. $A(\lambda u_0) = \lambda A(u_0)$
2. $A(u_0 + u_1) = A(u_0) + A(u_1)$.

The *operator norm* of A is given by $\|A\| := \sup_{u \in X} \left\{ \frac{\|Au\|}{\|u\|} \,\middle|\, \|u\| \neq 0 \right\}$. A is continuous if and only if A is bounded, i.e., there exists a constant $c \in \mathbb{R}$ with $\|A\| \leq c$.

We first define the *Fourier transform*, which is required for fundamental investigations of the convolution and Radon operator.

Definition 2.12 (Fourier transform)**.** Let $f \in L^1(\mathbb{R}^2)$. The map

$$(\mathcal{F}f)(\xi) = \frac{1}{2\pi} \int_{\mathbb{R}^2} f(x) e^{-ix\cdot\xi} \mathrm{d}x$$

denotes the *Fourier transform*. The *inverse Fourier transform* of $w \in L^1(\mathbb{R}^2)$ is given by

$$(\mathcal{F}^{-1}w)(x) = \frac{1}{2\pi} \int_{\mathbb{R}^2} w(x) e^{ix\cdot\xi} \mathrm{d}\xi.$$

2.2.1. Convolution operator

In the following, we focus on the case of convolution which was introduced in chapter 1.

Definition 2.13 (Convolution operator)**.** The operator $A : L^2(\mathbb{R}^2) \to L^2(\mathbb{R}^2)$ with

$$(Au)(x) = (k_A * u)(x) = \int_{\Omega} k_A(x - y) u(y) \mathrm{d}y$$

denotes the *convolution operator*. We call $k_A \in L^1(\mathbb{R}^2)$ the *integration kernel*.

Proposition 2.14. The convolution A is a linear continuous operator regarding to $L^2(\mathbb{R}^2)$.

Proof. Linearity follows directly from linearity of the integration operator.

One can show that continuity still applies

$$\begin{aligned}
\|Au\|^2_{L^2(\mathbb{R}^2)} &= \int_{\mathbb{R}^2} [(Au)(x)]^2 \mathrm{d}x \\
&= \int_{\mathbb{R}^2} \left(\int_{\mathbb{R}^2} \sqrt{k_A(z)}\sqrt{k_A(z)} u(x-z) \mathrm{d}z \right)^2 \mathrm{d}x \\
&\leq \int_{\mathbb{R}^2} \|k_A\|_{L^1(\mathbb{R}^2)} \left(\int_{\mathbb{R}^2} k_A(z) u(x-z)^2 \mathrm{d}z \right) \mathrm{d}x \\
&= \|k_A\|_{L^1(\mathbb{R}^2)} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} u(x-z)^2 \mathrm{d}x k_A(z) \mathrm{d}z \\
&= \|k_A\|^2_{L^1(\mathbb{R}^2)} \|u\|^2_{L^2(\mathbb{R}^2)} .
\end{aligned}$$

□

With the following theorem, the problem can be re-formulated in Fourier space.

Theorem 2.15 (Convolution theorem)**.** Consider $f, g \in L^1(\mathbb{R}^2)$. The *convolution theorem* states that

$$\mathcal{F}(f * g) = 2\pi(\mathcal{F}f)(\mathcal{F}g).$$

Proof. Let $\xi \in \mathbb{R}^2$. Then

$$\begin{aligned}
\mathcal{F}(f * g)(\xi) &= \frac{1}{2\pi} \int_{\mathbb{R}^2} e^{-ix\cdot\xi} \int_{\mathbb{R}^2} f(x-y) g(y) \mathrm{d}y \mathrm{d}x \\
&= \frac{1}{2\pi} \int_{\mathbb{R}^2} g(y) e^{-iy\cdot\xi} \int_{\mathbb{R}^2} f(x-y) e^{-iy(x-y)\cdot\xi} \mathrm{d}x \mathrm{d}y \\
&= 2\pi \left[(\mathcal{F}g)(\xi)(\mathcal{F}f)(\xi) \right] .
\end{aligned}$$

□

Based on these properties, the ill-posedness of the convolution operator can be revealed. The following theorem points out this difficulty by investigating an explicit kernel, the Gaussian kernel.

Theorem 2.16. Let $k_A(x) = \frac{1}{2\pi} e^{-\frac{1}{2}(x_1^2 + x_2^2)}$ be the Gaussian kernel and $A : L^2(\mathbb{R}^2) \to L^2(\mathbb{R}^2)$, with $Au = k_A * u$. The inverse problem $z = Au$

1. generally does not have a solution,
2. if a solution exists it is unique,
3. but not continuous in z.

Proof. 1. Since $k_A \in C^\infty(\mathbb{R}^2) \cap W^{\infty,1}(\mathbb{R}^2)$ and Young's inequality[2] holds,

$$u * (\partial^\alpha k_A) = \partial^\alpha (u * k_A)$$

is true for every multi-index α. Therefore, $Au \in C^\infty(\mathbb{R}^2)$ and if z is not differentiable, $Au = z$ does not have a solution.

2. The uniqueness applies, since

$$z = Au \;\Leftrightarrow\; \mathcal{F}(z) = \mathcal{F}(k_A)\mathcal{F}(u) \;\Leftrightarrow\; u = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(z)}{\mathcal{F}(k_A)}\right).$$

3. Let $\mathcal{F}(z_k) = \chi_{k\vec{e}_1+[0,1]^2}$. With the Plancherel theorem one can infer

$$\|z_k\|_{L^2(\mathbb{R}^2)} = \|\mathcal{F}(z_k)\|_{L^2(\mathbb{R}^2)} = 1 \quad \forall k \in \mathbb{N}.$$

The discontinuity for A^{-1} follows from

$$\|u_k\|_{L^2(\mathbb{R}^2)} = \|\mathcal{F}(u_k)\|_{L^2(\mathbb{R}^2)} = \left\|\frac{\mathcal{F}(z_k)}{\mathcal{F}(k_A)}\right\|_{L^2(\mathbb{R}^2)} \overset{k\to\infty}{\longrightarrow} \infty,$$

since $\mathcal{F}(k_A)(x) \overset{|x|\to\infty}{\longrightarrow} 0$.

□

2.2.2. Radon transform

The next step is a similar investigation of the Radon operator $\mathcal{R}$ of computerized tomography. This includes a discuss of the ill-posedness of such transformation on $\mathbb{R}^2$, based on references [12, 13]. Usually, the Radon transform is defined over smooth functions whose derivatives (including the function itself) decay at infinity faster than any power. The associated function space

$$S(\mathbb{R}^2) := \left\{u \in C^\infty(\mathbb{R}^2) \;:\; |x^\alpha D^\beta u| \leq C(\alpha,\beta) \quad \forall \alpha, \beta \in \mathbb{N}^2\right\}$$

is called *Schwartz space.*

Definition 2.17 (Radon transform)**.** Let $C := \{(\theta, s) : \theta \in S^1, s \in \mathbb{R}\}$. The map $\mathcal{R} : \mathcal{S}(\mathbb{R}^2) \to \mathcal{S}(C)$, with

$$(\mathcal{R}u)(\theta, s) = \int_{x\cdot\theta=s} u(x)\mathrm{d}\sigma(x),$$

is proper, linear, and continuous. We call $\mathcal{R}$ the *Radon transform.*

---

[2] $\|u * v\|_{L^r} \leq \|u\|_{L^p}\|v\|_{L^q}$ for $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$ with $p, q, r > 1$.

As described in section 1.1, we get integrals for every angle $\theta \in S^1$. In the following, we develop an inversion formula for $\mathcal{R}$. A first necessity is the L^2-adjoint operator, also called *back-projection.*

Theorem 2.18 (Back-projection)**.** The map $\mathcal{R}^* : \mathcal{S}(C) \to \mathcal{S}(\mathbb{R}^2)$, with

$$(\mathcal{R}^* v)(x) = \int_{S^1} v(\theta, x \cdot \theta) \mathrm{d}\theta$$

for $x \in \mathbb{R}^2$, is the L^2-adjoint of $\mathcal{R}$.

Proof. Let $u \in L^2(\mathbb{R}^2)$ and $v \in L^2(C)$. Then

$$\begin{aligned}\langle \mathcal{R}u, v \rangle_{L^2(C)} =& \int_{\mathbb{R}} \int_{S^1} v(\theta, s)(\mathcal{R}u)(\theta, s) \mathrm{d}\theta \mathrm{d}s \\ =& \int_{\mathbb{R}} \int_{S^1} v(\theta, s) \int_{x \cdot \theta = s} u(x) \mathrm{d}x \mathrm{d}\theta \mathrm{d}s \\ =& \int_{S^1} \int_{\mathbb{R}^2} u(x) v(\theta, x \cdot \theta) \mathrm{d}x \mathrm{d}\theta \\ =& \int_{\mathbb{R}^2} u(x) \int_{S^1} v(\theta, x \cdot \theta) \mathrm{d}\theta \mathrm{d}x \\ =& \langle u, \mathcal{R}^* v \rangle_{L^2(\mathbb{R}^2)}.\end{aligned}$$

☐

The name *back-projection* hints that one deals with a certain kind of inversion. This, and the difference to an entire inversion, is clarified in the following theorem.

Theorem 2.19. For $u \in \mathcal{S}(\mathbb{R}^2)$ holds

$$\mathcal{R}^* \mathcal{R} u = 2 \left(\frac{1}{\|\cdot\|} * u \right).$$

Proof. Let $x \in \mathbb{R}^2$. Then

$$\begin{aligned}(\mathcal{R}^* \mathcal{R} u)(x) =& \int_{S^1} \mathcal{R}u(\theta, x \cdot \theta) \mathrm{d}\theta \\ =& \int_{S^1} \int_{y \cdot \theta = x \cdot \theta} u(y) \mathrm{d}y \mathrm{d}\theta \\ =& \int_{S^1} \int_{y \cdot \theta = 0} u(x + y) \mathrm{d}y \mathrm{d}\theta \\ =& |S^0| \int_{\mathbb{R}^2} u(x + y) \frac{1}{\|y\|} \mathrm{d}y \\ =& 2 \left(\frac{1}{\|\cdot\|} * u \right)(x).\end{aligned}$$

☐

Note the convolution in the inversion formula of the Radon transform. First, two properties are needed before an adjustment to the back-projection can be constructed.

Theorem 2.20 (Fourier-slice theorem)**.** Let $u \in \mathcal{S}(\mathbb{R}^2)$, $\theta \in S^1$, and $s \in \mathbb{R}$. This yields

$$\mathcal{F}(\mathcal{R}u)(\theta,\sigma) = \frac{1}{2\pi}\mathcal{F}(u)(\sigma \cdot \theta).$$

Proof.

$$\begin{aligned}\mathcal{F}(\mathcal{R}u)(\theta,\sigma) =& \frac{1}{2\pi}\int_{\mathbb{R}}(\mathcal{R}u)(\theta,s)e^{-is\theta}\mathrm{d}s\\ =&\frac{1}{2\pi}\int_{\mathbb{R}}\int_{x\cdot\theta=s}u(x)e^{-is\sigma}\mathrm{d}x\mathrm{d}s\\ =&\frac{1}{2\pi}\int_{\mathbb{R}^2}u(x)e^{-i\sigma x\cdot\theta}\mathrm{d}x\\ =&\frac{1}{2\pi}(\mathcal{F}u)(\sigma\cdot\theta)\end{aligned}$$

□

In a next step, we discuss an important connection between the convolution and the Radon transform.

Lemma 2.21. For $h \in \mathcal{S}(C)$ and $u \in \mathcal{S}(\mathbb{R}^2)$ holds

$$(\mathcal{R}^*h) * u = \mathcal{R}^*(h * \mathcal{R}u).$$

See reference [12] for the proof.

With this, the convolution can be transferred from the image space to the data space. The so-called *filtered back-projection* is obtained.

Theorem 2.22 (Filtered back-projection)**.** For $u \in \mathcal{S}(\mathbb{R}^2)$,

$$u = \frac{1}{4\pi}\mathcal{R}^*\left(\mathcal{R}u * h\right)$$

is valid, where $\mathcal{F}(h)(x) = |x|$. The convolution kernel h is referred to as *Ramp filter*.

Proof. Let $x \in \mathbb{R}^2$, then

$$\begin{aligned}\frac{1}{4\pi}\mathcal{R}^*\left(h * \mathcal{R}u)\right)(x) =& \frac{1}{4\pi}\int_{S^1}\left(h * \mathcal{R}u)\right)(\theta, x\cdot\theta)\mathrm{d}\theta \\ =& \frac{1}{4\pi}\int_{S^1}\frac{1}{2\pi}\int_{\mathbb{R}}\left(2\pi|\sigma|\mathcal{F}(u)(\sigma)\right)e^{i\sigma x\theta}\mathrm{d}\sigma\mathrm{d}\theta \\ =& \frac{1}{2\pi}\int_{S^1}\int_{\mathbb{R}+}\sigma(\mathcal{F}(u)(\sigma))e^{i\sigma x\theta}\mathrm{d}\sigma\mathrm{d}\theta \\ =& \frac{1}{2\pi}\int_{\mathbb{R}^2}\mathcal{F}(u)(\xi)e^{ix\xi}\mathrm{d}\xi \\ =& u(x).\end{aligned}$$

□

Now an inversion formula is established with $\mathcal{R}^{-1}(\cdot) = \frac{1}{4\pi}\mathcal{R}^*\left((\mathcal{R}(\cdot) * h\right)$. By examining this equation, we obtain the following continuity property.

Theorem 2.23 (Ill-posedness of Radon transform)**.** Let $K \subset \mathbb{R}^2$ be compact. There exist some constants c_0 and c_1, with

$$c_0 \left\|u\right\|_{H^0(\mathbb{R}^2)} \leq \left\|\mathcal{R}u\right\|_{H^{1/2}(C)} \leq c_1 \left\|u\right\|_{H^0(\mathbb{R}^2)} \quad \forall u \in C_0^\infty(K).$$

Furthermore, for $z = \mathcal{R}u$,

$$\left\|\mathcal{R}^{-1}z\right\|_{L^2(\mathbb{R}^2)} \leq c_0 \left\|z\right\|_{H^{1/2}(C)} \leq \frac{c_1}{c_0}\left\|\mathcal{R}^{-1}z\right\|_{L^2(\mathbb{R}^2)}.$$

The inverse of the Radon transform $\mathcal{R}^{-1}$ is therefore continuous if and only if the data space norm contains derivatives of half order[3].

For an underlying proof based on the Sobolev theory, see references [12, 13].

Even if there exists a continuous inversion with respect to a certain Sobolev space, this is, as shown, not the case between general L^2-spaces. With the knowledge of the noise behavior in computerized tomography from chapter 1, one generally obtains a discontinuous inversion. Thus, the Radon transform violates the third Hadamard's law and creates an ill-posed problem. This already becomes visible in the reconstruction proof in theorem 2.22. The underlying noise of the data $\mathcal{F}(u)(\theta, \sigma)$ is amplified with magnitude $|\sigma|$ in the intermediate result. For large $|\sigma|$ we also get arbitrarily high reconstruction errors, even where the data noise is small.

In addition, other examples exist that deteriorate a reconstruction. A low-dose version that processes over a few equally distributed angles may be one example [14]. Another is an

---

[3] The Sobolev space $H^\alpha(\mathbb{R}^2)$ of real order α is defined by $H^\alpha := \{f \in \mathcal{S}^*(\mathbb{R}^2) : (1 + |\xi|)^{\alpha/2}\mathcal{F}(f) \in L^2(\mathbb{R}^2)\}$. See chapter 7 in reference [12] for an introduction to the underlying theory.

angle limitation where a certain range of measurements are not included in the sinogram [15].

2.3. Regularization

Both problems were shown to violate Hadamard's third rule of a continuous inverse. If only the existence condition is affected, the minimum-norm solution would suffice (remember Fig. 1.4). For our applications, however, we examine the optimization problem (1.2). In order to investigate specific problems of imaging, we will finally present an appropriate concept for the convex regularization functional G.

2.3.1. Functions of bounded variation

In chapter 1.2 we briefly discussed a suitable regularization variant for images in the discrete case: the *total variation* as the 1-norm of the gradient. Now we model the theoretical basis for this approach. Our target is the existence-proof of a minimizer of equation (1.2).

The common approach includes splitting the image $z \in L^2(\Omega)$ into two components, $z = u + v$. This splitting is a reminiscent of the initial problem (1.1), where u and η differ in similar whiteness, since u consists of the essential geometry of the image, whereas v and η contain high oscillations. In this model $u \in \mathrm{BV}(\Omega)$ represents a *function of bounded variation*, while $v := f - u$ defines the residual. This concept was pioneered by Rudin, Osher, and Fatemi in reference [16] and is known as the *ROF model*. In the following, we focus on the interaction between total variation and the space $\mathrm{BV}(\Omega)$.

The essential structure of the image as a function with bounded variation is defined with constraints on the weak derivative. The following theory of the space $\mathrm{BV}(\Omega)$ is based on reference [17]. Throughout this section we denote by Ω a generic open bounded set in $\mathbb{R}^2$.

Definition 2.24. A function $u \in L^1(\Omega)$ is called *function of bounded variation*, if the weak derivative can be expressed as a Radon measure $\mu \in \mathcal{M}(\operatorname{int}\Omega)$. In other words,

$$\int_\Omega \varphi'(x) \cdot u(x)dx = -\int_\Omega \varphi(x)\mathrm{d}\mu(x)$$

applies for all test functions $\varphi \in C_0^1 := \left\{\varphi \in C^1(\Omega) \;\middle|\; \varphi|_{\partial\Omega} = 0\right\}$.

If $u \in \mathrm{BV}(\Omega)$, the weak derivative is often denoted Du.

Definition 2.25 (Total variation)**.** The *total variation* for some $u \in L^1(\Omega)$ is given by

$$\mathrm{TV}(u) = \sup\left\{\int_\Omega u(x) \cdot \varphi'(x)dx \;\middle|\; \varphi \in C_0^1(\Omega), \|\varphi\|_{C^0(\Omega)} \leq 1\right\}.$$

Theorem 2.26. For $u \in L^1(\Omega)$ holds

$$u \in \mathrm{BV}(\Omega) \Leftrightarrow \mathrm{TV}(u) < \infty.$$

See reference [17] for the underlying proof.

As consequence, the definition for functions with bounded variations is often directly defined as

$$\mathrm{BV}(\Omega) = \{u \in L^1(\Omega) \mid \; \mathrm{TV}(u) < \infty\}, \qquad \|\cdot\|_{\mathrm{BV}} = \|\cdot\|_{L^1(\Omega)} + \mathrm{TV}(\cdot).$$

Main properties of the TV-functional, obtained in references [17, 18], can be summarized as followed:

1. $\mathrm{TV}(\cdot)$ is a semi-norm on the Banach space $(\mathrm{BV}(\Omega), \|\cdot\|_{\mathrm{BV}})$.
2. $\mathrm{TV}(\cdot)$ is proper, convex, and lower semi-continuous on each $L^p(\Omega)$, $1 \leq p < \infty$.
3. Let $u \in L^1(\Omega)$. Then $u \in \mathrm{BV}(\Omega)$ if and only if there exists a sequence $(u_k)_k \subset C^\infty(\Omega)$ converging to u in $L^1(\Omega)$ and satisfying

$$\mathrm{TV}(u) = \lim_{k\to\infty} \int_\Omega |\nabla u_k| \mathrm{d}x < \infty.$$

Another important property for the existence-proof of minimizers can be found in the next theorem, which is based on reference [19].

Theorem 2.27 (Poincaré-Wirtinger inequality in the BV-space)**.** Let $\Omega \subset \mathbb{R}^2$ be a Lipschitz open bounded set. Then, there exists a constant $c > 0$ such that

$$\|u - m(u)\|_{L^1(\Omega)} \leq c\,\mathrm{TV}(u) \quad \forall u \in \mathrm{BV}(\Omega),$$

where $m(u) := \frac{1}{|\Omega|}\int_\Omega u(x)\mathrm{d}x$ is the mean-value of u.

Proof. Let $u \in \mathrm{BV}(\Omega)$ and $(u_k)_k \subset C^\infty(\bar{\Omega})$ be a sequence such that

$$\lim_{k\to\infty} \|u_k - u\|_{L^1(\Omega)} = 0, \qquad \lim_{k\to\infty} \mathrm{TV}(u_k) = \mathrm{TV}(u).$$

Furthermore, $u_k \in W^{1,1}$, since Ω is bounded. In addition, $\lim_{k\to\infty} m(u_k) = m(u)$ so the Poincaré-Wirtinger inequality for the Soblev-space $W^{1,1}(\Omega)$ infers

$$\|u_k - m(u_k)\|_{L^1(\Omega)} \leq c\,\|\nabla u_k\|_{L^1(\Omega)} = \mathrm{TV}(u_k) \quad \forall\, k \in \mathbb{N}$$

for some $c > 0$. The result is obtained by going to the limit. □

We are able to prove the existence of a minimizer in the next theorem, following the scheme presented in reference [20].

Theorem 2.28. Let $\Omega, \Sigma \subset \mathbb{R}^2$ be an open bounded Lipschitz domain. For a linear bounded operator $A : L^2(\Omega) \to L^2(\Sigma)$ that does not annihilate constant functions the problem (with $\lambda > 0$)

$$\min_{u \in L^2(\Omega)} [\mathcal{E}(u) := \frac{\lambda}{2} \|Au - z\|^2_{L^2(\Sigma)} + \mathrm{TV}(u)]$$

has at least one solution. If A is injective, this solution is unique.

Proof. The functional $\mathcal{E}$ is convex and lower semi-continuous, since TV and the norm are both. In addition, A is a linear bounded operator. We proof the BV-coerciveness, i.e.

$$\left[\, \|u\|_{\mathrm{BV}(\Omega)} \to \infty \right] \Rightarrow \left[\mathcal{E}(u) \to \infty\right].$$

Let $u \in \mathrm{BV}(\Omega)$. The following decomposition holds true

$$u = w + v, \qquad \text{with} \qquad w = m(u)\chi_\Omega \qquad \text{and} \qquad \int_\Omega v \mathrm{d}x = 0.$$

We obtain

$$\begin{aligned} \|u\|_{\mathrm{BV}(\Omega)} &\leq \|w\|_{\mathrm{BV}(\Omega)} + \|v\|_{\mathrm{BV}(\Omega)} \\ &\leq \|w\|_{L^1(\Omega)} + \|v\|_{L^1(\Omega)} + \mathrm{TV}(v) \\ &\leq \|w\|_{L^1(\Omega)} + (1 + c_1)\, \mathrm{TV}(v) \end{aligned}$$

for the constant c_1 according to the Poincaré-Wirtinger inequality 2.27.

Since A does not annihilate constants, there exists $c_2 > 0$ independent of w such that

$$\|Aw\|_{L^2(\Sigma)} \geq c_2 \|w\|_{L^1(\Omega)}.$$

Further,

$$\begin{aligned} \mathcal{E}(u) &\geq \mathrm{TV}(v) + \|Aw + Av - z\|^2_{L^2(\Omega)} \\ &\geq \mathrm{TV}(v) + \|Aw\|_{L^2(\Sigma)} (\|Aw\|_{L^2(\Sigma)} - 2\|Av - z\|_{L^2(\Sigma)}) \\ &\geq \mathrm{TV}(v) + \|Aw\|_{L^2(\Sigma)} (c_2 \|w\|_{L^1(\Omega)} - 2(\|A\| \|v\|_{L^2(\Omega)} + \|z\|_{L^2(\Omega)})) \\ &\geq \mathrm{TV}(v) + \|Aw\|_{L^2(\Sigma)} (c_2 \|w\|_{L^1(\Omega)} - 2(\|A\| c_1 \mathrm{TV}(v) + \|z\|_{L^2(\Omega)})). \end{aligned}$$

We have to distinguish two cases:

1. $c_2 \|w\|_{L^1(\Omega)} - 2(\|A\| c_1 \mathrm{TV}(v) + \|z\|_{L^2(\Omega)}) \geq 1$.

Then $\mathcal{E}(u) \geq \mathrm{TV}(v) + \epsilon \left\| Aw \right\|_{L^2(\Sigma)}$ and

$$\left\| w \right\|_{L^1(\Omega)} \leq \frac{1}{\epsilon c_2} \mathcal{E}(u) \qquad \left\| u \right\|_{\mathrm{BV}} \leq \left(\frac{1}{\epsilon c_2} + 1 + c_1 \right) \mathcal{E}(u).$$

2. $c_2 \left\| w \right\|_{L^1(\Omega)} - 2(\left\| A \right\| c_1 \, \mathrm{TV}(v) + \left\| z \right\|_{L^2(\Omega)}) < 1$. Then

$$\begin{aligned} \left\| w \right\|_{L^1(\Omega)} &< \frac{1}{c_2} \left(1 + 2(\left\| A \right\| c_1 \, \mathrm{TV}(v) + \left\| z \right\|_{L^2(\Omega)}) \right), \\ \left\| u \right\|_{\mathrm{BV}} - \frac{1 + 2 \left\| z \right\|_{L^2(\Omega)}}{c_2} &\leq \left(\frac{2 \left\| A \right\| c_1}{c_2} + (1 + c_1) \right) \mathcal{E}(u). \end{aligned}$$

The combination of both cases shows the BV-coercivity.

Let $(u_k)_k$ be a minimizing sequence for $\mathcal{E}$. As shown, $(u_k)_k$ is BV-bounded. So there exists a subsequence $(u_{k_j})_j$ that converges weakly to some $\tilde{u} \in L^2(\Omega)$, since A is a continuous linear operator (the sequence $(Au_{k_j})_j$ converges weakly to $A\tilde{u}$, too). As convexity and lower semi-continuity of $\mathcal{E}$ is given, we obtain

$$\mathcal{E}(\tilde{u}) \leq \liminf_{j \to \infty} \mathcal{E}((u_{k_j})_j) = \inf_{u \in L^2(\Omega)} \mathcal{E}(u).$$

Since $\left\| \cdot \right\|^2_{L^2(\Sigma)}$ is strictly convex, the uniqueness directly follows for an injective A.

$\square$

We have proved the existence of an optimizer and the validity for a special version of equation (1.2).

Remark. The total variation can lead to so-called staircasing effects: The appearance of unintended edges can be observed. This can be prevented by considering higher derivatives in the TV-norm. In reference [18] this extended regularization is investigated. This is not required and, in addition, unnecessarily complicates numerical implementations. Therefore, we will mainly work with the first-order total variation. Fig. 2.2 shows the described effect.

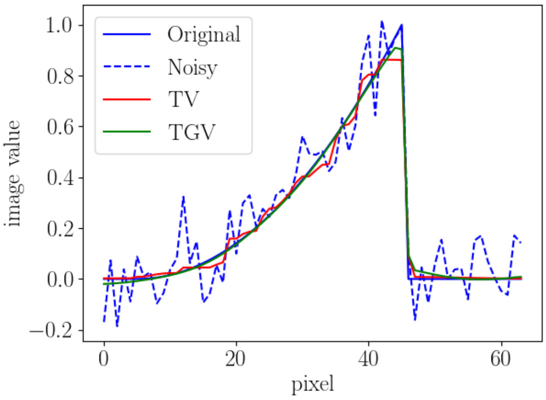


Figure 2.2.: Staircasing effect on total-variation-based regularization. A comparison of the underlying two-dimensional images can be found in Figure C.4.

2.3.2. Iterative Bregman divergence

Now, we consider an iterative extension of the regularization function. With this, we prevent the contrast loss of total-variation regularization (remember example 1.1). In this method, the iterative procedure converges from an over-regularized towards the degraded solution. This process is based on the regularization functional. The iterative procedure stops as soon as the spacing between solution and data becomes smaller than a threshold. Since this method applies to arbitrary convex regularizations, we speak once again of a general convex regularization function f in the following. The algorithm will later be reformulated in terms of the concrete TV-functional.

Definition 2.29 (Bregman divergence)**.** Let $f : \mathcal{H} \to \mathbb{R} \cup \{\infty\}$ be a proper, lower semi-continuous, and convex function. For $p \in \partial f(y)$, the so-called *Bregman divergence* between $u, v \in \operatorname{dom}(f)$ is defined by

$$D_f^p(u,v) := f(u) - f(v) - \langle p, u - v \rangle.$$

The *symmetric Bregman divergence* reads

$$D_f^{\text{symm}}(u,v) := D_f^p(u,v) + D_f^q(v,u) = \langle q - p, u - v \rangle,$$

where $p \in \partial f(v)$ and $q \in \partial f(u)$.

This concept (see Fig. 2.3) can be interpreted as a distance between f at point u and a linear extension in u developed around v. In particular

$$D_f^p(u,v) = f(u) - \underbrace{(f(v) + \langle p, u - v \rangle)}_{\text{linear extension}}.$$

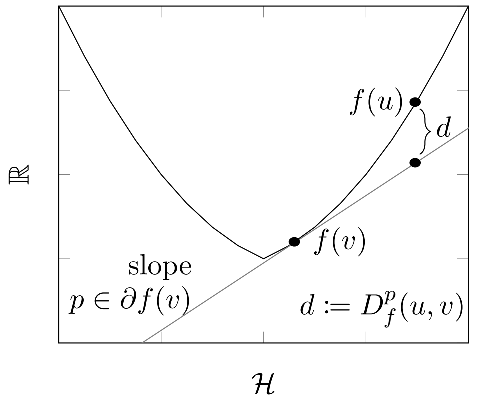


Figure 2.3.: Visualization of the Bregman divergence.

The slope of the linear extension is based on the subgradient $p \in \partial f(v)$. Intuitively, we get something similar to a distance between two points with respect to a convex functional. Several well-known divergences, e.g. the Kullback-Leibler divergence, are based on this idea [21]. Even if the Bregman divergence is not a metric in the narrow sense - since it is not symmetric and the triangle inequality does not hold - there are some similar properties [22]:

1. $D_f^p(u,v) \geq 0 \quad \forall u,v \in \mathcal{H}$.
2. $D_f^p(u,v) = 0 \;\Rightarrow\; u = v \quad \forall u,v \in \mathcal{H}$.
3. $D_f^p(u,v)$ is convex in the first argument but not necessarily in the second one.
4. For f strictly convex, $D_f^p(u,v)$ is also strictly convex in u for each v. As a consequence, conclusion 2 holds in both directions.

Therefore, D_f^p is often called *Bregman distance*.

The next step involves the concretization to our imaging case. We set the Bregman divergence in respect to the total variation, $f(\cdot) = \mathrm{TV}(\cdot)$. Investigations are now set up for the explicit fidelity term $F(u) = \frac{\lambda}{2}\left\|Au - z\right\|^2_{L^2(\Sigma)}$. We recall the same setting as in theorem 2.28, where $A : L^2(\Omega) \to L^2(\Sigma)$ is a linear bounded operator. The process is now iterated: We iterate from an initially over-regularized solution towards the disturbed version along the TV-functional. This results into the iteration rule

$$u^{(k+1)} = \underset{u \in L^2(\Omega)}{\arg\min} \left[Q^{(k+1)}(u) := \frac{\lambda}{2}\left\|Au - z\right\|^2_{L^2(\Sigma)} - \mathrm{TV}(u^{(k)}) - \langle p^{(k)}, u - u^{(k)} \rangle + \mathrm{TV}(u) \right],$$
$$\text{with} \quad p^{(k)} \in \partial\, \mathrm{TV}(u^{(k)}).$$

With respect to reference [23], we show the well-definedness of the iteration in the following theorem.

Theorem 2.30. Let $u_0 = 0$ and $p_0 = 0 \in \partial\,\mathrm{TV}(u_0)$. For each $k \in \mathbb{N}$ there exists a minimizer $u^{(k+1)}$ of $Q^{(k+1)}$. There exists a subgradient $p^{(k+1)} \in \partial\,\mathrm{TV}(u^{(k+1)})$ and $q^{(k+1)} = \partial\left[\frac{\lambda}{2}\left\|A\cdot - z\right\|^2_{L^2(\Sigma)}\right]$, such that

$$p^{(k+1)} = p^{(k)} - q^{(k+1)}.$$

The minimizer $u^{(k+1)}$ is unique if A is injective.

Proof. This can be proven via induction. Since $u_0 = 0$ and $p_0 = 0$, we have $Q^{(1)}(u) = \frac{\lambda}{2}\left\|Au - z\right\|^2_{L^2(\Sigma)} + \mathrm{TV}(u)$ and according to theorem 2.28 there exists a minimizer, as well as $p^{(1)} = q^{(1)}$. In addition, we introduce an auxiliary variable $r^{(1)} := \lambda(z - Au^{(1)})$, so that $p^{(1)} = A^* r^{(1)}$ applies.

Now we proceed from k to $k+1$ and assume that $p^{(k)} = A^* r^{(k)}$ for $r^{(k)} \in L^2(\Sigma)$. Under these assumptions, the functional $Q^{(k+1)}$ is convex and lower semi-continuous. The following estimation holds

$$\begin{aligned}
Q^{(k+1)} =& \frac{\lambda}{2}\left\|Au - z\right\|^2_{L^2(\Sigma)} - \langle p^{(k)}, u - u^{(k)}\rangle + \mathrm{TV}(u) - \mathrm{TV}(u^{(k)}) \\
=& \frac{\lambda}{2}\left\|Au - z\right\|^2_{L^2(\Sigma)} - \langle r^{(k)}, Au - Au^{(k)}\rangle + \mathrm{TV}(u) - \mathrm{TV}(u^{(k)}) \\
=& \frac{\lambda}{2}\left\|Au - z - \lambda^{-1}r^{(k)}\right\|^2_{L^2(\Sigma)} + \frac{1}{2}\langle Au - z - \lambda^{-1}r^{(k)}, r^{(k)}\rangle \\
&+ \frac{1}{2}\langle r^{(k)}, Au - z\rangle - \langle r^{(k)}, Au - Au^{(k)}\rangle + \mathrm{TV}(u) - \mathrm{TV}(u^{(k)}) \\
=& \frac{\lambda}{2}\left\|Au - z - \lambda^{-1}r^{(k)}\right\|^2_{L^2(\Sigma)} + \frac{1}{2}\langle -z - Au^{(k)} - \lambda^{-1}r^{(k)}, r^{(k)}\rangle \\
&+ \frac{1}{2}\langle r^{(k)}, z - Au^{(k)}\rangle + \mathrm{TV}(u) - \mathrm{TV}(u^{(k)}) \\
=& \frac{\lambda}{2}\left\|Au - \bar{z}\right\|^2_{L^2(\Sigma)} + \mathrm{TV}(u) - \langle r^{(k)}, z - Au^{(k)}\rangle + \frac{\left\|r^{(k)}\right\|^2_{L^2(\Sigma)}}{2\lambda} - \mathrm{TV}(u^{(k)})
\end{aligned}$$

for $\bar{z} = z + \lambda^{-1} r^{(k)} \in L^2(\Sigma)$.

Only $\frac{\lambda}{2}\left\|Au - \bar{z}\right\|^2_{L^2(\Sigma)}$ and $\mathrm{TV}(u)$ are not constant on the right-hand side of the inequality. Therefore, $Q^{(k+1)}$ is BV-coercive and the existence-proof follows the same way as shown in theorem 2.28.

By strict convexity of $\frac{\lambda}{2}\left\|A\cdot - z\right\|^2$, the uniqueness of the minimizer follows as before via the injectivity of A. With $\partial(-\langle p^{(k)}, \cdot\rangle) = \{-p^{(k)}\}$, the optimality condition for this problem implies

$$p^{(k)} \in \partial\,\mathrm{TV}(u^{(k+1)}) + \partial\left[\frac{\lambda}{2}\left\|A\cdot - z\right\|^2_{L^2(\Sigma)}\right](u^{(k+1)}),$$

which proofs the existence of $p^{(k+1)} \in \partial\,\mathrm{TV}(u^{(k+1)})$ and $q^{(k)} = \lambda A^*(Au^{(k+1)} - z)$. With $r^{(k+1)} := r^{(k)} - \lambda(Au^{(k+1)} - z) \in L^2(\Omega)$ and $p^{(k+1)} = A^* r^{(k+1)}$, we obtain

$$p^{(k+1)} = p^{(k)} - q^{(k+1)}.$$

□

A general convergence theorem can be found in reference [23]. With respect to our noisy data, we consider a specialized form: We suppose that z_{gt} is the true noise-free image and that $\tilde{u}$ is a minimizer with $\frac{\lambda}{2}\left\|A\tilde{u} - z_{gt}\right\|^2_{L^2(\Sigma)} = 0$. In addition, we model a positive number δ as the noise level with

$$\frac{\lambda}{2}\left\|A\tilde{u} - z\right\|^2_{L^2(\Sigma)} \leq \delta^2.$$

For this case, we conclude this section with two theorems, whose proofs are provided in reference [23].

Theorem 2.31 (Decreasing divergence)**.** Let $\tilde{u}, z$, and z_{gt} be such that $\tilde{u}$ is a minimizer of $\frac{\lambda}{2}\left\|A\cdot - z_{gt}\right\|^2_{L^2(\Sigma)}$ and $\frac{\lambda}{2}\left\|A\tilde{u} - z\right\|^2_{L^2(\Sigma)} \leq \delta^2$ holds. Then, as long as $\frac{\lambda}{2}\left\|Au^{(k)} - z\right\|^2_{L^2(\Sigma)} > \delta^2$, the Bregman divergence between $u^{(k)}$ and $\tilde{u}$ is decreasing. More precisely,

$$D^{p^{(k)}}_{\mathrm{TV}}(\tilde{u}, u^{(k)}) < D^{p^{(k-1)}}_{\mathrm{TV}}(\tilde{u}, u^{k-1}).$$

We set a stopping index with

$$k_*(\delta) := \max\left\{k \in \mathbb{N} \mid \frac{\lambda}{2}\left\|Au_k - z\right\|^2_{L^2(\Sigma)} \geq \tau\delta^2\right\},$$

where $\tau > 1$. The obvious choice of $\tau = 1$ is too severe to guarantee the boundedness of $\mathrm{TV}(u_{k_*})$.

Theorem 2.32 (Semi-convergence for noisy data)**.** Let $A\tilde{u} = z$. With the assumption of theorem 2.31 and stopping index $k_*(\delta)$, $\mathrm{TV}(u_{k_*(\delta)})$ is uniformly bounded in δ and as $\delta \to 0$, there exists a weak-* convergent subsequence $(u_{k_*(\delta_l)})_l$ in $\mathrm{BV}(\Omega)$. If the set $\{k_*(\delta)\}_{\delta\in\mathbb{R}^+}$ is unbounded, the limit of each weak-* convergent subsequence is a solution of $Au = z_{gt}$.

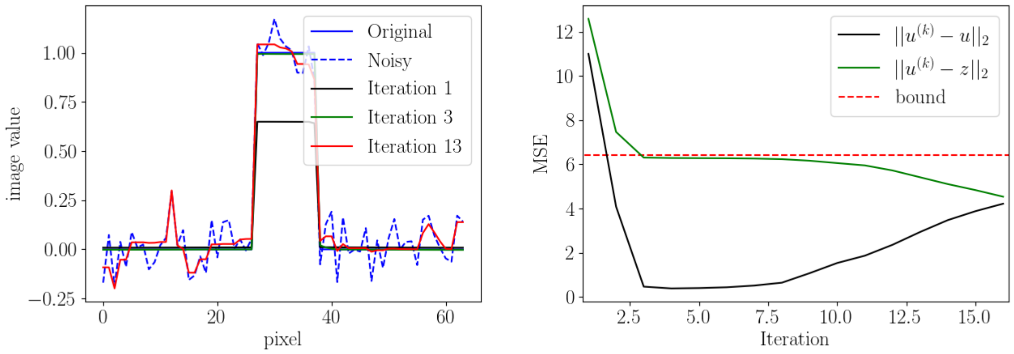


Figure 2.4.: Solution of the iterative Bregman method for the case of a square. The subfigure on the left shows the cross section over different iterations. The subfigure on the right displays the relation to the discrepancy principle based on the mean squared error (MSE). The underlying images are given in Fig. C.2.

We merge the results into an algorithm that we call BTV from here on.

```
Algorithm 2: Iterative Bregman algorithm with respect to total variation (BTV).
Input: $u^{(0)} = 0$, $p^{(0)} = 0$, noise-variance $\sigma^2$, weight $\lambda$.
while $\left\| Au^{(k)} - z \right\|_{L^2(\Sigma)} > \sigma\sqrt{|\Omega|}$ do
    Solve via Primal-Dual algorithm the discrete version of
        $$u^{(k+1)} = \operatorname*{arg\,min}_{u \in L^2(\Omega)} \frac{\lambda}{2} \|Au - z\|^2_{L^2(\Sigma)} + D^{p^{(k)}}_{\mathrm{TV}}(u, u^{(k)}).$$
    $p^{(k+1)}_{i.j} = p^{(k)}_{i.j} - \lambda A^*(Au^{(k+1)} - z)_{i.j}$
    $k = k + 1$
```

Solving via the Primal-Dual algorithm is possible with a simple adjustment of Prox_F, see example A.2. The application of the BTV algorithm to the square example 1.1 can be seen in Fig. 2.4.

CHAPTER 3

Spatially-Adapted Regularization

In this chapter we deal with another concept of enhanced regularization. With the Bregman method, we introduced an iteration which proceeds from an over-regularized solution to the degraded data along the TV-functional. A similar iterative approach will be used for the spatially-adapted version as well. We have already seen in example 1.1 that the weighting depends on the nature of an object. However, an image consists of many different objects:

> "The fact that in general images are comprised of multiple objects at different scales suggests that different values of λ localized at image features of different scales are desirable to obtain better restoration results. For this reason, the use of a spatially dependent regularization parameter $\lambda \in L^{\infty}(\Omega)$ was proposed [...]." [24]

The next two sections list a way which spatially adjusts the regularization parameter based on variance estimation. The sole change in the optimization problem is the weighting parameter λ which is promoted from a scalar to an element of $L^{\infty}(\Omega)$. In particular,

$$\operatorname*{arg\,min}_{u \in \mathrm{BV}(\Omega)} \frac{1}{2} \int_{\Omega} \lambda(x)(Au - z)^2(x)\mathrm{d}x + \mathrm{TV}(u). \tag{3.1}$$

To find an appropriate choice of $\lambda(x)$ will be the main quest in the following chapter. For the parameter selection, which is based on variance estimation, we first have to restrict the problem to the image processing case. For this special case, the fundamentals are described in order to extend them later to the Radon operator.

3.1. Formulation for Image Processing

We will first establish a constraint problem for describing the spatially-adapted regularization. Secondly, we demonstrate a connection to the form of Eq. (3.1) in order to solve the problem by the Primal-Dual algorithm. As a third point, we then transfer the regularization and the associated variance estimation to the discrete case.

To introduce the theory of local regularization, we restrict the linear operator A to the convolution special case as mentioned. Thus, the starting point agrees with the one of the references [24, 25], which allows us to introduce the basic concepts of their work. This type of operators has a significant advantage in considering the residual, which is the basis for variance-based spatially-adapted regularization.

Definition 3.1 (Smoothed residual)**.** Let $\omega \in L^\infty(\Omega \times \Omega)$ be a normalized filter and $\omega \geq 0$ with

$$\int_\Omega \int_\Omega \omega(x,y) \mathrm{d}y \mathrm{d}x = 1 \qquad \int_\Omega \int_\Omega \omega(x,y)\phi^2 \mathrm{d}y \mathrm{d}x \geq \epsilon \left\|\phi\right\|^2_{L^2(\Omega)} \quad \forall \phi \in L^2(\Omega).$$

The ω-smoothed version of the residual S is defined as

$$\begin{aligned} S : L^2(\Omega) &\to L^\infty(\Omega) \\ (Su)(x) &\mapsto \int_\Omega \omega(x,y)(Au-z)^2(y)\mathrm{d}y. \end{aligned}$$

Note that Su may be interpreted as a local variance. The inequality property for the normalized filter shown above must hold to guarantee boundedness in proposition 3.5. Based on this definition, one foresees the necessity of restricting the operator to the convolution case. A smoothed version of $(Au-z)^2$ only makes sense if local information of an object is bound together in a neighborhood. On the contrary, when considering the Radon operator, we have global information in every point of the data space. In section 3.2, we will formulate an extension that also allows the consideration of the Radon operator. Meanwhile, we stay with A as a convolution operator.

Lemma 3.2. The following properties of the local residual S are valid.

1. The Fréchet derivative of S in u with direction $v \in L^2(\Omega)$ is given by

$$\delta S(u;v) = 2 \int_\Omega \omega(x,y)[(Au-z)Av](y)\mathrm{d}y.$$

2. S is continuous and (pointwise) convex.

Proof. 1. Let $u \in L^2(\Omega)$. Then,

$$\begin{aligned} \delta S(u;v) &= \lim_{\alpha\to 0} \frac{1}{\alpha}(S(u+\alpha v) - S(u)) \\ &= \lim_{\alpha\to 0} \frac{1}{\alpha}\left(\int_\Omega \omega(x,y)[(Au-z)^2 + 2\alpha AuAv - 2\alpha Avz + \alpha^2 AvAv](y)\mathrm{d}y - S(u)\right) \\ &= \lim_{\alpha\to 0} \frac{1}{\alpha}\left(S(u) - S(u) + 2\alpha \int_\Omega \omega(x,y)[(Au-z)Av](y)\mathrm{d}y + \alpha^2 \int_\Omega \omega(x,y)AvAv(y)\mathrm{d}y\right) \\ &=2\int_\Omega \omega(x,y)[(Au-z)Av](y)\mathrm{d}y. \end{aligned}$$

Since $\|S(u+v) - S(u) - \delta S(u; v)\| = 0$, S is Fréchet differentiable at u with Fréchet differential $\delta S(u, \cdot)$.

2. The continuity follows from point 1. For the convexity: Let $u_0, u_1 \in L^2(\Omega)$ and $\lambda \in [0,1]$. Then,

$$\begin{aligned} S(\lambda u_0 + (1-\lambda)u_1) &= \int_\Omega \omega(x,y) \left(A[\lambda u_0 + (1-\lambda)u_1] - z\right)^2 (y)\mathrm{d}y \\ &= \int_\Omega \omega(x,y) \left[\lambda(Au_0 - z) + (1-\lambda)(Au_1 - z)\right]^2 (y)\mathrm{d}y \\ &\leq \lambda S(u_0) + (1-\lambda)S(u_1). \end{aligned}$$

The third inequality holds due to the convexity of $u \mapsto u^2$.

□

Unlike before, data fidelity is now described in a novel way. Based on the smoothed residual, the distance between reconstruction and measurement must not exceed the size of the noise variance.

Definition 3.3 (Local ROF-model)**.** For a linear bounded operator $A : L^2(\Omega) \to L^2(\Omega)$ and the local residual S, we define a new constrained formulation with

$$\operatorname*{arg\,min}_{u \in \mathcal{U}} \ \mathrm{TV}(u),$$

where $\mathcal{U} := \{u \in \mathrm{BV}(\Omega) : S(u) \leq \sigma^2 \text{ a.e. in } \Omega\}$.

In the following sections, the local ROF-model is investigated.

3.1.1. Existence of a solution

We will now prove the existence of a minimizer of the local ROF-model, which requires the following properties of $\mathcal{U}$.

Proposition 3.4. $\mathcal{U} \subset \mathrm{BV}(\Omega)$ is closed and convex.

Proof. $\mathcal{U}$ is closed since the set $\{u \in \mathrm{BV}(\Omega) : u \leq \sigma^2 \text{ a.e. in } \Omega\}$ is closed and S is continuous. The convexity is a direct consequence of the convexity of S from lemma 3.2. □

Proposition 3.5. Assume that A does not annihilate constant functions, i.e. $A\chi_\Omega \neq 0$. Then, the set $\mathcal{U} \subset \mathrm{BV}(\Omega)$ is bounded in $\mathrm{BV}(\Omega)$ or, more precisely with respect to BV-coerciveness,

$$\left[\|u\|_{\mathrm{BV}} \to \infty \right] \ \Rightarrow \ \left[\mathcal{J}(u) := \mathrm{TV}(u) + \int_\Omega \int_\Omega \omega(x,y)(Au - z)^2(y)\mathrm{d}y\mathrm{d}x \to \infty \right].$$

Proof. The proof of theorem 2.27 can be mostly taken over. Let us remember the decomposition for $u \in \mathrm{BV}(\Omega)$, which reads

$$u = w + v, \qquad \text{with} \qquad w = \left(\frac{1}{|\Omega|}\int_\Omega u \mathrm{d}x\right)\chi_\Omega \quad \text{and} \quad \int_\Omega v \mathrm{d}x = 0.$$

By the assumption of $\omega \in L^\infty(\Omega \times \Omega)$ in definition 3.1,

$$\begin{aligned}\mathcal{J}(u) =& \mathrm{TV}(u) + \int_\Omega \int_\Omega \omega(x,y)(Au - z)^2(y)\mathrm{d}y\mathrm{d}x \\ \geq& \mathrm{TV}(v) + \epsilon \left\|Aw + Av - z\right\|^2_{L^2(\Omega)} \\ \geq& \mathrm{TV}(v) + \epsilon \left\|Aw\right\|_{L^2(\Omega)} (c_2 \left\|w\right\|_{L^1(\Omega)} - 2(\left\|A\right\| c_1 \mathrm{TV}(v) + \left\|z\right\|_{L^2(\Omega)})).\end{aligned}$$

From here on, the statement follows with the proof of theorem 2.27. ☐

Thus, the preparations for an existence-proof of a minimizer is placed.

Theorem 3.6. Let $A : L^2(\Omega) \to L^2(\Omega)$ be a linear bounded operator that does not annihilate constant functions. Then the local ROF-model has a solution $\tilde{u}$.

Proof. We first recall that $\mathrm{TV}(\cdot)$ is bounded from below, convex, and lower semi-continuous. A is a continuous linear operator. We have shown that $\mathcal{U}$ is closed and convex. We choose a minimizing sequence $(u_k)_k$. Due to proposition 3.5, $(u_k)_k$ is bounded in $\mathrm{BV}(\Omega)$. Like before, $(u_{k_j})_j$ converges weakly to some $\tilde{u} \in L^2(\Omega)$. By the weak lower semi-continuity, one obtains

$$\mathrm{TV}(\tilde{u}) \leq \liminf_{j\to\infty} \mathrm{TV}((u_{k_j})_j) = \inf_{u\in\mathcal{U}} \mathrm{TV}(u).$$

Since A is a continuous linear operator, $(Au_{k_j})_j$ converges weakly to $A\tilde{u}$. Moreover, since $\mathcal{U}$ is closed and convex, $\tilde{u} \in \mathcal{U}$. ☐

The uniqueness of the minimizer depends on the filter ω. In the next assumption, we investigate the preconditions.

Assumption 3.7. Let $u_1, u_2 \in \mathrm{BV}(\Omega)$ denote two solutions of the local ROF-model, with $u_1 \neq u_2$. If there exist a $\delta > 0$ and $\Omega_\delta \subset \Omega$, with $|\Omega_\delta| > 0$, such that

$$\left(\frac{1}{2}A(u_1 + u_2) - z\right)^2 \leq \frac{1}{2}\left((Au_1 - z)^2 + (Au_2 - z)^2\right) - \delta \quad \text{a.e. in } \Omega_\delta,$$

then there exists ϵ_δ such that

$$S\left(\frac{u_1 + u_2}{2}\right) \leq \sigma^2 - \epsilon_\delta \quad \text{for almost all } x \in \Omega.$$

This allows to make the following statement about the uniqueness of the solution, which is proved in reference [25].

Theorem 3.8. Assume that $A : L^2(\Omega) \to L^2(\Omega)$ is a linear bounded operator that does not annihilate constant functions and suppose $A\mathbb{1} = \mathbb{1}$. In addition, we suppose that assumption 3.7 is satisfied and that

$$\inf_{c\in\mathbb{R}} \int_\Omega \omega(x,y)(c-z)^2(y)\mathrm{d}y \geq \sigma^2 \quad \text{a.e. in } \Omega$$

holds true. Then, for every solution $\tilde{u}$ of theorem 3.6, $A\tilde{u}$ has the same value.

An example for a suitable mean filter that satisfies that assumption 3.7 reads

$$\omega(x,y) = \begin{cases} \frac{1}{\omega_\epsilon^2} & \text{if } \|y-x\|_\infty \leq \frac{w}{2} \\ \epsilon_0 & \text{otherwise} \end{cases},$$

where $x \in \Omega$ is fixed, the essential width of the filter window $w > 0$ is sufficiently small, and $0 < \epsilon_0 \ll \min(1, \frac{1}{\omega_\epsilon^2})$. We choose ω_ϵ such that $\int_\Omega \int_\Omega \omega(x,y)dxdy = 1$. With respect to reference [25], we have $\epsilon_\delta = \epsilon_0 \delta |\Omega_\delta|$.

3.1.2. First-order optimality characterization

In order to solve the problem discretely, we demonstrate a characterization over Eq. (3.1) below. We obtain the connection by introducing a local penalty problem. We prove that this penalty problem converges to a minimizer of the local ROF-model. Furthermore, we show a connection via an equivalent optimality condition between this penalty problem and Eq. (3.1).

Definition 3.9 (Local penalty problem)**.** For the smoothed residual S, we define the local penalty problem by

$$\operatorname*{arg\,min}_{u\in \mathrm{BV}(\Omega)} \left[\mathcal{P}_\gamma(u) := \mathrm{TV}(u) + \frac{\gamma}{2}\int_\Omega \max(S(u)-\sigma^2, 0)^2 \mathrm{d}x \right].$$

Here, $\gamma > 0$ denotes the penalty parameter.

Proposition 3.10. Assume that $A : L^2(\Omega) \to L^2(\Omega)$ is a linear bounded operator that does not annihilate constant functions. Then, the local penalty problem has a solution $u_\gamma \in \mathrm{BV}(\Omega)$ for every $\gamma > 0$. For $\gamma \to \infty$, there exists a subsequence $(u_{\gamma_k})_k$, which converges weakly in $L^2(\Omega)$ to a solution of the local ROF-model.

Proof. Let $\tilde{u}$ be the solution of the local ROF-model.

Concerning the existence result:

First, $\mathcal{P}_\gamma : \mathrm{BV}(\Omega) \to \mathbb{R}$ is weakly lower semi-continuous due to continuity and pointwise convexity of S (as well as $\max(\cdot, 0)$) and the weak lower semi-continuity of $\mathrm{TV}(\cdot)$. Let $(u_k)_k \subset \mathrm{BV}(\Omega)$ denote a minimizing sequence. Then, for all sufficiently large k we have

$$\mathcal{P}_\gamma(u_k) \leq \mathcal{P}_\gamma(\tilde{u}) + 1 = \mathrm{TV}(\tilde{u}) + 1.$$

Since $S(u) \geq 0$ almost everywhere in Ω for any $u \in \mathrm{BV}(\Omega)$, there exists a constant c such that $\|S(u_k)\|_{L^2(\Omega)} \leq c$. By proposition 3.5, $(u_k)_k$ is bounded in $\mathrm{BV}(\Omega)$. The existence follows from the same reasoning as before.

Concerning the convergence result:

Firstly, $(u_\gamma)_\gamma$ is bounded in $\mathrm{BV}(\Omega)$. By lower semi-continuity, we have

$$\mathrm{TV}(\tilde{u}_\gamma) \leq \liminf_{\gamma \to \infty} \mathcal{P}_\gamma(u_\gamma) \leq \mathcal{P}_\gamma(\tilde{u}) = \mathrm{TV}(\tilde{u}) = \inf_{u \in \mathcal{U}} \mathrm{TV}(u),$$

where $\tilde{u}_\gamma$ is a weak limit of a subsequence $(u_{\gamma_k})_k$ in $L^2(\Omega)$.
It remains to show that $\tilde{u}_\gamma \in \mathcal{U}$. Because $\mathcal{P}_\gamma(u_\gamma) \leq \mathcal{P}_\gamma(\tilde{u})$ holds for all $\gamma > 0$,

$$\frac{\gamma}{2} \int_\Omega \max(S(u_\gamma) - \sigma^2, 0)^2 dx \leq \mathrm{TV}(\tilde{u}).$$

As consequence, we obtain

$$\int_\Omega \max(S(u_\gamma) - \sigma^2, 0)^2 \mathrm{d}x \overset{\gamma \to \infty}{\longrightarrow} 0.$$

With continuity of A, weak lower semi-continuity of $\max(\cdot, 0)$ and S, and Fatou's Lemma,

$$S(\tilde{u}_\gamma) \leq \sigma^2 \text{ a.e. in } \Omega$$

and therefore $\tilde{u}_\gamma \in \mathcal{U}$. □

We conclude that

$$\left\| \max(S(u_\gamma) - \sigma^2, 0) \right\|_{L^2(\Omega)} = \mathcal{O}\left(\frac{1}{\sqrt{\gamma}}\right), \tag{3.2}$$

where $\mathcal{O}((a_k)_k)/(a_k)_k \to 0$ for a sequence $(a_k)_k \subset \mathbb{R}_+$ with $a_k \overset{k \to \infty}{\longrightarrow} 0$.

For an arbitrarily fixed $\gamma > 0$, we define

$$\begin{aligned} \lambda_\gamma^\circ &:= \gamma \max(S(u_\gamma) - \sigma^2, 0) \\ \lambda_\gamma &:= \int_\Omega \omega(x, y) \lambda_\gamma^\circ \mathrm{d}x. \end{aligned}$$

A connection between the local penalty problem 3.9 and equation (3.1) becomes clear by observing the Fréchet derivative of both.

Proposition 3.11. The first-order optimality condition of $\mathcal{P}_\gamma$ coincides with that of

$$\frac{1}{2}\int_\Omega \lambda(x)(Au-z)^2(x)\mathrm{d}x + \mathrm{TV}(u),$$

for $2\lambda_\gamma = \lambda(x)$ and all $\gamma > 0$.

Proof. The calculation is directly possible by making use of the chain rule and the Fréchet derivative of S:

$$\begin{aligned}\delta\left[\frac{\gamma}{2}\int_\Omega \max(S(\cdot)-\sigma^2,0)^2 dx\right](u_\gamma, v) &= \int_\Omega \delta S(u_\gamma, v)\gamma \max(S(u_\gamma)-\sigma^2,0)\mathrm{d}y \\ &= \int_\Omega [2(Au_\gamma - z)Av](y) \\ &\qquad \cdot \int_\Omega \gamma\max(S(u_\gamma)-\sigma^2,0)(x)\omega(x,y)\mathrm{d}x\mathrm{d}y \\ &= \int_\Omega 2\lambda_\gamma(y)[(Au_\gamma - z)Av](y)\mathrm{d}y.\end{aligned}$$

The Fréchet derivative of $F(\cdot) = \frac{1}{2}\int_\Omega \lambda(x)(A\cdot - z)^2(x)\mathrm{d}x$ in direction $v \in \mathrm{BV}(\Omega)$ follows with

$$\begin{aligned}\delta F(u,v) &= \lim_{\alpha\to 0}\frac{1}{\alpha}(F(u+\alpha v) - F(u)) \\ &= \lim_{\alpha\to 0}\frac{1}{\alpha}\left(F(u) + 2\alpha\int_\Omega \lambda(x)[(Au-z)Av](x)\mathrm{d}x + \alpha^2\int_\Omega \lambda(x)(Av)^2(x)\mathrm{d}x - F(u)\right) \\ &= \int_\Omega \lambda(x)[(Au-z)Av](x)\mathrm{d}x.\end{aligned}$$

Setting $\lambda(x) = 2\lambda_\gamma(x)$ shows that the two conditions coincide.

□

By convexity of the penalty problem and equation (3.1), both problems have the same minimizer if it exists. As long as $\lambda(x)$ is bounded from below, equation (3.1) has a minimizer by virtue of theorem 2.28. We address this problem with the following theorem.

Theorem 3.12. Assume A is a linear bounded operator that does not annihilate constant functions. Let $\tilde{u}$ denote a weak limit point of $(u_{\gamma_n})_n$ as $\gamma_n \to \infty$. Moreover, we assume that there exists a constant $c > 0$ such that $\left\|\lambda_\gamma^\circ\right\|_{L^1(\Omega)} \leq c$. Then, there exists a $\tilde{\lambda} \in L^\infty(\Omega)$, a bounded Borel measure $\tilde{\lambda}^\circ$, and a subsequence $(\gamma_{n_k})_k$ such that the following conclusions hold true:

1. $\lambda_{\gamma_{n_k}}$ converges weakly to $\tilde{\lambda}$.
2. For all $\varphi \in C(\bar{\Omega})$ holds

$$\int_\Omega \varphi \lambda^\circ_{\gamma_{n_k}} \mathrm{d}x \to \int_\Omega \varphi \mathrm{d}\tilde{\lambda}^\circ, \quad \text{and} \quad \int_\Omega \lambda^\circ_{\gamma_n}(S(u_{\gamma_n}) - \sigma^2)\mathrm{d}x \to 0.$$

See reference [25] for an underlying proof.

If equation (3.2) holds true with $\mathcal{O}\left(\frac{1}{\sqrt{\gamma}}\right)$ replaced by $\mathcal{O}\left(\frac{1}{\gamma}\right)$, i.e. there exists a constant $\tilde{c}$ such that

$$\left\|\max(S(u_{\gamma_n}) - \sigma^2, 0)\right\|_{L^2(\Omega)} \leq \frac{\tilde{c}}{\gamma_n},$$

then $\lambda^\circ_{\gamma_n}$ is uniformly bounded in $L^2(\Omega)$. Therefore, there exists a weak limit $\tilde{\lambda}^\circ \in L^2(\Omega)$ and with theorem 3.12 follows

$$\tilde{\lambda}^\circ \geq 0 \text{ a.e. in } \Omega, \quad S(\tilde{u}) \leq \sigma^2 \text{ a.e. in } \Omega, \quad \lim_{n\to\infty} \int_\Omega \lambda^\circ_{\gamma_n}(S(u_{\gamma_n}) - \sigma^2)\mathrm{d}x = 0.$$

The last relation holds, since $\int_\Omega \tilde{\lambda}^\circ(S(\tilde{u}) - \sigma^2)\mathrm{d}x = 0$. We may equivalently write

$$\tilde{\lambda}^\circ \geq 0 \text{ a.e. in } \Omega, \quad \tilde{\lambda}^\circ = \tilde{\lambda}^\circ + \rho \max(S(\tilde{u}) - \sigma^2, 0) \tag{3.3}$$

for an arbitrary and fixed $\rho > 0$. The connection between the local ROF-model and equation 3.1 is presented but depends on a lower bound $0 < \epsilon \leq \tilde{\lambda}^\circ$ and the assumptions around theorem 3.12. A rigorous investigation for the case that the local ROF-model admits the existence of a multiplier $\tilde{\lambda}^\circ$ is beyond the scope of this work.

However, based on the statement of equation (3.3), we are able to characterize the spatially-dependent $\lambda(x)$ and construct a new iterative strategy, which will be the topic of the next section.

3.1.3. Discrete variance estimation

To derive a meaningful algorithm, we now switch to the discrete case. Hence, the discrete image domain Ω contains $N = N_1 \times N_2$ pixels. The discrete smoothed residual S^w depends on a $w \times w$ window and is defined by

$$S^w(u_{i,j}) := \frac{1}{w^2} \sum_{(s,t)\in\Omega^w_{i,j}} r_{s,t} \qquad r_{s,t} := (z_{s,t} - (Au)_{s,t})^2,$$

$$\Omega^w_{i,j} := \left\{(s+i, t+j) : -\frac{w-1}{2} \leq s, t \leq \frac{w-1}{2}\right\}.$$

The Neumann boundary condition is applied.

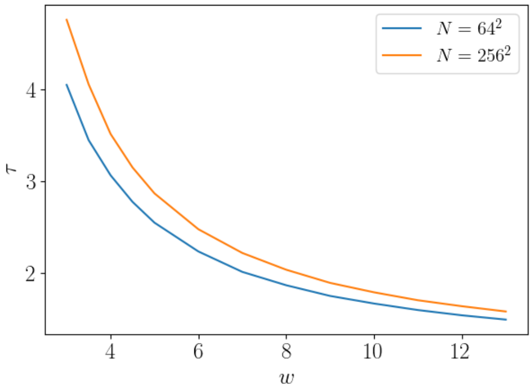


Figure 3.1.: Dependence between the value τ and the window size w with respect to the pixel quantity N to imply left information in the residual S^w.

The discrete version of the noise η in equation (1.1) is also described as array-like value of independent random variables so that $\eta_{i,j} \sim \mathcal{N}(0, \sigma^2)$, with $1 \leq i \leq N_1, 1 \leq j \leq N_2$. It is noticeable that for a small window size the process is not stable enough to require $S^w(\cdot) \approx \sigma^2$. Therefore, a boundary $B(w, N)$ is sought such that $S^w(\cdot) > B(w, N)$ implies left information in the residual. We consider

$$T^w_{i,j} := \frac{1}{\sigma^2} \sum_{(s,t)\in\Omega^w_{i,j}} \eta_{s,t}.$$

The sum of w^2-normally-distributed variables applies $T^\omega_{i,j} \sim \chi^2_{w^2}$.

We set

$$B(w, N) = \tau\sigma^2, \qquad \tau = \frac{1}{w^2} \left(\mathbb{E}(T_{\max}) + \mathrm{STD}(T_{\max})\right), \tag{3.4}$$

with respect to reference [25]. $T_{\max}$ denotes the maximum value of N observations distributed along $\chi^2_{w^2}$. An example calculation for τ in equation (3.4) for N normal distributed variables in $\eta_{i,j}$ can be found in reference [25] and is shown in Fig. 3.1.

With respect to the bound $B(w, N)$, a stable variance estimator $\tilde{S}^w$ is defined with

$$\tilde{S}^w(u_{i,j}) := \begin{cases} S^w(u_{i,j}) & \text{if } S^w(u_{i,j}) \geq B(w, N) \\ \sigma^2 & \text{otherwise} \end{cases}. \tag{3.5}$$

This ensures that the occurrence of information in the residual can be determined via $\tilde{S}^w_{i,j}$, adjusted to the window size. This is followed by the adaptation of the weight $\lambda^{(k)}$ according to the following scheme: $\lambda^{(0)}$ is assigned a small value so that much information remains in the residual. Therefore, $\tilde{S}^w(u^{(0)}_{i,j})$ contains much higher variances. $\lambda^{(1)}$ is increased gradually with respect to this value according to the following procedure:

$$\begin{aligned}\tilde{\lambda}_{i,j}^{(k+1)} &= \tilde{\lambda}_{i,j}^{(k)} + \rho^{(k)} \max(\tilde{S}^w(u_{i,j}^{(k)}) - \sigma^2, 0) = \tilde{\lambda}_{i,j}^{(k)} + \rho^{(k)}(\tilde{S}^w(u_{i,j}^{(k)}) - \sigma^2) \\ \lambda_{i,j}^{(k+1)} &= \frac{1}{w^2} \sum_{(s,t)\in\Omega_{i,j}^w} \tilde{\lambda}_{s,t}^{(k+1)},\end{aligned}$$

where $\rho^{(k)} > 0$. Since $\tilde{S}_{i,j}^w$ is always larger than σ^2, the maximum vanishes. We set $\rho^{(k)} = \frac{1}{\sigma^2}\left\|\tilde{\lambda}_{i,j}^{(k)}\right\|_\infty$ to maintain scaling of the new $\tilde{\lambda}_{i,j}^{(k+1)}$. The following algorithm is created, which adjusts the fidelity weight locally.

Algorithm 3: Spatially Adapted - Total Variation (SA-TV)

```
Data: Operator $A$, data $z$, pixel-wise noise distribution $\mathcal{N}(0,\sigma^2)$
Initialization : $\lambda_{i,j}^{(0)} = \tilde{\lambda}_{i,j}^{(0)} \in \mathbb{R}_+^{N_1\times N_2}$, step size $\zeta > 1$, $L \gg 0$, $k = 0$
    Solve discrete version of Eq. (3.1) with $\lambda(x) = \lambda_{i,j}^{(0)}$ and store solution to $\tilde{u}^{(0)} = u^{(0)}$.
while $\left\|Au^{(k)} - z\right\|_{L^2(\Omega)} > \sigma \cdot \sqrt{|\Omega|}$ do
    $v_{i,j}^{(k)} = (z - Au^{(k)})_{i,j}$
    $\tilde{\lambda}_{i,j}^{(k+1)} = \zeta \min\left(\tilde{\lambda}_{i,j}^{(k)} + \rho^{(k)}\left(\sqrt{\tilde{S}^w(u_{i,j}^{(k)})} - \sigma\right), L\right)$
    $\lambda_{i,j}^{(k+1)} = \frac{1}{w}\sum_{(s,t)\in\Omega_{i,j}^w} \tilde{\lambda}_{s,t}^{(k+1)}$
    Solve via Primal-Dual algorithm the discrete version of
        $\tilde{u}^{(k+1)} = \arg\min_{u\in \mathrm{BV}(\Omega)} \frac{1}{2}\int_\Omega \lambda^{(k)}(x)(Au - v^{(k)})^2(x)\mathrm{d}x + \mathrm{TV}(u).$
    $u_{i,j}^{(k+1)} = u_{i,j}^{(k)} + \tilde{u}_{i,j}^{(k+1)}$
    $k = k + 1$
Output $u^{(k)}$
```

The starting value $\lambda^{(0)} \in \mathbb{R}^{N_1\times N_2}$ should be chosen small in order to evaluate the residual effectively. While the step size ζ describes the continuous increase in each iteration, $\rho^{(k)}$ regulates the relationship from previous $\lambda^{(k)}$ to the information in the residual. $\zeta > 1$ is needed to guarantee convergence with respect to the boundary $B(\omega, N)$. In reference [25], $\zeta = 2$ is derived which comes from the notion of dyadic scales for the appropriate highlighting of residual regions with informations left. The large positive value L is necessary to ensure uniform boundedness of $\lambda^{(k)}$. Otherwise, the local regularization effect would vanish.

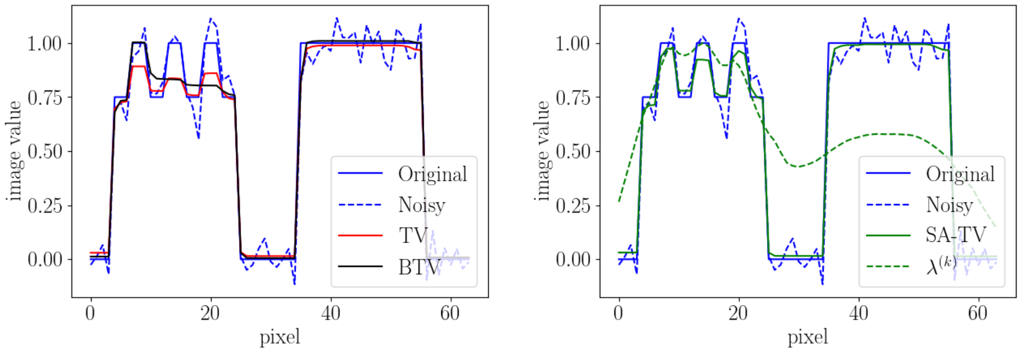


Figure 3.2.: One-dimensional image cross-section for a comparison between different regularization techniques (TV, BTV, SA-TV). $\lambda^{(k)}$ denotes a scaled version of space-different regularization weight. The underlying two-dimensional images are illustrated in Fig. C.3.

The assumption that the underlying noise is normally distributed is only used to calculate the bound $B(w, N)$. With an adjustment of this bound (and thus the variance estimation), other types of noise can be considered. In Fig. 3.2 we see a final result of the SA-TV algorithm. While the Bregman-based method completely eliminates smaller structures, the global weight in TV cannot highlight the specific area between pixels 5 to 25. This area is emphasized in the SA-TV algorithm. Therefore, the reconstruction is closer to the original image.

3.1.4. Combination with Bregman divergence

Note that we extend the TV-regularization in two different ways. On the one hand, we get rid of the loss of contrast with the Bregman iteration. On the other hand, the spatially-adapted version highlights fine structures. Both procedures operate iteratively from an over-regularized solution towards the initial data and stop due to the discrepancy principle. Nevertheless, both approaches differ by different problem resolutions and thus can only be compared with each other to a limited extent. While images with many constant surfaces are better reconstructed via the Bregman algorithm, the SA-TV algorithm provides a better solution for fine objects.

The question arises whether it is possible to obtain both: full contrast and fine structures by combining these techniques. The concept, which is further described under the name *Spatially Adapted - Bregman Total Variation* (SA-BTV), follows the idea: If we solve the regularization by means of the Bregman-based method, one obtains a contrast-faithful solution. From this solution, we measure the local variance using $\tilde{S}^w$ and calculate the possibility of left information in the residual. According to a similar rule, the weight $\lambda(x)$ is now adjusted. This is used to solve the Bregman problem once again but with the new weight $\lambda(x)$. This procedure is repeated until changes in the solution become negligible. We recall the iterative subdifferential

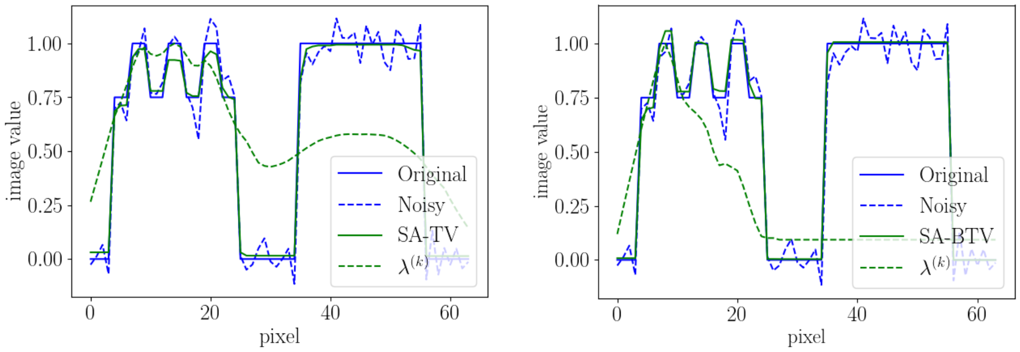
$$p^{(k+1)} = p^{(k)} - \lambda A^*(Au^{(k)} - z).$$

In areas with a higher value of λ, we move faster towards the original data and thus get more details. This leads to the new algorithm.

Algorithm 4: Spatially Adapted - Total Variation with Bregman method (SA-BTV).

```
Input: Maximal iteration count $l_0$, noise-variance $\sigma^2$, small $(\lambda^{(0)})_{i,j} \equiv c \in \mathbb{R}$
for $l = 0, \dots, l_0$ do
    Set $u^{(0)} = 0$, $p^{(0)} = 0, k = 0$
    while $\left\| Au^{(k)} - z \right\|_{L^2(\Omega)} > \sigma\sqrt{|\Omega|}$ do
        Solve via Primal-Dual algorithm the discrete version of
            $u_{\mathrm{BTV}}^{(k+1)} = \arg\min_{u \in \mathrm{BV}(\Omega)} \frac{1}{2} \int_\Omega \lambda^{(k)}(x)(Au - z)^2(x)\mathrm{d}x + D_{\mathrm{TV}}^{p^{(k)}}(u, u_{\mathrm{BTV}}^{(k)}).$
        $p_{i,j}^{(k+1)} = p_{i,j}^{(k)} + \lambda_{i,j}^{(l)} A^*(Au_{\mathrm{BTV}}^{(k+1)} - z)_{i,j}$
        $k = k + 1$
    $u^{(l)} = u_{\mathrm{BTV}}^{(k)}$
    calculate $\tilde{S}^\omega(u_{i,j}^{(l)})$ with (3.5) and $v_{i,j}^{(l)} = (z - Au^{(l)})_{i,j}^2$
    $(\tilde{\lambda}^{(l+1)})_{i,j} = \min\left(\tilde{\lambda}_{i,j}^{(l)} + \rho^{(l)}\left(\sqrt{\tilde{S}^\omega(u_{i,j}^{(l)})} - \sigma\right), L\right)$
    $\lambda_{i,j}^{(l+1)} = \frac{1}{w}\sum_{(s,t)\in\Omega_{i,j}^w} \tilde{\lambda}_{s,t}^{(l+1)}$
    $l = l + 1$
```

Since convergence is achieved via the Bregman iterations, increasing the level of $\lambda^{(k+1)}$ is no longer necessary. Therefore, ζ is omitted. Also, $\rho^{(k)} = \frac{\|\tilde{\lambda}^{(k)}\|_\infty}{\sigma}$ remains.



a) b)

Figure 3.3.: One-dimensional image cross-section for a comparison between different regularization techniques (SA-TV, SA-BTV). $\lambda^{(k)}$ denotes a scaled version of the space-different regularization weight (in subfigure a) for SA-TV, in subfigure b) for SA-BTV). The underlying two-dimensional images are illustrated in Fig. C.3.

In Fig. 3.3, we see the difference in the results of SA-BTV in contrast to SA-TV. The contrast-loss is not reflected in the weight of the SA-BTV $\lambda^{(k)}$. Thus, there is only an increase in the detail range from pixel 5 to 25. The method combines the contrast preservation of the Bregman method with the local highlighting of details from the SA-TV algorithm.

3.2. Spatially-Adapted Regularization for Radon Transform

While most of the results from the previous section are defined for general operators that do not annihilate constant functions, the variance estimation only holds when local structures are maintained in the data space, which is not the case for the Radon transform $\mathcal{R}$. Each area in the sinogram contains global information. Thus, we cannot separate image objects on the basis of the residual $(\mathcal{R}u - z)^2$.

In reference [26], this problem is solved by introducing a second variable $w \in L^2(\Omega)$. Thus, we formulate a new optimization problem over the pair $(u, w) \in \mathrm{BV}(\Omega) \times L^2(\Omega)$. On the one hand, w is the optimum between data fidelity and proximity to u. On the other, u optimizes the TV-regularization in conjunction to the proximity to w. Overall, the problem reads

$$\min_{u \in \mathrm{BV}(\Omega), w \in L^2(\Omega)} \left[\mathcal{L}(u, w) := \frac{\alpha}{2} \|\mathcal{R}w - z\|^2_{L^2(C)} + \frac{1}{2} \int_\Omega \lambda(x)(w-u)^2(x) \mathrm{d}x + \mathrm{TV}(u) \right], \quad (3.6)$$

where $0 < \epsilon_0 \leq \lambda(x) \leq L$.

Theorem 3.13 (Existence of a minimizer)**.** Let $\Omega \subset \mathbb{R}^2$ be an open bounded subset, $\mathcal{R} : L^2(\Omega) \to L^2(C)$ the Radon operator, and $\lambda \in L^\infty(\Omega)$ be bounded in $[\epsilon_0, L]$. Then, the functional $\mathcal{L} : \mathrm{BV}(\Omega) \times L^2(\Omega) \to \mathbb{R} \cup \{\infty\}$ has a unique minimizer $(\tilde{u}, \tilde{w})$.

Proof. $\mathcal{L}$ is convex and lower semi-continuous as a sum of convex and lower semi-continuous functions.

We focus on the BV-boundedness of $\mathcal{L}$ in u (independent of w). First, we obtain

$$\frac{\alpha \left\|\mathcal{R}\right\|^2}{\epsilon_0}\mathcal{L}(u,w) \geq \frac{\alpha \left\|\mathcal{R}\right\|^2}{2} \left\|u-w\right\|^2_{L^2(\Omega)} \geq \frac{\alpha}{2} \left\|\mathcal{R}u - \mathcal{R}w\right\|^2_{L^2(C)} .$$

Then,

$$\begin{aligned}\left(1+\frac{\alpha \left\|\mathcal{R}\right\|^2}{\epsilon_0}\right)\mathcal{L}(u,w) \geq& \frac{1}{2}\left(\sqrt{\alpha}\left\|\mathcal{R}w - z\right\|_{L^2(C)}\right)^2 + \frac{1}{2}\left(\sqrt{\alpha}\left\|\mathcal{R}u - \mathcal{R}w\right\|_{L^2(C)}\right)^2 + \mathrm{TV}(u)\\ \geq& \left(\frac{\sqrt{\alpha}}{2}\left(\left\|\mathcal{R}w - z\right\|_{L^2(C)} + \left\|\mathcal{R}u - \mathcal{R}w\right\|_{L^2(C)}\right)\right)^2 + \mathrm{TV}(u)\\ \geq& \frac{\alpha}{4}\left\|\mathcal{R}u - z\right\|^2_{L^2(C)} + \mathrm{TV}(u).\end{aligned}$$

The second inequality holds due to the convexity of $u \mapsto u^2$. With theorem 2.28 applies the BV-boundedness in u.

The L^2-boundedness of $\mathcal{L}$ in w (for $c = \left\|u\right\|_{L^2(\Omega)} < \infty$) is given, since

$$\frac{1}{\epsilon_0}\mathcal{L}(u,w) \geq \left\|w-u\right\|^2_{L^2(\Omega)} \geq \left(\left\|w\right\|_{L^2(\Omega)} - \left\|u\right\|_{L^2(\Omega)}\right)^2 \geq \left\|w\right\|_{L^2(\Omega)}\left(\left\|w\right\|_{L^2(\Omega)} - 2c\right) \geq \left\|w\right\|_{L^2(\Omega)}$$

for $\left\|w\right\|_{L^2(\Omega)} - 2c \geq 1$.
On the other hand, if $\left\|w\right\|_{L^2(\Omega)} - 2c < 1$, then

$$\left\|w\right\|_{L^2(\Omega)} < 1 + 2c$$

and hence

$$\left\|w\right\|_{L^2(\Omega)} - (1+2c) \leq \mathcal{L}(u,w).$$

Therefore, the coerciveness holds with

$$[\left\|(u,w)\right\|_{\mathrm{BV}(\Omega)\times L^2(\Omega)} = \left\|u\right\|_{\mathrm{BV}(\Omega)} + \left\|w\right\|_{L^2(\Omega)} \to \infty] \Rightarrow [\mathcal{L}(u,w) \to \infty].$$

Let $(u_k, w_k)_k$ be a minimizing sequence for $\mathcal{L}$. As shown, $(u_k, w_k)_k$ is $(\mathrm{BV} \times L^2(\Omega))$-bounded. Therefore, a subsequence $(u_{k_j}, w_{k_j})_j$ exists that converges weakly to some $(\tilde{u}, \tilde{w}) \in \mathrm{BV}(\Omega) \times$

$L^2(\Omega)$. As convexity and lower semi-continuity of $\mathcal{L}$ is given, we obtain

$$\mathcal{L}(\tilde{u}, \tilde{w}) \leq \liminf_{j \to \infty} \mathcal{L}((u_{k_j}, w_{k_j})_j) = \inf_{u \in \mathrm{BV}(\Omega), w \in L^2(\Omega)} \mathcal{L}(u, w).$$

Since $\|\cdot\|^2_{L^2(\Sigma)}$ is strictly convex, the uniqueness directly follows with the injectivity of $\mathcal{R}$. □

The idea behind the extension is the splitting of equation (3.6) into two subproblems:

$$u^{(k+1)} = \operatorname*{arg\,min}_{u \in \mathrm{BV}(\Omega)} \frac{1}{2} \int_\Omega \lambda(x)(w^{(k)} - u)^2(x) \mathrm{d}x + \mathrm{TV}(u), \tag{3.7}$$

$$w^{(k+1)} = \operatorname*{arg\,min}_{w \in L^2(\Omega)} \frac{\alpha}{2} \|\mathcal{R}w - z\|^2 + \frac{1}{2} \int_\Omega \lambda(x)(w^{(k)} - u)^2(x) \mathrm{d}x. \tag{3.8}$$

Theorem 3.14. Let $w^{(0)} \in L^2(\Omega)$ and $\lambda(x)$ be fixed and bounded in $[\epsilon_0, L]$. Then, the sequence $(u^{(k)}, w^{(k)})_k$ converges weakly to the unique minimizer of equation (3.6).

Proof. For a fixed $\lambda(x)$, $(u^{(k)}, w^{(k)})_k$ is a minimizing sequence, since

$$\mathcal{L}(u^{(k+1)}, w^{(k+1)}) \leq \mathcal{L}(u^{(k+1)}, w^{(k)}) \leq \mathcal{L}(u^{(k)}, w^{(k)}).$$

With strict convexity and boundedness of the objective function $\mathcal{L}$, we obtain the weak convergence of the algorithm to the unique minimizer $(\tilde{u}, \tilde{w})$. □

Thus, it is possible to apply a spatially-adapted algorithm to the operator-free part. In the other part, the adjusted weight λ is used in each case. Since the scales of both parts can lead to some problems, the data fidelity weight α is increased until the distance to the data, according to the discrepancy criterion, is reached. In order to guarantee convergence, a spatially-dependent optimization of the weighting only takes place in the first k_0 iterations. This results in an iteration sequence to regularize the inverse problem in the computerized tomography with local weighting parameter.

Variance estimation is now structurally possible as it operates on the image space. Content-wise, this is difficult because a desired noise-distance to w is not directly evident. Once again, reference [26] is the basis for the approach presented here: The variance estimation is based on the noise of the black background in the image structure. This means that errors of the underlying operator are also taken into account. Depending on the situation, this can also lead to problems. Spatially-dependent noise may be highlighted in this way. For this reason, the algorithm should be used with caution. There are also differences in the application areas around the Radon transform, which is shown in the simulations in the next chapter.

Algorithm 5: Versions of SA-TV and SA-BTV for Radon transform.

```
Input: Sinogram $z$, noise variance $\sigma^2$
Initialization: $k = 0$, $w^{(0)} \in L^2(\Omega)$
while $\left\|w^{(k+1)} - w^{(k)}\right\| / \left\|w^{(k)}\right\| + \left\|u^{(k+1)} - u^{(k)}\right\| / \left\|u^{(k)}\right\| > 2 \cdot tol$ do
    if $k < k_0$ then
        Calculate $u^{(k)}, \lambda(x)$ by solving discrete version of Eq. (3.7) via SA-TV or SA-BTV.
    else
        Calculate $u^{(k)}$ by solving discrete version of Eq. (3.7) via TV or BTV with fixed $\lambda(x)$.
    reset $\alpha$ to small value
    while $\left\|\mathcal{R}w^{(k)} - z\right\| > \sigma\sqrt{|\Omega|}$ do
        Calculate $w^{(k)}$ by solving discrete version of Eq. (3.8) via Primal-Dual algorithm with $\alpha$.
        increase $\alpha$
    $k = k + 1$
```

CHAPTER 4

Implementation and Numerical Experiments

A description of the TV-regularized implementation and the associated extensions are the topic of this chapter. In this context, the advantages and disadvantages of the four developed methods, i.e. TV, BTV, SA-TV, and SA-BTV, are discussed. For this purpose, the three different problems from chapter 1 are considered separately: the denoising, the deconvolution, and the reconstruction of the Radon transform.

Functionalities and properties are discussed via two particular images. The head phantom from reference [27] represents an ideal motif with many constant surfaces and objects of different scales. In addition, the black background enables the required variance estimation for the spatially-dependent weighting in case of the Radon operator. A normalized version with shape $(N_1, N_2) = (256, 256)$ and $N = 256^2$ pixels is presented. Since the head phantom is a piecewise constant image, we consider, in addition, a real object measured from a CT-scanner, which was provided by the European Institute for Molecular Imaging (EIMI). Both images are displayed in their original form in figure 4.1.

In order to compare the quality of different reconstructions, two performance quantities are determined

$$\begin{aligned}
\mathrm{PSNR}(u_{\text{original}}, u_{\text{approx}}) =& 20\log_{10}\left(\frac{\max(u_{\text{original}})}{\mathrm{MSE}(u_{\text{original}}, u_{\text{approx}})}\right), \\
\mathrm{MAE}(u_{\text{original}}, u_{\text{approx}}) =& \frac{1}{N_1 N_2}\sum_{i=0}^{N_1-1}\sum_{j=0}^{N_2-1} \left|u_{\text{original}}(i,j) - u_{\text{approx}}(i,j)\right|.
\end{aligned}$$

Since the peak signal-to-noise ratio (PSNR) is defined on the basis of the mean squared error[1], it penalizes high errors more strongly. In order to classify the loss of contrast in a suitable way, we additionally consider the mean absolute error (MAE). To ensure comparability, all of the following pictorial two-dimensional simulations were scaled to the

---

[1] $\mathrm{MSE}(u_{\text{original}}, u_{\text{approx}}) = \frac{1}{N_1 N_2}\sum_{i=0}^{N_1-1}\sum_{j=0}^{N_2-1}\left(u_{\text{original}}(i,j) - u_{\text{approx}}(i,j)\right)^2$

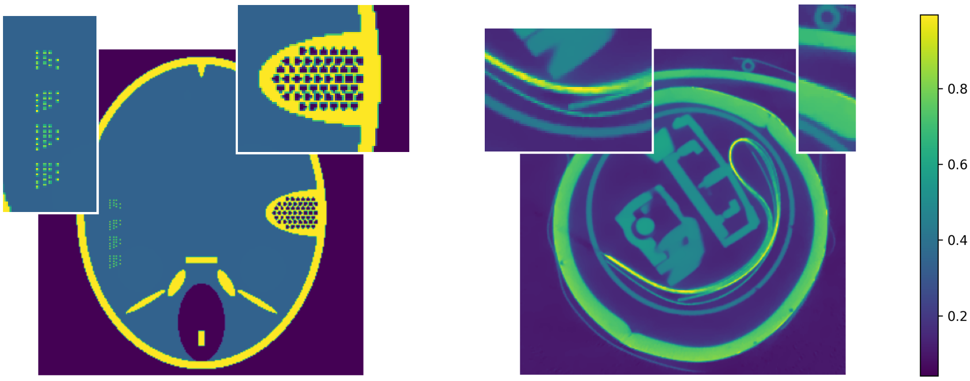


Figure 4.1.: Example figures for further investigations. A head phantom is shown on the left, whereas a toy scan is shown on the right. The upper corners show a section of a respective picture enlarged by a factor of 2.

range between zero and one, i.e.

$$u_{\text{plot}}(i,j) = \begin{cases} 1 & \text{if } u(i,j) > 1 \\ 0 & \text{if } u(i,j) < 0 \\ u(i,j) & \text{else} \end{cases}.$$

4.1. Denoising

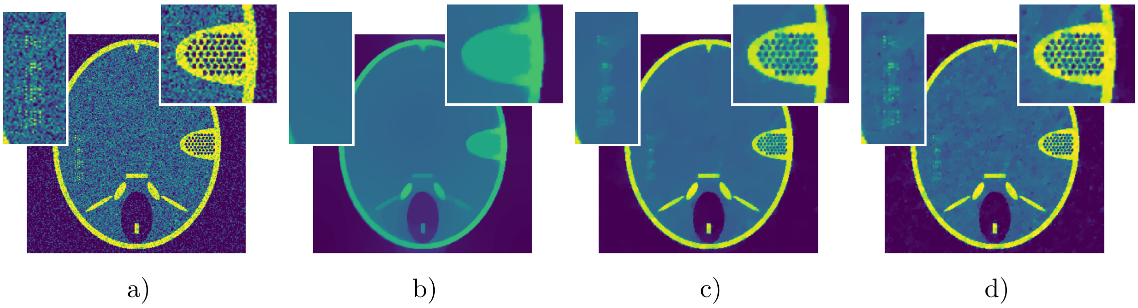
As before, the denoising is used as an introduction to the methods discussed in this work. On the basis of this problem, we describe the individual processes to ensure the fairest possible comparability. Overall, the corrupt image $z_{i,j} = u_{i,j} + \eta_{i,j}$ is examined with $\eta_{i,j} \sim \mathcal{N}(0, \sigma^2)$. First, we study the plain TV-regularization. The well-known optimization problem

$$\arg\min_u \frac{\lambda}{2} \|u - z\|_2^2 + \text{TV}(u)$$

is used. For the Primal-Dual algorithm, the following proximal operators are considered with respect to appendix A

$$\begin{aligned} u = \text{Prox}_{\tau(\frac{\lambda}{2}\|\cdot - z\|_2^2)}(x) \;&\Leftrightarrow\; u_{i,j} = \frac{x_{i,j} + \tau\lambda z_{i,j}}{1 + \tau\lambda}, \\ v = \text{Prox}_{\gamma(\|\cdot\|_1)^*}(x) \;&\Leftrightarrow\; v_{i,j} = \frac{x_{i,j}}{\max(1, |x_{i,j}|)}. \end{aligned}$$

This depends only on the weighting parameter λ, since the step-sizes τ and γ are chosen with respect to $\tau\gamma \|\nabla\|^2 \leq 1$. Figure 4.2 demonstrates the effect different weightings have


a) b) c) d)

	PSNR	MAE
a) Degraded	14.02	0.159
b) $\lambda = 1.0$	17.71	0.067
c) $\lambda = 4.5$	25.2	0.023
d) $\lambda = 7$	25.12	0.032

Figure 4.2.: Regularization with total variation for different weightings. Subfigure a) represents the initial degraded image with added Gaussian noise $\eta_{i,j} \sim \mathcal{N}(0, \frac{1}{25})$. The underlying weight λ in subfigures b)-d) can be found in the table above. The limit is marked in green where Morozov's discrepancy principle takes effect.

on the resulting image. In the following, the best possible one is chosen by systematically increasing an initially small λ until the discrepancy principle, i.e. $\|u - z\|_2 < 256\sigma$, takes effect.

We have already seen the effect of different weights on a square in example 1.1. The behavior observed in that example is confirmed on the head phantom as well. In summary, an over-regularized solution produces a few constant surfaces which no longer correspond to the original pixel-level. A high loss of contrast is the consequence. For an under-regularized reconstruction, some artifacts occur in the solution. Via Morozov's discrepancy principle, we receive a suitable balancing of both effects.

After the plain TV-regularization method, this procedure was extended by the Bregman scheme. Here, the algorithm iterates from an initial over-regularized solution towards the degraded image along the TV-functional. We defined the optimization formula per iteration with

$$\begin{aligned} u^{(k+1)} &= \arg\min_u \frac{\lambda}{2} \|u - z\|_2^2 - \mathrm{TV}(u^{(k)}) - \langle p^{(k)}, u - u^{(k)} \rangle + \mathrm{TV}(u), \\ p^{(k+1)} &= p^{(k)} - \lambda(u^{(k+1)} - z). \end{aligned}$$

a) b) c)

	PSNR	MAE
TV [Fig. 4.2 c)]	25.2	0.023
c) BTV	25.71	0.017

Figure 4.3.: Bregman iterations with respect to the TV-functional, with $\lambda = 0.5$. Subfigures a), b), and c) represent the output of iteration number one, three, and six, respectively. The initial degraded image corresponds to the one shown in Fig. 4.2.

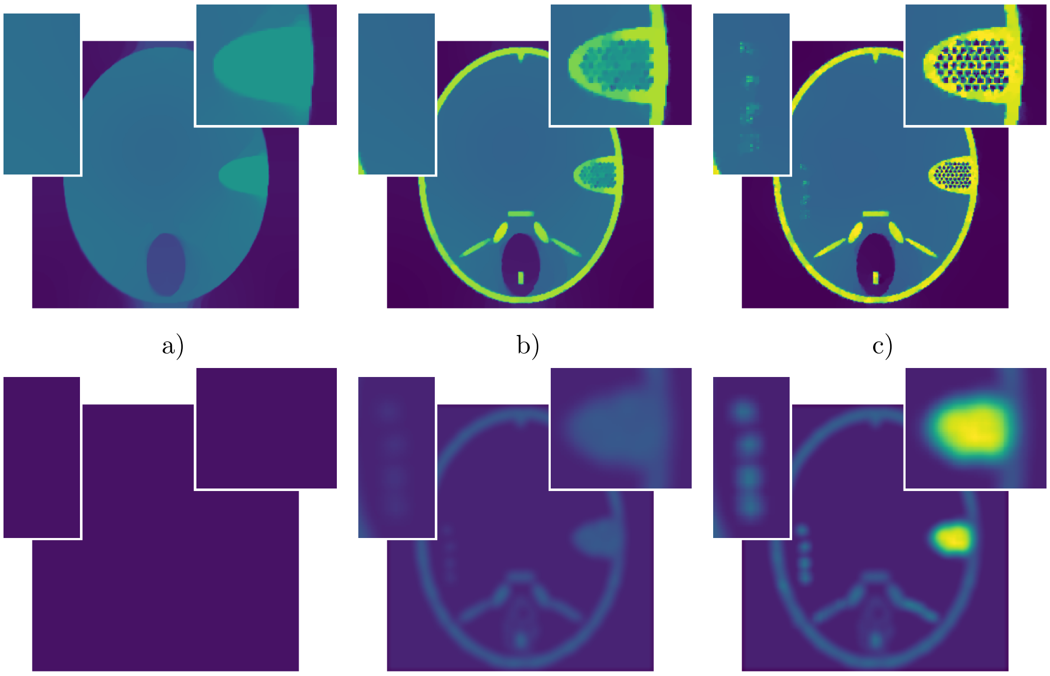
The proximal operators change slightly to

$$\begin{aligned} u = \operatorname{Prox}_{\tau(\frac{\lambda}{2}\|\cdot - z\|_2^2 - \mathrm{TV}(u^{(k)}) - \langle p^{(k)}, \cdot - u^{(k)}\rangle)}(x) \;&\Leftrightarrow\; u_{i,j} = \frac{x_{i,j} + \tau\lambda z_{i,j} + \tau p_{i,j}^{(k)}}{1 + \tau\lambda}, \\ v = \operatorname{Prox}_{\sigma(\|\cdot\|_1)^*}(x) \;&\Leftrightarrow\; v_{i,j} = \frac{x_{i,j}}{\max(1, |x_{i,j}|)}. \end{aligned}$$

Again, details can be found in appendix A. Figure 4.3 displays three different iteration steps of the Bregman scheme of the six iterations that were performed in total for this input. Morozov's discrepancy principle was reached after the sixth iteration. The number of iterations is determined by the weight λ, since in the calculation of $p^{(k)}$ the weight works similar to a step-size. To iterate closer to the limit of discrepancy, a relatively small λ is advised (here $\lambda = 0.5$ is chosen). Especially for larger images this can lead to long calculations until convergence is reached.

Overall, we see a clear improvement in contrast preservation. While the exact details are still missing after three iterations, the constant surfaces are already at the original level. This is further confirmed by the mean absolute error, which is reasonably lower than the one of the plain TV-regularization. Even though an additional increase in the PSNR, smaller details disappear almost completely compared to the previous reconstruction. Therefore, the Bregman method has its great advantage in preserving contrast but not details.

In chapter 3 we introduce the spatially-adapted regularization, which will be investigated as a next method of denoising. We set the step-size $\zeta = 2$ for all numerical tests of this chapter, which is in accordance with reference [25]. The amount of iterations can be determined by



a) b) c)

d) e) f)

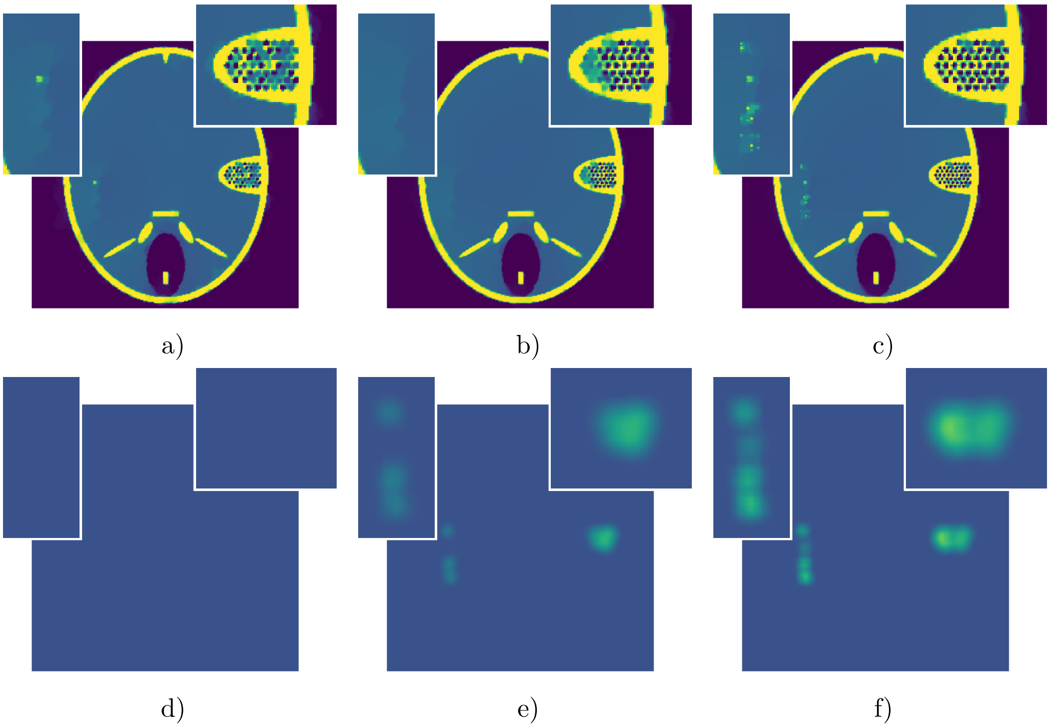
	PSNR	MAE
TV [Fig. 4.2 c)]	25.2	0.023
BTV [Fig. 4.3 c)]	25.71	0.017
c) SA-TV	26.48	0.02

Figure 4.4.: Solutions $u^{(k)}$ and weight $\lambda^{(k)}$ generated by the SA-TV algorithm. Subfigures a), b), and c) display the solution and of the first, second, and third iteration, while subfigures d), e), and f) shows the weight in the corresponding iterations, respectively. The initial degraded image corresponds to the one shown in Fig. 4.2.

the initial λ as stated before. Furthermore, we have observed a rather stable result for different starting values (see Fig. C.5).

Fig. 4.4 illustrates the different states of the SA-TV algorithm. As this simulation uses the same initial starting weight $\lambda^{(0)}$, the first iteration of this method yields the same result the first iteration of the Bregman algorithm. After the first iteration, the weight in the SA-TV method is updated to $\lambda^{(1)}$ and the resulting images begin to differ when compared to the Bregman scheme.

To compensate the loss of contrast, areas of small detail are also accompanied by a much higher weight λ. This phenomenon is shown in the Fig. 4.4. The bound $B(w, N)$ divides between noise and information effectively: no unnecessary weighting-region occurs. Altogether, the higher mean absolute error indicates an insufficient contrast preservation via variance estimation when compared to the Bregman method. Despite this, the PSNR



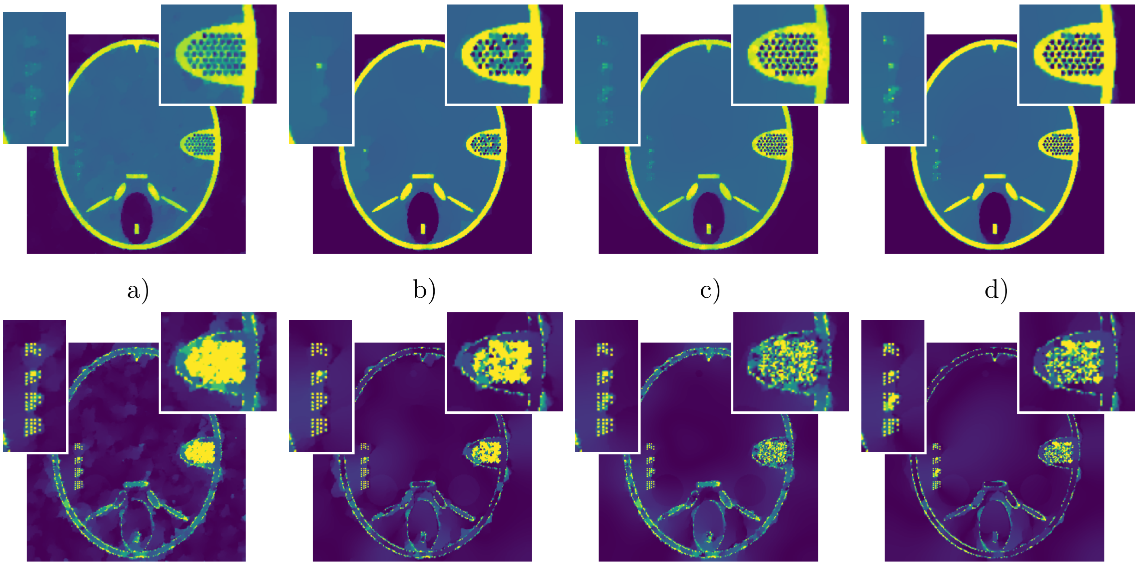
	PSNR	MAE
TV [Fig. 4.2 c)]	25.2	0.023
BTV [Fig. 4.3 c)]	25.71	0.017
SA-TV [Fig. 4.4 c)]	26.48	0.02
c) SA-BTV	27.11	0.015

Figure 4.5.: Solutions $u^{(k)}$ and weight $\lambda^{(k)}$ generated by the SA-BTV algorithm. Subfigures a), b), and c) display the solution and of the first, second, and fourth iteration, while subfigures d), e), and f) shows the weight in the corresponding iterations, respectively. The initial degraded image corresponds to the one shown in Fig. 4.2.

increases and it is recognizable that the reconstruction of detailed regions is much more effective. In summary, small structures can be well reconstructed and no new edges are created - at a loss of contrast.

In the SA-BTV algorithm, we have combined the contrast-preserving Bregman method with the detail-focused SA-TV algorithm. Fig. 4.5 illustrates the output of three iterations. Since the initial weight $\lambda^{(0)}$ coincides with that from the previous methods, the first iteration outputs the same image as in the BTV algorithm again. Based on the variance estimation of this image, a new weighting is calculated. This does not highlight any contrast region and focuses only on the two underrepresented regions, which is one of the main differences to the soley spatially-adaptive method. Further iterations emphasize this effect.

Even though both PSNR and MAE are improved compared to the other variants, it still



a) b) c) d)

e) f) g) h)

	$\sigma = 0.3$		$\sigma = 0.2$		$\sigma = 0.1$	
	PSNR	MAE	PSNR	MAE	PSNR	MAE
Degraded	10.5	0.238	14.02	0.159	20.04	0.079
a) TV	22.13	0.032	25.2	0.023	30.1	0.013
b) BTV	22.89	0.026	25.71	0.017	31.85	0.008
c) SA-TV	23.91	0.026	26.6	0.019	31.8	0.011
d) SA-BTV	24.06	0.022	27.11	0.015	32.42	0.008

Figure 4.6.: Overview for all regularization techniques and different noise rates $\eta_{i,j} \sim \mathcal{N}(0, \sigma^2)$. The illustrated images are the results corresponding to the second column ($\sigma = 0.2$) of the table above.

generates new edges on fine structures which is typical for Bregman-based solutions. Again, the number of iterations depends on the starting weight. However, by running the Bregman method repeatedly, a considerable amount of Primal-Dual iterations have to be executed.

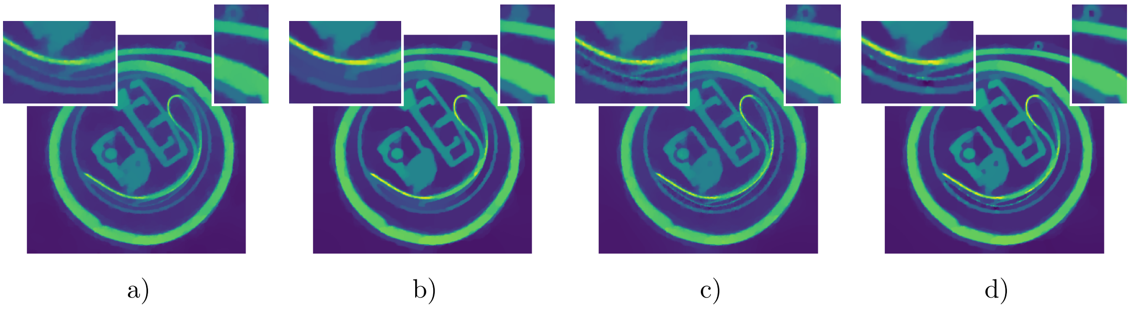
To finish the denoising, we compare all four outputs side-by-side. Fig. 4.6 shows the head phantom reconstruction for different noise levels. Based on the residual images[2], we can clearly observe the advantages of every individual extension. While the two Bregman-based methods map the surfaces much better (notice the yellow ring), the locally-adapted versions are stronger in highlighting details. The respective advantages can be seen for all tested noise levels in the PSNR and MAE.

Supplementary, Fig. 4.7 shows a reconstructed representation of 45 samples (15 different slices with 3 different seeds on each slice). The enlarged sections are particularly noteworthy

---

[2] The illustrated residual is chopped from above to visualize the loss of contrast in an easier way, i.e. $\min(|u_{\text{approx}} - u_{\text{original}}|, 0.2)$.

and show similar behaviors as before: The Bregman-based methods can reconstruct the bright areas more originally, whereas the spatial adapted ones can recover fine structures much easier. Furthermore, the described effects remain present for all tested σ across different images. Therefore, one may conclude that these effects hold generally, i.e. they do not originate from specifically selected images.


a) b) c) d)

	M_{PSNR}	σ_{PSNR}	M_{MAE}	σ_{MAE}
Degraded	20.01	0.02	0.079	0.0005
a) TV	28.32	0.85	0.024	0.0036
b) BTV	29.19	0.74	0.019	0.0024
c) SA-TV	29.76	0.79	0.021	0.0027
d) SA-BTV	29.92	0.81	0.018	0.0024

Figure 4.7.: Overview over all regularization techniques. Combination of 15 CT slices and 3 different random seeds. All of these slices were degraded by the noise ratio $\eta_{i,j} \sim \mathcal{N}(0, \frac{1}{100})$.

4.2. Deconvolution

Next, we consider the implementation and simulated results of the deconvolution problem. With respect to the forward operator A and associated kernel k_A, we obtain the optimization problem

$$\arg\min_u \frac{\lambda}{2} \|Au - z\|_2^2 + \mathrm{TV}(u).$$

Due to the properties of the Primal-Dual algorithm, an adjustment of the proximal operator suffices for the fidelity term. We obtain these via

$$u = \mathrm{Prox}_{\tau(\frac{\lambda}{2}\|A(\cdot)-z\|_2^2)}(x) \;\Leftrightarrow\; u = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(x) + \tau\lambda\mathcal{F}(z)\mathcal{F}(k_A)^*}{1+\tau\lambda\mathcal{F}(k_A)^2}\right),$$
$$v = \mathrm{Prox}_{\gamma(\|\cdot\|_1)^*}(x) \;\Leftrightarrow\; v = \frac{x}{\max(1, |x|)},$$

where $\mathcal{F}$ denotes the Fourier transform. The modification based on operator A is valid with respect to appendix A, the convolution theorem, and reference [11].

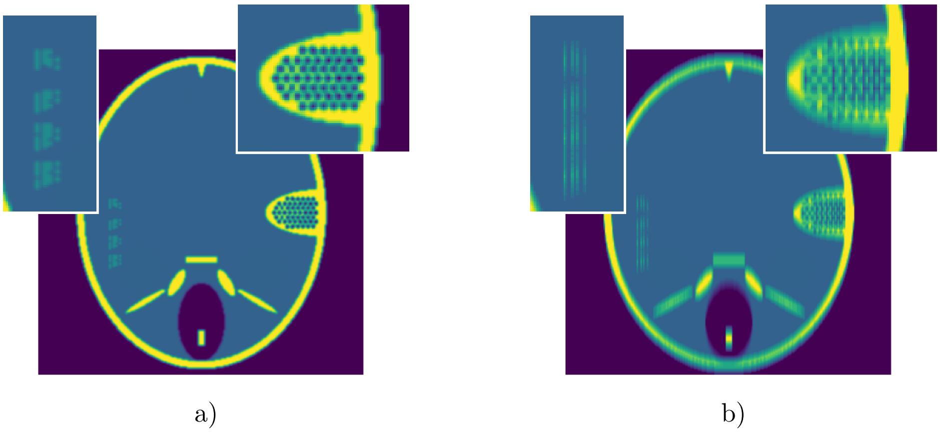


Figure 4.8.: Convolved images with k_{blurring} in subfigure a) and with k_{motion} in subfigure b). These images do not contain any noise.

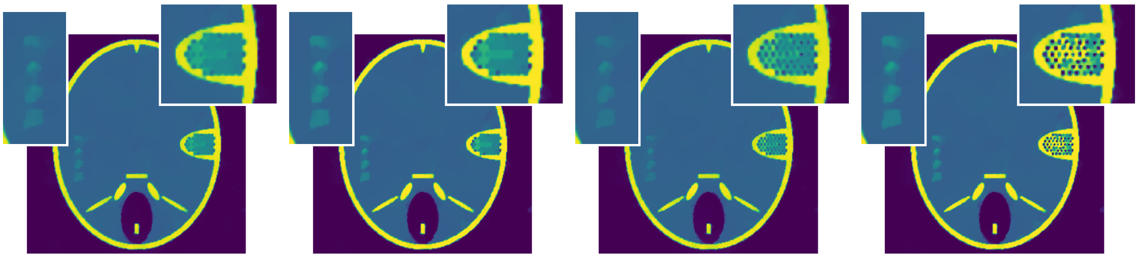
We investigate the deconvolution results for every algorithm on the basis of a blurring and a motion kernel (see the effect on the head phantom in Fig. 4.8). In particular, we consider

$$k_{\text{blurring}} = \frac{1}{16}\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix} \quad \text{and} \quad k_{\text{motion}} = \frac{1}{10}\begin{pmatrix} 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \end{pmatrix}.$$

First of all, we focus on the blurring kernel. Figure 4.9 shows the results for different noise magnitudes added to the deblurred head phantom. The resulting performance - we measure via PSNR and MAE - depends on the underlying noise level. It is noticeable that the Bregman-based methods perform worse for smaller noise: A decreasing standard deviation of the Gaussian noise increases the weight of the fidelity term. Therefore, contrasts are preserved better. For such cases, the advantage of the Bregman-based methods is omitted and a minimally worse result is obtained due to more interrelated areas. On the other hand, the mean absolute error shows a much better contrast retention for high levels of noise.

In contrast, the spatially-adapted weighting produces better results as long as the level of noise is not excessive. For a higher rate of degradation, it is no longer possible to distinguish between image details and noise. Thereby, the SA-TV and SA-BTV algorithms output an equivalent result compared to their global-weighting counterparts. For the illustrated results (degraded with $\sigma = 0.1$), there is a higher data fidelity in the detailed area on the right region. The visibility of the points on detailed area to the left side was, however, not improved due to their tiny size and the fact that the residual is determined by the blurred variant. The latter originates from the fact that we convolve in the smooth residual S again.

The described differences are confirmed in figure 4.10. Here, a stable performance is verified



a) b) c) d)

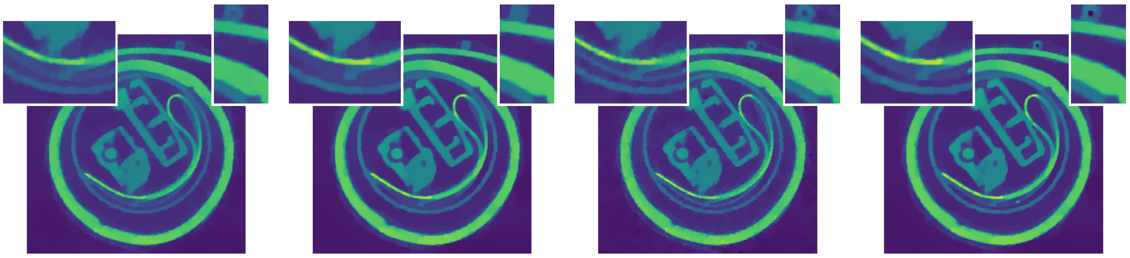
	$\sigma = 0.2$		$\sigma = 0.1$		$\sigma = 0.04$		$\sigma = 0.01$	
	PSNR	MAE	PSNR	MAE	PSNR	MAE	PSNR	MAE
Degraded	13.55	0.159	18.41	0.083	22.14	0.033	23.35	0.0198
a) TV	21.71	0.029	23.25	0.0197	25.79	0.0121	28.18	0.007
b) BTV	22.12	0.023	23.48	0.018	25.69	0.0113	27.92	0.0069
c) SA-TV	21.8	0.028	23.6	0.019	25.98	0.012	28.44	0.007
d) SA-BTV	22.11	0.022	24.02	0.017	25.35	0.011	27.84	0.0069

Figure 4.9.: Reconstruction from a convolved image with k_{blurring} and different noise $\eta_{i,j} \sim \mathcal{N}(0, \sigma^2)$. The presented subfigures demonstrate the outputs of the four different algorithms (identified in the table above), with $\sigma = 0.1$.

by the real toy-scan. Since the noise level $\sigma = 0.1$ is neither too small nor too large, the previously mentioned observations are confirmed: On the one hand we have an increased contrast due to the Bregman-expansion. On the other hand we obtain a higher level of detail with the spatially-adjusted weighting. Nonetheless, we can still notice a difference. While the points of the phantom motif all separated similarly poorly, lines in the real motif are reconstructed much better in SA-TV. For the SA-BTV algorithm the same are displayed as a surface (see the zoomed box on the left). This leads into an overall higher M_{PSNR} in the SA-TV algorithm.

The observed phenomenon cannot be adopted directly to the motion kernel k_{motion} (see figure 4.11). The spatial enhancement is even less pronounced. A higher data fidelity shows hardly any improvement compared to the standard methods. In comparison, the advantage of the Bregman-based methods, the preservation of contrast, remains recognizable for a decreasing data fidelity. This was derived from the MAE.

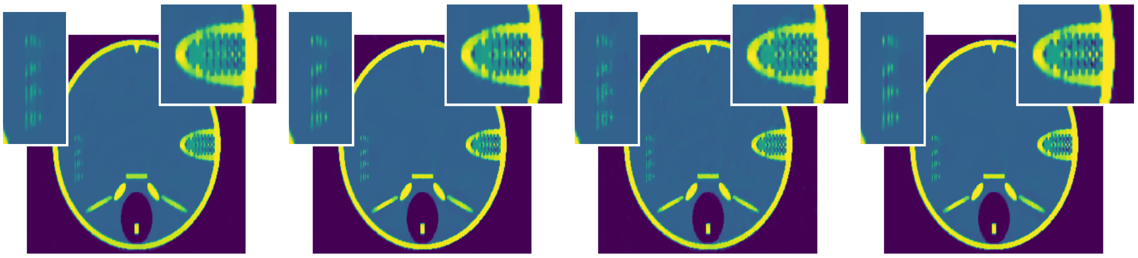
Overall, it can be concluded that Bregman-based methods are more effective for deconvolutions when data fidelity is low. For a higher rate of noise these approaches should receive more attention. Such a general statement cannot be made for the spatially-adapted methods SA-TV and SA-BTV as they are dependent on the operator (here on the underlying kernel) and the data space. If the degradation is too high, for example for the motion kernel, these methods are ill suited.



a) b) c) d)

	M_{PSNR}	σ_{PSNR}	M_{MAE}	σ_{MAE}
Degraded	19.78	0.048	0.082	0.0004
a) TV	27.38	0.82	0.023	0.0032
b) BTV	27.49	0.8	0.022	0.0030
c) SA-TV	28.04	0.79	0.023	0.0036
d) SA-BTV	27.89	0.82	0.021	0.0031

Figure 4.10.: Blurred image reconstruction with added Gaussian noise $\eta_{i,j} \sim \mathcal{N}(0, \frac{1}{100})$. Combination of 15 CT slices and 3 different random seeds. The subfigures represent the results of the algorithms a) TV, b) BTV, c) SA-TV, and d) SA-BTV.



a) b) c) d)

	$\sigma = 0.1$		$\sigma = 0.04$		$\sigma = 0.01$	
	PSNR	MAE	PSNR	MAE	PSNR	MAE
Degraded	16.27	0.087	18.14	0.053	18.58	0.038
a) TV	20.48	0.032	23.23	0.0192	28.73	0.0082
b) BTV	20.77	0.029	23.67	0.0167	28.67	0.0079
c) SA-TV	20.86	0.03	23.73	0.0188	28.79	0.0093
d) SA-BTV	20.75	0.029	23.74	0.0165	28.84	0.0076

Figure 4.11.: Motioned image reconstruction from convolved image with k_{motion} and different noise level $\eta_{i,j} \sim \mathcal{N}(0, \sigma^2)$. The subfigures represent the output of the algorithms: subfigure a) TV, b) BTV, c) SA-TV and d) SA-BTV with $\sigma = 0.04$ (column 2).

4.3. Radon Transform

We close the analysis with the inverse problem included in computerized tomography. For this purpose, the general optimization problem

$$\arg\min_{u} \frac{\lambda}{2} \left\| \mathcal{R}u - z \right\|_2^2 + \mathrm{TV}(u)$$

is solved, where $\mathcal{R}$ denotes to Radon transform, $\lambda > 0$, and z is a measured sinogram.

While the adaptation of the fidelity term (from a convolution to the Radon operator) is theoretically possible without any constraints, this turns out to be immensely difficult for the implementation. In the calculation of $\mathrm{Prox}_{\frac{\lambda}{2}\|\mathcal{R}\cdot - z\|_2^2}$, which can be found in the appendix, the inversion of the operator was assumed. In the discrete form, a true inverse is hardly possible due to a limited number of angles. With an explicit limitation of the limited-angle or low-dose computerized tomography, the inversion is completely impossible. To still be able to apply the Primal-Dual algorithm, we follow reference [28] and redefine F,G, and K. As consequence, we can no longer separate between fidelity and regularization terms.

The following space definitions are considered

$$X := \mathbb{R}^{N_1 \times N_2}, \quad Y := S \times V, \quad S := \mathbb{R}^{N_\theta \times N_s}, \quad V := \mathbb{R}^{2N_1 \times N_2}.$$

Here, X denotes the primal space, which defines the images as before. The dual space Y is now a product space of sinograms S and gradients V. The linear operator $K : X \to Y$ maps between those two spaces, with

$$K(x) \coloneqq (\mathcal{R}x, \nabla x), \qquad K^*(s, v) = \mathcal{R}^* s - \nabla^* v.$$

We define $F : X \to \mathbb{R} \cup \{\infty\}$ and $G : Y \to \mathbb{R} \cup \{\infty\}$, with

$$F(x) \coloneqq 0, \qquad G(s, v) \coloneqq \frac{\lambda}{2} \left\| s - z \right\|_2^2 + \left\| v \right\|_1.$$

In the case of a Bregman-based reconstruction, we divide the linear parts into F. Now the proximal operator of G^* must be calculated. Due to properties of the Fenchel-Legendre conjugates, it is possible to consider them separately

$$G^*(s, v) = \left(\frac{\lambda}{2} \left\| \cdot - z \right\|^2 \right)^* (s) + (\left\| \cdot \right\|_1)^*(v), \qquad \text{with} \quad \left(\frac{\lambda}{2} \left\| \cdot - z \right\|^2 \right)^* (s) = \frac{1}{2\lambda} \left\| s \right\|^2 + \langle s, z \rangle.$$

Thus, the following three proximal operators must be calculated each iteration

$$\begin{aligned} s = \operatorname{Prox}_{\gamma(\frac{\lambda}{2}\|\cdot - z\|_2^2)^*}(x) \;&\Leftrightarrow\; s_{i,j} = \frac{\lambda}{\lambda+\gamma}\left(x_{ij} - \gamma z\right), \\ v = \operatorname{Prox}_{\gamma(\|\cdot\|_1)^*}(x) \;&\Leftrightarrow\; v_{i,j} = \frac{x_{i,j}}{\max(1, |x_{i,j}|)}, \\ u = \operatorname{Prox}_0(x) \;&\Leftrightarrow\; u_{i,j} = x_{i,j}. \end{aligned}$$

Overall, the Primal-Dual algorithm now becomes:

Algorithm 6: Primal-Dual algorithm for regularized reconstruction of the Radon transform.

```
Data: Measurement $z \in S$.
Result: $\arg\min_u \frac{\lambda}{2}\|\mathcal{R}u - z\|_2^2 + \mathrm{TV}(u)$
Initialization : $\tau, \gamma > 0$,
        $(u^{(0)}, (s^{(0)}, v^{(0)})) \in X \times (S \times V)$ and set $\tilde{u}^{(0)} = u^{(0)}$.
while $\left\|u^{(k)} - u^{(k-1)}\right\| / \left\|u^{(k)}\right\| > tol$ do
    $s^{(k+1)} = \frac{\lambda}{\lambda+\gamma}\left(s^{(k)} + \gamma\mathcal{R}\tilde{u}^{(k)} - z\right)$
    $v^{(k+1)} = \frac{v^{(k)}+\gamma\nabla\tilde{u}^{(k)}}{\max(1,|v^{(k)}+\gamma\nabla\tilde{u}^{(k)}|)}$
    $u^{(k+1)} = u^{(k)} - \tau(\mathcal{R}^* s^{(k+1)} + \nabla^* v^{(k+1)})$
    $\tilde{u}^{(k+1)} = u^{(k+1)} + (u^{(k+1)} - u^{(k)})$
    $k = k + 1$
```

Therefore, we can bypass the inversion requirement and obtain a suitable alternative via the adjoint operator.

For further simulations, we consider $z_{i,j} = \mathcal{R}u_{i,j} + \eta_{i,j}$, with $\eta_{i,j} \sim \mathcal{N}(0, \sigma^2)$ and $\sigma = \frac{\|\mathcal{R}u)\|_\infty}{100}$. In order to get an overview of the advantages and disadvantages of the reconstruction algorithms, we consider the three different cases: a low-dose, a limited-angle, and the fully measured variant.

The different TV-regularization methods are compared with the filtered back-projection (FBP) and the Kaczmarz' method [3] [29]. The Astra Toolbox [30] was used for the calculation of the discrete (adjoint) Radon transform, since it satisfies

$$\langle \mathcal{R}x, \mathcal{R}x\rangle \approx \langle x, \mathcal{R}^*(\mathcal{R}x)\rangle \quad \forall x \in X,$$

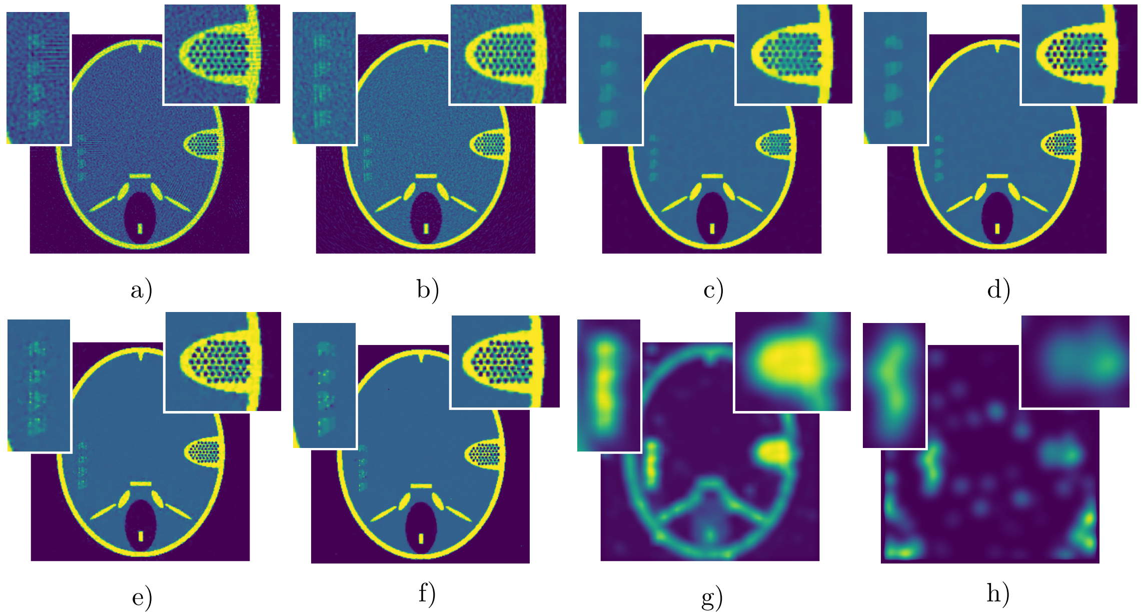
in addition to a performant runtime.

---

[3] Another iterative method named after the Polish mathematician Stefan Kaczmarz, often used in computerized tomography (here computed with 5 iterations).

4.3.1. Full-measured computerized tomography

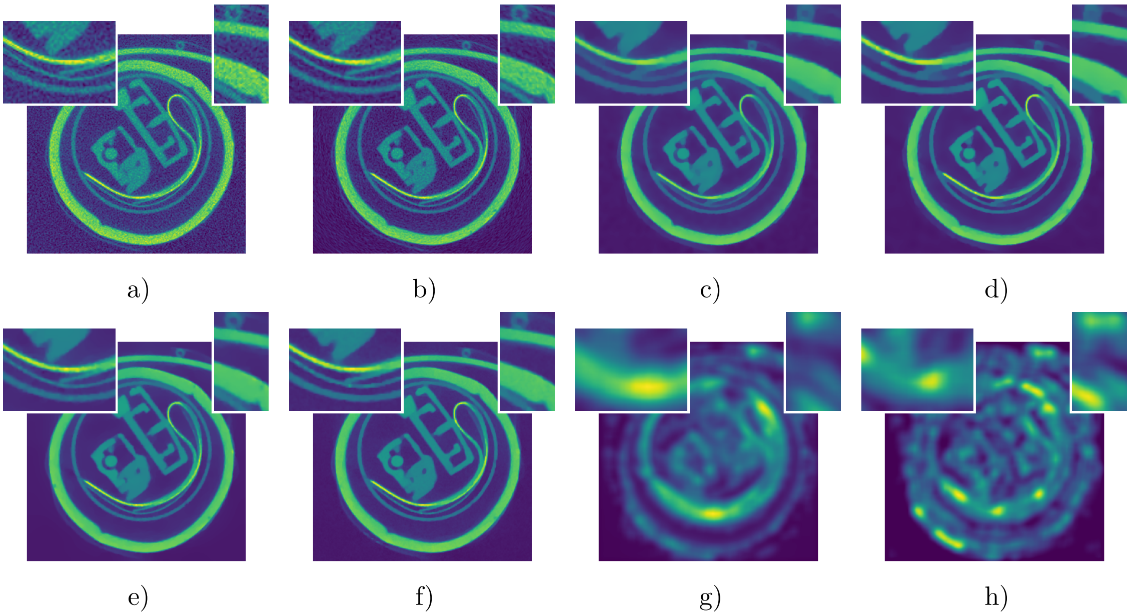
First, we consider the results for the fully measured Radon operator. In this case, an additive data-noise is the only constraint for the reconstruction. As before, while Fig. 4.12 demonstrates the simulation on the phantom, Fig. 4.13 illustrates the reconstructions of a slice from the toy motif. Both images follow the same behavior: There is a clear difference between regularized procedures and direct reconstruction. In the reconstruction of the filtered back-projection as well as in the Kaczmarz' method, noisy effects persist. Due to the discrepancy principle, noise is drastically reduced in the solutions of the regularized optimization approaches. A loss of contrast is also be recognizable here: Both Bregman-based methods, BTV and SA-BTV, perform slightly better when compared to the alternatives (note the mean absolute error).



	PSNR	MAE
a) FBP	18.55	0.092
b) Kaczmarz'	22.51	0.053
c) TV	25.86	0.015
d) BTV	26.6	0.137
e) SA-TV	27.13	0.015
f) SA-BTV	27.48	0.012

Figure 4.12.: Full-measured CT-reconstructions are shown with a measurement of 180 angles with additive Gaussian noise $\eta_{i,j} \sim \mathcal{N}(0, \frac{\|\mathcal{R}u)\|_\infty}{100})$. Subfigures a) - f) illustrate solutions of the algorithms named in the table above. Subfigure g) shows the weight of the SA-TV algorithm, while subfigure h) represent the weight in the SA-BTV algorithm.

If we take a closer look to the level of detail, we obtain similar results as before in the case of denoising. The emphasis on detailed regions is clearly visible for both motifs. Even in the head phantom, the pointy area on the left can be distinguished once again. This is due to the fact that the variance estimation of the residual takes place on the space of images. The weightings also indicate which areas have been focused on in the process. In SA-TV the loss of contrast is also noticeable. However, in SA-BTV only detailed areas are weighted higher. A high variance in the corners remains, noticeable in figure 4.12. This shows a danger in the variance estimation. Depending on the underlying operator and measured object, the error becomes also spatially-dependent. This is a vulnerability because any area above a specified boundary is highlighted. In this case, it is not yet relevant because the error structure is too disjointed. However, exactly this condition becomes important in the limited-angle variant. The assumption of a normal distributed error is not fulfilled and thus $B(w, N)$ is not sharp.

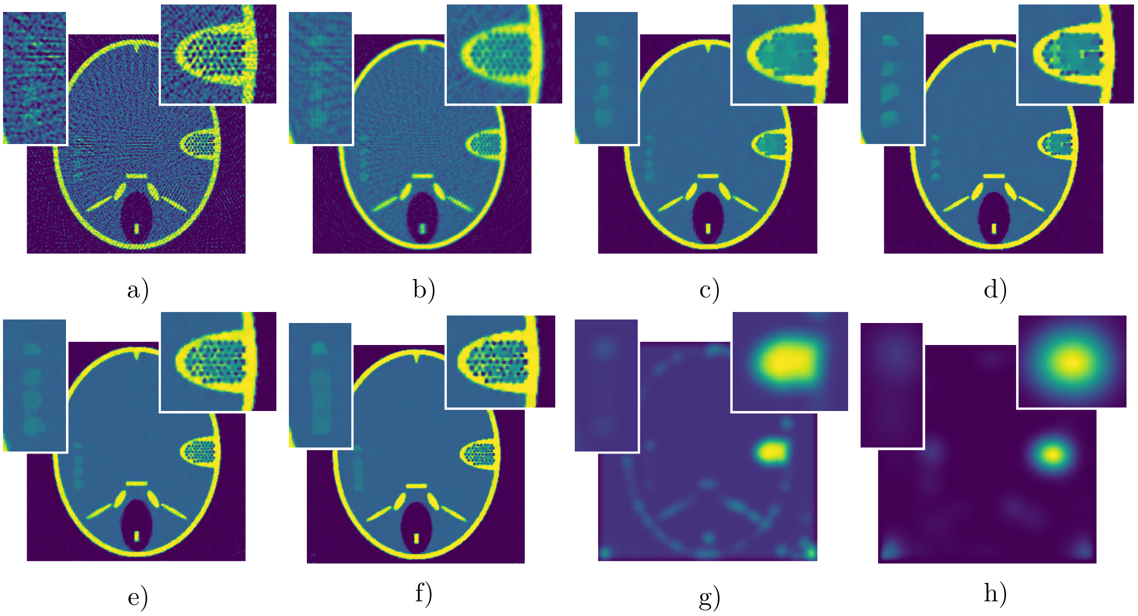

a) b) c) d)
e) f) g) h)

	M_{PSNR}	σ_{PSNR}	M_{MAE}	σ_{MAE}
a) FBP	21.24	0.94	0.069	0.0061
b) Kaczmarz’	26.58	0.89	0.0368	0.0033
c) TV	29.46	0.85	0.0176	0.0034
d) BTV	29.91	0.82	0.0156	0.0030
e) SA-TV	31.24	0.84	0.0174	0.0035
f) SA-BTV	31.77	0.82	0.0146	0.0031

Figure 4.13.: The full-measured CT-reconstruction for a combination of 15 CT slices and 3 different random seeds is shown for 180 angles and additive noise ratio $\eta_{i,j} \sim \mathcal{N}(0, \frac{\|\mathcal{R}u)\|_\infty}{100})$. Subfigures a) - f) shows the solutions named in the table above. Subfigure g) denotes the weight of the SA-TV algorithm, whereas subfigure h) denotes the weight of the SA-BTV algorithm.

4.3.2. Low-dose computerized tomography

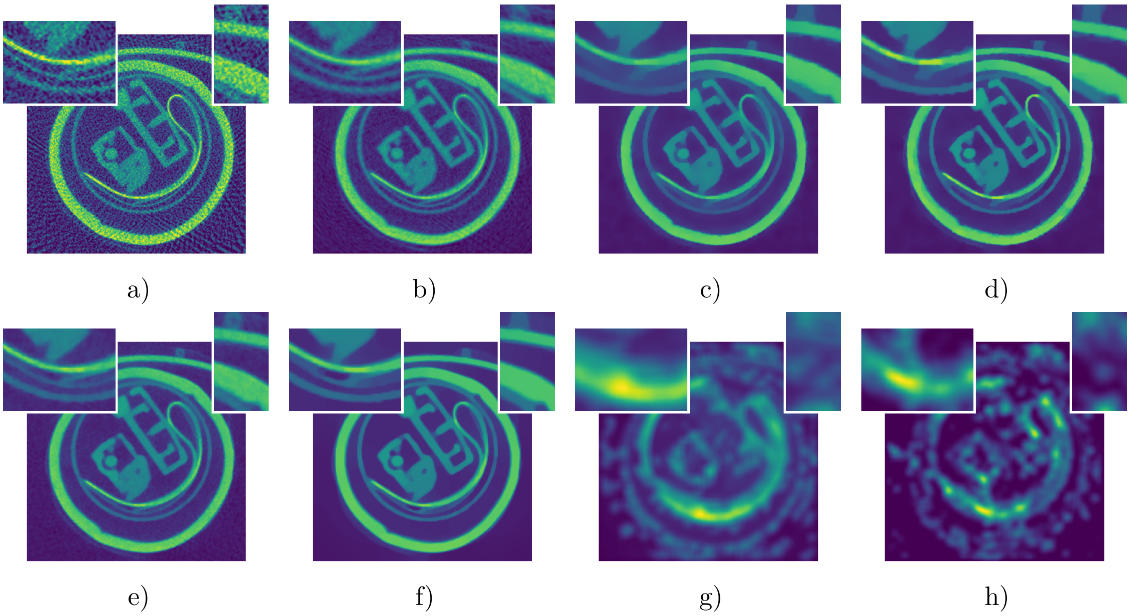
A very similar behavior can be identified in the low-dose version of a computerized-tomography reconstruction, displayed in figures 4.14 and 4.15. For this purpose, the measurement is limited to a third of the angles before. This leads to a higher deterioration of all results. However, the previous observations can be transferred to this situation, as well. While the Kaczmarz' method compensates the error in a much better way than the FBP, the regularized solution produce much cleaner surfaces. The advantage in the MAE for the Bregman-based methods is still apparent.


a) b) c) d)
e) f) g) h)

	PSNR	MAE
a) FBP	14.54	0.144
b) Kaczmarz'	20.8	0.054
c) TV	23.26	0.02
d) BTV	23.65	0.018
e) SA-TV	24.02	0.022
f) SA-BTV	24.31	0.018

Figure 4.14.: The low-dose CT-reconstruction was performed on a measurement of 60 uniform-distributed angles and noise level $\eta_{i,j} \sim \mathcal{N}(0, \frac{\|\mathcal{R}u)\|_\infty}{100})$. Subfigures a) - f) illustrate the solutions named in the table above. Subfigure g) denotes the weight of the SA-TV algorithm, while subfigure h) shows the weight in the SA-BTV scheme.

In the head phantom, we can again clearly see the difference between the detailed area to the left and to the right. The left-hand fine structures produces an insufficient variance in the residual to exceed the border $B(w, m)$. Therefore, there is a disadvantage in this region. This is different on the right-hand side. Here, the solution gets an upgrade. This creates an overall more detailed image. As before, there is an emphasis on the corners. The error remains incoherent in these areas and is thus not incorporated into the solution. These results are also confirmed on the toy subject. Especially the line in the left inset can be reconstructed in the SA-TV algorithm in a much better way. In addition, it is noticeable that especially the Kaczmarz' solution represents all important information well, despite an irregular reconstruction. Overall, the effects are confirmed although the differences are less pronounced.


a) b) c) d)
e) f) g) h)

	M_{PSNR}	σ_{PSNR}	M_{MAE}	σ_{MAE}
a) FBP	17.06	0.89	0.11	0.011
b) Kaczmarz'	26.1	0.826	0.035	0.0035
c) TV	26.46	0.81	0.026	0.0034
d) BTV	27.61	0.76	0.021	0.0032
e) SA-TV	27.56	0.79	0.0236	0.0033
f) SA-BTV	28.1	0.78	0.022	0.003

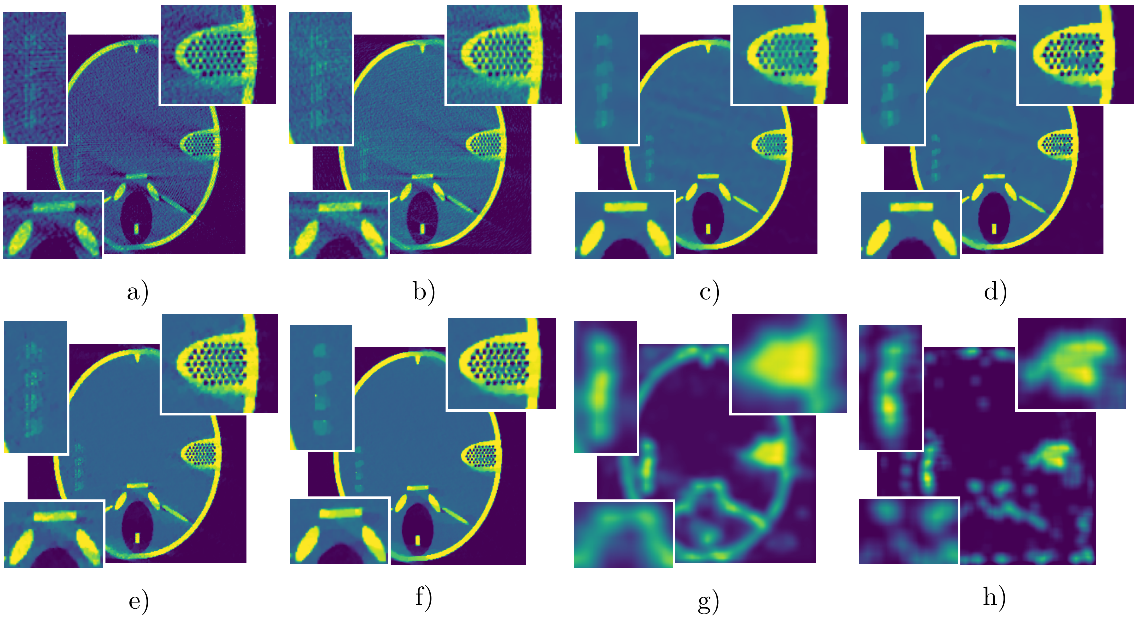
Figure 4.15.: The low-dose CT-reconstruction for a combination of 15 CT slices and 3 different random seeds was performed on a measurement at 60 angles with additive Gaussian noise $\eta_{i,j} \sim \mathcal{N}(0, \frac{\|\mathcal{R}u\|_\infty}{100})$. Subfigure a) - f) demonstrate the solution named in the table above. Subfigure g) denotes the weight of SA-TV, while subfigure h) the weight in SA-BTV.

4.3.3. Limited-angle computerized tomography

Last, a limited-angle measurement is reconstructed and shown in Fig. 4.16 and Fig. 4.17. In this case, the data is given by an incomplete sinogram of 150 different degrees. Here, for the first time, the global-weighted optimization algorithms give better results compared to the spatially-adapted ones. We can observe more included striae of the missing angles in those adjusted reconstructions, since they are evaluated as information in the residual. This phenomenon is not observed in normal TV-based reconstructions. In both motifs, an area is marked by an additional inset that falsely emphasizes these streaks.

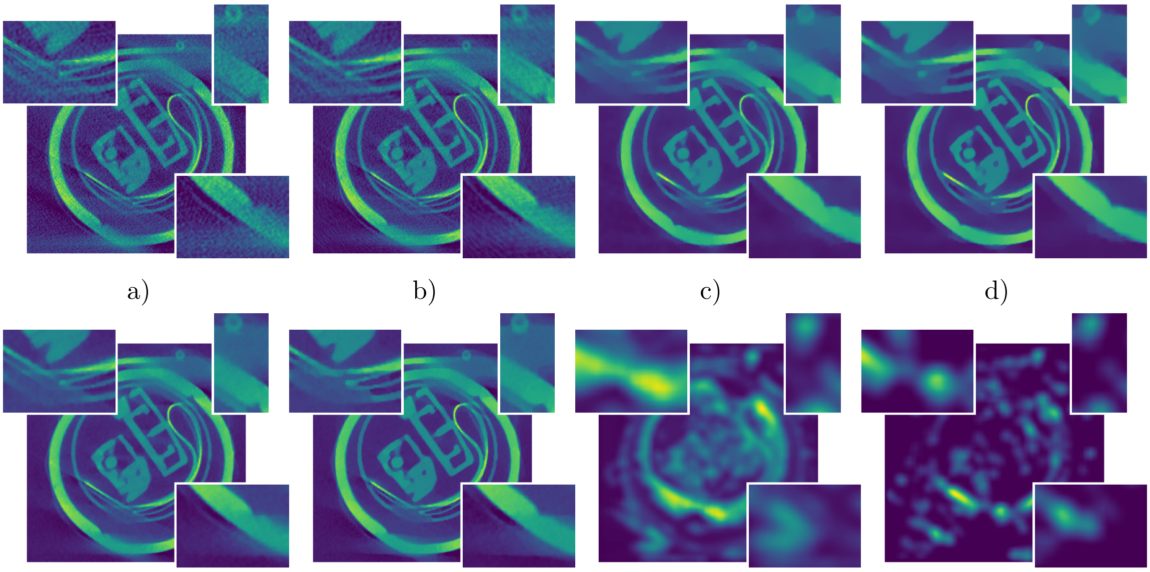
Next to the streaking problem, we already notice a suitable highlighting of the detailed areas. The contrast retention of the Bregman-based methods is still clearly visible and demonstrates the advantage of these regularization procedures.

Overall, we recognize once again the disadvantage of the variance estimation by the black background. Albeit still having a high reconstruction potential, a generalization still remains a challenging quest for the future.


a) b) c) d)
e) f) g) h)

	PSNR	MAE
a) FBP	16.04	0.12
b) Kaczmarz’	19.98	0.064
c) TV	23.68	0.023
d) BTV	24.14	0.02
e) SA-TV	23.59	0.025
f) SA-BTV	23.81	0.021

Figure 4.16.: The limited-angle CT-reconstruction was performed for 150 angles over 150 degrees and additive Gaussian noise $\eta_{i,j} \sim \mathcal{N}\left(0, \frac{\|\mathcal{R}u)\|_\infty}{100}\right)$. Subfigures a) - f) illustrate the solution with PSNR and MAE and are named in table above. Subfigure g) and h) denote the weight of the SA-TV and SA-BTV algorithms, respectively.



a) b) c) d)

e) f) g) h)

	M_{PSNR}	σ_{PSNR}	M_{MAE}	σ_{MAE}
a) FBP	18.85	0.75	0.083	0.007
b) Kaczmarz'	21.41	0.69	0.057	0.0041
c) TV	23.11	0.65	0.036	0.0036
d) BTV	23.43	0.62	0.032	0.0032
e) SA-TV	22.54	0.7	0.041	0.004
f) SA-BTV	22.69	0.69	0.039	0.0039

Figure 4.17.: The limited-angle CT-reconstruction was performed for a combination of 15 CT slices and 3 different random seeds and a measurement of 150 angles over 150 degrees and additive Gaussian noise $\eta_{i,j} \sim \mathcal{N}\left(0, \frac{\|\mathcal{R}u)\|_\infty}{100}\right)$. Subfigures a) - f) demonstrate the solutions of the algorithms, which are named in the table above. Subfigure g) denotes the weight of the SA-TV algorithm, while subfigure h) illustrate the weight in the SA-BTV algorithm.

CHAPTER 5

Conclusion and Outlook

The focus of this thesis was to find a suitable regularized solution method for imaging problems. The necessity of this procedure was discussed for two concrete situations: The inverse problems associated with the forward convolution and the Radon operator violate the third condition according to Hadamard. Therefore, a minimum-norm solution is not sufficient.

The studied optimization problem consists of two different parts, the fidelity and the regularization term. An enhanced weighting between both terms is described in this work. The decisive difference here is a local weight $\lambda \in L^\infty(\Omega)$ instead of the prior scalar version. This modification was motivated by an object-related sense, since an image is composed of many objects at different scales. The theory arising from this idea has been worked out in this thesis and confirmed by simulations: Fine structures can be reconstructed in more detail compared to the scalar-weighted TV-regularization. Beyond that a loss of contrast can also be observed, which increases for higher penalizations via total variation. To eliminate this phenomenon, an iterative method based on the Bregman divergence was developed. Since both enhancements are based on different problems, a comparison of these methods is only possible to a limited extent. Depending on the image characteristics, a better solution is obtained with one or the other approach. This circumstance leads to a novel possibility. The idea is to combine both concepts by considering variance estimation on basis of the Bregman solution. This method was named *Spatially Adapted - Bregman Total Variation* (SA-BTV) and studied in detail. Consequently, a new regularization technique, which adapts the spatially-dependent weight on every Bregman solution was developed as well. With this new technique, contrast preservation and highlighting domains are both possible and constitute a significant improvement over global methods.

On the basis of denoising, we first introduced and studied the different methods step by step. The respective improvement of the method was confirmed without further restrictions: Across different subjects and noise sizes, improvements of the contrast by Bregman-based methods and detailed region by the spatially-adapted methods were verified. An adaptation to the convolution operator is straightforwardly possible, as local structures in the data space are preserved. The simulations also showed a few limitations: The improvements depend on the noise level and the underlying kernel. Also, very fine structures can no longer be highlighted in the smoothed data residual.

A direct transfer of the spatially-adapted variant to the Radon transform is not possible. This is mainly due to the variance estimation, which requires local environments in the data. A sinogram, on the other hand, contains global information of the entire measurement. For this reason, the variance estimation was shifted from the data space to the image space by a new partial term. In the simulations, the following became clear: As long as the operator produces global errors - here, the black background-variance is a suitable indicator - the spatially-adapted reconstructions show the same improvements as for direct denoising. Therefore, those methods are well appropriate in the case of a full-measured or low-dose computerized-tomographic-reconstruction. For the limited-angle CT, the method is not suitable in its present version. Global techniques, such as TV- and Bregman-based TV-regularizations, achieve better results.

In future work, a more appropriate variance estimation should be developed. The version presented here is limited to normally-distributed noise. A more general version that addresses the perturbations of the underlying operator in a better way would be desirable. Furthermore, a local description of the reachable residual would also be possible as long as it could be approximated a priori. This would allow a generalization to many more inverse problems, thus increasing applications.

Overall, this work shows the great advantage of regularized optimization in the context of inverse and imaging problems. All the necessary foundations have been developed and implemented, allowing for a well-structured introduction to the subject. Furthermore, the advantages of spatially-adapted regularization techniques were confirmed and combined in a novel way with the contrast-preserving Bregman-based method and applied to real and simulated data.

Bibliography

[1] Jacques Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin. pp. 49–52.*, 1902.

[2] Thorsten Buzug. Einführung in die Computertomographie. Mathematisch-physikalische Grundlagen der Bildrekonstruktion, 2008.

[3] Virtanen and SciPy 1.0 Contributors. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature Methods*, 17(3):261–272, 2020.

[4] Stéfan van der Walt and the scikit-image contributors. scikit-image: image processing in Python. *PeerJ*, 2:e453, 2014.

[5] Guy Gilboa. Nonlinear eigenproblems in image processing and computer vision. In *Advances in Computer Vision and Pattern Recognition*, 2018.

[6] Ondřej Šindelář and Filip Sroubek. Image deblurring in smartphone devices using built-in inertial measurement sensors. *Journal of Electronic Imaging*, 22:1003, 2013.

[7] Neel Joshi, Richard Szeliski, and David Kriegman. PSF Estimation using Sharp Edge Prediction, 2008.

[8] Yurii Nesterov. Introductory lectures on convex optimization. *Applied Optimization volume 87*, 2004.

[9] N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in Optimization, 1(3):123-231*, 2014.

[10] Osman Güler. On the convergence of the proximal point algorithm for convex minimization. *SIAM J. Control Optim.*, 29(2):403–419, 1991.

[11] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems a first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[12] F. Natterer. The Mathematics of Computerized Tomography. *Society for Industrial and Applied Mathematics*, 2001.

[13] Jan-F. Pietschmann and F. Wübbeling. Inverse und schlecht gestellte Probleme (unpublished), 2012.

[14] Stephen J. Swensen, James R. Jett, Jeff A. Sloan, David E. Midthun, Thomas E. Hartman, Anne-Marie Sykes, Gregory L. Aughenbaugh, Frank E. Zink, Shauna L. Hillman, Gayle R. Noetzel, Randolph S. Marks, Amy C. Clayton, and Peter C. Pairolero. Screen-

ing for lung cancer with low-dose spiral computed tomography. *American journal of respiratory and critical care medicine*, 165:508–13, 2002.

[15] Mark E. Davison. The ill-conditioned nature of the limited angle tomography problem. *SIAM J. Appl. Math.*, 43(2):428–448, 1983.

[16] Rudin, L. I., Osher, S., Fatemi, E.. Nonlinear total variation based noise removal algorithms. Physica D. 60 (1–4): 259–268, 1992.

[17] L. Ambrosio, D. Pallara, and N. Fusco. Functions of bounded variation and free discontinuity problems, 2000.

[18] Kristian Bredies and Tuomo Valkonen. Inverse problems with second-order total generalized variation constraints. *Proceedings of SampTA 2011, 9th International Conference on Sampling Theory and Applications*, 2011.

[19] Maïtine Bergounioux. On poincare-wirtinger inequalities in spaces of functions of bounded variation, 2011.

[20] R Acar and C R Vogel. Analysis of bounded variation penalty methods for ill-posed problems. *Inverse Problems*, 10(6):1217–1229, 1994.

[21] Martin Adamčík. The information geometry of bregman divergences and some applications in multi-expert reasoning. *Entropy*, 16(12):6338–6381, 2014.

[22] Heinz Bauschke and Jonathan (Jon) Borwein. Joint and separate convexity of the bregman distance. *Studies in Computational Mathematics*, 8:23–36, 2001.

[23] Stanley Osher, Martin Burger, Donald Goldfarb, Jinjun Xu, and Wotao Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Model. Simul.*, 4(2):460–489, 2005.

[24] Kristian Bredies, Yiqiu Dong, and Michael Hintermüller. Spatially dependent regularization parameter selection in total generalized variation models for image restoration. *International Journal of Computer Mathematics - IJCM*, 90:1–15, 2012.

[25] Yiqiu Dong, Michael Hintermüller, and M. Monserrat Rincon-Camacho. Automated regularization parameter selection in multi-scale total variation models for image restoration. *Journal of Mathematical Imaging and Vision*, 40(1):82–104, 2011.

[26] Yiqiu Dong and Carola-Bibiane Schönlieb. Tomographic Reconstruction with Spatially Varying Parameter Selection, 2018.

[27] Zhicong Yu, Frédéric Noo, Frank Dennerlein, Adam Wunderlich, Günter Lauritsch, and Joachim Hornegger. Simulation tools for two-dimensional experiments in x-ray computed tomography using the forbild head phantom. *Physics in medicine and biology*, 57:N237–52, 2012.

[28] Kristian Bredies, Robert Nuster, and Raphael Watschinger. TGV-Regularized Inver-

sion of the Radon Transform for Photoacoustic Tomography. *Biomed. Opt. Express*, 11(2):994–1019, 2020.

[29] Stefan Kaczmarz. Angenäherte Auflösung von Systemen Linearer Gleichungen. *Bulletin International de l'Académie Polonaise des Sciences et des Lettres. Classe des Sciences Mathématiques et Naturelles. Série A, Sciences Mathématiques,*, 1937.

[30] Wim van Aarle, Willem Jan Palenstijn, Jan De Beenhouwer, Thomas Altantzis, Sara Bals, K. Joost Batenburg, and Jan Sijbers. The astra toolbox: A platform for advanced algorithm development in electron tomography. *Ultramicroscopy*, 157:35 – 47, 2015.

[31] Amir Beck. First-Order Methods in Optimization. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2017.

[32] I Do Imaging Wiki: https://wiki.idoimaging.com/.

APPENDIX A

Proximal Operator

We summarize some concrete proximal operators, which are important for this work. The determination of the proximal operator of the fidelity term is possible with the following theorem. The proof is based on reference [31].

Theorem A.1. Let $g : Y \to \mathbb{R} \cup \{\infty\}$ be a proper, lower semi-continuous, and convex function. Moreover, let $f(x) = g(Ax - z)$, where $b \in Y$ and $A : X \to Y$ is a linear operator with inverse A^{-1}. Then

$$\text{Prox}_f(x) = A^{-1}(\text{Prox}_g(Ax - z) + z)$$

for any $x \in X$.

Proof. By the definition of the proximal operator we obtain

$$\begin{aligned}
\text{Prox}_f(x) &= \operatorname*{arg\,min}_{u \in X} f(u) + \frac{1}{2} \|u - x\|^2 \\
&= \operatorname*{arg\,min}_{u \in X} g(Au - z) + \frac{1}{2} \|u - x\|^2 \\
&= \operatorname*{arg\,min}_{u \in X, y \in Y} \left\{ g(y) + \frac{1}{2} \|u - x\|^2 \mid y = Au - z \right\} \\
&= \operatorname*{arg\,min}_{u \in X} \left\{ \frac{1}{2} \|u - x\|^2 \mid Au = \tilde{y} + z \right\}.
\end{aligned}$$

Here $(\tilde{u}, \tilde{y})$ denotes the optimal solution, which exists and is unique by properties of g. Via strong duality transformations, there exists $y_0 \in Y$ with

$$\tilde{u} = \operatorname*{arg\,min}_{u \in X} \frac{1}{2} \|u - x\|^2 + \langle y_0, Au - \tilde{y} - z \rangle, \qquad A(\tilde{u}) = \tilde{y} + z.$$

By substitution of $A(\tilde{u}) = \tilde{y} + z$

$$\begin{aligned}
A(x - A^{-1} y_0) &= \tilde{y} + z \\
\Leftrightarrow \qquad y_0 &= Ax - \tilde{y} - z,
\end{aligned}$$

and therefore

$$\operatorname{Prox}_f(x) = \tilde{u} = x + A^{-1}(\tilde{y} - Ax + z).$$

Now substitute $u = \tilde{u}$ in the optimization problem above and the characterization for $\tilde{z}$ holds

$$\begin{aligned}\tilde{y} =& \arg\min_{y \in Y} g(y) + \frac{1}{2}\left\|x + A^{-1}(\tilde{y} - Ax + z) - x\right\|^2 \\ =& \arg\min_{y \in Y} g(y) + \frac{1}{2}\left\|\tilde{y} - Ax + z\right\|^2 \\ =& \operatorname{Prox}_g(Ax - z).\end{aligned}$$

Thus, by a further substituting of $\tilde{y}$, the statement holds true in total. □

Example A.2. The following four examples of proximal operators are of fundamental importance to this work.

1. Let $f : H \to \mathbb{R} \cup \{\infty\}$ with $f(x) = \frac{\lambda}{2}\|x\|^2$, the proximal operator follows with the characterization via the subdifferential

$$\begin{aligned}p = \operatorname{Prox}_f(x) &\Leftrightarrow x - p \in \partial f(p) = \{\lambda p\} \\ &\Leftrightarrow p = \frac{x}{1+\lambda}.\end{aligned}$$

2. Let $f : H \to \mathbb{R} \cup \{\infty\}$ with $f(x) = \frac{\lambda}{2}\|Au - z\|^2$ be the fidelity term. Considering theorem A.1 with $g(y) = \frac{\lambda}{2}\|y\|^2$,

$$\operatorname{Prox}_f(x) = A^{-1}\left(\frac{Ax - z}{1+\lambda} + z\right) = A^{-1}\left(\frac{Ax + \lambda z}{1+\lambda}\right).$$

3. For the Bregman fidelity term $f : H \to \mathbb{R} \cup \{\infty\}$ with $f(x) = \frac{\lambda}{2}\|Au - z\|^2 - \mathrm{TV}(c) - \langle p, u - c\rangle$ and $c \in \mathcal{H}$,

$$\operatorname{Prox}_f(x) = A^{-1}\left(\frac{A(x+p) + \lambda z}{1+\lambda}\right)$$

   holds true. This follows directly from the characterization of the subdifferential.

4. Let $f : H \to \mathbb{R} \cup \{\infty\}$ with $f(x) = \iota_{\overline{B(0,1)}}(x)$ the indicator function. Then

$$\begin{aligned}p = \operatorname{Prox}_f(x) &\Leftrightarrow x - p \in \partial f(p) \\ &\Leftrightarrow p = \begin{cases} x & \text{if } x \in \overline{B(0,1)} \\ \frac{x}{\|x\|} & \text{otherwise} \end{cases}.\end{aligned}$$

In example A.2.2. a real inverse is needed. For this reason, this proximal operator is not sufficient for the considerations of the Radon operator in section 4.3. However, we used this operator for the deconvolution problem in section 4.2 .

Theorem A.3. Example A.2.4. is given because it is the Fenchel-Legendre conjugate of $\|\cdot\|$, contained in the TV-term.

Proof. Let $g(x) = \|x\|$, then

$$g^*(u) = \sup_{x\in\mathcal{H}} \langle x, u\rangle - \|x\| .$$

For $\|u\| > 1$, we consider the sequence $x_k = u \cdot k$ and

$$f^*(u) \geq \limsup_{k\to\infty} (\|u\|^2 - \|u\|) \cdot k = \infty.$$

For $\|u\| \leq 1$, with the Cauchy-Schwarz inequality applies

$$f^*(u) \leq \sup_{x\in\mathcal{H}} (\|u\| \, \|x\| - \|x\|) = 0.$$

In summary

$$g^*(u) = \iota_{\overline{B(0,1)}}(u).$$

□

Theorem A.4. Let $f(x) = \frac{\lambda}{2}||Ax - z||^2$ with $\lambda > 0$. The subdifferential is defined with

$$\partial f(x) = \{\nabla f(x)\} = \{\lambda A^*(Ax - z)\}.$$

Proof.

$$\begin{aligned}
\lim_{\alpha\searrow 0} \frac{(f(x+\alpha\cdot y) - f(x))}{\alpha} &= \lim_{\alpha\searrow 0} \frac{\lambda}{2\alpha}(\|A(x+\alpha y) - z\|^2 - \|Ax - z\|^2) \\
&= \lim_{\alpha\searrow 0} \frac{\lambda}{2\alpha} \left(\langle A(x+\alpha y), A(x+\alpha y)\rangle - \langle A(x+\alpha y), z\rangle - \langle z, A(x+\alpha y)\rangle\right. \\
&\quad \left. - \langle Ax, Ax\rangle + \langle Ax, z\rangle + \langle z, Ax\rangle\right) \\
&= \lim_{\alpha\searrow 0} \frac{\lambda}{2\alpha} \left(\langle Ax, \alpha Ay\rangle + \langle \alpha Ay, Ax\rangle - \langle \alpha Ay, z\rangle - \langle z, \alpha Ay\rangle\right) \\
&= \lim_{\alpha\searrow 0} \frac{\lambda}{2\alpha} \left(\alpha\langle A^*Ax, y\rangle + \alpha\langle y, A^*Ax\rangle - \alpha\langle y, A^*z\rangle - \alpha\langle A^*z, y\rangle\right) \\
&= \frac{\lambda}{2} \left(\langle A^*Ax, y\rangle + \langle y, A^*Ax\rangle - \langle y, A^*z\rangle - \langle A^*z, y\rangle\right) \\
&= \frac{\lambda}{2} \left(\langle 2A^*Ax - 2A^*z, y\rangle\right) \\
&= \langle \lambda A^*(Ax - z), y\rangle
\end{aligned}$$

□

APPENDIX B

Source Code

The complete source code for the computed simulations and functions can be found at GitHub (`https://github.com/lucasplagwitz/recon`). We summarize a small list of the kernel elements in the following.

The submodule solver contains the Primal-Dual solvers, which can be seen in listing B.1.

```python
class PdHgm(object):
  """
  Primal-Dual solver.

  argmin G(x) + F(Kx)
  """

  ...

  def solve(self):

    ...

    while (self.tol < self.sens or self.k == 0) and (self.k < self.max_iter):

      self.x_prev = copy.copy(self.x)
      self.y_prev = copy.copy(self.y)

      # primal iteration
      self.x = self.G.prox(self.x - self.G.prox_param * (self.K.H * self.y))

      # dual iteration
      self.y = self.F_star.prox(self.y + self.F_star.prox_param *
          (self.K * (2 * self.x - self.x_prev)))

      if self.k % 200 == 0:
        self.update_sensivity()

    return self.x
```

Listing B.1: Primal-Dual algorithm

The sub-module terms contains various fidelity and regularization terms. Each term contains a __call__ and the prox-method. Exemplary, the listings B.2 and B.3 show the

L^2-fidelity and the indicator term. The latter is the Fenchel-Legendre conjugate to the 1-norm.

```
class DatanormL2(BaseDataterm):
  """Datanorm-L2

  This class is the basic form of the L2-Datanorm in terms of A: X -> Y.

  Function u(x):
  lambda/2 * ||Ax - f||_2^2

  Special Form with A=identity:
  lambda/2 * ||x - f||_2^2
  """

  ...



  def __call__(self, x):
    return self.lam/2*np.sum((self.operator*x-self.data)**2)

  def prox(self, x):
    """
      Proximal Operator

      prox(x) = A_inv * (( A*x + tau * lambda * f) / (1 + tau * lambda *
          diag_sampling))
    """
    u = self.inv_operator*(
      (self.operator*x + self.prox_param * self.lam * self.data) /
      (1 + self.prox_param * self.lam * self.diag_sampling)
    )

    return u
```

Listing B.2: L^2-Datanorm

```
class IndicatorL2(BaseRegTerm):
  """IndicatorL2

  This function is the indicator function of the convex set P ("union of pointwise L2 balls"):
  P := {p in Y: |p| <= upper_bound}.  |p| = max_ij|p_ij| = max_ij sqrt(p1_ij**2 + p2_ij**2)
  """

  ...

  def prox(self, f):
    """
      Proximal Operator of indicator Function iP -> Projection.

      prox(f) = f / max(1, ||f||)

      f_row stores row-wise: In case of f^(*) as Projection and K=Grad
      -> dual f_row[i,:] stores i-th directional derivative.
    """

    assert self._input_check(f)
    if self.times is None:
      norm_f = self._infty_abs(f)
      norm_f = np.array([norm_f]*self.derivate_dim).ravel() / self.upper_bound
      norm_f[norm_f < 1] = 1

    ...

    return f/norm_f
```

Listing B.3: L^2-Indicator

APPENDIX C

Imaging Results

The algorithms were illustrated in Chapter 2 using cross-sections. In the following, these cross sections are shown in their two-dimensional version.

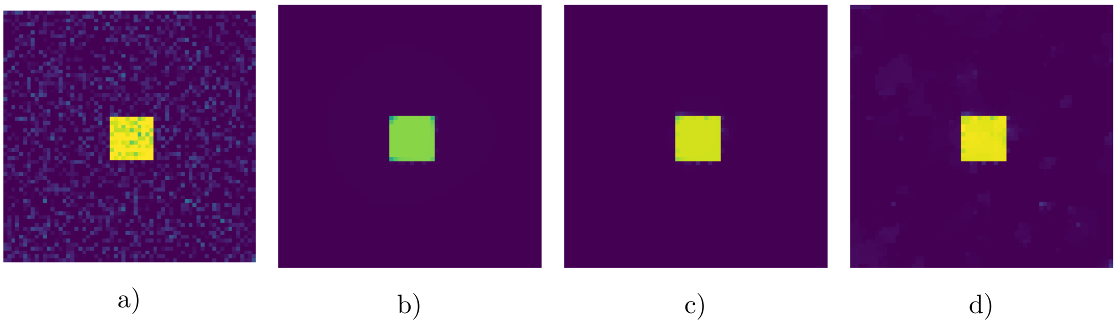


Figure C.1.: Imaging result for example 1.1 with width $2\delta = 10$. While subfigure a) demonstrates the noisy image, subfigures b), c), and d) show the solutions for the underlying weights $\lambda = 2$, $\lambda = 5.5$, and $\lambda = 13$.

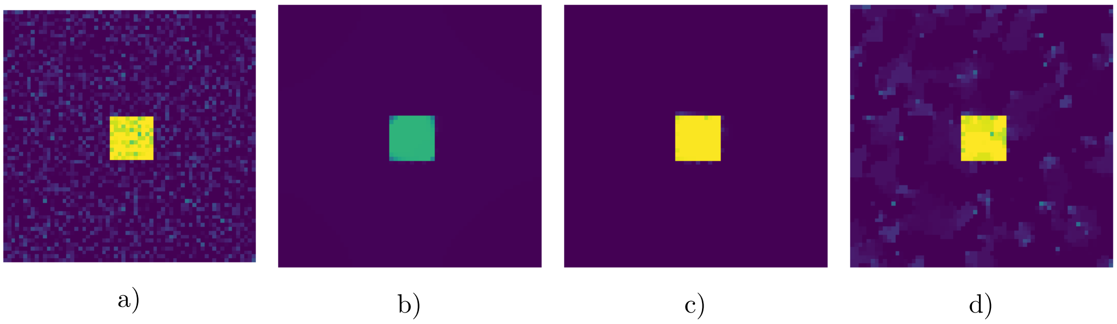
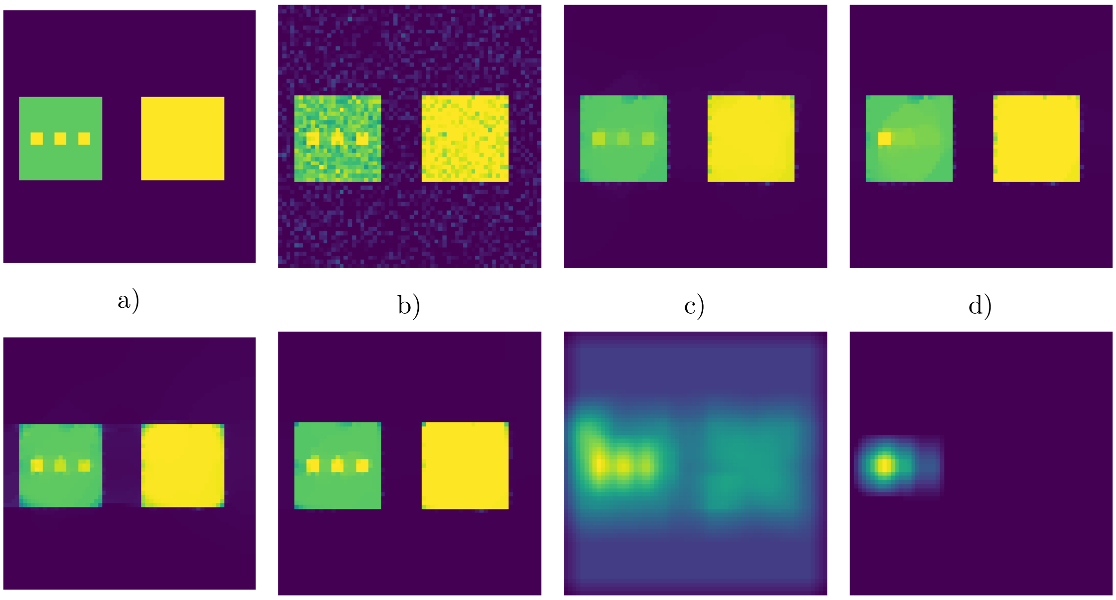


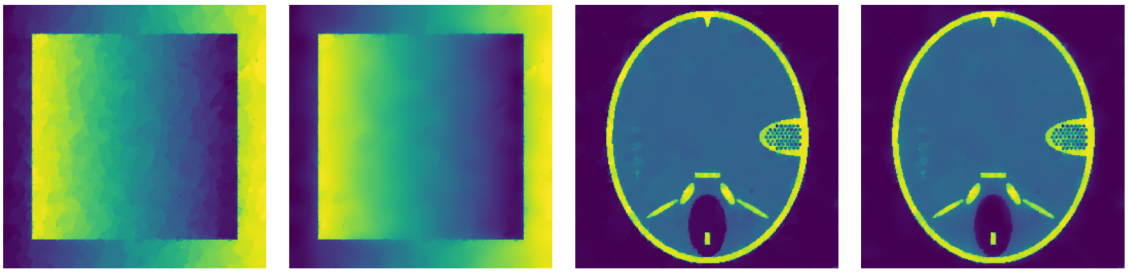
Figure C.2.: Imaging result for Bregman-based iterations. We see the noisy image in subfigure a), while subfigures b), c), and d) illustrate the solutions of iteration 1, 3, and 13, respectively. The discrepancy criterion would have stopped after 3 iterations.



a) b) c) d)

e) f) g) h)

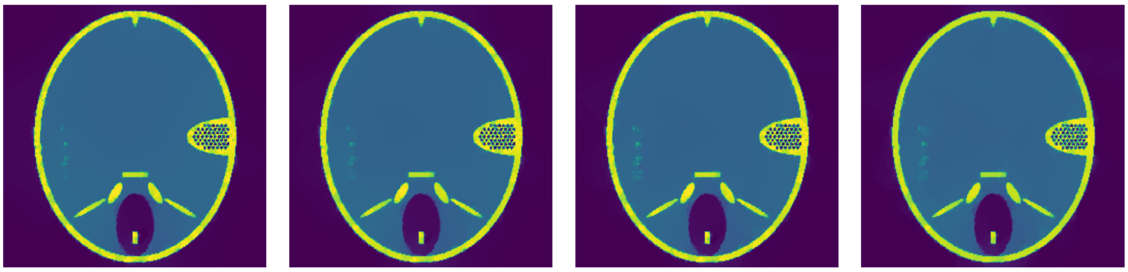
Figure C.3.: Results of the cross sections from chapter 3. The noisy version of the original (subfigure a)) is shown in subfigure b). In the placed order, Subfigure c), d), e), f) show the result of the TV, BTV, SA-TV, SA-BTV algorithms, respectively. Subfigure g) shows the determined weight λ from the SA-TV algorithm while subfigure h) shows λ from the SA-BTV algorithm.



a) b) c) d)

	PSNR	MAE
a) TV	32.13	0.016
b) TGV	32.84	0.013
c) TV	25.2	0.024
e) TGV	24.78	0.025

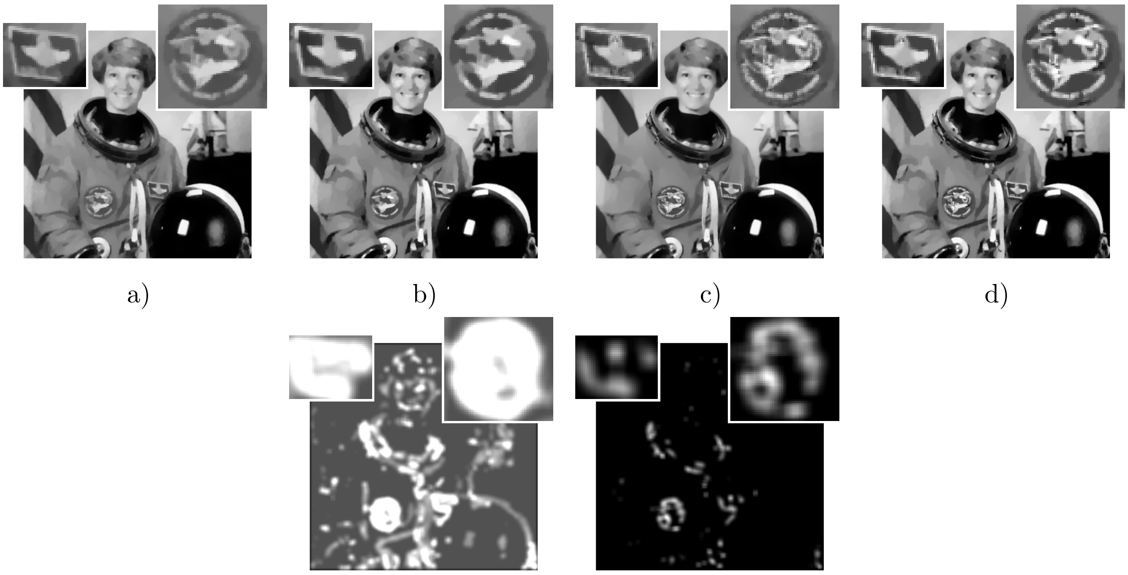
Figure C.4.: Differences between total generalized variation and first-order total variation are shown on the basis of two different image structures.


a) b) c) d)

	PSNR	MAE	Iterations
a) $\lambda = 0.1$	26.52	0.02	4
b) $\lambda = 0.5$	26.31	0.02	3
c) $\lambda = 1$	26.76	0.02	3
e) $\lambda = 2$	25.77	0.021	2

Figure C.5.: Solutions of the SA-TV algorithm for different initial weightings are shown.

Lastly, we illustrate the results for two further motifs. In Figure C.6 a real world motif is shown. The zoomed-boxes demonstrate a higher level of detail in the spatially-adapted solutions. Figure C.7 demonstrates the results of the different algorithms for a CT-body-scan. The TV-functional is not well suited here, since real edges hardly ever occur in the human body. Due to the large surfaces and rather small details, spatial adaptation hardly shows any advantages.

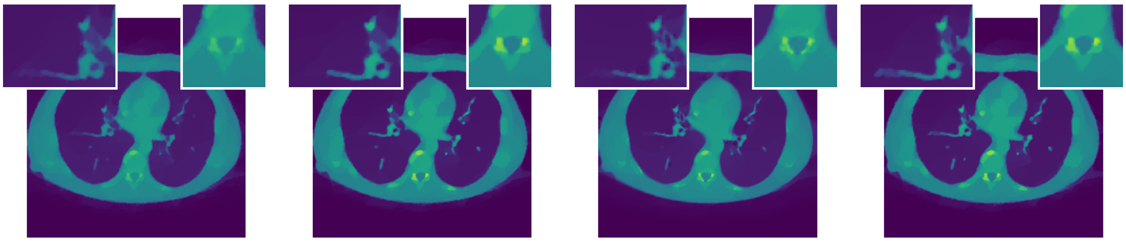


a) b) c) d)

e) f)

	PSNR	MAE
a) TV	26.62	0.029
b) BTV	27.01	0.027
c) SA-TV	27.11	0.029
d) SA-BTV	27.59	0.026

Figure C.6.: Overview over all regularization techniques on a denoised ($\eta_{i,j} \sim \mathcal{N}\left(0, \frac{1}{25}\right)$) real-world image. Subfigure e) illustrates the weight λ calculated in the SA-TV algorithm, subfigure f) shows λ in the SA-BTV algorithm. Image taken from Ref. [4].



a) b) c) d)

	M_{PSNR}	σ_{PSNR}	M_{MAE}	σ_{MAE}
Degraded	20.4	0	0.079	0
a) TV	30.45	1.049	0.017	0.0021
b) BTV	30.76	0.984	0.0156	0.0019
c) SA-TV	30.56	0.879	0.017	0.002
d) SA-BTV	30.87	0.987	0.0156	0.0019

Figure C.7.: An overview over all regularization techniques applied to a combination of 30 abdomen CT-scans is shown. Denoising from an image with noise $\eta_{i,j} \sim \mathcal{N}(0, \frac{1}{100})$. Data taken from [32].

Declaration of Academic Integrity

I hereby confirm that this thesis on *Spatially-Adapted Regularization for Inverse Problems* is solely my own work and that I have used no sources or aids other than the ones stated. All passages in my thesis for which other sources, including electronic media, have been used, be it direct quotes or content references, have been acknowledged as such and the sources cited.

Münster, April 1, 2021

(Lucas Plagwitz)