UNIVERSITÄT ZU LÜBECK
INSTITUTE OF MATHEMATICS AND
IMAGE COMPUTING

Image reconstruction with non-parametric noise models

Bildrekonstruktion mit nichtparametrischen Rauschmodellen

Bachelorarbeit

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

Vorgelegt von
Felix Kastner

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

Mit Unterstützung von
Dr. Yury Korolev
Institute of Mathematics and Image Computing

Lübeck, den 14.11.2017

Eidesstattliche Erklärung

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

Lübeck,
14.11.2017

Felix Kastner

Abstract

In many practical situations, such as medical applications, images of interest cannot be measured directly. Therefore, indirect measurement methods are employed, which measure related quantities. Then special reconstruction methods are used to infer about the quantities of interest from the data. These data also usually contain noise, thus the task of image reconstruction is most often complemented with the task of denoising. While current methods try to model the statistical distribution of the noise, in many practical situations the true noise distribution may be unknown or exact modelling may not be computationally feasible. Therefore there is a need for approximate methods that do not assume any particular noise distribution. Such models may be referred to as non-parametric noise models.

The goal of this thesis is a numerical study of a recently proposed approach that models errors in the data using intervals in a suitable partial order. From the statistical point of view, this corresponds to using confidence intervals for the unknown exact measurements rather than point estimates. In this thesis the performance of this method is studied in problems with unbounded noise and compared to existing approaches based on, e.g., Gaussian noise assumptions. To this end first a short introduction into variational methods for the solution of inverse problems as well as the basics of Banach lattices will be given.

Kurzfassung

In vielen praktischen Situationen, wie zum Beispiel medizinischen Anwendungen, können Bilder nicht direkt aufgenommen werden. Deswegen werden indirekte Bildgebungsverfahren verwendet, die verwandte Parameter messen. Dann werden spezielle Rekonstruktionsverfahren verwendet, um Informationen über die gesuchte Größe aus den Daten herzuleiten. Da diese Daten üblicherweise verrauscht sind, wird die Bildrekonstruktion meist durch Entrauschen ergänzt. Während aktuelle Methoden versuchen, die Verteilung des Rauschens exakt zu modellieren, ist diese in vielen Anwendungen oft unbekannt oder die numerische Umsetzung zu aufwändig. Deshalb werden approximative Methoden benötigt, die keine spezielle Verteilung des Rauschens annehmen. Diese Modelle können als nichtparametrische Rauschmodelle bezeichnet werden.

Das Ziel dieser Arbeit ist eine numerische Studie eines kürzlich vorgestellten Ansatzes, welcher die Fehler in den Daten mit Hilfe von Intervallen in einer geeigneten Halbordnung modelliert. Aus einem statistischen Blickwinkel entspricht dies der Verwendung von Konfidenzintervallen für die unbekannten exakten Messwerte anstelle von Punktschätzern. In dieser Arbeit wird die Leistung dieser Methode anhand von Problemen mit unbeschränktem Rauschen untersucht und mit existierenden Ansätzen, die zum Beispiel auf der Gauss-Annahme beruhen, verglichen. Hierfür wird zunächst eine kurze Einführung in Variationsmethoden zur Lösung inverser Probleme und die Grundlagen sogenannter Banach-Verbände vorgestellt.

Contents

Introduction

Image reconstruction is the task of recovering the original image from a measurement of a related quantity. Typically this measurement is also corrupted, e.g., by noise or blur. There are many important applications where only noisy or blurry images are available, e.g., in astronomy where the target is far away and a lot of signals interfere with the observation.

Image reconstruction fits into the broader class of problems where we want to get information about a quantity of interest that cannot be measured directly. We then need to reconstruct the information from observable data that is related to the quantity of interest. We call this *inverse problems* because we want to reconstruct a "cause" from an observed "effect". The corresponding *forward problem* would be to calculate the effect from a given cause.

Inverse problems are often ill-posed and therefore need regularisation. To construct a regularisation algorithm we need to have some knowledge of the existing noise, because as was first shown in [5] there cannot exist regularisation algorithms for ill-posed problems that are independent of the noise level in the data. Current approaches either explicitly or implicitly assume a specific noise distribution. For example in variational regularisation the L^2 fidelity term is tailored to Gaussian noise, the L^1 fidelity to impulse noise and the Kullback-Leibler divergence to Poisson noise. These are all convex but for some noise distributions, like the Rician distribution, the exact fidelity term is more complex and non-convex. In this case or if the true noise model is unknown often the standard assumption of a Gaussian distribution is made [52, 53].

In this thesis we will study a new fidelity term that makes only minimal assumptions about the existing noise and can thus be used in situations in which the noise distribution is unknown or too complex. Our goal is to study numerically whether this new approach is better suited to reconstruct images corrupted by unknown or complex noise than models using a false but simple assumption on the noise distribution.

The thesis consists of three chapters. In Chapter 1 we will first specify the notion of inverse problems. Then we will see how variational methods can be used to regularise ill-posed problems. We will also discuss the choice of regularisation and fidelity terms in special cases.

In Chapter 2 we will give a brief introduction into Banach lattices and present the partial order based approach. We will discuss the choice of order intervals in problems with unbounded noise and study the role of the so-called quantile parameter, which determines the width of the intervals, as a regularisation parameter that balances the influence of the data and the regulariser on the reconstruction.

In the last Chapter 3 we will study the new model in conjunction with the total variation regulariser and then look at the numerical performance of the method in application to image denoising. We compare the new approach to the ROF model and the TV-L^1 model.

Chapter 1: Inverse problems

In [36] J. B. Keller defined *inverse problems* as follows:

> "We call two problems **inverses** of one another if the formulation of each involves all or part of the solution of the other. Often, for historical reasons, one of the two problems has been studied extensively for some time, while the other is newer and not so well understood. In such cases, the former is called the **direct problem**, while the latter is called the **inverse problem**."

An example of an inverse problem is inferring the shape of an object by looking at the shadow it casts on a wall. The corresponding forward problem is predicting the shape of the shadow if the light is known to come from a particular direction.

Forward problems are typically well-posed in the sense of Hadamard [30, 31], i.e. they fulfil the following three conditions:

1. a solution exists (existence),
2. the solution is unique (uniqueness),
3. and the solution depends continuously on the data (stability),

whereas the corresponding inverse problem is often ill-posed, meaning that one or more of these conditions are not satisfied. In this thesis we consider inverse problems for which the first two conditions are satisfied, but the third one may be violated.

One approach to overcome the instability is to approximate the problem by a parametrised family of well-posed problems, the solutions of which converge to the solution of the ill-posed problem as the parameter approaches some limit value. This is known as *regularisation* [21].

The truncated singular value decomposition [32] and regularisation by projection [21] are examples of regularisation methods for ill-posed inverse problems. However, perhaps the most widely used approach is variational regularisation, which is the framework we use in this thesis. In this chapter we first present the variational approach to inverse problems and regularisation, introducing the concepts of fidelity and regularisation functionals. In Section 1.2 we will discuss some commonly used regularisers and their properties and in Section 1.3 we will present a statistical motivation for the choice of fidelity terms.

1.1 Variational Methods

The main idea of variational regularisation is to approximate the solution of an ill-posed inverse problem by a solution of a suitable optimisation problem. In this section we will recall three different but connected approaches, namely Tikhonov regularisation, the residual method and the method of quasi-solutions.

Inverse problems are typically formulated as operator equations

$$Au = f, \quad u \in \mathfrak{U},\ f \in \mathfrak{F}, \tag{1.1}$$

where $\mathfrak{U}$ and $\mathfrak{F}$ are vector spaces, typically Hilbert or Banach spaces. $A\colon \mathfrak{U} \to \mathfrak{F}$ is called the *forward operator* and models the relationship between the quantity of interest u and the measurement f. In practice, the data f as well as the operator A will often be noisy. The noise in the measurements typically comes from the finite precision of measurement devices, while errors in the operator are due to the discrepancy between the real-world phenomena and the mathematical models used or to the uncertainty in some model-related parameters. We will denote by $A^\star$ and $f^\star$ the unknown exact operator and measurement respectively. In this work $\mathfrak{U}$ and $\mathfrak{F}$ are assumed to be Banach spaces and we will only consider operators $A^\star$ that are linear, bounded and injective. We assume that a solution exists and the injectiveness of $A^\star$ guarantees that it is unique. We will call this solution $u^\star$.

A popular approach to solving inverse problems is to use variational methods to find an approximation to $u^\star$ as a minimiser of a suitable functional. This functional typically consists of a *fidelity term* $\mathcal{F}\colon \mathfrak{F} \times \mathfrak{F} \to \mathbb{R}$ and a *regularisation functional* $\mathcal{R}\colon \mathfrak{U} \to \mathbb{R}$:

$$\min_{u \in \mathfrak{U}} \mathcal{F}(Au, f) + \alpha \mathcal{R}(u), \tag{1.2}$$

where α is the *regularisation parameter*, this approach is referred to as *Tikhonov regularisation* [33]. Here $\mathcal{F}$ measures how good u explains the data and $\mathcal{R}$ measures how good u satisfies a prior assumption like smoothness or sparseness.

There exist different fidelity terms to model different notions of closeness between the predicted and measured data. The choice of the fidelity term is typically motivated by the statistics of the underlying noise in the data. Some typical fidelities will be discussed in Section 1.3. There are also different regularisation functionals that promote different properties of u. Ideally $\mathcal{R}(u)$ should be small for $u = u^\star$ and big for solutions u with unfavourable structure. We will discuss some regularisers in more detail in Section 1.2.

The regularisation parameter α balances the influence of the fidelity term against the regulariser. The bigger we choose α the more we force the solution to satisfy the prior assumptions and the smaller we choose α the more we favour closeness of Au to the data f.

In practice it is desirable to choose this parameter automatically, depending not only on the noise level in the data, but also on the data themselves. Suppose we know that the error in the data can be bounded as $\|f^\star - f\| \leq \delta$. Then it can be shown that the minimisers of (1.2) converge to the exact solution $u^\star$ as $\delta \to 0$ and simultaneously the regularisation parameter is chosen as a function of δ such that $\alpha(\delta) \xrightarrow{\delta \to 0} 0$ and $\frac{\delta^2}{\alpha(\delta)} \xrightarrow{\delta \to 0} 0$ hold [21]. This is known as an a-priori parameter choice rule. Optimal a-priori parameter choice rules require knowledge that is often not available. Thus a-posteriori parameter choice rules like the discrepancy principle [21] are widely used.

An alternative approach is known as the *residual method* [33]. It consists in minimising the regulariser subject to fidelity constraints:

$$\min_{u \in \mathfrak{U}} \mathcal{R}(u) \quad \text{s.t. } \mathcal{F}(Au, f) \leq \delta, \tag{1.3}$$

for some estimate of the noise level $\delta > 0$. Among all elements $u \in \mathfrak{U}$ that satisfy this fidelity constraint we choose the one that best fits our prior knowledge. If we choose

the parameter $\alpha(\delta)$ in (1.2) according to the discrepancy principle [21] then the two approaches are equivalent.

Yet another approach consists in minimising the fidelity subject to a regularity constraint in terms of $\mathcal{R}$. It is known as the *method of quasi-solutions* [33] and results in the following optimisation problem:

$$\min_{u \in \mathfrak{U}} \mathcal{F}(Au, f) \quad \text{s.t. } \mathcal{R}(u) \leq \tau. \tag{1.4}$$

This method is used less often but can be helpful if a bound for the regularity of the solution is known a priori [19].

A good choice of the fidelity term and the regulariser is crucial for the performance of the method and depends on the task one would like to solve. In the next two sections we will present some common choices for these functionals.

1.2 Regularisers

If the solution space $\mathfrak{U}$ is a reflexive Banach space then a possible choice of the regulariser is the norm in $\mathfrak{U}$ [33]. Restricting our solution, e.g., to the Sobolev space $\mathfrak{U} = W^{1,2}(\Omega)$, $\Omega \subset \mathbb{R}^2$ we could use the norm $\|u\|_{W^{1,2}(\Omega)} := \|u\|_{L^2(\Omega)} + \|\nabla u\|_{L^2(\Omega)}$ and obtain smooth solutions with weak first derivatives in $L^2(\Omega)$.

In natural images and imaging applications, however, sharp edges are common and thus other regularisers are used that allow to preserve jumps. Without claiming completeness, we list some regularisers that are widely used in image reconstruction. In this section $\Omega \subset \mathbb{R}^2$ is assumed to be a two-dimensional Lipschitz domain.

Total Variation

The *total variation* (TV) is one of the most widely used regularisers. It was first introduced into the field of image reconstruction in [47]. For a two-dimensional Lipschitz domain $\Omega \subset \mathbb{R}^2$ and a function $u \in L^1(\Omega)$ the isotropic total variation is defined as

$$\mathrm{TV}(u) := \sup \left\{ \int_\Omega u(x) \operatorname{div} \varphi(x) dx \colon \varphi \in \mathcal{C}_c^1\left(\Omega, \mathbb{R}^2\right), \|\varphi\|_\infty \leq 1 \right\}, \tag{1.5}$$

where $\operatorname{div} \varphi = \frac{\partial \varphi_1}{\partial x_1} + \frac{\partial \varphi_2}{\partial x_2}$ and $\|\varphi\|_\infty = \sup_{x \in \Omega} \|\varphi(x)\|_2$. For $u \in \mathcal{C}^1$ this is equivalent to

$$\mathrm{TV}(u) = \int_\Omega \|\nabla u(x)\|_2 dx. \tag{1.6}$$

All functions $u \in L^1(\Omega)$ with $\mathrm{TV}(u) < \infty$ form the space of *functions of bounded variation*

$$\mathrm{BV}(\Omega) := \left\{ u \in L^1(\Omega) \colon \mathrm{TV}(u) < \infty \right\}, \tag{1.7}$$

which becomes a Banach space with the norm

$$\|u\|_{\mathrm{BV}(\Omega)} := \|u\|_{L^1(\Omega)} + \mathrm{TV}(u). \tag{1.8}$$

It was shown in [3, Corollary 3.49 with Prop. 3.21] that the space $\mathrm{BV}(\Omega)$ can be continuously embedded in $L^2(\Omega)$ and compactly embedded in $L^p(\Omega)$ for $1 \leq p < 2$. See [23] for more background on functions of bounded variation and [2] for an analysis of TV-regularised inverse problems.

The reason for the popularity of TV as a regulariser is its ability to preserve edges. In contrast to the Sobolev norm mentioned above it favours piecewise constant and thus discontinuous solutions. For smooth regions in the exact solution, however, this leads to a "stair-like" reconstruction. This effect is called *staircasing* and considered the main weakness of TV in terms of reconstruction quality [14, 20, 42, 44, 46]. To reduce this effect extensions of TV to include higher order derivatives have been proposed [11, 14, 15, 48]. Perhaps the most recognised model is the total generalised variation introduced in [11].

Total Generalised Variation

The second order *total generalised variation* [11] (TGV) for $u \in L^1(\Omega)$ is defined as:

$$\mathrm{TGV}^2_{\alpha,\beta}(u) \coloneqq \sup \left\{ \int_\Omega u(x) \operatorname{div}^2 \varphi(x) dx \colon \varphi \in \mathcal{C}^2_c \left(\Omega, S^{2\times 2}\right), \|\varphi\|_\infty \leq \alpha, \|\operatorname{div} \varphi\|_\infty \leq \beta \right\}, \tag{1.9}$$

where $\alpha, \beta > 0$, $S^{2\times 2}$ is the space of symmetric 2×2 matrices, the divergence for these matrices is defined as

$$\operatorname{div} \varphi = \begin{pmatrix} \frac{\partial \varphi_{1,1}}{\partial x_1} + \frac{\partial \varphi_{1,2}}{\partial x_2} \\ \frac{\partial \varphi_{2,1}}{\partial x_1} + \frac{\partial \varphi_{2,2}}{\partial x_2} \end{pmatrix}$$

and

$$\operatorname{div}^2 \varphi = \frac{\partial^2 \varphi_{1,1}}{\partial x_1^2} + \frac{\partial^2 \varphi_{2,2}}{\partial x_2^2} + 2 \frac{\partial^2 \varphi_{1,2}}{\partial x_1 \partial x_2}.$$

It was shown that $\mathrm{TGV}^2(u) = 0$ if and only if u is a polynomial of degree less than two and thus TGV^2 promotes piecewise affine solutions, effectively eliminating the staircasing effect [11]. The space of *functions of bounded second order generalised variation* (BGV)

$$\mathrm{BGV}^2(\Omega) \coloneqq \left\{ u \in L^1(\Omega) \colon \mathrm{TGV}^2_{\alpha,\beta}(u) < \infty \right\} \tag{1.10}$$

was shown to be independent of the weights α and β [11]. In [9] the authors proved that the space $\mathrm{BGV}^2(\Omega)$ is isomorphic to $\mathrm{BV}(\Omega)$.

TV-L^p infimal convolution

A disadvantage of TGV is that it uses higher order derivatives. To overcome this issue, in [13] the authors proposed a family of infimal convolution type regularisers that use only first order derivatives. They are defined as follows:

$$\mathrm{TVL}^p_{\alpha,\beta}(u) \coloneqq \min_{w \in L^p(\Omega)} \alpha \|Du - w\|_{\mathcal{M}} + \beta \|w\|_{L^p(\Omega)}, \tag{1.11}$$

where Du is the distributional derivative of u, which is a finite Radon measure for $\mathrm{TV}(u) < \infty$, $\|\cdot\|_{\mathcal{M}}$ is the Radon norm of a measure and $\mathrm{TV}(u) = \|Du\|_{\mathcal{M}}$. The similarities and differences to the TV and TGV functionals are best seen in the dual formulation [13]:

$$\mathrm{TVL}^p_{\alpha,\beta}(u) = \sup\left\{\int_\Omega u(x)\,\mathrm{div}\,\varphi(x)dx \colon \varphi \in \mathcal{C}^1_c\left(\Omega, \mathbb{R}^2\right), \|\varphi\|_\infty \leq \alpha, \|\varphi\|_{L^q(\Omega)} \leq \beta\right\}, \tag{1.12}$$

where $\alpha, \beta > 0$, $1 < p \leq \infty$ and q is the Hölder conjugate of p satisfying $\frac{1}{p} + \frac{1}{q} = 1$. For sufficiently high values of p and suitable choice of α and β this model was also shown to reduce staircasing significantly. Additionally it was shown that $\mathrm{TVL}^p_{\alpha,\beta}(u) < \infty$ if and only if $\mathrm{TV}(u) < \infty$ and therefore the domain of this functional is exactly the space of functions of bounded variation $\mathrm{BV}(\Omega)$.

1.3 Relationship between fidelity terms and noise statistics

Following [17, 34] we will discuss the relationship between fidelity functionals and notions of closeness of the measured and predicted data motivated by the statistics of the noise. Taking for a moment the Bayesian point of view, we will discuss *maximum a-posteriori* (MAP) estimates and show, for illustrative purposes, that a Gaussian noise assumption leads to the widely used L^2 fidelity term in (1.2).

Consider the measurement and the unobservable quantity of interest as random variables F and U respectively. We are interested in the conditional probability

$$P\left(U = u \,|\, F = f\right).$$

In this setting we regard the measurements f as discrete images drawn from a probability distribution that depends on the unobservable quantity of interest u. This dependency is given by the *data model* $P\left(F = f \,|\, U = u\right)$, which is the conditional probability that the measurement F would be f if u was indeed the state of the quantity U. The regularisers from Section 1.2 can then be seen as a *prior model* $P\left(U = u\right)$, assigning a probability to all possible values of the unobservable quantity U.

In order to maximise the a-posteriori probability we use Bayes' theorem and get

$$\tilde{u} = \arg\max_{u \in \mathfrak{U}} P\left(u \,|\, f\right) = \arg\max_{u \in \mathfrak{U}} \frac{P\left(f \,|\, u\right) P\left(u\right)}{P\left(f\right)} = \arg\max_{u \in \mathfrak{U}} P\left(f \,|\, u\right) P\left(u\right), \tag{1.13}$$

where $P\left(f\right)$ is a normalization term independent of u and therefore can be dropped from the maximisation problem. Let $f \colon \Omega \to \mathbb{R}$ be the observation where $\Omega \subset \mathbb{N}$ is a discrete image domain and u is the unknown value of the quantity of interest. If we assume independent additive noise with a zero-mean Gaussian distribution with standard deviation σ for every pixel $x \in \Omega$ then the joint probability is

$$P\left(f \,|\, u\right) = \prod_{x \in \Omega} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(f(x) - Au(x))^2}{2\sigma^2}}, \tag{1.14}$$

where the constant factors independent of u can be dropped as they do not influence the maximisation problem. As a prior model we want to use log-concave priors as they are widely used and studied [12]:

$$P(u) = e^{-\beta \mathcal{R}(u)}, \tag{1.15}$$

where $\mathcal{R}(u)$ is a convex functional. Here a higher probability corresponds to a small value of the regulariser $\mathcal{R}(u)$. Substituting into (1.13) we get

$$\begin{aligned} \tilde{u} &= \arg\max_{u \in \mathfrak{U}} \prod_{x \in \Omega} e^{\frac{-(f(x)-Au(x))^2}{2\sigma^2}} \cdot e^{-\beta \mathcal{R}(u)} \\ &= \arg\min_{u \in \mathfrak{U}} \|f - Au\|_2^2 + \alpha \mathcal{R}(u), \end{aligned}$$

where we used that the logarithm is a strictly increasing function and we combined the constants σ and β into one parameter $\alpha = 2\sigma^2\beta$. This corresponds to (1.2) with $\mathcal{F}(Au, f) = \|Au - f\|_2^2$, showing that the Gaussian assumption leads to an L^2 fidelity term in (1.2).

This connection between the noise assumption and the fidelity term is very important as it allows us to construct fidelities tailored to specific noise models. In [39] it was shown that for signal dependent Poisson noise the appropriate fidelity term is as follows:

$$\mathcal{F}(Au, f) = \int_\Omega Au(x) - f(x) \log Au(x) dx, \tag{1.16}$$

which is related to the Kullback-Leibler divergence. For salt & pepper or general impulse noise the L^1 norm $\mathcal{F}(Au, f) = \|Au - f\|_1$ was shown to be a suitable fidelity term [43].

However, for some distributions the correct fidelity terms may be quite complicated. For example, correct modelling of the Rician noise distribution results in a non-convex fidelity term [22], which required the development of different convex relaxations [18, 22]. The situation can become even more complicated if the distribution of the noise is not exactly known. This motivates the development of simple approximations of the fidelity term, which do not assume any specific noise distribution. For this reason, [25] suggests the use of confidence intervals for the unknown exact measurement as a fidelity term. From the functional-analytic point of view this leads to the need to consider inverse problems in functional spaces with partial order – Banach lattices. We will review this approach in the next chapter.

Chapter 2: Partial order based methods for inverse problems

Partial order can be introduced in many functional spaces common in image processing, such as L^p spaces, turning them into Banach lattices. The richer structure of Banach lattices as compared to Banach spaces allows for a more detailed description of errors in the data and the forward operator of an inverse problem using order intervals. In [37, 38] a partial order based feasible set was proposed and a lattice analogue of the residual method has been formulated.

In this chapter we will first review some definitions from the theory of Banach lattices (Section 2.1) and then discuss inverse problems in Banach lattices (Section 2.2). The relationship between this approach and some existing interval based methods will be discussed in Section 2.3. The choice of the order intervals describing errors in the data and the forward operator, crucial for practical applications, will be discussed in Section 2.4, where we will also make a connection to confidence intervals announced in Chapter 1.

2.1 Banach lattices

Banach lattices are vector spaces that are equipped with a partial order as well as a norm, which in a certain sense agree with each other. In this section we give a brief introduction to Banach lattices following [1].

A vector space X equipped with a partial order $\geq_X$ is called an *ordered vector space* if the order fulfils the following two conditions which ensure that it is compatible with the algebraic structure of X:

$$\begin{aligned} &\forall x, y, z \in X\colon && x \geq_X y \implies x + z \geq_X y + z, \\ &\forall x, y \in X\, \forall \alpha \geq 0\colon && x \geq_X y \implies \alpha x \geq_X \alpha y. \end{aligned} \tag{2.1}$$

We also write $y \leq_X x$ if $x \geq_X y$. If every pair $x, y \in X$ has a least upper bound or *supremum* and a greatest lower bound or *infimum* with respect to the order $\geq_X$ then X together with $\geq_X$ is called a *vector lattice* or *Riesz space*. The supremum is typically denoted by $x \vee y$ and the infimum by $x \wedge y$. These operations give rise to the notion of an elements' *positive part*, *negative part* and *absolute value*:

$$x^+ := x \vee 0, \quad x^- := (-x) \vee 0 \quad \text{and} \quad |x| := x \vee (-x).$$

Then also the following two equalities hold:

$$x = x^+ - x^-, \quad |x| = x^+ + x^-.$$

If additionally the space X is equipped with a monotone norm $\|\cdot\|\colon X \to \mathbb{R}^+$, i.e. one that fulfils the following condition

$$\forall x, y \in X\colon |x| \geq_X |y| \implies \|x\| \geq \|y\|,$$

then X is called a *normed Riesz space*. If X is complete with respect to $\|\cdot\|$, i.e. if X is a Banach space, then X is called a *Banach lattice*.

We now turn to operators acting between Banach lattices and some of their properties. A linear operator $A\colon X \to Y$ where X and Y are Banach lattices is called *positive* and we write $A \geq 0$ if

$$\forall x \in X\colon x \geq_X 0 \implies Ax \geq_Y 0$$

and it is called *regular* if it can be written as the difference of two positive operators. We denote the space of all regular operators between two Banach lattices X and Y by $\mathcal{L}_r(X, Y)$. Note that this space becomes itself an ordered vector space with the order relation defined as follows:

$$\forall A, B \in \mathcal{L}_r(X, Y)\colon A \geq_{\mathcal{L}_r(X,Y)} B \iff A - B \geq 0.$$

It also holds that every regular operator between two Banach lattices is continuous and hence bounded [1, Thm. 1.31].

In the following sections we will drop the subscript on the order relation whenever it is clear which order is meant.

2.2 Inverse problems in Banach lattices

In this section we will review the approach taken in [37, 38]. Let $\mathfrak{U}$ and $\mathfrak{F}$ be Banach lattices, $u^\star \in \mathfrak{U}$, $f^\star \in \mathfrak{F}$, $A^\star \in \mathcal{L}_r(\mathfrak{U}, \mathfrak{F})$. Let further f be the noisy measurement of $f^\star$ and A be an approximation of the exact operator $A^\star$. Then the inverse problem we want to solve is

$$Au = f. \tag{2.2}$$

To quantify the uncertainty in the data and the operator we use intervals given in the partial order:

$$\begin{aligned} f^l \leq f^\star \leq f^u, \\ A^l \leq A^\star \leq A^u \end{aligned} \tag{2.3}$$

where $A^l, A^u \in \mathcal{L}_r(\mathfrak{U}, \mathfrak{F})$ are regular operators and $f^l, f^u \in \mathfrak{F}$. Assuming $u^\star \geq 0$ one can show the inclusion of $u^\star$ into the following feasible set [37]:

$$\mathcal{Z}_{\text{order}} := \left\{ u \geq 0\colon A^u u \geq f^l, A^l u \leq f^u \right\}. \tag{2.4}$$

Indeed, we have that

$$\begin{aligned} A^u u^\star \geq A^\star u^\star = f^\star \geq f^l, \\ A^l u^\star \leq A^\star u^\star = f^\star \leq f^u. \end{aligned}$$

This set is convex since

$$A^u\left(\lambda u + (1-\lambda)v\right) \overset{A \text{ linear}}{=} \lambda A^u u + (1-\lambda)A^u v \overset{(2.1)}{\geq} \lambda f^l + (1-\lambda)f^l = f^l$$

and

$$A^l\left(\lambda u + (1-\lambda)v\right) \overset{A \text{ linear}}{=} \lambda A^l u + (1-\lambda)A^l v \overset{(2.1)}{\leq} \lambda f^u + (1-\lambda)f^u = f^u$$

and $\lambda u + (1-\lambda)v \geq 0$ for every $u, v \in \mathcal{Z}_{\text{order}}, \lambda \in [0,1]$.

The residual method (1.3) with the feasible set (2.4) leads to the following minimisation problem:

$$\min_{u \in \mathfrak{U}} \mathcal{R}(u) \quad \text{s.t. } u \geq 0, A^u u \geq f^l, A^l u \leq f^u \tag{2.5}$$

where $\mathcal{R}(u)$ is a regularisation functional. We will call (2.5) the partial order-based or interval-based model or interval model for short. Note that if $\mathcal{R}(\cdot)$ is convex, then (2.5) is a convex optimisation problem.

The question now is whether this is indeed a good approximation to the inverse problem (2.2). In classic methods where the error is bounded in the norm as $\|f^\star - f_\delta\| \leq \delta$, the asymptotic behaviour of the solutions is analysed as $\delta \to 0$. The analogue in the lattice framework is to analyse the convergence as the bounds (2.3) get tighter. Therefore we consider sequences of bounds

$$f_1^l \leq f_2^l \leq \ldots \leq f^\star \leq \ldots \leq f_2^u \leq f_1^u, \tag{2.6}$$

$$A_1^l \leq A_2^l \leq \ldots \leq A^\star \leq \ldots \leq A_2^u \leq A_1^u \tag{2.7}$$

such that the intervals converge in norm to a single point:

$$\left\| f_n^u - f_n^l \right\| \xrightarrow{n\to\infty} 0 \tag{2.8}$$

$$\left\| A_n^u - A_n^l \right\| \xrightarrow{n\to\infty} 0. \tag{2.9}$$

Then we consider the sequence of minimisers of (2.5):

$$u_n \in \arg\min_{u \in \mathfrak{U}} \mathcal{R}(u) \quad \text{s.t. } u \geq 0, A_n^u u \geq f_n^l, A_n^l u \leq f_n^u. \tag{2.10}$$

The following theorem states that the sequence of minimisers u_n indeed converges to the exact solution $u^\star$ of the inverse problem (2.2).

Theorem 1 (Convergence [37, Thm. 2])**.**
Suppose that the regularisation functional $\mathcal{R}$ fulfils the following assumptions:

1. *$\mathcal{R}$ is bounded from below,*
2. *$\mathcal{R}$ is lower semi-continuous and*
3. *the non-empty sublevel sets of $\mathcal{R}$ are sequentially compact in $\mathfrak{U}$.*

Then for u_n as defined in (2.10) $u_n \xrightarrow{n\to\infty} u^\star$ *strongly in* $\mathfrak{U}$.

This result guarantees that the model (2.5) is a reasonable approximation to the problem (2.2). The performance of this method in practice largely depends on the choice of the bounds f^l, f^u, A^l and A^u. The choice of the bounds for the data will be discussed in Section 2.4. The choice of the bounds for the operator is not the topic of this work.

The fact that the optimisation problem (2.5) is convex even if errors in the operator are present is a favourable feature of the approach. This distinguishes it from the standard norm based methods. To see this, consider the case when the inverse problem (2.2) is formulated in normed spaces. Then one can assume that an upper bound for the noise is given in terms of norms:

$$\begin{aligned} \|f^\star - f_\delta\| &\leq \delta, \\ \|A^\star - A_\eta\| &\leq \eta \end{aligned} \tag{2.11}$$

where f_δ and A_η are the inexact, noisy data and operator, $f^\star$ and $A^\star$ are the unknown exact data and operator respectively and δ and η are the bounds. Now a feasible set, i.e. a subset of $\mathfrak{U}$ of elements that solve the problem (2.2) within the tolerances allowed by (2.11), can be defined as follows [33]

$$\mathcal{Z}_{\text{norm}} := \{u \in \mathfrak{U} \colon \|A_\eta u - f_\delta\| \leq \eta\|u\| + \delta\}\,. \tag{2.12}$$

It can be verified easily that the exact solution $u^\star$ belongs to $\mathcal{Z}_{\text{norm}}$:

$$\begin{aligned} \|A_\eta u^\star - f_\delta\| &= \|(A_\eta u^\star - A^\star u^\star) + (f^\star - f_\delta)\| \\ &\leq \|A_\eta u^\star - A^\star u^\star\| + \|f^\star - f_\delta\| \\ &\leq \eta\|u^\star\| + \delta. \end{aligned}$$

The residual method (1.3) in this case leads to the following optimisation problem:

$$\min_{u \in \mathfrak{U}} \mathcal{R}(u) \quad \text{s.t. } \|A_\eta u - f_\delta\| \leq \eta\|u\| + \delta.,$$

where $\mathcal{R}(u)$ is a regularisation functional. Unless $\eta = 0$ the set (2.12) is non-convex and therefore does not allow us to use the well studied techniques of convex optimisation.

2.3 Relation to existing methods

The approach presented in Section 2.2 is not the only interval-based method in imaging. In this section we would like to highlight the connection between this approach and the approach taken in [8, 10] for TV- or TGV-based decompression of JPEG images. As we shall see, the approach of [8, 10] is a special case of the framework presented in Section 2.2.

In [8] the authors proposed a variational model for JPEG decompression based on minimizing the Total Variation over a special set of approximate solutions defined by the data contained in the JPEG image. Further analysis of the general model with Total Generalised Variation regularisation was presented in [10].

The JPEG algorithm compresses images by computing the coefficients of the blockwise discrete cosine transform (BDCT) and then rounding them to integers. This quantisation leads to a loss of information and during the reconstruction we in fact only know that the true BDCT coefficient lies between two consecutive integers. Using this information, the authors of [8, 10] defined a feasible set as follows:

$$U \coloneqq \left\{ u \in L^2(\Omega) \colon (Au)_n \in J_n \; \forall\, n \in \mathbb{N} \right\},$$

where $A \colon L^2(\Omega) \to \ell^2$ is the BDCT forward operator and $(J_n)_{n\in\mathbb{N}} \coloneqq ([l_n, r_n])_{n\in\mathbb{N}}$ are the lower and upper quantisation bounds for the BDCT coefficients (integers).

For decompression they propose to solve the following optimisation problem:

$$\min_{u \in L^2(\Omega)} \mathrm{TV}(u) + \mathcal{I}_U(u), \tag{2.13}$$

where $\mathcal{I}_U$ is the convex indicator functional associated with the set U.

Let us see how this approach fits into the framework of Section 2.2. The space $L^2(\Omega)$ becomes a Banach lattice if it is equipped with the following partial order

$$\forall f, g \in L^2(\Omega) \colon f \geq_{L^2(\Omega)} g \iff f(x) \geq g(x) \text{ for almost every } x \in \Omega, \tag{2.14}$$

whereas ℓ^2 becomes a Banach lattice with the following partial order

$$\forall\, (x_n)_{n\in\mathbb{N}}, (y_n)_{n\in\mathbb{N}} \in \ell^2 \colon (x_n)_{n\in\mathbb{N}} \geq_{\ell^2} (y_n)_{n\in\mathbb{N}} \iff x_n \geq y_n \; \forall\, n \in \mathbb{N}. \tag{2.15}$$

Let l, r be the two sequences defined by the lower and upper bounds of the intervals $(J_n)_{n\in\mathbb{N}}$:

$$l \coloneqq (l_n)_{n\in\mathbb{N}} \text{ and } r \coloneqq (r_n)_{n\in\mathbb{N}}\,. \tag{2.16}$$

Although A is acting $L^2(\Omega) \to \ell^2$ the requirement that the bounds l, r are in ℓ^2 might be too restricting. Therefore we require that

$$l, r \in \ell^\infty$$

and consider A acting from $L^2(\Omega)$ into ℓ^∞ as well, which is a Banach lattice with the same partial order (2.15). Using the natural assumption that the true image is positive, i.e. $u^\star \geq 0$, we can define the following feasible set:

$$U = \left\{ u \in L^2(\Omega) \colon u \geq 0, l \leq_{\ell^\infty} Au \leq_{\ell^\infty} r \right\}, \tag{2.17}$$

which can be seen as the feasible set (2.4) with exact operator $A^l = A^u = A$.

The problem (2.13) can be reformulated as follows:

$$\min_{u \in L^2(\Omega)} \mathrm{TV}(u) \quad \text{s.t. } l \leq Au \leq r. \tag{2.18}$$

The advantage of translating the problem (2.13) into the framework of Section 2.2 is that we can now apply Theorem 1, i.e., the solutions are guaranteed to converge as the bounds get tighter. In the setting of [8, 10] this corresponds to convergence as the quantization values for the BDCT coefficients get smaller.

2.4 Choice of the intervals

Theorem 1 guarantees convergence of regularised solutions (2.10) to the exact solution $u^\star$ as the intervals (2.3) containing the exact right-hand side and operator get tighter. In this thesis we only study the choice of the intervals for the data and assume the operator to be exactly known. Section 2.3 gives an example of an applied problem where these intervals arise naturally during quantisation and convergence of these bounds is understood as successive refinement of this quantisation.

There are other cases when the intervals can be chosen rather straightforwardly. One is the case of additive noise with finite support. Suppose that the measurement f is corrupted by additive noise ε supported on the interval $[\varepsilon^l, \varepsilon^u]$:

$$f = f^\star + \varepsilon, \quad \varepsilon^l \leq \varepsilon \leq \varepsilon^u,$$

where $f^\star = Au^\star$ is the exact measurement. Then, given f and the pair $(\varepsilon^l, \varepsilon^u)$, we can find an interval containing the exact measurement $f^\star$ as follows:

$$f - \varepsilon^u \leq f^\star \leq f - \varepsilon^l.$$

Convergence of these bounds is implied by the convergence of the support of the noise $[\varepsilon^l, \varepsilon^u]$ to a single point in the sense that $\|\varepsilon^u - \varepsilon^l\| \to 0$.

If the noise is unbounded, the choice of the intervals is less straightforward. [25] proposed to use confidence intervals for the unknown right-hand side as intervals $[f^l, f^u]$ in (2.3). If we assume additive zero-mean noise, then the measurement f is a random variable with mean $f^\star$. If we had a series of repeated measurements, we could construct confidence intervals for the true mean of f that would converge as the number of repeated measurements would go to infinity [25].

Interpreting the intervals $[f^l, f^u]$ in (2.3) as confidence intervals means the inequalities $f^l \leq f^\star \leq f^u$ hold only with a certain probability that diminishes as the number of pixels in the image grows [25]. We will come back to this observation later in Chapter 3. The confidence level in this case can play the role of a regularisation parameter. It determines the width of the intervals $[f^l, f^u]$ and therefore balances the influence of the fidelity term and the regulariser.

Since in practice we do not have a series of repeated measurements of the same object, we have to come up with tricks to estimate the lower and upper bounds f^l and f^u from a single measurement. Assuming that the noise is additive and independent and identically distributed (i.i.d.) in different pixels and also assuming that we know the ground truth in some background region of the image, we can obtain as many samples of the noise as there are pixels in the background region. Using a histogram of this noise and fixing a parameter $\beta \in (0, 1)$, we can find an interval $[\varepsilon^l, \varepsilon^u]$ containing $(1-\beta)\cdot 100\%$ of the noise samples. We use the $\frac{\beta}{2}$-quantile of the noise distribution to find ε^l and the $(1 - \frac{\beta}{2})$-quantile to find ε^u. We will refer to the parameter β as the *quantile parameter*.

In Chapter 3 we will study numerically the effect of the quantile parameter β on the reconstruction quality and the competitiveness of the interval-based approach in the case of unbounded noise.

Chapter 3: Numerical study of the interval-based model for unbounded noise

In Section 2.4 we analysed how the intervals (2.3) can be chosen if the noise is not finitely supported. In this chapter we will now investigate how the choice of the quantile parameter impacts the reconstruction quality. We will also study whether the partial order based approach can outperform standard approaches on images with non-Gaussian noise.

In section Section 3.1 we will explain the set-up of the study. Section 3.2 is dedicated to some properties of the partial order based model in conjunction with the total variation regularisation functional. After a brief detour into how the optimisation problems can be solved numerically in Section 3.3 we will explain how we chose the parameters for the experiments in Section 3.4. In Section 3.6 we will then present the results of the study and discuss how the choice of the quantile parameter influences the reconstruction quality. Lastly we will draw some conclusions in Section 4.

3.1 Set-up

In this chapter we will work in a finite-dimensional setting where $u \in \mathfrak{U} = I^n$ and $f, f^l, f^u \in \mathfrak{F} = I^m$ are discrete images represented as column vectors, where I is the set of possible pixel values. In order to concentrate on the choice of the quantile parameter for the data we only study the case of denoising images, i.e., where the forward operator A is the identity matrix and therefore $n = m$. For all experiments greyscale images with values between 0 and 255 were used, i.e., $I = [0, 255] \cap \mathbb{Z}$.

Regularisation models

For the numerical study we use the model (2.5) with total variation regularisation (1.5)

$$\min_{u \in \mathfrak{U}} \mathrm{TV}(u) \quad \text{s.t. } f^l_\beta \leq u \leq f^u_\beta, \tag{3.1}$$

where $f^l_\beta = f - \varepsilon^u$ and $f^u_\beta = f - \varepsilon^l$. ε^l and ε^u are, respectively, the $\frac{\beta}{2}$-quantile and the $(1 - \frac{\beta}{2})$-quantile from the histogram of the background noise as discussed in Section 2.4. We use this regulariser because it is widely used and so we can get results that are comparable to standard models. We will compare the results with two other well-known total variation based approaches, namely the Rudin-Osher-Fatemi (ROF) model [47]

$$\min_{u \in \mathfrak{U}} \|u - f\|_2 + \alpha \, \mathrm{TV}(u) \tag{3.2}$$

and the TV-L^1 model [16]

$$\min_{u \in \mathfrak{U}} \|u - f\|_1 + \alpha \, \mathrm{TV}(u). \tag{3.3}$$

Quality metrics

To evaluate the reconstruction quality we calculate the *peak signal-to-noise ratio* (PSNR) which measures absolute errors in every pixel as well as the *structural similarity* (SSIM) index [55] which assesses the quality based on changes in the structural information. The PSNR is defined as follows:

$$\mathrm{PSNR}(x, y) = 10 \times \log_{10}\left(\frac{L^2}{\mathrm{MSE}(x, y)}\right), \tag{3.4}$$

where L is the peak signal level of the images, in our case $L = 255$, and $\mathrm{MSE}(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2$ is the mean squared error. A higher PSNR indicates better reconstruction quality. PSNR has been criticised as not agreeing with the human perception of image quality [54], however, it is still widely used to assess reconstruction quality, e.g., [13, 49].

The SSIM uses a combination of three different characteristics of the compared images. The luminance is estimated using the mean intensities μ_x, μ_y, the contrast is estimated using the standard deviations σ_x, σ_y and the structure is estimated using the covariance σ_{xy}. The SSIM is then defined as:

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)\,(2\sigma_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)}, \tag{3.5}$$

where $C_1 = (0.01 \times L)^2$ and $C_2 = (0.03 \times L)^2$ are regularisation constants to provide stability in regions where the mean or standard deviation are close to zero, and $L = 255$ is the peak signal level. The SSIM is usually computed locally in every pixel and then averaged to give one single quality measure of the entire image. The local SSIM is computed over an isotropic Gaussian weighted window. The SSIM values range from 0 to 1, where 1 corresponds to an exact reconstruction.

The Rician distribution

As the true noise model we chose Rician distributed additive noise for the following reasons. First, it often arises in image reconstruction. For example, in magnetic resonance imaging (MRI), where the image is the magnitude of a complex signal with normally distributed components and therefore follows a Rician distribution [6, 29]. The probability density function of a Rician distributed variable is as follows:

$$P\left(x \,|\, \nu, \sigma\right) = \frac{x}{\sigma^2} e^{\frac{-(x^2+\nu^2)}{2\sigma^2}} I_0\left(\frac{x\nu}{\sigma^2}\right), \tag{3.6}$$

where $I_0(x)$ is the modified Bessel function of the first kind. Another reason for the choice of Rician noise is that for small values of ν the distribution departs significantly from the Gaussian distribution and therefore is not modelled accurately by the standard L^2 fidelity, leaving room for improvement. An important feature of the Rician distribution is that it is supported on the interval $[0, \infty)$ and thus admits only positive values.

An exact fidelity term for Rician noise in images was proposed in [22]. This fidelity term is non-convex and subsequently different convex relaxations were developed [18, 22]. In

Figure 1. The six images used in the experiments: bird, cameraman, circles, horiz, lena, squares (from top left to bottom right). Each image has a resolution of 256×256px.

this work, we do not compare our approach to these models, since our goal is not to model Rician noise, but to propose a simple approximate fidelity term that is independent of the noise distribution. We will compare our approach to the L^2 and L^1 fidelity terms, which are often used due to their simplicity also in situations where the necessary assumptions on the noise distribution do not hold.

Test images

As test images we used three often-used real images and three synthetic images to assess different aspects of the new approach. The original images (without noise) are shown in Figure 1.

3.2 The TV regularised interval model

In this section we will discuss the model (3.1) and its properties. We will see that the minimiser of (3.1) is not necessarily unique and introduce a small correction term to ensure uniqueness and justify the use of quantitative image quality measures such as PSNR and SSIM. We will then show that the used regularisation functionals fulfil the assumptions made in Theorem 1.

Non-uniqueness

Since $\mathcal{R}_0(u) = \mathrm{TV}(u)$ is not a strictly convex functional, we cannot guarantee the minimiser of (3.1) to be unique. Indeed consider the following simple example.

Example 2.
Suppose there are constants $c_1 < c_2$ such that $f_i^l \leq c_1$ and $c_2 \leq f_i^u$ for all $i = 1, \dots, n$. Then for every $c_1 \leq c \leq c_2$ the constant function $u \equiv c$ is a minimiser of (3.1) as $\mathrm{TV}(u) = 0$ and $u \in \mathcal{Z}_{\mathrm{order}}$.

In order to ensure uniqueness, we need to modify the regulariser. Since BV is a subset of $L^1(\Omega)$ it is natural to consider $\mathcal{R}_1(u) = \mathrm{TV}(u) + \|u\|_1$ as a regulariser. However, as in the standard TV-L^1 model [16], the minimisers of (3.1) with $\mathcal{R}_1(u)$ as regulariser are not unique either. This can be seen in the following example.

Example 3.
Suppose we have d-dimensional data and the indices $I = \left\{ i_j \,\middle|\, j = 1, \dots, (2d)^d \right\}$ form a d-dimensional cube with side length $2d$. Now let $f^u \equiv c$ and $f_i^l = \begin{cases} c, & i \notin I \\ c-1, & i \in I \end{cases}$. Let us now consider the data defined by $(u_\tau)_i = \begin{cases} c, & i \notin I \\ c-\tau, & i \in I \end{cases}$ for $0 < \tau \leq 1$. This leads to $\mathcal{R}_1(u_\tau) = 2d \times (2d)^{d-1} \times \tau + (nc - |I|\tau) = nc$, which shows that in this case the value $\mathcal{R}_1(u_\tau)$ is independent of τ and thus every u_τ is a minimiser.

To ensure uniqueness of the minimiser in our experiments we amended the TV functional with an L^2 term:

$$\min_{u \in \mathfrak{U}} \mathrm{TV}(u) + \gamma \left\| u - \bar{f} \right\|_2 \quad \text{s.t. } f^l \leq u \leq f^u, \tag{3.7}$$

where $\bar{f}$ is the mean of f and γ is a small constant because we don't want the additional term to significantly influence the outcome of the regularisation other than to guarantee the uniqueness of the solution. In our experiments we set $\gamma = 10^{-4}$. The minimiser of this problem exists and is unique because the feasible set $\mathcal{Z}_{\mathrm{order}}$ is closed, bounded and convex and the employed regularisation functional, which is the sum of a convex and a strictly convex functional, is strictly convex.

Convergence of the minimisers

Now we will analyse whether the functionals discussed above fulfil the conditions of Theorem 1. In this paragraph we will work in the infinite-dimensional setting and set $\mathfrak{U} = \mathfrak{F} = L^1(\Omega)$ where $\Omega \subseteq \mathbb{R}^n$ is a compact image domain. We consider the regularisers as functionals $\mathcal{R} \colon \mathfrak{U} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$.

Starting with the BV norm as regulariser $\mathcal{R}_1(u) = \mathrm{TV}(u) + \|u\|_1$ we first observe that it is bounded from below. Since TV is lower semi-continuous in the $L^1_{loc}(\Omega)$ norm topology [3, Prop. 3.6] and for compact Ω it holds that $L^1_{loc}(\Omega) = L^1(\Omega)$ we get that TV is lower semi-continuous with respect to the $L^1(\Omega)$ norm topology. Therefore the second assumption of Theorem 1 holds as well. It was shown in [3, Prop. 3.23] that any sequence in the sublevel sets of the BV norm has a subsequence that weakly* converges in BV. By [3, Prop. 3.13] this implies that the subsequence converges strongly in $L^1(\Omega)$ showing that the sublevel sets of $\mathcal{R}_1(u)$ are sequentially compact in $L^1(\Omega)$. Thus it satisfies all assumptions of Theorem 1.

In fact, since in the absence of a forward operator the L^1 norm is bounded on the feasible set $\mathcal{Z}_{\text{order}}$, the condition $\text{TV}(u) \leq \lambda$ implies the condition $\text{TV}(u) + \|u\|_1 \leq \lambda_1$ for some $\lambda_1 \geq \lambda$ for every $u \in \mathcal{Z}_{\text{order}}$. Therefore, also $\mathcal{R}_0(u) = \text{TV}(u)$ satisfies the assumptions of Theorem 1.

Now we will show that also $\mathcal{R}_2(u) = \text{TV}(u) + \|u\|_2$ fulfils the assumptions of Theorem 1. The first two assumptions hold analogously to the reasoning above. It remains to show that the sublevel sets are sequentially compact in $L^1(\Omega)$. Since closed subsets of compact sets are also compact we will show that $S_2 := \{u \in L^2 \,|\, \text{TV}(u) + \|u\|_2 \leq \lambda_2\}$ is a subset of $S_1 := \{u \in L^1 \,|\, \text{TV}(u) + \|u\|_1 \leq \lambda_1\}$ for some appropriately chosen λ_1, as we have already shown S_1 to be compact. Closedness of S_2 follows from the fact that both TV and the norm are lower semi-continuous [3, Prop. 3.6]. If the domain Ω has finite measure $\mu(\Omega) < \infty$ then the space $L^2(\Omega)$ can be continuously embedded in $L^1(\Omega)$ as can be seen using Hölder's inequality for $u \in L^2(\Omega)$ and $I \in L^2(\Omega)$ where $I(x) = 1$ a.e. in Ω:

$$\|u\|_1 = \|Iu\|_1 \leq \|I\|_2 \|u\|_2 = \mu(\Omega)^{\frac{1}{2}} \|u\|_2.$$

For $u \in S_2$ we now get the following estimate:

$$\begin{aligned} & & \text{TV}(u) + \|u\|_1 &\leq \text{TV}(u) + \mu(\Omega)^{\frac{1}{2}} \|u\|_2 & \\ &\text{if } \mu(\Omega)^{\frac{1}{2}} \geq 1 & &\leq \mu(\Omega)^{\frac{1}{2}} \left(\text{TV}(u) + \|u\|_2\right) & \leq \mu(\Omega)^{\frac{1}{2}} \lambda_2 \\ &\text{if } \mu(\Omega)^{\frac{1}{2}} < 1 & &\leq \text{TV}(u) + \|u\|_2 & \leq \lambda_2. \end{aligned}$$

Setting $\lambda_1 = \max\left\{\mu(\Omega)^{\frac{1}{2}} \lambda_2, \lambda_2\right\}$ we see that $S_2 \subset S_1$. From this we can conclude that the sublevel sets of $\mathcal{R}_2(u)$ are sequentially compact in $L^1(\Omega)$ and we get strong convergence of the minimisers (2.10) in $L^1(\Omega)$ by Theorem 1.

3.3 Optimisation methods

Since the problem (3.7) is convex, we can use the well-studied methods of convex optimisation [7]. Convexity ensures that every local minimiser is also a global minimiser. In the case of strict convexity we find that the minimiser is also unique. A main advantage of convex problems is that they can be solved efficiently in polynomial time. Polynomial complexity is achieved by, e.g., interior-point methods [41]. Interior-point methods have been developed first for linear programming problems [35] and were extended to handle convex problems through the use of conic formulations in [41].

For the experiments we used the modelling framework CVX [27] to translate the problem into a conic formulation and then used MOSEK [4, 40] to solve it. In this section we want to recall the setting of conic problems and see how a conic reformulation of (3.1) can be achieved. Then we will look at the CVX package and how it can be used to solve medium-scale convex optimisation problems.

Conic problems

Let us first recall what a cone is. A subset K of a vector space V is called a *cone* if

$$\forall x \in K \, \forall \lambda \geq 0 \colon \lambda x \in K.$$

A cone K is called:

closed if K is a closed subset of V,

convex if $\forall x, y \in K$ we have $x + y \in K$,

pointed if $K \cap -K = \{0\}$,

solid if K has non-empty interior,

proper if K is closed, convex, pointed and solid.

Every proper cone K allows us to define a partial order $\geq_K$ on V as follows:

$$\forall x, y \in V : x \geq_K y \iff x - y \in K.$$

A *conic program* can be formulated as [7]

$$\begin{aligned} &\min c^\top x \quad \text{s.t. } Ax = b, x \geq_K 0 \\ \Leftrightarrow &\min c^\top x \quad \text{s.t. } Ax = b, x \in K \end{aligned} \tag{3.8}$$

Note that for two cones $K_1 \subseteq V_1, K_2 \subseteq V_2$ the direct product $K_1 \times K_2 \subseteq V_1 \times V_2$ is also a cone in the product space. It is therefore possible to restrict different entries of x in (3.8) to different cones

$$K = K_1 \times \cdots \times K_l, x = (x_1, \ldots, x_l) : x \in K \Leftrightarrow x_i \in K_i,\ i = 1, \ldots, l.$$

Three important cones for convex optimisation are the *positive cone*

$$\mathbb{R}^n_+ := \{x \in \mathbb{R}^n : x_i \geq 0, i = 1, \ldots, n\},$$

the *second-order cone* or *Lorentz cone*

$$Q^n := \left\{x \in \mathbb{R}^n : x_1^2 + x_2^2 + \cdots + x_{n-1}^2 \leq x_n^2, x_n \geq 0\right\}$$

and the *positive semi-definite cone*

$$S^n_+ := \left\{A \in \mathbb{R}^{n \times n} : A = A^\top, x^\top A x \geq 0 \, \forall x \in \mathbb{R}^n\right\},$$

which are all proper cones in $\mathbb{R}^n$ and $\mathbb{R}^{n \times n}$, respectively. The whole space $\mathbb{R}^n$ can also be considered a cone but it is not proper because it is not pointed.

TV discretisation

In our experiments we treated the two-dimensional images as column vectors by stacking the columns of the image on top of one another. Assuming, for simplicity, square images with $n \times n$ pixels, we represent the two-dimensional images as column vectors $u \in \mathbb{R}^{n^2}$. Then we define the gradient in x and y directions as follows

$$\begin{aligned}
(\partial_x^+ u)_{i,j} &:= u_{i+1+(j-1)n} - u_{i+(j-1)n} && \forall\, i = 1, \dots, n-1, && j = 1, \dots, n \\
(\partial_y^+ u)_{i,j} &:= u_{i+jn} - u_{i+(j-1)n} && \forall\, i = 1, \dots, n, && j = 1, \dots, n-1
\end{aligned}$$

with zero boundary conditions

$$\begin{aligned}
(\partial_x^+ u)_{n,j} &:= 0 \quad \forall\, j = 1, \dots, n \\
(\partial_y^+ u)_{i,n} &:= 0 \quad \forall\, i = 1, \dots, n.
\end{aligned}$$

We define the full gradient in each point as the vector

$$(\partial^+ u)_{i,j} = \begin{pmatrix} (\partial_x^+ u)_{i,j} \\ (\partial_y^+ u)_{i,j} \end{pmatrix}.$$

Now we can define the discretisation of the total variation functional in two different ways. The isotropic variant

$$\mathrm{TV}(u) := \sum_{1 \le i,j \le n} \|(\partial^+ u)_{i,j}\|_2 = \sum_{1 \le i,j \le n} \left(((\partial_x^+ u)_{i,j})^2 + ((\partial_y^+ u)_{i,j})^2 \right)^{\frac{1}{2}}, \tag{3.9}$$

and the anisotropic variant

$$\mathrm{TV}_{\mathrm{anisotr}}(u) := \sum_{1 \le i,j \le n} \|(\partial^+ u)_{i,j}\|_1 = \sum_{1 \le i,j \le n} (\partial_x^+ u)_{i,j} + (\partial_y^+ u)_{i,j}, \tag{3.10}$$

which is a special case of the general anisotropic TV first introduced and studied in [45]. In our experiments we used the isotropic TV variant as in the original ROF model [47].

With these definitions a conic formulation (3.8) of the model (3.1) can be written as follows [24]:

$$\begin{aligned}
\min \quad & \sum_{1 \le i,j \le n} t_{i,j} \\
\text{s.t.} \quad & u - f^l \in \mathbb{R}^{n^2}_+ \\
& f^u - u \in \mathbb{R}^{n^2}_+ \\
& \begin{pmatrix} (\partial^+ u)_{i,j} \\ t_{i,j} \end{pmatrix} \in Q^3 \quad \forall 1 \le i,j \le n,
\end{aligned}$$

where $t_{i,j}$ and u are the variables. For more details on conic formulations of the ROF model and other TV regularised models as well as their dual forms see [24].

CVX

CVX is a MATLAB package that translates a convex problem into a conic one and calls an interior-point optimiser to solve it [26, 27]. CVX can handle problem formulations that follow a set of rules, called the *disciplined convex programming* (DCP) ruleset [28]. These are rules that specify how convex functions can be combined while preserving convexity. Together with a library of known convex functions CVX can automatically verify that the given problem is convex and automate the following analysis and solution.

Another feature of CVX is a new approach to modelling smooth and non-smooth functions as parametrised convex programs. These are called *graph implementations* [26]. Here functions are point-wise represented by a convex optimisation problem, specifically they exploit the relation between a convex function and its epigraph:

$$f(x) \equiv \inf \{y \,|\, (x, y) \in \operatorname{epi} f\},$$

where the epigraph is defined as

$$\operatorname{epi} f := \{(x, y) \in X \times \mathbb{R} \,|\, f(x) \leq y\}.$$

The absolute value of x can, e.g., be modelled as the solution to

$$|x| = \min y \quad \text{s.t. } x \leq y, -x \leq y.$$

This allows CVX to also transform non-smooth problems into smooth conic problems. The drawback of this approach is that often many new variables and constraints are introduced in the translation process, therefore it is not applicable to very large problems.

To solve the resulting transformed problem, several external interior-point optimisers are supported [40, 50, 51].

3.4 Choice of the parameters

In this section we will briefly explain how we chose the regularisation parameters in the experiments. Since we want to see whether the interval-based model can perform better than the ROF and TV-L^1 models, we are interested in the best possible reconstructions obtained with each of the latter two models. We optimised the reconstruction once with respect to the PSNR of the reconstruction and once with respect to the SSIM.

To find the optimal value of the regularisation parameter, we first identified a value h_0^u at which the images are completely homogeneous. For the ROF model (3.2) we found that in our examples $h_0^u = 1$ is sufficient and $h_0^u = 10$ is sufficient for the TV-L^1 model (3.3). Then we set $h_0^l = 0$ and evaluated the optimisation problems at 11 equidistant parameter values $\alpha_0 = h_0^l, \alpha_1, \ldots, \alpha_{10} = h_0^u$. Let α_j be the value where the PSNR/SSIM of the reconstruction was highest and define $h_1^l = \begin{cases} \alpha_{j-1} & \text{for } j > 0 \\ \alpha_0 & \text{for } j = 0 \end{cases}$ and $h_1^u = \begin{cases} \alpha_{j+1} & \text{for } j < 10 \\ \alpha_{10} & \text{for } j = 10 \end{cases}$. We iterated this procedure until the range $[h_i^l, h_i^u]$ was smaller than some predefined tolerance level $h_i^u - h_i^l \leq \tau$.

In our experiments we chose $\tau = 5 \times 10^{-4}$. This led to 5 iterations for the ROF model, resulting in 55 evaluations of the optimisation problem, and 7 iterations for TV-L^1, resulting in 77 evaluations.

The quantile parameter β for the interval-based model was chosen analogously with the starting range $[h_0^l, h_0^u] = [0, 1]$ and $\tau = 5 \times 10^{-4}$, resulting in 55 evaluations to find the best reconstruction.

3.5 Reconstruction artefacts

Since we use confidence intervals for the noise in every pixel to obtain the lower and upper bounds f^l and f^u in (3.1), we expect there will be pixels where the actual value of the noise will be outside the confidence interval and therefore the true image will not belong to the interval $[f^l, f^u]$. Since this interval enters the optimisation problem (3.1) as a hard constraint, there is no chance of recovering the exact image in those pixels. Therefore we expect such pixels to appear as outliers in the reconstruction, i.e. as isolated pixels having a different value from the surrounding ones. The number of these outliers depends on the quantile parameter β. This will be made clear by the following proposition.

Proposition 4 (Probability of outliers)**.**
Let $u^\star \in \mathbb{R}^{m\times n}$ be the exact solution and $f \in \mathbb{R}^{m\times n}$ the measured data. Suppose the measurement is corrupted by additive noise with known probability density function $g(x)$ such that $f_{ij} = u^\star_{ij} + \varepsilon_{ij}$ where $i = 1, \dots, m$, $j = 1, \dots, n$ and ε_{ij} are independent and identically distributed.

For $\beta \in [0, 1]$ let us define $f^l_{ij} := f_{ij} - Q_g\left(1 - \frac{\beta}{2}\right)$ and $f^u_{ij} := f_{ij} - Q_g\left(\frac{\beta}{2}\right)$, where Q_g is the quantile function[1]. Then the probability that the value $u^\star_{ij}$ in pixel (i, j) of the exact solution lies outside the interval $[f^l_{ij}, f^u_{ij}]$ is exactly β.

Proof. Begin by calculating the probability that $u^\star_{ij}$ is less than the lower bound:

$$\begin{aligned} P\left(u^\star_{ij} < f^l_{ij}\right) &= P\left(f_{ij} - \varepsilon_{ij} < f_{ij} - Q_g\left(1 - \frac{\beta}{2}\right)\right) \\ &= P\left(\varepsilon_{ij} > Q_g\left(1 - \frac{\beta}{2}\right)\right) \\ &= 1 - P\left(\varepsilon_{ij} \le Q_g\left(1 - \frac{\beta}{2}\right)\right) \\ &= 1 - \left(1 - \frac{\beta}{2}\right) \\ &= \frac{\beta}{2} \end{aligned}$$

[1]The quantile function of g is defined as $Q_g \colon [0, 1] \to \mathbb{R}, p \mapsto \inf\left\{x \in \mathbb{R} \colon \int_{-\infty}^{x} g(y)dy \ge p\right\}$. Note that this implies $P(\varepsilon_{ij} \le Q_g(p)) = p$.

Table 1. PSNR/SSIM values of the noisy and reconstructed images. Rician noise with $\nu = 0$ and $\sigma = 10$. The parameters were chosen to give the best possible PSNR.

		Reconstruction model		
Images	Noisy	ROF	TV-L^1	Interval
bird	25.1337/0.7025	25.9295/0.9382	26.2029/0.9265	36.8625/0.9212
cameraman	25.1399/0.7492	25.6870/0.8964	25.5367/0.8329	34.4167/0.9018
circles	25.2771/0.6378	26.2369/0.9485	26.6582/0.9498	40.8764/0.9679
horiz	25.1340/0.5901	26.0748/0.8952	26.4317/0.8874	40.5875/0.9771
lena	25.1337/0.8063	25.6598/0.9145	25.6621/0.8872	34.4891/0.9153
squares	25.1337/0.6119	26.1749/0.9806	26.7853/0.9804	46.6163/0.9930

Table 2. PSNR/SSIM values of the noisy and reconstructed images. Rician noise with $\nu = 0$ and $\sigma = 10$. The parameters were chosen to give the best possible SSIM.

		Reconstruction model		
Images	Noisy	ROF	TV-L^1	Interval
bird	25.1337/0.7025	25.9256/0.9394	26.1873/0.9301	36.1587/0.9301
cameraman	25.1399/0.7492	25.6612/0.9002	25.1114/0.8511	33.7467/0.9188
circles	25.2771/0.6378	26.2341/0.9488	26.6479/0.9521	31.6435/0.9964
horiz	25.1340/0.5901	26.0662/0.8967	26.4073/0.8921	38.7219/0.9902
lena	25.1337/0.8063	25.6506/0.9170	25.5992/0.8909	34.1939/0.9209
squares	25.1337/0.6119	26.1749/0.9806	26.6435/0.9818	41.5186/0.9969

Similarly one can show that $P\left(u^\star_{ij} > f^u_{ij}\right) = \frac{\beta}{2}$.
Thus $P\left(u^\star_{ij} \notin \left[f^l_{ij}, f^u_{ij}\right]\right) = P\left(u^\star_{ij} < f^l_{ij}\right) + P\left(u^\star_{ij} > f^u_{ij}\right) = \beta$. □

To see how large the effect can be consider a standard image with 100×100 pixels. The probability that every point of the exact solution lies inside the constructed intervals is $P\left(f^l \leq u^\star \leq f^u\right) = (1-\beta)^{m \times n}$. For reasonable values of β in the order of 10^{-2} and 10^{-1} this is just 2×10^{-44} and 3×10^{-458}, respectively.

3.6 Numerical Experiments

In this section we will present the results of our numerical experiments. The original images shown in Figure 1 were corrupted by Rician noise with parameters $\nu = 0$ and $\sigma = 10$ as well as $\sigma = 30$ and then reconstructed using the ROF model (3.2), the TV-L^1 model (3.3) and the interval-based model (3.7). The PSNR and SSIM values of the reconstructions can be seen in Tables 1 to 4. The reconstructions are shown in Figures 2 to 7. We do not show all reconstructions with all combinations of parameters, concentrating on the most revealing ones.

Table 3. PSNR/SSIM values of the noisy and reconstructed images. Rician noise with $\nu = 0$ and $\sigma = 30$. The parameters were chosen to give the best possible PSNR.

		Reconstruction model		
Images	Noisy	ROF	TV-L^1	Interval
bird	15.6489/0.2711	16.5782/0.8514	17.0139/0.8424	30.5621/0.7898
cameraman	15.6708/0.3704	16.4419/0.7235	16.6737/0.6610	27.3512/0.7308
circles	16.5245/0.3459	17.4363/0.7970	17.8473/0.8002	23.4186/0.4859
horiz	15.6034/0.1708	16.5762/0.6225	17.1142/0.6348	31.5209/0.8591
lena	15.6620/0.4172	16.4471/0.7538	16.7731/0.7333	27.5896/0.7491
squares	15.5980/0.1648	16.6382/0.8867	17.2755/0.8895	36.8142/0.9449

Table 4. PSNR/SSIM values of the noisy and reconstructed images. Rician noise with $\nu = 0$ and $\sigma = 30$. The parameters were chosen to give the best possible SSIM.

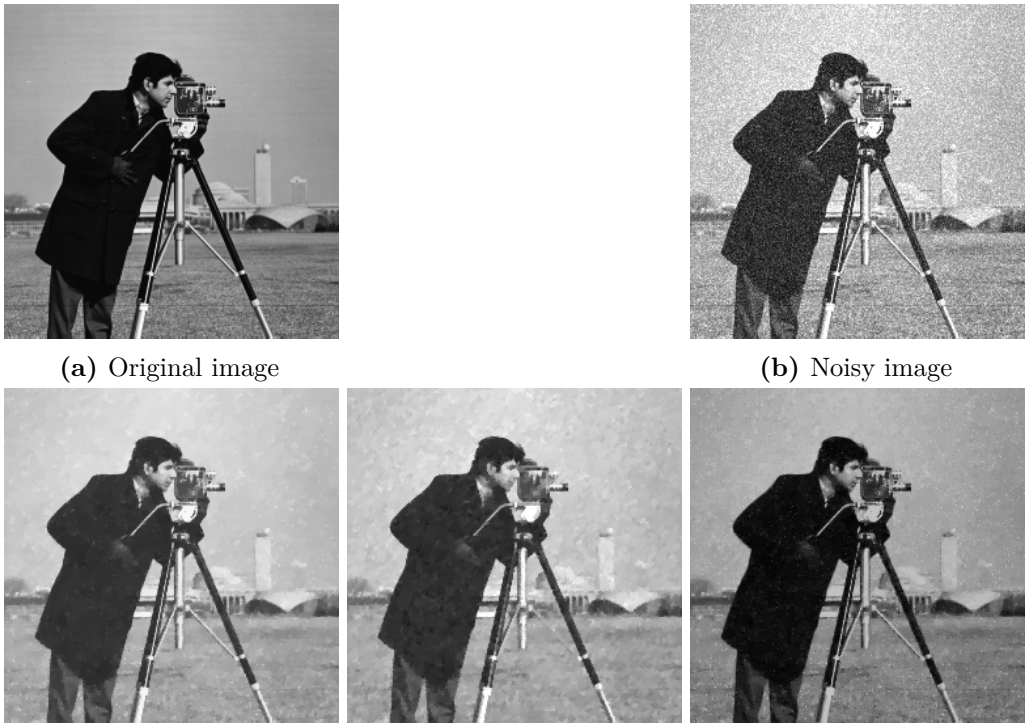
		Reconstruction model		
Images	Noisy	ROF	TV-L^1	Interval
bird	15.6489/0.2711	16.5738/0.8558	17.0064/0.8448	27.7865/0.8495
cameraman	15.6708/0.3704	16.4258/0.7334	16.6291/0.6923	26.3029/0.7911
circles	16.5245/0.3459	17.4195/0.8065	17.7719/0.8147	17.1736/0.9447
horiz	15.6034/0.1708	16.5699/0.6286	17.1138/0.6349	28.9245/0.9425
lena	15.6620/0.4172	16.4384/0.7610	16.7702/0.7349	26.7800/0.7769
squares	15.5980/0.1648	16.6379/0.8868	17.2096/0.8942	33.2705/0.9748

Visual inspection

The first observation to make is that the colours in the interval-based reconstructions are closer to the original than in the ROF and TV-L^1 reconstructions. With ROF and TV-L^1 the reconstructions look too bright, as if the images were overexposed (see, e.g., Figure 2). This can be explained by the different assumptions about the noise that are made by the reconstruction models. ROF as well as TV-L^1 implicitly assume a noise distribution with zero mean while the bounds in the interval-based model provide the information that the Rician noise is only positive. Therefore the ROF and TV-L^1 reconstructed grey values are approximately shifted by the mean of the noise.

One typical feature of the total variation regulariser is the staircasing that we briefly discussed in Section 1.2. This effect can be seen most prominently in the reconstructions of the 'bird' image (Figure 3). It is also visible in the reconstructions of the 'cameraman' and 'lena' image as these contain smooth non-constant regions which are reconstructed with piecewise-constant solutions. Staircasing can be observed with each optimisation model we used, since they all incorporate the TV regulariser.

In the TV-L^1 reconstruction of the 'squares' image, we see that the middle square is completely removed. This dependence on the geometry of the squares and not on their contrast is a well-known feature of the TV-L^1 model. The contrast of objects is preserved



(a) Original image **(b)** Noisy image

(c) ROF, $\alpha = 0.003$ **(d)** TV-L^1, $\alpha = 0.641$ **(e)** Interval, $\beta = 0.233$

Figure 2. Reconstruction of the 'cameraman' image (optimised for PSNR) with Rician noise and $\sigma = 30$. The colours in the interval-based reconstruction are closer to the original while ROF and TV-L^1 produce 'overexposed' reconstructions.

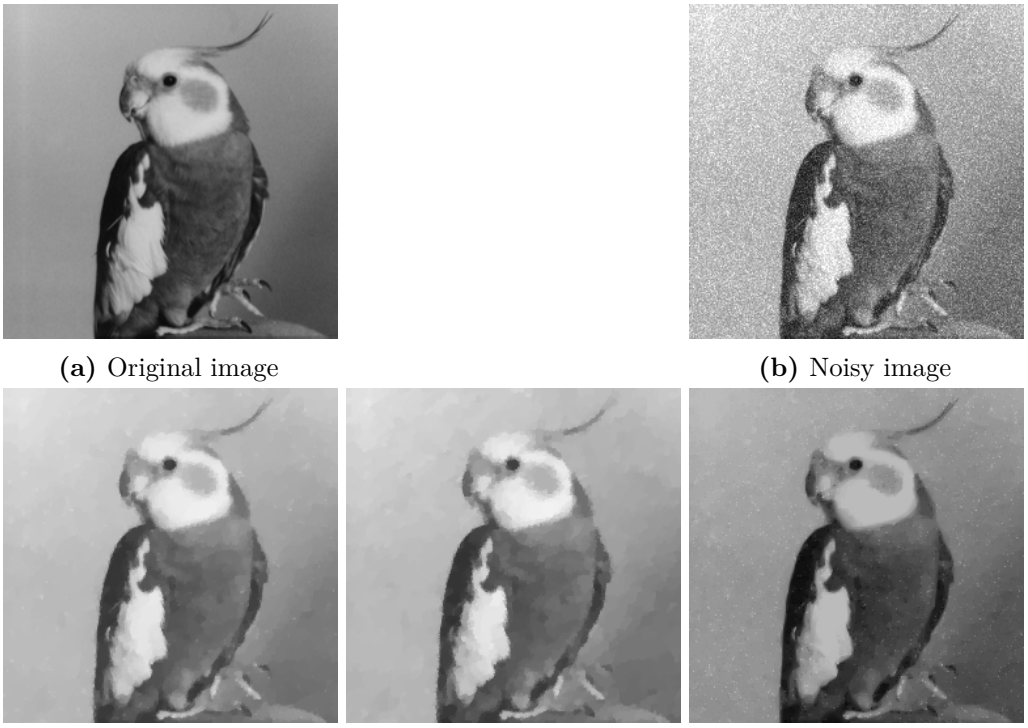
until they are completely removed. This effect has been extensively studied in [16].

Let us now examine the noise removal characteristics of the three optimisation models. Looking, for instance, at Figure 2 we see that all three methods are able to remove most of the noise. The nature of the remaining noise, however, is different for the three models. In the ROF reconstruction noise appears as small blocky artefacts while the TV-L^1 model condenses the noise into bigger patches of the same colour. In the interval-based reconstructions we observe artefacts in the form of isolated pixels that differ from the background as anticipated in Section 3.5.

Quality metrics

Looking at the results in the Tables 1 to 4 we observe that the reconstructions with the interval-based model usually have a significantly higher PSNR than the ROF and TV-L^1 model whereas the SSIM is only slightly better and sometimes even slightly worse.

This can be explained by the previously mentioned effect that the ROF and TV-L^1 reconstructions are shifted towards higher intensities, because they assume zero-mean noise. Pixel-wise differences in intensity are the only source of information used by the PSNR, whilst for the SSIM this is only one aspect of closeness, along with structural information like the local covariance. This explains why the effect of incorrect brightness



(a) Original image **(b)** Noisy image

(c) ROF **(d)** TV-L^1 **(e)** Interval

Figure 3. Reconstruction of the 'bird' image (optimised for PSNR) with Rician noise and $\sigma = 30$. The staircasing effect of the total variation regulariser is apparent in all three reconstructions.

is less important for SSIM than PSNR.

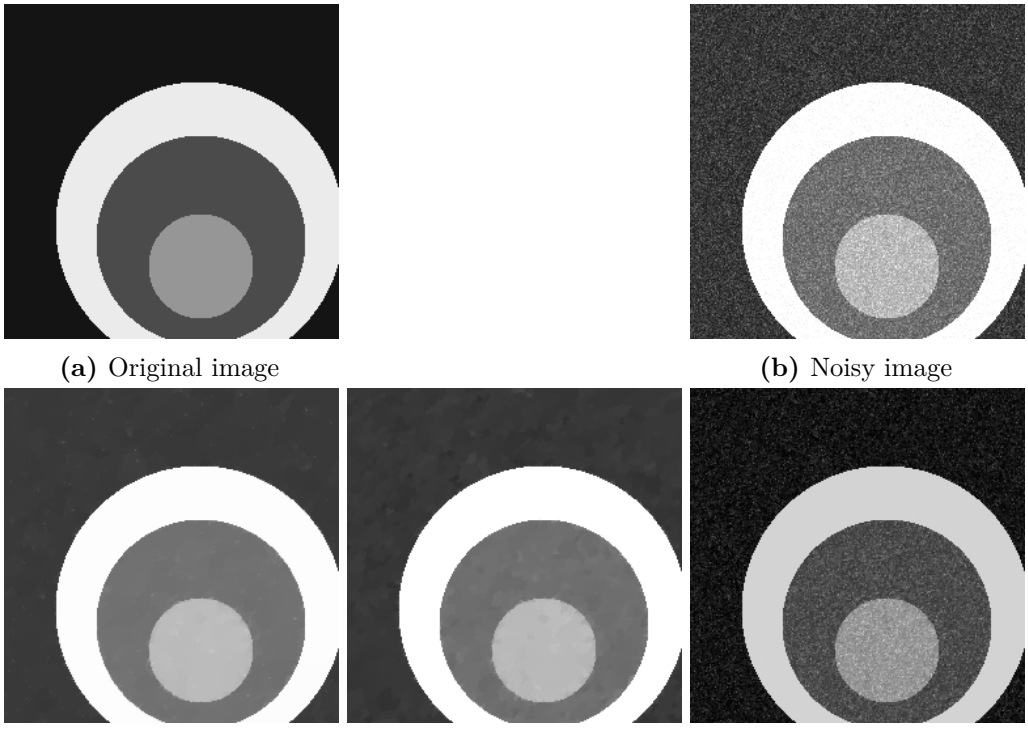
The artefacts mentioned earlier also have an effect on the SSIM values of the interval-based reconstructions. Introducing 'structure' at locations where the ground truth is homogeneous, they have an effect on the local covariance and result in lower SSIM values.

In reconstructions with high noise level the interval-based model yields a much better SSIM (when optimised for SSIM) than the ROF and TV-L^1 model on synthetic images. This effect is not so significant in reconstructions with a small noise level, suggesting that the interval-based model is more robust with respect to the increasing noise level.

Information loss due to thresholding

Let us have a close look at the 'circles' example (Figures 4 and 5). In Table 3 we see that the SSIM of the PSNR-optimised interval-based reconstruction of 'circles' is much smaller compared to the other images, while in the SSIM-optimised reconstruction the PSNR is very low (Table 4). Looking at the corresponding reconstructed images in Figure 4e (PSNR-optimised) and Figure 5e (SSIM-optimised) we see that in the former image the reconstruction is still very noisy but the original colours are restored whereas in the latter one the noise has been removed but the bright circles are greyed out.

The reason for such inaccurate reconstruction of bright regions of the image is that in the regions that are bright in the original image adding positive noise leads to pixel



(a) Original image **(b)** Noisy image

(c) ROF, $\alpha = 0.004$ **(d)** TV-L^1, $\alpha = 0.840$ **(e)** Interval, $\beta = 0.688$

Figure 4. Reconstruction of the 'circles' image (optimised for PSNR) with Rician noise and $\sigma = 30$. Optimising the interval-based model for PSNR yields good reconstruction of the colours but leaves a considerable amount of noise.

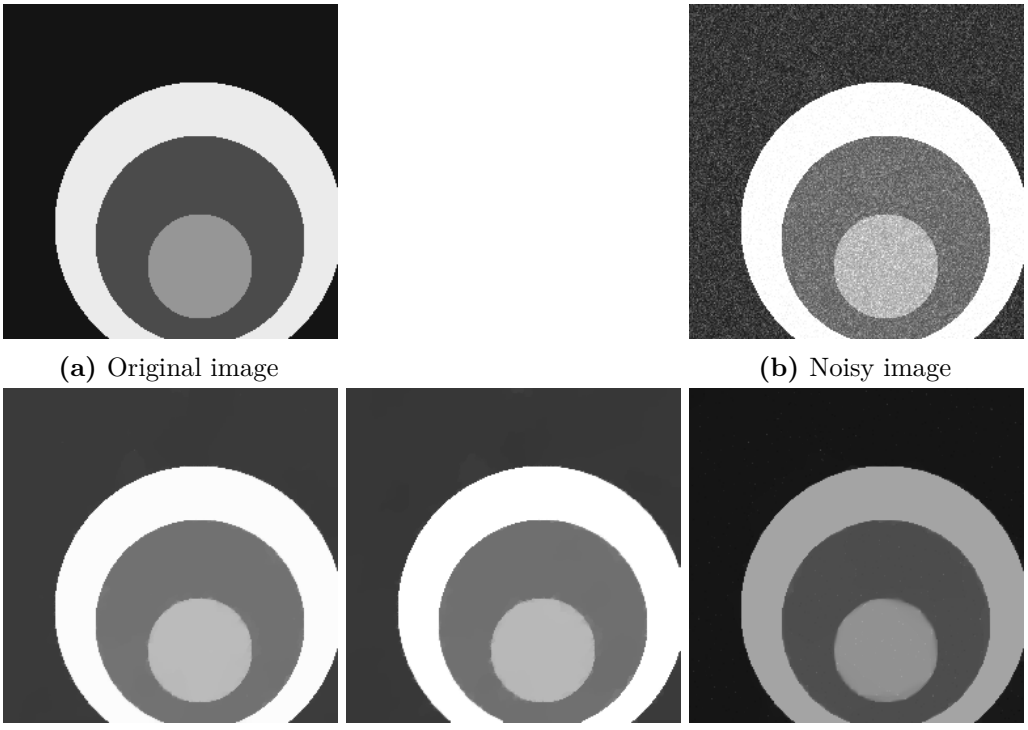
values greater than 255. Since in practice an image sensor cannot register more than 100% intensity we cut off these values at 255. This has an effect on the lower and upper bounds: the noisy pixel value is assumed to be 255 whereas in fact it should be much higher. Thus the bounds are constructed too low, in extreme cases even excluding the original image. This results in the poor reconstruction of bright colours.

Therefore, in regions where the original colours were very bright they can only be reconstructed using narrow intervals, thus leaving much of the noise. To remove the noise we need wider intervals in which the regulariser can identify the solution, which results in a loss of contrast. We will talk more about the choice of the width of the intervals towards the end of this section.

The same effect of reconstructing bright areas too dark can be seen in the reconstruction of the 'lena' image in Figure 6e, where the bright parts of the hat and in the background are somewhat darkened.

Influence of the regulariser

Let us take a closer look at the reconstruction of the 'squares' image (Figure 7). We observe that in the interval-based reconstruction the corners of the squares are rounded off, while in the two other reconstructions they are sharper. Let us analyse why this



(a) Original image **(b)** Noisy image

(c) ROF, $\alpha = 0.006$ **(d)** TV-L^1, $\alpha = 1.296$ **(e)** Interval, $\beta = 0.019$

Figure 5. Reconstruction of the 'circles' image (optimised for SSIM) with Rician noise and $\sigma = 30$. Optimising the interval-based model for SSIM removes the noise but results in poor colour reconstruction in bright regions.

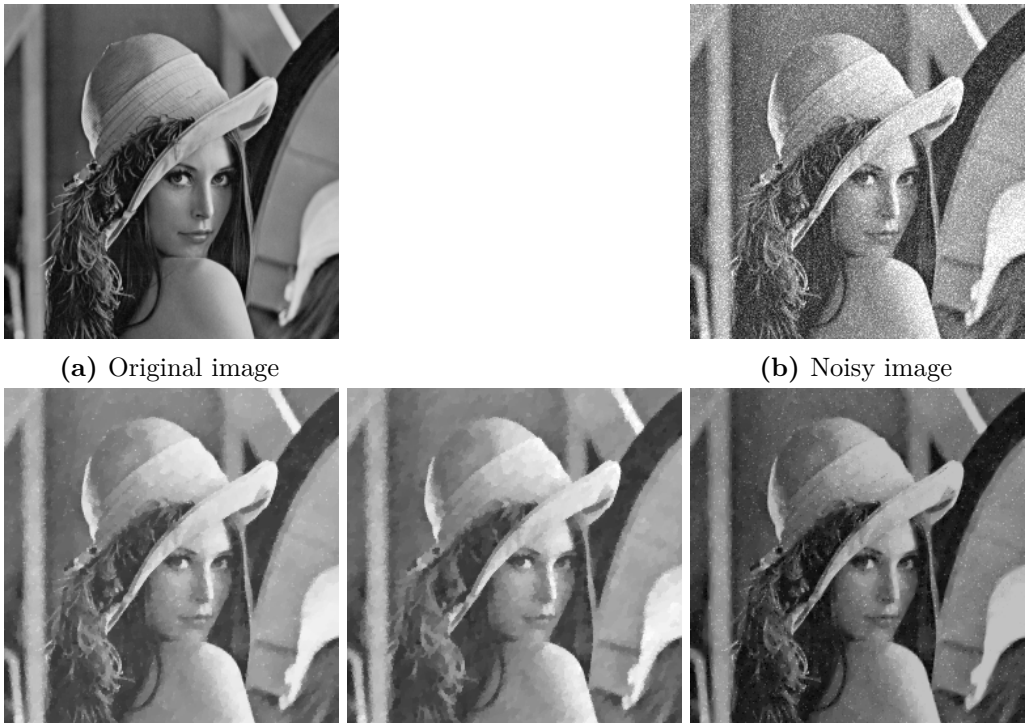
happens.

In Section 2.2 we discussed the similarity between the interval-based method and the classical residual method. The fidelity term in the residual method does not differentiate between the elements in the feasible set, i.e., only the regulariser is responsible for choosing the best element from that set. Therefore, features specific to the regulariser are much more apparent in the residual-type interval-based reconstructions. Since we use isotropic TV, whose characteristic shape is the unit circle [45], the rounding of sharp corners is natural.

Optimal values of the regularisation parameters

In our experiments, optimal values of the regularisation parameter α for the ROF model (3.2) typically are between 0.003 and 0.004 for the real images and slightly higher for the synthetic images that consist mostly of homogeneous regions. For the TV-L^1 model (3.3) the values of α are in the range of 0.4 to 1.0 for the real images and 0.7 to 4.6 for the synthetic images.

The optimal values of the quantile parameter β for the interval-based model (3.1) vary quite significantly. Note that the quantile parameter is a priori confined to the interval $[0, 1]$. The optimal values for the real images, which have more fine structure, range from



(a) Original image **(b)** Noisy image

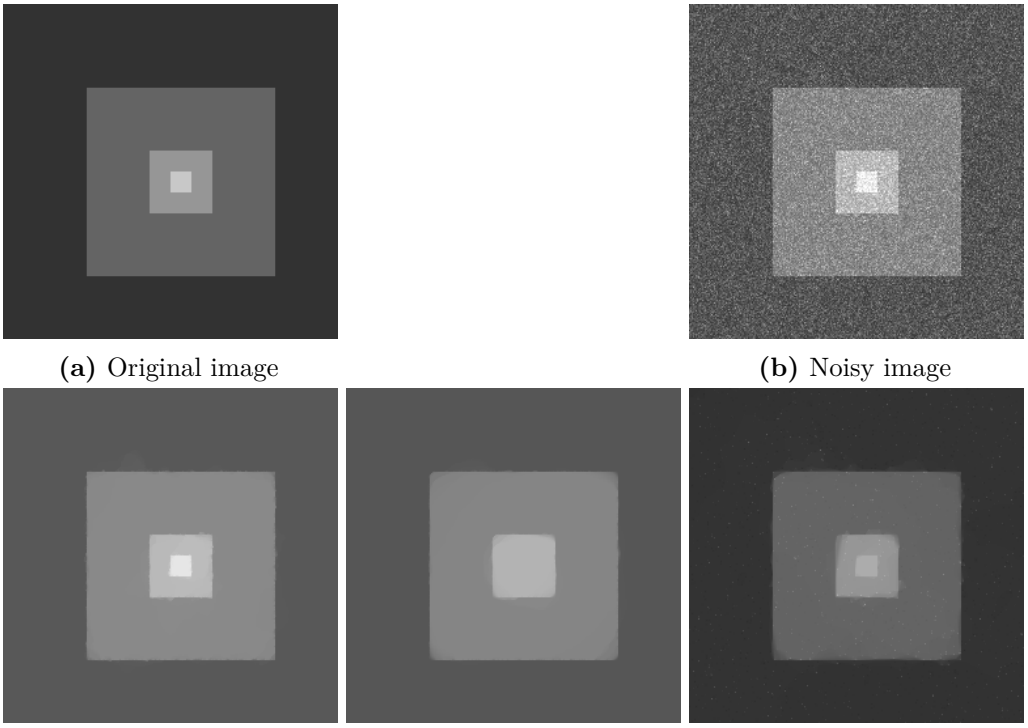
(c) ROF, $\alpha = 0.003$ **(d)** TV-L^1, $\alpha = 0.747$ **(e)** Interval, $\beta = 0.250$

Figure 6. Reconstruction of the 'lena' image (optimised for PSNR) with Rician noise and $\sigma = 30$. Bright spots at the hat and in the background are greyed out in the reconstruction with the interval-based model, but overall the colours are closer to the original than with ROF and TV-L^1.

0.038 to 0.383, whereas the optimal values for the synthetic images are generally smaller, ranging from 0.007 to 0.111. An outlier here is the 'circles' image, which we discussed in the previous paragraph, with an optimal value of 0.688 for the PSNR-optimised reconstruction with $\sigma = 30$ (Figure 4e) and 0.019 for the SSIM-optimised reconstruction with $\sigma = 30$ (Figure 5e).

In Figure 8 we show plots of the PSNR and SSIM in dependence of the quantile parameter for two images: 'lena' and 'circles'. We omit the other plots as they look very similar and don't offer additional insight. It seems as if the graphs always have a unique maximum although we couldn't prove such a relationship. Also we can't predict where this maximum will occur, as it varies from image to image. We included the graphs of the 'circles' image (Figures 8c and 8d) because we can see clearly the aforementioned effect, that the two quality metrics prefer opposite values of the quantile parameter due to the thresholding at high values.

We also did not find any clear trend for a dependence of the quantile parameter β on the noise level σ. However, it appears that for the real images generally the optimal quantile parameter is lower as the noise level increases. For the synthetic images, that are close to the kernel of the regulariser, no such effect is observed.



(a) Original image **(b)** Noisy image

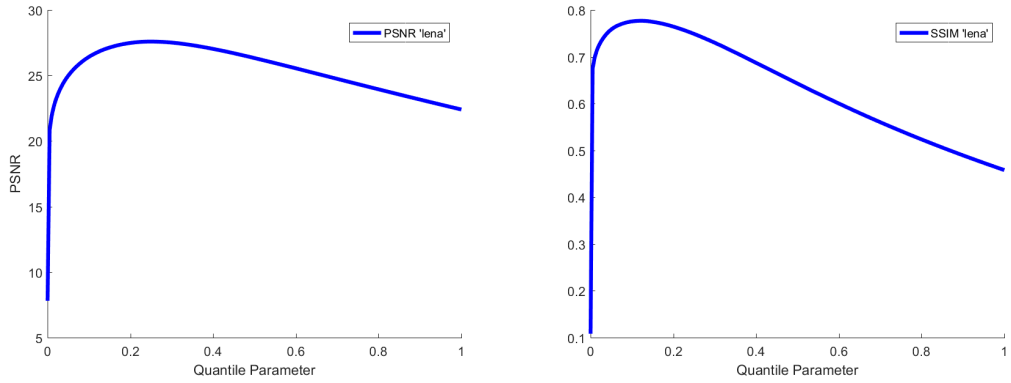
(c) ROF, $\alpha = 0.007$ **(d)** TV-L^1, $\alpha = 4.652$ **(e)** Interval, $\beta = 0.045$

Figure 7. Reconstruction of the 'squares' image (optimised for PSNR) with Rician noise and $\sigma = 30$. The middle square is removed completely with the TV-L^1 model. In the interval-based reconstruction the corners of the squares are rounded off. The contrast looks better in the interval-based reconstruction.

For PSNR-optimised reconstructions the optimal values of the quantile parameter β are typically higher than for SSIM-optimised reconstructions, while the optimal regularisation parameter α is lower when optimising for PSNR. This indicates that the SSIM prefers more regularised images.

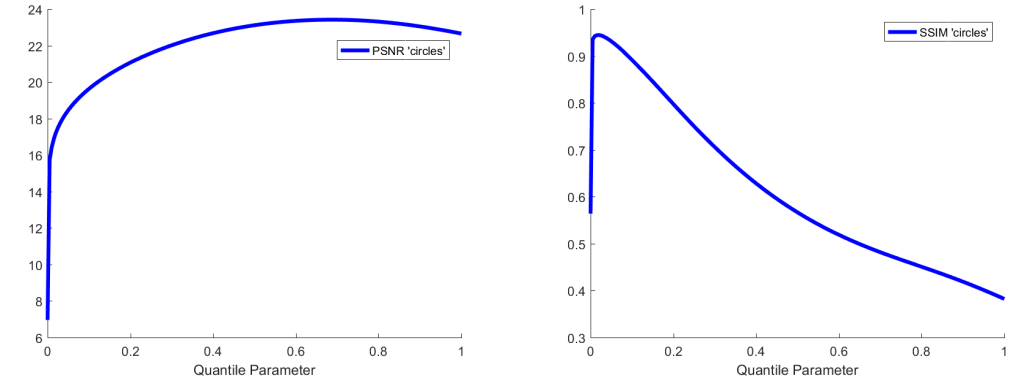
We can see that the quantile parameter β plays the role of the regularisation parameter in the interval-based model. A low value corresponds to wide intervals $[\varepsilon^l, \varepsilon^u]$ and $[f^l, f^u]$ and thus gives more choice for the regulariser, resulting in more regularisation. Looking at the limiting case $\beta \to 0$, the intervals become infinitely wide and the regulariser will be minimised over the whole space $\mathbb{R}^{nm}$. The reconstruction will then be an element from the kernel of the regulariser since no other restrictions apply.

A high value of the quantile parameter on the other hand corresponds to a narrow interval $[f^l, f^u]$, therefore restricting the influence of the regulariser and ensuring that the reconstruction does not depart too much from the measurement. In the limit $\beta \to 1$ the interval $[\varepsilon^l, \varepsilon^u]$ converges to the median of the noise distribution and thus the interval $[f^l, f^u]$ converges to the noisy data minus the median of the noise. Therefore the feasible set consists of only one element that is purely determined by the noisy measurement. Since there is only one element in the feasible set, the regulariser has no choice and the



(a) Example PSNR(β) curve for 'lena' with Rician noise and $\sigma = 30$.

(b) Example SSIM(β) curve for 'lena' with Rician noise and $\sigma = 30$.



(c) Example PSNR(β) curve for 'circles' with Rician noise and $\sigma = 30$.

(d) Example SSIM(β) curve for 'circles' with Rician noise and $\sigma = 30$.

Figure 8. Example curves showing the dependence of PSNR and SSIM on the quantile parameter. We can see that each curve has a unique maximum.

reconstruction will be the noisy input data minus the median of the noise.

Chapter 4: Conclusions and Outlook

We studied the interval-based regularisation model for inverse problems with applications in image reconstruction and denoising. The main advantage of this approach is that it does not make any assumptions about the noise distribution in the data. This is particularly important if the noise distribution is unknown or far from easy-to-model distributions such as the Gaussian distribution. In our experiments with additive Rician noise we observed that the interval-based model yields reconstructions with significantly improved PSNR, while the improvement in SSIM was only significant at high noise levels in synthetic images that are close to the kernel of the regulariser.

Furthermore we examined the use of the quantile parameter as a regularisation parameter. By determining the widths of the intervals it is able to balance the influence of the regulariser against the influence of the fidelity term. We can also use the quantile parameter to control the amount of artefacts in the reconstruction. These appear as isolated pixels in contrast to, e.g., the ROF model, where the noise tends to form larger block-like structures. Since the artefacts in the interval-based model are caused by the hard constraints in the model formulation, we expect that the number of artefacts can be reduced by a relaxation that allows pixels to be outside the interval but introduce a penalty based on how far they exceed the bounds. We expect this also to increase the SSIM of the reconstructions.

Another feature of the interval-based model is that the characteristics of the regulariser are more pronounced in the reconstructions. This effect occurs because the regulariser has more freedom in the residual-type interval-based model than in methods based on a weighted sum of a fidelity term and a regulariser.

We think the interval-based model should be studied further. As mentioned above a relaxation of the hard constraints has the potential to improve the reconstruction quality. The model is also capable of dealing with uncertainty in the operator, which is an aspect that we did not study in this work. This is for example important if we don't know the exact mathematical model that describes a physical relation or some model-related parameters are only known approximately. Another open question is the automatic choice of the quantile parameter. We would like to have a method to predict the optimal or a near-optimal value of the quantile parameter from knowledge about the structure of the image or the noise distribution. One possible approach could be to develop an analogue to the discrepancy principle for the interval-based model.

References

[1] Y. A. Abramovich and C. D. Aliprantis. *An invitation to operator theory*, volume 1. American Mathematical Soc., 2002.

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

[3] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems*, volume 254 of *Oxford Mathematical Monographs*. Clarendon Press Oxford, 2000.

[4] E. D. Andersen, C. Roos, and T. Terlaky. On implementing a primal-dual interior-point method for conic quadratic optimization. *Mathematical Programming*, 95(2):249–277, 2003.

[5] A. Bakushinskii. Remarks on choosing a regularization parameter using the quasi-optimality and ratio criterion. *USSR Computational Mathematics and Mathematical Physics*, 24(4):181 – 182, 1984.

[6] M. A. Bernstein, D. M. Thomasson, and W. H. Perman. Improved detectability in low signal-to-noise ratio magnetic resonance images by means of a phase-corrected real reconstruction. *Medical Physics*, 16(5):813–817, 1989.

[7] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[8] K. Bredies and M. Holler. A total variation-based jpeg decompression model. *SIAM Journal on Imaging Sciences*, 5(1):366–393, 2012.

[9] K. Bredies and M. Holler. Regularization of linear inverse problems with total generalized variation. *Journal of Inverse and Ill-posed Problems*, 22(6):871–913, 2014.

[10] K. Bredies and M. Holler. A TGV-based framework for variational image decompression, zooming, and reconstruction. part I: Analytics. *SIAM Journal on Imaging Sciences*, 8(4): 2814–2850, 2015.

[11] K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526, 2010.

[12] M. Burger and F. Lucka. Maximum a posteriori estimates in linear inverse problems with log-concave priors are proper bayes estimators. *Inverse Problems*, 30(11):114004, 2014.

[13] M. Burger, K. Papafitsoros, E. Papoutsellis, and C.-B. Schönlieb. Infimal convolution regularisation functionals of bv and spaces: Part I: The finite p case. *Journal of Mathematical Imaging and Vision*, 55:343, 2016.

[14] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(2):167–188, 1997.

[15] T. Chan, A. Marquina, and P. Mulet. High-order total variation-based image restoration. *SIAM Journal on Scientific Computing*, 22(2):503–516, 2000.

[16] T. F. Chan and S. Esedoglu. Aspects of total variation regularized L1 function approximation. *SIAM Journal on Applied Mathematics*, 65(5):1817–1837, 2005.

[17] T. F. Chan and J. Shen. *Image processing and analysis: variational, PDE, wavelet, and stochastic methods*. SIAM, 2005.

[18] L. Chen and T. Zeng. A convex variational model for restoring blurred images with rician noise. *Journal of Mathematical Imaging and Vision*, 53(1):92–111, Sep 2015.

[19] P. L. Combettes and J.-C. Pesquet. Image restoration subject to a total variation constraint. *IEEE transactions on image processing*, 13(9):1213–1222, 2004.

[20] D. C. Dobson and F. Santosa. Recovery of blocky images from noisy and blurred data. *SIAM Journal on Applied Mathematics*, 56(4):1181–1198, 1996.

[21] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.

[22] P. Getreuer, M. Tong, and L. A. Vese. A variational model for the restoration of mr images corrupted by blur and rician noise. In *International Symposium on Visual Computing*, pages 686–698. Springer, 2011.

[23] E. Giusti. Minimal surfaces and functions of bounded variation. *Monogr. Math.*, 80, 1984.

[24] D. Goldfarb and W. Yin. Second-order cone programming methods for total variation-based image restoration. *SIAM Journal on Scientific Computing*, 27(2):622–645, 2005.

[25] A. Gorokh, Y. Korolev, and T. Valkonen. Diffusion tensor imaging with deterministic error bounds. *Journal of Mathematical Imaging and Vision*, 56(1):137–157, 2016.

[26] M. Grant and S. Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008.

[27] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1, Mar. 2014. URL `http://cvxr.com/cvx`.

[28] M. Grant, S. Boyd, and Y. Ye. Disciplined convex programming. In *Global optimization*, pages 155–210. Springer, 2006.

[29] H. Gudbjartsson and S. Patz. The rician distribution of noisy mri data. *Magnetic resonance in medicine*, 34(6):910–914, 1995.

[30] J. Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton university bulletin*, pages 49–52, 1902.

[31] J. Hadamard. *Lectures on Cauchy's problem in linear partial differential equations*. Courier Corporation, 2014.

[32] P. C. Hansen. The truncated SVD as a method for regularization. *BIT Numerical Mathematics*, 27(4):534–553, 1987.

[33] V. K. Ivanov, V. V. Vasin, and V. P. Tanana. *Theory of linear ill-posed problems and its applications*, volume 36. Walter de Gruyter, 2002.

[34] J. Kaipio and E. Somersalo. *Statistical and computational inverse problems*, volume 160. Springer Science & Business Media, 2005.

[35] N. Karmarkar. A new polynomial-time algorithm for linear programming. In *Proceedings of the sixteenth annual ACM symposium on Theory of computing*, pages 302–311. ACM, 1984.

[36] J. B. Keller. Inverse problems. *The American Mathematical Monthly*, 83(2):107–118, 1976.

[37] Y. Korolev. Making use of a partial order in solving inverse problems: II. *Inverse Problems*, 30(8):085003, 2014.

[38] Y. Korolev and A. Yagola. Making use of a partial order in solving inverse problems. *Inverse Problems*, 29(9):095012, 2013.

[39] T. Le, R. Chartrand, and T. J. Asaki. A variational approach to reconstructing images corrupted by poisson noise. *Journal of mathematical imaging and vision*, 27(3):257–263, 2007.

[40] MOSEK ApS. The MOSEK optimization software, Aug. 2017. URL `https://www.mosek.com`.

[41] Y. Nesterov and A. Nemirovskii. *Interior-point Polynomial Algorithms in Convex Programming*. Studies in Applied Mathematics. Society for Industrial and Applied Mathematics, 1994.

[42] M. Nikolova. Local strong homogeneity of a regularized estimator. *SIAM Journal on Applied Mathematics*, 61(2):633–658, 2000.

[43] M. Nikolova. A variational approach to remove outliers and impulse noise. *Journal of Mathematical Imaging and Vision*, 20(1):99–120, 2004.

[44] M. Nikolova. Weakly constrained minimization: application to the estimation of images and signals involving constant regions. *Journal of Mathematical Imaging and Vision*, 21(2): 155–175, 2004.

[45] S. Osher and S. Esedoḡlu. Decomposition of images by the anisotropic rudin-osher-fatemi model. *Communications on pure and applied mathematics*, 57(12):1609–1626, 2004.

[46] W. Ring. Structural properties of solutions to total variation regularization problems. *ESAIM: Mathematical Modelling and Numerical Analysis*, 34(4):799–810, 2000.

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

[48] O. Scherzer. Denoising with higher order derivatives of bounded variation and an application to parameter estimation. *Computing*, 60(1):1–27, Mar 1998.

[49] S. M. Spann, K. S. Kazimierski, C. S. Aigner, M. Kraiger, K. Bredies, and R. Stollberger. Spatio-temporal TGV denoising for ASL perfusion imaging. *NeuroImage*, 2017.

[50] J. F. Sturm. Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones. *Optimization methods and software*, 11(1-4):625–653, 1999.

[51] K.-C. Toh, M. J. Todd, and R. H. Tütüncü. SDPT3—a Matlab software package for semidefinite programming, version 1.3. *Optimization methods and software*, 11(1-4):545–581, 1999.

[52] T. Valkonen, K. Bredies, and F. Knoll. TGV for diffusion tensors: A comparison of fidelity functions. *Journal of Inverse and Ill-Posed Problems*, 21(3):355–377, 2013.

[53] T. Valkonen, K. Bredies, and F. Knoll. Total generalized variation in diffusion tensor imaging. *SIAM Journal on Imaging Sciences*, 6(1):487–525, 2013.

[54] Z. Wang and A. C. Bovik. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE signal processing magazine*, 26(1):98–117, 2009.

[55] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600–612, 2004.