UNIVERSITY OF TWENTE.

**Faculty of Electrical Engineering,
Mathematics & Computer Science**

Higher Order Variational Methods for Photoacoustic Tomography

Marinus Jan Lagerwerf
M.Sc. Thesis
September 2015

Assessment committee:
prof. dr. ir. S.A. van Gils
dr. ir. C. Brune
dr. ir. S. Manohar
dr. ir. J.W. Polderman

Supervisors:
dr. ir. C. Brune
dr. ir. S. Manohar

Internship:
prof. dr. ir. C.-B. Schönlieb (Cambridge)

Applied Analysis & Mathmatical Physics
Applied Mathmatics
Faculty of Electrical Engineering,
Mathematics and Computer Science
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

Acknowledgements

This thesis presents the work I have done at the Applied Analysis research group in close collaboration with the Biomedical Photonic Imaging research group both at the University of Twente. In this time I worked with various people who contributed their time and effort to my research. For which I am grateful to all. However I want to thank some of them specifically.

Christoph Brune, for being my supervisor, during my final project and my internship in Cambridge, and pushing me to get the most out of my work. Without him this thesis could not have been completed in its current form.

Srirang Manohar, for bringing me into his photoacoustic tomography project, always being enthusiastic about what I did and asking the right questions.

Carola-Bibliane Schönlieb and Martin Benning, my internship supervisors in Cambridge, for the fun meetings, the great time in Cambridge and motivating me to pursue doing research in the imaging field.

Peter van Es, for always taking time to answer my questions and sparring about my ideas.

Loes Knoben, Yoeri Boink and Joep Peeters, for being my rubber ducks in time of need.

Finally all my friends and family for their continued support, not only during my thesis, but during the entirity of my education.

Abstract

The goal of this work is to model, implement and test higher order variational methods for photoacoustic tomography. Photoacoustic tomography is a novel imaging method, which is used in breast cancer and rheumatism diagnosis. The challenges in this tomography problem are the robust of handling noise and subsampling on the data and reconstructing difficiult data structures. To addres those the variational methods using Total Variation and Total Generalized Variation with Bregman iteration are implemented. Moreover, a preconditioned Primal-Dual Hybrid Gradient algorithm is derived to effeciently solve the non-smooth convex minimization problems introduced by these reconstruction models. To show the effectiveness of these methods a careful study is done on both synthetic and experimental data.

Keywords: photoacoustic tomography, inverse problems, variational methods, total variation, total veneralized variation, bregman iteration, preconditioned primal-dual hybrid gradient algorithm.

Figure 1: Photoacoustic tomography system
This figure shows the photoacoustic tomography system doing a measurement on a human finger. A schematic version of this setup is later shown in **Figure 3** *&* **Figure 6**.

1 Introduction

Accurate image reconstruction is a fundamental challenge in many scientific fields, such as biomedicine and microscopy. The challenges with reconstructed images are that they can easily suffer from artefacts due to noise, measurement designs (e.g. sparsity, sampling), or not accurate use of prior information. Photoacoustic tomography is a novel image reconstruction method, with these kind of challenges.

Photoacoustic tomography is a method which combines optical and acoustical properties of an object to image the optical absorption distribution inside this object. To get the optical absorption distribution the object is illuminated with a lightsource, creating a distribution of optical energy in this object (see **Figure 1**). This optical energy is transformed into outward traveling pressure waves (ultrasound waves) via fast heat release and thermal expansion, this physical process is called the photoacoustic effect. The strength of this method is that it works around the weaknesses of photonic imaging, *i.e.* relative high scattering through soft tissue, and ultrasound imaging, *i.e.* low penetration through small soft tissue. This offers usefull applications in diagnosis methods, such as breast cancer diagnosis and rheaumatism diagnosis. In rheumatism diagnosis one wants to detect small inflammations around finger joints. Some initial research on this topic is done in [WCJ07] and [vEBM+14]. In **Figure 2** one of the results from [vEBM+14] is shown. This is a healthy finger, shown from sideview and certain slices around the joints. The challenges for this problem are readily seen in the these slice views of the finger, *i.e.* the data has sparse structures, low contrast and noise in the background. Moreover the streaking artefacts show that there is also a challenge in the subsampling of the data. This shows the strong need for an image reconstruction method which is robust against noise and subsampling and can handle difficult data structures. From the theory it is known that variational methods such as Total Variation (TV) regularization and the higher order version, Total Generalized Variation (TGV) regularization, with certain extensions can handle such challenges. The goal of this thesis is the implementation and testing of the image reconstruction models using TV and TGV for the photoacoustic image reconstruction problem. For the implementation of these models an algorithm to solve a non-smooth minimization problem is needed, the proposed algorithm in this thesis is a Primal-Dual Hybrid Gradient algorithm, which is chosen for its general applicability and efficient update rules. For the testing of the models synthetic and experimental data is used. These data sets are specifically chosen to show how the reconstruction models handle challenges such as noise, subsampling and difficult data structures.

This thesis will first focus on introducing the forward model for photoacoustic tomography, this is done in **Chapter 2**. Next in **Chapter 3** this forward model is translated to a mathematical problem and the mathmatical models to solve this problem are introduced. To implement these mathmatical models there is an algorithm needed, in **Chapter 4** an algorithm is introduced and validated. Using this algorithm the results for the reconstruction models are shown in **Chapter 5**. The thesis concludes with a summary and the conclusions of the research, in **Chapter 6**, and an outlook on future research, in **Chapter 7**.
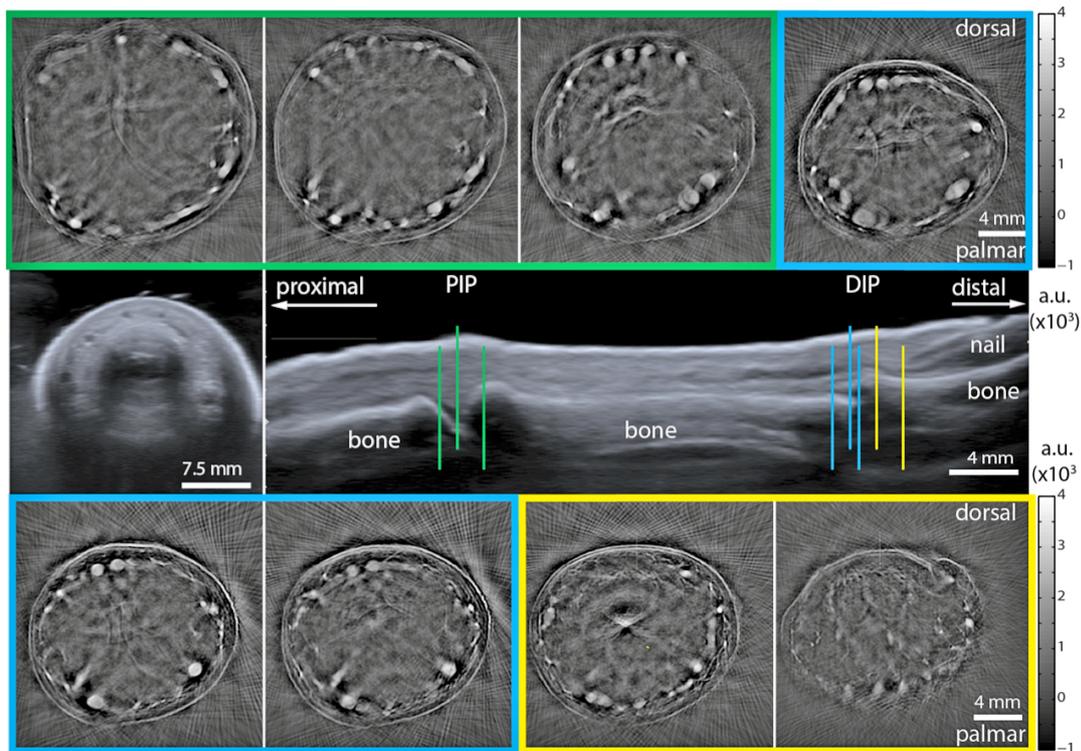

Figure 2: Initial imaging results on an healthy finger.
The two figures on the second row are respectively a transversal and longitudinal ultrasound reconstruction of the finger. Whereas the slices of the finger are photoacoustic image reconstructions, made around the joints and the nail of the finger.
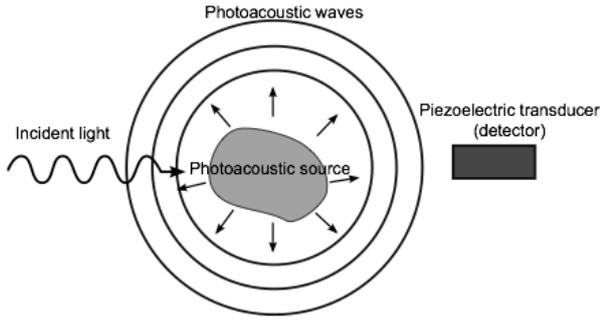
Figure 3: Schematic visualization of the photoacoustic effect
By illuminating a photoacoustic source with a light, the source heats up. This process can be modelled with a heat conduction equation (see (1)). This heating results in thermal expansion creating outward traveling photoacoustic wave, for which an explicit relation is shown in (8). These waves are detected by detectors, resulting in a measurement.

2 Photoacoustic tomography

As said before photoacoustic tomography is based on the photoacoustic effect, which can be defined as the conversion of absorbed optical energy to acoustic energy. Many physical effects influence this conversion, but the most efficient conversion, therefore most influential, is the thermoelastic effect. The thermal expansion, caused by absorption of optical energy, generates an initial pressure distribution, which results in outward traveling pressure waves (acoustic energy). A schematic visualization of this process is shown in **Figure 3**. Using this knowledge a forward model is created, which describes the physical process of the photoacoustic image reconstruction.

2.1 Forward model for photoacoustic tomography

The relation between the heating through absorption of optical energy and the temperature distribution is modeled by the heat conduction equation,

$$\rho C_p \frac{\partial T(\mathbf{r},t)}{\partial t} = \lambda T(\mathbf{r},t) + H(\mathbf{r},t), \tag{1}$$

where $T(\mathbf{r},t)$ [K] and $H(\mathbf{r},t)$ [J/(m^{3}s)] are respectively the temperature distribution and the heating function which are spatially and time dependent, ρ [kg/m^3] is the density of the material, C_p [J/(K kg)] is the specific heat and λ is the thermal conductivity. Assuming the pulse of optical energy is small enough (in the temporal sense) the heat diffusion can be neglected, resulting in

$$\rho C_p \frac{\partial T(\mathbf{r},t)}{\partial t} = H(\mathbf{r},t). \tag{2}$$

This assumption of 'short enough' can be quantified in the condition of thermal and stress confinement (see [WW12]) and can easily be met in practice.
Furthermore the excess temperature distribution can be related to the acoustic pressure $p(\mathbf{r},t)$,

$$\nabla^2 p(\mathbf{r},t) - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} p(\mathbf{r},t) = -\beta\rho\frac{\partial^2}{\partial t^2} T(\mathbf{r},t), \tag{3}$$

where c [m/s] is the speed of sound and β [K^{-1}] is the volume thermal expansion coefficient. Combining these two relations and considering that the heating function can be seen as a product of a spatially dependent optical absorption distribution of the heat, $A(\mathbf{r})$, and a time profile of the illuminating optical energy source, $I(t)$, one arrives at the following equation:

$$\nabla^2 p(\mathbf{r},t) - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} p(\mathbf{r},t) = -\frac{\beta}{C_p}\frac{\partial I(t)}{\partial t} A(\mathbf{r}). \tag{4}$$

3

This can be seen as a wave equation with a source term and can be solved with the aid of Green's function (see [KLA$^+$95]),

$$p(\mathbf{r}, t) = \frac{\beta}{C_p} \iint_{\mathbb{R}^d} A(\mathbf{r}') \frac{\partial I(t')}{\partial t'} \frac{\delta(t - \frac{1}{c} \|\mathbf{r} - \mathbf{r}'\| - t')}{\|\mathbf{r} - \mathbf{r}'\|} d\mathbf{r}' dt' \tag{5}$$

$$= \frac{\beta}{4\pi C_p} \int_{\mathbb{R}^d} \frac{A(\mathbf{r}')}{\|\mathbf{r} - \mathbf{r}'\|} \frac{\partial}{\partial t} I(t - \frac{1}{c} \|\mathbf{r} - \mathbf{r}'\|) d\mathbf{r}'. \tag{6}$$

Changing to spherical coordinates and $t' = \frac{\|\mathbf{r} - \mathbf{r}'\|}{c}$ gives

$$p(\mathbf{r}, t) = \frac{\beta}{4\pi C_p} \int \left(\frac{1}{t'} \iint A(ct', \theta, \phi)(ct')^2 \sin(\theta) \, d\theta \, d\phi \right) \frac{\partial I(t - t')}{\partial t} dt' \tag{7}$$

$$= \frac{\beta}{4\pi C_p} \left(\frac{1}{t} \iint_{\|\mathbf{r} - \mathbf{r}'\| = ct} A(\mathbf{r}') d\mathbf{r}' \right) *_t \frac{\partial I(t)}{\partial t}. \tag{8}$$

To measure the generated pressure wave, consider an ultrasound transducer at a certain position $\mathbf{r}$. The measured signal can be expressed as:

$$\tilde{p}(\mathbf{r}, t) = p(\mathbf{r}, t) * h_{IR}(\mathbf{r}, t), \tag{9}$$

where $h_{IR}(\mathbf{r}, t)$ is the impulse response of the transducer. To simplify the model the diffraction effects are ignored, This simplification leads to a pure time dependent impulse response of the transducer. Now one can reconstruct the optical absorption distribution $A(\mathbf{r})$ from the measurements $\tilde{p}(\mathbf{r}, t)$, given there is a description of the illumination profile $I(t)$ and the impulse response $h_{IR}(t)$. In practice these descriptions are difficult to measure. To solve this, a practical solution is proposed in [WXZC04]. The idea is to consider an optical absorption distribution which one has control over and derive an expression for $\frac{\partial I(t)}{\partial t} *_t h_{IR}(t)$. Take a photoacoustic point source at $\mathbf{r_p}$ this will result in a simple absorption distribution, $A(\mathbf{r}) = \delta(\mathbf{r} - \mathbf{r_p})$. Plugging this in (8) gives:

$$p_\delta(\mathbf{r}, t) = \frac{\beta}{4\pi C_p} \int_{\mathbb{R}^d} \frac{\delta(\mathbf{r}' - \mathbf{r_p})}{\|\mathbf{r} - \mathbf{r}'\|} \frac{\partial}{\partial t} I(t - \frac{1}{c} \|\mathbf{r} - \mathbf{r}'\|) d\mathbf{r}', \tag{10}$$

$$= \frac{\beta}{4\pi C_p \|\mathbf{r} - \mathbf{r_p}\|} \frac{\partial}{\partial t} I(t - \frac{1}{c} \|\mathbf{r} - \mathbf{r_p}\|), \tag{11}$$

rearranging this expression,

$$\frac{\partial I(t)}{\partial t} = \frac{4\pi C_p \|\mathbf{r} - \mathbf{r_p}\|}{\beta} p_\delta(\mathbf{r}, t - \frac{1}{c} \|\mathbf{r} - \mathbf{r_p}\|), \tag{12}$$

convolving both sides with $h_{IR}(t)$ and using (6),

$$\frac{\partial I(t)}{\partial t} *_t h_{IR}(t) = \frac{4\pi \|\mathbf{r} - \mathbf{r_p}\| C_p}{\beta} \tilde{p}_\delta(\mathbf{r}, t - \frac{1}{c} \|\mathbf{r} - \mathbf{r_p}\|). \tag{13}$$

Combining (8), (6) and (13) gives

$$\tilde{p}(\mathbf{r}, t) = \|\mathbf{r} - \mathbf{r_p}\| \left(\frac{1}{t} \iint_{\|\mathbf{r} - \mathbf{r}'\| = ct} A(\mathbf{r}') d\mathbf{r}' \right) *_t \tilde{p}_\delta(\mathbf{r}, t'), \tag{14}$$

where $t' = t - \frac{1}{c} \|\mathbf{r} - \mathbf{r_p}\|$. Since in this thesis only the 2-d case is considered the physical model symplifies to

$$\tilde{p}(\mathbf{r}, t) = \|\mathbf{r} - \mathbf{r_p}\| \left(\frac{1}{t} \int_{\|\mathbf{r} - \mathbf{r}'\| = ct} A(\mathbf{r}') d\mathbf{r}' \right) *_t \tilde{p}_\delta(\mathbf{r}, t') \tag{15}$$

This physical model is the basis for photoacoustic image reconstruction.

3 Modeling

In this chapter the mathematical modeling of the photoacoustic tomography problem is adressed. First the mathematical structure of the problem is studied and how this problem will be treated. Next the current methods are introduced and at last the new modeling suggestion is introduced.

3.1 Coupled problem

Most image reconstruction problems can be described in the form of an inverse problem, $Ku = f$, where K is the forward operator defined through the physical process, u is the desired image and f is the measured data. To represent the image reconstruction problem in this way there has to be an efficient way to describe K. In theory one can just concatenate the operations, in practice this concatenation results in a nonsparse linear operator $K \in \mathbb{R}^{n_{mp} \times n_p}$ (where n_{mp} is the number of measurement points and n_p is the number of pixels in the reconstructed image), which is far to big for practical use. Therefore the problem is split into a multistage inverse problem.

$$K_2 K_1 u = f, \tag{16}$$

where

$$K_1 x \sim \frac{\|\mathbf{r} - \mathbf{r_p}\|}{t} \int\limits_{\|\mathbf{r}-\mathbf{r'}\|=ct} x(\mathbf{r'})d\mathbf{r'}, \qquad\qquad K_2 x \sim \tilde{p}_\delta(\mathbf{r},t') *_t x(\mathbf{r},t), \tag{17}$$

$$u \sim A(\mathbf{r}), \qquad\qquad f \sim \tilde{p}(\mathbf{r},t). \tag{18}$$

This modeling structure gives way to two solving procedures. The first would be considering both inverse problems simultateneously. As seen in [BMPS14] and [BBLS15] such a strategy can improve the reconstruction results in certain cases, however in this paper the other approach will be considered. This approach is simply solving the two inverse problems sequantially, *i.e.* first solving the outer problem and plugging that solution in the inner inverse problem

$$K_2 \tilde{f} = f, \qquad\qquad K_1 u = \tilde{f}. \tag{19}$$

By splitting the problems for both inverse problems a model can be designed for both problems separately and if both of these problems can be solved very precisely the reconstruction itself is very precise. However a problem for this solving method is that if one of these two problems cannot be solved precisely the whole reconstruction suffers.

3.2 Direct inversion models

The current image reconstruction model solves the coupled inverse problems in series with two direct methods.
For the outer inverse problem the convolution theorem is used. The convolution theorem reads,

$$f * g = \mathcal{F}^{-1}\left\{\mathcal{F}\{f\} \cdot \mathcal{F}\{g\}\right\}. \tag{20}$$

Applying (20) to (15) gives a direct way of solving the outer problem,

$$\mathcal{F}^{-1}\left\{\frac{\mathcal{F}\{\tilde{p}(\mathbf{r},t)\}}{\mathcal{F}\{\tilde{p}_\delta(\mathbf{r},t')\}}\right\} = \|\mathbf{r} - \mathbf{r_p}\| \frac{1}{t} \int\limits_{\|\mathbf{r}-\mathbf{r'}\|=ct} A(\mathbf{r'})d\mathbf{r'}. \tag{21}$$

However this expression might be undefined when $\mathcal{F}\{\tilde{p}_\delta(\mathbf{r}, t')\}$ vanishes. Moreover any noise on the denominator can have big influence on the value of $\frac{\mathcal{F}\{\tilde{p}(\mathbf{r},t)\}}{\mathcal{F}\{\tilde{p}_\delta(\mathbf{r},t')\}}$ when the denominator is small. Therefore instead of this simple division, the following method is used,

$$\mathcal{F}^{-1}\left\{\frac{\overline{\mathcal{F}\{\tilde{p}_\delta(\mathbf{r},t')\}}\mathcal{F}\{\tilde{p}(\mathbf{r},t)\}}{|\mathcal{F}\{\tilde{p}_\delta(\mathbf{r},t')\}|^2+\epsilon}\right\} \approx \|\mathbf{r}-\mathbf{r_p}\|\frac{1}{t}\int\limits_{\|\mathbf{r}-\mathbf{r'}\|=ct} A(\mathbf{r'})d\mathbf{r'}, \tag{22}$$

where $|\mathcal{F}\{\tilde{p}_\delta(\mathbf{r},t')\}|^2 \gg \epsilon > 0$. In the form of the inverse problem this becomes,

$$K_1 u = K_{deconv}f = \tilde{f}, \tag{23}$$

where $\tilde{f}$ is the input for the inner inverse problem.

For the inner inverse problem the filtered backprojection (FBP) model is used. The idea is to apply the backprojection K_1^T to the measurements. Since the backprojection is not the inverse operator of K_1 there will be artefacts resulting from this process. To limit this artefacts a linear filtering operator K_{filt} is introduced, *i.e.*

$$u = K_1^* K_{\text{filt}}\tilde{f}, \tag{24}$$

where u is then the reconstructed image. Notice that since K_{filt} is a linear filter, any filtering done to remove noise from the data will result in losing significant edges, and the other way around, preserving the edges will result in not handling the noise in any way. Moreover the structure of this direct inversion model does not give enough room for imposing more sophisticated regularization on the reconstruction.

These modeling choices give an easy to implement and fast direct way of reconstruction. However due to the ill-posedness of the inverse problems the simplicity of these direct methods lose quality in their reconstructions. Therefore more precise reconstruction models are proposed in this paper.

3.3 Variational methods and regularizers

In this section the proposed models are discussed. These models focus on solving the inner inverse problem, *i.e.* $K_1 u = \tilde{f}$, more precisely.

Whereas FBP is a direct method, variational methods are indirect approaches. Instead of directly computing the reconstruction, a minimization problem is constructed based on the priors of the problem. The general structure of such a problem is

$$\underset{u}{\operatorname{argmin}}\left\{\lambda D(K_1 u, \tilde{f}) + R(u)\right\}, \tag{25}$$

where $D(K_1 u, \tilde{f})$ is the data fidelity, $R(u)$ is the regularization term and λ is the regularization parameter. The data fidelity represents the distance between the reconstructed image after applying the forward problem and the noisy data. The regularization term is a penalty term which grows when certain prior assumptions are violated. By minimizing these two terms an image u close to the data is reconstructed following the priors imposed through the regularization. Through the choices of priors a variational method can be tailored to fit a certain problem.

A widely used variational method, which is based on the findings in Rudin, Osher and Fatemi [ROF92], is the L^2/TV model, the popularity of this model is due to the fact that it can handle noise and still allows the reconstructed image to have edges. The L^2 in L^2/TV refers to the choice of data fidelity, this data fidelity has as an underlying assumption that the noise on the data is additive Gaussian white noise. For all the variational methods considered in this thesis the L^2 data fidelity will be chosen, therefore the L^2 will be omitted in future references.

3.3.1 Reconstruction model with Total Variation regularization

This reconstruction model is based on the Total Variation regularizer, this regularizer has as underlying assumption that the true image is piecewise constant. As said before these kind of models are defined through a minimization problem. The minimization problem related to this model is:

$$u = \operatorname*{argmin}_{u \in \mathrm{BV}(\Omega)} \left\{ \frac{\lambda}{2} \left\| K_1 u - \tilde{f} \right\|_{L^2(\Omega)}^2 + |u|_{BV(\Omega)} \right\}, \tag{26}$$

where $\mathrm{BV}(\Omega)$ denotes the space of functions with bounded variation on Ω and $|\cdot|_{BV(\Omega)}$ denotes the BV seminorm, formally given by

$$|u|_{BV(\Omega)} = \sup \left\{ \int_\Omega u \operatorname{div} v \, dx \Big| v \in \mathcal{C}_c^1(\Omega, \mathrm{Sym}^1(\mathbb{R}^d)), \|v\|_\infty \le 1 \right\}, \tag{27}$$

When $u \in W^{1,1}$ this formal definition simplifies to

$$|u|_{BV(\Omega)} = \int_\Omega |\nabla u| \, dx, \tag{28}$$

with $|g| = \sqrt{g_1^2 + g_2^2 + ... + g_2^2}$. This $|u|_{BV(\Omega)}$ is sometimes called the *total variation* of u. This last expression (28) for the *total variation* of u indicates how this regularization penalizes non piecewise constant structures, since this term is minimized and therefore the gradient of u is also minimized. Note that a piecewise constant u cannot be an element of $W^{1,1}$, however within $W^{1,1}$ on can get very close to a piecewise constant reconstruction. The technical details around this, the space of BV and the BV seminorm are beyond the scope of this thesis, for further reference one could look at [AV94], [CKP99] and [ROF92].

3.3.2 Reconstruction model with Total Generalized Variation regularization

A more general version of Total Variation is Total Generalized Variation (TGV). In [BKP10] this model was introduced, to also allow higher order regularity in the reconstructed images. The idea of total generalized variation is creating regularization term that penalizes not only the first order derivatives but also higher order derivatives. Looking at the formal definition

$$\mathrm{TGV}_\beta^2(u) = \sup \left\{ \int_\Omega u \operatorname{div}^k v \, dx \Big| v \in \mathcal{C}_c^k(\Omega, \mathrm{Sym}^k(\mathbb{R}^d)), \|\mathrm{div}^l v\|_\infty \le \beta_l, l = 0, ..., k-1 \right\}, \tag{29}$$

one can see that the choice of k influences how much regularity is allowed in the reconstructions and the choice of β_l influences the priority put on a certain regularity level. Note that by taking $k = 1, \beta = 1$ the definition of TGV coincides with the definition of TV. From the set of regularizers defined through this definition, TGV_β^2 is the one used in this thesis, since from practice there is no reason to expect higher order regularity needed than the linear regularity provided in this case. Rewriting the formal definition for $\mathrm{TGV}_\beta^2(u)$ gives

$$\mathrm{TGV}_\beta^2(u) = \min_{w,v \in \mathrm{BV}(\Omega), \nabla u = v + w} \left\{ \int_\Omega |w| \, dx + \beta \int_\Omega |\mathcal{E}(v)| \, dx \right\}, \tag{30}$$

where $|\cdot|$ is defined as done in the previous section and $\mathcal{E}(\cdot)$ is the symmetric gradient. This splitting of $\nabla u = v + w$ shows how TGV_β^2 still penalizes but also allows a part of the gradient (v) to be nonzero. Moreover, by penalizing the derivative of v a balance between linear structure

and piecewise constant structures is ensured. This balance can be influenced by the choice of β. Combining this with the data fidelity, the minimization problem related to the TGV_{β}^2 reconstruction model is

$$u = \operatorname*{argmin}_{u,v,w \in \mathrm{BV}(\Omega), \nabla u = v + w} \left\{ \frac{\lambda}{2} \left\| K_1 u - \tilde{f} \right\|_{L^2(\Omega)}^2 + \int_{\Omega} |w| \, dx + \beta \int_{\Omega} |\mathcal{E}(v)| \, dx \right\}. \tag{31}$$

3.3.3 Simultaneous contrast enhancement via iterative regularization

A well known problem for TV and TGV_{β}^2 models is that with strongly imposed regularization the reconstructions suffer from significant contrast loss. An inverse scale space method solving this problem is presented in [OBG$^+$05], which was later also implemented for the TGV_{β}^2 case in [BBBM13], however only for the case of denoising, *i.e.* $K_1 = \mathrm{Id}$.

Bregman iterative procedure Consider a general variational problem (25) the contrast enhancement procedure for this general problem is shown in **Model** 1. Instead of solving the minimization problem once, a sequence of minimization problems is solved and per step one has more information about the noise on the image, which is used in the next iteration.

Initialization:
$u_0 = 0$ and $p_0 = 0 \in \partial R(u_0)$
for $k = 0, 1, 2, \cdots$ *until stopping criterion is fulfilled* **do**
$\quad\mid\quad u_k = \operatorname{argmin}_u \{\lambda D(u, f) + R(u) - \langle u, p_{k-1} \rangle\}$
$\quad\mid\quad$ Update $p_k \in \partial R(u_k)$
end

      Model 1: Iterative regularization via a sequence of variational methods.

Without specifying $R(u)$, besides being convex and 1-homogeneous, and taking $D(K_1 u, \tilde{f}) = \frac{1}{2} \left\| K_1 u - \tilde{f} \right\|_2^2$ as data fidelity one can simplify the Bregman iterative procedure for the general $\mathrm{L}^2/R(u)$ model, just as done in [Bru10]. The minimization problem in this case would be,

$$\operatorname*{argmin}_u \left\{ \frac{\lambda}{2} \left\| K_1 u - \tilde{f} \right\|_2^2 + R(u) - \langle u, p_{k-1} \rangle \right\}. \tag{32}$$

First of all an update rule for p_k is needed. From the first order optimality condition of (32) an expression for $\partial R(u_k)$ is derived,

$$0 \in \partial R(u_k) - p_{k-1} + \lambda K_1^* \left(K_1 u_k - \tilde{f} \right), \qquad p_{k-1} - \lambda K_1^* \left(K_1 u_k - \tilde{f} \right) \in \partial R(u_k), \tag{33}$$

which gives a simple an update rule for p_k,

$$p_k = p_{k-1} - \lambda K_1^* \left(K_1 u_k - \tilde{f} \right), \tag{34}$$

Substituting $p_k := \lambda K_1^*(f_k - \tilde{f})$, the auxiliary variable p_k gets encorperated in the data f_k, giving the minimization problem

$$u_k = \operatorname*{argmin}_u \left\{ R(u) + \frac{\lambda}{2} \left\| K_1 u - \tilde{f} \right\|_2^2 - \langle u, \lambda K_1^*(f_{k-1} - \tilde{f}) \rangle \right\} \tag{35}$$

$$= \operatorname*{argmin}_u \left\{ R(u) + \frac{\lambda}{2} \left(\langle K_1 u, K_1 u \rangle - \langle K_1 u, f \rangle - \langle f, K u \rangle + \langle \tilde{f}, \tilde{f} \rangle \right) - \langle u, \lambda K_1^* \left(f_{k-1} - \tilde{f} \right) \rangle \right\}, \tag{36}$$

$$= \operatorname*{argmin}_u \left\{ R(u) + \frac{\lambda}{2} \| K_1 u - f_{k-1} \|_2^2 \right\}, \tag{37}$$

and the update rule,

$$f_k = f_{k-1} + \tilde{f} - K u_k. \tag{38}$$

The resulting algorithm is shown in **Model 2** where a stopping criterion is still not specified, this will be discussed in 5.4.

Initialization:
$u_0 = 0$ and $f_0 = \tilde{f}$
for $k = 0, 1, 2, \cdots$ *until stopping criterion is fulfilled* **do**
$\quad\quad u_k = \operatorname{argmin}_u \left\{ R(u) + \frac{\lambda}{2} \| K_1 u - f_{k-1} \|_2^2 \right\}$
$\quad\quad f_k = f_{k-1} + \tilde{f} - K_1 u_k.$
end
Model 2: Simplification of the iterative regularization via a sequence of variational methods for $\mathrm{L}^2/R(u)$ reconstruction model

---

This chapter discussed and introduced the reconstruction models. However, to actually implement these models, an algorithm that can effciently solve the non-smooth convex minimization problems shown in (26), (31) and **Models 1 & 2**, has to be derived. In the next chapter such an algorithm is derived, discussed and validated.

4 Numerical Methods

Now that the mathmatical models are introduced, the implementation of these models has to be discussed. The current reconstruction method is staightforward in its implementation and will therefore not be discussed. However the proposed models TV and TGV^2_β do need some additional attention. These variational methods have an intermediate step, the minimization of a convex non-smooth functional, which is a challenging problem on itself. The general strategy, within the image reconstruction field, is to solve these kind of minimization problems through rewriting it as a sadde point problem and use a primal-dual procedure to solve it. In this paper a primal-dual hybrid gradient (PDHG) alogrithm, as introduced in [ZC08], is used. There are several reasons for chosing this specific algorithm. First of all the applicability of this method, due to the simple general setup of the method (see 4.1), this method can be applied to a broad spectrum of saddle point problems. Also the modular structure of this method, gives way to easily adapt the algorithm implemented for a certain problem to an algorithm for a similar problem (see [CP11], [SJP12]). Comparing this method to, alternating direction method of multipliers (ADMM) [Gab83] and forward backward splitting (FBS) [Tse91], it has the advantage that the update steps are 'easier', *i.e.* ADMM and FBS solve stepwise simpler minimization problems where PDHG has a direct solution every step (see **Algorithm 4 & 5**). Furthermore it can be shown that PDHG has a convergence rate of at least $\mathcal{O}(n)$ and can even be $\mathcal{O}(n^2)$ (see [CP11], [PC11], [LP14]).

In the remainder of this chapter a specific PDGH will be explained and algorithms for the TV and TGV^2_β model will be derived and analyzed.

4.1 Primal-Dual Hybrid Gradient methods

The specific version of a Primal-Dual Hybrid Gradient algorithm used in this paper is explained in [CP11].

Let two finite-dimensional real vector spaces be equipped with an inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\| = \langle \cdot, \cdot \rangle^{\frac{1}{2}}$. The map $A : X \to Y$ is a continuous linear operator with induced norm

$$\|A\| = \max \left\{ \|Ax\| : x \in X \text{ with } \|x\| \leq 1 \right\}. \tag{39}$$

The general saddle point problem is then

$$\min_{x \in X} \max_{y \in Y} \langle Ax, y \rangle + G(x) - F^*(y), \tag{40}$$

where $G : X \to [0, +\infty]$ and $F : Y \to [0, +\infty]$ are proper, convex, lower-semicontinuous (l.s.c.) functions, F^* being itself the convex conjugate of a convex l.s.c. function F. Note that this saddle point problem is the formulation of the nonlinear primal problem

$$\min_{x \in X} F(Ax) + G(x), \tag{41}$$

in combination with the corresponding dual problem

$$\max_{y \in Y} - \left(G^*(-A^*y) + F^*(y) \right). \tag{42}$$

Considering a general problem of this form the PDHG algorithm would be as shown in **Algorithm 3**. For simple subproblems resolvent operators $R^H_\gamma(z)$ can be obtained through the following proximal point method:

$$R^H_\gamma := (I + \gamma \partial H)^{-1} (z) = \operatorname*{argmin}_{z'} \left\{ H(z') + \frac{\|z - z'\|^2_2}{2\gamma} \right\}. \tag{43}$$

The convex conjugate of an operator can be computed through the Legendre transform (see [Roc15]) and the original function can be recovered by applying the transformation again.

$$H^*(z) = \max_{z'} \left\{ \langle z, z' \rangle - H(z) \right\}. \tag{44}$$

Initialization:
Choose $(x^0, y^0) \in X \times Y$, $\tau, \sigma > 0$, s.t. $\sigma \tau L^2 < 1$, with $L = \|A\|$ and set $\bar{x}^0 = x^0$.
for $n = 0, 1, 2, \cdots$ **do**

$$\begin{cases} y^{n+1} & = (I + \sigma \partial F^*)^{-1} (y^n + \sigma A \bar{x}^n) \\ x^{n+1} & = (I + \tau \partial G)^{-1} (x^n - \tau A^* y^{n+1}) \\ \bar{x}^{n+1} & = 2x^{n+1} - x^n \end{cases}$$

end

Algorithm 3: General Primal-Dual Hybrid Gradient algorithm

4.1.1 Primal-Dual Hybrid Gradient algorithm for the TV model

To derive an algorithm the minimization problem corresponding to the TV model has to be written in the same form, recall (26), the discrete version in $\mathbb{R}^2$ would be

$$\min_u \frac{\lambda}{2} \left\| K_1 u - \tilde{f} \right\|_2^2 + \|\nabla u\|_{1,2}, \tag{45}$$

with

$$\|u\|_2^2 = \langle u, u \rangle, \qquad \|\nabla u\|_{1,2} = \sum_{i,j} |(\nabla u)_{i,j}|, \qquad |(\nabla u)_{i,j}| = \sqrt{((\nabla u)_{i,j}^1)^2 + ((\nabla u)_{i,j}^2)^2}. \tag{46}$$

In a likewise manner as done in [SJP12], F, G and A are set as follows

$$F(Ax) = \frac{\lambda}{2} \left\| K_1 x - \tilde{f} \right\|_2^2 + \|\nabla x\|, \qquad G(x) = 0, \qquad A = \begin{bmatrix} K_1 \\ \nabla \end{bmatrix}. \tag{47}$$

For the computation of the resolvent operators, split F into F_1 and F_2

$$F(y_1, y_2) = \frac{\lambda}{2} \left\| y_1 - \tilde{f} \right\|_2^2 + \|y_2\| = F_1(y_1) + F_2(y_2), \tag{48}$$

and consider the resolvent operators for F_1 and F_2 separately. Before the resolvent operator can be computed, the convex conjugates have to be computed, first for F_1

$$F_1^*(z) = \max_{z'} \left\{ \langle z, z' \rangle - \frac{\lambda}{2} \left\| z' - \tilde{f} \right\|_2^2 \right\}. \tag{49}$$

Computing the first order optimality condition gives

$$0 = z - \lambda \left(z' - \tilde{f} \right), \implies z' = \frac{z}{\lambda} + \tilde{f}. \tag{50}$$

Checking the second order optimality condition, shows that the z' found maximizes (52), plugging in the result gives:

$$F_1^*(z) = \frac{1}{2\lambda} \|z\|_2^2 + \left\langle y, \tilde{f} \right\rangle. \tag{51}$$

11

Computing the convex conjugate for F_2

$$F_1^*(z) = \max_{z'} \left\{ \langle z, z' \rangle - \|z'\| \right\}. \tag{52}$$

Note that

$$0 \geq \langle x, y \rangle - \|x\|_1, \text{ when } |y| \leq 1, \tag{53}$$

$$0 < \langle x, y \rangle - \|x\|_1, \text{ when } |y| > 1. \tag{54}$$

Therefore F_2^* becomes

$$F_2^* = \begin{cases} 0 & \text{if } |y| \leq 1, \\ \infty & \text{if } |y| > 1. \end{cases} \tag{55}$$

Now the resolvent operators can be computed for F_1^* and F_2^*, first for F_1^*

$$(I + \gamma \partial F_1^*)^{-1}(z) = \operatorname*{argmin}_{z'} \left\{ \frac{1}{2\lambda} \|z\|_2^2 + \left\langle y, \tilde{f} \right\rangle + \frac{\|z - z'\|_2^2}{2\gamma} \right\}. \tag{56}$$

Computing the first order optimality condition and checking the second order optimality condition gives

$$0 = \frac{z' - z}{\gamma} + \frac{z'}{\lambda} + \tilde{f}, \implies (I + \gamma \partial F_1^*)^{-1}(z) = \frac{z - \gamma \tilde{f}}{1 + \frac{\gamma}{\lambda}}. \tag{57}$$

Using equation (55) the resolvent operator for F_2^* can be computed as

$$(I + \gamma \partial F_2^*)^{-1}(z) = \operatorname*{argmin}_{z'} \left\{ F_2^*(z') + \frac{\|z - z'\|_2^2}{2\gamma} \right\}, \tag{58}$$

$$= \operatorname*{argmin}_{|z'| \leq 1} \left\{ \frac{\|z' - z\|_2^2}{2\gamma} \right\}, \tag{59}$$

$$= \frac{z}{\max(\mathbf{1}, |z|)}. \tag{60}$$

It remains to compute the resolvent operator for G

$$(I + \gamma \partial G)^{-1}(z) = \operatorname*{argmin}_{z'} \left\{ \frac{\|z - z'\|_2^2}{2\gamma} \right\} = z. \tag{61}$$

Using (57), (60) and (61) and **Algorithm 3** the PDHG algorithm for the TV model is derived (see **Algorithm 4**).

4.1.2 Primal-Dual Hybrid Gradient algorithm for the TGV_β^2 model

In a likewise manner as done in the previous paragraph a CP-PDHG algorithm can be derived for the TGV_β^2 model. First of all recall (31) and consider its discrete version,

$$\min_{u,v} \frac{\lambda}{2} \left\| K_1 u - \tilde{f} \right\|_2^2 + \|\nabla u - v\|_{1,2} + \beta \|\mathcal{E}v\|_{1,2}, \tag{62}$$

Initialization:
Choose $\left(u^0, p^0, q^0\right)$, $\tau, \sigma > 0$, s.t. $\sigma\tau L^2 < 1$, with $L = \|A\|$ and set $\bar{u}^0 = u^0$.

for $n = 0, 1, 2, \cdots$ **do**

$$
\begin{cases}
p^{n+1} & = (I + \sigma\partial F_1^*)^{-1} \left(p^n + \sigma K_1 \bar{u}^n\right) \\
q^{n+1} & = (I + \sigma\partial F_2^*)^{-1} \left(q^n + \sigma\nabla\bar{u}^n\right) \\
u^{n+1} & = (I + \tau\partial G)^{-1} \left(u^n - \tau\left(K_1^* p^{n+1} - \operatorname{div} q^{n+1}\right)\right) \\
\bar{u}^{n+1} & = 2u^{n+1} - u^n
\end{cases}
$$

end

Algorithm 4: Primal-Dual Hybrid Gradient algorithm for the TV model

where $\mathcal{E}$ is the discretised symmetric gradient and the constraint $\nabla u = w + v$ is incorporated in the problem. Setting F, G and A as follows

$$
F(Ax) = \frac{\lambda}{2} \left\|K_1 x_1 - \tilde{f}\right\|_2^2 + \|\nabla x_1 - x_2\|_{1,2} + \|\mathcal{E}x_2\|_{1,2}, \quad G(x) = 0, \quad A = \begin{bmatrix} K_1 & 0 \\ \nabla & -I \\ 0 & \mathcal{E} \end{bmatrix}, \quad (63)
$$

and repeating the reasoning for the resolvent operators, one gets

$$
(I + \gamma\partial F_1^*)^{-1}(z) = \frac{z - \gamma\tilde{f}}{1 + \frac{\gamma}{\lambda}}, \qquad\qquad (I + \gamma\partial G_1)^{-1}(z) = z, \qquad (64)
$$

$$
(I + \gamma\partial F_2^*)^{-1}(z) = \frac{z}{\max(\mathbf{1}, |z|)}, \qquad\qquad (I + \gamma\partial G_2)^{-1}(z) = z, \qquad (65)
$$

$$
(I + \gamma\partial F_3^*)^{-1}(z) = \frac{\beta z}{\max(\beta\mathbf{1}, |z|)}. \qquad (66)
$$

With these resolvent operators and the general algorithm a PDHG algorithm for the TGV_β^2 model is derived (see **Algorithm 5**).

Initialization:
Choose $\left(u^0, v^0, p^0, q^0, r^0\right)$, $\tau, \sigma > 0$, s.t. $\sigma\tau L^2 < 1$, with $L = \|A\|$ and set $\bar{u}^0 = u^0, \bar{v}^0 = v^0$.

for $n = 0, 1, 2, \cdots$ **do**

$$
\begin{cases}
p^{n+1} & = (I + \sigma\partial F_1^*)^{-1} \left(p^n + \sigma K_1 \bar{u}^n\right) \\
q^{n+1} & = (I + \sigma\partial F_2^*)^{-1} \left(q^n + \sigma\left(\nabla\bar{u}^n - \bar{v}^n\right)\right) \\
r^{n+1} & = (I + \sigma\partial F_3^*)^{-1} \left(r^n + \sigma\mathcal{E}\bar{v}^n\right) \\
u^{n+1} & = (I + \tau\partial G_1)^{-1} \left(u^n - \tau\left(K_1^* p^{n+1} - \operatorname{div} q^{n+1}\right)\right) \\
v^{n+1} & = (I + \tau\partial G_2)^{-1} \left(v^n - \tau\left(\mathcal{E}^* r^{n+1} - q^{n+1}\right)\right) \\
\bar{u}^{n+1} & = 2u^{n+1} - u^n \\
\bar{v}^{n+1} & = 2v^{n+1} - v^n
\end{cases}
$$

end

Algorithm 5: Primal-Dual Hybrid Gradient algorithm for the TGV_β^2 model

4.1.3 Parameter selection and Preconditioning

The last point, which has to be discussed before the validation of the algorithms, is the choice of stepsize parameters τ and σ. In the basic algorithm those are chosen spatially constant, and

moreover they must satisfy $\tau\sigma L < 1$ for the algorithm to converge. The combination of these constraints does not give a lot of room to make an optimal choice for the stepsize parameters. In [PC11] a smarter strategy for chosing the values of these stepsize parameters is introduced. Instead of taking a spatially constant parameter, the choice is now spatially dependent on the weight given through the continous linear operator A, *i.e.*

Definition $T = diag(\tau)$, *where* $\tau = (\tau_1, ..., \tau_n)$ *and* $\Sigma = diag(\sigma)$, *where* $\sigma = (\sigma_1, ..., \sigma_m)$. *In particular,*

$$\tau_j = \frac{1}{r\sum_{i=1}^{m}|A_{i,j}|}, \qquad\qquad \sigma_j = \frac{r}{\sum_{j=1}^{n}|A_{i,j}|}, \qquad (67)$$

where $r > 0$.

Note that the size of these parameters are still limited, this has again to do with assuring that the algorithm converges. For the more technical details see [PC11].
In **Algorithm 6** the implementation of this new parameter choice strategy is shown. This new algorithm is called the preconditioned PDHG, since matrices T and Σ function as preconditioners for the linear operators A and A^*.

> **Initialization:**
> Choose $(x^0, y^0) \in X \times Y$, $r > 0$, and set $\bar{x}^0 = x^0$.
> **for** $n = 0, 1, 2, \cdots$ **do**
> $$\begin{cases} y^{n+1} &= (I + \Sigma\partial F^*)^{-1}(y^n + \Sigma A\bar{x}^n) \\ x^{n+1} &= (I + T\partial G)^{-1}(x^n - TA^*y^{n+1}) \\ \bar{x}^{n+1} &= 2x^{n+1} - x^n \end{cases}$$
> **end**

Algorithm 6: General preconditioned Primal-Dual Hybrid Gradient algorithm

4.2 Algorithm validation

This section discusses the validation of the algorithms. Whereas the PDHG algorithm and the preconditioned PDHG are discussed in the previous section, in this section only the validation of the preconditioned version will be shown. This is due to the fact that only the preconditioned version is used for the experiments. Note that the normal PDHG works fine, however it is significantly slower, which makes it less useful in practice.

4.2.1 Validation criteria

To validate the algorithm there are two critera used. To see if the variables converge the relative residual is used and to see if the primal-dual problem is solved the primal-dual gap is considered.

Residuals Given a sequence of variables x^n the residual x_{res}^n is defined as

$$x_{res}^n = \frac{\left\|x^{n-1} - x^n\right\|_2}{\left\|x^{n-1}\right\|}. \qquad (68)$$

If $x_{res}^n \to 0$ when $n \to \infty$ then $x^{n-1} \to x^n$ and therefore the sequence x^n converges.

Primal-dual gap From the theory it is known that the primal-dual problem is solved when the solution of (41) value coincides with (42). This gives reason to consider the difference between these two values for every step, *i.e.* the primal-dual gap PD_{gap}^n

$$PD_{gap}^n = F(Ax^n) + G(x^n) - (-G^*(-A^*y^n) - F^*(y^n)).\tag{69}$$

So if the sequence PD_{gap}^n converges to 0, if $n \to \infty$, then the primal-dual problem would be solved.

Using (44) and the results in 4.1, explicit expressions for the primal-dual gaps are derived, for L^2/TV

$$PD_{gap}^n = \frac{\lambda}{2} \left\| K_1 u^n - \tilde{f} \right\|_2^2 + \|\nabla u^n\|_{1,2} + \frac{1}{2\lambda} \|p^n\|_2^2 + \langle p^n, f \rangle + G^*(\text{div } q^n - K_1^* p^n) + F_2^*(q^n),\tag{70}$$

and for L^2/TGV_β^2

$$PD_{gap}^n = \frac{\lambda}{2} \left\| K_1 u^n - \tilde{f} \right\|_2^2 + \|\nabla u^n - v^n\|_{1,2} + \beta \|\mathcal{E}v^n\|_{1,2} + \frac{1}{2\lambda} \|p^n\|_2^2 + \langle p^n, f \rangle \tag{71}$$

$$+ G^*(\text{div } q^n - K_1^* p^n) + G^*(q^n - \mathcal{E}^* r^n) + F_2^*(q^n) + F_3^*(r^n),\tag{72}$$

with

$$G^*(z) = \begin{cases} 0 & \text{if } z = 0 \\ \infty & \text{if } z \neq 0 \end{cases}, \quad F_2^*(z) = \begin{cases} 0 & \text{if } |z| \leq 1 \\ \infty & \text{if } |z| > 1 \end{cases}, \quad F_3^*(z) = \begin{cases} 0 & \text{if } |z| \leq \beta \\ \infty & \text{if } |z| > \beta \end{cases}.\tag{73}$$

where the functions G^*, F_2^*, F_3^* can be considered as constraints that need to be fulfilled. So instead of considering the primal-dual gap as defined above, a conditional primal-dual gap (cPD_{gap}^n) will be considered, wherein the contributions of G^*, F_2^*, F_3^* are ommited and the constraints relating to these functions will be monitored separately.

4.2.2 Convergence behaviour

In **Figures 4 & 5** the behaviour of the residuals, conditional primal-dual gap and constraints are shown for the L^2/TV and L^2/TGV_β^2 method. First of all note that in all of the cases the primal and dual variables converge, the primal-dual gap vanishes (these values shown are relatively small compared to the functional values of $\pm 3.2 \cdot 10^4$) and the constraints are satisfied as the number of iterations n is big enough. Most notable is the oscillating behaviour for small n, this is due to the preconditioning and has also been observed in [PC11] and [LP14].

Something not observed in **Figures 4 & 5** is the influence of certain parameters on the convergence behaviour. By changing the regularization parameters λ and β the amount of iterations needed to convergence changes, *i.e.* the more emphasis on the regularization terms the bigger n needs to be to see convergence. Which is not surprising, since the more emphasis on the non-smooth terms the more ill-posed the minimization problem becomes. Furthermore the choice of r greatly influences the behaviour primal-dual gap. This can be explained by interpreting the functionality of r in the context of this algorithm. For both the primal and dual problem there is a stepsize needed. However due to an implied size constraint (see **Lemma 2** in [PC11]) in the preconditioning matrices the stepsizes for the primal and dual problem cannot be chosen separately. Considering this, r can be interpreted as the relative stepsize between

the primal and dual problem, which explains that the choice of this parameter r influences the behavior of the primal-dual gap.

Although the convergence behaviour is not the same in every case, the general behavior which implies that the algorithms converge to the solution of the problem can be seen in all of these different cases.
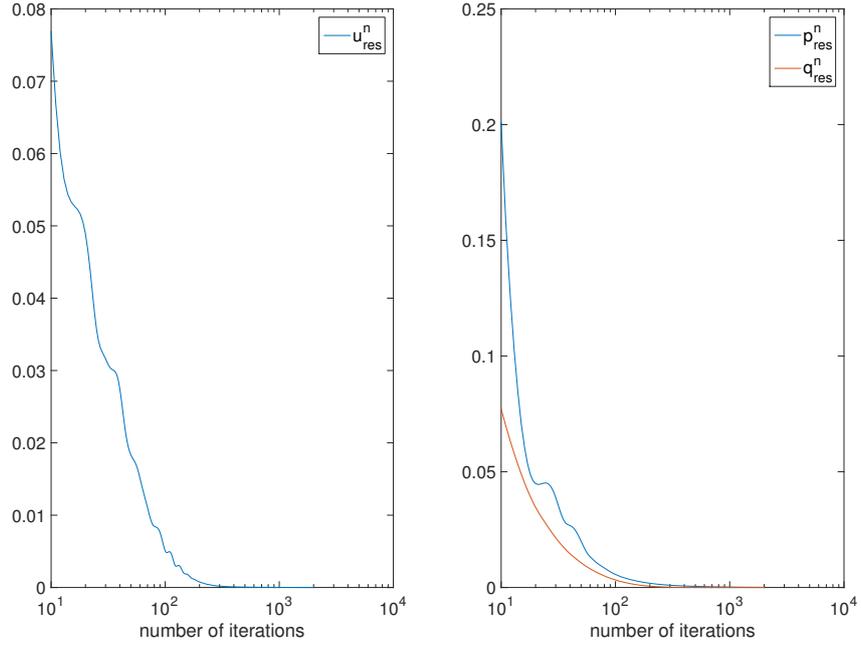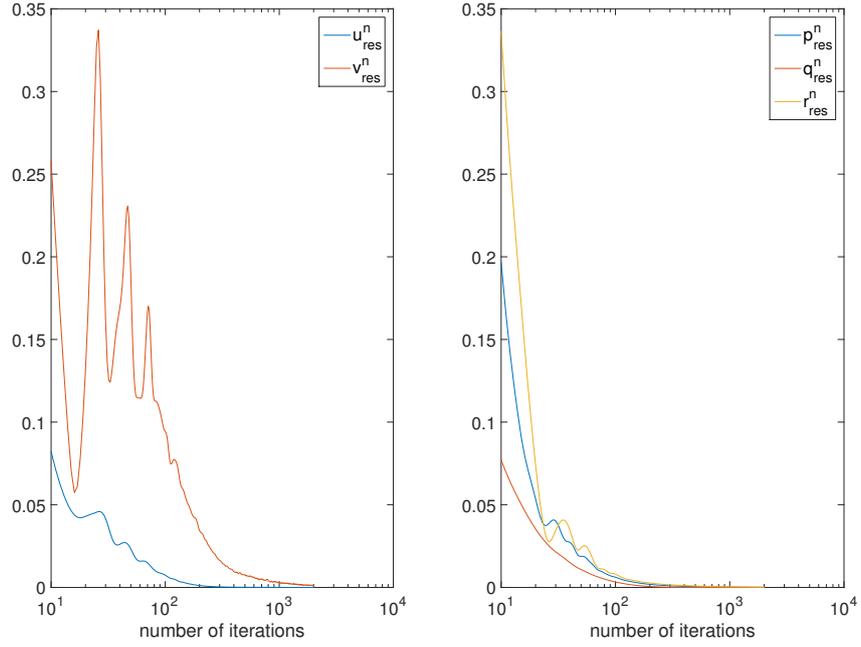
(a) Residuals
$u_{res}^{2000} = 3.96 \cdot 10^{-6}$, $p_{res}^{2000} = 8.05 \cdot 10^{-5}$,
$q_{res}^{2000} = 1.19 \cdot 10^{-6}$.
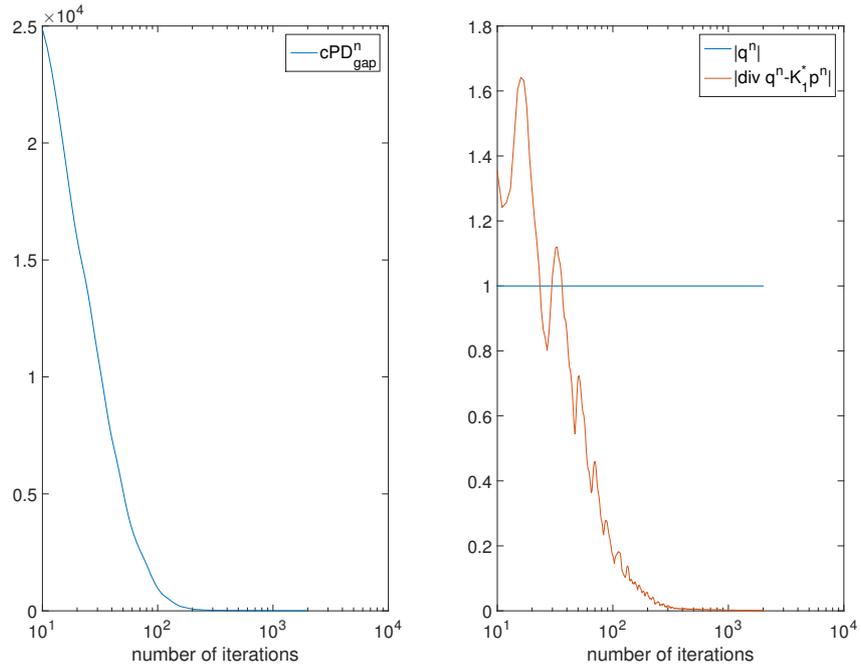
(b) Conditonal primal-dual gap & constraints
$cPD_{gap}^{2000} = 1.66$, $\left| q^{2000} \right| = 1$,
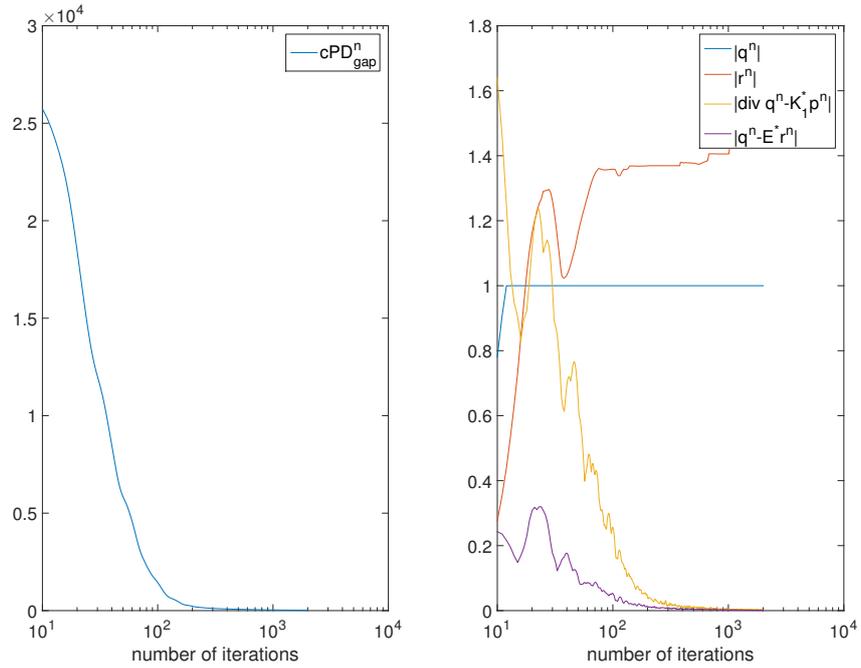$\left| \text{div } q^{2000} - K_1^* p^{2000} \right| = 5.71 \cdot 10^{-4}$.

Figure 4: Convergence results for the TV model.
These figures show the convergence results for the TV model applied to a 200×200 Shepp-Logan phantom with added Gaussian noise (with a standard deviation 5% of the maximum value of the input data). The regularization parameter is set at $\lambda = 50$. Furthermore the amount of iterations done is set at $n = 2000$ and the relative stepsize parameter is $r = 1$.

(a) Residuals
$u_{res}^{2000} = 1.33 \cdot 10^{-5}$, $v_{res}^{2000} = 1.20 \cdot 10^{-3}$,
$p_{res}^{2000} = 1.12 \cdot 10^{-4}$, $q_{res}^{2000} = 3.51 \cdot 10^{-6}$,
$r_{res}^{2000} = 1.80 \cdot 10^{-4}$.

(b) Conditonal primal-dual gap & constraints
$cPD_{gap}^{2000} = 17.76$, $\left| q^{2000} \right| = 1$, $\left| q^{2000} \right| = 1.45$
$\left| \text{div } r^{2000} - K_1^* p^{2000} \right| = 1.20 \cdot 10^{-3}$,
$\left| q^{2000} - \mathcal{E}^* p^{2000} \right| = 6.56 \cdot 10^{-6}$.

Figure 5: Convergence results for the TGV_β^2 model.
These figures show the convergence results for the TGV_β^2 model applied to a 200×200 Shepp-Logan phantom with added Gaussian noise (with a standard deviation 5% of the maximum value of the input data). The regularization parameters are set at $\lambda = 50, \beta = 1.5$. Furthermore the amount of iterations done is set at $n = 2000$ and the relative stepsize parameter is $r = 1$.

5 Results

The results and experiments are mostly focussed on the inner inverse problem. Sections 5.2, 5.3 and 5.4 focus on showing the limitations of the models discussed in **Chapter 3** for the inner inverse problem. In 5.5 the outer inverse problem is also considered, a realistic synthetic data set is introduced and the reconstructions with FBP, TV and TGV models on this set are discussed. But before the results are shown, the experimental setup and several other technicalities need to be adressed.

5.1 Experimental setup

In this section the complete experimental setup is discussed. To ensure control over all the uncertainties within the problem, all the assets in the experimental setup will be computer modelled. A synthetic data model will be given to study the inner inverse problem independently of the outer inverse problem. Furthermore the data structures, reconstruction criteria and parameter estimation methods will be analyzed.

5.1.1 Synthetic data model

By applying the forward model to a ground truth figure (also 'computer phantom') and adding noise a model for synthetic data model is created, *i.e.*

$$f_{syn} = K_2 K_1 u_{GT} + \epsilon, \tag{74}$$

where u_{GT} is the ground truth figure and ϵ is the added noise to the measurement. This noise is modeled as Gaussian noise with zero mean and variance based on the maximum values of the ground truth figure

$$\epsilon \sim s \max\left(K_2 K_1 u_{GT}\right) \mathcal{N}(0,1), \tag{75}$$

with $s \in [0,1]$.
Using (74), the outer inverse problem can be tested without any influence of the inner inverse problem. Moreover the combination of the outer and inner inverse problem can be tested. However the goal was to create a synthetic data model, such that the inner inverse problem could be studied independently. To solve this problem assume there is a perfect way to solve the outer inverse problem (K_2^{-1}) and apply this to (16) to get

$$K_2^{-1} f_{syn} = K_2^{-1}\left(K_2 K_1 u_{GT} + \epsilon\right), \tag{76}$$

$$\tilde{f}_{syn} = K_1 u_{GT} + \tilde{\epsilon}., \tag{77}$$

with $\tilde{\epsilon} \sim s \max\left(K_1 u_{GT}\right) \mathcal{N}(0,1)$, where $\mathcal{N}(0,1)$ denotes a normal distribution with zero mean and standard deviation 1 and $s \in [0,1]$. This way the synthetic data is not dependent on K_2 and gives a method of purely testing the models for the inner inverse problem.

Using synthetic data with this kind of procedure is called 'commiting an inverse crime'. This will result in reconstructions that are better than those that can be achieved in a more realistic situation in general. However this will still give a good indication of the limitations of the methods.
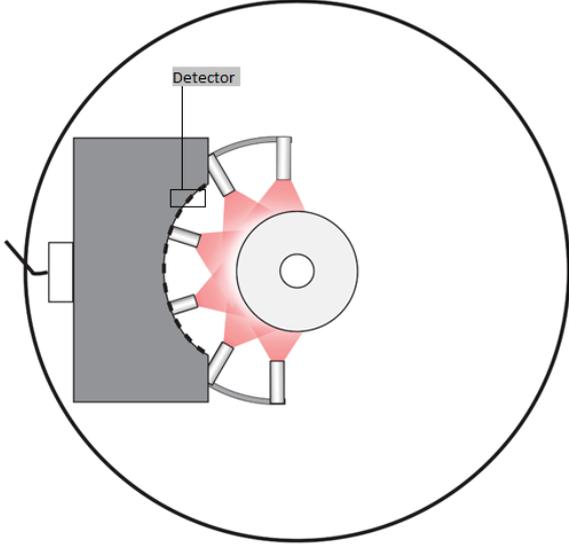
Figure 6: Schematic visualization of a measurement partition

A sinogram is built up from 12 measurement partitions, i.e. this measurement setup does a measurement from this angle, then rotates 30 degrees does a new measurement, creating a full 360 degree view of the object.

5.1.2 Data

In the experiments there will be different data sets used, the ground truth images of these sets are shown in **Figure 7** and the corresponding sinograms are shown in **Figure 8**. To test the limitations of the image reconstruction models the influence of uncertainty on these data sets is tested. First of all on a simple case as **Figure 7a**, to limit the influence of structure in the data (see 5.2). In 5.3 the datasets shown in **Figures 7b, 7c, 7d** are used to show the influence of a challenging structure, the need for higher order models and contrast enhancement. In 5.5 realistic data will be considered, however this will not be discussed here.

Before discussing the data sets itself, it is good to understand how the measurements are done and related to the sinograms. The measurement method considered in this thesis consists of 12 smaller measurements done with 32 detectors. These smaller measurements (partitions) are identical in structure except for having a different angle view on the object (**Figure 6** shows a schematic visualization of a partition of the measurement). In a measurement partition the 32 detectors collect data every timestep for a fixed amount of time. Combining these 12 partitions creates a sinogram, which is used as input for the reconstruction model.

There are several ways of considering uncertainty in these sinograms, in these experiments there are two main uncertainties considered, namely additive Guassian noise and subsampling.

The Gaussian noise is defined as in (75), here the s is the standard deviation of the noise and it is a fraction of the maximum value of the data structure itself. In **Figure 9** several partial sinograms are shown with different noise intensities.

Subsampling is having less than the maximal obtainable data. Most subsampling is periodic subsampling, *e.g.* leaving certain data periodically out of the data set on purpose to limit the size of the data set or doing a measurement with a few broken detectors (see **Figure 10a**). Another common subsampling case is limited angle subsampling. This is where a coherent part of the data cannot be considered, in this case it means that certain partitions would not be considered (see **Figure 10b**).

Since the focus does not lie on the handling of all possible versions of uncertainties on the data, a limited amount of cases is considered here. The cases chosen here are done to fit the data fidelity choice, other noise types could lead to data fidelities. More restrictive subsampling strategies would imply new strategies on the use of forward operators and regularization, for
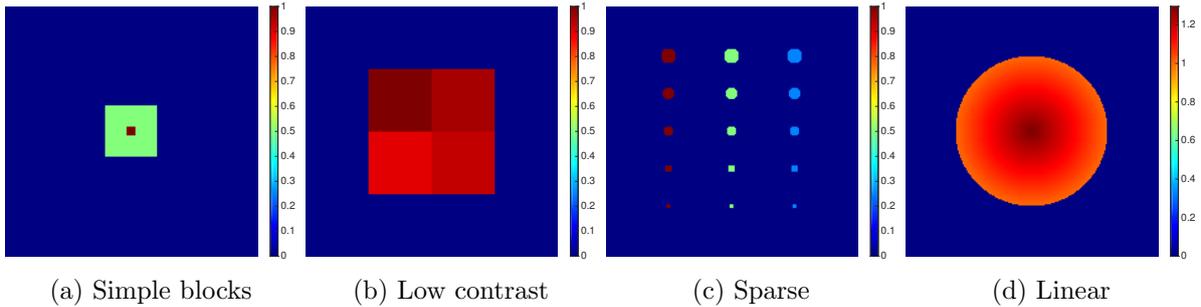
(a) Simple blocks   (b) Low contrast   (c) Sparse   (d) Linear

Figure 7: Different data structures
In these figures the ground truths for the considered data are shown.

(a) Simple blocks   (b) Low contrast   (c) Sparse   (d) Linear
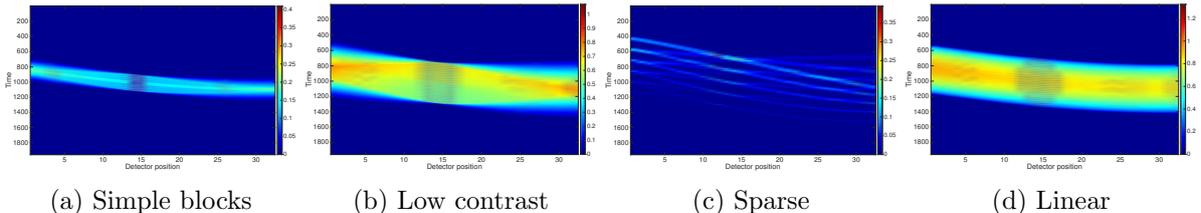
Figure 8: Partitions of the sinograms
In these figures the first measurement partition from the sinogram is shown, the full sinogram consists of 12 of these measurement partitions.

more information on this subject one could look at [BBLS15]. This work goes deeper into the limited angle subsampling topic and suggests an interesting strategy to tackle this problem.

5.1.3 Reconstruction criteria and parameter estimation

To compare the reconstructions there are criteria needed to judge the quality of the reconstruction. One of these criteria is obviously the reconstructed figure itself, but in some cases this is not suffiicient. Therefore two additional criteria are considered, which quantify the reconstruction quality. First of all the relative distance to the groundtruth, defined as

$$d_{GT}^n = \frac{\|u_{GT} - u_2^n\|}{u_{GT}}. \tag{78}$$

The smaller this relative distance to the ground truth, the better the reconstruction.
The last criteria is the peak signal-to-noise ratio (PSNR), which shows the ratio between the maximum possible power of a signal and the power of distorting noise that affects the quality of its representation expressed in dB,

$$\text{PSNR}_n = 10 \log_{10} \left(\frac{(\max(u_{GT}))^2}{\text{MSE}(u^n)} \right), \qquad \text{MSE}(u^n) = \frac{\|u_{GT} - u^n\|_2^2}{n_{pixels}} \tag{79}$$

where $\text{MSU}(u^n)$ is the mean squared error of the reconstruction and n_{pixels} is the amount of pixels in the image. Note that a lower mean squared error implies a better reconstruction, therefore a higher PSNR implies a better reconstruction.

5.1.4 L-curve parameter estimation

The choice of regularization parameter for the variational methods influences the quality of the reconstructions greatly, therefore there has to be a structural way to chose these parameters.
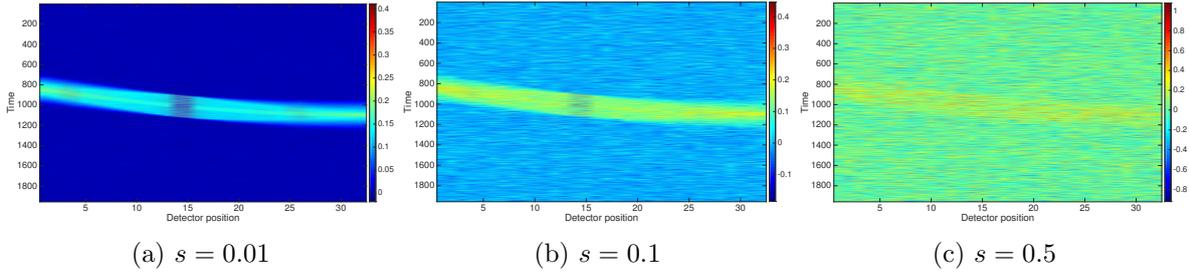
(a) $s = 0.01$ (b) $s = 0.1$ (c) $s = 0.5$

Figure 9: Sinograms with added noise

For visibility only the first measurement partition of the sinogram is shown. This is a sinogram related to **Figure 7a**. *Furthermore due to the huge amount of timesteps compared to the amount of detectors, the width of the elements in the x-direction is elongated, resulting in the strange discontinuities.*
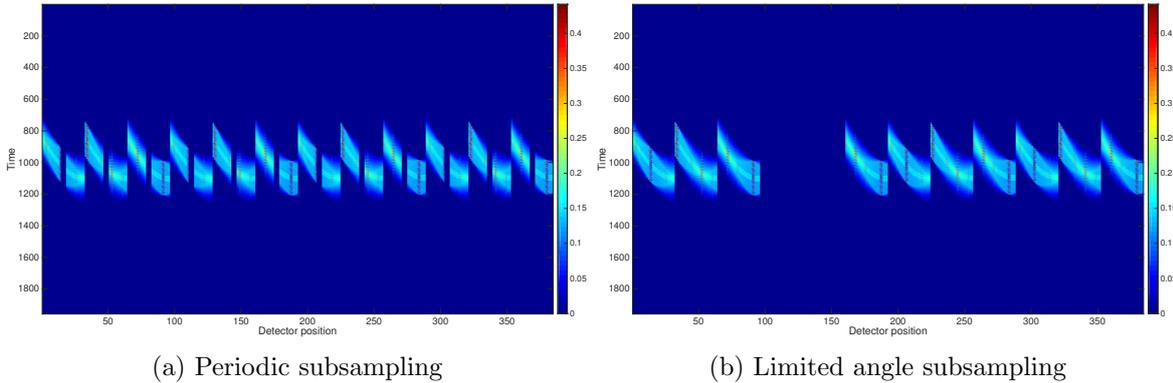

(a) Periodic subsampling   (b) Limited angle subsampling

Figure 10: Subsampling strategies

Periodic subsampling as seen in **Figure 10a** *is when certain detectors are not working anymore, in this case detector 15-18 are not working. Limited angle subsampling is when for example certain measurement partitions are not considered, in* **Figure 10b** *measurement partitions 4 and 5 are not considered. (see 1 for more information on the measurements)*

For this the L-curve parameter estimation is chosen. In this method one simply computes the reconstructions for several parameter values and looks for the optimal choice in this case, for this the quantative criteria are used. Due to the nature of the minimization problems, the results of this method show an L-shaped curve (see **Figure 11**).

This method will most likely not give the optimal choice of the parameter, but in practice this method shows to be sufficient.

5.2 Influence of uncertain data

In this section the influence of uncertainty in the data is tested. The results are split up into a part about the noise and a part about subsampling.

5.2.1 Noisy data

To test the influence of the noise a simple data structure is used (see **Figure 7a**). Since this is a piecewise constant figure only the TV and FBP models will be considered in the results. In **Table 1** the quantative results for the noise experiment are shown. An interesting thing to see in these results is that the relative difference in the FBP results is a lot smaller than for the TV model. Moreover the noise-free FBP has worse quantative results than the noisiest case for
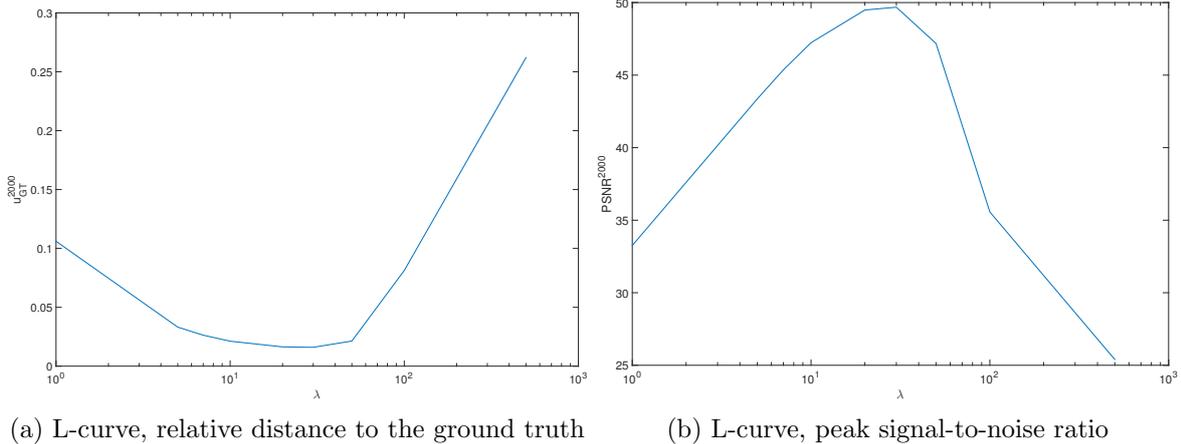
(a) L-curve, relative distance to the ground truth

(b) L-curve, peak signal-to-noise ratio

Figure 11: *L-curve parameter estimation for the TV model applied to* **Figure 7a** *with* $s = 0.05$ *(see* (75)*) and* $n = 2000$. *In this case* $\lambda = 30$ *would be the best choice. For the* TGV_β^2 *model the L-curve method would give a two dimensional version of these figures (since there are two parameters to estimate)*

	d_{GT}	$PSNR$		d_{GT}	$PSNR$
L^2/TV			FBP		
$s = 0, \lambda = 1000$	$8.30 \cdot 10^{-4}$	81.04	$s = 0$	0.835	21.06
$s = 0.01, \lambda = 100$	$7.25 \cdot 10^{-3}$	62.33	$s = 0.01$	0.835	21.06
$s = 0.05, \lambda = 30$	$2.21 \cdot 10^{-2}$	52.55	$s = 0.05$	0.840	21.01
$s = 0.1, \lambda = 25$	$4.04 \cdot 10^{-2}$	47.30	$s = 0.1$	0.854	20.86
$s = 0.2, \lambda = 12$	$7.51 \cdot 10^{-2}$	41.91	$s = 0.2$	0.911	20.31
$s = 0.3, \lambda = 7$	0.109	38.67	$s = 0.3$	0.995	19.54
$s = 0.4, \lambda = 5$	0.127	37.36	$s = 0.4$	1.11	18.63
$s = 0.5, \lambda = 4$	0.163	35.20	$s = 0.5$	1.23	17.68

Table 1: Reconstruction results for TV and FBP applied to noisy data

TV. In **Figure 12a** the noise-free FBP reconstruction is shown. Looking at this reconstruction the reason for the low quantative results is the contrast loss in combination with the streaking artefacts. What the quantitative results did not indicate is that the contours are reconstructed very well. Moreover comparing the reconstruction to the noisiest reconstruction for the TV model (see **Figure 13c**), there is much to say for prefering the FBP reconstruction. Which implies that the quantative reconstructions should not always be leading in the determination of the quality of a reconstruction. Obviously this is an unfair comparison and comparing the FBP reconstructions to the TV reconstructions for the same noise levels shows that the TV model clearly outperforms FBP.

5.2.2 Subsampling data

Looking at the sinogram structure (see **Figure 14a**) there is a lot of information in every partition and the partitions overlap in information. This indicates that the methods can probably handle signifcant subsampling. In **Figure 14b** and **Figure 14c** respectively a significant periodic and limited angle subsampling strategy are shown. Here in the case of periodic subsampling all partitions are considered, but only one detector is used. In limited angle subsampling
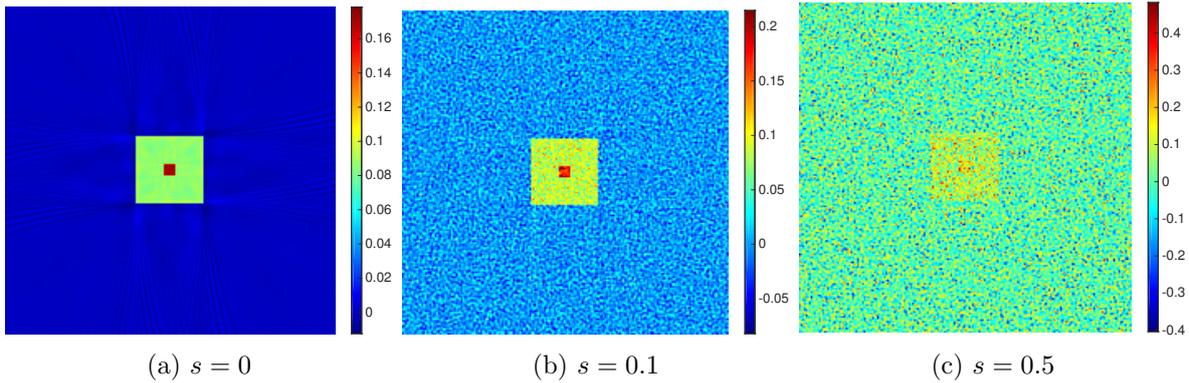
(a) $s = 0$ (b) $s = 0.1$ (c) $s = 0.5$

Figure 12: Reconstruction results for FBP applied to noisy data

As is known from filtered backprojection it creates streaking artefacts, which can cleary be seen in **Figure 12a**, *and it has no efficient way to handle noise, which can be seen in* **Figure 12b** *and* **Figure 12c**.

(a) $s = 0, \lambda = 1000$ (b) $s = 0.1, \lambda = 25$ (c) $s = 0.5, \lambda = 4$
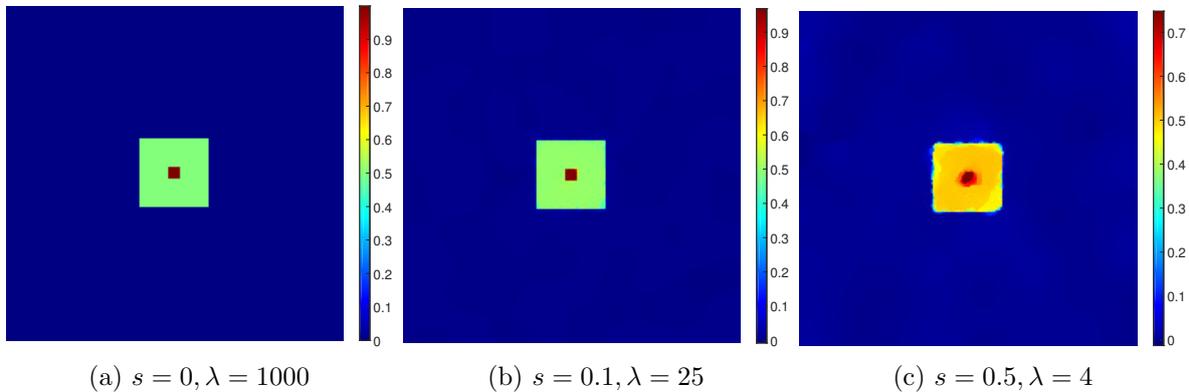
Figure 13: Reconstruction results for TV applied to noisy data

Looking at these figures shows that the TV model can handle noise in an efficient way if the regularization parameter is chosen correctly. However the more regularization that is needed the more the well known contrast loss problem occurs, which can best be seen in **Figure 13c**.

all detectors are considered, but only the last partition is. In **Figure 15** and **Figure 16** the reconstruction results for both of these strategies are shown for FBP and TV. There is a strong difference between the FBP and the TV models, for both subsampling strategies the TV has enough information to do a good reconstruction, whereas FBP loses a lot of contrast, creates significant artefacts and cannot reconstruct the contours. It only remains to see how much information TV needs to still do an adequate reconstruction.

In **Table 2** the reconstruction results for both subsampling strategies are shown, where for the periodic subsampling the amount of partitions used is lowered and with limited angle subsampling the amount of detectors used is lowered. As can be seen in the results again roughly 70% of this data can be thrown away before the reconstructions are showing significant artefacts (see **Figure 17**).

5.3 Difficult data structures

In the previous sections it is shown that for simple data structures a lot of uncertainty can be handled, in this section three challenging data structures tested.
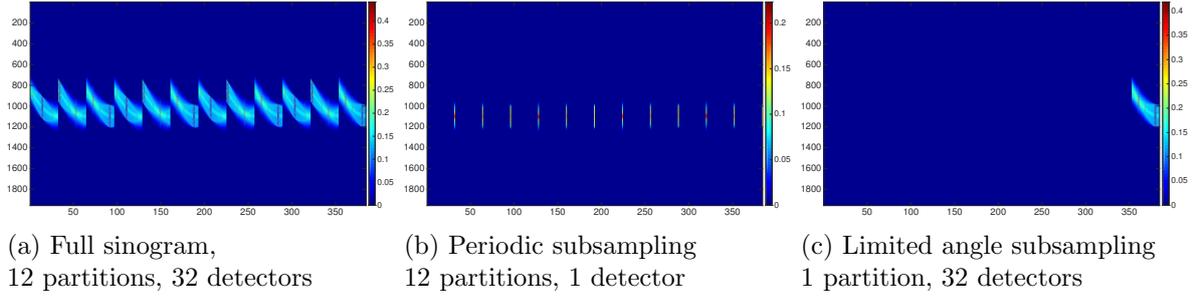
(a) Full sinogram,
12 partitions, 32 detectors

(b) Periodic subsampling
12 partitions, 1 detector

(c) Limited angle subsampling
1 partition, 32 detectors

Figure 14: Sinograms with subsampling strategies
Figure 14a *shows a full sinogram of a the simple block structure (see* **Figure 7a** *without any subsampling.* **Figure 14b** *&* **Figure 14c** *show significant subsampling for the periodic and limited angle subsampling strategies.*

One detector	d_{GT}	$PSNR$	One partition	d_{GT}	$PSNR$
# partitions			# detectors		
6	$4.72 \cdot 10^{-2}$	45.94	11	$4.81 \cdot 10^{-2}$	45.78
5	$5.90 \cdot 10^{-2}$	44.01	9	$4.94 \cdot 10^{-2}$	45.55
4	$8.90 \cdot 10^{-2}$	41.26	7	$5.43 \cdot 10^{-2}$	44.73
3	0.120	37.86	5	0.145	36.21
2	0.514	25.21	3	0.728	22.18
1	0.883	20.51	1	0.883	20.51

Table 2: Reconstruction results for TV applied to subsampled data
In these reconstructions a small noise ($s = 0.01$) is added to have less influence of the inverse crime. For all the reconstructions $\lambda = 1000$ is chosen, compared to the previous section this seems not a lot, however since there is significant less information in the data, the data fidelity in the functional is smaller in absolute values, therefore the regularization has relatively more influence.

5.3.1 Low contrast and sparse data structures

Figure 7b and **Figure 7c** show the data structures considered in this section. The challenge for both of these structures is that if there is significant noise it is hard to distinguish the noise from the data. Both **Figure 18** and **Figure 19** illustrate this problem well. In **Figure 18b** the lines in the block start fading due to this problem. It is not strange that this is happening, since there is a 3% contrast difference between these blocks (going clockwise from the top left) in the ground truth and the noise here has a standard deviation of 20% of the maximum value of the data. In **Figure 19b** the problem is that the sparse dot in the bottom left behaves the same as the noise, *i.e.* a small deviation of the surrounding, not significant in size and not significant in intensity. In both cases the problem with $s = 0.1$ noise gives a good reconstruction.

5.3.2 Linear data structure

Since the TV model has the assumption that the ground truth data is piecewise constant, this sort of data should be very difficult to reconstruct. In **Figure 20a** the result is shown. The 'staircasing' artefacts can be clearly seen, other than that the reconstruction is very clean. Looking at the TGV^2_β reconstruction (see **Figure 20b**) the linear part inside the figure is reconstructed very good. However at the contour there seems to be a kind of reflection, creating
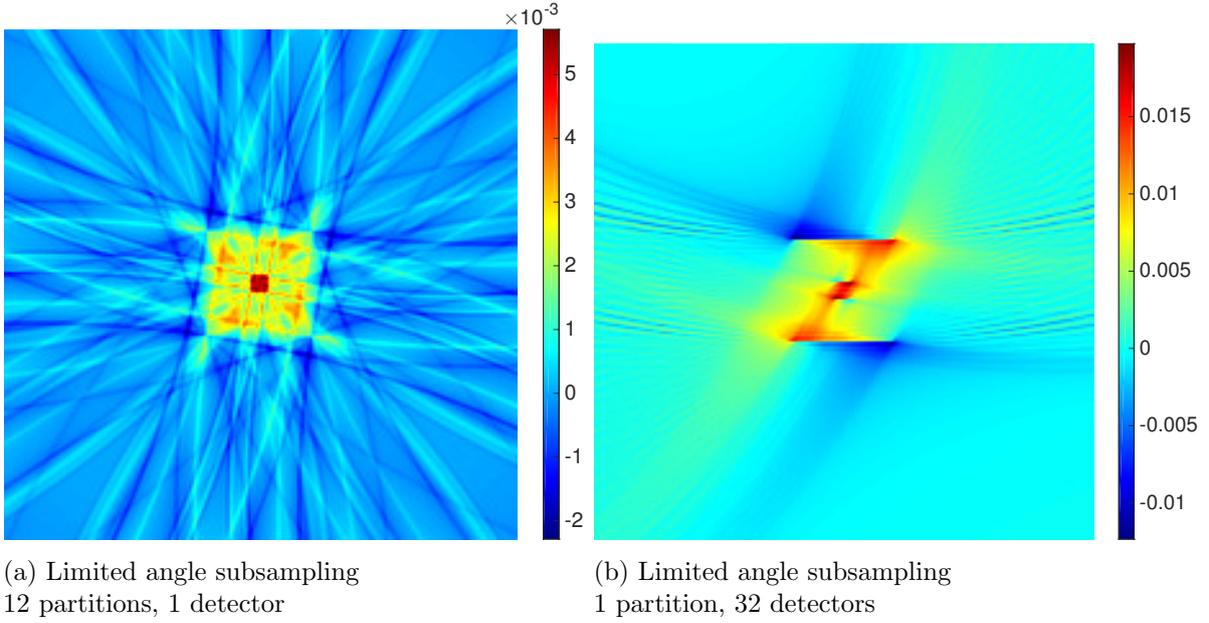
(a) Limited angle subsampling
12 partitions, 1 detector

(b) Limited angle subsampling
1 partition, 32 detectors

Figure 15: FBP subsampling

(a) Limited angle subsampling
12 partitions, 1 detector

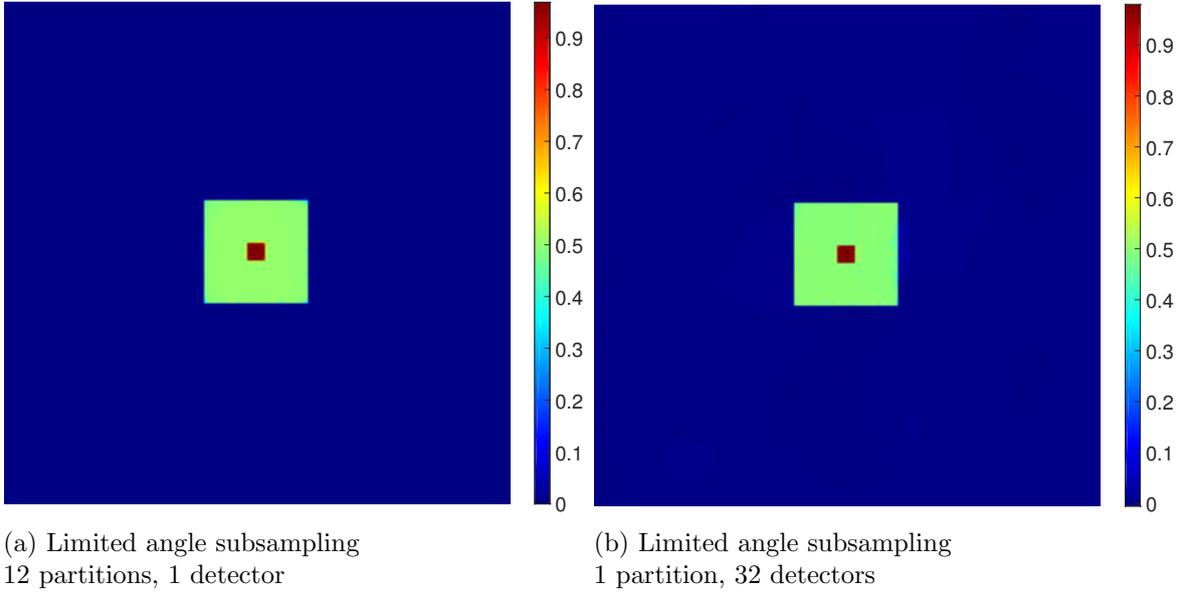(b) Limited angle subsampling
1 partition, 32 detectors

Figure 16: Subsampling TV

Figure 15 & **Figure 16** *show the reconstructed images for respectively FBP and TV ($\lambda = 1000$) with the significantly subsampled data and a little noise ($s = 0.01$) as shown in* **Figure 14b** *and* **Figure 14c**. *It shows that FBP cannot handle such significant subsampling, whereas TV still performs very well.*

a circle around the figure with more intensity than the background and less than the contour itself. This kind of artefacts are also seen in real data sets, so this could be an artefact created by the operator. This type of artefact is not seen that significantly in the TV reconstructions since these reconstructions do not allow this gradual change of intensity. However looking back at all the TV reconstructions all the cases, where not a perfect reconstruction is shown, the contours and their surroundings go gradually down.
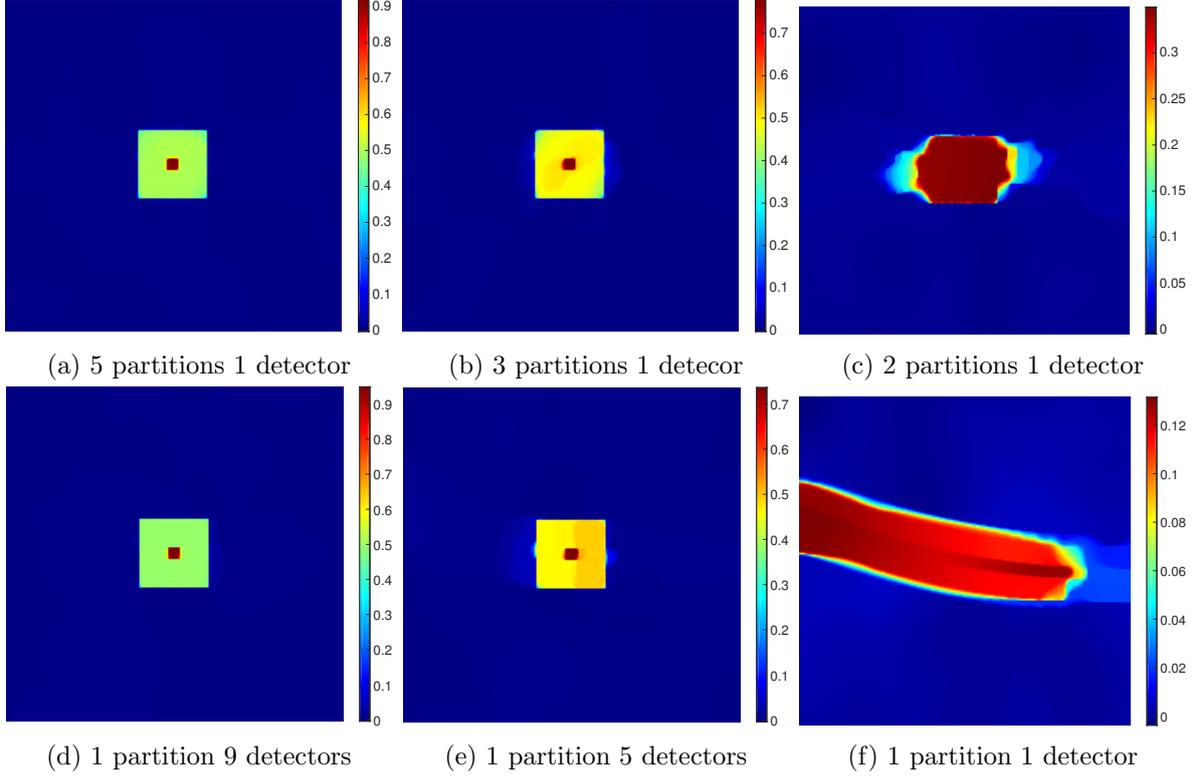
26

(a) 5 partitions 1 detector	(b) 3 partitions 1 detecor	(c) 2 partitions 1 detector
(d) 1 partition 9 detectors	(e) 1 partition 5 detectors	(f) 1 partition 1 detector

Figure 17: Reconstruction results for TV applied to highly subsampled data
These figures show that at some point the TV model will introduce artefacts from subsampling, first the contrast is lowered and later on the structure itself is lost. Note that in the cases where the TV model breaks down there is approximately 1.5% of the original data left.

5.4 Contrast loss

As known from the theory and seen in the previous results TV & TGV_β^2 have problems with contrast loss when a high amount of regularization is needed. To tackle this problem the Bregman iterative procedure was introduced in [OBG$^+$05]. Earlier in this thesis (see 3.3.3) this procedure was worked out for a general $\mathrm{L}^2/R(u)$ method, but without a specified stopping criterion. In [OBG$^+$05] a stoppig criterion based on the amount of noise on the problem is introduced, the so-called discrepancy principle. In practice it might be a challenge to use such a stopping criterion, however in this case the amount of noise is readily avaiable. The TV Bregman iterative procedure with discrepancy principle as stopping criterion is shown in **Model 7**. In **Figure 21** an example illustrating the effectiveness of this contrast enhancement is shown.

Initialization:
$u_0 = 0$ and $f_0 = f$
while $\ \|f - K_1 u_k\|_2^2 > \|f - K_1 u_{GT}\|_2^2$ **do**
$\quad \left| \ u_k = \mathrm{argmin}_u \left\{ R(u) + \frac{\lambda}{2} \|K_1 u - f_{k-1}\|_2^2 \right\} \right.$
$\quad \left| \ f_k = f_{k-1} + f - K_1 u_k. \right.$
$\quad \left| \ k = k + 1 \right.$
end

Model 7: Bregman iterative procedure for TV reconstruction model

(a) $s = 0.1$, $\lambda = 5$
$d_{GT} = 2.90 \cdot 10^{-2}$, $PSNR = 37.17$

(b) $s = 0.2$, $\lambda = 3$
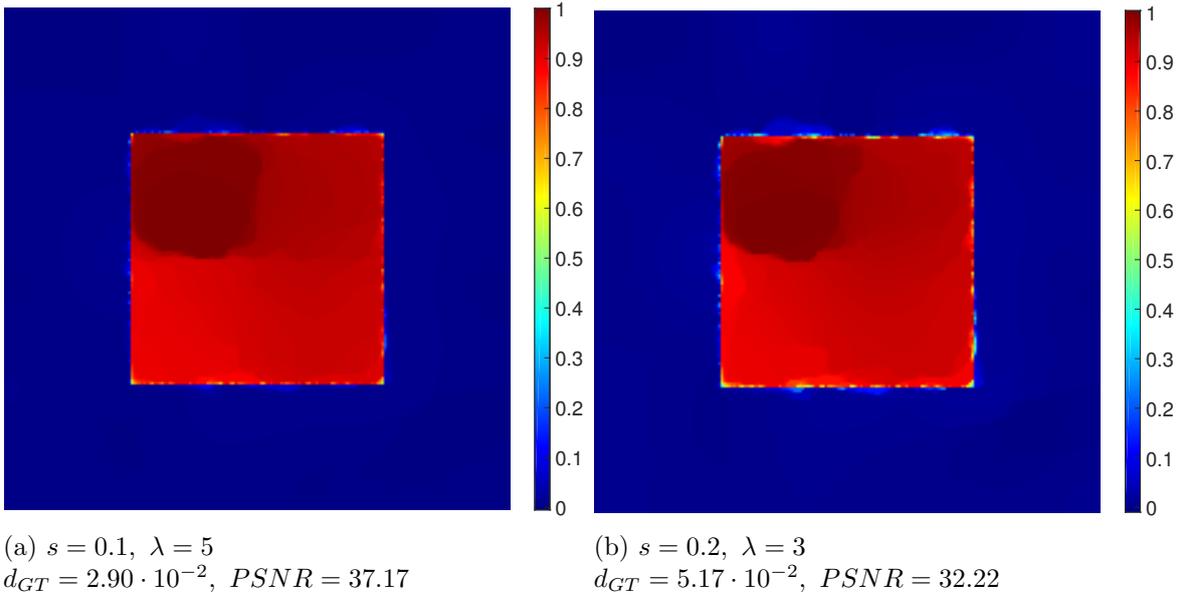$d_{GT} = 5.17 \cdot 10^{-2}$, $PSNR = 32.22$

Figure 18: Reconstruction results for TV applied to low contrast data
The reconstructions in these figures show noise artefacts at the contours, moreover the hard discontinuities in the ground truth are mostly lost in these reconstructions.

(a) $s = 0.1$, $\lambda = 20$
$d_{GT} = 0.120$, $PSNR = 39.16$

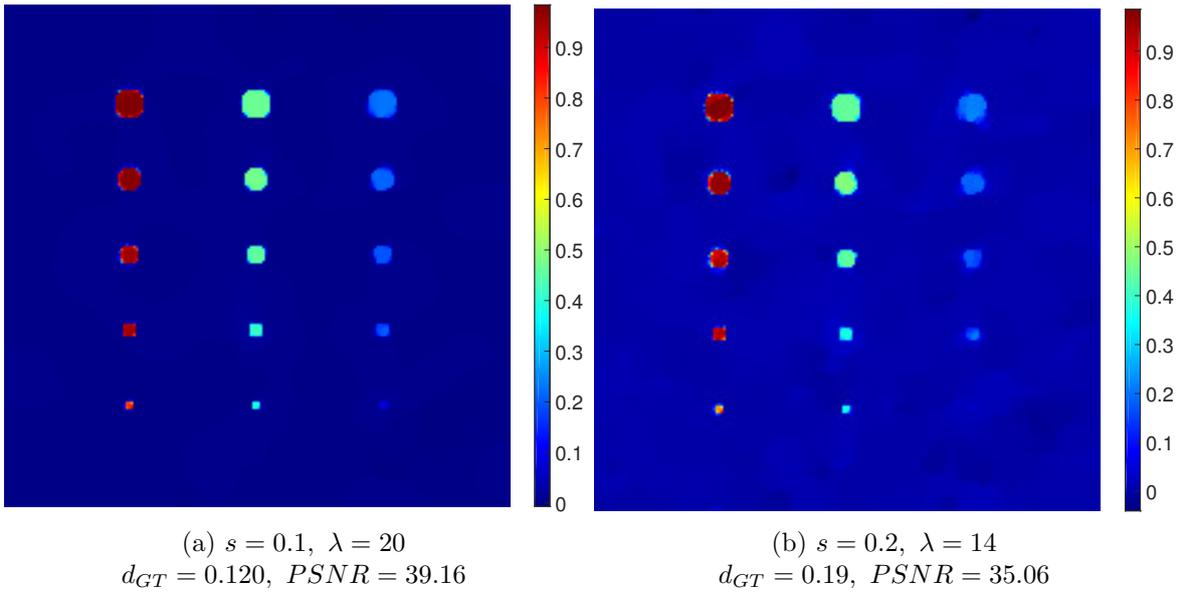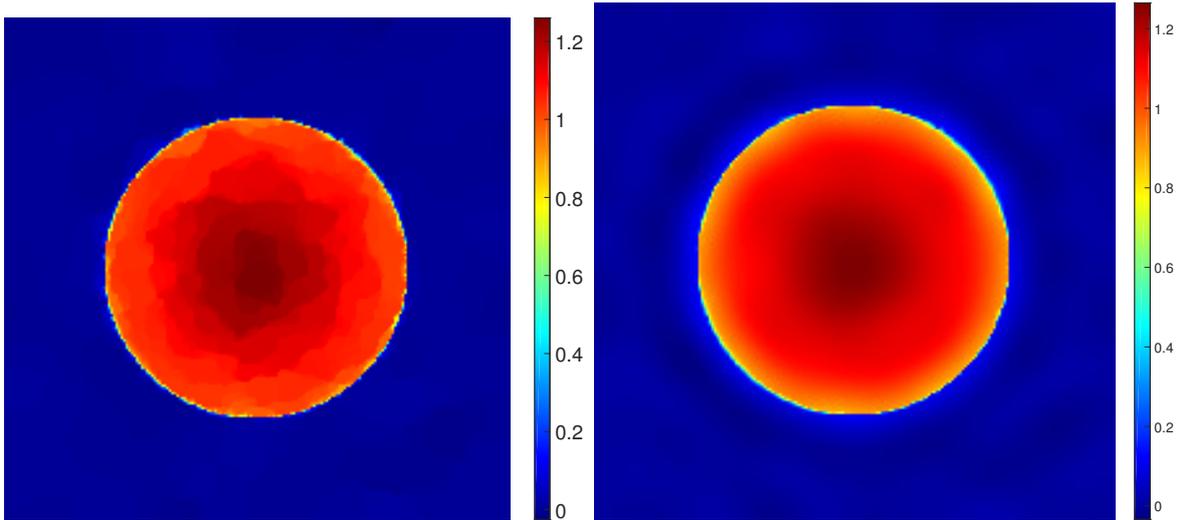(b) $s = 0.2$, $\lambda = 14$
$d_{GT} = 0.19$, $PSNR = 35.06$

Figure 19: Reconstruction results for TV applied to sparse data
The contours in these reconstructions are recovered very well in this cases, however the contrast loss is clearly visible in the smaller dots, up to even the bottom right dot even disappearing in the background (see **Figure 19b***).*

5.5 Realistic data

This section focussed on a more realistic case. The idea is to take a real data set reconstructed with FBP, then creating a synthetic data set that looks like this real data set when it is reconstructed with FBP. Obviously to make this comparison fair the outer inverse problem should also be considered here. Therefore the synthetic data is now computed from (74). So first the outer inverse problem needs to be solved, this is done with the current solution strategy (see

(a) Reconstruction result for TV applied to linear data

$s = 0.1, \ \lambda = 20$

$d_{GT} = 4.75 \cdot 10^{-2}, \ PSNR = 31.15$

(b) Reconstruction result for TGV_β^2 applied to linear data

$s = 0.2 \ \lambda = 14$

$d_{GT} = 0.19, \ PSNR = 35.06$

Figure 20: Linear data

Interestingly enough the quantative results indicate that the TV recosntruction is better than the TGV_β^2 reconstruction. This is due to the 'staircasing' artefacts in TV reconstruction being less influential than this big 'reflection' artefact in the TGV_β^2 reconstruction on the quantative reconstruction criteria.
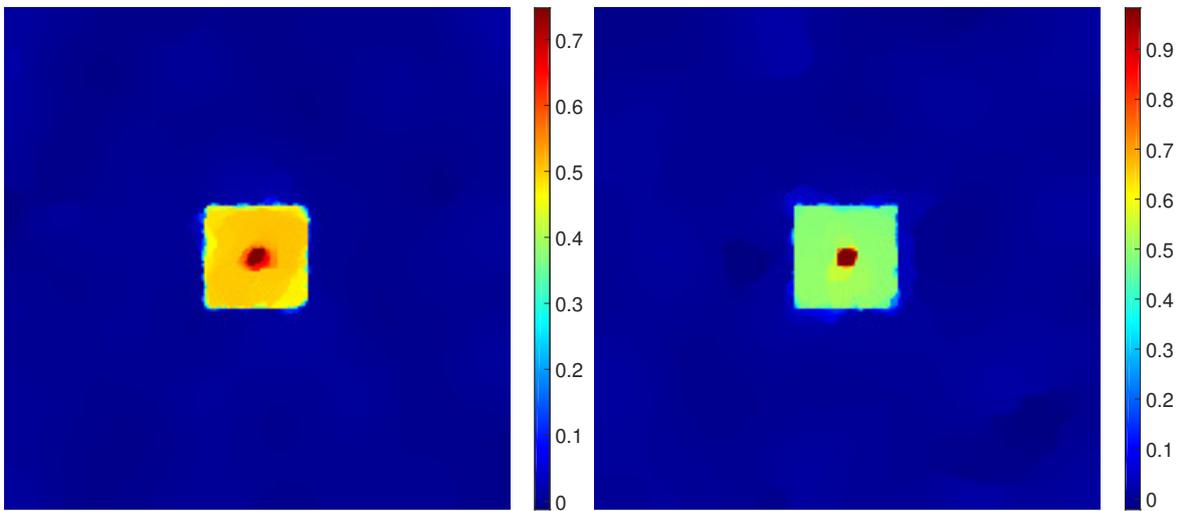
(a) TV reconstruction without Bregman, $s = 0.5, \ \lambda = 8$

(b) TV reconstruction with Bregman iteration $s = 0.5, \ \lambda = 3, \ \beta = 5$

Figure 21: TV reconstruction for strong noise data with and without Bregman

*Comparing these reconstructions to the ground truth, **Figure 7a**, the reconstruction with the Bregman iterative procedure shows almost the correct intensities and is far closer to the correct relative intensity.*

3.2). In **Figure 22** the solution of the outer inverse problem is shown without and with low

noise ($s = 0.01$). Notice that very little noise on the original data introduces a lot of noise on the solution of the outer problem, moreover the noise structure is more coherent than the simple additive Gaussian noise. This has to be taken into account in finding this synthetic realistic data set.

In **Figure 23** the realistic synthetic data and the real finger data are shown. The synthetic data has a lot of similarities to the real data, except for the finer structures inside the finger itself. Looking at this data set, it is quite challenging, it has thin low contrast lines, small high intensity points and a bit of linearity in the background.

Figure 24 shows the reconstructions of the different methods on the synthetic realistic data set. Looking at these reconstructions the same behavior is observed as in the previous experiments. However in the TV and TGV^2_β there are more 'patchy' artefacts. This is most likely due to how the noise on the original data is propogated in the outer inverse problem. These 'patchy' artefacts are more seen in the TGV^2_β reconstruction, this is due to the method interpreting these patches as a linear structure and therefore leaving them in.

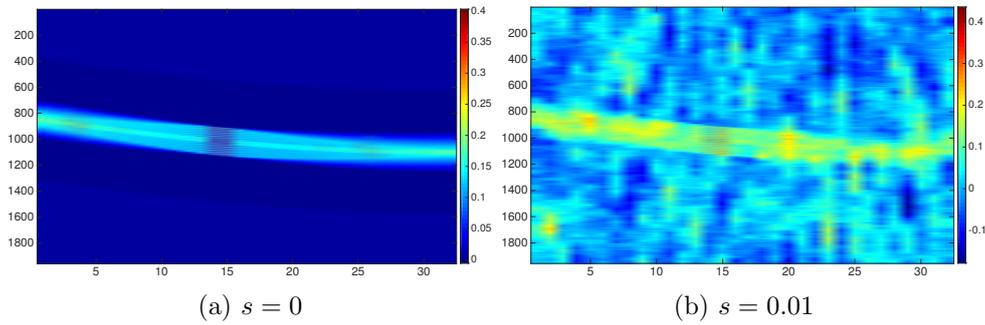(a) $s = 0$ (b) $s = 0.01$

Figure 22: First partition of the solution to the outer problem
In this figure the solutions to the outer inverse problem are shown for the simple blocks data set (see **Figure 7a** *and* **Figure 8a***). Notice that the case without noise shows already an artefact around the data and the case with very little noise is full of artefacts.*

(a) Ground truth   (b) FBP reconstruction   (c) Real finger data
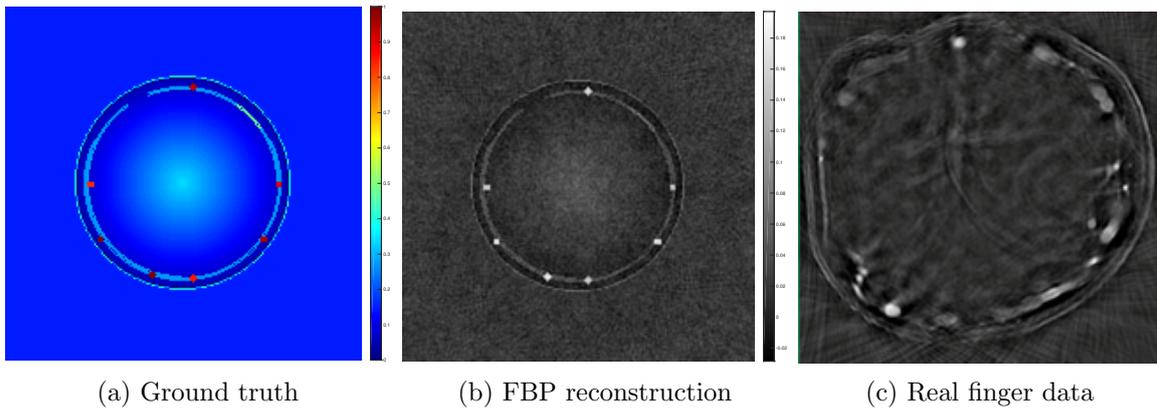
Figure 23: Creation of the realistic data set
The synthetic data is very similar to the real data, except for the fine structures inside the finger itself. This reconstruction is done with solving the outer and inner inverse problem.
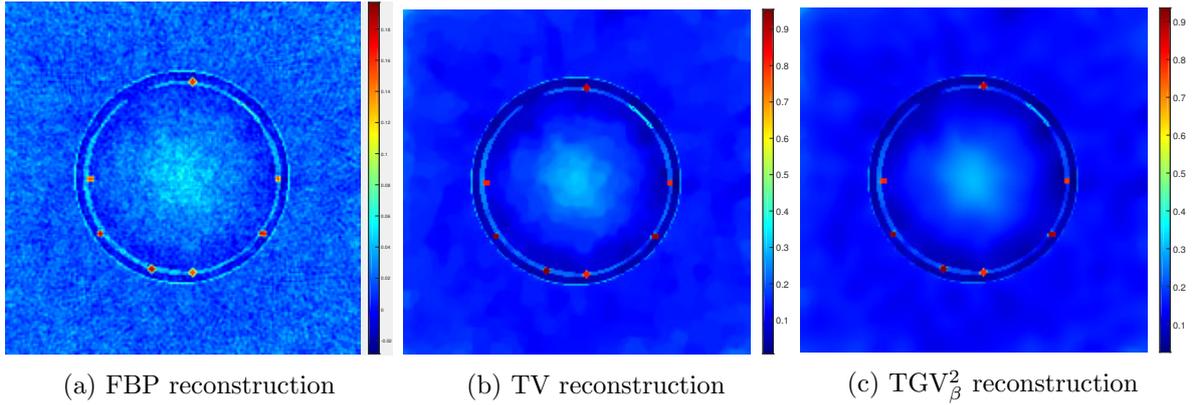
| (a) FBP reconstruction | (b) TV reconstruction | (c) TGV_β^2 reconstruction |

Figure 24: Reconstruction results on realistic synthetic data
Looking at these reconstructions the FBP reconstruction shows significant contrast loss and not handling the noise. The TV reconstruction shows not a lot of influence of the noise, but inside the finger there are 'staircasing' artefacts. At last the TGV_β^2 reconstruction does not show any artefacts, but it does show a bit more influence of noise.

6 Summary and Conclusions

In this thesis L^2/TV and $\mathrm{L}^2/\mathrm{TGV}_\beta^2$ are implemented in the reconstruction model for photoacoustic tomography. Moreover the limitations of these models are carefully tested for problems such as uncertain data and difficult data structures.

In the first part the forward model describing the physical process for photoacoustic tomography and the mathematical model are introduced. Where the structure of the forward model, *i.e.* the measured data being the time convolution of the calibration measurement with the projection integral

$$\tilde{p}(\mathbf{r}, t) = \tilde{p}_\delta(\mathbf{r}, t') *_t \left(\frac{\|\mathbf{r} - \mathbf{r_p}\|}{t} \int\limits_{\|\mathbf{r}-\mathbf{r}'\|=ct} A(\mathbf{r}')d\mathbf{r}' \right), \tag{80}$$

gives way to the coupled inverse problem mathmatical model. This model has several solution strategies. The solution strategy considered in this thesis seperates the coupled inverse problems and uses the solution of the outer problem as input for the inner problem. The inner problems is perfectly suitable for implementing variational methods such as L^2/TV and $\mathrm{L}^2/\mathrm{TGV}_\beta^2$.
The big challenge with this kind of nonlinear variational methods is to implement them in an efficient way. This is due to the nonsmooth nature of the saddle point problems introduced in these models. In this thesis a Primal-Dual Hybrid Gradient algorithm is chosen to implement the L^2/TV and $\mathrm{L}^2/\mathrm{TGV}_\beta^2$ models. This algorithm is chosen for its general applicability for this kind of problems, giving room for improvements on the model. Simple updating rules making the implementing easier and the possibility of using preconditioning on the algorithms. As seen in 4.2 this algorithm solves the saddle point problems efficiently and correctly.
At last the reconstruction models are tested on cases that closely relate to problems in practice. Recalling **Figure 2** the reconstructions showed sparse, low contrast data with noise and streaking artefacts possibly from subsampling. In **Chapter 5** it is shown that the L^2/TV model can handle noise efficiently up to contrast loss (see **Figure 13**). Moreover combining the L^2/TV model with the Bregman iterative procedure results in reconstructions without the ealier observed contrast loss (see **Figure 21**). Also the experiments for subsampling showed

that the L^2/TV model can handle a lot of subsampling. In the cases where the FBP reconstructions seemed to completely break down (see **Figure 15**) with significant subsampling, the L^2/TV reconstructions showed no artefacts (see **Figure 16**). The L^2/TV reconstructions for low contrast and sparse data structures showed difficulties distinguishing data from noise when the noise was strong enough (see **Figure 18** & **Figure 19**). However, in the case where the L^2/TV still showed good reconstructions the standard deviation of the noise was already higher than significant parts of the data. As expected the L^2/TGV$^2_\beta$ reconstruction outperformed the L^2/TV reconstruction when reconstructing a linear data structure (see **Figure 20**), since the L^2/TV model has as underlying assumption that the true image is piecewise constant. The last results shown are on a more realistic data set. This data set was created to have likewise features as real data. These last results, show precisely what all the previous experiments showed, *i.e.* the variational methods clearly outperform the FBP. Moreover the L^2/TGV$^2_\beta$ model shows better results for linear data structures than the L^2/TV model.

The final conclusion of this thesis is that variational methods, such as L^2/TV and L^2/TGV$^2_\beta$, can be implemented in the reconstruction model for photoacoustic tomography and the initial results are promising for improving the quality of the reconstructions.

7 Outlook

All those results are done through generating synthetic data with the same forward operator as used in the reconstruction. This implies that the reconstructions will most likely not show all the artefacts that the mathematical forward operator might introduce. Therefore the next step would be to create the synthetic data with another forward operator, such as the K-WAVE toolbox [TC10], and performing the reconstructions with the models suggested in this paper. This gives the same controlled environment for the experiments, but works around the 'inverse crime'.

In the realistic data results (5.5) it is shown that the current method to solve the outer inverse problem is not optimal. Finding a method that solves this problem more precisely will help the overall reconstructions greatly. As suggested in 3.1 an interesting way to solve this outer problem is trying to solve it simultaneously with the inner problem. Another strategy that might give great results is finding an effecient way to concatenate the outer and inner inverse problem operator, to simplify the inverse problem, if there is a way to do this, the methods for the current inner problem can probably be implemented for this new simpler inverse problem without a lot of change of the current framework.

The suggested models have a well known problem that in practice the choice of regularization parameter is not very straightforward. Therefore it is important to find a consistent way of determining those.

Something not yet discussed in this paper is that in the real data there are problems with the bone reflecting the light back to the detector. This might result in objects shown in the reconstructions that should not be there. A way to handle this is to determine where the bone is located and removing all the data in the sinograms that might be data reflected by the bone and trying to inpaint (interpolate) this data, in a likewise manner as done in [BBLS15].

References

[AV94]   Robert Acar and Curtis R Vogel, *Analysis of bounded variation penalty methods for ill-posed problems*, Inverse problems **10** (1994), no. 6, 1217. 7

[BBBM13]   Martin Benning, Christoph Brune, Martin Burger, and Jahn Müller, *Higher-order TV methods enhancement via Bregman iteration*, Journal of Scientific Computing **54** (2013), no. 2-3, 269–310. 8

[BBLS15]   Martin Benning, Christoph Brune, Marinus Jan Lagerwerf, and Carola-Bibliane Schönlieb, *TGV sinogram inpainting for limited angle tomography*, Tech. report, University of Cambridge, 01 2015. 5, 21, 33

[BKP10]   Kristian Bredies, Karl Kunisch, and Thomas Pock, *Total generalized variation*, SIAM Journal on Imaging Sciences **3** (2010), no. 3, 492–526. 7

[BMPS14]   Martin Burger, Jahn Müller, Evangelos Papoutsellis, and Carola-Bibiane Schönlieb, *Total variation regularization in measurement and image space for PET reconstruction*, Inverse Problems **30** (2014), no. 10, 105003. 5

[Bru10]   Christoph Brune, *4d imaging in tomography and optical nanoscopy*, Ph.D. thesis, University of Münster, 2010. 8

[CKP99]   Eduardo Casas, Karl Kunisch, and Cecilia Pola, *Regularization by functions of bounded variation and applications to image enhancement*, Applied Mathematics and Optimization **40** (1999), no. 2, 229–257. 7

[CP11]   Antonin Chambolle and Thomas Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision **40** (2011), no. 1, 120–145. 10

[Gab83]   Daniel Gabay, *Chapter ix applications of the method of multipliers to variational inequalities*, Studies in mathematics and its applications **15** (1983), 299–331. 10

[KLA+95]   Robert A Kruger, Pingyu Liu, C Robert Appledorn, et al., *Photoacoustic ultrasound reconstruction tomography*, Medical physics **22** (1995), no. 10, 1605–1609. 4

[LP14]   Dirk A Lorenz and Thomas Pock, *An inertial forward-backward algorithm for monotone inclusions*, Journal of Mathematical Imaging and Vision **51** (2014), no. 2, 311–325. 10, 15

[OBG+05]   Stanley Osher, Martin Burger, Donald Goldfarb, Jinjun Xu, and Wotao Yin, *An iterative regularization method for total variation-based image restoration*, Multiscale Modeling & Simulation **4** (2005), no. 2, 460–489. 8, 27

[PC11]   Thomas Pock and Antonin Chambolle, *Diagonal preconditioning for first order primal-dual algorithms in convex optimization*, Computer Vision (ICCV), 2011 IEEE International Conference on, IEEE, 2011, pp. 1762–1769. 10, 14, 15

[Roc15]   Ralph Tyrell Rockafellar, *Convex analysis*, Princeton university press, 2015. 11

[ROF92]   Leonid I Rudin, Stanley Osher, and Emad Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D: Nonlinear Phenomena **60** (1992), no. 1, 259–268. 6, 7

[SJP12]   Emil Y Sidky, Jakob H Jørgensen, and Xiaochuan Pan, *Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle–Pock algorithm*, Physics in medicine and biology **57** (2012), no. 10, 3065. 10, 11

[TC10]   Bradley E Treeby and Benjamin T Cox, *k-wave: Matlab toolbox for the simulation and reconstruction of photoacoustic wave fields*, Journal of biomedical optics **15** (2010), no. 2, 021314–021314. 33

[Tse91]   Paul Tseng, *Applications of a splitting algorithm to decomposition in convex programming and variational inequalities*, SIAM Journal on Control and Optimization **29** (1991), no. 1, 119–138. 10

[vEBM⁺14]   Peter van Es, Samir K Biswas, Hein J Bernelot Moens, Wiendelt Steenbergen, and Srirang Manohar, *Initial results of finger imaging using photoacoustic computed tomography*, Journal of biomedical optics **19** (2014), no. 6, 060501–060501. 1

[WCJ07]   Xueding Wang, David L Chamberland, and David A Jamadar, *Noninvasive photoacoustic tomography of human peripheral joints toward diagnosis of inflammatory arthritis*, Optics letters **32** (2007), no. 20, 3002–3004. 1

[WW12]   Lihong V Wang and Hsin-i Wu, *Biomedical optics: principles and imaging*, John Wiley & Sons, 2012. 3

[WXZC04]   Yi Wang, Da Xing, Yaguang Zeng, and Qun Chen, *Photoacoustic imaging with deconvolution algorithm*, Physics in medicine and biology **49** (2004), no. 14, 3117. 4

[ZC08]   Mingqiang Zhu and Tony Chan, *An efficient primal-dual hybrid gradient algorithm for total variation image restoration*, UCLA CAM Report (2008), 08–34. 10