VARIATIONAL REGULARIZATION STRATEGY FOR ATMOSPHERIC TOMOGRAPHY

Dissertation for the award of degree "Doctor rerum naturalium" (Dr.rer.nat) of the Georg-August-Universität-Göttingen within the doctoral program mathematics of the Georg-August University School of Science (GAUSS)

submitted by

Erdem Altuntac

from Istanbul-Turkey

Göttingen, March 2016

to the memory of *Necdet Altuntac* ...

Everything should be made as simple as possible, but not simpler.

Albert Einstein

Don't only practice your art,
but force your way into its secrets;
art deserves that,
for it and knowledge can raise man to the Divine.

Ludwig van Beethoven

Acknowlegdements

I am indepted to my adviser Prof. Dr. D Russell Luke for giving me the full access to his software SAMSARA for the correction and the design of my three dimensional total variation functional. It is still owing to SAMSARA that reasonable reconstructions were produced. Still my gratitude to Prof. Dr. D. Russell Luke for teaching me how to work on the ideas with strict discipline. His words of wisdom *'State of the art in research is to master what you have in hand first'* will linger on. I am also grateful to him for sharing a scientific moment when we enjoy solar eclipse simply with a pencil and two papers.

I am grateful to my second adviser Prof. Dr. Thorsten Hohage for the strategic discussions on the modelling of the tomography problem and on Chapter 3. His guidance also made me aware of the recent developments in convex variational regularization for inverse ill-posed problems. I acquired great scientific knowledge as well as experience in nested loops in regularization strategies while I was taking his lecture Inverse Problem-III.

Valuable discussions with my colleagues, or even comrades-in-arms, Housen Li, Dr. Henrik Schumacher, Frederic Weidling, Dr. Patrick Neumann have advanced the value of my work and my knowledge. The scientific board of German Weather Service (Deutscher Wetterdienst - DWD) and Prof. Dr. Roland Potthast together with Gerhard Paul deserve all the credits for the commencement of my doctoral studies. I familiarized myself with the engineering background of this special tomographic problem with the help of Dr. Michael Bender.

Last, but not least, very special thanks to Prof. Dr. Ina Kersten and Prof. Dr. Stefan Halverscheid for their support in helping me to get through the challenging moments.

This work is dedicated to the memory of my name-father Nejdet Altuntac whose early passing shocked us deeply.

Contents

vi

List of Figures

Abstract

The main focus of this dissertation is to establish the necessary theory with numerical illustrations for solving an atmospheric tomography problem. The inverse problem is the reconstruction of some volume data from sparsely and non-uniformly distributed tomographic measurement. For a given linear, compact and injective forward operator $\mathcal{T}$ defined between some reflexive Banach space $\mathcal{V}$ and Hilbert space $\mathcal{H}$, $\mathcal{T} : \mathcal{V} \to \mathcal{H}$, we formulate the following convex minimization problem,

$$\varphi_\alpha^\delta \in \arg \min_{\varphi \in \mathcal{V}} F_\alpha(\varphi, f^\delta)$$

with the general Tikhonov functional

$$F_\alpha(\varphi, f^\delta) := \frac{1}{2} ||\mathcal{T}\varphi - f^\delta||_{\mathcal{H}}^2 + \alpha J(\varphi),$$

where the smooth, *i.e.* Fréchet differentiable, and convex penalizer is $J : \mathcal{V} \to \mathbb{R}_+$. We analyse the stable convergence of the expected regularized solution to the true solution by *a posteriori* strategy for the choice of the regularization parameter, $\alpha = \alpha(\delta, f^\delta)$. Of particular interest in terms of the penalty term *J*, we choose *smooth total variation* functional

$$J_\beta^{TV}(\varphi) := \int_\Omega \sqrt{|\nabla \varphi(x)|_2^2 + \beta} dx,$$

where $0 < \beta < 1$ and $| \cdot |_2$ is the usual Euclidean norm. We estimate a new lower bound for the Bregman distance particularly associated with the penalty term J_β^{TV}. We further investigate the impact of TV regularization on inverse ill-posed problems and convey the phenomenon of loss of contrast in TV regularization.

We demonstrate our regularization on simulated data, employing a novel reverse-communication large-scale nonlinear optimization software and also compare the result against traditional algorithms.

Chapter 1

INTRODUCTION

1.1 Summary of This Dissertation

The aim of this work is to develop the theory and numerical methods for an atmospheric tomography problem called *GPS-tomography*, [**15, 67, 93**]. GPS-tomography involves the reconstruction of some quantity, pointwise within a volume (e.g. humidity) from measurements transmitted by non-uniformly distributed transducers (satellites). These measurements are collected by non-uniformly distributed receivers on the ground (ground stations) as illustrated in Figure 1.2. As with conventional tomography, the task here is the reconstruction of the density profile of a layer in the atmosphere from a set of line integrals.

In the discretized form of the problem, it is assumed that each station receives equal number of signals transmitted by the satellites. Also for the sake of simplicity, we ignore any deviations from the shortest path between transmitters and receivers due to atmospheric inhomogeneity. The received signal is then modelled as a line integral along the shortest path between the satellites and the ground stations.

We study the variational formulation of the inverse problem: minimizing a least squares functional with twice continuously Fréchet differentiable convex penalty. We are tasked with reconstructing some volume data from sparsely distributed tomographic measurement. Since it is known that total variation regularizer promotes sparsity, [**16, p. 2**], we make a particular choice of the penalty which is smooth total variation penalty. Choice of a proper regularizer in variational regularization depending on the smoothness of the target function has been discussed in [**16, p. 2**]. In our numerical illustrations, we have simulated a data with smooth intensity. The regularizer of choice is total variation, or rather a smooth approximation thereof. By this approach, we also introduce new way of solving this large scale problem.

Organization of this work is as follows;

- **Chapter 1:** In the Section 1.2 of this chapter, we model our specific tomography problem with discussing data collection procedure.

- **Chapter 2:** In this chapter, we introduce the foundations of the work that are the base of entire mathematical development. In Subsection 2.3.2, we introduce Bregman distance as the necessary tool for our convergence theory. We express the Bregman distance as the remainder term of the 2nd order Taylor expansion. By this expression, Hessian of twice continuously Fréchet differentiable functional will actually be the Bregman distance associated with the functional itself.

- **Chapter 3:** In this chapter, we emphasize the importance of variational inequalities in order to be able to show the convergence convex minimization problem. We give a special choice of a concave index function that has been introduced in [**51, Eq (3.2)**]. We discuss *a posteriori* strategy for the choice of the regularization parameter $\alpha = \alpha(\delta, f^\delta)$. Under this *a posteriori* strategy and the considered deterministic noise model, $f^\delta \in \mathcal{B}_\delta(f^\dagger)$, in the measurement space, the following rates will be able to be quantified;

  1. $\mathcal{T}\varphi^\delta_{\alpha(\delta,f^\delta)} \in \mathcal{B}_{\mathcal{O}(\delta)}(\mathcal{T}\varphi^\dagger)$; norm of the discrepancy between $\mathcal{T}\varphi^\delta_{\alpha(\delta,f^\delta)}$ and $\mathcal{T}\varphi^\dagger$ by the rate of $\mathcal{O}(\delta)$, *i.e.* $\|\mathcal{T}\varphi^\delta_{\alpha(\delta,f^\delta)} - \mathcal{T}\varphi^\dagger\|_{\mathcal{L}^2(\mathcal{Z})} = \mathcal{O}(\delta)$.
  2. $D_J(\varphi^\delta_{\alpha(\delta,f^\delta)}, \varphi^\dagger) = \mathcal{O}(\Psi(\delta))$; upper bound for the Bregman distance D_J.
  3. $\varphi^\delta_{\alpha(\delta,f^\delta)} \in \mathcal{B}_{\mathcal{O}(\Psi(\delta))}(\varphi^\dagger)$; convergence of the regularized solution $\varphi^\delta_{\alpha(\delta,f^\delta)}$ to the true solution $\varphi^\dagger$ by the rate of the noise amount $\mathcal{O}(\Psi(\delta))$.

  In Section 3.4, we finally give necessary interpretation of the convex regularization theory for our specific choice of the regularizer smoothed version of the total variation. Particularly in Subsection 3.4.1, we summarise our new convergence theory for the smoothed total variation regularization in one chain of inequalities.

- **Chapter 4:** This chapter is only devoted for numerical illustration. After reviewing fundamental knowledge about the gradient based algorithms, we demonstrate our illustrations in different algorithms. Regularization as a result of *limited memory BFGS (L-BFGS)* on simulated data is demonstrated by employing a novel reverse-communication large-scale nonlinear optimization software SAMSARA, [**64**]. In this software two popular penalty terms that are of quadratic Tikhonov and smoothed-TV are tested.

1.2 Physical Problem: From Propagation in Time to Propagation in Space

Let γ be the minimal path between any two points in $\mathbb{R}^3$. Then a light beam takes the least time T with speed c along this path

$$T = \frac{1}{c} \int_\gamma n(\mathbf{r}) d\mathbf{r}, \tag{1.1}$$

where n is index of refraction. The linear relation between the refractivity profile N and the refractive index n is expressed by $N = 10^6(n-1)$, [15, 67, 93]. Thus, if one chooses the refractivity profile as the frame of reference, then (1.1) reads,

$$T = \frac{1}{c} \int_\gamma (10^{-6} N(\mathbf{r}) + 1) d\mathbf{r}.$$

The task here is to reconstruct the density function from some noisy travel time tomographic measurement which is provided by the set of line integrals over the signals between non-uniformly and sparsely located transducers and again non-uniformly and sparsely located ground stations.

We now introduce a ground-based model for the physical problem with its geometry. Denote by g some Lipschitz continuous function with its Lipschitz constant $L_g \in \mathbb{R}_+$ for the surface of the earth,

$$g : [0, S] \times [0, P] \to [0, h^\infty), \tag{1.2}$$

and denote by $\mathcal{G}$ the graph of the surface function g

$$\mathcal{G} = \text{graph}(g) = \{(x, y, z) \mid (x, y) \in [0, S] \times [0, P], z = g(x, y) \geq 0\}. \tag{1.3}$$

Then a ground station $\mathbf{s}$ is a set of points in $\mathbb{R}^3$ located on earth with the coordinate points (x_s, y_s, z_s),

$$\mathbf{s} := (x_s, y_s, z_s) \in \mathcal{G}, \tag{1.4}$$

and likewise emitters $\mathbf{e}$ that are all located at the same altitude h^∞ is also set of points in $\mathbb{R}^3$,

$$\mathbf{e} := \{(x_e, y_e, z_e) \mid (x_e, y_e) \in [0, S] \times [0, P] \text{ and } z_e = h^\infty\}. \tag{1.5}$$

Let (ρ_s, σ_s) be the polar angles of the station **s** as inclination and azimuth respectively. Then in spherical coordinates, the location of any station **s** is given by

$$\mathbf{s} = R \left(\cos(\rho_s)\cos(\sigma_s), \cos(\rho_s)\sin(\sigma_s), \sin(\rho_s) \right)^T,$$

where $R = \sqrt{x_s^2 + y_s^2 + z_s^2}$. Our area of interest is a compact subdomain, *i.e.* $\Omega \subset \Omega_o$,

$$\Omega_o := \{(x, y, z) \in [0, S] \times [0, P] \times [0, h^\infty) \mid z > g(x, y)\}. \tag{1.6}$$

The unknown density function $\varphi : \Omega \to \mathbb{R}$ is integrable and, by assumption, vanishes outside the area of interest Ω. This is explained by introducing a step function as such

$$\tilde{\varphi}(x) := \begin{cases} \varphi(x) & , \text{ for } x \in \Omega \\ 0 & , \text{ for } x \in \Omega_o \backslash \Omega. \end{cases} \tag{1.7}$$

Following [**68, Ch. 2**] and [**72, p. 45**], the ray direction $\vec{\theta}$ is reparametrized by

$$\vec{\theta} = - \left(\cos(\rho_r + \rho_s)\cos(\sigma_r + \sigma_s), \cos(\rho_r + \rho_s)\sin(\sigma_r + \sigma_s), \sin(\rho_r + \rho_s) \right)^T,$$

where the inclination and the azimuth of the ray $\gamma_{[\mathbf{s}, \vec{\theta}]} : [z_s, h^\infty] \to \Omega_o \subset \mathbb{R}^3$ according to the surface are denoted by $(\rho_r + \rho_s, \sigma_r + \sigma_s) = (\tilde{\rho}, \tilde{\sigma})$, see Figure 1.1 for this angular parameterization. We present the domain of any integrated measurement f through some ray $\gamma_{[\mathbf{s}, \vec{\theta}]}$ by

$$f : \mathcal{Z} \subset \mathcal{Z}_o \subset \mathcal{G} \times \mathbb{S}^2 \to \mathbb{R}_+, \tag{1.8}$$

where the maximal set of measurement is defined by

$$\mathcal{Z}_o := \{(\mathbf{s}, \vec{\theta}) \mid \tilde{\rho} \geq |\arctan(L_g)| \text{, and } \pi - \tilde{\rho} > 0\}.$$

Physically, there exist many rays in various directions $\vec{\theta} \in \mathbb{S}^2$. Denote by

$$\mathbb{S}_\mathbf{s} := \{\vec{\theta} \in \mathbb{S}^2 | (\mathbf{s}, \vec{\theta}) \in \mathcal{Z}\}, \tag{1.9}$$

the set of directions. Then, in fact, the measured data f is obtained only for $\vec{\theta} \in \mathbb{S}_\mathbf{s}$. We must also denote the directions that do not necessarily belong to the set of directions $\mathbb{S}_\mathbf{s}$ by $\vec{\theta}_o \in \mathbb{S}^2 \backslash \mathbb{S}_\mathbf{s}$. Note that $\mathbb{S}_\mathbf{s} \subset \mathbb{S}^2$ which is the partial information case.

5

Thus, there could be formulated a parameter function $t(\epsilon) := \frac{\epsilon}{\sin(\tilde{\rho})}$ such that a ray in $\mathbb{R}^3$ starting from the station $\mathbf{s}$ in the direction $\vec{\theta} \in \mathbb{S}_{\mathbf{s}}$ is defined by

$$\gamma_{[\mathbf{s},\vec{\theta}]}(\epsilon) := \left\{ \mathbf{s} + \frac{\epsilon}{\sin(\tilde{\rho})}\vec{\theta} \,\middle|\, \epsilon \in [z_s, h^\infty] \right\}. \tag{1.10}$$

It is well known that the incompleteness of data causes nonuniqueness issue in inverse problems. See [**70, p. 144**] for tomographic interpretation. The assumption of that Ω is a strict subset of Ω_o is essential for the uniqueness of the inverse problem.

Assumption 1.2.1. *[**Geometric assumption**] Area of interest $\Omega \subset \Omega_o$ is chosen such that, there exists a relatively open subset $\tilde{\mathcal{G}} \subset \mathcal{G}$ such that*

$$\mathfrak{C} := \bigcap_{\mathbf{s} \in \mathcal{G}} \mathbb{S}_{\mathbf{s}} \subset \mathbb{S}^2,$$

is open, and for each $\vec{\theta}_o \in \mathfrak{C}$ there exists $\mathbf{s}_o \in \mathcal{G}$ such that $(\mathbf{s}_o, \vec{\theta}_o) \in \mathcal{Z}$ with

$$\gamma_{[\mathbf{s}_o, \vec{\theta}_o]} \cap \Omega = \emptyset. \tag{1.11}$$

To obtain measurement f, we apply fan-beam projection operator along the ray $\gamma_{[\mathbf{s},\vec{\theta}]}$ on some density profile defined by $\varphi := \frac{1}{c}(10^{-6}N + 1) = \frac{n}{c}$. This operation is formulated by

$$f(\mathbf{s}, \vec{\theta}) = \mathcal{T}_{\mathbf{s}}\varphi(\vec{\theta}) = (\mathcal{T}\varphi)(\mathbf{s}, \vec{\theta}) = \int_{\gamma_{[\mathbf{s},\vec{\theta}]}} \varphi(\mathbf{r})d\mathbf{r}, \text{ for } \mathbf{r} \in \gamma_{[\mathbf{s},\vec{\theta}]} \subset \Omega_o. \tag{1.12}$$

Also, with the angular parameterization above, we then have,

$$\mathcal{T}_{\mathbf{s}}\varphi(\vec{\theta}) = \mathcal{T}\varphi(\tilde{\rho}, \tilde{\sigma}), \text{ for } (\tilde{\rho}, \tilde{\sigma}) \in (0, \pi) \times (0, 2\pi).$$

The discretized integration from one point to the next one along the ray γ is carried out via the parameter function $t(\epsilon) = \frac{\epsilon}{\sin(\tilde{\rho})}$, for all $\epsilon \in [z_s, h^\infty]$, see Figure 1.3. In the continuum form, we use ray transform in the direction $\vec{\theta}(\tilde{\rho}, \tilde{\sigma}) \in \mathbb{S}_{\mathbf{s}}$ for any angle pairs $(\tilde{\rho}, \tilde{\sigma})$, on the density function $\varphi : \Omega_o \to \mathbb{R}_+$ where $\mathbf{r} \in \gamma_{[\mathbf{s},\vec{\theta}]}$, as such

$$f(\mathbf{s}, \vec{\theta}) = \mathcal{T}_{\mathbf{s}}\varphi(\vec{\theta}) = (\mathcal{T}\varphi)(\mathbf{s}, \vec{\theta}) = \int_{\mathbf{r} \in \gamma_{[\mathbf{s},\vec{\theta}]}} \varphi(\mathbf{r})d\mathbf{r} = \int_{\mathbf{z_s}}^{\mathbf{h}^\infty} \varphi(\gamma_{[\mathbf{s},\vec{\theta}]}(\epsilon))|\gamma'_{[\mathbf{s},\vec{\theta}]}(\epsilon)|\mathbf{d}\epsilon. \tag{1.13}$$

6

So as a linear operator equation, we have $\mathcal{T}\varphi = f$ where $\mathcal{T}$ represents the line integration operating on the density profile φ to obtain measurement f.

We, below, only give the injectivity of the forward operator for the verification of uniqueness principle.

Theorem 1.2.2. [Injectivity of the forward operator $\mathcal{T}$] [45, Theorems 5.1 - 5.6] and [70, Theorem 6.2]
  Under the Assumption 1.2.1, the operator $\mathcal{T}$ is injective.

It is not possible to reconstruct the unknown function φ exactly from finite of measurements. However, [**45, Theorems 5.1 - 5.6**] show that arbitrarily good approximation can be obtained, (cf. [**45, Section 6.**]).
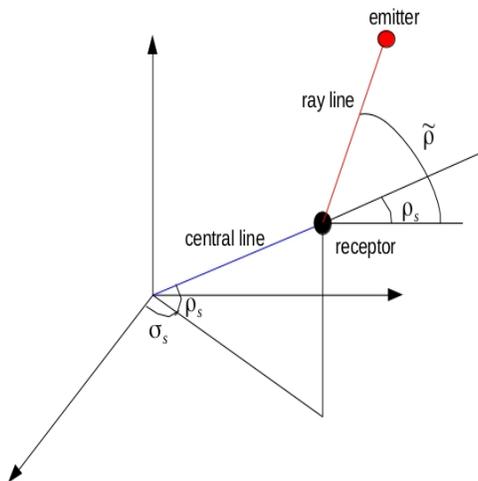
Figure 1.1: The sketch of angular parameterization. Intersection point between the central and the ray lines is the ground station with the angles (ρ_s, σ_s).
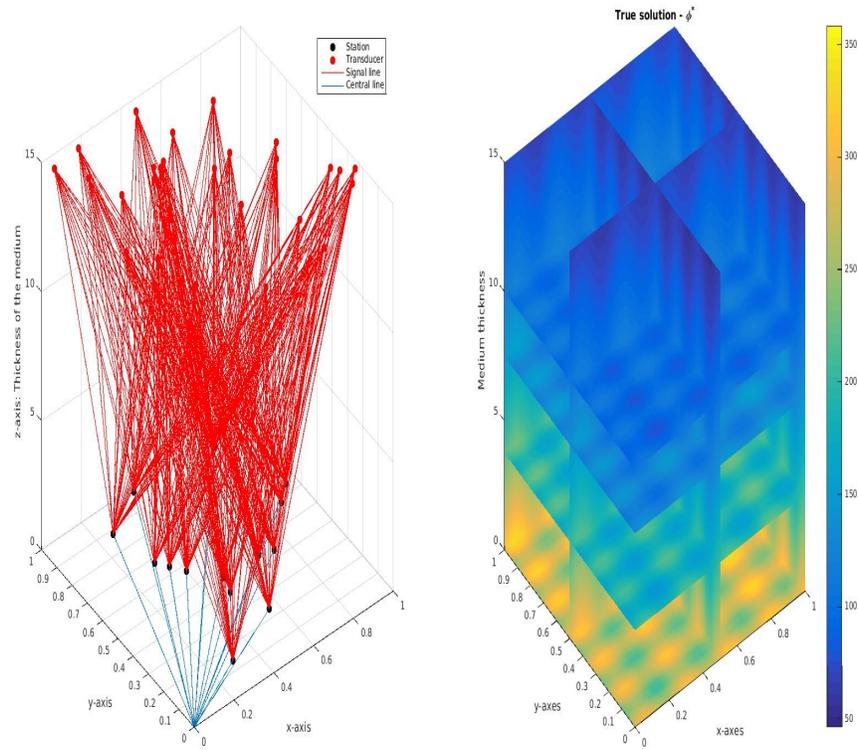
Figure 1.2: A 3-D network together with simulated data φ, as the true solution $\varphi^\dagger$, is illustrated over a non-uniformly scaled domain. Black dots indicate stations whilst signals penetrate the area of interest through red dots. In this illustration, 15 ground stations (receiver) intercept signals emitted by 30 transducers and all are randomly distributed over 27000 cubes.
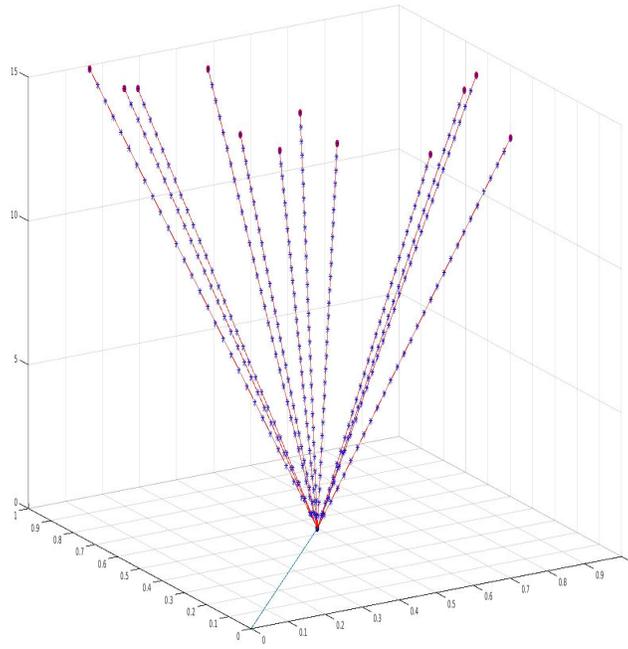
Figure 1.3: Finite number of points denoted by blue stars along 12 rays are illustrated. These points are found via the parameter function $t(\epsilon) = \frac{\epsilon}{\sin(\bar{\rho})}$, for $\epsilon \in [z_s, h^\infty]$ where h^∞ is the upper bound of the line integral in (1.13).

Chapter 2

PREREQUISITE KNOWLEDGE AND PRELIMINARY RESULTS

We present the foundations of this work which are the base for the entire mathematical development.

2.1 Vector Calculus and Functional Analysis Notations

It is our task to reconstruct a non-negative, differentiable, scalar function defined on a compact and convex subset Ω of $\mathbb{R}^3$, *i.e.* $\varphi(\mathbf{x}) : \Omega \subset \mathbb{R}^3 \to \mathbb{R}_+$ where the spatial coordinate is $\mathbf{x} = (x, y, z)$. By our specific choice of the regularization strategy, we must review basics from the vector calculus. The gradient of a scalar valued function φ is regarded as a vector with the components

$$\nabla \varphi = \left(\begin{array}{ccc} \frac{\partial \varphi}{\partial x}, & \frac{\partial \varphi}{\partial y}, & \frac{\partial \varphi}{\partial z} \end{array} \right)^T,$$

and the magnitude of the gradient in the Euclidean sense is then given by

$$|\nabla \varphi|_2 = \left(\left| \frac{\partial \varphi}{\partial x} \right|^2 + \left| \frac{\partial \varphi}{\partial y} \right|^2 + \left| \frac{\partial \varphi}{\partial z} \right|^2 \right)^{1/2}. \tag{2.1}$$

Divergence of a vector valued mapping $\Phi(\mathbf{x}) : \Omega \subset \mathbb{R}^3 \to \mathbb{R}^3$ is also given by

$$\mathrm{div}\Phi = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}) \cdot (\Phi_x, \Phi_y, \Phi_z) = \frac{\partial \Phi_x}{\partial x} + \frac{\partial \Phi_y}{\partial y} + \frac{\partial \Phi_z}{\partial z}.$$

We denote by $\mathcal{L}^p(\Omega)$ the class of all measurable functions φ, defined on $\Omega \subset \mathbb{R}^3$, for which

$$||\varphi||_{\mathcal{L}^p(\Omega)} := \left(\int_\Omega |\varphi(\mathbf{x})|^p d\mathbf{x} \right)^{1/p} < \infty. \tag{2.2}$$

$\mathcal{L}^p(\Omega)$ are complete normed spaces, *i.e.* Banach spaces. For $p \in (1, \infty)$, they are *reflexive Banach spaces*, [**62, p. 242**]. General Hilbert spaces will be denoted throughout by $\mathcal{H}$, in particular $\mathcal{H} = \mathcal{L}^2(\Omega)$. Different from usual $\mathcal{L}^2$ space, we need to define weighted form of it as we will need establish some convergence theory in the weighted Hilbert space. Let the positive definite $w : \Omega \to \mathbb{R}_+$ be weight functional. Then, we denote weighted $\mathcal{L}^2(\Omega)$ by $\mathcal{L}^2(\Omega, w) = \mathcal{L}^2_w(\Omega)$, which is endowed with the norm,

$$||\varphi||_{\mathcal{L}^2_w(\Omega)} := \left(\int_\Omega |\varphi(x)|^2 w(x) dx \right)^{1/2} < \infty. \tag{2.3}$$

Let $\mathcal{C}(\Omega)$ be the space of continuous functions that are defined on the compact domain Ω with its Lipschitz boundary $\partial\Omega$, denote by D^σ, where $|\sigma| \le k$ for $k \ge 1$, the generalized derivative [**92, Definitions 21.1-2 of Section 21.1**]. Then, the space of continuously differentiable functions $\mathcal{C}^k(\Omega)$ is defined by

$$\mathcal{C}^k(\Omega) := \{\varphi \in \mathcal{C}(\Omega) | D^\sigma(\varphi) \in \mathcal{C}(\Omega), \forall \sigma \in \mathbb{N} \text{ with } |\sigma| \le k\}.$$

Recall from the subsection 2.1 that the target function φ is differentiable which is in other words φ is $\mathcal{C}^1(\Omega)$ smooth. We also denote the space of compactly supported functions by

$$\mathcal{C}_c(\Omega) := \{\varphi | \text{ supp}(\varphi) \text{ is compact }\}.$$

We will also need to work with Sobolev spaces, [**2**]. Sobolev space, for $p \ge 1$ and $k \ge 1$, is defined by,

$$\mathcal{W}^{k,p}(\Omega) := \{\varphi \in \mathcal{C}(\Omega) | D^\sigma(\varphi) \in \mathcal{L}^p(\Omega), \forall \sigma \in \mathbb{N} \text{ with } |\sigma| \le k\}.$$

Additionally, functions of Sobolev space with zero boundary value is defined by

$$\mathcal{W}^{k,p}_0(\Omega) := \{\varphi \in \mathcal{C}(\Omega) | D^\sigma(\varphi) \in \mathcal{L}^p(\Omega), \forall \sigma \in \mathbb{N} \text{ with } |\sigma| \le k, \text{ and } \varphi(x) = 0 \text{ for } x \in \partial\Omega\}.$$

It is now worthwhile to recall the density argument, [**37, Subsection 5.2.2**], in $\mathcal{W}^{k,p}(\Omega)$,

$$\overline{C_c^\infty(\Omega)} = \mathcal{W}^{k,p}_0(\Omega).$$

Specifically, it is also known that $\overline{C_c^1(\Omega)} = \mathcal{W}^{1,2}(\Omega)$, see [**92, p. 240-241**] for details. Furthermore, for the functions $\varphi \in \mathcal{W}^{1,1}(\Omega)$ a generalized form of Poincaré inequality states a relationship between the spaces $\mathcal{W}^{1,1}$ and $\mathcal{L}^1$.

11

Theorem 2.1.1. *[Poincaré - Wirtinger Inequality][18, Theorem 3.1] Over the compact domain Ω and with its volume $|\Omega|$, there exists a constant C_Ω such that for any $\varphi \in \mathcal{W}^{1,1}(\Omega)$,*

$$\left\| \varphi - \frac{1}{|\Omega|} \int_\Omega \varphi(x)dx \right\|_{\mathcal{L}^1(\Omega)} \leq C_\Omega \|\nabla\varphi\|_{\mathcal{L}^1(\Omega)}. \tag{2.4}$$

To find the unique solution for our specific application problem, we formulate a convex minimization problem. This formulation will be introduced in the following subsection. It is this work's aim to find the solution to the minimization problem in the class of *Bounded Variation* (BV) space. We now review some basics about BV space.

2.2 Functions of Bounded Variation Space

In order to achieve a reconstruction in $\mathcal{L}^1$ space, we focus on the TV type regularization and its relation with BV space functions. As has been stated by the problem formulation, we aim to reconstruct a function in a Banach space from sparse tomographic data. It has been conveyed that the use of *TV* promotes sparsity of the gradient, [16], which is one of the scientific motivations of this work that we focus on the total variation (TV) of a a function. In our numerical illustrations, we have simulated a data with smooth intensity, [16]. It is another reason why we choose to formulate a smooth minimization problem which will be studied in the following section. TV of a function that is defined over the compact domain Ω is given below.

Definition 2.2.1. *[$TV(\varphi, \Omega)$][6, Definition 3.4], [79, Definition 9.64], [86, Definition 8.9] Over the compact domain Ω, total variation of a function $\varphi \in \mathcal{L}^1(\Omega)$ is defined in the weak sense as follows ,*

$$TV(\varphi, \Omega) := \sup_{\Phi \in \mathcal{C}_c^1(\Omega;\mathbb{R}^3)} \left\{ \int_\Omega \varphi(x)\mathrm{div}\ \Phi(x)dx\ :\ ||\Phi||_\infty \leq 1 \right\}. \tag{2.5}$$

Total variation type regularization targets the reconstruction of bounded variation (BV) class of functions that are defined by

$$BV(\Omega) := \{ \varphi \in \mathcal{L}^1(\Omega) : TV(\varphi, \Omega) < \infty \} \tag{2.6}$$

with the norm

$$||\varphi||_{BV} := ||\varphi||_{\mathcal{L}^1} + TV(\varphi, \Omega). \tag{2.7}$$

BV function spaces are Banach spaces, [86]. Furthermore, the Sobolev space $\mathcal{W}^{1,1}$ is also a subspace of $BV(\Omega)$, [1, Theorem 2.1]. This fact is formulated below.

12

Proposition 2.2.2. *[$\mathcal{W}^{1,1}(\Omega)$ is a subspace of $BV(\Omega)$][**86, Proposition 8.13**] If $\varphi \in \mathcal{W}^{1,1}(\Omega)$, then*

$$TV(\varphi) = \int_\Omega |\nabla\varphi(x)|_2 dx < \infty.$$

By the result in [**1, Theorem 2.1**], it is known that one can arrive, with a proper choice of $\Phi \in \mathcal{C}_c^1(\Omega)$, at the following from (2.5),

$$TV(\varphi) = \int_\Omega |\nabla\varphi(x)|_2 dx \cong \int_\Omega \left(|\nabla\varphi(x)|_2^2 + \beta \right)^{1/2} dx, \qquad (2.8)$$

where $0 < \beta < 1$ is fixed. We also refer [**24, 27, 33, 77, 87**] where (2.8) has appeared. By Definition 2.2.1, it is easy to obtain another TV form for the computerized environment, see [**1, Proof of Theorem 2.1**]. Observe that, by Gauss' divergence theorem [**91, Eq. 22**] , with the Neumann boundary condition $\frac{\partial\varphi}{\partial\eta}|_{\partial\Omega} = 0$,

$$\int_\Omega \varphi(x)\mathrm{div}\,\Phi(x)dx = \int_{\partial\Omega} \frac{\partial\varphi(x)}{\partial\eta}\Phi(x)dS - \int_\Omega \nabla\varphi(x)\Phi(x)dx = \int_\Omega \nabla\varphi(x)(-\Phi(x))dx.$$

Again by Gauss' divergence theorem and the Cauchy-Schwarz inequality,

$$\int_\Omega \varphi(x)\mathrm{div}\,\Phi(x)dx = -\int_\Omega \nabla\varphi(x)\Phi(x)dx \leq \left| \int_\Omega \nabla\varphi(x)\Phi(x)dx \right| \leq \int_\Omega |\nabla\varphi(x)|_2|\Phi(x)|_2 dx.$$

Now taking the supremum of the both sides will yield

$$\sup_{\Phi \in \mathcal{C}_c^1(\Omega;\mathbb{R}^3)} \left\{ \int_\Omega \varphi(x)\mathrm{div}\,\Phi(x)dx \right\} \leq \sup_{\Phi \in \mathcal{C}_c^1(\Omega;\mathbb{R}^3)} \left\{ \int_\Omega |\nabla\varphi(x)|_2|\Phi(x)|_2 dx \right\} \leq \int_\Omega |\nabla\varphi(x)|_2 dx,$$

since $||\Phi||_\infty \leq 1$. Now, let us define the direction as $\Phi := -\frac{\nabla\varphi}{|\nabla\varphi|_2}$. Then,

$$\int_\Omega \varphi(x)\mathrm{div}\,\Phi(x)dx = \int_\Omega \nabla\varphi(x)\frac{\nabla\varphi(x)}{|\nabla\varphi(x)|_2}dx = \int_\Omega |\nabla\varphi(x)|_2 dx.$$

Thereof, throughout this work, we will define TV of a function $\varphi \in \mathcal{C}^1(\Omega)$ by,

$$TV(\varphi, \Omega) := \int_\Omega |\nabla\varphi(x)|_2 dx. \qquad (2.9)$$

In the following sections, we will formulate a convex, unconstrained minimization problem associated with the smoothed form of (2.9). For a moment, let us define this smooth TV by

$$TV_\beta(\varphi, \Omega) := \int_\Omega \sqrt{|\nabla\varphi(x)|_2^2 + \beta}dx.$$

It has been shown in [**1, Theorem 2.2**] that for Ω bounded and for any $\varphi \in BV(\Omega)$

$$\lim_{\beta\to 0}TV_\beta(\varphi, \Omega) = TV(\varphi, \Omega). \tag{2.10}$$

This limit with the smoothing parameter $\beta \to 0$, smooth TV maintains the same properties as nonsmooth TV. Moreover, still for Ω bounded and for any $\varphi \in BV(\Omega)$, the following relation holds,

$$TV(\varphi, \Omega) \le TV_\beta(\varphi, \Omega) \le TV(\varphi, \Omega) + \sqrt{\beta}|\Omega|. \tag{2.11}$$

2.3 General Regularization Theory

Henceforth, we will be working with Fréchet differentiable functionals. Fréchet derivative is a derivative defined between normed spaces. In compliance with our theoretical establishment, we formally formulate the Fréchet derivative between Banach spaces below.

Definition 2.3.1. *[**Fréchet Derivative**] Let $\mathcal{V}$ and $\mathcal{Y}$ be Banach spaces. The Fréchet derivative of an operator $\mathcal{P} : \mathcal{V} \to \mathcal{Y}$ at the point $a \in \mathcal{V}$ is a bounded linear operator $\mathcal{P}'[a] : \mathcal{V} \to \mathcal{Y}$ satisfying the following limit relation,*

$$\lim_{h\to 0}\frac{\|\mathcal{P}(a + h) - \mathcal{P}(a) - \mathcal{P}'[a]\cdot(h)\|}{\|h\|} = 0.$$

Remark 2.3.2. *[[**28, Theorem 3.5-3**], [**86, Definition 2.32**], [**91, Theorem 2.E**]] According to Riesz representation theorem, if $\mathcal{V}$ is a Hilbert space, then for a smooth functional $\mathcal{P} : \mathcal{V} \to \mathbb{R} \cup \{\infty\}$, then we can represent the Fréchet derivative in the following form,*

$$\mathcal{P}'[u^*]\cdot(u - u^*) = \langle\nabla\mathcal{P}(u^*), u - u^*\rangle,$$

with some $\nabla\mathcal{P}(u^*) \in \mathcal{V}$ called the gradient of $\mathcal{P}$ at u^*.

14

As a restriction of the Radon transform our forward operator $\mathcal{T} : \mathcal{D}(\mathcal{T}) \subset \mathcal{V} \to \mathcal{H}$ from a Hilbert space $\mathcal{V}$ to a Hilbert space $\mathcal{H}$ is compact, and by Theorem 1.2.2 it is injective. Recall from the Subsection 1.2 that the image space of our forward operator is in fact $\mathcal{H} = \mathcal{L}^2(\mathcal{Z})$ where $\mathcal{Z} \subset \mathcal{G} \times \mathbb{S}_{\mathbf{s}}$, see (1.3) and (1.9) for the details. As well known in the theory of integral equations (cf. [**44, Section 3**]), one may also introduce generalized solutions of a linear operator equation $\mathcal{T}\varphi = f^\delta$, such that $f^\delta \in \mathcal{R}(\mathcal{T}) \oplus \mathcal{R}(T)^\perp$, by formulating the following least squares minimization problem

$$\varphi^\delta \in \arg\min_{\varphi \in \mathcal{V}} \frac{1}{2}||\mathcal{T}\varphi - f^\delta||^2_{\mathcal{L}^2(\mathcal{Z})}. \tag{2.12}$$

Then the generalized minimizer is given by $\varphi^\delta = \mathcal{T}^\dagger f^\delta$ where $\mathcal{T}^\dagger := (\mathcal{T}^*\mathcal{T})^{-1}\mathcal{T}^*$ denotes the Moore-Penrose inverse. Let $\{\mu_j; u_j, v_j\}_{j=1}^\infty$ be a singular system for our forward operator $\mathcal{T}$. The singular values of $\mathcal{T}$, for each $j \in \mathbb{N}$, are ordered from largest to smallest satisfying,

$$\mu_1 \geq \mu_2 \geq \cdots \geq \mu_j \geq \cdots \geq 0.$$

The SVD representation of the pseudoinverse, (cf. [**35, Theorem 2.8**] and [**44, Eq. (9)**]) is given by

$$\mathcal{T}^\dagger f = \sum_{j=1}^\infty \frac{1}{\mu_j} \langle f, v_j \rangle u_j \ , \text{ for all } f \in \mathcal{D}(\mathcal{T}^\dagger). \tag{2.13}$$

Rapid decay of the singular values μ_j causes instability of the solution to the problem (2.12). Therefore, in order to avoid this instability, we include a penalty term J to the problem (2.12) and formulate our main variational minimization problem,

$$\varphi^\delta_\alpha \in \arg\min_{\varphi \in \mathcal{V}} \left\{ \frac{1}{2}||\mathcal{T}\varphi - f^\delta||^2_{\mathcal{L}^2(\mathcal{Z})} + \alpha J(\varphi) \right\}. \tag{2.14}$$

Here, $J : \mathcal{V} \to \mathbb{R}_+ \cup \{\infty\}$ is the lower semi-continuous, Fréchet differentiable and convex penalty term, and the regularization parameter is denoted by $\alpha > 0$. Of particular interest to this work in terms of the penalty term, we choose a smooth functional J. Discussion over the choice of penalty term in variational regularization has been made in [**16, p. 2**].

Solving (2.14) amounts to finding an approximation of the unbounded Moore-Penrose inverse operator $\mathcal{T}^\dagger : \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp \to \mathcal{V}$ by a family of operator $R_\alpha : \mathcal{H} \to \mathcal{V}$ parameterized by $\alpha > 0$ which is named as *regularization operator*, see [**29, Definition 4.3**] and [**63, Theorem 2.2**].

15

We prove convergence of the considered regularization strategy by showing that some regularization strategy is admissible. The definition of an admissible regularization strategy in the general sense is given below.

Definition 2.3.3. *[59, Definition 2.3] Consider a compact forward operator defined from a reflexive Banach space $\mathcal{V}$ to some Hilbert space $\mathcal{H}$, $\mathcal{T} : \mathcal{V} \to \mathcal{H}$. A regularization strategy, consisting of a family of operator $R_\alpha : \mathcal{H} \to \mathcal{V}$ parameterized by $\alpha > 0$ and a parameter choice strategy $\underline{\alpha} : \mathbb{R}_+ \times \mathcal{H} \to \mathbb{R}_+$, is called admissible if*

$$\sup \left\{ ||R_{\underline{\alpha}(\delta,f^\delta)} f^\delta - \varphi^\dagger||_{\mathcal{V}} : f^\delta \in \mathcal{H}, ||\mathcal{T}\varphi^\dagger - f^\delta||_{\mathcal{H}} \leq \delta \right\} \to 0, \ \ as \ \delta \to 0,$$

for every $\varphi^\dagger \in \mathcal{V}$.

2.3.1 First order optimality condition

It is a fact that the regularized solution φ_α^δ for the problem (2.14) satisfies the *first order optimality condition*. As has been formulated by the problem (2.14), we seek minimizer for the objective functional given below,

$$
\begin{aligned}
F_\alpha(\varphi, f^\delta) : \mathcal{V} \times \mathcal{L}^2(\mathcal{Z}) &\longrightarrow \mathbb{R}_+, \\
(\varphi, f^\delta) &\longmapsto F_\alpha(\varphi, f^\delta) := \frac{1}{2}||\mathcal{T}\varphi - f^\delta||_{\mathcal{L}^2(\mathcal{Z})}^2 + \alpha J(\varphi).
\end{aligned}
$$

The Fréchet derivative of the functional F_α attains zero at the regularized solution φ_α^δ, for some $\alpha > 0$. This fact, from the Fréchet derivative, is given by

$$F_\alpha'[\varphi_\alpha^\delta] \cdot (\varphi - \varphi_\alpha^\delta) = 0 \ , \text{ for all } \varphi \in \mathcal{V}.$$

If $\mathcal{V}$ is a Hilbert space, then in the light of Remark 2.3.2, for all $\varphi \in \mathcal{V}$, this can represented by

$$F_\alpha'[\varphi_\alpha^\delta] \cdot (\varphi - \varphi_\alpha^\delta) = \langle \nabla F_\alpha(\varphi_\alpha^\delta), \varphi - \varphi_\alpha^\delta \rangle = 0.$$

From here the first order optimality condition is derived as such

$$\langle \mathcal{T}^*(f^\delta - \mathcal{T}\varphi_\alpha^\delta), \varphi - \varphi_\alpha^\delta \rangle = \alpha \langle \nabla J(\varphi_\alpha^\delta), \varphi - \varphi_\alpha^\delta \rangle, \text{ for all } \varphi \in \mathcal{V}. \tag{2.15}$$

2.3.2   Convergence analysis and a preliminary result

In this work, the norm convergence estimations will be obtained by means of the Bregman distance associated with the smooth potential J, [22]. Convergence of the regularized solution φ_α^δ for the problem (2.14) to the true solution $\varphi^\dagger$ requires one to define a rule for the choice of the regularization parameter α, see Definition 2.3.3. As we rather focuse on *a posteriori* strategy for the choice of regularization parameter $\alpha = \alpha(\delta, f^\delta)$, *Morozov's discrepancy principle* (MDP) is the strategy for this work to choose the regularization parameter, [7]. Under this principle, one must be able to choose such a regularization parameter $\alpha = \alpha(\delta, f^\delta)$ for the regularized solution $\varphi_{\alpha(\delta, f^\delta)}^\delta$ of the problem (2.14) that,

$$\underline{\tau}\delta \leq ||\mathcal{T}\varphi_{\alpha(\delta, f^\delta)}^\delta - f^\delta||_{\mathcal{L}^2(\mathcal{Z})} \leq \overline{\tau}\delta, \text{ for some fixed } 1 < \underline{\tau} \leq \overline{\tau} < \infty,$$

holds. More information about the choice of regularization parameter through MDP can be found in [**35, Eq. (4.57) and (4.58)**], [**59, Definition 2.3**]. We always work with the given perturbed data $f^\delta \in \mathcal{B}_\delta(f^\dagger)$ and introduce the convergence rates in terms of the noise level δ. Note that we formulate Bregman distance as associated with some Fréchet differentiable and convex functional.

Definition 2.3.4. *[Bregman distance][**79, Definitions 3.28 and 10.31**] Let $\mathcal{P} : \mathcal{V} \to \mathbb{R} \cup \{\infty\}$ be a convex and Fréchet differentiable functional. Then, for $u, u^* \in \mathcal{D}(\mathcal{P}) \subseteq \mathcal{V}$, the Bregman distance $D_\mathcal{P} : \mathcal{V} \times \mathcal{V} \to R_+$ associated with the functional $\mathcal{P}$ is given by*

$$D_\mathcal{P}(u, u^*) = \mathcal{P}(u) - \mathcal{P}(u^*) - \langle \nabla\mathcal{P}(u^*), u - u^* \rangle. \tag{2.16}$$

The following formulation emphasizes the functionality of the Bregman distance in proving the norm convergence of the minimizer φ_α^δ of the minimization problem (2.14) to the true solution $\varphi^\dagger$.

Definition 2.3.5. *[Total convexity][**21, Definition 1**], [**51, Section A.2**]*
Let $\mathcal{P} : \mathcal{V} \to \mathbb{R} \cup \{\infty\}$ be a Fréchet differentiable convex functional. Then $\mathcal{P}$ is called totally convex in $u^* \in \mathcal{V}$, if,

$$\mathcal{P}(u) - \mathcal{P}(u^*) - \langle \nabla\mathcal{P}(u^*), u - u^* \rangle \to 0 \Rightarrow ||u - u^*||_\mathcal{V} \to 0.$$

It is said that $\mathcal{P}$ is q-convex in $u^* \in \mathcal{V}$ with a $q \in [2, \infty)$, if for all $M > 0$ there exists a $c^* > 0$ such that for all $||u - u^*||_\mathcal{V} \leq M$ we have

$$\mathcal{P}(u) - \mathcal{P}(u^*) - \langle \nabla\mathcal{P}(u^*), u - u^* \rangle \geq c^*||u - u^*||_\mathcal{V}^q. \tag{2.17}$$

17

In addition to the traditional definition of Bregman distance in (2.17), the *symmetric Bregman distance* is also given below, (cf. [**41**, **Definition 2.1**]),

$$D_{\mathcal{P}}^{\mathrm{sym}}(u, u^*) := D_{\mathcal{P}}(u, u^*) + D_{\mathcal{P}}(u^*, u). \tag{2.18}$$

From here, one can easily observe that

$$D_{\mathcal{P}}^{\mathrm{sym}}(u, u^*) \geq D_{\mathcal{P}}(u, u^*). \tag{2.19}$$

Same also holds if one replaces the right hand side by $D_{\mathcal{P}}(u^*, u)$. In the next chapter, we will derive convergence theorems based on the lower bound for the Bregman distance $D_{\mathcal{P}}(u, u^*)$, or equivalently for $D_{\mathcal{P}}(u^*, u)$. By the relation (2.19), this lower bound will surely be the lower bound for $D_{\mathcal{P}}^{\mathrm{sym}}(u, u^*)$.

According to the estimation (2.17), to establish an $\mathcal{L}^2$ norm convergence, the functional $\mathcal{P}$ must be then $2-$convex functional. If the functional $\mathcal{P}$ is strongly (or, strictly) convex, such constant can also be estimated by a lower bound of the Hessian of the functional. We will see this in the next chapter. To this end, it is necessary to express the Bregman distance as the remainder term of the 2nd order Taylor expansion as formulated below.

Proposition 2.3.6. *Let $\mathcal{V}$ be a Hilbert space and let $\mathcal{P} : \mathcal{V} \to \mathbb{R} \cup \{\infty\}$ be some convex and twice continuously differentiable functional. Then for all $u, v \in \mathcal{V}$ there exists w on the line segment connecting u and v such that*

$$D_{\mathcal{P}}(u, v) = \frac{1}{2}\mathcal{P}''[w](u - v, u - v). \tag{2.20}$$

Proof. Consider the Taylor expansion of $f(t) := \mathcal{P}(v + t(u - v))$ with Lagrangian remainder term for some $\tau \in (0, 1)$ below,

$$f(1) = f(0) + f'(0) + \frac{1}{2}f''(\tau).$$

This translates into

$$\mathcal{P}(u) = \mathcal{P}(v) + \mathcal{P}'[v] \cdot (u - v) + \frac{1}{2}\mathcal{P}''[u + \tau(v - u)](u - v, u - v),$$

from which the assertion follows.

$\square$

Chapter 3

VARIATIONAL CONVEX REGULARIZATION

For the given linear, compact forward operator $\mathcal{T} : \mathcal{V} \to \mathcal{H}$, defined from a Banach space $\mathcal{V}$ into a Hilbert space $\mathcal{H}$, and some noisy data $f^\delta \in \mathcal{H}$ with the classical deterministic noise model $f^\delta \in \mathcal{B}_\delta(f^\dagger)$, *i.e.* $||f^\dagger - f^\delta||_\mathcal{H} \leq \delta$, we seek the minimizer for the general Tikhonov type objective functional $F_\alpha : \mathcal{V} \times \mathcal{H} \to \mathbb{R}_+$ given by

$$F_\alpha(\varphi, f^\delta) := \frac{1}{2}||\mathcal{T}\varphi - f^\delta||^2_\mathcal{H} + \alpha J(\varphi), \tag{3.1}$$

where the positive real valued penalty term $J \in \mathcal{C}^2(\mathcal{V})$ is convex. In the Subsection 1.2, it has been introduced that the image space of the forward operator $\mathcal{T}$ is in fact $\mathcal{H} = \mathcal{L}^2(\mathcal{Z})$, see (1.8). Thus, in what follows, the data fidelity term in the objective functional F_α will be amended accordingly. Of particular interest, we are rather concerned with solving the following problem,

$$\varphi^\delta_\alpha \in \arg\min_{\varphi \in \mathcal{V}} \left\{ \frac{1}{2}||\mathcal{T}\varphi - f^\delta||^2_{\mathcal{L}^2(\mathcal{Z})} + \alpha J^{\text{TV}}_\beta(\varphi) \right\}. \tag{3.2}$$

Here, the assumed solution space is $\mathcal{V} = \mathcal{L}^p(\Omega)$, for $1 \leq p \leq d/(d-1)$ where $d = 3$ since this work focuses on three dimensional reconstruction, and the smooth-TV penalty term, [**1**, **27**, **33**], is defined by,

$$J^{\text{TV}}_\beta(\varphi) := \int_\Omega \sqrt{|\nabla\varphi(x)|^2_2 + \beta}dx, \tag{3.3}$$

where $|\cdot|_2$ is the Euclidean norm defined by (2.1).

3.1   Existence and Uniqueness of the Minimizer

The motivation above leads us to seeking the regularized solution φ_α^δ on the space $\mathcal{V} = \mathcal{W}^{1,2}$. Note that for Ω bounded the inclusion $\mathcal{W}^{1,2}(\Omega) \subset \mathcal{W}^{1,p}(\Omega)$ holds, where $1 \leq p \leq d/(d-1)$ with $d = 3$. Existence of the minimizer has been studied extensively in [**1, 50, 80, 86**].

In order to ensure the existence and uniqueness of the regularizer, one must be able to ensure the BV-coercivity of the objective functional F_α. In [**1, Lemma 4.1**], BV-coercivity of F_α has been proven only for the case of $\beta = 0$ due to the inequalities in (2.11) and under the certain conditions on the forward operator $\mathcal{T}$, one of which is $\mathcal{T}$ does not annihilate constant functions. This condition holds true for our tomographic problem since the Radon transformation transforms any non-zero constant function into another non-zero constant function. On the other hand, it has been formulated in (1.11) that there are also rays with empty intersection. So, the definition of $\mathcal{T}$ and the Assumption 1.2.1 convey the following facts about the forward operator,

$$\mathcal{T}1 \geq 0 \text{ and } \mathcal{T}1 \neq 0. \tag{3.4}$$

Furthermore $\mathcal{T}$ is injective, as we have indicated by Theorem 1.2.2, then the objectively functional (3.1) is strictly convex.

Let the mean value functional $MV : \mathcal{L}^1(\Omega) \to \mathbb{R}$ be defined by

$$MV[u] := \frac{1}{|\Omega|} \int_\Omega u(x)dx. \tag{3.5}$$

It has been stated in [**1, Eq. (4.3)**] that, for some bounded domain Ω, any $u \in BV(\Omega)$ has the following decomposition

$$u = \tilde{u} + MV[u]1, \tag{3.6}$$

where

$$MV[\tilde{u}] = 0. \tag{3.7}$$

It is due to this decomposition and the properties of our forward operator given in (3.4) that the BV-coercivity of the functional (3.2) with $\beta = 0$ is guarenteed, [**1, Lemma 4.1**]. That is

20

$$F_\alpha(\varphi, f^\delta) \to \infty, \text{ as } ||\varphi||_{\mathrm{BV}} \to \infty.$$

BV-coercivity of the functional F_α in conjunction with the aforementioned facts of the forward operator implies the existence of the minimizer. Uniqueness of the minimizer is also another provided fact since the functional (3.1) is strictly convex. Hence, the regularized minimizer φ_α^δ to the problem (3.2) exists and is unique in the space $\mathcal{W}^{1,2}(\Omega)$ according to [**1, Theorems 3.1 and 4.1**].

3.2 Convexity of the Smooth TV Functional

Not only is this section dedicated to some discussion over the convexity of the smooth TV functional (3.3), but also for a motivation for the convergence analysis which will be developed in the later sections, a non-negative lower bound for the Hessian of J_β^{TV} will be beneficial for showing convergence. Convexity of J_β^{TV} has been guaranteed in [**1, Theorem 2.4**]. As has been stated by Proposition 2.3.6, the Hessian of J_β^{TV} is necessary to establish the convergence theory in the sense of Bregman distance. The Hessian will be calculated following from the bilinear form of the Fréchet derivative of J_β^{TV} in the direction $\Phi \in \mathcal{C}_c^1(\Omega)$ as such,

$$
\begin{aligned}
(J_\beta^{\mathrm{TV}})'[\varphi](\Phi) = \frac{d}{dt} J_\beta^{\mathrm{TV}}(\varphi + t\Phi)|_{t=0} &= \int_\Omega \frac{(\nabla\varphi(x) + t\nabla\Phi(x)) \cdot \nabla\Phi(x)}{(|\nabla\varphi(x) + t\nabla\Phi(x)|_2^2 + \beta)^{1/2}} dx \Bigg|_{t=0} \\
&= \int_\Omega \frac{\nabla\varphi(x) \cdot \nabla\Phi(x)}{(|\nabla\varphi(x)|_2^2 + \beta)^{1/2}} dx. \qquad (3.8)
\end{aligned}
$$

Then the second order Fréchet derivative in the direction $\Psi \in \mathcal{C}_c^1(\Omega)$, *i.e.* the Hessian, follows from (3.8)

$$(J_\beta^{\mathrm{TV}})''[\varphi](\Phi, \Psi) = \frac{d}{ds}(J_\beta^{\mathrm{TV}})'[\varphi + s\Psi](\Phi)\Bigg|_{s=0} \overset{1}{=} \frac{d}{ds}\int_\Omega \frac{(\nabla\varphi(x) + s\nabla\Psi(x)) \cdot \nabla\Phi(x)}{(|\nabla\varphi(x) + s\nabla\Psi(x)|_2^2 + \beta)^{1/2}} dx \Bigg|_{s=0},$$

which is,

$$\frac{d}{ds}(J_\beta^{\text{TV}})'[\varphi + s\Psi](\Phi)\Big|_{s=0} = \int_\Omega \frac{(\nabla\Psi(x) \cdot \nabla\Phi(x))\,(|\nabla\varphi(x)|_2^2 + \beta) - |\nabla\varphi(x) \cdot \nabla\Phi(x)|_2^2}{(|\nabla\varphi(x)|_2^2 + \beta)^{3/2}}dx.$$

Here, the sign '·' inside the integral indicates the inner product in Euclidean space. By the choice of $\Psi = \Phi$, the Hessian of the smooth-TV functional is obtained,

$$(J_\beta^{\text{TV}})''[\varphi](\Phi, \Phi) = \int_\Omega \frac{|\nabla\Phi(x)|_2^2\,(|\nabla\varphi(x)|_2^2 + \beta) - |\nabla\varphi(x) \cdot \nabla\Phi(x)|_2^2}{(|\nabla\varphi(x)|_2^2 + \beta)^{3/2}}dx. \tag{3.9}$$

From here, we are only able to estimate a non-negative lower bound for the Hessian of J_β^{TV} in the weighted $\mathcal{L}^2$ norm, recall that the weighted $\mathcal{L}^2$ norm definition introduced in (2.3).

Theorem 3.2.1. *[**Strong convexity of the functional** J_β^{TV}] Over the bounded domain $\Omega \subset \mathbb{R}^N$, where $N = 2$, or 3, let the function $\varphi : \Omega \to \mathbb{R}$ be in the class of $\mathcal{W}^{1,1}(\Omega)$. Then, there can be derived a weight functional $m_\beta : \mathcal{W}^{1,1}(\Omega) \to \mathbb{R}_+$ such that the Hessian of the smooth TV functional $J_\beta^{\text{TV}} \in \mathcal{C}^2$ in the direction $\Phi \in \mathcal{C}_c^1(\Omega)$ can be bounded below by,*

$$(J_\beta^{\text{TV}})''[\varphi](\Phi, \Phi) \geq ||\nabla\Phi||_{\mathcal{L}^2_{m_\beta}(\Omega)}^2.$$

Proof. Consider the basic inequality $|\nabla\varphi \cdot \nabla\Phi|_2^2 \leq |\nabla\varphi|_2^2|\nabla\Phi|_2^2$ in (3.9) to arrive at the following estimation,

$$\begin{aligned}(J_\beta^{\text{TV}})''[\varphi](\Phi, \Phi) &\geq \int_\Omega \frac{|\nabla\Phi(x)|_2^2\,(|\nabla\varphi(x)|_2^2 + \beta) - |\nabla\varphi(x)|_2^2|\nabla\Phi(x)|_2^2}{(|\nabla\varphi(x)|_2^2 + \beta)^{3/2}}dx \\ &= \int_\Omega \frac{\beta|\nabla\Phi(x)|_2^2}{(|\nabla\varphi(x)|_2^2 + \beta)^{3/2}}dx. \end{aligned} \tag{3.10}$$

---

[1]More precisely this equals to,

$$\int_\Omega \left(\frac{(\nabla\Psi(x) \cdot \nabla\Phi(x))(|\nabla\varphi(x) + s\nabla\Psi(x)|_2^2 + \beta)^{1/2} - ((\nabla\varphi(x) + s\nabla\Psi(x)) \cdot \nabla\Phi(x)) \cdot \left(\frac{(\nabla\varphi(x)+s\nabla\Psi(x))\cdot\nabla\Psi(x)}{(|\nabla\varphi(x)+s\nabla\Psi(x)|_2^2+\beta)^{1/2}}\right)}{(|\nabla\varphi(x) + s\nabla\Psi(x)|_2^2 + \beta)} \right) dx \Big|_{s=0}$$

22

Here let us set the weight functional by

$$m_\beta(\varphi, x) := \frac{\beta}{(|\nabla\varphi(x)|_2^2 + \beta)^{3/2}}, \text{ for all } x \in \Omega, \tag{3.11}$$

where $0 < \beta < 1$ is fixed. Hence, we arrive at the desired lower bound,

$$(J_\beta^{\mathrm{TV}})''[\varphi](\Phi, \Phi) \geq \int_\Omega m_\beta(\varphi, x)|\nabla\Phi(x)|_2^2 dx = ||\nabla\Phi||_{\mathcal{L}_{m_\beta}^2(\Omega)}^2. \tag{3.12}$$

$\square$

With a further assumption on the function φ, this lower bound can be reformulated as such.

Corollary 3.2.2. *Over the bounded domain $\Omega \subset \mathbb{R}^N$, where $N = 2$, or 3, suppose that $\varphi \in \mathcal{W}^{1,\infty}(\Omega)$. Then, there exists a functional $l_\beta : \mathcal{W}^{1,\infty}(\Omega) \to \mathbb{R}_+$ defined by*

$$l_\beta(\varphi) := \min_{x \in \Omega}\{m_\beta(\varphi, x)\}.$$

With this functional and a constant C_Ω, in the direction $\Phi \in \mathcal{C}_c^1(\Omega)$, the following lower bound holds,

$$(J_\beta^{\mathrm{TV}})''[\varphi](\Phi, \Phi) \geq l_\beta(\varphi)\frac{C_\Omega}{|\Omega|}\left\|\Phi - \frac{1}{|\Omega|}\int_\Omega \Phi(x)dx\right\|_{\mathcal{L}^1(\Omega)}^2.$$

Proof. Following from the weight functional defined by (3.11) in the proof of Theorem 3.2.1, let, over the bounded domain Ω, the functional $l_\beta : \mathcal{W}^{1,\infty}(\Omega) \to \mathbb{R}_+$ be defined by

$$l_\beta(\varphi) := \min_{x \in \Omega}\{m_\beta(\varphi, x)\} = \frac{\beta}{(\max_{x \in \Omega}|\nabla\varphi(x)|_2^2 + \beta)^{3/2}} \geq \frac{\beta}{(|\nabla\varphi(x)|_\infty^2 + \beta)^{3/2}}. \tag{3.13}$$

So, according to (3.12),

$$(J_\beta^{\mathrm{TV}})''[\varphi](\Phi, \Phi) \geq l_\beta(\varphi)\int_\Omega |\nabla\Phi(x)|_2^2 dx = l_\beta(\varphi)\|\nabla\Phi\|_{\mathcal{L}^2(\Omega)}^2. \tag{3.14}$$

Hence, after considering the Poincaré-Wirtinger inequality (2.4) in Theorem 2.1.1 and using the fact that $||\cdot||_{\mathcal{L}^1(\Omega)} \leq \sqrt{\Omega}||\cdot||_{\mathcal{L}^2(\Omega)}$ for the bounded domain Ω, we have

23

$$\begin{aligned}
(J_\beta^{\mathrm{TV}})''[\varphi](\Phi, \Phi) &\geq l_\beta(\varphi)\|\nabla\Phi\|^2_{\mathcal{L}^2(\Omega)} \\
&\geq l_\beta(\varphi)\frac{1}{|\Omega|}\|\nabla\Phi\|^2_{\mathcal{L}^1(\Omega)} \\
&\geq l_\beta(\varphi)\frac{C_\Omega}{|\Omega|}\left\|\Phi - \frac{1}{|\Omega|}\int_\Omega \Phi(x)dx\right\|^2_{\mathcal{L}^1(\Omega)}.
\end{aligned} \tag{3.15}$$

$\square$

Remark 3.2.3. *Note that we have obtained this lower bound estimation under the condition that denominator can be bounded only if the target function $\varphi \in \mathcal{W}^{1,\infty}(\Omega)$.*

3.2.1 Geometric form of Hessian operator of TV functional

For the functions $\varphi \in \mathcal{W}^{1,1}(\Omega)$ and $\Phi \in \mathcal{C}_c^1(\Omega)$, where $\Omega \subset \mathbb{R}^N$ and $N = 2$, or 3 is the bounded domain, (3.9) can also be represented in a geometric form.

$$\begin{aligned}
(J_\beta^{\mathrm{TV}})''[\varphi](\Phi, \Phi) &= \int_\Omega \frac{\beta|\nabla\Phi(x)|_2^2 + |\nabla\Phi(x)|_2^2|\nabla\varphi(x)|_2^2\left(1 - \frac{|\nabla\varphi(x)\cdot\nabla\Phi(x)|_2^2}{|\nabla\Phi(x)|_2^2|\nabla\varphi(x)|_2^2}\right)}{\left(|\nabla\varphi(x)|_2^2 + \beta\right)^{3/2}}dx \\
&= \int_\Omega \frac{|\nabla\Phi(x)|_2^2\left(\beta + |\nabla\varphi(x)|_2^2(1 - \cos^2\measuredangle(\nabla\varphi(x); \nabla\Phi(x)))\right)}{\left(|\nabla\varphi(x)|_2^2 + \beta\right)^{3/2}}dx \\
&= \int_\Omega \frac{|\nabla\Phi(x)|_2^2\left(\beta + |\nabla\varphi(x)|_2^2\sin^2\measuredangle(\nabla\varphi(x); \nabla\Phi(x))\right)}{\left(|\nabla\varphi(x)|_2^2 + \beta\right)^{3/2}}dx. \tag{3.16}
\end{aligned}$$

Now, let us take $\beta = 0$ and recall by the physical properties of the unknown function that $\nabla\varphi \neq 0$. Then, we obtain

$$\begin{aligned}
(J_0^{\mathrm{TV}})''[\varphi](\Phi, \Phi) &= \int_\Omega \frac{|\nabla\Phi(x)|_2^2|\nabla\varphi(x)|_2^2\sin^2\measuredangle(\nabla\varphi(x); \nabla\Phi(x))}{|\nabla\varphi(x)|_2^3}dx \\
&= \int_\Omega \frac{|\nabla\Phi(x)|_2^2\sin^2\measuredangle(\nabla\varphi(x); \nabla\Phi(x))}{|\nabla\varphi(x)|_2}dx
\end{aligned} \tag{3.17}$$

Before deriving the consequence from this geometric representation, let us perceive the meaning of directional function Φ. If $\mathcal{Q} : \mathbb{R} \to \mathbb{R}$ is some scaling function, then the directional function is $\Phi = \varphi - \psi$, where $\psi = \mathcal{Q} \circ \varphi$.

Corollary 3.2.4. *In (3.17), if $\measuredangle\big(\nabla\varphi(x);\nabla\Phi(x)\big) \in \{0,\pi\}$, for any $x \in \Omega$, that is if $\nabla\Phi \parallel \nabla\varphi$, then,*

$$(J_0^{\mathrm{TV}})''[\varphi](\Phi,\Phi) = 0,$$

which, by Proposition 2.3.6, implies

$$D_{J_0^{\mathrm{TV}}}(\mathcal{Q}\circ\varphi, \varphi) = 0. \tag{3.18}$$

Remark 3.2.5. *From the result (3.18), we see that if $\nabla\Phi$ and $\nabla\varphi$ are linearly dependent, i.e. $\nabla\Phi = \lambda\nabla\varphi$ for some $\lambda \in \mathbb{R}\backslash\{0\}$, then the magnitudes of these vector fields will be proportional, i.e. $|\nabla\Phi|_2 \propto |\nabla\varphi|_2$. Thus, so are their TV norms, $TV(\Phi) \propto TV(\varphi)$. Nevertheless, distance between φ and Φ under the mapping TV in the Bregman sense equals to zero when $\nabla\Phi \parallel \nabla\varphi$.*

3.3 Variational Convergence Analysis with Variational Source Condition

In this section, we will rather focus on unconstraint, smooth, convex minimization problem formulated in the general form below

$$\varphi_\alpha^\delta \in \arg\min_{\varphi\in\mathcal{V}} \left\{ \frac{1}{2}||\mathcal{T}\varphi - f^\delta||_\mathcal{H} + \alpha J(\varphi) \right\}. \tag{3.19}$$

Here, the forward operator $\mathcal{T} : \mathcal{V} \to \mathcal{H}$ is linear and compact, and $J \in \mathcal{C}^2(\mathcal{V})$ is the convex penalty term.

As alternative to well established Tikhonov regularization, [**82, 83**], studying convex regularization strategy has been initiated by introducing a new image denoising method named as *total variation* (TV), [**77**]. Due to nonlinearity in the Fréchet derivative of TV penalty term, variational inequalities in convergence analysis for the minimization problems in the form of (3.19) are useful. In the work [**23**], Bregman distance has been introduced for stability analysis of the regularized solution of the problem (3.19). Henceforth, usage of the Bregman distance has been adapted and developed in several works, [**23, 40, 41, 42, 53, 63**] and references therein.

In a work by Hofmann and Mathé *et al.* 2012, [**51**], *a priori* and *a posteriori* strategies for the choice of regularization parameter in Banach spaces under the variational source condition to determine the total error estimation

$$E(\varphi^{\delta}_{\alpha(\delta,f^{\delta})}, \varphi^{\dagger}) := ||\varphi^{\delta}_{\alpha(\delta,f^{\delta})} - \varphi^{\dagger}||_{\mathcal{V}}, \tag{3.20}$$

have been studied extensively. The objective of this section is to investigate the stable convergence and convergence rates for the error measure $E(\varphi^{\delta}_{\alpha(\delta,f^{\delta})}, \varphi^{\dagger})$ as the noise amount $\delta \to 0$. Convergence rates results for some general operator $\mathcal{T}$ can be obtained by formulating *variational inequality* which uses the concept of index functions. A function $\Psi : [0, \infty) \to [0, \infty)$ is called *index function* if it is continuous, monotonically increasing and $\Psi(0) = 0$, [**40, Eq. (1)**], [**51, Eq. (1.5)**], [**53, Eq. (2)**]. Stability analysis for the problem (3.19) with some general convex penalty term $J : \mathcal{V} \to \mathbb{R}_{+} \cup \{\infty\}$ has been conducted extensively in the literature. Specifically in [**51, Eq. (1.5)**] variational inequality assumption has been introduced together with a concept of *domain of validity* $\mathcal{M}$ wherein the regularized $\varphi^{\delta}_{\alpha(\delta,f^{\delta})}$ and the true $\varphi^{\dagger}$ solutions assume to exist. A similar argument had been discussed in [**79, Assumption 3.13**] before the work by Hofmann & Mathé *et. al* 2012. The domain of validity $\mathcal{M}$ is structured from the level sets of the penalty term J according to [**51**], or from the functional itself F_{α} according to [**79, Assumption 3.13**]. However, a simpler form of variational inequality assumption, which is the conventional form, still leads to the desired convergence/stability results. In a more recent work by Hohage *et al.* 2015, [**53, Eq. (2)**] and references therein, a conventional variational source condition with a logarithmic index function Ψ has been derived for an inverse scattering problem. In analogous to the works mentioned above and in accordance with the theoretical aim of this work, we give the variational source condition associated with the Bregman distance in the following assumption.

Assumption 3.3.1. *[**Variational Source Condition**] There exists some constant $\sigma \in (0, 1]$ and a concave index function Ψ such that*

$$\frac{\sigma}{2} D_J(\varphi, \varphi^{\dagger}) \leq J(\varphi) - J(\varphi^{\dagger}) + \Psi\left(||\mathcal{T}\varphi - \mathcal{T}\varphi^{\dagger}||_{\mathcal{L}^2(\mathcal{Z})}\right), \text{ for all } \varphi \in \mathcal{V}. \tag{3.21}$$

Recall the fact that classical deterministic noise model $f^{\delta} \in \mathcal{B}_{\delta}(f^{\dagger})$ is taken into account throughout our analysis. Moreover, the regularized solution must satisfy the first order optimality condition given by (2.15). As we have listed in Subsection 2.3, under some *a posteriori* strategy for the choice of regularization parameter $\alpha = \alpha(\delta, f^{\delta})$ together with the aforementioned assumptions, it is possible to be able to quantify the following rates;

1. $\mathcal{T}\varphi^{\delta}_{\alpha(\delta,f^{\delta})} \in \mathcal{B}_{\mathcal{O}(\delta)}(\mathcal{T}\varphi^{\dagger})$; norm of the discrepancy between $\mathcal{T}\varphi^{\delta}_{\alpha(\delta,f^{\delta})}$ and $\mathcal{T}\varphi^{\dagger}$ by the rate of $\mathcal{O}(\delta)$, *i.e.* $||\mathcal{T}\varphi^{\delta}_{\alpha(\delta,f^{\delta})} - \mathcal{T}\varphi^{\dagger}||_{\mathcal{L}^2(\mathcal{Z})} = \mathcal{O}(\delta)$.

2. $D_J(\varphi^\delta_{\alpha(\delta,f^\delta)}, \varphi^\dagger) = \mathcal{O}(\Psi(\delta))$; upper bound for the Bregman distance D_J.

3. $\varphi_{\alpha(\delta,f^\delta)} \in \mathcal{B}_{\mathcal{O}(\Psi(\delta))}(\varphi^\dagger)$; convergence of the regularized solution $\varphi_{\alpha(\delta,f^\delta)}$ to the true solution $\varphi^\dagger$ by the rate of the noise amount $\mathcal{O}(\Psi(\delta))$.

3.3.1 Choice of the regularization parameter with Morozov's discrepancy principle

We are concerned with asymptotic properties of the regularization parameter α for the Tikhonov-regularized solution obtained by Morozov's discrepancy principle. Morozov's discrepancy principle (MDP) serves as an *a posteriori* parameter choice rule for the Tikhonov type objective functionals (3.1) and has certain impact on the convergence of the regularized solution for the problem in (3.19) with some general convex penalty term *J*. As has been introduced in [**5, Theorem 3.10**] and [**7**], we will make use of the following set notations in the theorem formulations that are necessary to prove the norm convergence of the solution $\varphi_{\alpha(\delta,f^\delta)}$ to the true solution $\varphi^\dagger$ for the problem (3.19).

$$\overline{S} := \left\{\alpha : ||\mathcal{T}\varphi^\delta_{\alpha(\delta,f^\delta)} - f^\delta||_{\mathcal{L}^2(\mathcal{Z})} \leq \overline{\tau}\delta \text{ for } \varphi^\delta_{\alpha(\delta,f^\delta)} \in \arg\min_{\varphi \in \mathcal{V}}\{F_\alpha(\varphi, f^\delta)\}\right\} \quad (3.22)$$

$$\underline{S} := \left\{\alpha : \underline{\tau}\delta \leq ||\mathcal{T}\varphi^\delta_{\alpha(\delta,f^\delta)} - f^\delta||_{\mathcal{L}^2(\mathcal{Z})} \text{ for } \varphi^\delta_{\alpha(\delta,f^\delta)} \in \arg\min_{\varphi \in \mathcal{V}}\{F_\alpha(\varphi, f^\delta)\}\right\} \quad (3.23)$$

where $1 < \underline{\tau} \leq \overline{\tau} < \infty$ are fixed. Analogously, as well known from [**35, Eq. (4.57) and (4.58)**], [**59, Definition 2.3**], in order to obtain tight rates of convergence for $||\varphi^\delta_{\alpha(\delta,f^\delta)} - \varphi^\dagger||_\mathcal{V}$ we are interested in such a regularization parameter $\alpha(\delta, f^\delta)$ with some fixed $1 < \underline{\tau} \leq \overline{\tau} < \infty$, that

$$\alpha(\delta, f^\delta) \in \{\alpha > 0 \; : \; \underline{\tau}\delta \leq ||\mathcal{T}\varphi^\delta_{\alpha(\delta,f^\delta)} - f^\delta||_{\mathcal{L}^2(\mathcal{Z})} \leq \overline{\tau}\delta\} = \overline{S} \cap \underline{S} \text{ for given } (\delta, f^\delta). \quad (3.24)$$

In particular, if $\alpha(\delta, f^\delta) \in \overline{S}$, then one can immediately conclude that,

$$||\mathcal{T}\varphi^\delta_{\alpha(\delta,f^\delta)} - \mathcal{T}\varphi^\dagger||_{\mathcal{L}^2(\mathcal{Z})} \leq (\overline{\tau} + 1)\delta. \quad (3.25)$$

Furthermore, according to [**51, Corollary 2**], the regularization parameter $\alpha(\delta, f^\delta) \in \underline{S}$ can be bounded below by,

$$\alpha(\delta, f^\delta) \geq \frac{1}{4} \frac{\underline{\tau}^2 - 1}{\underline{\tau}^2 + 1} \frac{\delta^2}{\Psi((\underline{\tau} - 1)\delta)}, \tag{3.26}$$

where Ψ is the concave index function appeared in Assumption 3.3.1. With this lower bound, the possible singularity is avoided as $\alpha \to 0$.

Lemma 3.3.2. *Let the regularization parameter $\alpha = \alpha(\delta, f^\delta)$, for the minimizer $\varphi^\delta_{\alpha(\delta, f^\delta)} \in \mathcal{D}(F_\alpha)$ of the objective functional F_α in (3.1), be chosen according to the disrepancy principle $\alpha(\delta, f^\delta) \in \overline{S} \cap \underline{S}$ where the given data $f^\delta \in \mathcal{B}_\delta(f^\dagger)$. Then this a posteriori rule for the choice of the regularization parameter stabilises the following J difference*

$$J(\varphi^\delta_{\alpha(\delta, f^\delta)}) - J(\varphi^\dagger) = \mathcal{O}(\Psi(\delta)). \tag{3.27}$$

Proof. Let $\alpha > 0$. Since $\varphi^\delta_\alpha \in \mathcal{D}(F_\alpha)$ is the minimizer of the objective functional F_α, for any $\varphi \in \mathcal{D}(F_\alpha)$ it holds that $F_\alpha(\varphi^\delta_\alpha) \leq F_\alpha(\varphi)$. This implies the following,

$$J(\varphi^\delta_\alpha) - J(\varphi^\dagger) \leq \frac{\delta^2}{2\alpha}. \tag{3.28}$$

Here, the decrease in α will cause a blow-up on the right hand side. This is controlled by the choice of the regularization parameter $\alpha = \alpha(\delta, f^\delta) \in \underline{S}$. Thus, we make use of the lower bound for the regularization parameter given in (3.26) to have a stable upper bound by using the facts that Ψ is a concave and increasing function,

$$\frac{\delta^2}{2\alpha} \leq 2\frac{\underline{\tau}^2 + 1}{\underline{\tau}^2 - 1}\Psi((\underline{\tau} - 1)\delta) \overset{1}{\leq} 2\frac{\underline{\tau}^2 + 1}{\underline{\tau}^2 - 1}\Psi((\underline{\tau} + 1)\delta) \overset{2}{\leq} 2\frac{\underline{\tau}^2 + 1}{\underline{\tau} - 1}\Psi(\delta) \tag{3.29}$$

Hence, this control over the trade-off between δ^2 and α yields the desired result. $\square$

Counterparts to the following two results have been formulated in [**51, Theorem 2**] and in [**79, Theorem 3.26**]. As has been discussed above, both works involve the concept of domain of validity developing from squentially pre-compact sublevel sets which is not necessarily needed to develop the necessary bounds. Aside from this, the introduced sublevel sets in the regarding work are not weakly sequentially compact for TV functional.

---

[1]Since Ψ is an increasing function, then $\Psi((\underline{\tau} - 1)\delta) \leq \Psi((\underline{\tau} + 1)\delta)$.

[2]Due to the concavity of Ψ, $\Psi((\underline{\tau} + 1)\delta) \leq (\underline{\tau} + 1)\Psi(\delta)$ holds, see [**51, Eq. 2.3 of Proposition 1**].

EXAMPLE 3.3.3 (A counterexample showing that the sublevel sets for J_β^{TV} are not sequentially pre-compact). *According to [**51, p. 2**], the sublevel sets for J_β^{TV} are defined below,*

$$\mathcal{M}_R^{J_\beta^{\text{TV}}} := \{\varphi \in \mathcal{W}^{1,2}(\Omega) : J_\beta^{\text{TV}}(\varphi) \leq R\}, \text{ for } R > 0.$$

Obviously $\mathcal{M}_R^{J_\beta^{\text{TV}}} \subset BV(\Omega)$. To ensure that the sublevel sets are weakly sequentially compact, one must show that every sequence $\varphi_n \in \mathcal{M}_R^{J_\beta^{\text{TV}}}$ has a weakly convergent subsequence with the limit in $\mathcal{M}_R^{J_\beta^{\text{TV}}}$, i.e. the sequence $\varphi_n \in \mathcal{M}^{J_\beta^{\text{TV}}}$ has a subsequence $\{\varphi_{n_k}\}_{k=1}^\infty \subset \{\varphi_n\}_{n=1}^\infty$ such that $\varphi_{n_k} \rightharpoonup \varphi^*$ as $k \to \infty$ where $\varphi^* \in \mathcal{M}_R^{J_\beta^{\text{TV}}}$. However, it can be shown that the sublevel sets $\mathcal{M}_R^{J_\beta^{\text{TV}}}$ contain sequence which does not have weakly convergent subsequence. To do so, for some real function $\tilde{\varphi} \in \mathcal{M}_R^{J_\beta^{\text{TV}}}$, consider the sequence $\varphi_n(x) = \tilde{\varphi}(x) + n1$, where $x \in \Omega$. Although, for the defined sequence $J_\beta^{\text{TV}}(\varphi_n) = J_\beta^{\text{TV}}(\tilde{\varphi}) \leq R$ holds for any $n \in \mathbb{N}$ the sequence φ_n cannot have weakly convergent subsequence in $BV(\Omega)$ since

$$
\begin{aligned}
||\varphi_n - \varphi_m||_{BV(\Omega)} &\geq ||\varphi_n - \varphi_m||_{\mathcal{L}^1(\Omega)} \\
&\geq |n - m||\Omega| \to \infty \text{ as } n \to \infty.
\end{aligned}
\tag{3.30}
$$

Theorem 3.3.4. *Let the true solution $\varphi^\dagger$ satisfy Assumption 3.3.1. Under the same conditions in Lemma 3.3.2, we have*

$$D_J(\varphi_{\alpha(\delta,f^\delta)}^\delta, \varphi^\dagger) = \mathcal{O}(\Psi(\delta)), \tag{3.31}$$

as $\delta \to 0$.

Proof. Since the true solution $\varphi^\dagger$ satisfies Assumption 3.3.1,

$$
\begin{aligned}
D_J(\varphi_{\alpha(\delta,f^\delta)}^\delta, \varphi^\dagger) &\leq J(\varphi_{\alpha(\delta,f^\delta)}^\delta) - J(\varphi^\dagger) + \Psi\left(||\mathcal{T}\varphi_{\alpha(\delta,f^\delta)}^\delta - \mathcal{T}\varphi^\dagger||_{\mathcal{L}^2(\mathcal{Z})}\right). \\
&\leq 2\frac{\underline{\tau}^2 + 1}{\underline{\tau} - 1}\Psi(\delta) + \Psi\left((\overline{\tau} + 1)\delta\right) \\
&\leq 2\frac{\underline{\tau}^2 + 1}{\underline{\tau} - 1}\Psi(\delta) + (\overline{\tau} + 1)\Psi(\delta).
\end{aligned}
\tag{3.32}
$$

The first term on the right hand side, the bound for the J difference, comes from Lemma 3.3.2. As in the estimation (3.29) of Lemma 3.3.2, Ψ is concave function, thus $\Psi\left((\overline{\tau} + 1)\delta\right) \leq (\overline{\tau} + 1)\Psi(\delta)$. $\qquad\square$

29

Theorem 3.3.5. *Let the regularization parameter $\alpha = \alpha(\delta, f^\delta)$, for the minimizer $\varphi^\delta_{\alpha(\delta,f^\delta)} \in \mathcal{D}(F_\alpha)$ of the objective functional F_α in (3.1), be chosen according to the disrepancy principle $\alpha(\delta, f^\delta) \in \overline{S} \cap \underline{S}$ where the given data $f^\delta \in \mathcal{B}_\delta(f^\dagger)$. Suppose that the true solution $\varphi^\dagger \in \mathcal{V}$, where $\mathcal{T}\varphi^\dagger = f^\dagger$, satisfies Assumption 3.3.1 with the concave and monotonically increasing index function $\Psi : [0, \infty) \to [0, \infty)$. Then, this* a posteriori *rule for the choice of regularization parameter yields the following rate,*

$$D_J(\varphi^\dagger, \varphi^\delta_{\alpha(\delta,f^\delta)}) = \mathcal{O}(\Psi(\delta)), \tag{3.33}$$

as $\delta \to 0$.

Proof. Firstly, by Assumption 3.3.1, it can easily be observed that,

$$J(\varphi^\dagger) - J(\varphi^\delta_{\alpha(\delta,f^\delta)}) \leq \Psi\left(||\mathcal{T}\varphi^\delta_{\alpha(\delta,f^\delta)} - \mathcal{T}\varphi^\dagger||_{\mathcal{L}^2(\mathcal{Z})}\right). \tag{3.34}$$

From the early observation (3.25) and since Ψ is a monotonically increasing, concave function, we obtain,

$$J(\varphi^\dagger) - J(\varphi^\delta_{\alpha(\delta,f^\delta)}) \leq \Psi\left((\overline{\tau}+1)\delta\right) \leq (\overline{\tau}+1)\Psi(\delta). \tag{3.35}$$

Regarding the aimed upper bound for the Bregman distance, use the estimation (3.35) for $\alpha(\delta, f^\delta) \in \overline{S}$ and observe the following,

$$
\begin{aligned}
D_J(\varphi^\dagger, \varphi^\delta_{\alpha(\delta,f^\delta)}) &= J(\varphi^\dagger) - J(\varphi^\delta_{\alpha(\delta,f^\delta)}) - \langle \nabla J(\varphi^\delta_{\alpha(\delta,f^\delta)}), \varphi^\dagger - \varphi^\delta_{\alpha(\delta,f^\delta)}\rangle \\
&= J(\varphi^\dagger) - J(\varphi^\delta_{\alpha(\delta,f^\delta)}) + \langle \nabla J(\varphi^\delta_{\alpha(\delta,f^\delta)}), \varphi^\delta_{\alpha(\delta,f^\delta)} - \varphi^\dagger\rangle \\
&\leq (\overline{\tau}+1)\Psi(\delta) + \langle \nabla J(\varphi^\delta_{\alpha(\delta,f^\delta)}), \varphi^\delta_{\alpha(\delta,f^\delta)} - \varphi^\dagger\rangle
\end{aligned} \tag{3.36}
$$

Since the regularized solution $\varphi^\delta_{\alpha(\delta,f^\delta)}$ satisfies first order optimality condition (2.15), we then have,

$$D_J(\varphi^\dagger, \varphi^\delta_{\alpha(\delta,f^\delta)}) \leq (\overline{\tau}+1)\Psi(\delta) + \frac{1}{\alpha(\delta, f^\delta)}\langle \mathcal{T}^*(f^\delta - \mathcal{T}\varphi^\delta_{\alpha(\delta,f^\delta)}), \varphi^\delta_{\alpha(\delta,f^\delta)} - \varphi^\dagger\rangle.$$

Now apply the Cauchy-Schwarz inequality and take into account the estimation (3.25), $||\mathcal{T}\varphi^\delta_{\alpha(\delta,f^\delta)} - f^\delta||_{\mathcal{L}^2(\mathcal{Z})} \leq \overline{\tau}\delta$ for the choice of regularization parameter $\alpha(\delta, f^\delta) \in \overline{S}$, to arrive at

$$
\begin{aligned}
D_J(\varphi^\dagger, \varphi^\delta_{\alpha(\delta, f^\delta)}) \;\leq\;& (\bar{\tau} + 1)\Psi(\delta) + \frac{1}{\alpha(\delta, f^\delta)}||\mathcal{T}\varphi^\delta_{\alpha(\delta, f^\delta)} - f^\delta||_{\mathcal{L}^2(\mathcal{Z})}||\mathcal{T}\varphi^\delta_{\alpha(\delta, f^\delta)} - \mathcal{T}\varphi^\dagger||_{\mathcal{L}^2(\mathcal{Z})} \\
\leq\;& (\bar{\tau} + 1)\Psi(\delta) + \frac{1}{\alpha(\delta, f^\delta)}\bar{\tau}\delta||\mathcal{T}\varphi^\delta_{\alpha(\delta, f^\delta)} - \mathcal{T}\varphi^\dagger||_{\mathcal{L}^2(\mathcal{Z})} \\
\leq\;& (\bar{\tau} + 1)\Psi(\delta) + \frac{1}{\alpha(\delta, f^\delta)}\bar{\tau}(\bar{\tau} + 1)\delta^2 \\
\leq\;& (\bar{\tau} + 1)\Psi(\delta) + 2\frac{\tau^2 + 1}{\underline{\tau} - 1}\Psi(\delta).
\end{aligned}
$$

Here, again, the lower bound for the regularization parameter $\alpha(\delta, f^\delta)$ given in (3.26) has controlled the trade-off between δ^2 and α. Hence, this yields the stable upper bound (3.33). $\qquad\square$

Upper bounds obtained in the theorems 3.3.4 and 3.3.5 provide upper bound for the symmetric Bregman distance defined in (2.18).

Corollary 3.3.6. *From the theorems 3.3.4 and 3.3.5, it is concluded that*

$$
D_J^{\mathrm{sym}}(\varphi^\delta_{\alpha(\delta, f^\delta)}, \varphi^\dagger) = \mathcal{O}(\Psi(\delta)), \;\; as \; \delta \to 0. \tag{3.37}
$$

3.4 Interpretation of Convex Variational Regularization for Smooth-TV Functional

Eventually, in this section, we will be able to establish some convergence theory specific to TV regularization by relying on the regularity property of the target function φ. Unlike the available literature [1, 9, 11, 24, 25, 27, 33, 34, 87], we will arrive at a new lower bound for the Bregman distance particularly associated with the smooth-TV functional J_β^{TV}. To this end, we will make use of the positive lower bound for the Hessian of the functional J_β^{TV} that has been found in Theorem 3.2.1. This lower bound will provide the convergence with its rates. Firstly, we need to formulate necessary lemmata revealing the relation between function, image of the function and function's mean value.

Lemma 3.4.1. *For all $u \in \mathcal{L}^1(\Omega)$, the left hand side of the Poincaré-Wirtinger inequality (2.4) can be bounded below by*

$$\frac{1}{2}||u||^2_{\mathcal{L}^1(\Omega)} - ||MV[u]||^2_{\mathcal{L}^1(\Omega)} \leq ||u - MV[u]||^2_{\mathcal{L}^1(\Omega)}. \tag{3.38}$$

Proof. Assertion simply follows from an immediate consequence of the binomial formula, which is $2ab \leq a^2 + b^2$ for $a, b \in \mathbb{R}_+$,

$$
\begin{aligned}
||u||^2_{\mathcal{L}^1(\Omega)} &\leq \left(||u - MV[u]||_{\mathcal{L}^1(\Omega)} + ||MV[u]||_{\mathcal{L}^1(\Omega)} \right)^2 \\
&\leq 2 \left(||u - MV[u]||^2_{\mathcal{L}^1(\Omega)} + ||MV[u]||^2_{\mathcal{L}^1(\Omega)} \right). \tag{3.39}
\end{aligned}
$$

$\square$

Lemma 3.4.2. *Assume there exists some γ satisfying*

$$\gamma := \sup_{u \perp 1} \left| \langle \frac{\mathcal{T}u}{||\mathcal{T}u||_{\mathcal{H}}}, \frac{\mathcal{T}1}{||\mathcal{T}1||_{\mathcal{H}}} \rangle \right| < 1. \tag{3.40}$$

Then, for all $u \in \mathcal{L}^2(\Omega)$, and for any operator $\mathcal{T} \in \mathfrak{L}(\mathcal{L}^2(\Omega), \mathcal{H})$ satisfying $\mathcal{T}1 \neq 0$, there could be found some constant $C_\gamma > 0$ such that

$$||\mathcal{T}u||^2_{\mathcal{H}} \geq C_\gamma |MV[u]|^2 \tag{3.41}$$

holds.

Proof. Assign the mean value functional of u to $\overline{u} := MV[u]$ and observe the following

$$
\begin{aligned}
||\mathcal{T}u||^2_{\mathcal{H}} &= ||\mathcal{T}(u - \overline{u}) + \mathcal{T}\overline{u}||^2_{\mathcal{H}} \\
&= ||\mathcal{T}(u - \overline{u})||^2_{\mathcal{H}} + 2\langle \mathcal{T}(u - \overline{u}), \mathcal{T}\overline{u} \rangle + ||\mathcal{T}\overline{u}||^2_{\mathcal{H}} \\
&\geq ||\mathcal{T}(u - \overline{u})||^2_{\mathcal{H}} - 2\gamma ||\mathcal{T}(u - \overline{u})||_{\mathcal{H}} ||\mathcal{T}\overline{u}||_{\mathcal{H}} + \gamma^2 ||\mathcal{T}\overline{u}||^2_{\mathcal{H}} + (1 - \gamma^2) ||\mathcal{T}\overline{u}||^2_{\mathcal{H}} \\
&= \left(||\mathcal{T}(u - \overline{u})||_{\mathcal{H}} - \gamma ||\mathcal{T}\overline{u}||_{\mathcal{H}} \right)^2 + (1 - \gamma^2) ||\mathcal{T}\overline{u}||^2_{\mathcal{H}} \\
&\geq (1 - \gamma^2) ||\mathcal{T}\overline{u}||^2_{\mathcal{H}} = (1 - \gamma^2) ||\mathcal{T}1||^2_{\mathcal{H}} MV[u]^2.
\end{aligned}
$$

Hence, we have proven the assertion by defining the constant below

$$C_\gamma := (1 - \gamma^2) ||\mathcal{T}1||^2_{\mathcal{H}}. \tag{3.42}$$

$\square$

Remark 3.4.3. *Note that if a constant function is a singular vector of $\mathcal{T}$, then (3.40) holds true with $\gamma = 0$. In Radon transform the zeroth singular vector is the constant function times the weight function, [66], [70, p. 145-146]. In other words, the constant function is a singular vector up to a positive weight function. Thus the assumption (3.40) seems reasonable for our tomography problem.*

It follows from the lemmata 3.4.1 and 3.4.2 that

$$\frac{1}{2}||u||^2_{\mathcal{L}^1(\Omega)} - \frac{|\Omega|}{C_\gamma}||Tu||^2_{\mathcal{H}} \leq ||u - MV[u]||^2_{\mathcal{L}^1(\Omega)}. \tag{3.43}$$

Together with this last estimation, the lemmata established above will be beneficial to the later development for the norm convergence. In order for preparation, we give the following result.

Corollary 3.4.4. *Let the regularization parameter $\alpha(\delta, f^\delta) \in \overline{S}$ be chosen* a posteriori *for the regularized minimizer $\varphi^\delta_{\alpha(\delta,f^\delta)}$ of the problem (3.2). Then from the result (3.41) and the estimation (3.25), we obtain the following estimation,*

$$C_\gamma |MV[\varphi^\delta_{\alpha(\delta,f^\delta)} - \varphi^\dagger]|^2 \leq (\overline{\tau} + 1)^2 \delta^2. \tag{3.44}$$

Likewise, from Lemma 3.4.1,

$$\begin{aligned}\frac{1}{2}||\varphi^\delta_{\alpha(\delta,f^\delta)} - \varphi^\dagger||^2_{\mathcal{L}^1(\Omega)} &- \frac{|\Omega|}{C_\gamma}(\overline{\tau} + 1)^2 \delta^2 \leq \\ &\leq ||\varphi^\delta_{\alpha(\delta,f^\delta)} - \varphi^\dagger - MV[\varphi^\delta_{\alpha(\delta,f^\delta)} - \varphi^\dagger]||^2_{\mathcal{L}^1(\Omega)} \quad (3.45)\end{aligned}$$

3.4.1 Uniform convergence for smooth-TV regularization via a new lower bound for $D_{J^{\mathrm{TV}}_\beta}$

It is the desired conclusion of this work to show the norm convergence of the regularized solution φ^δ_α of the problem (3.2) to the true solution $\varphi^\dagger$. In order to achieve this by means of Bregman distance, lower and upper bound estimations must be established. The upper bound for the Bregman distance has already been provided by the general convex variational regularization in Theorem 3.3.5. As for the lower bound, we have estimated a lower bound particulary for the Hessian of the smooth-TV operator J^{TV}_β in Theorem 3.2.1, and in Corollary 3.2.2, under certain conditions. Note that, as we will formulate in the following first result, the lower bound is valid for

33

some regularization parameter $\alpha > 0$. However, by theorems 3.3.4 and 3.3.5, the upper bound is valid only for *a posteriori* chosen regularization parameter $\alpha = \alpha(\delta, f^\delta)$. Our first result below just due to the TV functional is about the distance between the gradients of the functions and we obtain this result in the weighted norm.

Theorem 3.4.5. *Denote, for some $\alpha > 0$ and the bounded domain $\Omega \subset \mathbb{R}^N$ where $N = 2$ or 3, by $\varphi_\alpha^\delta, \varphi^\dagger \in \mathcal{W}^{1,2}(\Omega) \subset \mathcal{W}^{1,1}(\Omega)$ the regularized and the true solutions to the problem (3.2) respectively. Let us define $w := \varphi^\dagger + \tau(\varphi_\alpha^\delta - \varphi^\dagger)$ and the weight functional $m_\beta : \mathcal{W}^{1,1}(\Omega) \to \mathbb{R}_+$ by,*

$$m_\beta(w, x) := \frac{\beta}{(|\nabla w(x)|_2^2 + \beta)^{3/2}}, \ \text{for all } x \in \Omega, \tag{3.46}$$

where $0 < \beta < 1$ is fixed. Then,

$$D_{J_\beta^{\mathrm{TV}}}(\varphi^\dagger, \varphi_\alpha^\delta) \geq \frac{1}{2}||\nabla(\varphi^\dagger - \varphi_\alpha^\delta)||_{\mathcal{L}_{m_\beta}^2(\Omega)}^2. \tag{3.47}$$

Proof. According to the setting $w := \varphi^\dagger + \tau(\varphi_\alpha^\delta - \varphi^\dagger)$ for some $\tau \in (0, 1)$, we, by Proposition 2.3.6, have

$$D_{J_\beta^{\mathrm{TV}}}(\varphi^\dagger, \varphi_\alpha^\delta) = \frac{1}{2}(J_\beta^{\mathrm{TV}})''[w](\varphi^\dagger - \varphi_\alpha^\delta, \varphi^\dagger - \varphi_\alpha^\delta). \tag{3.48}$$

We now take into account the estimation (3.12) in the proof of Theorem 3.2.1 as such,

$$\begin{aligned}(J_\beta^{\mathrm{TV}})''[w](\varphi^\dagger - \varphi_\alpha^\delta, \varphi^\dagger - \varphi_\alpha^\delta) &\geq \int_\Omega m_\beta(w, x)||\nabla\left(\varphi^\dagger(x) - \varphi_\alpha^\delta(x)\right)||_2^2 dx \\ &= ||\nabla(\varphi^\dagger - \varphi_\alpha^\delta)||_{\mathcal{L}_{m_\beta}^2(\Omega)}^2.\end{aligned}$$

Hence, the desired result follows from plugging this into (3.48). $\square$

This result only yields that the difference between the gradients of the functions $\varphi^\dagger$ and φ_α^δ can be sufficiently small in the weighted norm sense. However, this does not immediately imply the difference of the functions, which will bring us the norm convergence of φ_α^δ to $\varphi^\dagger$. Recall that in Corollary 3.2.2 the term $||\nabla(\varphi^\dagger - \varphi_\alpha^\delta)||_{\mathcal{L}_{m_\beta}^2(\Omega)}$ has been bounded below by the Poincaré-Wirtinger inequality with the further assumption $w \in \mathcal{W}^{1,\infty}(\Omega)$. Also in the second assertion of Corollary 3.4.4, a lower bound for the Poincaré-Wirtinger inequality has been developed depending on *a posteriori* rule for the choice of the regularization parameter. These establishments stated above allow one to be able formulate the norm convergence theorem. In agreement with Theorem 3.3.5, we, then in the light of corollaries 3.2.2 and 3.4.4, improve the result in Theorem 3.4.5 below.

Theorem 3.4.6. *Let the assumption (3.40) in Lemma 3.4.2 be hold for some constant C_γ defined by (3.42). Furthermore, let the regularization parameter $\alpha(\delta, f^\delta) \in \overline{S} \cap \underline{S}$ be chosen* a posteriori *and let $\varphi^\delta_{\alpha(\delta,f^\delta)}, \varphi^\dagger \in \mathcal{W}^{1,\infty}(\Omega) \subset \mathcal{W}^{1,1}(\Omega)$ be the regularized and the true solutions that are defined over the bounded domain $\Omega \subset \mathbb{R}^N$ where $N = 2$ or 3. Then, there can be derived a functional $l_\beta : \mathcal{W}^{1,\infty}(\Omega) \to \mathbb{R}_+$,*

$$l_\beta(w) := \min_{x \in \Omega}\{m_\beta(w, x)\},$$

which implies the following convergence,

$$\|\varphi^\dagger - \varphi^\delta_{\alpha(\delta,f^\delta)}\|^2_{\mathcal{L}^1(\Omega)} = \mathcal{O}(\Psi(\delta)), \ as \ \delta \to 0. \tag{3.49}$$

Proof. It immediately follows from (3.14) and (3.15) that

$$(J^{\mathrm{TV}}_\beta)''[w](\varphi^\dagger - \varphi^\delta_{\alpha(\delta,f^\delta)}, \varphi^\dagger - \varphi^\delta_{\alpha(\delta,f^\delta)}) \geq l_\beta(w)\|\nabla(\varphi^\dagger - \varphi^\delta_{\alpha(\delta,f^\delta)})\|^2_{\mathcal{L}^2(\Omega)}$$

$$\geq l_\beta(w)\frac{1}{|\Omega|}\left(\left\|\varphi^\dagger - \varphi^\delta_{\alpha(\delta,f^\delta)} - \frac{1}{|\Omega|}\int_\Omega \left(\varphi^\dagger(x) - \varphi^\delta_{\alpha(\delta,f^\delta)}(x)\right)dx\right\|^2_{\mathcal{L}^1(\Omega)}\right),$$

where $l_\beta(w)$ has been defined by (3.13) with $w := \varphi^\dagger + \tau(\varphi^\delta_{\alpha(\delta,f^\delta)} - \varphi^\dagger)$ for some $\tau \in (0, 1)$. This, again by Proposition 2.3.6, implies,

$$D_{J^{\mathrm{TV}}_\beta}(\varphi^\dagger, \varphi^\delta_{\alpha(\delta,f^\delta)}) \geq \frac{C_\Omega}{2|\Omega|}l_\beta(w)\left(\left\|\varphi^\dagger - \varphi^\delta_{\alpha(\delta,f^\delta)} - \frac{1}{|\Omega|}\int_\Omega \left(\varphi^\dagger(x) - \varphi^\delta_{\alpha(\delta,f^\delta)}(x)\right)dx\right\|^2_{\mathcal{L}^1(\Omega)}\right),$$

where the constant C_Ω is from the Poincaré-Wirtinger inequality (2.4). Now the second assertion in Corollary 3.4.4 yields the following

$$D_{J^{\mathrm{TV}}_\beta}(\varphi^\dagger, \varphi^\delta_{\alpha(\delta,f^\delta)}) \geq \frac{C_\Omega}{2|\Omega|}l_\beta(w)\left(\frac{1}{2}\|\varphi^\delta_{\alpha(\delta,f^\delta)} - \varphi^\dagger\|^2_{\mathcal{L}^1(\Omega)} - \frac{|\Omega|}{C_\gamma}(\overline{\tau} + 1)^2\delta^2\right). \tag{3.50}$$

Here the constant C_γ on the right hand side comes from Lemma 3.4.2. In the general sense, in Theorem 3.3.4 or equivalently in Theorem 3.3.5, Bregman distance has been bounded above by the specified choice of the regularization parameter. After rearranging the last inequality, (3.50) reads

$$\frac{C_\Omega}{4|\Omega|}l_\beta(w)\|\varphi^\delta_{\alpha(\delta,f^\delta)} - \varphi^\dagger\|^2_{\mathcal{L}^1(\Omega)} \leq \mathcal{O}(\Psi(\delta)) + \frac{C_\Omega}{2|\Omega|C_\gamma}l_\beta(w)(\overline{\tau} + 1)^2\delta^2. \tag{3.51}$$

Hence, the rate of convergence asserted in (3.49) is obtained since the term $\mathcal{O}(\Psi(\delta))$ on the right hand side dominates the second term. $\qquad \square$

35

Remark 3.4.7. *[**Invertibility of the functional** $l_\beta(w)$ **in** (**3.51**)*] *The proof reveals that in the second assertion of Thoerem 3.4.6, the term $\mathcal{O}(\Psi(\delta))$ also contains $\frac{1}{l_\beta(w)}$, where $w := \varphi^\dagger + \tau(\varphi_\alpha^\delta - \varphi^\dagger)$ for some $\tau \in (0, 1)$. Invertibility of $l_\beta(w)$ is guaranteed by the regularity statement $\varphi_\alpha^\delta, \varphi^\dagger \in \mathcal{W}^{1,\infty}(\Omega)$.*

Remark 3.4.8. *[**Phenomenon of loss of contrast in TV regularization for inverse problems**] According to Remark 3.2.3, with the given further regularity assumption $\varphi_\alpha^\delta, \varphi^\dagger \in \mathcal{W}^{1,\infty}(\Omega)$, the denominator in (3.13) has been bounded and the functional l_β has been derived. In the regions with moderate gradients, the reconstruction error is expected to be small up to an additive constant by the Poincáre - Wirtinger inequality (cf. [4]),*

$$MV[\varphi^\dagger - \varphi_{\alpha(\delta, f^\delta)}^\delta] = \frac{1}{|\Omega|} \int_\Omega \left(\varphi^\dagger(x) - \varphi_{\alpha(\delta, f^\delta)}^\delta(x) \right) dx. \tag{3.52}$$

However, although edge preservation is not in the focus of this work, in the regions that are enclosed by edges this constant may not be neglectible. This is the phenomenon of loss of contrast in TV regularization for inverse problems. Refer to recent works [3, 39] whereby this phenomenon has been demonstrated experimentally.

Although theorems 3.4.5 and 3.4.6 have been stated for $D_{J_\beta^{\mathrm{TV}}}(\varphi^\dagger, \varphi_{\alpha(\delta, f^\delta)}^\delta)$, the same results still hold for $D_{J_\beta^{\mathrm{TV}}}(\varphi_{\alpha(\delta, f^\delta)}^\delta, \varphi^\dagger)$ under the regularity property $\varphi_{\alpha(\delta, f^\delta)}^\delta, \varphi^\dagger \in \mathcal{W}^{1,\infty}(\Omega)$. Thus, due to definition given in (2.18), we can give the following twofold results below.

Corollary 3.4.9. *Over the bounded domain $\Omega \subset \mathbb{R}^N$ where $N = 2$ or 3, let the regularization parameter $\alpha = \alpha(\delta, f^\delta) \in \underline{S} \cap \overline{S}$ be chosen according to the discrepancy principle, and denote by $\varphi_{\alpha(\delta, f^\delta)}^\delta \in \mathcal{W}^{1,2}(\Omega) \subset BV(\Omega)$ the regularized solution for the problem (3.2). Also, let $\varphi^\dagger \in \mathcal{V} \subseteq \mathcal{W}^{1,1}(\Omega)$ be the true solution satisfying the variational source condition in Assumption 3.3.1. Then, in the light of Theorem 3.4.5, following chain of inequalities holds*

$$\frac{1}{2} ||\nabla(\varphi^\dagger - \varphi_{\alpha(\delta, f^\delta)}^\delta)||_{\mathcal{L}_{m_\beta}^2(\Omega)}^2 \leq D_{J_\beta^{\mathrm{TV}}}^{\mathrm{sym}}(\varphi_{\alpha(\delta, f^\delta)}^\delta, \varphi^\dagger) = \mathcal{O}(\Psi(\delta)), \ as \ \delta \to 0. \tag{3.53}$$

Furthermore, with the additional regularity property on the data, $\varphi_\alpha^\delta, \varphi^\dagger \in \mathcal{W}^{1,\infty}(\Omega) \subset \mathcal{W}^{1,1}(\Omega)$, and with the assumption (3.40) in Lemma 3.4.2, the following holds true by Theorem 3.4.6,

$$\frac{1}{2|\Omega|} ||\varphi^\dagger - \varphi_{\alpha(\delta, f^\delta)}^\delta||_{\mathcal{L}^1(\Omega)}^2 = \mathcal{O}(\Psi(\delta)), \ as \ \delta \to 0. \tag{3.54}$$

Chapter 4

NUMERICAL RESULTS

We devote this chapter solely to the illustration of our numerical results. We will employ some well-known gradient based algorithms associated with convex objective functional F_α in the problem (3.19). Impact of the regularization parameter on the convergence analysis together with a detailed benchmarking against different algorithms are provided.

To be able produce the optimal solution $\varphi_\alpha^{\nu+1}$, we implement different gradient-based algorithms, with a suitable step length η, (cf. [**13**] and [**14, Eq 2.1**]),

$$\varphi_\alpha^{\nu+1} = \varphi_\alpha^\nu - \eta \nabla F_\alpha(\varphi_\alpha^\nu). \tag{4.1}$$

Recall that the objective functional F_α in the problem (3.19) is convex. We observe sufficient decay in the following components that we can claim the optimum solution as a result of any algorithm.

- $F(\varphi_\alpha^\nu, f^\delta)$; the functional value at every updated point φ_α^ν,

- $\frac{\|\varphi_\alpha^\nu - \varphi^\dagger\|}{\|\varphi^\dagger\|}$; the relative error value of the reconstruction against the true solution $\varphi^\dagger$,

- $\|\nabla F(\varphi_\alpha^\nu)\|$; the norm of the gradient value of the functional at every updated point φ_α^ν.

- $\|\mathcal{T}\varphi_\alpha^\nu - f^\delta\|$; the discrepancy of the image of the solution against the given data f^δ.

4.1 Generation of the Synthetic Profile

We introduce the true function $\varphi^\dagger$ for synthetically generated refractivity profile. One can find the physics of atmosphere in [**60, Ch. 4**]. The vertical profile of the refractivity $\varphi^\dagger$ can be approximated by an exponential function with given scale heights H_{1c} and H_{2c},

$$\varphi^\dagger(h) = \frac{N_0}{2} \left(\exp\left\{ -\frac{h}{H_{1c}} \right\} + \exp\left\{ -\frac{h}{H_{2c}} \right\} \right). \tag{4.2}$$

Linear functions of x and y would introduce gradients along these axes. Periodical variations are modelled to define horizontal profile,

$$\varphi^\dagger(x,y) = N_0 + \frac{N_x x}{\Delta_x} + \frac{N_y y}{\Delta_y} + N_1 \sin\left(\frac{2\pi \mu_x x}{\Delta_x} \right) + N_2 \cos\left(\frac{2\pi \mu_y y}{\Delta_y} \right), \tag{4.3}$$

where $\Delta_x = x_{\max} - x_{\min}$ and $\Delta_y = y_{\max} - y_{\min}$, N_1 and N_2 are the amplitudes of the periodic variations, μ_x and μ_x are the corresponding frequencies which are normalized to the x and y intervals. Combining everything one gets a three dimensional refractivity field with number of parameters

$$\varphi^\dagger(x,y,h) = \frac{N_0}{2} \left[N_0 + \frac{N_x x}{\Delta_x} + \frac{N_y y}{\Delta_y} + N_1 \sin\left(\frac{2\pi \mu_x x}{\Delta_x} \right) + N_2 \cos\left(\frac{2\pi \mu_y y}{\Delta_y} \right) \right]$$
$$\left(\exp\left\{ -\frac{h}{H_{1c}} \right\} + \exp\left\{ -\frac{h}{H_{2c}} \right\} \right). \tag{4.4}$$

For the parameters defined as $\mu_x = 4$, $\mu_y = 6$, $N_0 = 350$, $H_{1c} = 1$, $H_{2c} = 7$, $N_x = 30$, $N_y = 50$, N_1 and N_2 can be chosen in a way $N_0 - N_1 - N_2 \geq 200$ and $N_0 + N_1 + N_2 \leq 400$. Below in Figure 4.1, true and noisy functions can be seen for the numerical experiments.

4.2 Discretized Form of the Minimization Problem and the Toy Model Setup

Up to this point, we have analysed the optimum solution to our minimization problem (3.19) in the infinite dimension. However, in the computerized environment we
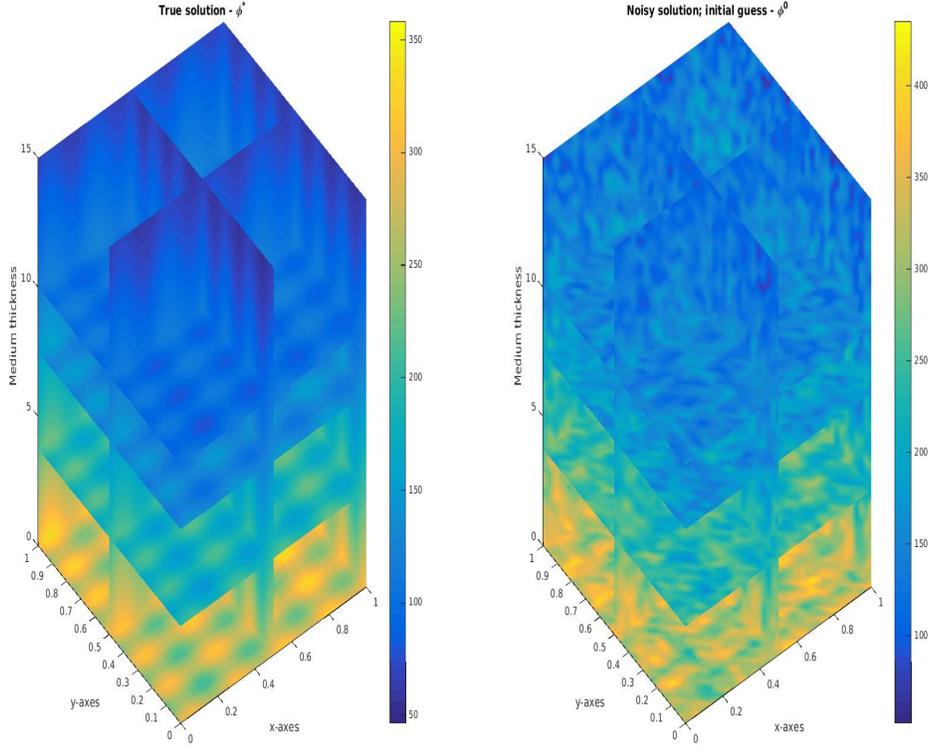
Figure 4.1: Simulated true and noisy solutions for the numerical experiments. The domain Ω has been discretized by $30 \times 30 \times 30$ points.

always work with finite dimensional setup. So, we now introduce our tomographic application and the minimization problems with their components in the finite dimension. We consider the domain $\Omega = [0,1] \times [0,1] \times [0,15]$ and the meshsize $\Delta_{\mathbf{x}} = 1/(N-1)$ with some determined mesh point number $N \in \mathbb{N}$ for any point $\mathbf{x} = (x,y,z) \in \Omega \subset \mathbb{R}^3$. Note that, here $h^{\infty} = 15$ according to (1.13). Within our compact domain $\Omega \subset \mathbb{R}^3$, we then generate a point-to-point discretization by starting from some point $\mathbf{x}_{i-1} \in \Omega \subset \mathbb{R}^3$ and iterating onward as such

$$\mathbf{x}_i = \mathbf{x}_{i-1} + \Delta_{\mathbf{x}}, \text{ for each } i = 1, \cdots, N.$$

In our experiments, we have developed $N_x = N_y = N_z = 30$ nodes to have $N = 27000$ nodes. The speed of light in (1.1) is taken $c = 1$, in order to be able to measure the

propagation of the light beams in space instead of in time. Since this measurement procedure takes place between finitely number of source and receiver setup, we collect discrete data. Recall from the Section 1.2 by (1.13) that the electromagnetic signals with the angles (ρ_r, σ_r) arrive in any receiver $\mathbf{s}$ with the polar angles (ρ_s, σ_s) in various directions $\theta \in \mathbb{S}_\mathbf{s} \subset \mathbb{S}^2$, where $\mathbb{S}_\mathbf{s}$ has been introduced in (1.9). So the ray in $\mathbb{R}^3$ is the set

$$\gamma_{[\mathbf{s},\theta]}(\epsilon) := \left\{ \mathbf{s} + \frac{\epsilon}{\sin(\tilde{\rho})}\theta \,\middle|\, \epsilon \in [z_s, h^\infty] \right\}, \tag{4.5}$$

and the integral transformation that is used for data collection

$$f(\mathbf{s}, \vec{\theta}) = \mathcal{T}_\mathbf{s}\varphi(\vec{\theta}) = (\mathcal{T}\varphi)(\mathbf{s}, \vec{\theta}) = \int_{\mathbf{r} \in \gamma_{[\mathbf{s},\tilde{\theta}]}} \varphi(\mathbf{r})d\mathbf{r} = \int_{\mathbf{z_s}}^{\mathbf{h}^\infty} \varphi(\gamma_{[\mathbf{s},\tilde{\theta}]}(\epsilon)) |\gamma'_{[\mathbf{s},\tilde{\theta}]}(\epsilon)| \mathbf{d}\epsilon.$$

The full path of the signal is the sum of the paths in the intercepted voxels. The model can be interpreted as a system of linear equations. Let us denote the discretized integration by T. With additive white Gaussian noise model vector $z_j \sim \mathcal{N}(0,1)$ (cf. [57]) and some known noise level δ, we produce measurement by

$$[T\varphi]_j = \sum_{i=1}^{N} \varphi_i w_{i,j} = f_j^\dagger + \delta z_j = f_j^\delta, \tag{4.6}$$

where $j = 1, 2, \cdots S$, S is the total number of signal paths from all visible satellites in the network at a fixed time instant, N is the total number of pixel nodes, $w_{i,j}$ is the length of jth ray passing through the node i, the φ_i are considered as density of the corresponding node i, [67].

The parameter function $t(\epsilon) = \frac{\epsilon}{\sin(\tilde{\rho})}$ in (4.5) permits one to determine the points along each signal for any $\epsilon \in [z_s, h^\infty]$ where h^∞ is the upper boundary of the medium as well as the line integral in (1.13), see Figure 4.3.

Regarding the discretized form of our minimization problem (3.19) with its components, we are provided with the compact forward operator $T : \mathbb{R}^N \to \mathbb{R}^M$ and the measurement vector $f^\delta \in \mathbb{R}^M$. As illustrated in Figure 4.2, 240 measurements are distributed nonuniformly and sparsely over evenly meshed 27000 nodes, Fig. 4.4. With this information, our objective functional is then $F_\alpha(\varphi, f^\delta) : \mathbb{R}^{M \times N} \to \mathbb{R}_+$, and we seek for the optimum solution to the problem

$$\varphi_\alpha^{\nu+1} \in \underset{\varphi^\nu \in \mathbb{R}^N}{\arg\min} \left\{ F_\alpha(\varphi^\nu, f^\delta) = \frac{1}{2}||T\varphi^\nu - f^\delta||_2^2 + \alpha J(\varphi^\nu) \right\}. \tag{4.7}$$

Since we have focused on the smoothed total variation regularization in our analysis, we then define the smoothed-TV penalty by

$$J_\beta^{\mathrm{TV}}(\varphi^\nu) := \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \sum_{k=1}^{n_z} \Gamma_\beta \left((D_{ijk}^x \varphi^\nu)^2 + (D_{ijk}^y \varphi^\nu)^2 + (D_{ijk}^z \varphi^\nu)^2 \right) \Delta_x \Delta_y \Delta_z, \qquad (4.8)$$

where the smoothing functional $\Gamma_\beta(\Phi) := \sqrt{|\Phi|_2^2 + \beta}$ for some fixed $\beta \in (0,1)$ and the discretized spatial derivatives according to the central difference form

$$D_{ijk}^x \varphi = \frac{\varphi_{i+1,j,k}^\nu - \varphi_{i-1,j,k}^\nu}{2\Delta_x}, \quad D_{ijk}^y \varphi = \frac{\varphi_{i,j+1,k}^\nu - \varphi_{i,j-1,k}^\nu}{2\Delta_y}, \quad D_{ijk}^z \varphi = \frac{\varphi_{i,j,k+1}^\nu - \varphi_{i,j,k-1}^\nu}{2\Delta_z}.$$
$$(4.9)$$

The optimum solution must satisfy the first optimality conditions in (2.15). That is

$$0 = \nabla F_\alpha(\varphi^\nu, f^\delta) = T^*(T\varphi_\alpha^\nu - f^\delta) + \alpha \nabla J_\beta^{\mathrm{TV}}(\varphi_\alpha^\nu).$$

Here $\nabla J_\beta^{\mathrm{TV}}(\varphi)$ is calculated by $\frac{d}{dt} J_\beta^{\mathrm{TV}}(\varphi + t\Psi)|_{t=0}$ in the direction $\Psi \in \mathcal{C}_c^1(\Omega)$ such that $||\Psi|| \leq 1$. In analogous to the calculations in (3.8), it can be observed that $\nabla J_\beta^{\mathrm{TV}}(\varphi) = L(\varphi)\varphi$ with the nonlinear term $L(\varphi)$,

$$
\begin{aligned}
L(\varphi) &= D_x^T diag(\Gamma'(\varphi))D_x + D_y^T diag(\Gamma'(\varphi))D_y + D_z^T diag(\Gamma'(\varphi))D_z \\
&= \begin{pmatrix} D_x^T & D_y^T & D_z^T \end{pmatrix} \begin{pmatrix} diag(\Gamma'(\varphi)) & 0 & 0 \\ 0 & diag(\Gamma'(\varphi)) & 0 \\ 0 & 0 & diag(\Gamma'(\varphi)) \end{pmatrix} \begin{pmatrix} D_x \\ D_y \\ D_z \end{pmatrix}.
\end{aligned}
$$

4.3 Gradient-Based Algorithms

We will employ different gradient based algorithms in order for numerical illustration. A gradient-based algorithm scheme, with a suitable step length η (cf. [13] and [14, Eq 2.1]), is given by

$$\varphi_\alpha^{\nu+1} = \varphi_\alpha^\nu - \eta \nabla F_\alpha(\varphi_\alpha^\nu). \qquad (4.10)$$

The gradient method takes at each iteration a step along the negative gradient direction, which is steepest descent direction. Below, we review the essentials of any gradient based algorithm which are step length, search directions and trust region.
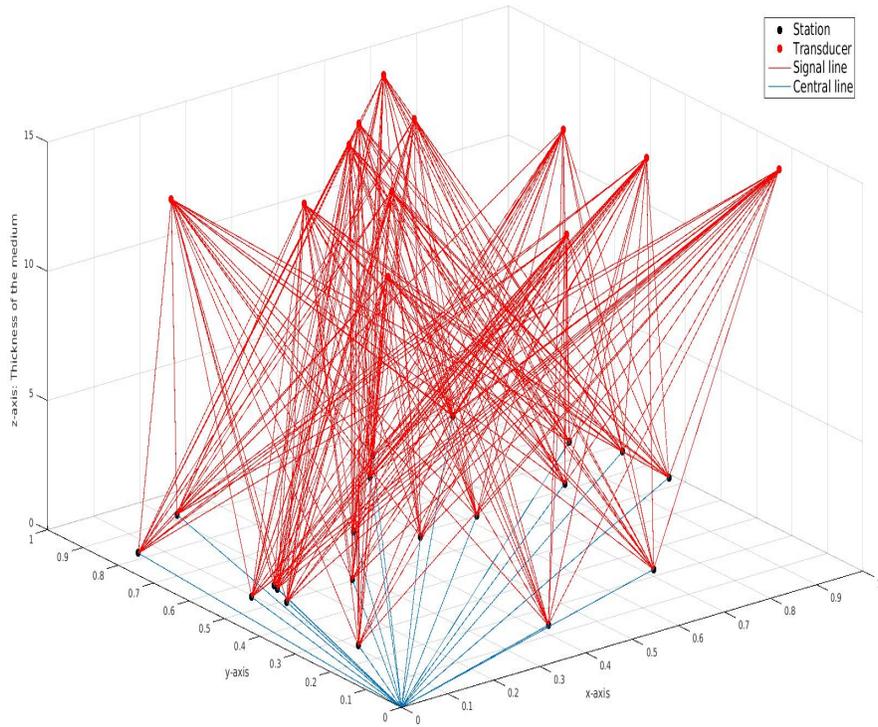
Figure 4.2: A 3-D network is illustrated over a non-uniformly scaled domain. Black dots indicate stations whilst signals penetrate the area of interest through red dots. In this illustration, 20 ground station (receiver) intercept signals emitted by 12 transducers and all are randomly distributed over 27000 cubes.

4.4   Step Length

In line search strategy, the algorithm produces the update φ_α^ν by moving along the chosen direction p^ν. By solving the following problem with respect to the step length η, we determine how far the algorithm should move along p^ν,

$$\min_{\eta>0}\{F_\alpha(\varphi_\alpha^\nu + \eta p^\nu)\}.$$

At the new update φ^ν, the functional value must decrease in comparison to the previous step,

$$F_\alpha(\varphi_\alpha^\nu) \geq F_\alpha(\varphi_\alpha^{\nu-1}).$$

At each iteration step ν, new search direction and step length are computed until a sufficient decrease is obtained in the objective functional F_α. This process stops when certain termination conditions for the line search algorithm are satisfied. The most popular ones are that of the *Wolfe conditions.* According to the first requirement of the conditions, the decrease in the objective function $F_\alpha(\varphi_\alpha^\nu)$ should be proportional to both step length η and the directional derivative $\nabla F_\alpha(\varphi_\alpha^\nu)^T p^\nu$. That is

$$F_\alpha(\varphi_\alpha^\nu + \eta p^\nu) \leq F_\alpha(\varphi_\alpha^\nu) + c_1 \eta \nabla F_\alpha(\varphi_\alpha^\nu)^T p^\nu, \tag{4.11}$$

for some constant $c_1 \in (0,1)$. In practice, c_1 can be chosen as quite small, say $c_1 = 10^{-4}$. This condition is sometimes called the *Armijo condition.* The second requirement is called the *curvature condition,* which requires η to satisfy

$$\nabla F_\alpha(\varphi_\alpha^\nu + \eta p^\nu)^T p^\nu \geq c_2 \nabla F_\alpha(\varphi_\alpha^\nu)^T p^\nu, \tag{4.12}$$

for some another constant $c_2 \in (c_1, 1)$.

4.5 Search Directions for Line Search Methods

Taylor expansion of the functional $F_\alpha : \mathbb{R}^N \to \mathbb{R}_+$ for any search direction p and step-length parameter η,

$$F_\alpha(\varphi_\alpha^\nu + \eta p) = F_\alpha(\varphi_\alpha^\nu) + \eta p^T \nabla F_\alpha(\varphi_\alpha^\nu) + \frac{1}{2}\eta^2 p^T \nabla^2 F_\alpha^\nu(\varphi_\alpha^\nu + tp), \tag{4.13}$$

for some $t \in (0, \eta)$. Here $p^T \nabla F_\alpha(\varphi_\alpha^\nu)$ is the rate of change in F_α along the direction p at φ_α^ν. Hence, p is the solution to the problem

$$\min_p \{p^T \nabla F_\alpha(\varphi_\alpha^\nu)\}, \text{ subject to } ||p|| = 1. \tag{4.14}$$

By definition of inner product, [**62, pp. 634, Subsection 3.2**],

$$p^T \nabla F_\alpha(\varphi_\alpha^\nu) = ||p|| ||\nabla F_\alpha(\varphi_\alpha^\nu)|| \cos\theta,$$

where θ is the angle between p and $\nabla F_\alpha(\varphi_\alpha^\nu)$. We have $p^T \nabla F_\alpha(\varphi_\alpha^\nu) = ||\nabla F_\alpha(\varphi_\alpha^\nu)|| \cos\theta$, from $||p|| = 1$. So the objective in (4.14) is minimized when $\cos\theta$ takes on its minimum value of -1 at $\theta = \pi$ radians, [**71, pp. 22**]. Hence, the solution to (4.14) is

$$p = -\frac{\nabla F_\alpha(\varphi_\alpha^\nu)}{||\nabla F_\alpha(\varphi_\alpha^\nu)||}. \tag{4.15}$$

Later we will also define another strategy named as *trust region.* Trust region is a ball defined by $||p||_2 \leq \Delta$, where the scalar $\Delta > 0$ is called the trust region radius.

4.5.1 Newton search direction

Newton direction is one of the most important search direction method. It is derived from the second-order Taylor expansion. Setting the step length parameter $\eta = 1$ in (4.13) brings,

$$F_\alpha(\varphi_\alpha^\nu + p) = F_\alpha(\varphi_\alpha^\nu) + p^T \nabla F^\nu + \frac{1}{2} p^T \nabla^2 F_\alpha(\varphi_\alpha^\nu) p, \tag{4.16}$$

By simply setting the derivative with respect to p of this expansion to zero, one obtains

$$p = -\frac{1}{2} [\nabla^2 F_\alpha(\varphi_\alpha^\nu)]^{-1} \nabla F_\alpha(\varphi_\alpha^\nu) \tag{4.17}$$

the Newton step. The Newton direction can be used in a line search method when $\nabla^2 F_\alpha$ is positive definite, [**71, p. 23**].

4.5.2 Quasi-Newton search direction and approximate Hessian

Unlike in the Newton search direction, quasi-Newton direction does not require the calculation of the Hessian $\nabla^2 F_\alpha$. In the quasi-Newton line search strategy, in place of the Hessian, we use an approximation matrix B_ν. Along the determined search direction, information about the second derivative of F_α is gained after updating the matrix B_ν at each iteration step $\nu = 1, 2, \cdots$.

Below we will review and employ some quasi-Newton methods for the numerical illustration of our tomographic problem.

4.6 Lagged Diffusivitiy Fixed Point Iteration - (LDFP)

LDFP is also in the class of quasi-Newton search direction algorithm. We have tested the LDFP algorithm with two different penalty terms which are quadratic Tikhonov

and the smoothed-TV forms. In Section A.1 of Appendix A, the results of LDFP associated with the quadratic Tikhonov type objective functional can be seen.

4.6.1 LDFP with smooth-TV objective functional

The favourite regularization strategy of this work is TV regularization. Therefore, we would like to begin with one of the simplest algorithm to find approximate solution for the problem (3.2). In (4.10), when the functional $F_\alpha(\varphi, f^\delta)$ is defined by

$$F_\alpha(\varphi, f^\delta) = \frac{1}{2}||\mathcal{T}\varphi - f^\delta||_2^2 + \alpha \int_\Omega \sqrt{|\nabla\varphi(x)|_2^2 + \beta}dx,$$

then well-known *lagged diffusivitiy fixed point iteration*, [**86, 87**], is given by the following scheme

$$
\begin{aligned}
\varphi_\alpha^{\nu+1} &= \varphi_{\alpha,\nu} + (\mathcal{T}^*\mathcal{T} + \alpha L(\varphi_\alpha^\nu))^{-1}\nabla F_\alpha(\varphi, f^\delta) \\
&= (\mathcal{T}^*\mathcal{T} + \alpha L(\varphi_\alpha^\nu))^{-1}\mathcal{T}^*f^\delta = \mathcal{R}_\alpha(\varphi_\alpha^\nu), \quad \nu = 0, 1, \cdots, \quad (4.18)
\end{aligned}
$$

where,

$$L(\varphi_\alpha^\nu) := -\nabla^* \cdot \left(\frac{\nabla}{(|\nabla\varphi_\alpha^\nu|_2^2 + \beta)^{1/2}}\right).$$

Direct implementation of this scheme would still be a costly iteration procedure since $L(\varphi)$ is highly nonlinear. Then, according to [**86, Algorithm 8.2.3**], the update $\varphi_\alpha^{\nu+1}$ is produced after the following linearization steps;

LDFP algorithm with smooth-TV penalty:

1. Compute $L^\nu := L(\varphi_\alpha^\nu)$ anisotropic Laplacian;
2. Compute $g^\nu := \mathcal{T}^*(\mathcal{T}\varphi_\alpha^\nu - f^\delta) + \alpha L^\nu\varphi_\alpha^\nu$ gradient step;
3. Compute $K_\alpha = \mathcal{T}^*\mathcal{T} + \alpha L^\nu$ approximate Hessian;
4. Solve $K_\alpha s^{\nu+1} = -g^\nu$ quasi-Newton step;
5. Update $\varphi_\alpha^{\nu+1} = \varphi_\alpha^\nu + s^{\nu+1}$;

In our experiments, we use usual CGNE for solving the inner system $K_\alpha s^\nu = g^\nu$, see [**46**].

In the Figure 4.5, we present the numerical results of LDFP algorithm with smoothed-TV functional per different number of the measurements. We run the algorithm only for 30 iteration steps to understand its behaviour.

4.7 BFGS Algorithm

Theorem 4.7.1. *[**71**, **Theorem 2.1**]*
    Suppose that $F_\alpha : \mathbb{R}^N \to \mathbb{R}$ is continuously differentiable and that $p \in \mathbb{R}^N$. Then we have that

$$F_\alpha(\varphi_\alpha + p) = \nabla F_\alpha(\varphi_\alpha) + \int_0^1 \nabla^2 F_\alpha(\varphi_\alpha + tp)p\, dt, \tag{4.19}$$

and that

$$\nabla F_\alpha(\varphi_\alpha + p) = \nabla F_\alpha(\varphi_\alpha) + \nabla F_\alpha(\varphi_\alpha)^T p + \frac{1}{2} p^T \nabla^2 F_\alpha(\varphi_\alpha + tp)p, \tag{4.20}$$

for some $t \in (0,1)$.

    From this statement, we have by adding and subtracting $\nabla^2 F_\alpha(\varphi_\alpha)p$ in (4.20) that

$$\nabla F_\alpha(\varphi_\alpha + p) = \nabla F_\alpha(\varphi_\alpha) + \nabla^2 F_\alpha(\varphi_\alpha)p + \int_0^1 [\nabla^2 F_\alpha(\varphi_\alpha + tp) - \nabla^2 F_\alpha(\varphi_\alpha)]p\, dt.$$

Since our functional F_α is smooth,

$$\nabla F_\alpha^{(\nu+1)} = \nabla F_\alpha^{(\nu)} + \nabla^2 F_\alpha^{(\nu)}(\varphi_\alpha^{(\nu+1)} - \varphi_\alpha^{(\nu)}) + o(||\varphi_\alpha^{(\nu+1)} - \varphi_\alpha^{(\nu)}||).$$

We can also write that

$$\nabla^2 F_\alpha^{(\nu)}(\varphi_\alpha^{(\nu+1)} - \varphi_\alpha^{(\nu)}) \approx \nabla F_\alpha^{(\nu+1)} - \nabla F_\alpha^{(\nu)}.$$

The new Hessian approximation B_ν is defined in a way that the *secant equation*

$$B_\nu s_\nu = y_\nu \tag{4.21}$$

holds, [**65**, **p. 91**], [**71**, **p. 24-25**]. Here,

$$s_\nu = \varphi_\alpha^{(\nu+1)} - \varphi_\alpha^{(\nu)}, \text{ and } y_\nu = \nabla F_\alpha^{(\nu+1)} - \nabla F_\alpha^{(\nu)}.$$

The most well-known formulae for updating the Hessian approximation B_ν are that of the *symmetric-rank-one* (SR1), [**71**, **p. 25**, **Eq. 2.17**], defined by

$$B_{\nu+1} = B_\nu + \frac{(y_\nu - B_\nu s_\nu)(y_\nu - B_\nu s_\nu)^T}{(y_\nu - B_\nu s_\nu)^T s_\nu}, \tag{4.22}$$

and the *Broyden-Fletcher-Goldfarb-Shanno (BFGS) formula*, [**71, p. 25, Eq. 2.18**], also given by

$$B_{\nu+1} = B_\nu - \frac{B_\nu s_\nu s_\nu^T B_\nu}{s_\nu^T B_\nu s_\nu} + \frac{y_\nu y_\nu^T}{y_\nu^T s_\nu}. \tag{4.23}$$

It is known that, (cf. [**71, p. 25**]), the BFGS update (4.23) generates positive definite approximation as long as the initial approximation B_0 is positive definite and $s_\nu^T y_\nu > 0$.

The quasi-Newton search direction is given by using B_ν in place of the exact Hessian in (4.17), that is,

$$p_\nu = -B_\nu^{-1} \nabla F_\alpha^\nu. \tag{4.24}$$

Consider the quadratic convex objective function

$$m_\nu(p) = F_\alpha^\nu + \nabla F_\alpha^T p + \frac{1}{2} p^T B_\nu p. \tag{4.25}$$

Here B_ν is an $N \times N$ symmetric positive definite matrix and it is updated at every iteration. This convex quadratic objective function attains its minimum at

$$p_\nu = -[B_\nu]^{-1} \nabla F_\alpha^\nu. \tag{4.26}$$

The mimizer p_ν is used as the search direction,

$$\varphi_\alpha^{\nu+1} = \varphi_\alpha^\nu + \eta_\nu p_\nu, \tag{4.27}$$

where the step length η_ν satisfies the Wolfe conditions (4.11) and (4.12).

In order to be able to determine B_ν uniquely, we solve the following problem

$$\min_B ||B - B_\nu|| \tag{4.28}$$

$$\text{subject to } B = B^T, \ Bs_\nu = y_\nu, \tag{4.29}$$

47

where $s_\nu^T y_\nu > 0$. With these preparations having made, we can now introduce the BFGS method.

BFGS method:

1. Given starting point φ^0, convergence tolerance $\epsilon > 0$, inverse Hessian approximation B_0;
2. Start iteration while $||\nabla F_\alpha^\nu|| > \epsilon$;
3. Compute search direction in (4.26);
4. Set $\varphi_\alpha^{\nu+1} = \varphi_\alpha^\nu + \eta p_\nu$, where η is the step length satisfying the Wolfe conditions (4.11) and (4.12);
5. Define $s_\nu = \varphi_\alpha^{\nu+1} - \varphi_\alpha^\nu$ and $y_\nu = \nabla F_\alpha^{\nu+1} - \nabla F_\alpha^\nu$;
6. Compute B_ν by means of (4.23);

4.8 Large-Scale Quasi-Newton

The quasi-Newton methods cannot be directly applicable to large optimization problems because their approximations to the Hessian or its inverse are usually dense. The storage and computational requirements grow in proportion to N^2, and become excessive for large N. In order to overcome this difficulty, limited-memory quasi-Newton methods have been introduced, [**65, 71**]. Here, we particularly focus on *limited memory BFGS* (L-BFGS) algorithm.

4.8.1 Limited memory BFGS (L-BFGS)

We begin description of the L-BFGS method with recalling the early knowledge about BFGS. At each iteration step in BFGS, the update $\varphi_\alpha^{\nu+1}$ is produced by

$$\varphi_\alpha^{\nu+1} = \varphi_\alpha^\nu - \eta B_\nu \nabla F_\alpha^\nu, \ \nu = 0, 1, 2, \cdots. \tag{4.30}$$

Here the step length is denoted by η and the matrix B_ν by means of

$$B_{\nu+1} = V_\nu^T B_\nu V_\nu + \rho_\nu s_\nu s_\nu^T \tag{4.31}$$

where

$$\rho_\nu = \frac{1}{y_\nu s_\nu}, \ V_\nu = I - \rho_\nu y_\nu s_\nu^T \tag{4.32}$$

and

$$s^{\nu+1} = \varphi_\alpha^{\nu+1} - \varphi_\alpha^\nu, \ y_\nu = \nabla F_\alpha^{\nu+1} - \nabla F_\alpha^\nu. \tag{4.33}$$

The matrix $B_{\nu+1}$ is obtained by updating B_ν using the pair $\{s_\nu, y_\nu\}$. Limited memory techniques amount to generating at each iteration the BFGS matrix from the m (usually $m \in [5, 10]$) most recent of the pairs $\{s_\nu, y_\nu\}$ and the initial matrix B_0. The choice of B_0 is usually carried on by $B_0 = \mu I$ where I is the identity matrix and μ is some scaling, see [**65, Subsection 5.2.2 on p. 93**] and [**81**]. According to [**71, Eq (9.6)**], the scaling factor that attempts to estimate the size of the true Hessian matrix is given by,

$$\mu = \frac{s_\nu^T y_\nu}{y_\nu^T y_\nu}.$$

The L-BFGS algorithm can be stated formally as follows.

L-BFGS method:

1. Choose a starting point φ_0, integer $m > 0$;
2. Initiate iteration $\nu = 0$;
3. Compute p_ν;
4. Compute $\varphi_\alpha^{(\nu+1)} = \varphi_\alpha^{(\nu+1)} + \eta^\nu p_\nu$;

if $\nu > m$, **then**

    5. Discard the vector pair $\{s_{\nu-m}, y_{\nu-m}\}$ from storage;

end if

6. Compute and save $s^{\nu+1} = \varphi_\alpha^{\nu+1} - \varphi_\alpha^\nu, y_\nu = \nabla F_\alpha^{\nu+1} - \nabla F_\alpha^\nu$;

We provide optimized solution from L-BFGS algorithm by employing a novel reverse-communication large-scale nonlinear optimization software SAMSARA, [**64**].

4.8.2   TV gradient step in L-BFGS

We run tests in SAMSARA, [**64**], with smoothed-TV penalty term

$$J_\beta^{\text{TV}}(\varphi) := \int_\Omega \sqrt{|\nabla \varphi(x)|_2^2 + \beta} dx.$$

The Fréchet derivative of the objective functional (3.1) with the smooth-TV penalty term, also the application of the first order optimality condition (2.15) for the smooth-

TV functional, leads to solving the following operator equation

$$(-\alpha\nabla^* \cdot \left(\frac{\nabla}{(\beta + |\nabla\varphi_\alpha^\delta|_2^2)^{1/2}}\right) + \mathcal{T}^*\mathcal{T})\varphi_\alpha^\delta = \mathcal{T}^*f^\delta \tag{4.34}$$

to produce $\varphi_\alpha^\delta \in \mathcal{V}$. We demonstrate different solution per different measurement number, $\{1, 50, 100, 240, 360, 450\}$, in the figures 4.7, 4.8, 4.9, 4.10. It is observed better and more stable convergence rate in the pre-image space with the more measurement number in the image space. Furthermore, the figures 4.11 and 4.12 demonstrate convergence in the pre-image/image spaces with varying amount of noise, $\delta \in \{20\%, 10\%, 5\%, 2\%, 0.1\%, 0.005\%, 0.001\%\}$. As a common expectation from an inverse ill-posed problem, the less amount of noise in the image space provides better and stable convergence rates in the pre-image space.

4.9 Benchmark: LDFP vs SAMSARA with Smooth-TV Penalty

A CPU time based benchmark test between SAMSARA and LDFP both associated with the smoothed-TV gradient step has been conducted, see the figures 4.13, 4.14 and 4.15.
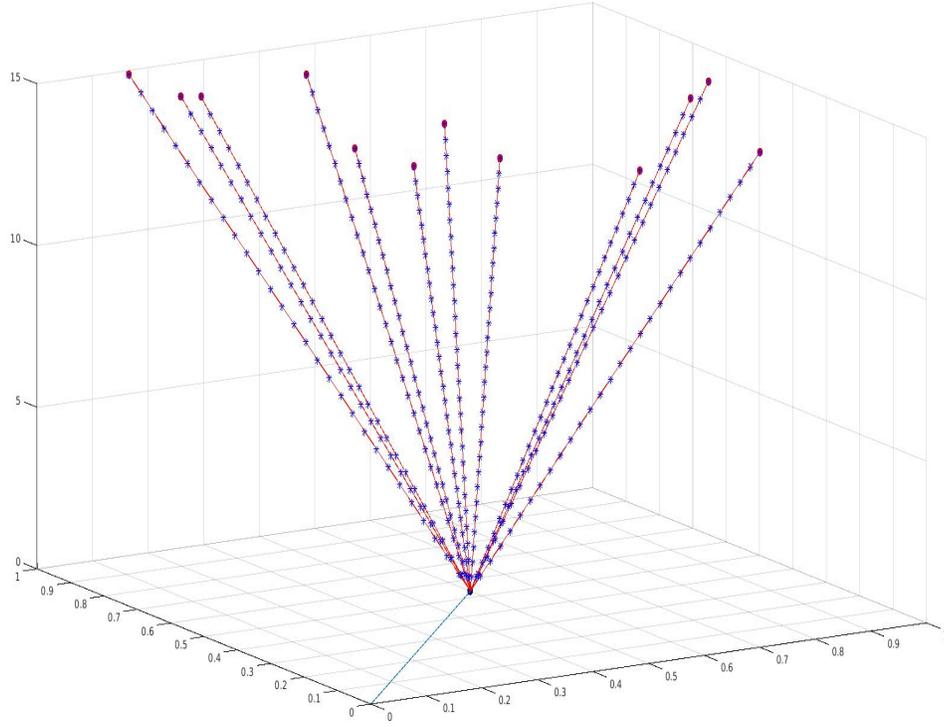
Figure 4.3: Finitely number of points denoted by blue stars along 12 rays are illustrated. These points are found via the parameter function $t(\epsilon) = \frac{\epsilon}{\sin(\tilde{\rho})}$, for $\epsilon \in [z_s, h^\infty]$ where h^∞ is the upper boundary of the line integral in (1.13).
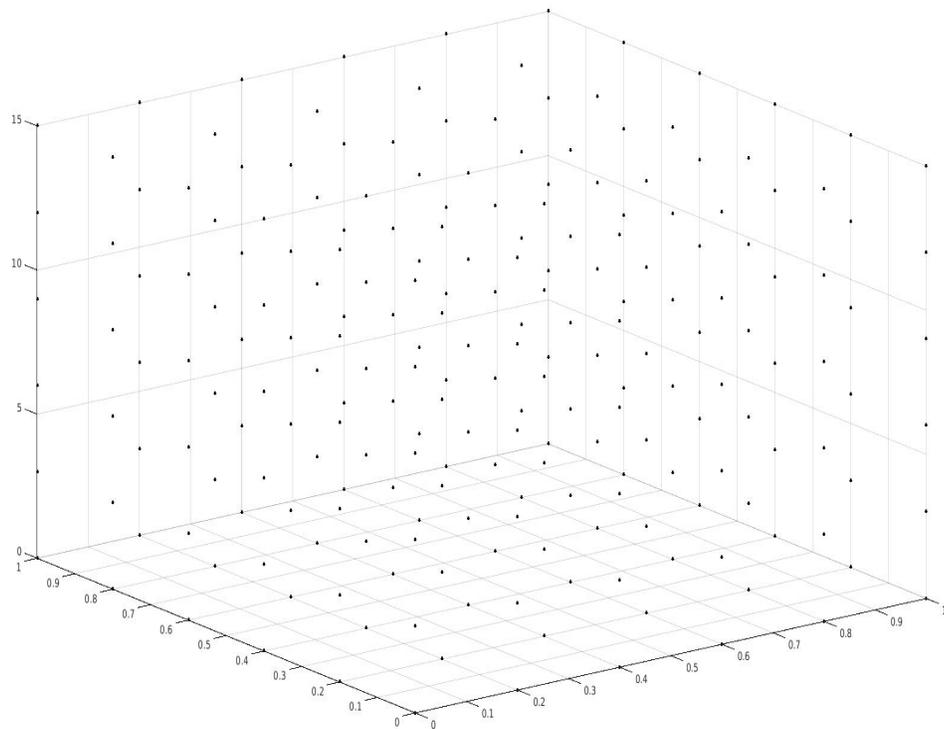
Figure 4.4: Discretization of our area of interest. For illustration purpose, we only present $6 \times 6 \times 6$ pixel nodes.
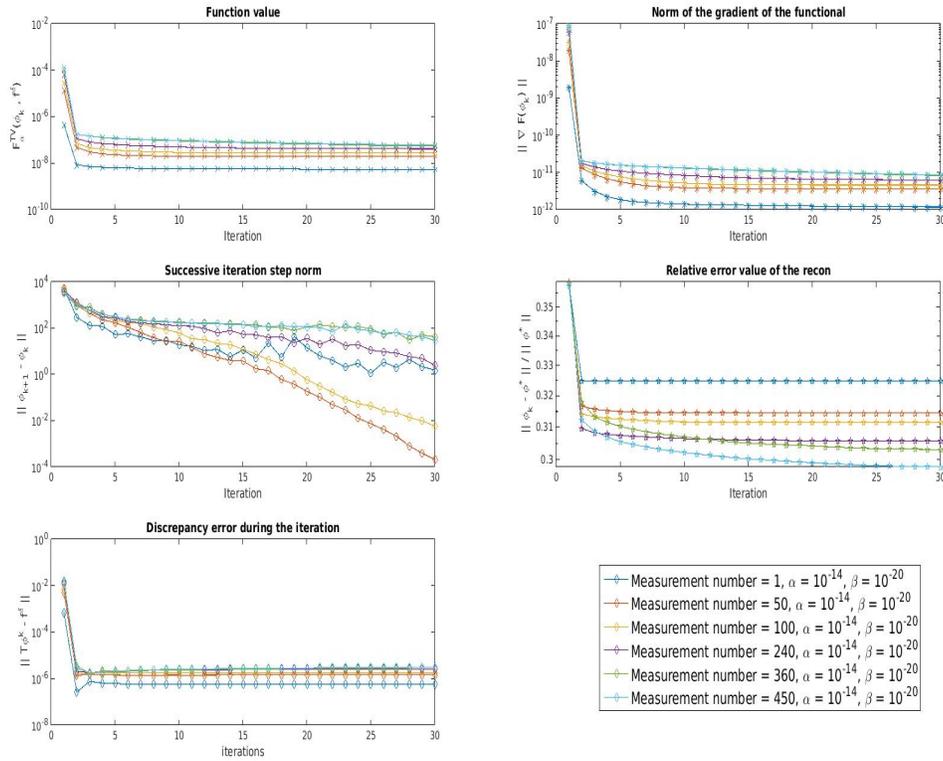
Figure 4.5: LDFP algorithm with smoothed-TV gradient step numerical convergence results per different number of the measurements, $\{1, 50, 100, 240, 360, 450\}$. Regularization parameter is chosen according to the stable behaviour of the discrepancy $||\mathcal{T}\varphi_\alpha^\nu - f^\delta||$ after each iteration step $\nu = 1, 2, 3, \cdots$.
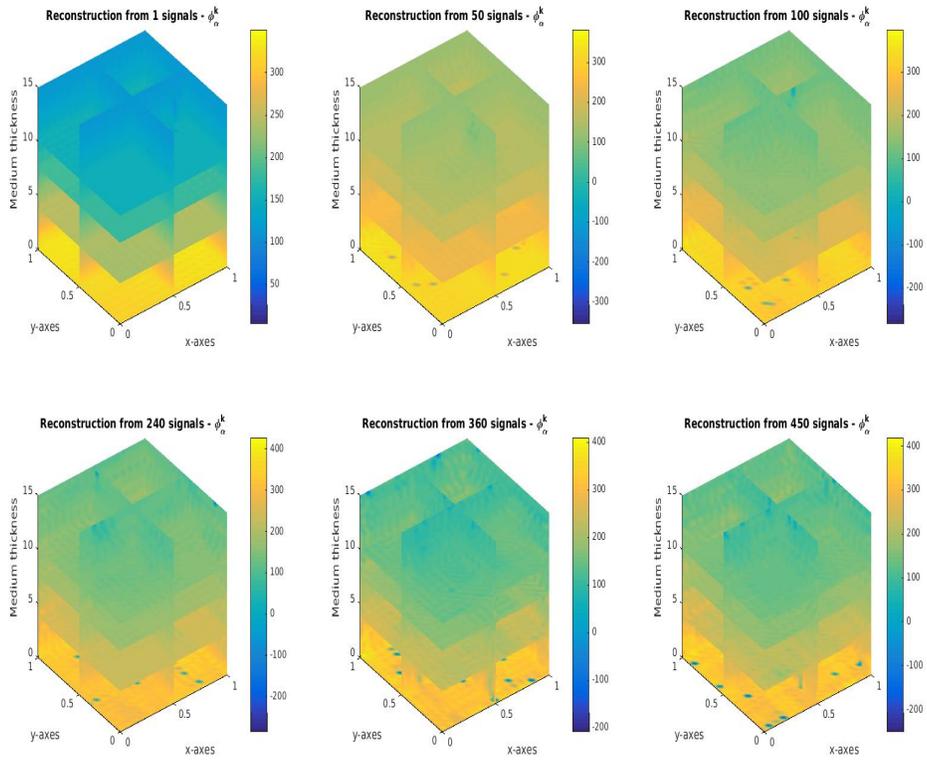
Figure 4.6: LDFP algorithm with smoothed-TV gradient step numerical reconstruction results per different number of the measurements, $\{1, 50, 100, 240, 360, 450\}$.
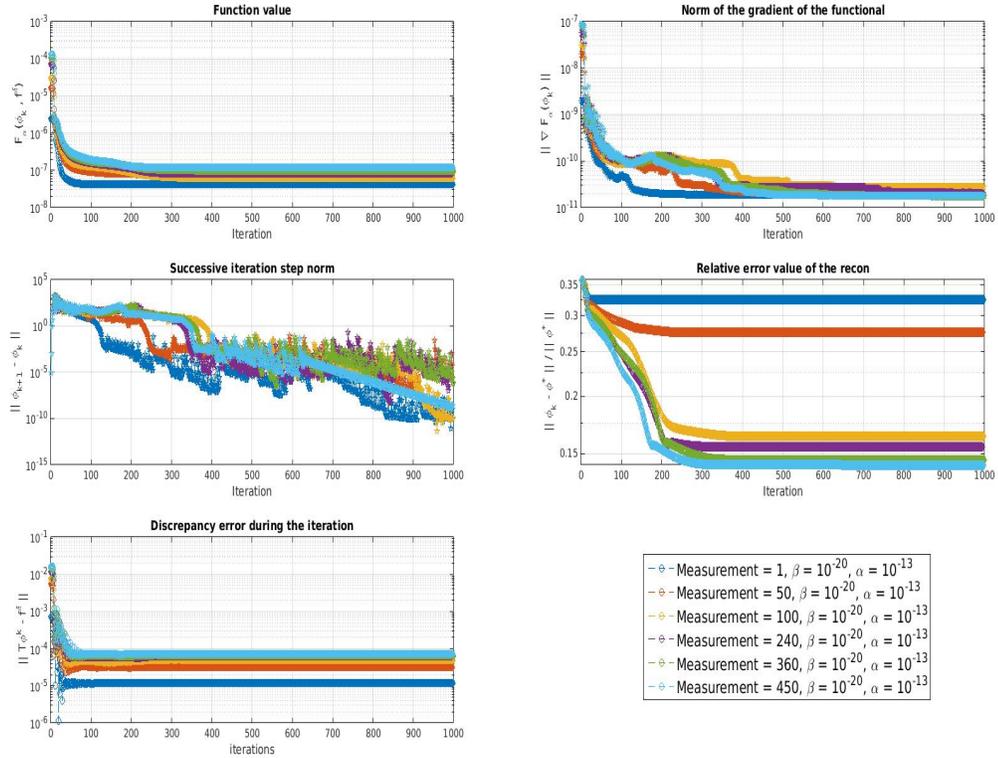
Figure 4.7: SAMSARA with TV gradient step numerical results per measurement. We have conducted our experiment in the software SAMSARA for the measurement number $\{1, 50, 100, 240, 360, 450\}$.
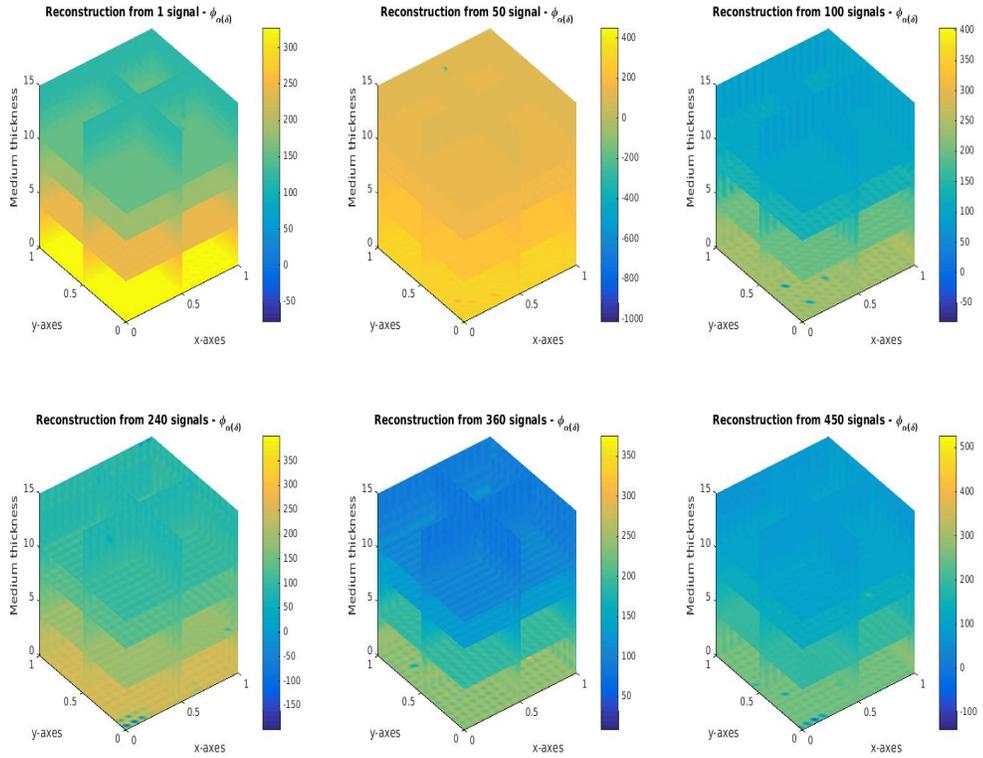
Figure 4.8: SAMSARA with TV gradient step numerical reconstruction results per measurement $\{1, 50, 100, 240, 360, 450\}$. Fixed regularization parameter $\alpha = 10^{-13}$ has been determined according to the behaviour in the discrepancy.
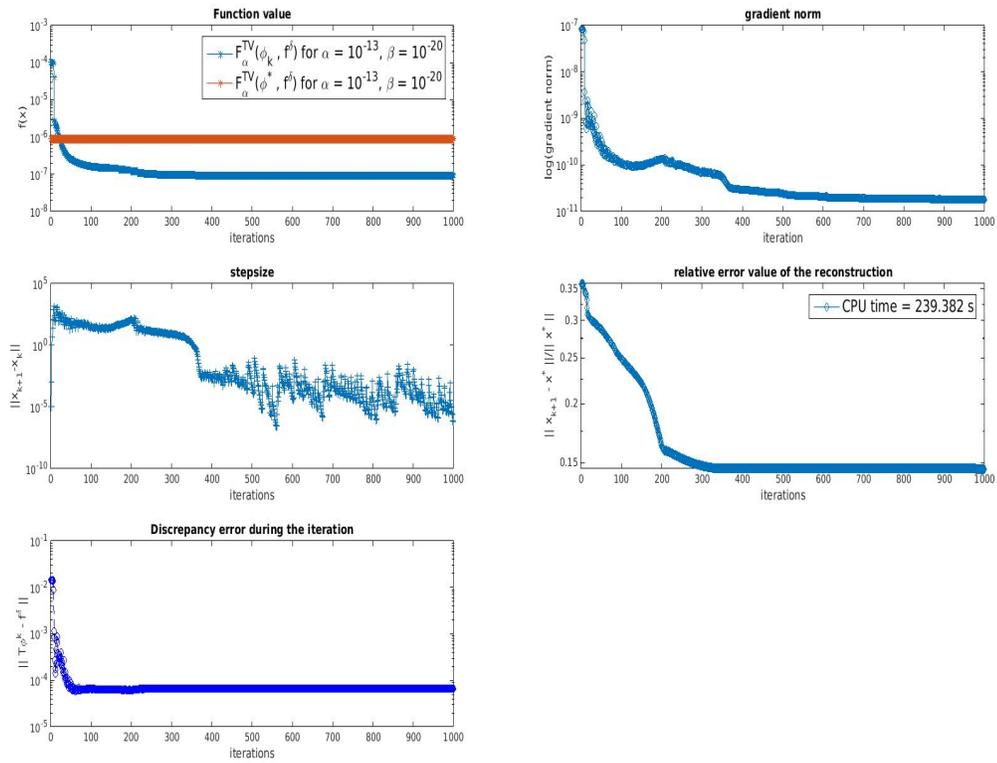
Figure 4.9: SAMSARA with TV gradient step numerical convergence from 360 signals.
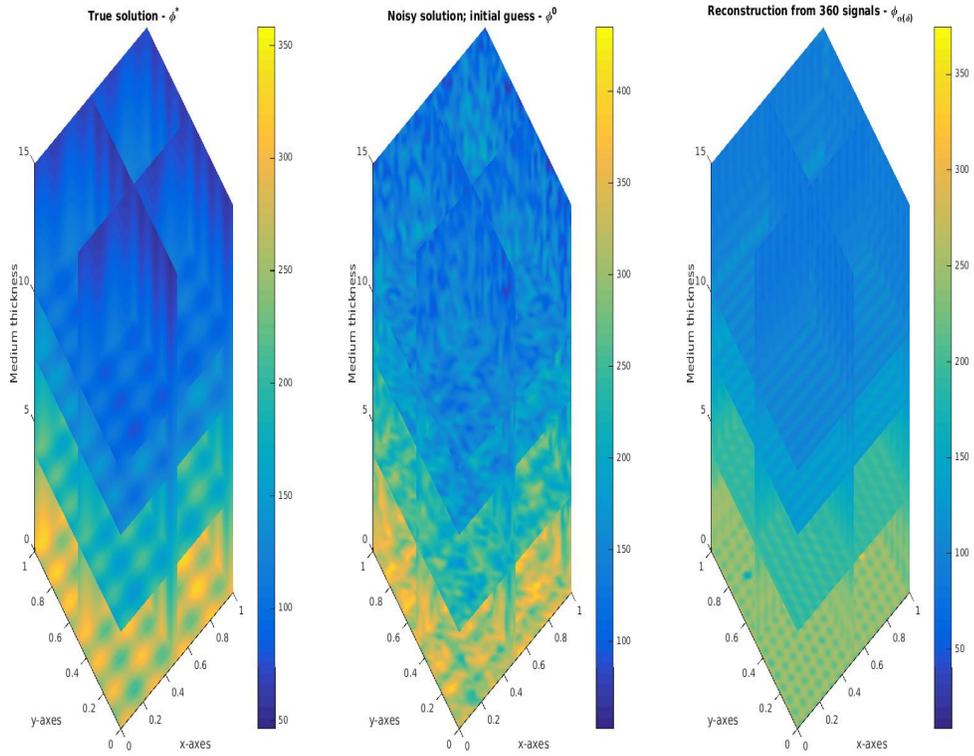
Figure 4.10: SAMSARA with TV gradient step numerical reconstruction from 360 signals.
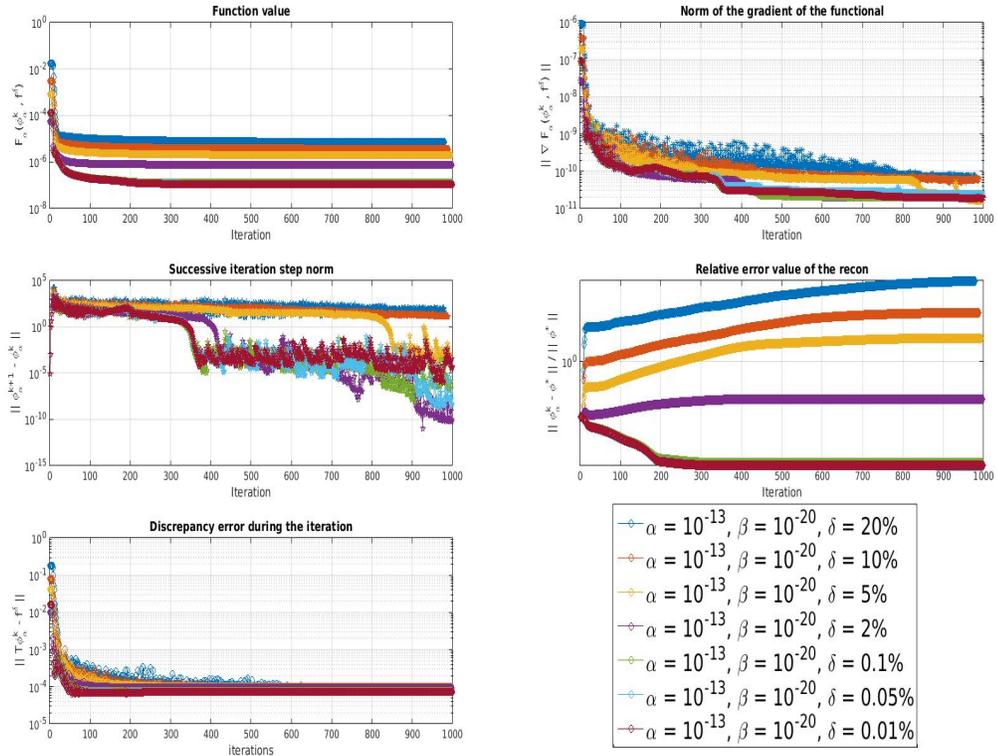
Figure 4.11: SAMSARA with TV gradient step convergence analysis per different noise amount $\{20\%, 10\%, 5\%, 2\%, 0.1\%, 0.005\%, 0.001\%\}$. Convergence in the pre-image space begins with $\delta \leq 2\%$.
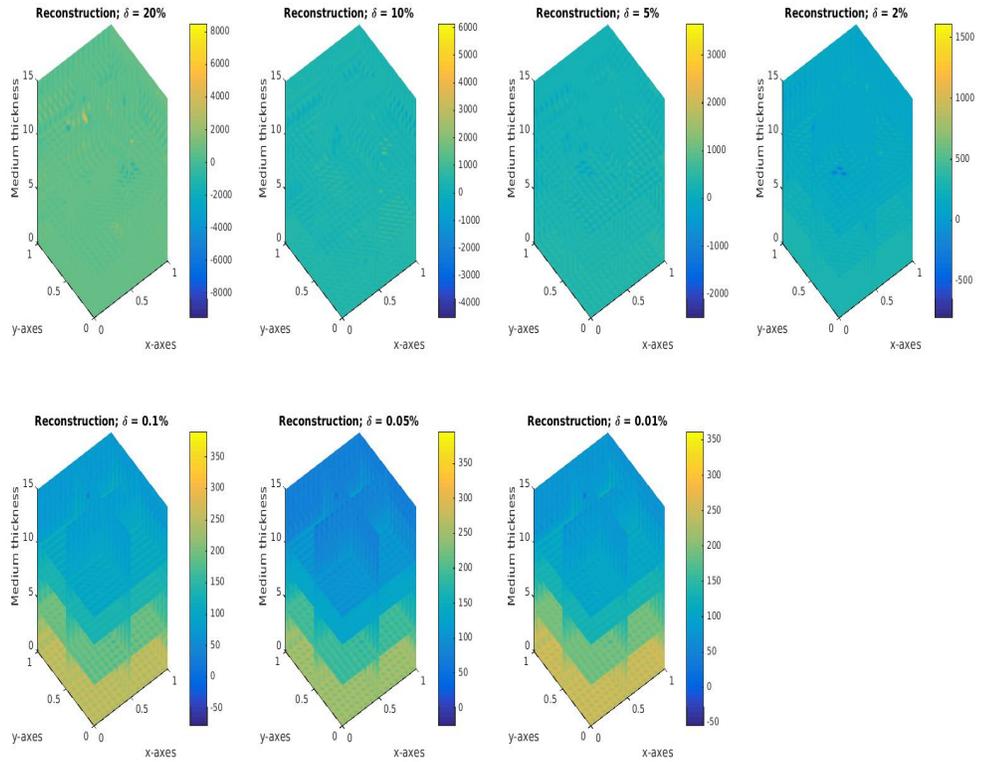
Figure 4.12: SAMSARA with TV gradient step different reconstruction per different noise amount. Corresponding numerical convergence analysis has been provided in Figure 4.11.
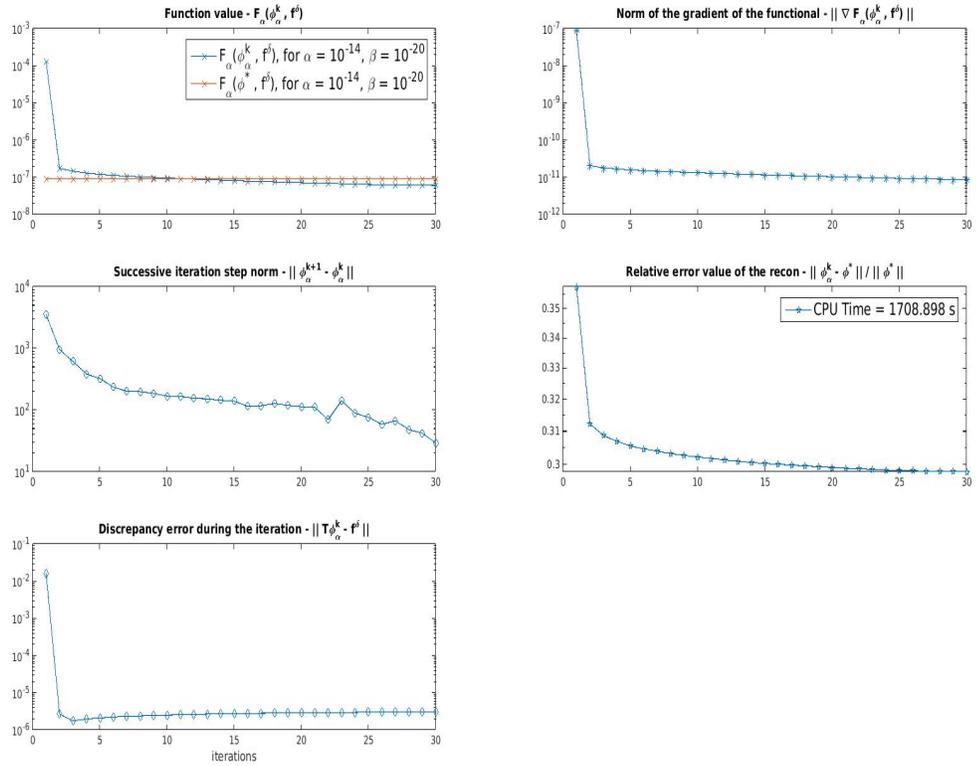
Figure 4.13: LDFP with smoothed-TV gradient convergence analysis; This result has been obtained from 450 measurements.
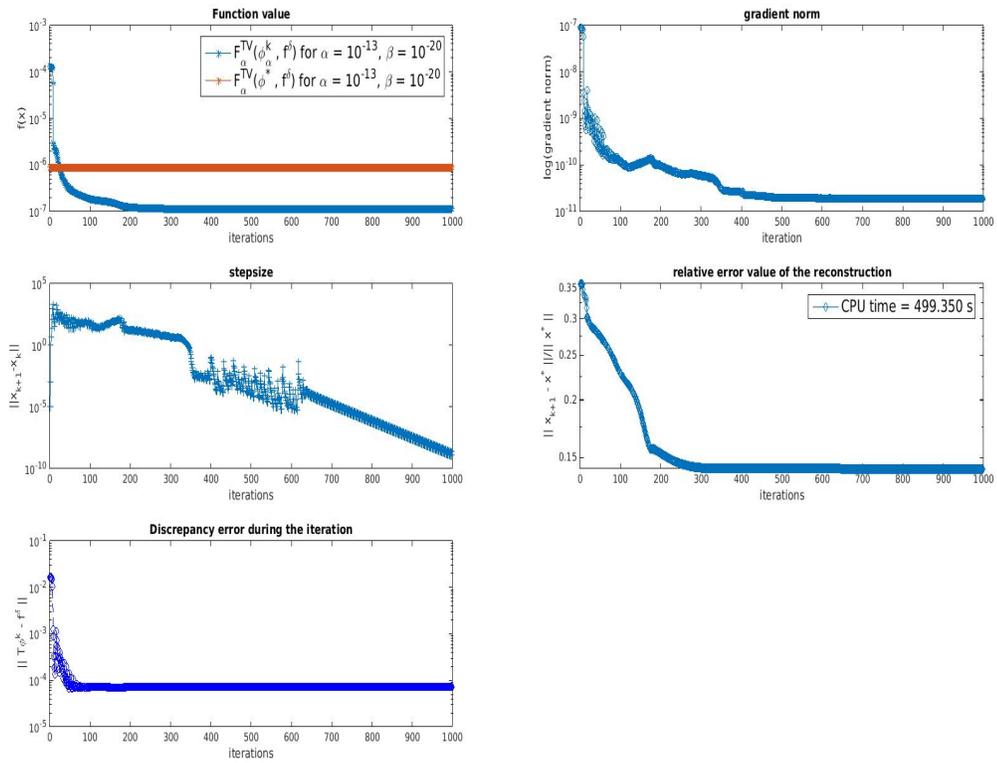
Figure 4.14: SAMSARA with smoothed-TV gradient numerics from 450 measurements.
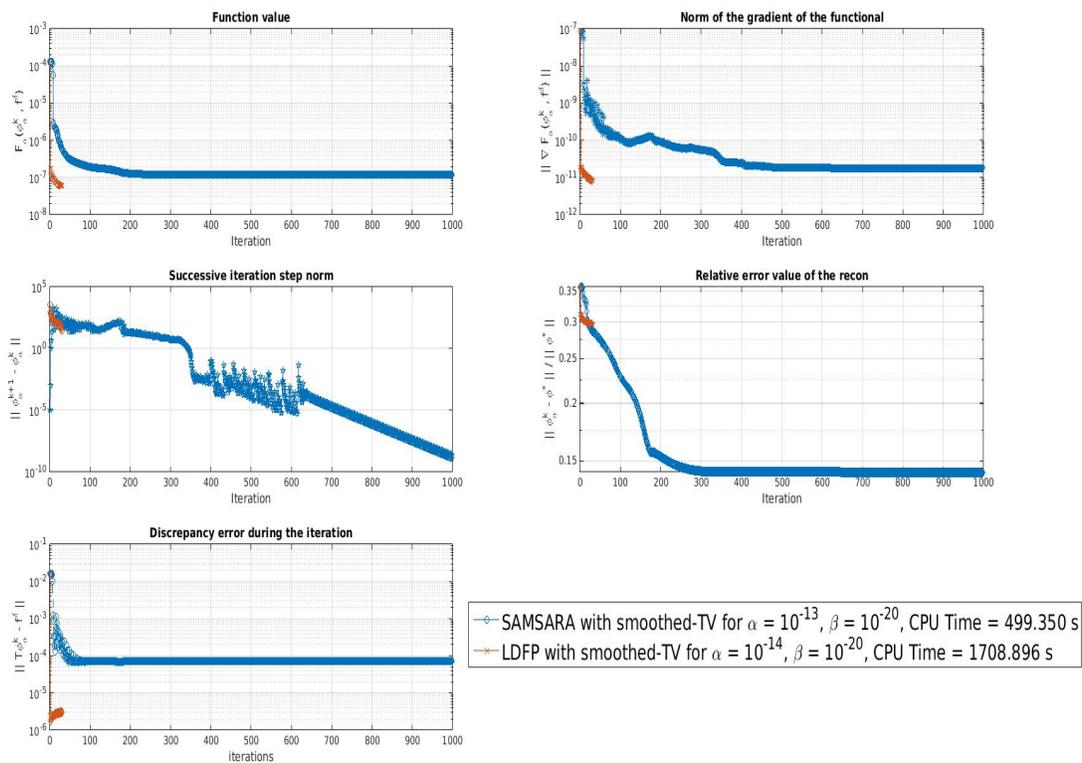
Figure 4.15: Benchmark: SAMSARA and LDFP both with smoothed-TV gradient numerics from 450 measurements. Recall from the figures 4.13 and 4.14 that LDFP got executed only for 30 iteration steps whereas SAMSARA 1000 iteration steps. Comparing the CPU times for each tests, it would take SAMSARA only 14.9805 seconds to iterate 30 times. In other words, SAMSARA works 33 times faster than usual LDFP algorithm with smoothed-TV gradient step.

Chapter 5

CONCLUSION AND FURTHER DISCUSSION

We have considered solving an atmospheric tomography problem, and we have analysed this solution in the infinite dimensional sense. Although in reality, this is a time dependent problem we only have considered fixed time step. That means we have assumed to be provided with certain number of measurement at some fixed time step, and our task is to provide an efficient method to reconstruct three dimensional density profile of a medium. In time dependent case, one must keep in mind that the distribution and the dimension of the measurement vector vary at each time step. Thus, only a finite dimensional analysis enables us to understand the problem and its solution. As for future prospect, a time dependent forward operator could be formulated in order to observe this case. Furthermore, although non-negativity constraint is the case of this problem, we have only focused on analysing stable reconstruction by considering to solve unconstrained smooth minimization problem formulated in (3.19).

As a common characteristic of inverse ill-posed problem, this problem has also incomplete data case. This issue causes nonuniqueness and instability in the solution. The developed analysis in Chapter 3 cannot overcome this issue. It is not possible to reconstruct the unknown function φ exactly from finitely number of measurements. However, [**45, Theorems 5.1 - 5.6**] show that arbitrarily good approximation can be obtained, see Assumption 1.2.1 with Theorem 1.2.2 for details. Since we have focused on TV regularization for an inverse ill-posed problem, it has been the task of this work to analyse the impact of this specific regularization strategy on the given problem. Thus, starting with the assumption that we have been provided with sufficient measurement is reasonable. Theoretical results have been obtained and

associated with the smooth-TV regularization only by some specific choice of the regularization parameter which is Morozov's discrepancy principle (MDP), Subsection 3.3.1. It could be worthwhile to develop an analagous work for Lepskiĭ principle which is another *a posteriori* strategy for the choice of regularization parameter.

It is well known that the efficient result by TV (or smoothed-TV) regularization usually comes from image processing where the compact operator is mostly considered to be identity operator, *i.e.,* $\mathcal{T} = \mathcal{I}$. Lagged diffusivity fixed point iteration (LDFP) is the easily implementable algorithm in order to approximate the solution for the problem (3.2), [**27, 86, 87**]. The LDFP algorithm is still widely used, see [**48**] for an EIT application. The convergence of this algorithm has been shown only for the case of $\mathcal{T} = \mathcal{I}$, [**8, 27**]. A tomographic application of total variation regularization with some compact forward operator has been recently studied in [**49**].

As has been displayed in Figure 4.15, SAMSARA works 33 times faster than usual LDFP algorithm. Although we have arrived at this empirical result only for some certain number of measurement with certain dimension of Hessian, reader can also observe by other figures for the efficiency of each algorithm.

Appendix A

FURTHER RECONSTRUCTIONS AND BENCHMARK

In the sections A.1 and A.2, we test our gradient based algorithms with quadratic Tikhonov objective functional

$$F_\alpha(\varphi, f^\delta) = \frac{1}{2}||\mathcal{T}\varphi - f^\delta||_2^2 + \alpha\frac{1}{2}||\varphi - \varphi^0||_2^2,$$

with some given initial guess φ^0.

A.1 LDFP with Quadratic Tikhonov Functional

Now, in (4.10), we define the quadratic Tikhonov functional as our objective

$$F_\alpha(\varphi, f^\delta) := \frac{1}{2}||\mathcal{T}\varphi - f^\delta||_2^2 + \alpha\frac{1}{2}||\varphi - \varphi^0||_2^2,$$

with some given initial guess φ^0. Then LDFP algorithm with Tikhonov penalty term is formulated below

LDFP method with Tikhonov:

1. Compute $g^\nu := \mathcal{T}^*(\mathcal{T}\varphi_\alpha^\nu - f^\delta) + \alpha(\varphi_\alpha^\nu - \varphi^0)$, gradient step;
2. Compute $K_\alpha = \mathcal{T}^*\mathcal{T} + \alpha I$, approximate Hessian;
3. Solve $K_\alpha^{-1}s^{\nu+1} = -g^\nu$, quasi-Newton step;

4. Update $\varphi_\alpha^{\nu+1} = \varphi_\alpha^\nu + s^{\nu+1}$;

We present our numerical results in the Figures A.1 - A.2. Because of the elapsed CPU time, we find it sufficient to iterate the algorithm only 5 times.

A.2 SAMSARA with Quadratic Tikhonov Gradient Step

Here, we present numerical results produced by SAMSARA, [**64**], with quadratic Tikhonov type objective functional. Recall the usual Tikhonov type objective functional,

$$F_\alpha(\varphi, f^\delta) = \frac{1}{2}||\mathcal{T}\varphi - f^\delta||^2_{\mathcal{L}^2(\mathcal{Z})} + \alpha\frac{1}{2}||\varphi - \varphi^{(0)}||^2_{\mathcal{L}^2(\Omega)}. \tag{A.1}$$

Then the gradient step to be implemented is

$$\nabla F_\alpha(\varphi, f^\delta) = \mathcal{T}^*(\mathcal{T}\varphi - f^\delta) + \alpha(\varphi - \varphi^{(0)}). \tag{A.2}$$

We run our tests with different number of measurements $\{1, 50, 100, 240, 360, 450\}$. Numerical convergence for each reconstruction is presented in Figure A.3. Each reconstruction is presented in Figure A.4.
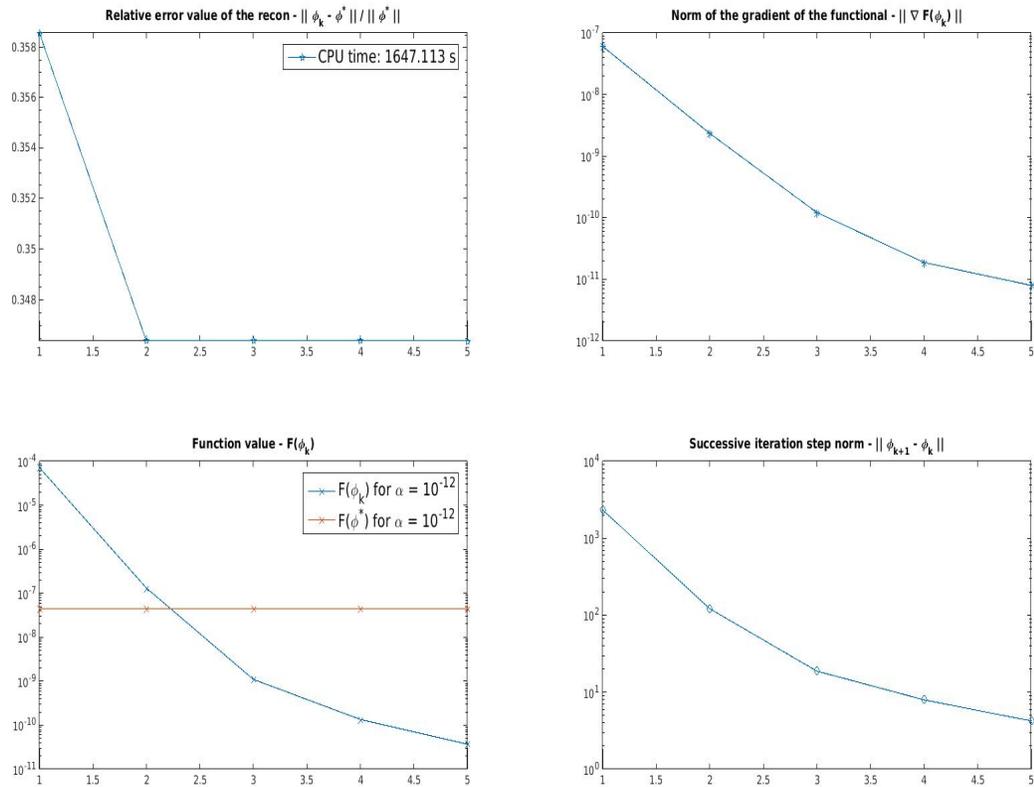
Figure A.1: LDFP numerics with Tikhonov functional. This optimum solution has been obtained with $\alpha = 10^{-12}$ in LDFP algorithm with Tikhonov functional. A comparison of the functional value during the iteration to its truth is also given.
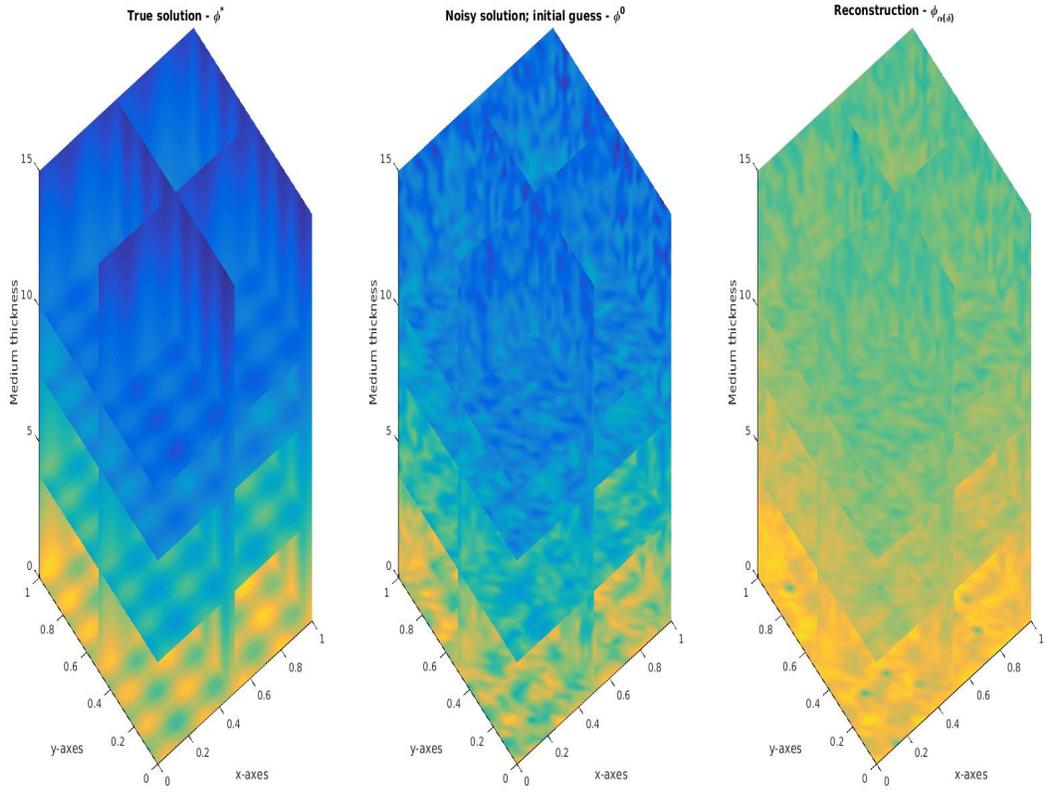
Figure A.2: LDFP with Tikhonov penalty term; Three dimensional visualization of the optimum solution, when $\alpha = 10^{-12}$, presented in Figure A.1.
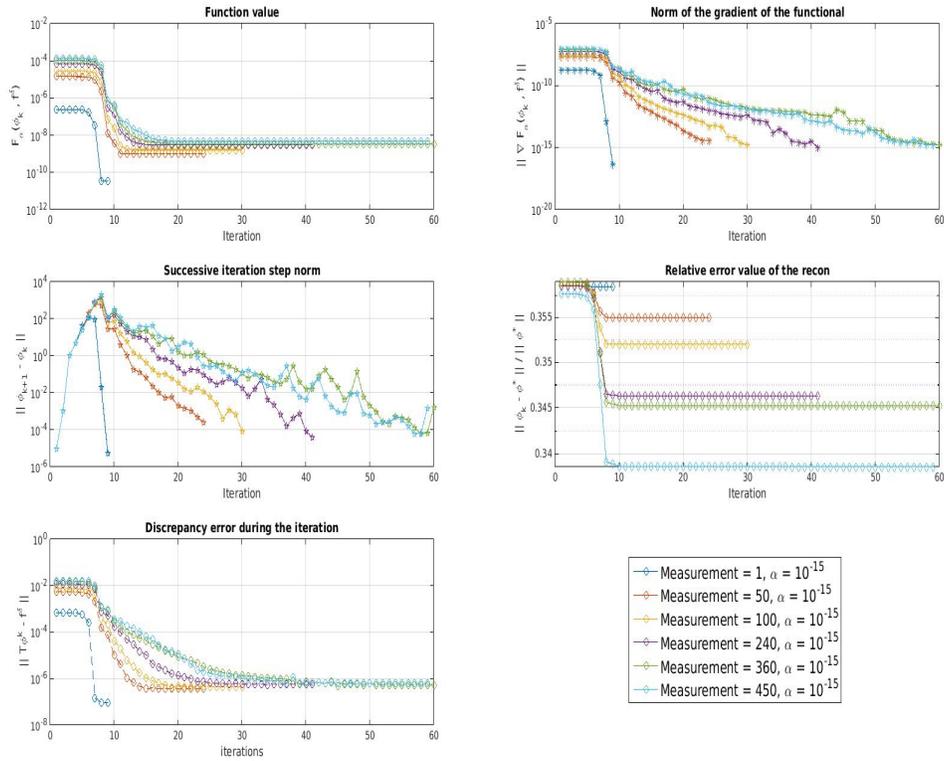
Figure A.3: SAMSARA with Tikhonov gradient step numerical results per measurement. We have conducted our experiment in the software SAMSARA for the measurement number $\{1, 50, 100, 240, 360, 450\}$. Fixed regularization parameter $\alpha = 10^{-15}$ has been determined according to the behaviour in the discrepancy after each iteration step $\nu = 1, 2, \cdots$.
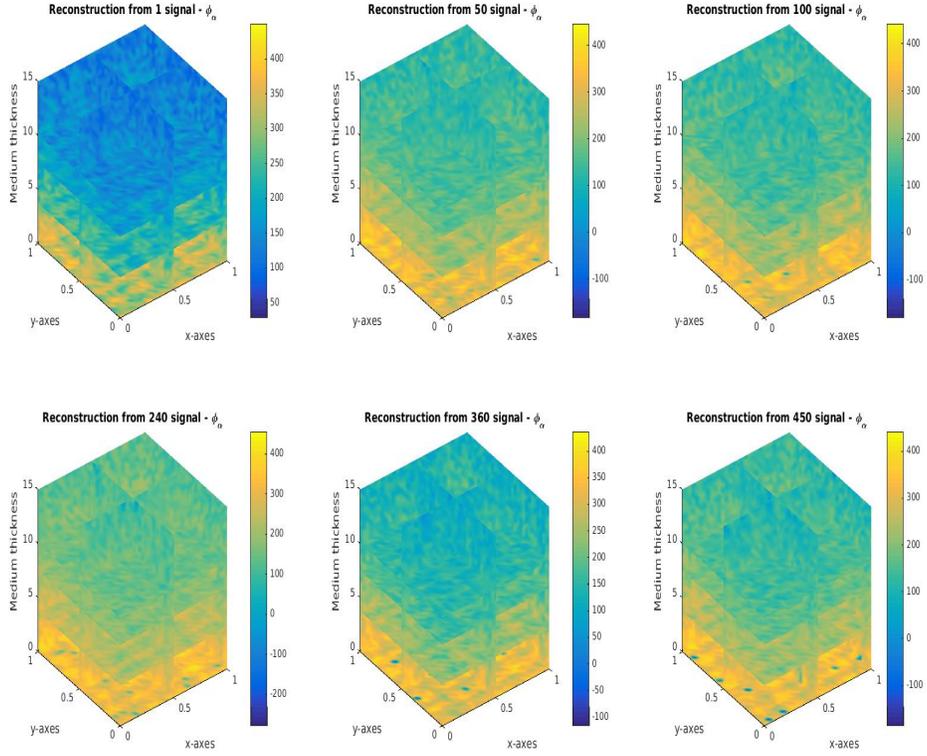
Figure A.4: SAMSARA with Tikhonov gradient step numerical reconstruction results per measurement $\{1, 50, 100, 240, 360, 450\}$ for the fixed regularization parameter $\alpha = 10^{-15}$

71

Bibliography

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

[2] R. A. Adams. *Sobolev Spaces.* Academic Press, INC., 1975.

[3] U. Albocher, P.E. Barbone, M. S. Richards, A.A. Oberai and I. Harari. *Approaches to accommodate noisy data in the direct solution of inverse problems in incompressible plane strain elasticity.* Inverse Probl. Sci. Eng., 22, 8, 1307-1328, 2014.

[4] G. Alessandrini, A. Morassi and E. Rosset. *The linear constraints in Poincaré and Korn type inequalities.* Forum Math., 20, 3, 557 - 569, 2008.

[5] S. Anzengruber and R. Ramlau. *Morozov's discrepancy principle for Tikhonov-type functionals with nonlinear operators.* Inverse Problems, 26, 025001 (17pp), 2010.

[6] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems.* Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, 2000.

[7] S. Anzengruber and R. Ramlau. *Convergence rates for Morozov's discrepancy principle using variational inequalities.* Inverse Problems, 27, 105007 (18pp), 2011.

[8] J.F. Aujol. *Some first-order algorithms for total variation based image restoration.* J. Math. Imaging Vision., 34, 3, 307-327, 2009.

[9] M. Bachmayr and M. Burger. *Iterative total variation schemes for nonlinear inverse problems.* Inverse Problems, 25, 105004 (26pp), 2009.

[10] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces.* Springer New York, 2011.

[11] J. M. Bardsley and A. Luttman. *Total variation-penalized Poisson liklehood estimation for ill-posed problems.* Adv. Comput. Math., 31, 25-59, 2009.

[12] A. Beck and S. Sabach. *Weiszfeld's Method: Old and New Results.* to appear in J. Optim. Theory and Appl., DOI 10.1007/s10957-014-0586-7.

[13] A. Beck and M. Teboulle. *Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems.* IEEE Trans. Image Process., 18, 11, 2419-2434, 2009.

[14] A. Beck and M. Teboulle. *A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems.* SIAM J. Imaging Sciences, 2, 1, 183-202, 2009.

[15] M. Bender, G. Dick, M. Ge, Z. Deng, J. Wickert, H.G. Kahle, A. Raabe and G. Tetzlaff. *Development of a GNSS water vapour tomography system using algebraic reconstruction technique.* Adv. Space. Res., 47, 10, 1704-1720, 2011.

[16] M. Benning, L. Gladden, D. Holland, C.-B. Schönlieb and T. Valkonen. *Phase reconstruction from velocity-encoded MRI measurements - a survey of sparsity-promoting variational approaches.* Journal of Magnetic Resonance, 238, 26 - 43, 2014.

[17] M. Bergounioux and E. Trélat. *A variational method using fractional order Hilbert spaces for tomographic reconstruction of blurred and noised binary images.* J. Funct. Anal., 259, 9, 2296-2332, 2010.

[18] M. Bergounioux. *On Poincaré-Wirtinger inequalities in space of functions of bounded variatoin.* Control Cybernet., 40, 4, 921-29, 2011.

[19] M. Bevis, S. Businger, T.A. Herring, C. Rocken, R.A. Anthes and R.H. Ware. *GPS Meteorology: Remote sensing of atmospheric water vapor using the global positioning system.* Journal of Gephysical Research, 97(D14), 15, 787-801, 1992.

[20] T. Bonesky, K. S. Kazimierski, P. Maass, F. Schöpfer and T. Schuster. *Minimization of Tikhonov Functionals in Banach Spaces.* Abstr. Appl. Anal., Art. ID 192679, 19 pp, 2008 .

[21] K. Bredies. *A forward-backward splitting algorithm for the minimization of nonsmooth convex functionals in Banach space.* Inverse Problems, 25, 1, 015005, 20 pp, 2009.

[22] L.M. Bregman. *The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming.* Ž. Vyčisl. Mat. i Mat. Fiz., 7, 620 - 631, 1967.

[23] M. Burger and S. Osher. *Convergence rates of convex variational regularization.* Inverse Problems, 20, 5, 1411 - 1421, 2004.

[24] A. Chambolle and P.L. Lions. *Image recovery via total variation minimization and related problems.* Numer. Math., 76, 167 - 188, 1997.

[25] T. F. Chan and K. Chen. *An optimization-based multilevel algorithm for total variation image denoising.* Multiscale Model. Simul., 5, 2, 615-645, 2006.

[26] T. F. Chan and P. Mulet. *On the convergence of the lagged diffusivity fixed point method in total variation image restoration.* SIAM J. Numer. Anal., 36, 2, 354-367, 1999.

[27] T. Chan, G. Golub and P. Mulet. *A nonlinear primal-dual method for total variation-baes image restoration.* SIAM J. Sci. Comp, 20, 1964-1977, 1999.

[28] P. G. Ciarlet. *Linear and nonlinear functional analysis with applications.* Society for Industrial and Applied Mathematics, Philadelphia, PA, 2013.

[29] D. Colton and R. Kress. *Inverse Acoustic and Electromagnetic Scattering Theory.* Springer Verlag Series in Applied Mathematics, 93, Third Edition, 2013.

[30] L. Debnath and P.Mikusiński, *Introduction to Hilbert spaces with applications.* 3rd Edition, Elsevier Academic Press, 2005.

[31] M. Defrise, C. Vanhove, and X. Liu. *An algorithm for total variation regularization in high-dimensional linear problems.* Inverse Problems, 27, 065002 (16pp), 2011.

[32] K. Deimling. *Nonlinear Functional Analysis.* Springer-Verlag, Berlin, xiv+450 pp., 1985.

[33] D. Dobson and O. Scherzer. *Analysis of regularized total variation penalty methods for denoising.* Inverse Problems, 12, 5, 601 - 617, 1996.

[34] D. C. Dobson and C. R. Vogel. *Convergence of an iterative method for total variation denoising.* SIAM J. Numer. Anal., 34, 5, 1779 - 1791, 1997.

[35] H. W. Engl, M. Hanke and A. Neubauer. *Regularization of inverse problems.* Math. Appl., 375. Kluwer Academic Publishers Group, Dordrecht, 1996.

[36] I. Ekeland. *On the variational principle.* J. Math. Anal. Appl., 47, 324 - 353, 1974.

[37] L. C. Evans. *Partial differential equations.* Graduate Studies in Mathematics, 19. American Mathematical Society, Providence, RI, 1998.

[38] J. M. Fowkes, N. I. M. Gould and C. L. Farmer. *A branch and bound algorithm for the global optimization of Hessian Lipschitz continuous functions.* J. Glob. Optim., 56, 1792 - 1815, 2013.

[39] L. Frerking, M. Burger, D. Vestweber and C. Brune. *TGV-based flow estimation for 4D leukocyte transmigration.* Proceedings of the Inverse Problems from Theory to Applications Conference (IPTA2014), IOP Publishing, 79 -83, 2014.

[40] M. Grasmair. *Generalized Bregman distances and convergence rates for non-convex regularization methods.* Inverse Problems, 26, 11, 115014, 16pp, 2010.

[41] M. Grasmair. *Variational inequalities and higher order convergence rates for Tikhonov regularisation on Banach spaces.* J. Inverse Ill-Posed Probl., 21, 379-394, 2013.

[42] M. Grasmair, M. Haltmeier and O. Scherzer. *Necessary and sufficient conditions for linear convergence of ℓ^1-regularization.* Comm. Pure Appl. Math., 64, 2, 161-182, 2011.

[43] C. W. Grötsch. *Inverse Problems in the Mathematical Sciences.* Vieweg, 1993.

[44] C. W. Grötsch. *Integral equations of the first kind, inverse problems and regularization: a crash course.* J. Phys.: Conf. Ser., 73, 012001, 2007.

[45] C. Hamaker, K. T. Smith, D. C. Solmon and S. L. Wagner. *The divergent beam x-ray transform,* Rocky Mountain J. Math., 10, 1, 253 - 283, 1980.

[46] M. Hanke. *Conjugate gradient type methods for ill-posed problems.* Pitman Research Notes in Mathematics Series, 327. Longman Scientific & Technical, Harlow, 1995.

[47] P. C. Hansen. *Rank deficient and discrete ill-posed problems: Numerical aspects of linear inversion.* SIAM Monogr. Math. Model. Comput., Philadelphia, 1998.

[48] L. Harhanen, N. Hyvönen, H. Majander and S. Staboulis. *Edge-enhancing re-construction algorithm for three dimensional electrical impedance tomography.* Siam. J. Sci. Comput., 37, 1, B60 - B78, 2015.

[49] Hämäläinen K, Harhanen L, Hauptmann A, Kallonen A, Niemi E and Siltanen S, *Total variation regularization for large-scale X-ray tomography.* International Journal of Tomography and Simulation, 25, 1, 1-25, 2014.

[50] M. Hintermüller, C.N. Rautenberg, J. Hahn. *Functional-analytic and numerical issues in splitting methods for total variation-based image reconstruction.* Inverse Problems, 30, 055014(34pp), 2014.

[51] B. Hofmann and P. Mathé. *Parameter choice in Banach space regularization under variational inequalities.* Inverse Problems 28, 104006 (17pp), 2012.

[52] B. Hofmann and M. Yamamoto. *On the interplay of source conditions and vari-ational inequalities for nonlinear ill-posed problems.* Appl. Anal., 89, 11, 1705 - 1727, 2010.

[53] T. Hohage and F. Weidling. *Verification of a variational source condition for acoustic inverse medium scattering problems.* Inverse Problems, 31, 075006 (14pp), 2015.

[54] D. Hohmann and H. Holzmann. *Weighted angle Radon transform: Convergence rates and efficient estimation.* to appear in Statistica Sinica, January 2016.

[55] V. Isakov. *Inverse problems for partial differential equations.* Second edition. Applied Mathematical Sciences, 127. Springer, New York, 2006.

[56] I. Jovanović, L. Sbaiz and M. Vetterli. *Acoustic tomography for scalar and vector fields: Theory and Application to temperature and wind estimation.* J. Atmos. Oceanic Technol., 26, 8, 1475 - 1492, 2009.

[57] H. Kekkonen, M. Lassas and S. Siltanen. *Analysis of regularized inversion of data corrupted by white Gaussian noise.* Inverse Problems, 30, 045009, 18pp, 2014.

[58] D. Kinderlehrer and G. Stampacchia. *An Introduction to Variational Inequali-ties and Their Applications.* Pure Appl. Math. 88, Academic Press, New York-London, 1980.

[59] A. Kirsch. *An Introduction to the Mathematical Theory of Inverse Problems.* Second edition. Applied Mathematical Sciences, 120. Springer, New York, 2011.

[60] F. Kleijer. *Troposphere Modeling and Filtering for Precise GPS, Leveling Mathematical Geodesy and Positioning.* Dissertation, ISBN: 90-804147-3-5, NUGI: 816, 2004.

[61] R. Kress. *Linear integral equations.* Second edition. Appl. Math. Sci., 82, Springer-Verlag, New York, 1999.

[62] E. Kreyszig. *Introductory Functional Analysis with Applications.* Wiley Classics Library. John Wiley & Sons, Inc., New York, 1989.

[63] D. A. Lorenz. *Convergence rates and source conditions for Tikhonov regularization with sparsity constraints.* J. Inv. Ill-Posed Problems, 16, 463-478, 2008.

[64] D. R. Luke. *SAMSARA, a reverse communication optimization toolbox for MATLAB.* http://num.math.uni-goettingen.de/~r.luke/publications/SAMSARA_MATLAB.tar.gz

[65] D. R. Luke. *Analysis of optical wavefront reconstruction and deconvolution in adaptive optics.* Dissertation, Department of Applied Mathematics, University of Washington, June 2001.

[66] P. Maaß. *Singular value decompositions for Radon transforms.* Mathematical methods in tomography (Oberwolfach, 1990), 6-14, Lecture Notes in Math., 1497, Springer, Berlin, 1991.

[67] P. Miidla, K. Rannat and P. Uba. *Tomographic approach for tropospheric water vapor detection.* Comput. Methods. Appl. Math., 8, 3, 263-278, 2008.

[68] F. Natterer. *The mathematics of computerized tomography.* SIAM, Classics Appl. Math., 32, 2001.

[69] F. Natterer. X-ray Tomography. Inverse Problems and Imaging, Springer, Berlin, 17 - 34, 2008.

[70] F. Natterer and F. Wübbeling. *Mathematical methods in image reconstruction.* SIAM Monogr. Math. Model. Comput., 05, 2001.

[71] J. Nocedal and S. J. Wright. *Numerical Optimization.* Springer Series in Operations Research. Springer-Verlag, New York, xxii+636 pp, 1999.

[72] G. Ólafsson and E. T. Quinto. *The Radon transform, inverse problems and tomography.* Proceedings of Symposia in Applied Mathematics, 63, American Mathematical Society Short Course, January 3-4, 2005, Atlanta, Georgia, 2006.

[73] S. S. Orlov. *Theory of three dimensional reconstruction II: The recovery operator.* Soviet Phys. Crystallogr., 20, 429 - 433, 1976.

[74] J. Radon. *Über die Bestimmung von Funktionnen durch ihre Integralwerte längs gewisser Mannigfaltikeiten.* Ber. Verh. Sachs. Akad. Wiss. Leipzig-Math.-Natur. Kl., 69, 262 - 277, 1917.

[75] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis.* Fundamental Principles of Mathematical Sciences, 317. Springer-Verlag, Berlin, 1998.

[76] H. L. Royden. *Real Analysis.* Macmillan, 1969.

[77] L. I. Rudin, S. J. Osher and E. Fatemi. *Nonlinear total variation based noise removal algorithms.* Physica D, 60, 259-268, 1992.

[78] W. Rudin. *Principles of Mathematical Analysis.* Third edition, International Series in Pure and Applied Mathematics. McGraw-Hill Book Co., New York-Auckland-Düsseldorf, 1976.

[79] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier and F. Lenzen. *Variational Methods in Imaging.* Applied Mathematical Sciences, 167, Springer, New York, 2009.

[80] S. Setzer. *Operator splittings, Bregman methods and frame shrinkage in image processing.* Int. J. Comput. Vis., 92, 265-80, 2011.

[81] D. F. Shanno and K. Phua. *Matrix conditioning and nonlinear optimization.* Math. Prog., 14, 149 - 160, 1978.

[82] A. N. Tikhonov. *On the solution of ill-posed problems and the method of regularization.* Dokl. Akad. Nauk SSSR, 151, 501-504, 1963.

[83] A. N. Tikhonov and V. Y. Arsenin. *Solutions of ill-posed problems.* Transl. from Russian, John Wiley & Sons, New-York - Toronto, 1977.

[84] M. R. Troller. *GPS based Determination of the Integrated and Spatially Distributed Water Vapor in the Troposphere.* Diss. ETH No. 15513, 2004.

[85] C. R. Vogel. *A limited memory BFGS method for an inverse problem in atmospheric imaging.* Lecture Notes in Earth Sciences, Methods and Applications of Inversion, 92, pp 292-304, 2005.

[86] C. R. Vogel. *Computational methods for inverse problems.* Frontiers Appl. Math., 23, 2002.

[87] C. R. Vogel and M. E. Oman. *Iterative methods for total variation denoising.* SIAM J. SCI. COMPUT., 17, 1, 227-238, 1996.

[88] Y. Wang, J. Yang, W. Yin, and Y. Zhang. *A new alternating minimization algorithm for total variation image reconstruction.* SIAM J. Imaging Sciences, 1, 3, 248-272, 2008.

[89] F. Werner and T. Hohage. *Convergence rates in expectation for Tikhonov-type regularization of inverse problems with Poisson data.* Inverse Problems, 28, 10, 104004, 15 pp., 2012.

[90] Z. B. Xu and G. F. Roach. *Characteristic ineqaulities of uniformly convex and uniformly smooth Banach spaces.* J. Math. Anal. Appl., 157, 1, 189-210, 1991 .

[91] Zeidler E. *Applied functional analysis applications to mathematical physics.* Applied Mathematical Sciences, 108. Springer-Verlag, New York, 1995.

[92] Zeidler E. *Nonlinear functional analysis and its applications. II/A. Linear monotone operators.* Springer-Verlag, New York, xviii+467 pp., 1990.

[93] F. Zus, M. Bender, Z. Deng, G. Dick, S. Heise, M. Shang-Guan and J. Wickert. *A methodology to compute GPS slant total delays in a numerical weather model.* Radio Science, 47, RS2018., 2012