Discrete Seismic Tomography

ISBN: 978-94-6375-692-1

Cover: Frontpage is a 3D impression of a level-set function for the salt model SEG/EAGE. Back cover depicts the projection of the level-set function where the contour of the salt is visible. The salt model is approximately 3.5 km deep and 15 km wide. Although the model has a resolution of 500×1000, it requires only 90 radial basis functions to represent its level-set function.

Print: Ridderprint | www.ridderprint.nl.

Discrete Seismic Tomography

Discrete Seismische Tomografie

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

maandag 16 december 2019 des middags te 4.15 uur

door

Ajinkya Anil Kadu

geboren op 20 april 1993 te Bhilai, India

Promotoren:
Prof. dr. K.J. Batenburg
Prof. dr. J.E. Frank
Prof. dr. W.A. Mulder

Copromotor:
Dr. T. van Leeuwen

This work is part of the Industrial Partnership Program 'Computational sciences for energy research' of the Foundation for Fundamental Research on Matter, which is part of the Netherlands Organization for Scientific Research (NWO). This research programme was co-financed by Shell Global Solutions International B.V.

Contents

CHAPTER 1

Introduction

" Science is built up of facts, as a house is built of stones;
but an accumulation of facts is no more a science than a
heap of stones is a house."
- Henri Poincaré

1.1 Tomographic imaging

The word 'tomography,' derived from the ancient Greek words 'tomos' (slice or section) and 'graphein' (write down), is the procedure to image an object of interest from its projections. For example, X-ray computed tomography provides a 3D-structure of an object by measuring the intensities of the X-rays that passed through them. Due to its non-destructive nature, tomography has found significance in many fields of sciences and engineering.

The last century has seen many contributions in the field of tomographic imaging. We can broadly categorize them into physics, mathematics, and applications. In physics, the significant contributions include the discovery of X-rays by Wilhelm Röntgen [155], development of the electromagnetic lens by Hans Busch [42] and the discovery of nuclear magnetic resonance by Isidor Rabi and Felix Bloch [28]. Some of the major early contributions to the mathematics of imaging were made by Johann Radon (radon transform [150]), Stefan Kaczmarz (practical algorithm for ray-based tomography [109]), and Max Born (linearization of scattering [30]). The applications include the X-ray tomographic scanner by Hounsfield [93], the magnetic resonance scanner by Paul Lauterbur [118], the electron microscope by Ernst Ruska and Max Knoll [113], and the cryo-electron microscope by Jacques Dubochet [64].

Tomographic imaging has found applications on all scales; from a resolution of nanometers to light-years (see Figure 1.1). For example, electron tomography is widely used to retrieve a detailed 3D reconstruction of sub-cellular macro-molecules like structurally heterogeneous protein complexes [92]. In the dairy industry, tomography is used to find holes in cheese [168]. In biomedical tomography, researchers are aiming to determine the concentration of drugs in mice to measure the performance of medicines [209, 116]. At the scale of the human body, techniques like X-ray computed tomography, magnetic resonance imaging, and ultrasound tomography are regularly used for medical diagnosis

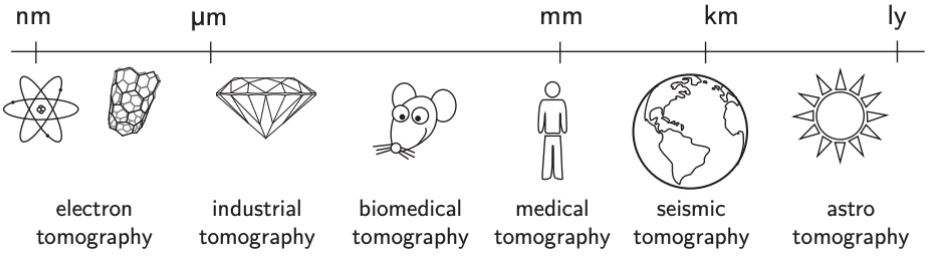


Figure 1.1: Applications of Tomography at various scales

[165, 138, 2, 75]. In the geophysical industry, tomography helps in locating and characterizing hydrocarbon reservoirs [53, 36]. Tomography also has a significant presence in astrophysics, for example, in detecting the downflows beneath sunspots and plages from the travel-time measurements of acoustic waves [65].

We can distinguish tomography broadly into two types: *(i)* scattering-based, and *(ii)* ray-based. In each experiment of scattering-based tomography, a set of sources transmits energy (in the form of waves) into the medium in order to illuminate the object. The energy interacts with the object resulting in a scattered wavefield that is then measured by receivers at specific locations. This interaction is governed by a wave equation, whose coefficients are dependent on the impedance of the object. By changing the positions of sources and receivers, the experiment continues until we obtain enough measurements. Figure 1.2(a) illustrates this setup. The tomographic imaging then aims to find an image x (*i.e.*, a map of the spatially varying impedance of the object) by fitting the data, y, to a model, $\mathcal{F}$, of the experiment. That is, mathematically, we solve

$$\mathcal{F}(x) = y, \tag{1.1}$$

where $\mathcal{F}$ is the forward operator that maps from an image to the data. The operator $\mathcal{F}$ accounts for the (nonlinear) interaction of the waves with the object described by a wave equation. Here, we refer to $\mathcal{F}(x)$ as synthetic data since we obtain it from the forward model that approximates the governing physics. This nonlinear inverse problem, identified by the name of Full-Waveform Inversion, is generally cast using a least-squares approach [77]. That is, it finds an estimate

$$x^\star = \underset{x}{\operatorname{argmin}} \quad \|\mathcal{F}(x) - y\|^2 \tag{1.2}$$

that minimizes the residual (or error) of the synthetic data, $\mathcal{F}(x)$, with the actual data y. This least-squares optimization problem is typically solved using a gradient-based iterative method, whose convergence depends on the initial estimate. The last three decades have seen an increase in this area of research mainly due to improvements in high-performance computing. For an overview, please refer to [204].

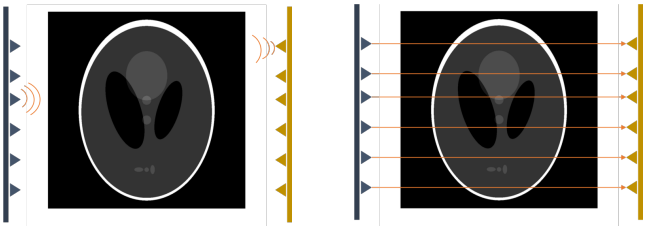
Many efforts have been made in the early 20^{th} century to linearize the interaction between the waves and the object. In particular, Max Born introduced a series expansion for the scattering process [30]. The first-order term in the series gives an approximation of the scattered wavefield that only accounts for single scattering, while the higher-order terms include multiple scattering. Approximating the interaction by single scattering, referred to as the Born approximation, leads to a linear inverse problem [94]. This approximation holds if the object is semi-transparent (low impedance), and the wavelengths are long compared to the size of the object.

Let us now look at the ray-based tomography. A typical setup for this tomography consists of source and detector panels, with a source emitting a beam of high-energy radiation that propagates along straight rays and a detector measuring the intensity of the ray after their interaction with the object (see Figure 1.2(b)). Generally, the sources and detectors are moved around the object (or, altenatively, the object is moved) to get enough measurements for the description of the object. This setup

usually occurs in X-ray computed tomography and electron tomography. With the assumption of *(i)* no diffraction and refraction (rays travel along a straight line), and *(ii)* monochromatic rays (same frequency), the interaction is characterized by Beer's law. It states that each material has a characteristic linear attenuation coefficient for rays emitted at a given frequency. As a result, dividing the measured intensity for a ray by the intensity at the source and taking the negative logarithm yields the integrated attenuation coefficient along the ray. Hence, the measurements are represented using a Radon transform [60], and the reconstruction of the object from these measurements leads to a linear inverse problem

$$x^\star = \underset{x}{\operatorname{argmin}} \quad \|Ax - y\|^2, \tag{1.3}$$

where A represents the radon transform. This problem is solved readily either using filtered back-projection (FBP [128]) or an iterative reconstruction algorithm [109] like the algebraic reconstruction technique (ART [80]). These methods numerically approximate the inverse map from the data to an image and require enough measurements with low noise to find the solution accurately. Please refer to [108] for an overview of ray-based tomography.



(b) Scattering-based (a) straight-ray

Figure 1.2: Tomography classified based on the interaction. On the left, we have sources, and on the right of the object are the receivers (transmission setup). Another popular acquisition scheme is a reflection where the sources and receivers are located on the same side of the object.

1.2 Regularization and optimization

Solving the inverse-problem (1.1) is not trivial. French scientist Hadamard studied these problems and provided conditions for well-posedness [85]. He stated that well-posed problems must satisfy three conditions: *existence*, *uniqueness*, and *stability* of the solution. The *existence* condition prescribes that for every set of measurements, y, an image must exist for which $\mathcal{F}(x) = y$, while *uniqueness* requires that it must be unique. The *stability* requires that a small change in the data should not produce substantial changes in the image. In other words, the inverse map from the data to an image must be continuous. If at least one of these conditions does not hold, then the problem is termed as an *ill-posed* problem.

In tomographic imaging, the *existence* condition fails if the measurement noise is not in the range of the forward operator. A limited number of projections may lead

to multiple images satisfying the data (non-uniqueness of the solution). The stability is related to the conditioning of the operator. Suppose $\mathcal{F}$ is a linear operator with singular values decreasing exponentially (with maximum singular value 1), then the inverse-operator will have high singular values ($\gg 1$). Hence, a small change in the data y will lead to a significant difference in the image x. We enlist the main reasons for the ill-posedness of tomography problems in the following:

- *Existence*
  - **Noise**: an improper model of the physics introduces non-random, coherent noise. Hence, the noisy data may not be in the range of the forward model.
  - **Nonlinearity**: the relationship between the waves and the object may be nonlinear.
- *Uniqueness*
  - **Partial information**:
    - In scattering-based tomography, scattered waves are measured at only a few locations. In some cases, these measurements are restricted to only one side (reflection measurements) due to limited access around the object.
    - In ray-based tomography, the projections are limited due to either the time constraints or the destructive nature of the rays. Some tomographic setup has a limited angular range for movement, hence making it difficult to measure the projections for certain angles.
  - **Sharp contrasts**: large discontinuities in the image give rise to complex interaction resulting in many local minima.

The condition of existence is relaxed by casting the inverse problem into a least-squares data-fitting procedure (see (1.2)). To address the ill-posedness (mainly the uniqueness and stability), Russian mathematician A.N. Tikhonov introduced regularization in 1963 [191]. It is one of the first instances where a stable solution to an ill-posed problem has been obtained. *Regularization* is the procedure of adding information to solve an ill-posed problem. It introduces prior information about the structure of an image into the inverse procedure. The regularized problem generates an estimate

$$x_{\mathcal{R}} = \underset{x}{\operatorname{argmin}} \Big\{ \|\mathcal{F}(x) - y\|^2 + \alpha \mathcal{R}(x) \Big\}, \tag{1.4}$$

where the residual between the synthetic and observed data, $\mathcal{F}(x) - y$, is measured in some norm, $\mathcal{R}$ is a regularization function that incorporates the prior knowledge about the image x, and α is a parameter that balances the misfit and the regularization. We classify regularization methods according to the way they enforce prior information, *i.e.*, implicit or explicit. Typically, explicit regularizations enforce strong prior information compared to implicit ones. This thesis mostly deals with explicit regularization methods.

Implicit regularization: It deals with constructing a function that incorporates the structure or characteristics of an image. Popular implicit regularization includes Tikhonov ($\mathcal{R}(x) = \|x\|^2$), ℓ_1 ($\mathcal{R}(x) = \|x\|_1$), and total variation ($\mathcal{R}(x) = \|\nabla x\|_1$). Standard Tikhonov regularization finds a unique image that fits the data

approximately in the case where either multiple solutions exist or where the inverse operator is unstable [191]. Its generalized variant finds a smooth image that approximates the data. In reality, the images may not be smooth. Around the 1990s, the idea of sparsity emerged, which states that the image can be represented using only a subset of a large class of functions. This sparsity on the image or its transformed domain is promoted through an ℓ_1-regularization, which is a convex approximation to an ℓ_0 norm [154]. In recent years, total variation (TV) regularization, introduced in [156], became very popular due to its ability to find a piecewise-constant image (for example, see [71] for application of TV in tomographic imaging).

Explicit regularization: These techniques represent images obeying the prior information explicitly and hence restrict the space of appropriate images. For example, if we are looking for an image that takes only a few *known* discrete values, we can separate the spatial region of the image into several components (whose grey levels are known) and try to find the shape of these components. Enforcing this discrete prior in the reconstruction is the subject of discrete tomography. A particular case of discrete images is a binary image. A binary image x constitutes only two distinct (homogeneous) regions (with values u_0 and u_1)

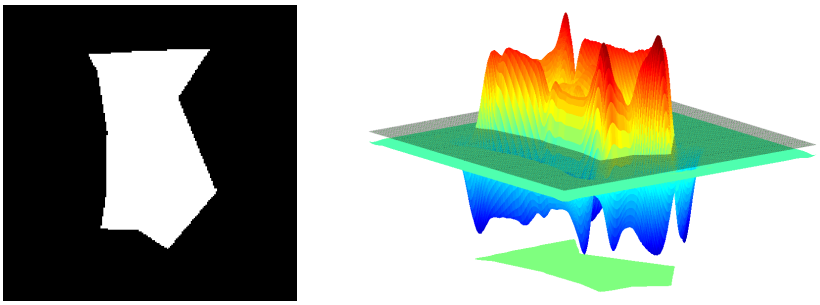
$$x(\mathbf{r}) = \begin{cases} u_1 & \mathbf{r} \in \Omega \\ u_0 & \mathbf{r} \notin \Omega \end{cases},$$

where Ω denotes the spatial region corresponding to the u_1. Without loss of generality, we can scale the image to $\{0, 1\}$, and search for the solution in this space. Since the search space is discrete, the reconstruction methods rely either on sampling the discrete space [126, 74] or making use of heuristics [19]. An overview of the contributions can be found in [89, 90]. In practice, a method known as Discrete Algebraic Reconstruction Technique (DART) is commonly used because of its ability to compute accurate reconstructions in almost real-time [18]. It exploits the discrete nature of the object by alternating iteratively between the segmentation and the reconstruction on the boundary of the segmented image. It has shown success in the various applications of computerized tomography [194] and electron tomography [196, 17]. DART suffers from the noise in the projections and requires many parameters to be specified by the user. To combat noise issue, SDART uses a penalty matrix that spreads the noise across the whole image domain [26]. Another method named TVR-DART produces more accurate reconstructions than DART under a noisy limited data scenario and requires significantly less parameters [215]. It uses the total variation framework to penalize the discrete reconstructions. We note that these heuristic algorithms are limited to the linear tomographic problems, and their extension to the nonlinear case is not trivial.

Instead of heuristics, we can write a continuous formulation using the level-set of a characteristic function ϕ,

$$x(\mathbf{r}) = \Big(1 - h\big(\phi\left(\mathbf{r}\right)\big)\Big)u_0 + h\big(\phi\left(\mathbf{r}\right)\big)u_1,$$

where $\mathbf{r}$ denotes the spatial coordinate, and h is a smooth approximation of the Heaviside function. With such explicit representation of the image x, we can solve



(a) True image (b) Level-set function

Figure 1.3: Level-set illustration for image (a). The image consists of only two grey levels; hence it is a binary image. The corresponding level-set function is given in (b). The gray-color plane denotes the zero-level-set plane. The level-set function represents an object shown at the bottom of the sub-figure (b).

for the level-set of function ϕ, and later retrieve the corresponding image. Figure 1.3 gives an example of the level-set reconstruction of a binary image. The level-set method, proposed in [142] to capture the fronts in fluid propagating with curvature–dependent speed, has been extensively used to solve inverse problems involving binary images [40]. This method suffers from many issues, including the reinitialization, and the use of a signed distance function. These issues can be addressed by parametrizing the level-set using radial basis functions and thereby reducing the dimensionality of the problem [1].

Another class of images, called partially discrete images, are images containing inclusions (with known grey levels) surrounded by a heterogeneous medium. There is a limited amount of research focused on the reconstruction of these images from their tomographic projections. PDART, an extension of DART to partially discrete images, segments the high-contrast inclusions embedded in heterogeneous composition from their linear tomographic projections [153].

Unlike direct inversion processes such as ART and FBP, regularization techniques require that the problem be solved by optimization methods. These methods differ significantly depending on the properties of the data-misfit function and the regularization function. In this thesis, we consider solving the problems of form (1.4), where the misfit function is always differentiable with its convexity dependent on the linearity of the forward operator $\mathcal{F}$. The implicit regularization functions are generally convex, but can be differentiable (*e.g.*, Tikhonov) or not (*e.g.*, ℓ_1 and TV). The level-set method, an explicit regularizer, is non-convex. We make this function differentiable by approximating the Heaviside function by a smooth function. Table 1.1 summarizes the optimization methods for various classes of regularized inverse problems.

1.3 Contributions and outline

The main aim of this thesis is to develop algorithms to incorporate discrete or partially discrete priors for solving ill-posed tomographic problems. The existing

		data-misfit function	
		convex	non-convex
Regularization	convex, smooth	Krylov-based methods [137]	gradient-based methods [34]
	convex, non-smooth	Primal-Dual [49], ADMM [33]	Proximal Newton-type [120], ADMM [33]
	non-convex	gradient-based methods [34]	

Table 1.1: Optimization methods for various classes of regularized inverse problems

(successful) algorithms are mostly based on heuristics and limited to the linear setting. We borrow ideas from convex optimization and level-set theory to develop a different class of methods to retrieve the discrete (or partially discrete) objects in a linear as well as nonlinear setup. Our algorithms compare favorably to the existing methods in the linear setup and outperform them in the nonlinear cases on numerical phantoms.

To make the outline simpler, we distribute chapters 3 to 7 to fall in one of the categories of tomography, as shown in Table 1.2. Chapter 2 provides an overview of the forward operators, regularization, and optimization. Chapter 8 concludes the thesis and provide future directions. We summarize our contributions in the following points:

- **Discrete tomography:** We consider a discrete tomography problem with two grey levels, known as binary tomography. We derive a novel convex formulation of this problem, which turns out to be an ℓ_1-regularized least-squares problem (LASSO). Based on small-scale experiments, we conjecture that if the binary tomography problem admits a unique solution, it can be recovered using the proposed formulation. In the case of multiple solutions, the convex program gives the intersection of the solutions. The proposed algorithm compares favorably to total variation and DART on numerical phantoms. Chapter 3 provides more details about the method.
- **Partially discrete tomography**: In a reconstruction of partially discrete images from its tomographic projections, we take a level-set approach. We separate the homogeneous inclusion from the heterogeneous materials and represent its shape using the level-set of the composition of radial basis functions. Since the tomographic system is linear, the resulting problem is linear in terms of the heterogeneous material and non-linear in terms of the level-set function characterizing the homogeneous region. We propose a bi-level optimization framework to solve this problem. The proposed method outperforms total variation, DART, and PDART. We discuss the details in Chapter 4.
- **Discrete waveform inversion**: Discrete waveform inversion is a particular instance of the full-waveform inversion problem where the image is discrete.

We considered a geophysical imaging problem of retrieving a salt-structure (high impedance object) in the subsurface. For simplicity, we work with an acoustic model and assume that the salt is homogeneous. We further assume that the impedance of the salt and the earth layers (background model) are known a priori. We use a level-set method to obtain the geometry of the salt by parametrizing with radial basis functions (RBF). By fixing the centers and the widths of RBFs, we try to find their coefficients by fitting the synthetic waveforms to the measured seismic data. We numerically show that we can reconstruct the salt quite accurately from the reflection measurements at the surface of the earth. Further discussions are available in Chapter 5.

- **Partially discrete waveform inversion**: We consider the reconstruction of a salt body embedded in a heterogeneous medium. This problem differs from the discrete waveform inversion problem because we do not assume any prior information about the heterogeneous medium. We use a parametric level-set approach along with an alternating strategy to solve the resulting problem. This is discussed in Chapter 6.
- **High-contrast imaging**: We consider the estimation of high-contrast materials from reflection measurements. The problem resembles the partially discrete waveform inversion, but it differs in the sense that the latter does not assume prior information about the grey levels of the high-contrast materials. We develop a total variation regularized inversion approach, which makes use of the multiple frequencies in the data. The proposed approach outperforms the classical methodologies on the numerical phantoms. We discuss this method in Chapter 7.

		Linear	**Non-linear**
	Discrete	**Chapter 3**: Binary Tomography	**Chapter 5**: Discrete Waveform Inversion
Partially	**Discrete**	**Chapter 4**: Partially discrete tomography	**Chapter 6**: Partially discrete waveform inversion **Chapter 7**: High-contrast reflection tomography

Table 1.2: Outline of the thesis.

CHAPTER 2

Preliminaries

" The essence of mathematics lies in its freedom."
- Georg Cantor

This chapter provides a theoretical background on the problem

$$\min_x \quad \|\mathcal{F}(x) - y\|^2 + \alpha\mathcal{R}(x), \tag{2.1}$$

where $\mathcal{F}$ is a forward operator, y is the data, x is the impedance (*i.e.*, an image), $\mathcal{R}$ is a regularization function, and α is a regularization parameter. In Section 2.1, we discuss various forward operators for tomography. We give a brief overview of regularization methods in Section 2.2. We then provide background on shape-optimization to regularize a discrete prior in Section 2.3. Finally, we discuss how to optimize the problem in Section 2.4.

2.1 Forward operators

We first discuss the forward model for scattering-based tomography. The scattering theory describes the relationship between the scattered wavefield and the impedance of the medium. First, we consider the exact relation, governed by the wave equation, to describe the interaction. Following this, we discuss the Born approximation that makes the forward operator linear. Finally, we discuss the model for straight-ray tomography.

2.1.1 Scattering

Consider an object in a domain $\Omega \subset \mathbb{R}^d$ located in a homogeneous surrounding, as shown in Figure 2.1. Let $\mathbf{r}$ represent the spatial coordinate. A scalar wave equation

$$\begin{aligned} \nabla^2 u(\mathbf{r},t) - \frac{1}{c^2(\mathbf{r})}\frac{\partial^2 u}{\partial t^2}(\mathbf{r},t) - \eta(\mathbf{r})\frac{\partial u}{\partial t}(\mathbf{r},t) + q(\mathbf{r},t) = 0, \\ \text{with} \qquad u(\mathbf{r},0) = 0 \quad \text{and} \quad \frac{\partial u}{\partial t}(\mathbf{r},0) = 0 \quad \forall \mathbf{r} \in \Omega \end{aligned} \tag{2.2}$$

provides a mathematical framework to relate a wavefield u that represents a physical quantity like acoustic pressure or the electric field strength, to the coefficients c and η characterizing the medium in which the wave propagates and the source function q. We restrict ourself to scalar waves, but it is possible to extend this framework to vector wavefields [31, 55]. We assume that the source function q is supported in the domain Γ_s, which we call the source domain. A medium is *homogeneous* if the impedance parameters, c, and η, do not depend on the position; otherwise, a medium is *heterogeneous*. For a lossless media, $\eta = 0$.

Closely related to the wave equation is its frequency-domain counterpart known as the Helmholtz equation. By applying a Fourier transform to the above wave equation, we arrive at the Helmholtz equation

$$\nabla^2 \hat{u}(\mathbf{r},\omega) - \gamma^2(\mathbf{r},\omega)\hat{u}(\mathbf{r},\omega) + \hat{q}(\mathbf{r},\omega) = 0, \tag{2.3}$$

where $\gamma = \sqrt{i\omega(i\omega + c^2\eta)}/c$ is a propagation coefficient with real part greater than 0, and ω is the angular frequency. For lossless media ($\eta = 0$), we have $\gamma = i\omega/c$. To

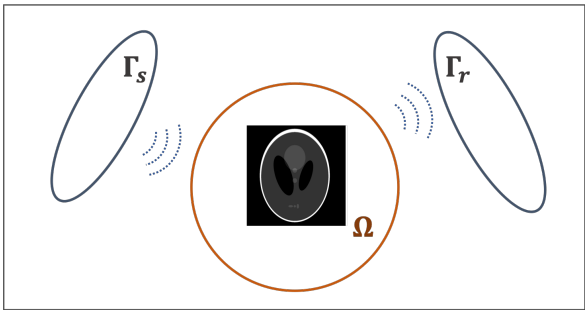


Figure 2.1: Scattering tomographic imaging setup. Γ_s denotes the source domain, Γ_r is the receiver domain, and Ω is an object domain.

solve this uniquely, we need to supplement this equation with boundary conditions. The Sommerfeld radiation condition is one such condition which states that the energy must scatter to infinity for a bounded source, *i.e.*,

$$\lim_{|\mathbf{r}|\to\infty} |\mathbf{r}|^{(d-1)/2} \left(\frac{\partial}{\partial |\mathbf{r}|} - i\omega \right) \hat{u}(\mathbf{r}) = 0.$$

Integral equations

Let us first consider the case of a homogeneous medium where the impedance is spatially invariant. We can apply a spatial Fourier transform to (2.3) to arrive at the simplified Helmholtz equation

$$(k^2 + \gamma^2)\tilde{u}(\mathbf{k}, \omega) = \tilde{q}(\mathbf{k}, \omega),$$

where $k = \|\mathbf{k}\|$ with $\|\cdot\|$ denoting the Euclidean norm. Introducing the spectral domain Green's function $\tilde{G} = 1/(k^2 + \gamma^2)$, the solution in the spatial Fourier domain is given by $\tilde{u} = \tilde{G}\tilde{q}$. Denote the spatial counterpart of $\tilde{G}$ by $\widehat{G}(\mathbf{r}, \omega)$. From the convolution theorem, the frequency-domain wavefield turns out to be

$$\hat{u}(\mathbf{r}, \omega) = \int_{\mathbf{y}\in\Gamma_s} \widehat{G}(\mathbf{r} - \mathbf{y}, \omega)\hat{q}(\mathbf{y}, \omega) \mathrm{d}\mathbf{y},$$

where we have used the fact that $\hat{q}$ vanishes outside the source domain Γ_s. For a given external source $\hat{q}$, the above expression provides the wavefield $\hat{u}$ in the frequency-domain as long as we know the Green's function $\widehat{G}$. The Green's functions are

$$\widehat{G}(\mathbf{r}, \omega) = \begin{cases} \frac{1}{2\gamma} \exp(-\gamma\|\mathbf{r}\|) & d = 1 \\ -\frac{i}{4} H_0^{(2)} \left(\gamma\|\mathbf{r}\| \exp\left(-i\pi/2\right)\right) & d = 2 \\ \frac{1}{4\pi\|\mathbf{r}\|} \exp(-\gamma\|\mathbf{r}\|) & d = 3 \end{cases},$$

where $H_0^{(2)}$ is the zero$^{\text{th}}$ order Hankel function of the second kind. Although we mainly work in the frequency domain, it is useful to write down the time-domain

wavefield $u(\mathbf{r}, t)$:

$$u(\mathbf{r}, t) = \int_{\mathbf{y} \in \Gamma_s} \int_{\tau=0}^{t} G(\mathbf{r} - \mathbf{y}, t - \tau) q(\mathbf{y}, \tau) \mathrm{d}\tau \, \mathrm{d}\mathbf{y}.$$

Now, let's turn our attention to the heterogeneous case. The total wavefield $\hat{u}$ is the superposition of an incident wavefield $\hat{u}_i$ and a scattered wavefield $\hat{u}_s$

$$\hat{u}(\mathbf{r}, \omega) = \hat{u}_i(\mathbf{r}, \omega) + \hat{u}_s(\mathbf{r}, \omega) \tag{2.4}$$

where $\hat{u}_i$ is defined as the field in absence of the object, while $\hat{u}_s$ accounts for the object. The incident wavefield satisfies the Helmholtz equation (2.3) with the background medium represented by $\gamma = \gamma_b$. We can separately write the equation for the domain Ω and outside of this domain:

$$\begin{aligned} \nabla^2 \hat{u}_i(\mathbf{r}, \omega) - \gamma_b^2(\mathbf{r}, \omega) \hat{u}_i(\mathbf{r}, \omega) &= 0, && \forall \, \mathbf{r} \in \Omega \\ \nabla^2 \hat{u}_i(\mathbf{r}, \omega) - \gamma_b^2(\mathbf{r}, \omega) \hat{u}_i(\mathbf{r}, \omega) &= -\hat{q}(\mathbf{r}, \omega), && \forall \, \mathbf{r} \notin \Omega \end{aligned}$$

Similarly, we can express the Helmholtz equation for the total wavefield. Combining the Helmholtz equations for input and total wavefield, we get the Helmholtz equation for the scattered wavefield as follows:

$$\begin{aligned} \nabla^2 \hat{u}_s(\mathbf{r}, \omega) - \gamma_b^2(\mathbf{r}, \omega) \hat{u}_s(\mathbf{r}, \omega) &= -(\gamma_b - \gamma^2) \hat{u}(\mathbf{r}, \omega), && \forall \, \mathbf{r} \in \Omega \\ \nabla^2 \hat{u}_s(\mathbf{r}, \omega) - \gamma_b^2(\mathbf{r}, \omega) \hat{u}_s(\mathbf{r}, \omega) &= 0, && \forall \, \mathbf{r} \notin \Omega. \end{aligned}$$

By introducing the scattering source q_s and the contrast function ξ as

$$q_s(\mathbf{r}, \omega) = \gamma_b^2 \xi(\mathbf{r}) \hat{u}(\mathbf{r}, \omega), \qquad \text{where} \qquad \xi(\mathbf{r}) = \begin{cases} 1 - \left(\frac{\gamma(\mathbf{r})}{\gamma_b} \right)^2 & \mathbf{r} \in \Omega \\ 0 & \mathbf{r} \notin \Omega \end{cases},$$

we get a single equation for the scattered wavefield:

$$\begin{aligned} & \nabla^2 \hat{u}_s(\mathbf{r}, \omega) - \gamma_b^2(\mathbf{r}, \omega) \hat{u}_s(\mathbf{r}, \omega) = \hat{q}_s(\mathbf{r}, \omega). && \forall \, \mathbf{r} \in \mathbb{R}^d \\ \text{(integral)} \qquad & \hat{u}_s(\mathbf{r}, \omega) = \underbrace{\gamma_b^2 \int_{\mathbf{r}' \in \Omega} \widehat{G}(\mathbf{r} - \mathbf{r}', \omega) \xi(\mathbf{r}) \hat{u}(\mathbf{r}', \omega) \, \mathrm{d}\mathbf{r}'}_{\widehat{G}\{q_s\}} && \forall \, \mathbf{r} \in \mathbb{R}^d \end{aligned}$$

where Γ_r is the receiver domain. $\widehat{G}\{q_s\}$ is the compact notation to denote the convolution of Green's function $\widehat{G}$ with contrast source q_s over region Ω. By substituting the above expression of scattered wavefield in the equation (2.4), we arrive at the *Lippmann-Schwinger* equation

$$\hat{u}(\mathbf{r}, \omega) - \underbrace{\gamma_b^2 \int_{\mathbf{r}' \in \Omega} \widehat{G}(\mathbf{r} - \mathbf{r}', \omega) \xi(\mathbf{r}) \hat{u}(\mathbf{r}', \omega) \, \mathrm{d}\mathbf{r}'}_{\widehat{G}\{\xi \hat{u}\}} = \hat{u}_i(\mathbf{r}, \omega), \tag{2.5}$$

where the incident wavefield is computed from the source function using

$$\hat{u}_i(\mathbf{r},\omega) = \int_{\mathbf{r}'\in\Gamma_s} \widehat{G}(\mathbf{r}-\mathbf{r}',\omega)\hat{q}(\mathbf{r}',\omega)\,\mathrm{d}\mathbf{r}'.$$

The Lippmann-Schwinger equation is also known as the *state* equation since it provides the total wavefield (states) in terms of the contrast of the object ξ. The receiver domain Γ_r measures the scattered wavefield $\hat{u}_s$ in the receiver domain and is being retrieved using

$$\hat{u}_s(\mathbf{r},\omega) = \gamma_b^2 \int_{\mathbf{r}'\in\Omega} \widehat{G}(\mathbf{r}-\mathbf{r}',\omega)\xi(\mathbf{r})\hat{u}(\mathbf{r}',\omega)\,\mathrm{d}\mathbf{r}' \qquad \forall\,\mathbf{r}\in\Gamma_s. \tag{2.6}$$

The three forward operators based on the scattering theory are summarized in Table 2.1. The operator P is the restriction operator that restricts the measurements to the receiver domain.

Type	x	$\mathcal{F}(x)$
Wave-equation	c,η	$\nabla^2 u(\mathbf{r},t) - \frac{1}{c^2(\mathbf{r})}\frac{\partial^2 u}{\partial t^2}(\mathbf{r},t) - \eta(\mathbf{r})\frac{\partial u}{\partial t}(\mathbf{r},t) = -q(\mathbf{r},t)$ $\mathcal{F}(x) = Pu$
Helmholtz	γ	$\nabla^2\hat{u}(\mathbf{r},\omega) - \gamma^2(\mathbf{r},\omega)\hat{u}(\mathbf{r},\omega) + \hat{q}(\mathbf{r},\omega) = 0$ $\mathcal{F}(x) = Pu$
Integral	ξ	$\hat{u} - \widehat{G}\{\xi\hat{u}\} = \hat{u}_i$ $\mathcal{F}(x) = \widehat{G}\{\xi\hat{u}\}$

Table 2.1: Scattering-based forward models

Born approximation

To linearize the interaction of waves with an object, Max Born introduced a series expansion for the Lippmann-Schwinger equation based on the Neumann series. Suppose $\mathcal{A}$ is a bounded linear operator, then the inverse of the operator $(\mathcal{I}-\mathcal{A})$ where $\mathcal{I}$ is an Identity operator, is given by a Neumann series,

$$(\mathcal{I}-\mathcal{A})^{-1} = \sum_{k=0}^{\infty}\mathcal{A}^k = \mathcal{I} + \mathcal{A} + \mathcal{A}^2 + \dots.$$

This series converges if the operator norm is bounded by 1 (*i.e.*, $\|\mathcal{A}\| < 1$). Let us now express the total-wavefield in the Lippmann-Schwinger equation as follows,

$$\left(\mathcal{I} - \hat{G}\{\xi\}\right)\hat{u} = \hat{u}_i \qquad \Longrightarrow \qquad \hat{u} = \left(\mathcal{I} - \hat{G}\{\xi\}\right)^{-1}\hat{u}_i.$$

By using the Neumann series to approximate the inverse, we arrive at the following Born series:

$$\hat{u} = \hat{u}_i + \hat{G}\{\xi\hat{u}_i\} + \hat{G}\{\xi\hat{u}_i\}\hat{G}\{\xi\hat{u}_i\} + \dots$$

The convergence of this series depends on the operator $\hat{G}\{\xi\}$ (the operator norm must be below 1). This condition generally holds for low-frequency waves traveling through an object with maximum contrast below 1.

The Born series carries the physics of multiple scattering:

$$\hat{u} = \underbrace{\hat{u}_i}_{\text{(incident)}} + \underbrace{\hat{G}\{\xi\hat{u}_i\}}_{\text{(single scattering)}} + \underbrace{\hat{G}\{\xi\hat{u}_i\}\hat{G}\{\xi\hat{u}_i\}}_{\text{(double scattering)}} + \dots$$

If we ignore the second and the higher order terms in the series, we get an approximate solution for the total wavefield. This is called the Born approximation:

$$\hat{u} \approx \hat{u}_i + \hat{G}\{\xi\hat{u}_i\} \tag{2.7}$$

Hence, it approximates the scattered wavefield $\hat{u}_s$ solely using the input wavefield ($\hat{u}_s = \hat{G}\{\xi\hat{u}_i\}$). This approximation results in the following inverse problem (posed as a least-squares optimization problem):

$$\min_{\xi} \quad \|\hat{G}\{\hat{u}_i\}\,\xi - y\|^2.$$

This problem is known as least-squares reverse-time migration in geophysical imaging [20]. It assumes that the interaction of the wave with an object can be explained with single scattering. Hence, this approach works well in the absence of multiple scattering (or if the multiple scattering is very weak).

2.1.2 Straight-ray

We introduce straight-ray tomography for the common case of X-ray imaging. X-rays consist of a flux of high-energy electromagnetic radiation, and they are characterized by their high frequency (de Broglie's equation). A geometric wave propagation approximation holds for X-rays, and we can model it as a vector field $\mathbf{I} = I(\mathbf{r})$ for $\mathbf{r} \in \mathbb{R}^d$ with $d = 2, 3$. The quantity $\|I(\mathbf{r})\|$ indicates the intensity of the flux at location $\mathbf{r}$, and similarly, $I(\mathbf{r})/\|I(\mathbf{r})\|$ indicates the direction of flux. Let us now describe the forward model, a description of the interaction between X-rays and matter. First, we make the following assumptions:

- **No diffraction nor refraction**: The X-rays travel along straight lines that do not bend when interacting with an object.
- **Monochromatic rays**: The X-rays all have the same frequency.
- **Beer's law**: Each material has a characteristic linear attenuation coefficient μ for rays emitted at a given frequency. That is,

  $$\frac{dI}{ds} = -\mu\big(x(s)\big)I(s),$$

  where s is the arc length along the trajectory described by the ray.

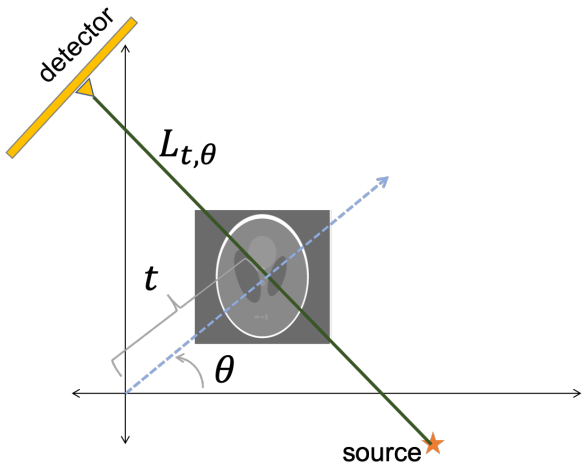


Figure 2.2: Straight-ray tomography setup.

- **Non-diverging source**: The intensity of X-rays do not decay due to beam spreading. Hence, the attenuation is explained solely by interaction with matter.

Under these assumptions, we can write an expression for Beer's law

$$I_{\mathbf{r}_0}(s) = I_{\mathbf{r}_0}(0) \exp\Big(- \int_0^s \mu\left(\mathbf{r}_0 + \tau \mathbf{v}\right) \, \mathrm{d}\tau \Big),$$

where we have parametrized a ray with a source at $\mathbf{r}_0$ as $\mathbf{r}(s, \mathbf{r}_0) = \mathbf{r}_0 + s\mathbf{v}$ with $s > 0$ and $\mathbf{v}$ ($\|\mathbf{v}\| = 1$) is the direction of propagation. By measuring the intensity at two points $\mathbf{r}_{\text{in}} = \mathbf{r}_0 + s_{\text{in}}\mathbf{v}$ and $\mathbf{r}_{\text{out}} = \mathbf{r}_0 + s_{\text{out}}\mathbf{v}$, we get

$$\log\left(\frac{I_{\mathbf{r}_0}\left(\mathbf{r}_{\text{out}}\right)}{I_{\mathbf{r}_0}\left(\mathbf{r}_{\text{in}}\right)}\right) = - \int_{s_{\text{in}}}^{s_{\text{out}}} \mu\left(\mathbf{r}_0 + \tau \mathbf{v}\right) \, \mathrm{d}\tau.$$

The change in the energy along a ray encodes information about the coefficients of the absorption of the object. We can rewrite the above expression as

$$\underbrace{\log\left(\frac{I_{\text{out}}}{I_0}\right)}_{y} = \underbrace{- \int_L \mu(\mathbf{r}) \, dS(\mathbf{r})}_{\mathcal{F}(\mu)},$$

where L parametrizes the line. Next, we study the projection (*i.e.*, the right-hand side in the above equation) in the following subsection. Here, we restrict ourself to 2D ($d = 2$) for simplicity.

Radon transform

For a square intergrable function $\mu \in L^2\left(\mathbb{R}^2\right)$, its Radon transform (also known as projection along a line) is

$$\mathcal{R}\mu\left(t,\theta\right) = \int_{L_{t,\theta}} \mu(\mathbf{r})\,\mathrm{d}S(\mathbf{r}) = \int\int \mu(\mathbf{r})\delta\left(\langle \mathbf{r}, \mathbf{n}_\theta\rangle - t\right)\,\mathrm{d}\mathbf{r}, \tag{2.8}$$

where $\mathbf{n}_\theta = (\cos\theta, \sin\theta)$ is the unit vector with angle θ, and δ is a Dirac-delta function. $\mathcal{R}\mu\left(t,\theta\right)$ is the integral of μ along the line $L_{t,\theta}$ whose direction is perpendicular to $\mathbf{n}_\theta$, and whose distance from the origin is t:

$$L_{t,\theta} = \left\{\mathbf{r} \,:\, \langle \mathbf{r}, \mathbf{n}_\theta\rangle = t\right\}.$$

Refer to Figure 2.2 for straight-ray tomography setup. We take $\theta \in [0,\pi)$ since the integral in (2.8) gives $\mathcal{R}\mu\left(t,\theta+k\pi\right) = \mathcal{R}\mu\left((-1)^k t,\theta\right)$. The Radon transform is well-defined for smooth functions that decay rapidly at infinity (Hence, it works for compactly supported μ). We discuss inverting the Radon transform in the next two subsections.

Backprojection

The value $\mu(\mathbf{r})$ contributes to the projection $\mathcal{R}\mu\left(t,\theta\right)$ for $\mathbf{r} \in L_{t,\theta}$. In absence of any additional information, we will assume $\mu(\mathbf{r})$ contributes equally in all cases of t and θ. Therefore, one way to estimate the value of $\mu(\mathbf{r})$ is by averaging the values of $\mathcal{R}\mu\left(t,\theta\right)$ along all lines that contain $\mathbf{r}$. From the definition, $\mathbf{r} \in L_{t,\theta}$ if and only if $t = \langle \mathbf{r}, \mathbf{n}_\theta\rangle$, and hence, the lines containing $\mathbf{r}$ are parametrized by θ. We thus define the backprojection of μ by

$$\tilde{\mu}(\mathbf{r}) = \frac{1}{\pi}\int_0^\pi \mathcal{R}\mu\left(\langle \mathbf{r}, \mathbf{n}_\theta\rangle, \theta\right)\,\mathrm{d}\theta.$$

However, we should note that $\tilde{\mu}$ is not equal to μ in general since it is not an inversion formula.

Radon inversion formula

We use the connection of Radon transform to the Fourier transform to deduce its inversion formula. The result is termed as the projection-slice theorem.

Theorem 2.1.1. *For any $\theta \in [0,\pi)$, the Fourier transform of the projection $\mathcal{R}\mu(\cdot,\theta)$ satisfies*

$$\int \mathcal{R}\mu(t,\theta)e^{-i\omega t}\,\mathrm{d}t = \widehat{\mu}(\omega\cos\theta, \omega\sin\theta),$$

where $\widehat{\mu}$ is a Fourier transform of function μ. That is, measuring the Radon transform is equivalent to acquiring the Fourier transform of μ along radial lines.

Proof. From direction computations, we have

$$\begin{aligned}\int \mathcal{R}\mu(t,\theta)e^{-i\omega t}\,\mathrm{d}t &= \int\int\int \mu(\mathbf{r})\delta\left(\langle \mathbf{r}, \mathbf{n}_\theta\rangle - t\right)e^{-i\omega t}\,\mathrm{d}\mathbf{r}\,\mathrm{d}t \\ &= \int\int \mu(\mathbf{r})e^{-i\omega\langle \mathbf{r}, \mathbf{n}_\theta\rangle}\,\mathrm{d}\mathbf{r} \\ &= \widehat{\mu}(\omega\mathbf{n}_\theta)\end{aligned}$$

□

The above theorem says that we can recover μ from its radon transform with the information of its Fourier transform $\widehat{\mu}$ along radial lines. This observation leads to an expression of filtered backprojection (or the Radon inversion formula) similar to the backprojection formula.

Theorem 2.1.2. *A filtered backprojection of a function μ is given by*

$$\mu(\mathbf{r}) = \frac{1}{(2\pi)^2} \int_0^\pi \big(\mathcal{R}\mu(\cdot,\theta) \star h\big)(\langle \mathbf{r}, \mathbf{n}_\theta \rangle)\, \mathrm{d}\theta,$$

where h is such that $\widehat{h}(\omega) = |\omega|$, and $\star$ denotes convolution.

The term "filtered" corresponds to the fact that the above inversion formula has the same form as the backprojection formula, but instead of directly averaging the contributions, we first filter the Radon transform along the radial direction with convolution kernel h.

Proof. Using the inverse Fourier transform, we have

$$\begin{aligned}
\mu(\mathbf{r}) &= \frac{1}{(2\pi)^2} \int\int \widehat{\mu}(\boldsymbol{\omega}) e^{i<\boldsymbol{\omega},\mathbf{r}>}\, \mathrm{d}\boldsymbol{\omega}, \\
&= \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^\infty \widehat{\mu}(\rho \mathbf{n}_\theta) e^{i\rho<\mathbf{n}_\theta,\mathbf{r}>} \rho\, \mathrm{d}\rho\, \mathrm{d}\theta, \\
&= \frac{1}{(2\pi)^2} \int_0^{\pi} \int_0^\infty \widehat{\mu}(\rho \mathbf{n}_\theta) e^{i\rho<\mathbf{n}_\theta,\mathbf{r}>} \rho\, \mathrm{d}\rho\, \mathrm{d}\theta \\
&\quad + \frac{1}{(2\pi)^2} \int_\pi^{2\pi} \int_0^\infty \widehat{\mu}(\rho \mathbf{n}_\theta) e^{i\rho<\mathbf{n}_\theta,\mathbf{r}>} \rho\, \mathrm{d}\rho\, \mathrm{d}\theta,
\end{aligned}$$

where we have transformed to polar coordinates. The second integral can be rewritten as

$$\begin{aligned}
\int_\pi^{2\pi} \int_0^\infty \widehat{\mu}\left(\rho \mathbf{n}_\theta\right) e^{i\rho<\mathbf{n}_\theta,\mathbf{r}>} \rho\, \mathrm{d}\rho\, \mathrm{d}\theta &= \int_\pi^{2\pi} \int_0^{-\infty} \widehat{\mu}\left(-\rho \mathbf{n}_\theta\right) e^{-i\rho<\mathbf{n}_\theta,\mathbf{r}>} (-\rho)\, \mathrm{d}\rho\, \mathrm{d}\theta \\
&= \int_\pi^{2\pi} \int_{-\infty}^0 \widehat{\mu}\left(\rho \mathbf{n}_{\theta-\pi}\right) e^{i\rho<\mathbf{n}_{\theta-\pi},\mathbf{r}>} (-\rho)\, \mathrm{d}\rho\, \mathrm{d}\theta \\
&= \int_0^{\pi} \int_{-\infty}^0 \widehat{\mu}\left(\rho \mathbf{n}_\theta\right) e^{i\rho<\mathbf{n}_\theta,\mathbf{r}>} (-\rho)\, \mathrm{d}\rho\, \mathrm{d}\theta,
\end{aligned}$$

from which it follows that

$$\begin{aligned}
\mu(\mathbf{r}) &= \frac{1}{(2\pi)^2} \int_0^\pi \int_{-\infty}^\infty \widehat{\mu}(\rho \mathbf{n}_\theta) e^{i\rho<\mathbf{n}_\theta,\mathbf{r}>} |\rho|\, \mathrm{d}\rho\, \mathrm{d}\theta \\
&= \frac{1}{(2\pi)^2} \int_0^\pi \big(\mathcal{R}\mu(\cdot,\theta) \star h\big)(\langle \mathbf{r}, \mathbf{n}_\theta \rangle)\, \mathrm{d}\theta.
\end{aligned}$$

The last equality is obtained by the convolution theorem with $\widehat{h}(\rho) = |\rho|$, applied alongside the projection-slice theorem. □

Of course, in practice, it is not possible to acquire information arriving from rays in all of the continuum directions, and hence, one needs to resort to discretize in both t and θ. This means that we acquire only a fraction of the radial lines in the Fourier domain (refer to projection-slice theorem). Therefore with this partial information about the Fourier transform of μ, we approximate μ by summing the corresponding filtered backprojections

$$\mu^{\text{approx}}(\mathbf{r}) = \sum_k \big(\mathcal{R}\mu(\cdot, \theta_k) \star h\big)(\langle \mathbf{r}, \mathbf{n}_{\theta_k} \rangle).$$

2.2 Regularization theory

In this section, we describe three regularization techniques commonly encountered in the variational formulation of inverse problems. These techniques are Tikhonov regularization, ℓ_1 regularization, and total variation. These techniques are implicit in the sense that they promote a particular structure of the solution without explicitly defining it.

2.2.1 Tikhonov regularization

For Tikhonov regularization, one poses the optimization problem as

$$x_{\text{tr}} \triangleq \underset{x}{\operatorname{argmin}} \left\{ \|\mathcal{F}(x) - y\|^2 + \alpha \|x\|^2 \right\}, \tag{2.9}$$

with $\alpha > 0$. Here, x_{tr} is the optimal solution to the problem. This regularization addresses the ill-posedness and restores existence, uniqueness and stability. To see this, let's consider a linear forward operator, *i.e.*, $\mathcal{F}(x) = Ax$, where $A : \mathbb{R}^n \to \mathbb{R}^m$ is a matrix with rank $k < \min(m, n)$. This problem does not have a unique solution, as the null-space of the matrix A is not empty. A particular unregularized solution is

$$x^\star = \left(A^T A\right)^\dagger A^T y = V \Sigma^\dagger U^T y = \sum_{i=1}^{\min(m,n)} \frac{1}{\sigma_i} \left(u_i^T y\right) v_i,$$

where † denotes the pseudo-inverse, $A = U\Sigma V^T$ is the singular value decomposition of A with σ_i as singular values, and u_i and v_i denote the column vectors of the matrix U and V respectively. For an unregularized problem, the stability of the solution depends on the singular values of A. Since the matrix has rank k, that is, the singular values $\sigma_i = 0$ for $i = k+1, \ldots, \min(m, n)$, the unregularized solution $x^\star$ is unstable. Let us now look at the Tikhonov-regularized solution

$$x_{\text{tr}} = \left(A^T A + \alpha I\right)^{-1} A^T y = V \left(\Sigma^T \Sigma + \alpha I\right)^{-1} \Sigma U^T y = \sum_{i=1}^{\min(m,n)} \frac{\sigma_i}{\sigma_i^2 + \alpha} \left(u_i^T y\right) v_i.$$

This solution is stable since the zero singular values are shifted by a constant $\alpha > 0$.

A generalized Tikhonov regularization solves the problem

$$x_{\text{gtr}} \triangleq \underset{x}{\operatorname{argmin}} \left\{ \|\mathcal{F}(x) - y\|^2 + \alpha \|Kx\|^2 \right\},$$

where K is a linear operator. The idea here is that, instead of penalizing the image x as in the standard Tikhonov regularization ($K = I$, where I is an identity operator), the generalized variant penalizes the transformation of x represented by a linear operator K. For example, if we penalize the image gradient (*i.e.*, $K = \nabla$), then the solution will be smooth. If the forward operator is linear, we get an explicit expression for the solution to generalized Tikhonov regularization:

$$x_{\text{gtr}} = \left(A^T A + \alpha K^T K\right)^{-1} A^T y.$$

Tikhonov regularized least-squares problem can be solved by gradient-descent methods. If the forward operator is linear, then Krylov-based methods like conjugate-gradient (CG), generalized minimum residual (GMRES), are more efficient.

2.2.2 ℓ_1 regularization

An ℓ_1-regularized least-squares problem finds a solution

$$x_{\ell_1} \triangleq \underset{x}{\operatorname{argmin}} \left\{ \|\mathcal{F}(x) - y\|^2 + \alpha \|x\|_1 \right\}, \tag{2.10}$$

where $\|\cdot\|_1$ represents the Manhattan norm, i.e., $\|x\|_1 = \sum_i |x_i|$. The above optimization problem is known as the Least Absolute Shrinkage and Selection Operator (LASSO) when the forward operator is linear. The method was introduced in 1992 [162] and started receiving attention in 1996 [190] due to its ability to generate a sparse solution.

Let us consider a linear forward operator $\mathcal{F}(x) = Ax$ where $A \in \mathbb{R}^{m \times n}$ is a full-rank matrix ($\mathbf{rank}(A) = m$) with $m < n$. For a given m-dimensional vector y, there are multiple solutions that satisfy $Ax = y$ due to the large null-space of A. We might be interested in a solution with only a few active entries, *i.e.*, in a *sparse* solution. Sparsity can be measured using an ℓ_0 norm, which denotes the cardinality of the vector (the number of non-zero entries). Since an ℓ_0-norm is not convex, we instead use an ℓ_1-norm, the closest convex proxy to an ℓ_0-norm. It has been shown that the ℓ_1 regularized problem indeed provides the solution with maximum sparsity (*i.e.*, the solution with the minimum number of active entries) if matrix A satisfies the restricted isometry property (RIP) [43]. The RIP states that if there exists a constant $\delta_s \in (0, 1)$ such that for every $m \times s$ submatrix A_s of A and for every s-dimensional vector x,

$$(1 - \delta_s)\|x\|^2 \leq \|A_s x\|^2 \leq (1 + \delta_s)\|x\|^2$$

then the matrix A is said to satisfy the s-RIP with restricted isometry constant δ_s.

In many scenarios, an image itself may not be sparse. Instead, it is composed of a linear combination of only a few active orthonormal vectors (that are unknown, but the complete orthonormal basis is known). This means, if K represents an orthonormal basis ($K^T K = I$), then an image has a sparse coefficient in basis K (minimum $\|z\|_1$ such that $z = K\beta$). With such prior information, we get a generalized version of the ℓ_1-regularized least-squares problem:

$$x_{\ell_1} \triangleq \underset{x}{\operatorname{argmin}} \left\{ \|\mathcal{F}(x) - y\|^2 + \alpha \|K^T x\|_1 \right\},$$

This framework has found many applications due to the description of images using Fourier, wavelet or similar basis [58, 181].

ℓ_1-regularized problems can be solved with proximal methods. However, if the forward operator is linear, we can use a primal-dual method. ADMM can also be used to solve these problems.

2.2.3 Total variation

For a function $u : \Omega \to \mathbb{R}$, the total variation (TV)

$$TV(u) = \sup_{\|p\|_\infty \leq 1} \int_\Omega u(x) \operatorname{div} p(x) \, \mathrm{d}x,$$

where div is the divergence operator. p, a continuous function, plays a central role in variational formulation due to its numerous favorable properties: *i)* it preserves discontinuities since a sharp transition from a to b has the same cost as a smooth monotone transition, *ii)* it is convex and thus amenable to globally optimal solutions, *iii)* efficient algorithms exist to solve the problem. This regularization was introduced initially to denoise images [156]. It was observed that such regularization gives a piecewise-constant approximation of the noisy image (later proved rigorously in [193]). For C^1-continuous functions, the TV has a following more straightforward expression:

$$TV(u) = \int_\Omega |\nabla u(x)| \, \mathrm{d}x = \|\nabla u\|_1.$$

Hence, a TV-regularized least-squares problem finds a solution

$$x_{tv} \triangleq \operatorname*{argmin}_x \left\{ \|\mathcal{F}(x) - y\|^2 + \alpha \|\nabla x\|_1 \right\},$$

where $\alpha > 0$ is a regularization parameter. In essence, this approach tries to produce a piecewise constant image by penalizing the edges (captured by the gradient). Here, the ℓ_1-norm promotes sparsity in the gradient and hence reduces the blurry effects (the smooth transitions) in the image.

TV-regularized problems can be solved numerically using proximal methods. If the forward operator is linear, then the resulting problem is solved using a primal-dual method. ADMM is also useful in solving these problems.

2.3 Discrete prior - shape reconstruction

In this section, we consider a particular class of explicit regularization technique that incorporates a discrete prior. A *discrete* image consists of a few homogeneous regions (or domains). If the impedance (or grey level) of these regions is known, we call this information a discrete prior. In many tomographic applications, we know the impedance of the materials in the object: for example, in electron tomography, the types of atoms present in the molecule under investigation is known. Hence, if a discrete prior is available in tomographic reconstructions, we can cast the problem

as finding the shape of these homogeneous regions. The procedure of finding the optimal shape of a domain that minimizes or maximizes a given criterion (*i.e.*, the objective function) is known as *shape optimization*. Here, we briefly discuss the level-set based approach for shape optimization. As such, shape optimization has a long history. Therefore, there is a vast literature in this field, and we refer to the articles [179, 4, 22, 119, 3, 178, 87, 198], and references therein.

A *discrete* image x consists of k distinct homogeneous regions, each of grey level u_i $(i = 1, \ldots, k)$,

$$x(\mathbf{r}) = \begin{cases} u_1 & \mathbf{r} \in \Theta_1 \\ \vdots & \vdots \\ u_k & \mathbf{r} \in \Theta_k \end{cases}, \qquad \Theta = \Theta_1 \cup \cdots \cup \Theta_k, \qquad \Theta_i \cap \Theta_j = \emptyset \, \forall \, i \neq j,$$

where $\Theta \in \mathbb{R}^d$ is an entire region of the image. We restrict our discussion to a *binary* image, that consists of only two grey levels, but it can be easily extended to discrete images. A binary image x has two regions Θ_0 and Θ_1 of grey levels u_0 and u_1 respectively. Generally, we regard the domain with grey-level u_0 as *background* and the area with grey-level u_1 as *inclusion*. It is easy to observe that $\Theta_0 = \Theta \setminus \Theta_1$. Hence, to retrieve a binary image (with known region Θ, and grey levels u_0 and u_1), it is sufficient to find the region Θ_1. Therefore, a tomographic reconstruction with the *binary prior* aims to retrieve the shape of the inclusion. Let us denote the inclusion Θ_1 by a characteristic function χ_{Θ_1} defined as

$$\chi_{\Theta_1}(\mathbf{r}) = \begin{cases} 1 & \mathbf{r} \in \Theta_1, \\ 0 & \mathbf{r} \notin \Theta_1. \end{cases}$$

The binary image of greylevels u_0 and u_1 is then mathematically represented as

$$x(\mathbf{r}) = u_0 \left(1 - \chi_{\Theta_1}(\mathbf{r})\right) + u_1 \chi_{\Theta_1}(\mathbf{r}). \tag{2.11}$$

Hence, finding a binary image amounts to finding a characteristic function χ_{Θ_1}. This leads to an integer programming problem, an optimization problem over the domain of integers (or discrete values). This problem is shown to be NP-complete [110], meaning it cannot be guaranteed that an exact solution is found in polynomial time with respect to the number of variables. Although the problem is NP-complete, we can still find an approximate solution by various methods. We discuss one such approach, called the level-set method, in the next subsection.

2.3.1 Level-set method

A binary image is implicitly related to the domain Θ_1 (see equation (2.11)). Rather than finding the domain Θ_1 or its characteristic function χ_{Θ_1}, we need a more explicit way to relate the domain to a binary image. Here is where the level-set proves to be useful. Let us denote the boundary of the domain Θ_1 by $\partial\Theta_1$, and represent it as the zero level-set of a continuous function $\phi : \Theta \to \mathbb{R}$. The zero

level-set relates the function ϕ to the domain Θ_1 and its boundary $\partial\Theta_1$ as

$$\begin{cases} \phi(\mathbf{r}) > 0 & \forall\, \mathbf{r} \in \Theta_1 \\ \phi(\mathbf{r}) = 0 & \forall\, \mathbf{r} \in \partial\Theta_1 \\ \phi(\mathbf{r}) < 0 & \forall\, \mathbf{r} \notin \Theta_1 \end{cases}.$$

We refer to the function ϕ as a level-set function for our convenience. With such level-set function ϕ, we can represent a binary image as

$$x(\mathbf{r}) = u_0 \left(1 - h\left(\phi\left(\mathbf{r}\right)\right)\right) + u_1 h\left(\phi\left(\mathbf{r}\right)\right),$$

where h is the Heaviside function, *i.e.*, $h(x) = 0.5(1 + \text{sign}(x))$. Hence, an inverse problem regularized by a level-set takes the following constrained least-squares form:

$$\begin{aligned} \min_{x,\phi} \quad & \|\mathcal{F}(x) - y\|^2 \\ \text{subject to} \quad & x = u_0 \left(1 - h\left(\phi\right)\right) + u_1 h\left(\phi\right), \end{aligned} \tag{2.12}$$

Here, the constraint acts as a regularization on the image x. This binary constraint restricts the values of x to u_0 and u_1, and hence, we call this regularization *explicit*. There are multiple ways to solve the above problem in two variables. A standard approach is to get rid of one of the variables and solve for the remaining. This approach is called a *reduced* approach, and the corresponding problem as the reduced problem. In the level-set case, we can get rid of x by substituting it, in terms of ϕ, in the cost function. Note that the optimization problem becomes unconstrained by getting rid of the constraints.

Two kinds of inter related methods are available to solve the reduced problem. In the traditional approach, we evolve the level-set function in such a way that the evolved form of the function minimizes the cost function. To take the concept of evolution into account, we define an artificial time where a level-set function ϕ at every time-frame $t \geq 0$ is rewritten as $\phi(\mathbf{r}, t)$. The zero level-set of $\phi(\mathbf{r}, t)$ denotes the boundary of domain Θ_1 at time t, *i.e.*, $\partial\Theta_1$. The Hamilton-Jacobi equation represents this evolution process

$$\frac{\partial\phi}{\partial t}(\mathbf{r}, t) + \mathbf{v}(\mathbf{r}, t) \cdot \nabla\phi(\mathbf{r}, t) = 0 \qquad \Longrightarrow \qquad \frac{\partial\phi}{\partial t}(\mathbf{r}, t) + V(\mathbf{r}, t)|\nabla\phi(\mathbf{r}, t)| = 0,$$

where $\mathbf{v}$ denotes the velocity of the evolution. The second equation indicates the more straightforward form that assumes the movement of the interface in the normal direction. Hence, the velocity $\mathbf{v}(\mathbf{r}, t) = V(\mathbf{r}, t)\vec{\mathbf{n}}(\mathbf{r}, t)$ with V being a scalar speed function, and $\vec{\mathbf{n}} = \nabla\phi/|\nabla\phi|$ the unit outward vector on $\partial\Theta_1$. The obvious choice for the speed function V is the gradient of the cost function with respect to the characteristic function $\chi = h(\phi)$ [63, 41]. Since the gradient is only valid on the boundary of the domain, a velocity extension to the entire domain is required to evolve the level-set function globally [141].

Another approach relies on the relaxation of the Heaviside function. If we approximate the Heaviside function using a smooth function, we can obtain the

gradient and the Hessian of the cost function with respect to the level-set function. A first-order optimization scheme that solely relies on the gradient of the function, is similar to the traditional approach that involves a Hamilton-Jacobi equation. The iterates of the first-order scheme can be thought of as the time samples of the level-set function evolved via a Hamilton-Jacobi equation with appropriate step-size. The second-order scheme that uses the gradient and the Hessian converges faster than the first order scheme by providing a better estimation of the velocity (in the Hamilton-Jacobi equation).

2.4 Optimization

For simplicity, we express the optimization problem in (2.1) as follows:

$$\min_{x} \quad \mathcal{C}(x) = g\,(x) + h(Kx), \tag{2.13}$$

$$\text{where} \quad g(x) = \|\mathcal{F}(x) - y\|^2, \quad h(Kx) = \alpha\mathcal{R}(x),$$

and we introduce a linear operator K for generalization. For standard Tikhonov and ℓ_1 regularization, $K = I$, an identity matrix, while for total variation, $K = D$, a discrete operator that approximates the spatial derivatives. We first provide a quick overview on the convex optimization theory [34]. We later discuss various algorithms to solve problems of the form (2.13).

2.4.1 Convex set and functions

A set X is convex if the line segment between any two points in X lies in X, i.e., if for any $x_1, x_2 \in X$ and any θ with $0 \leq \theta \leq 1$, we have $\theta x_1 + (1-\theta)x_2 \in X$.

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if the domain of f, denoted by $\mathbf{dom}\, f$, is a convex set and if

$$f\,(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y). \qquad \forall x, y \in \mathbf{dom}\, f, 0 \leq \theta \leq 1 \tag{2.14}$$

Geometrically, it means that the line segment between $(x, f(x))$ and $(y, f(y))$ lies above the graph of f. A strict inequality makes the function f strictly convex. We say f is concave if $-f$ is convex.

First-order condition

Suppose f is differentiable (*i.e.*, its gradient ∇f exists at each point in $\mathbf{dom}\, f$, which is open). Then f is convex if and only if

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

holds for all $x, y \in \mathbf{dom}\, f$. The above inequality states that for a convex function, the first order Taylor approximation is in fact a *global underestimator* of the function. It also shows that from local information about a convex function (*i.e.*, its derivative at a point) we can derive global information (*i.e.*, a global underestimator of it).

Second-order condition

We now assume that f is twice differentiable, i.e., its Hessian or second derivative $\nabla^2 f$ exists at each point in $\mathbf{dom}\, f$, which is open. Then f is convex if and only if its Hessian is positive semidefinite:

$$\nabla^2 f(x) \succeq 0,$$

for all $x \in \mathbf{dom}\, f$. If $\nabla^2 f(x) \succ 0$ for all $x \in \mathbf{dom}\, f$, then f is strictly convex.

The conjugate function

Given $f : \mathbb{R}^n \to \mathbb{R}$, its *conjugate* function $f^\star : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$f^\star(y) = \sup_{x \in \mathbf{dom}\, f} \left\{ y^T x - f(x) \right\}.$$

Since $f^\star$ is the pointwise supremum of a family of convex functions of y, it is a convex function. $f^\star$ is convex even if f is not. From the above definition, we get the Fenchel's inequality which states that for all x, y,

$$f(x) + f^\star(y) \geq x^T y$$

holds. The conjugate of conjugate of convex function is the original function, *i.e.*, $f^{\star\star} = f$.

Subgradient of a function

g is a subgradient of f at x if

$$f(y) \geq f(x) + g^T(y - x) \quad \text{for all } y,$$

If f is convex and differentiable, $\nabla f(x)$ is a subgradient of f at x. A set of all subgradients of f at x is called the subdifferential of f at x, denoted $\partial f(x)$:

$$\partial f = \left\{ g \,:\, f(y) \geq f(x) + g^T(y - x) \quad \forall\, y \right\}.$$

For any f, subdifferential $\partial f(x)$ is a closed convex set. If f is a convex function, then $\partial f(x)$ is non-empty, and furthermore, if f is differentiable, then $\partial f(x) = \{\nabla f(x)\}$.

2.4.2 Convex optimization

An optimization problem finds an x that minimizes the cost function $\mathcal{C}(x)$ among all x :

$$x^\star = \underset{x}{\operatorname{argmin}} \left\{ \mathcal{C}(x) = g\,(x) + h(Kx) \right\},$$

where $x \in \mathbb{R}^n$ is the optimization variable and the function $\mathcal{C} : \mathbb{R}^n \to \mathbb{R}$ is the total objective function or the cost function. An optimization problem is called *convex optimization* problem only when the cost function $\mathcal{C}$ is a convex function. $x^\star$ is the optimal solution for which the function C attains its minimum value.

2.4.3 Proximal operators

A multivalued or set-valued operator $\mathcal{H} : \mathbb{R}^n \to \mathbb{R}^n$ maps a point $x \in \mathbb{R}^n$ to sets $\mathcal{H}(x) \in \mathbb{R}^n$. The operator is *monotone* if

$$(\mathcal{H}(x) - \mathcal{H}(\hat{x}))^T (x - \hat{x}) \geq 0 \quad \forall\, x, \hat{x}.$$

A monotone operator $\mathcal{H}$ is considered *maximal monotone* if there is no other monotone operator that properly contains it. The operator $(\mathcal{I} + \gamma\mathcal{H})^{-1}$, with $\gamma > 0$ is called the resolvent of the operator $\mathcal{H}$. $\mathcal{I}$ is an identity operator that maps a point $x \in \mathbb{R}^n$ to itself. The resolvent $(\mathcal{I} + \gamma\mathcal{H})^{-1}(y)$ exists and is unique for all $y \in \mathbb{R}^n$ if $\mathcal{H}$ is maximal monotone [35, Proposition 2.2]. We should note that the value $x = (\mathcal{I} + \gamma\mathcal{H})^{-1}(y)$ of the resolvent is the unique solution of the monotone inclusion

$$y \in x + \gamma\mathcal{H}(x).$$

A resolvent of a monotone operator is a non-expansive operator. An operator $\mathcal{H}$ is non-expansive if

$$\|\mathcal{H}(x) - \mathcal{H}(y)\| \leq \|x - y\|, \quad \forall\, x, y.$$

A *proximal operator* (also known as *prox-operator*) of a closed convex function h is the resolvent with $\mathcal{H} = \partial h$, a sub-differential of a function h. The *prox-operator* reads as

$$\mathbf{prox}_{\gamma h} = (\mathcal{I} + \gamma\partial h)^{-1},$$

and it maps $x \in \mathbb{R}^n$ to the unique solution of the optimization problem

$$x^\star = \underset{x}{\operatorname{argmin}} \left\{ h(x) + \frac{1}{2\gamma}\|y - x\|^2 \right\}.$$

The prox-operator of a function and its conjugate are related by the *Moreau identity*,

$$\mathbf{prox}_{\gamma h^\star}(x) + \gamma\, \mathbf{prox}_{h/\gamma}\left(\frac{x}{\gamma}\right) = x. \tag{2.15}$$

This property is useful in terms of computing the prox-operator of a function from the prox-operator of its conjugate, and the other way around.

2.4.4 Gradient-based methods

We discuss mainly two gradient-based schemes: Newton and the Quasi-Newton method. For a twice-differentiable function $\mathcal{C}(x)$, the Taylor series around point x_0 gives

$$\mathcal{C}(x) = \mathcal{C}(x_0) + \nabla\mathcal{C}(x_0)^T (x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2\mathcal{C}(x_0)\,(x - x_0) + O\left(\|x - x_0\|^3\right).$$

The Newton and the Quasi-Newton methods generate a sequence based on the above quadratic approximation to the fuction $\mathcal{C}$ at every iterate of the sequence.

The procedure is as follows:

$$\begin{aligned} s^{(k)} &= \underset{s}{\operatorname{argmin}} \left\{ s^T \nabla \mathcal{C}\left(x^{(k)}\right) + \frac{1}{2} s^T H_k s \right\}, \\ \alpha_k &= \underset{\alpha}{\operatorname{argmin}} \left\{ \mathcal{C}\left(x^{(k)} + \alpha s^{(k)}\right) \right\}, \\ x^{(k+1)} &= x^{(k)} + \alpha_k s^{(k)}. \end{aligned} \tag{2.16}$$

Here, $H_k = \nabla^2 \mathcal{C}(x^k)$ is the Hessian for the Newton method, while we approximate the Hessian for the Quasi-Newton method. The standard algorithm for an approximation, a BFGS scheme [73], updates H_{k+1} as follows:

$$H_{k+1} = H_k - \frac{H_k p_k p_k^T H_k}{p_k^T H_k p_k} + \frac{y_k y_k^T}{y_k^T p_k},$$

where $p_k = x^{(k)} - x^{(k-1)}$ and $y_k = \nabla \mathcal{C}(x^{(k)}) - \nabla \mathcal{C}(x^{(k-1)})$. It is easy to observe that these updates are rank-one. The BFGS Hessian can also be represented as

$$H_{k+1} = H_0 - QRQ^T \tag{2.17}$$

where

$$Q = \begin{bmatrix} H_0 P_k & Y_k \end{bmatrix}, \qquad R = \begin{bmatrix} P_k^T H_0 P_k & L_k \\ L_k^T & -D_k \end{bmatrix}^{-1}, \qquad D_k = \begin{bmatrix} p_0^T y_0 & & \\ & \cdots & \\ & & p_k^T y_k \end{bmatrix},$$

$$P_k = \begin{bmatrix} p_0 \, p_1 \, \ldots \, p_k \end{bmatrix}, \qquad Y_k = \begin{bmatrix} y_0 \, y_1 \, \ldots \, y_k \end{bmatrix},$$

and the matrix L is a lower triangular matrix with elements

$$l_{ij} = \begin{cases} p_{i-1}^T y_{j-1} & \text{if } i > j, \\ 0 & \text{otherwise.} \end{cases}$$

The initial approximation to the Hessian is set to a constant diagonal matrix, *i.e.*, $H_0 = \beta I$, where $\beta = y_0^T p_0 / p_0^T p_0$. As is evident from (2.17), the matrix H_k becomes denser as $k \to \infty$, and computationally expensive. To reduce the computations, we use limited-memory BFGS. Rather than working with all the gradients, it relies on the information from the m most recent gradients. The sizes of matrix Q and R are reduced to $n \times 2m$ and $2m \times 2m$ respectively.

Here, the *linesearch* refers to the strategy to compute α_k. There are various ways to perform *linesearch*, the prominent one being backtracking. This linesearch procedure is based on finding a value $\beta \in (0, 1]$ such that the Armijo rule is satisfied, *i.e.*,

$$\mathcal{C}\left(x^{(k)} + \beta s^{(k)}\right) \leq \mathcal{C}\left(x^{(k)}\right) + \beta \sigma \nabla \mathcal{C}\left(x^{(k)}\right)$$

for a $\sigma \in (0, 0.5)$. If $\mathcal{C}$ is a convex function, then the Newton and Quasi-Newton method converge to a global minimum. If $\mathcal{C}$ is non-convex, these methods can only guarantee the convergence to a local optimum. Algorithm 1 describes the Newton and/or Quasi-Newton method.

Algorithm 1 Newton/Quasi-Newton Method

for $k = 0$ to $k_{\max}$ **do**
    compute the gradient $\nabla \mathcal{C}(x^{(k)})$
    compute the (exact/approximate) Hessian H_k
    $s^{(k)} = H_k^{-1} \nabla \mathcal{C}(x^{(k)})$
    $\alpha_k = \text{linesearch}_\alpha \mathcal{C}\left(x^{(k)} + \alpha_k s^{(k)}\right)$
    $x^{(k+1)} = x^{(k)} + \alpha_k s^{(k)}$
    check optimality conditions
end for

2.4.5 Proximal methods

In this section, we discuss the proximal version of gradient-based methods. When the cost function $\mathcal{C}(x)$ is non-differentiable (because of the non-differentiability of the function h), we can no longer compute the gradient of the cost function. With the Taylor series expansion of g around x_0, we can express the cost function as

$$\begin{aligned}\mathcal{C}(x) = g(x_0) + \nabla g(x_0)^T (x - x_0) + (x - x_0)^T \nabla^2 g(x_0) (x - x_0) \\ + h(Kx) + O(\|x - x_0\|^2).\end{aligned}$$

The proximal method generates a sequence of iterates based on the above quadratic approximation of the cost function at the current iterate.

$$\begin{aligned} s^{(k)} &= \underset{s}{\text{argmin}} \left\{ s^T \nabla g\left(x^{(k)}\right) + \frac{1}{2} s^T H_k s + h\left(K\left(x^{(k)} + s\right)\right) \right\}, \\ \alpha_k &= \underset{\alpha}{\text{argmin}} \left\{ \mathcal{C}\left(x^{(k)} + \alpha s^{(k)}\right) \right\}, \\ x^{(k+1)} &= x^{(k)} + \alpha_k s^{(k)}. \end{aligned} \tag{2.18}$$

The steps in (2.18) can be summarized as follows: Step (a) finds a search direction $\mathbf{s}^k$. It involves the minimization of quadratic approximation of f at $\mathbf{x}^{(k)}$, ensuring that it satisfies the constraints. In step (b), we define a function $\hat{\mathbf{x}} : \mathbb{R} \to \mathbb{R}$ which is a proximal of the iterate $\mathbf{x}^k + \alpha \mathbf{s}^{(k)}$ with respect to g. The function $\hat{\mathbf{x}}$ ensures that the step length, α, must satisfy the constraints. Step (c) does a line-search with respect to the feasible α. Once we obtain the correct α, we update our variable of interest $\mathbf{x}$ in step (d).

The minimization problem in step (a) of (2.18) is a convex minimization problem. The cost function is the sum of two convex functions. To solve this minimization problem, we use the first-order primal-dual method described in the next subsection. Please refer to [120] for detailed discussions.

2.4.6 Primal-dual method

This method assumes that the function g is a differentiable closed convex function and h is a closed non-differentiable convex function. It is based on a fixed-point method. A fixed point of an operator $\mathcal{T} : \mathbb{R}^n \to \mathbb{R}^n$ is defined as the set of points

Algorithm 2 Proximal Newton/Quasi-Newton Method

1: **for** $k = 0$ to k_{max} **do**
2: compute the gradient $\nabla g(x^{(k)})$
3: compute the (exact/approximate) Hessian H_k
4: compute step size by solving step (1) of (2.18)
5: $\alpha_k = \text{linesearch}_\alpha \, \mathcal{C}\left(x^{(k)} + \alpha_k s^{(k)}\right)$
6: $x^{(k+1)} = x^{(k)} + \alpha_k s^{(k)}$
7: check optimality conditions
8: **end for**

$z \in \mathbb{R}^n$ such that $\mathcal{T}(z) = z$. A fixed point method finds one such point by generating a sequence of iterates $z^{(k)}$ with $k = 1, \ldots, n$ of form

$$z^{(k+1)} = \mathcal{T}\left(z^{(k)}\right),$$

given an initial point $z^{(0)}$. The iterates converge to one of the fixed points if $\mathcal{T}$ is a non-expansive operator. Now recall that the resolvent of a monotone operator is a non-expansive operator. Also, it can be easily seen that the zeros of the maximal monotone operator are the fixed points of its resolvent. Hence, the fixed point iterations take the following form to find the zeros of a monotone operator $\mathcal{H}$:

$$z^{(k+1)} = (\mathcal{I} + \alpha\mathcal{H})^{-1} z^{(k)}.$$

A more efficient scheme to find the zero of $\mathcal{H}$ is a preconditioned fixed-point method. This iteration scheme generates a sequence

$$z^{(k+1)} = \left(\mathcal{I} + \mathcal{P}^{-1}\mathcal{H}\right)^{-1} z^{(k)},$$

with $\mathcal{P}$ being a symmetric positive-definite linear operator.

The optimality conditions for (2.13) states that a zero-vector must be in the subdifferential of the cost function, *i.e.*,

$$0 \in \nabla g(\mathbf{x}) + K^T \partial h(Kx), \tag{2.19}$$

where $\partial h : \mathbb{R}^m \to \mathbb{R}^m$ is the subdifferential of function h. Let's consider variable $u \in \mathbb{R}^m$ in the subdifferential of h *i.e.*,

$$u \in \partial h(Kx), \qquad \implies 0 \in \partial h^\star(u) - Kx, \tag{2.20}$$

where $h^\star$ is the convex conjugate of the function h. From equations (2.19) and (2.20), we can write the optimality conditions in the form of a following system

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \underbrace{\begin{bmatrix} \nabla g & K^T \\ -K & \partial h^\star \end{bmatrix}}_{\mathcal{H}} \underbrace{\begin{bmatrix} x \\ u \end{bmatrix}}_{\mathbf{z}}. \tag{2.21}$$

It is easy to show that the operator $\mathcal{H}$ is a monotone operator. Considering a preconditioner operator

$$\mathcal{P} = \begin{bmatrix} \frac{1}{\gamma}\mathcal{I} & -K^T \\ -K & \frac{1}{\gamma}\mathcal{I} \end{bmatrix},$$

with $\gamma > 0$, the fixed point iteration scheme results in the following system:

$$\begin{bmatrix} \frac{1}{\gamma}\mathcal{I} + \nabla g & 0 \\ -2K & \frac{1}{\gamma}\mathcal{I} + \partial h^\star \end{bmatrix} \begin{bmatrix} x^{(t+1)} \\ u^{(t+1)} \end{bmatrix} = \begin{bmatrix} \frac{1}{\gamma}\mathcal{I} & -K^T \\ -K & \frac{1}{\gamma}\mathcal{I} \end{bmatrix} \begin{bmatrix} x^{(t)} \\ u^{(t)} \end{bmatrix}.$$

Now each row can be expressed in terms of the resolvent and can be computed explicitly. Hence, we obtain the following primal-dual algorithm

$$\begin{aligned} x^{(t+1)} &= (\mathcal{I} + \gamma \nabla g)^{-1} \left(x^{(t)} - \gamma K^T u^{(t)} \right) \\ u^{(t+1)} &= (\mathcal{I} + \gamma \partial h^\star)^{-1} \left(u^{(t)} - \gamma K \left(x^{(t)} - 2x^{(t+1)} \right) \right) \end{aligned} \tag{2.22}$$

If the proximal operators of functions g and h are simple, then each iteration can be computed efficiently. For the proximal operator of the conjugate $h^\star$, we use the Moreau identity (2.15).

The convergence of primal-dual algorithm depends on the single parameter γ. This parameter must be chosen such that the preconditioner operator is positive semi-definite, *i.e.*, $\mathcal{P} \succeq 0$. The positive definiteness of the operator is related to its Schur complement: 'If operator M is positive definite then so is its Schur complement. For a block operator $M = \begin{bmatrix} A & B \,;\, C & D \end{bmatrix}$ with appropriate sizes of operators A, B, C, and D, the schur complement of block D of the matrix M is $M/D = A - BD^{-1}C$. Hence, for a preconditioner operator $\mathcal{P}$ with $D = \mathcal{I}/\gamma$, we get a schur complement $\mathcal{P}/D = \frac{1}{\gamma}\mathcal{I} - \gamma K^T K$. Hence, the condition on γ is $0 < \gamma \leq 1/\sqrt{\rho\left(K^T K\right)}$, where ρ denotes the spectral radius. Algorithm 3 describes the primal-dual method. Please refer to [157] for detailed discussion regarding this method.

Algorithm 3 Primal-Dual Method

1: **for** $k = 0$ to k_{max} **do**
2: $\quad x^{(k+1)} = \mathbf{prox}_{\gamma g}\left(x^{(k)} - \gamma K^T u^{(k)}\right)$
3: $\quad u^{(k+1)} = \mathbf{prox}_{\gamma h^\star}\left(u^{(k)} - \gamma K\left(x^{(k)} - 2x^{(k+1)}\right)\right)$
4: $\quad$ check optimality conditions
5: **end for**

2.4.7 Alternating direction method of multipliers

We can rewrite the minimization problem in (2.13) as

$$\begin{aligned} \min_{x,z} \quad & g(x) + h(z), \\ \text{subject to} \quad & z = Kx, \end{aligned} \tag{2.23}$$

by introducing the variable $z \in \mathbb{R}^m$. An augmented Lagrangian for this problem reads

$$\mathcal{L}_\rho(x, z, u) = g(x) + h(z) + u^T(z - Kx) + \frac{\rho}{2}\|z - Kx\|^2$$

with $\rho > 0$ a penalty parameter and $u \in \mathbb{R}^m$ a Lagrange multiplier corresponding to the equality constraint $z = Kx$. The saddle point of this augmented Lagrangian corresponds to the optimal solution of (2.23). Hence, to find a saddle point of the augmented Lagrangian

$$(x^\star, z^\star, u^\star) = \max_u \left(\min_{x,z} \mathcal{L}_\rho(x, z, u) \right),$$

the alternating direction method of multipliers (ADMM) suggests a scheme to generate a sequence of iterates that optimizes the augmented Lagrangian with respect to each variable. The scheme leads to the following:

for $k = 0, \ldots, k_{max}$

$$\begin{aligned}
x^{(k+1)} &= \operatorname*{argmin}_x \left\{ g(x) + \left(u^{(k)}\right)^T \left(z^{(k)} - Kx\right) + \frac{\rho}{2}\|z^{(k)} - Kx\|^2 \right\}, \\
z^{(k+1)} &= \operatorname*{argmin}_z \left\{ h(z) + \left(u^{(k)}\right)^T \left(z - Kx^{(k+1)}\right) + \frac{\rho}{2}\|z - Kx^{(k+1)}\|^2 \right\}, \\
u^{(k+1)} &= u^{(k)} + \rho\left(z^{(k+1)} - Kx^{(k+1)}\right).
\end{aligned}$$

In the description of Algorithm 4, we have replaced the update in z by the proximal operation. Please refer to [33] for more information about this method.

Algorithm 4 ADMM

1: **for** $k = 0$ to $k_{\max}$ **do**
2: $\quad x^{(k+1)} = \operatorname{argmin}_x \left\{ g(x) + \rho/2\|Kx - z^{(k)} + u^{(k)}/\rho\|^2 \right\}$
3: $\quad z^{(k+1)} = \mathbf{prox}_{\rho h}\left(Kx^{(k+1)} + u^{(k)}/\rho\right)$
4: $\quad u^{(k+1)} = u^{(k)} + \rho\left(z^{(k+1)} - Kx^{(k+1)}\right)$
5: $\quad$ check optimality conditions
6: **end for**

CHAPTER 3

Binary tomography

Abstract - *Binary tomography is concerned with the recovery of binary images from a few of their projections (i.e., sums of the pixel values along various directions). To reconstruct an image from noisy projection data, one can pose it as a constrained least-squares problem. As the constraints are non-convex, many approaches for solving it rely on either relaxing the constraints or heuristics. We propose a novel convex formulation, based on the Lagrange dual of the constrained least-squares problem. The resulting problem is a generalized LASSO problem which can be solved efficiently. It is a relaxation in the sense that it can only be guaranteed to give a feasible solution; not necessarily the optimal one. In exhaustive experiments on small images (*2×2*,* 3×3*,* 4×4*) we find, however, that if the problem has a unique solution, our dual approach finds it. In the case of multiple solutions, our approach finds the commonalities between the solutions. Further experiments on realistic numerical phantoms and an experiment on an X-ray dataset show that our method compares favorably to Total Variation and DART.*

This chapter is based on the following publications:

A Kadu and T van Leeuwen. A convex formulation for binary tomography. *IEEE Transactions on Computational Imaging*, pages 1–1, 2019.

" Truth is ever to be found in simplicity, and not in the multiplicity and confusion of things."
- Isaac Newton

3.1 Introduction

Discrete tomography is concerned with the recovery of discrete images (i.e., images whose pixels take on a small number of prescribed grey values) from a few of their projections (i.e., sums of the pixel values along various directions). Early work on the subject mostly deals with the mathematical analysis, combinatorics, and geometry. Since the 1970s, the development of algorithms for discrete tomography has become an active area of research as well [90]. It has found applications in image processing and computer vision [163, 171], atomic-resolution electron microscopy [159, 45], medicine imaging [170, 47] and material sciences [130, 10, 196, 17].

3.1.1 Mathematical formulation

The discrete tomography problem may be mathematically formulated as follows. We represent an image by a grid of $N = n \times n$ pixels taking values $x_j \in \mathcal{U} = \{u_0, u_1, \ldots, u_K\}$. The projections are linear combinations of the pixels along m different (lattice) directions. We denote the linear transformation from image to projection data by

$$\mathbf{y} = \mathbf{A}\mathbf{x},$$

where x_j denotes the value of the image in the j^{th} cell, y_i is the (weighted) sum of the image along the i^{th} ray, and a_{ij} is proportional to the length of the i^{th} ray in the j^{th} cell[1].

The goal is to find a solution to this system of equations with the constraint that $x_j \in \mathcal{U}$, i.e.,

$$\text{find} \quad \mathbf{x} \in \mathcal{U}^N \quad \text{such that} \quad \mathbf{A}\mathbf{x} = \mathbf{y}.$$

When the system of equations does not have a unique solution, finding one that only takes values in $\mathcal{U}^N$ has been shown to be an NP-hard problem for more than 3 directions, *i.e.*, $m \geq 3$ [76].

Due to the presence of noise, the system of equations may not have a solution, and the problem is sometimes formulated as a constrained least-squares problem

$$\min_{\mathbf{x} \in \mathcal{U}^N} \quad \tfrac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2. \tag{3.1}$$

Obviously, the constraints are non-convex and solving (3.1) exactly is not trivial. Next, we briefly discuss some existing approaches for solving it.

3.1.2 Literature review

Methods for solving (3.1) can be roughly divided into four classes: algebraic methods, stochastic sampling methods, (convex) relaxation and (heuristic/greedy) combinatorial approaches.

[1] We note that other projection models exist and can be similarly represented by a_{ij}.

The *algebraic* methods exploit the algebraic structure of the problem and may give some insight into the (non-) uniqueness and the required number of projections [211, 86]. While theoretically very elegant, these methods are not readily generalized to realistic projection models and noisy data.

The *stochastic sampling methods* typically construct a probability density function on the space of discrete images, allowing one to sample images and use Markov Chain Monte Carlo type methods to find a solution [127, 50, 74]. These methods are very flexible but may require a prohibitive number of samples when applied to large-scale datasets.

Relaxation methods are based on some form of convex or non-convex relaxation of the constraint. This allows for a natural extension of existing variational formulations and iterative algorithms [48, 202, 44, 169, 192]. While these approaches can often be implemented efficiently and apply to large-scale problems, getting them to converge to the correct discrete solution can be challenging. Another variant of convex relaxation includes the linear-programming based method [114]. This method works well on small-scale images and noise-free data.

The *heuristic* algorithms, finally, combine ideas from combinatorial optimization and iterative methods. Such methods are often efficient and known to perform well in practice [13, 18].

A more extensive overview of various methods for binary tomography and variants thereof (e.g., with more than two grey levels) are discussed in [89].

3.1.3 Contributions and outline

We propose a novel, convex, reformulation for discrete tomography with two grey values $\{u_0, u_1\}$ (often referred to as *binary* tomography). Starting from the constrained least-squares problem (3.1) we derive a corresponding Lagrange dual problem, which is convex by construction. Solving this problem yields an image with pixel values in $\{u_0, 0, u_1\}$. Setting the remaining zero-valued pixels to u_0 or u_1 generates a *feasible* solution of (3.1) but not necessarily an *optimal* one. In this sense, our approach is a relaxation. Exhaustive enumeration of small-scale ($n = 2, 3, 4$) images with few directions ($m = 2, 3$) show that if the problem has a unique solution, then solving the dual problem yields the correct solution. When there are multiple solutions, the dual approach finds the common elements of the solutions, leaving the remaining pixels undefined (zero-valued). We conjecture that this holds for larger n and m as well. This implies that we can only expect to usefully solve problem instances that allow a unique solution and characterize the non-uniqueness when there are a few solutions.

For practical applications, the most relevant setting is where the equations alone do not permit a unique solution, but the constrained problem does. Otherwise, more measurements or prior information about the object would be needed in order to usefully image it. With well-chosen numerical experiments on synthetic and real data, we show that our new approach is competitive for practical applications in X-ray tomography as well.

The outline of the chapter is as follows. We first give an intuitive derivation of the dual problem for invertible $\mathbf{A}^T\mathbf{A}$ before presenting the main results for general $\mathbf{A}$. We then discuss two methods for solving the resulting convex problem. Then,

we offer the numerical results on small-scale binary problems in Section 3.3 to support our conjecture. Numerical results on numerical phantoms and real data are presented in Section 3.4. Finally, we conclude the chapter in Section 3.5.

3.2 Dual problem

For the purpose of the derivation, we assume that the problem has pixel values ± 1. The least-squares *binary* tomography problem can then be formulated as:

$$\mathbf{x}^\star \triangleq \underset{\mathbf{x}}{\operatorname{argmin}} \left(\inf_{\boldsymbol{\phi}} \quad \tfrac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 \right), \qquad \text{subject to} \quad \mathbf{x} = \operatorname{sign}(\boldsymbol{\phi}), \tag{3.2}$$

where $\boldsymbol{\phi} \in \mathbb{R}^N$ is an auxiliary variable, and $\operatorname{sign}(\cdot)$ denotes the elementwise signum function. In our analysis, we consider the signum function such that $\operatorname{sign}(0) = 0$. The Lagrangian for this problem is defined as

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\phi}, \boldsymbol{\nu}) \triangleq \tfrac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 + \boldsymbol{\nu}^T \left(\mathbf{x} - \operatorname{sign}(\boldsymbol{\phi})\right). \tag{3.3}$$

The variable $\boldsymbol{\nu} \in \mathbb{R}^N$ is the Lagrange multiplier associated with the equality constraint $\mathbf{x} = \operatorname{sign}(\boldsymbol{\phi})$. We refer to this variable as the *dual* variable in the remainder of the chapter. We define the dual function $g(\boldsymbol{\nu})$ corresponding to the Lagrangian (3.3) as

$$g(\boldsymbol{\nu}) \triangleq \inf_{\mathbf{x}, \boldsymbol{\phi}} \mathcal{L}(\mathbf{x}, \boldsymbol{\phi}, \boldsymbol{\nu})$$

The primal problem (3.2) has a dual optimization problem expressed as

$$\boldsymbol{\nu}^\star \triangleq \underset{\boldsymbol{\nu}}{\operatorname{argmax}}\, g(\boldsymbol{\nu}).$$

As the dual function is *always* concave [152], this provides a way to define a convex formulation for the original problem. We should note two important aspects of duality theory here: *i)* we are not guaranteed in general that maximizing $g(\boldsymbol{\nu})$ yields a solution to the primal problem; *ii)* the reformulation is only computationally useful if we can efficiently evaluate g. The conditions under which the dual problem yields a solution to the primal problem are known as Slater's conditions [177] and are difficult to check in general unless the primal problem is convex. We will later show, by example, that the dual problem *does not always* solve the primal problem. Classifying under which conditions we *can* solve the primary problem via its dual is beyond the scope of this chapter.

It turns out we can obtain a closed-form expression for g. Before presenting the general form of g, we first present a detailed derivation for invertible $\mathbf{A}^T\mathbf{A}$ to provide some insight.

3.2.1 Invertible $\mathbf{A}^T\mathbf{A}$

The Lagrangian is *separable* in terms of $\mathbf{x}$ and $\boldsymbol{\phi}$. Hence, we can represent the dual function as the sum of two functions, $g_1(\boldsymbol{\nu})$ and $g_2(\boldsymbol{\nu})$.

$$g(\boldsymbol{\nu}) = \underbrace{\inf_{\mathbf{x}} \left\{ \tfrac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 + \boldsymbol{\nu}^T\mathbf{x} \right\}}_{g_1(\boldsymbol{\nu})} + \underbrace{\inf_{\boldsymbol{\phi}} \left\{ -\boldsymbol{\nu}^T \operatorname{sign}(\boldsymbol{\phi}) \right\}}_{g_2(\boldsymbol{\nu})} \tag{3.4}$$

First, we consider $g_1(\boldsymbol{\nu})$. Assuming $\mathbf{A}^T\mathbf{A}$ to be a non-singular matrix, we find the unique minimizer by setting the gradient to zero:

$$\mathbf{x}^\star = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\left(\mathbf{A}^T\mathbf{y} - \boldsymbol{\nu}\right). \tag{3.5}$$

Substituting $\mathbf{x}^\star$ back in the expression and re-arranging some terms we arrive at the following expression for g_1:

$$g_1(\boldsymbol{\nu}) = -\tfrac{1}{2}\|\mathbf{A}^T\mathbf{y} - \boldsymbol{\nu}\|^2_{(\mathbf{A}^T\mathbf{A})^{-1}} + \tfrac{1}{2}\mathbf{y}^T\mathbf{y}, \tag{3.6}$$

where $\|\mathbf{z}\|^2_{\mathbf{W}} = \mathbf{z}^T\mathbf{W}\mathbf{z}$ is a weighted ℓ_2-norm.

Next, we consider $g_2(\boldsymbol{\nu})$. Note that this function is *separable* in terms of ν_i

$$g_2(\boldsymbol{\nu}) = \inf_{\boldsymbol{\phi}} \left\{ -\sum_{i=1}^{N} \left(\nu_i \operatorname{sign}(\phi_i)\right) \right\}.$$

The function $-\nu \operatorname{sign}(\phi)$ achieves its smallest value for $\phi = \nu$ when $\nu \neq 0$. This solution is not unique of course, but that does not matter as we are only interested in the sign of ϕ. When $\nu_i = 0$ the function takes on value 0 regardless of the value of ϕ_i. We thus find

$$g_2(\boldsymbol{\nu}) = -\|\boldsymbol{\nu}\|_1. \tag{3.7}$$

Hence, the dual function for the Lagrangian in (3.3) takes the following explicit form:

$$g(\boldsymbol{\nu}) = -\tfrac{1}{2}\|\mathbf{A}^T\mathbf{y} - \boldsymbol{\nu}\|^2_{(\mathbf{A}^T\mathbf{A})^{-1}} - \|\boldsymbol{\nu}\|_1 + \tfrac{1}{2}\mathbf{y}^T\mathbf{y} \tag{3.8}$$

The maximizer to dual function (3.8) is found by solving the following minimization problem:

$$\boldsymbol{\nu}^\star = \underset{\boldsymbol{\nu}\in\mathbb{R}^N}{\operatorname{argmin}} \ \tfrac{1}{2}\|\boldsymbol{\nu} - \mathbf{A}^T\mathbf{y}\|^2_{(\mathbf{A}^T\mathbf{A})^{-1}} + \|\boldsymbol{\nu}\|_1. \tag{3.9}$$

This optimization problem lies in the class of *least absolute shrinkage and selection operator* (LASSO) [190]. The primal solution can be synthesized from the solution of the dual problem via $\mathbf{x}^\star = \operatorname{sign}(\boldsymbol{\phi}^\star) = \operatorname{sign}(\boldsymbol{\nu}^\star)$.

It is important to note at this point that the solution of the dual problem only determines those elements of the primal problem, x_i, for which $\nu_i \neq 0$. The remaining degrees of freedom in $\mathbf{x}$ need to be determined by alternative means.

The resulting solution is a *feasible* solution of the primal problem, but not necessarily the *optimal* one.

To gain some insight into the behaviour of the dual objective, consider a one-dimensional example with $\mathbf{A} = 1$:

$$x^\star = \underset{x\in\{-1,1\}}{\operatorname{argmin}} \tfrac{1}{2}(x-y)^2. \tag{3.10}$$

The solution to this problem is given by $x^\star = \operatorname{sign}(y)$. The corresponding dual problem is

$$\nu^\star = \underset{\nu\in\mathbb{R}}{\operatorname{argmin}} \tfrac{1}{2}(\nu-y)^2 + |\nu|, \tag{3.11}$$

the solution of which is given by $\nu^* = \max(|y|-1, 0)\operatorname{sign}(y)$. Hence, for $|y| > 1$, the solution of the dual problem yields the desired solution. For $|y| \leq 1$, however, the dual problem yields $\nu^\star = 0$ in which case the primal solution $x^* = 0$. We will see in Section 3.2.3 that when using certain iterative methods to solve the dual problem, the iterations will naturally approach the solution $\nu^\star = 0$ from the correct side, so that the sign of the approximate solution may still be useful.

3.2.2 Main results

We state the main results below. The proofs for these statements are provided in Section 3.6.

Proposition 3.2.1. *The dual objective of* (3.2) *for general* $\mathbf{A} \in \mathbb{R}^{m\times N}$ *is given by*

$$g(\boldsymbol{\nu}) = \begin{cases} -\frac{1}{2}\|\boldsymbol{\nu} - \mathbf{A}^T\mathbf{y}\|^2_{(\mathbf{A}^T\mathbf{A})^\dagger} - \|\boldsymbol{\nu}\|_1 + \frac{1}{2}\mathbf{y}^T\mathbf{y} & \boldsymbol{\nu} \in \mathcal{R}_{\mathbf{A}^T}, \\ -\infty & \textit{otherwise} \end{cases}$$

where † *denotes the pseudo-inverse and* $\mathcal{R}_{\mathbf{A}^T}$ *is the range of* $\mathbf{A}^T$*. This leads to the following optimization problem*

$$\boldsymbol{\nu}^\star = \underset{\boldsymbol{\nu}\in\mathcal{R}_{\mathbf{A}^T}}{\operatorname{argmin}} \tfrac{1}{2}\|\boldsymbol{\nu} - \mathbf{A}^T\mathbf{y}\|^2_{(\mathbf{A}^T\mathbf{A})^\dagger} + \|\boldsymbol{\nu}\|_1. \tag{3.12}$$

Remark. In case $m \geq N$ and $\mathbf{A}$ has full rank (*i.e.*, $\mathcal{R}_{\mathbf{A}^T} = \mathbb{R}^n$), $\mathbf{A}^T\mathbf{A}$ is invertible and the general form (3.12) simplifies to (3.9).

Corollary. *The minimization problem* (3.12) *can be restated as*

$$\boldsymbol{\mu}^\star = \underset{\boldsymbol{\mu}\in\mathbb{R}^m}{\operatorname{argmin}} \tfrac{1}{2}\|\mathbf{A}\mathbf{A}^\dagger(\boldsymbol{\mu} - \mathbf{y})\|^2 + \|\mathbf{A}^T\boldsymbol{\mu}\|_1, \tag{3.13}$$

and the primal solution is recovered through $\mathbf{x}^\star = \textit{sign}(\mathbf{A}^T\boldsymbol{\mu}^\star)$.

Remark. For $m \leq N$ and $\mathbf{A}$ full row rank, we have $\mathbf{A}\mathbf{A}^\dagger = \mathbf{I}$ and the formulation (3.13) simplifies to

$$\boldsymbol{\mu}^\star = \underset{\boldsymbol{\mu}\in\mathbb{R}^m}{\operatorname{argmin}} \quad \tfrac{1}{2}\|\boldsymbol{\mu} - \mathbf{y}\|^2 + \|\mathbf{A}^T\boldsymbol{\mu}\|_1. \tag{3.14}$$

This form implicitly handles the constraints on the search space of $\boldsymbol{\mu}$ in Proposition 3.2.1. It allows us to use the functional form for matrix $\mathbf{A}$ thereby reducing the storage and increasing the computational speed to find an optimal dual variable $\boldsymbol{\mu}^\star$.

Proposition 3.2.2. *The dual problem for a binary tomography problem with grey levels $u_0 < u_1$ is given by:*

$$\boldsymbol{\nu}^\star = \underset{\boldsymbol{\nu}\in\mathcal{R}_{\mathbf{A}^T}}{\operatorname{argmin}} \tfrac{1}{2}\|\boldsymbol{\nu} - \mathbf{A}^T\mathbf{y}\|^2_{(\mathbf{A}^T\mathbf{A})^\dagger} + p(\boldsymbol{\nu}), \tag{3.15}$$

where $p(\boldsymbol{\nu}) = \sum_i |u_0|\max(-\nu_i, 0) + |u_1|\max(\nu_i, 0)$ is an asymmetric one-norm. The primal solution is obtained using

$$\mathbf{x}^\star = u_0\mathbf{1} + (u_1 - u_0)H(\boldsymbol{\nu}^\star),$$

where $H(\cdot)$ denotes the Heaviside function.

We summarize the procedure in algorithm 5 for finding the optimal solution via solving the dual problem. In practical applications, the formulation in step 8 is very useful since the projection matrix $\mathbf{A}$ generally has a low rank, *i.e.*, $\mathbf{rank}(\mathbf{A}) < \min(m, N)$.

Algorithm 5 Dual problem for various cases

Input: $\mathbf{A} \in \mathbb{R}^{m\times N}$, $\mathbf{y} \in \mathbb{R}^m$
Output: $\mathbf{x}^\star \in \{-1, 0, 1\}^N$
1: **if** $\mathbf{rank}(\mathbf{A}) = \min(m, N)$ **then**
2: **if** $m > N$ **then**
3: $\boldsymbol{\nu}^\star \triangleq \operatorname{argmin}_{\boldsymbol{\nu}} \tfrac{1}{2}\|\boldsymbol{\nu} - \mathbf{A}^T\mathbf{y}\|^2_{(\mathbf{A}^T\mathbf{A})^{-1}} + \|\boldsymbol{\nu}\|_1$ **return** $\mathbf{x}^\star = \operatorname{sign}(\boldsymbol{\nu}^\star)$
4: **else**
5: $\boldsymbol{\nu}^\star \triangleq \operatorname{argmin}_{\boldsymbol{\nu}} \tfrac{1}{2}\|\boldsymbol{\nu} - \mathbf{y}\|^2 + \|\mathbf{A}^T\boldsymbol{\nu}\|_1$ **return** $\mathbf{x}^\star = \operatorname{sign}\left(\mathbf{A}^T\boldsymbol{\nu}^\star\right)$
6: **end if**
7: **else**
8: $\boldsymbol{\nu}^\star \triangleq \operatorname{argmin}_{\boldsymbol{\nu}} \tfrac{1}{2}\|\mathbf{A}\mathbf{A}^\dagger(\boldsymbol{\nu} - \mathbf{y})\|^2 + \|\mathbf{A}^T\boldsymbol{\nu}\|_1$ **return** $\mathbf{x}^\star = \operatorname{sign}\left(\mathbf{A}^T\boldsymbol{\nu}^\star\right)$
9: **end if**

Remark. The realistic tomographic data contains Poisson noise. In such a case, the binary tomography problem takes the constrained weighted least-squares form [72, 164]:

$$\mathbf{x}^\star = \underset{\mathbf{x}}{\operatorname{argmin}}\left(\inf_{\boldsymbol{\phi}} \tfrac{1}{2}\|\boldsymbol{y} - \mathbf{A}\mathbf{x}\|^2_{\boldsymbol{\Lambda}}\right), \quad \text{subject to} \quad \mathbf{x} = \operatorname{sign}(\boldsymbol{\phi}), \tag{3.16}$$

where $\boldsymbol{\Lambda} \in \mathbb{R}^{m\times m}$ is a diagonal matrix with elements $\Lambda_i > 0$ representing the least-squares weight per projection. The dual objective of (3.16) is given by

$$g(\boldsymbol{\nu}) = \begin{cases} -\frac{1}{2}\|\boldsymbol{\nu} - \mathbf{A}^T\boldsymbol{\Lambda}\mathbf{y}\|^2_{\mathbf{B}} - \|\boldsymbol{\nu}\|_1 + \frac{1}{2}\mathbf{y}^T\mathbf{y} & \boldsymbol{\nu} \in \mathcal{R}_{\mathbf{A}^T}, \\ -\infty & \text{otherwise}, \end{cases}$$

where $\mathbf{B} \triangleq \left(\mathbf{A}^T\boldsymbol{\Lambda}\mathbf{A}\right)^\dagger$. This leads to optimization problem

$$\nu^\star = \underset{\boldsymbol{\nu}\in\mathcal{R}_{\mathbf{A}^T}}{\operatorname{argmin}} \tfrac{1}{2}\|\boldsymbol{\nu} - \mathbf{A}^T\boldsymbol{\Lambda}\mathbf{y}\|^2_{(\mathbf{A}^T\boldsymbol{\Lambda}\mathbf{A})^\dagger} + \|\boldsymbol{\nu}\|_1. \tag{3.17}$$

If $\mathbf{rank}(\mathbf{A}) = m$ with $m \leq N$, the problem (3.17) reduces to

$$\mu^\star = \underset{\boldsymbol{\mu} \in \mathbb{R}^m}{\operatorname{argmin}} \tfrac{1}{2}\|\boldsymbol{\mu} - \boldsymbol{\Lambda}^{1/2}\mathbf{y}\|^2 + \|\mathbf{A}^T\boldsymbol{\Lambda}^{1/2}\boldsymbol{\mu}\|_1, \tag{3.18}$$

and the primal solution is recovered from $\mathbf{x}^\star = \operatorname{sign}(\mathbf{A}^T\boldsymbol{\Lambda}^{1/2}\boldsymbol{\mu}^\star)$.

3.2.3 Solving the dual problem

When $\mathbf{A}^T\mathbf{A}$ is invertible, the dual formulation (3.12) can be, in principle, solved[2] using a proximal gradient algorithm ([59, 21]):

$$\boldsymbol{\nu}_{k+1} \triangleq S_{L^{-1}}\left(\boldsymbol{\nu}_k - L^{-1}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\left(\mathbf{A}^T\mathbf{y} - \boldsymbol{\nu}_k\right)\right), \tag{3.19}$$

where $L = \|\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\|$ and the soft thresholding operator $S_\tau(\cdot) = \max(|\cdot| - \tau, 0)\operatorname{sign}(\cdot)$ is applied component-wise to its input. We can interpret this algorithm as minimizing subsequent approximations of the problem, as illustrated in Figure 3.1.

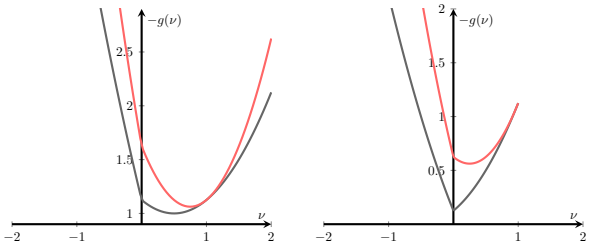


Figure 3.1: Plot of the dual function g (gray line) corresponding the the primal objective $(x-y)^2$ for $y = 1.5$ (left) and $y = 0.5$ (right) and its approximations (red line) at $x = 1$.

An interesting note is that, when starting from $\boldsymbol{\nu}_0 = \mathbf{0}$, the first iteration yields a thresholded version of $\mathbf{A}^\dagger\mathbf{y}$. As such, the proposed formulation is a natural extension of a naive segmentation approach and allows for segmentation in a data-consistent manner.

If $\mathbf{A}\mathbf{A}^T$ is invertible we have $\mathbf{A}\mathbf{A}^\dagger = \mathbf{I}$ and it seems more natural to solve (3.14) instead. Due to the appearance of $\mathbf{A}^T$ in the one-norm, it is no longer straightforward to apply a proximal gradient method. A possible strategy is to replace the one-norm with a smooth approximation of it, such as $|\cdot| = \sqrt{(\cdot)^2 + \epsilon}$. As illustrated in Figure 3.2, this will slightly shift the minimum of the problem. Since we are ultimately only using the sign of the solution, this may not be a problem. The resulting objective is smooth and can be solved using any gradient-descent algorithm.

We also note that splitting methods can be used to solve (3.14). For example, the alternating direction method of multipliers (ADMM) [33] and/or split-Bregman method [79]. Another class of method that can solve (3.14) are the primal-dual methods (e.g., Arrow-Hurwicz primal-dual algorithm [8], Chambolle-Pock

[2] In practice, computing the inverse of $\mathbf{A}^T\mathbf{A}$ for large-scale problems is difficult.

algorithm [49]). These methods rely on the proximal operators of functions and iterate towards finding the saddle point of the problem. If the proximal operators are simple, these are computationally faster than the splitting methods. Recently, proximal gradient methods have been developed to solve problems of the form (3.14) [82].

The dual problem (3.15) for binary tomography problem with grey levels $u_0 < u_1$ is also solved using proximal gradient method. We provide the proximal operator for an asymmetric one-norm in the following proposition.

Proposition 3.2.3. *The proximal operator for an asymmetric one-norm function*

$$p(\mathbf{x}) = \sum_{i=1}^{N} |u_0| \max(-x_i, 0) + |u_1| \max(x_i, 0)$$

with $u_0 < u_1$, is given by

$$\mathcal{P}_{p,\lambda}(\mathbf{z}) \triangleq \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \tfrac{1}{2}\|\mathbf{x} - \mathbf{z}\|^2 + \lambda p(\mathbf{x}) \right\} = \mathcal{S}_{\lambda u_0 < \lambda u_1}(\mathbf{z}),$$

where $\lambda > 0$, and $\mathcal{S}_{a<b}(\cdot)$ is an asymmetric soft-thresholding function

$$\mathcal{S}_{a<b}(t) = \begin{cases} t - |b| & t \geq |b| \\ 0 & -|a| < t < |b| \\ t + |a| & t \leq -|a| \end{cases}.$$

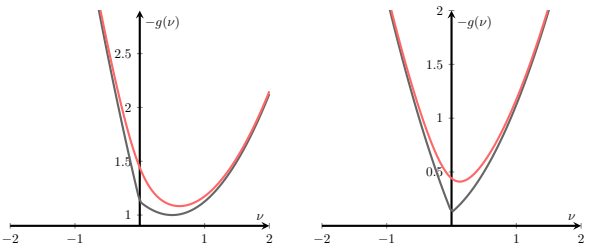


Figure 3.2: Plot of the dual function g (gray line) corresponding the the primal objective $(x - y)^2$ for $y = 1.5$ (left) and $y = 0.5$ (right) and its smooth approximation (red line) using $|\cdot| \approx \sqrt{(\cdot)^2 + \epsilon}$ with $\epsilon = 0.1$.

3.3 Numerical experiments - binary tomography

To illustrate the behavior of the dual approach, we consider the simple setting of reconstructing an $n \times n$ image from its sums along m lattice directions (here restricted to the horizontal, vertical and two diagonal directions, so $m \in [2, 4]$). For $m, n > 2$ the problem is known to be NP-hard. For small n we can simply enumerate all possible images, find all solutions in each case by a brute-force search and compare these to the solution obtained by the dual approach. To this

end, we solved the resulting dual problem (3.13) using the CVX package in Matlab [81]. This yields an approximate solution, and we set elements in the numerically computed dual solution smaller than 10^{-9} to zero. We then compare the obtained primal solution, which has values $-1, 0, 1$ to the solution(s) of the binary tomography problem. From performing these computations for $n = 2, 3, 4$ and $m = 2, 3, 4$ we conclude the following:

- If the problem has a unique solution, then the dual approach retrieves it.
- If the problem has multiple solutions, then the dual approach retrieves the intersection of all solutions. The remaining pixels in the dual solution are undetermined (have value zero).

An example is shown in Figure 3.3.

Table 3.1: Summary of complete enumeration experiments.

	n	**total**	**unique**	**multiple**
$m = 2$	2	16	14/14	2/2
	3	512	230/230	282/282
	4	65536	6902/6902	58541/58634*
$m = 3$	2	16	16/16	0/0
	3	512	496/496	16/16
	4	65536	54272/54272	10813/11264*
$m = 4$	2	16	16/16	0/0
	3	512	512/512	0/0
	4	65536	65024/65024	512/512

A summary of these results is presented in Table 3.1. The table shows the number of cases with a unique solution where the dual approach gave the correct solution and, in case of multiple solutions, the number of cases where the dual approach correctly determined the intersection of all solutions). In a few instances with multiple solutions, CVX failed to provide an accurate solution (denoted with * in the table).

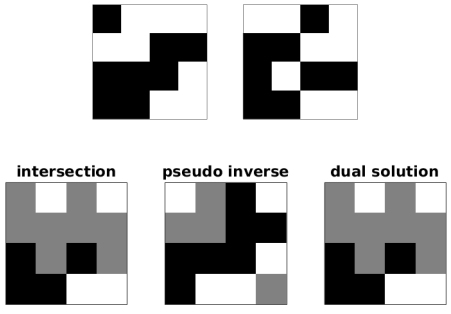


Figure 3.3: Example for $n = 4$ and $m = 3$ (using directons $(0, 1)$, $(1, 0)$ and $(1, 1)$). Two images with the same projections are shown in the top row while the intersection and the results obtained by the pseudo-inverse and the dual problem are shown in the bottom row.

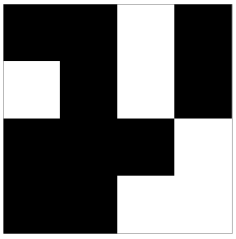


Figure 3.4: Example of a 4×4 binary image that is not *h,v,d*-convex but does permit a unique solution.

Based on these experiments, we conjecture that there is a subclass of the described binary tomography problem that is not NP-hard. We should note that, as n grows the number of cases that have a unique solution grows smaller unless m grows accordingly. It has been established that binary images that are *h,v,d*-convex[3] can be reconstructed from their horizontal, vertical and diagonal projections in polynomial time [14, 13]. However, we can construct images that are *not h,v,d*-convex but still permit a unique solution, see Figure 3.4. Such images are also retrieved using our dual approach.

3.4 Numerical experiments - X-ray tomography

In this section, we present numerical results for limited-angle X-ray tomography on a few numerical phantoms and an experimental X-ray dataset. First, we describe the phantoms and the performance measures used to compare our proposed dual approach (abbreviated as DP) to several state-of-the-art iterative reconstruction techniques. We conclude this section with results on an experimental dataset. All experiments are performed using Matlab in conjunction with the ASTRA toolbox [195].

3.4.1 Phantoms

For the synthetic tests, we consider four phantoms shown in Figure 3.5. All the phantoms are binary images of size 128 $\times$ 128 pixels. The grey levels are $u_0 = 0$ and $u_1 = 1$. The detector has 128 pixels, and the distance between the adjacent detectors is the same as the pixel size of the phantoms. We consider a parallel beam geometry for the acquisition of the tomographic data in all the simulation experiments.

3.4.2 Tests

We perform three different tests to check the robustness of the proposed method. First, we consider the problem of sparse projection data. For all the phantoms, we

[3]For *h,v,d*-convexity one uses the usual definition of convexity of a set but considers only line segments in the horizontal, vertical and diagonal directions.

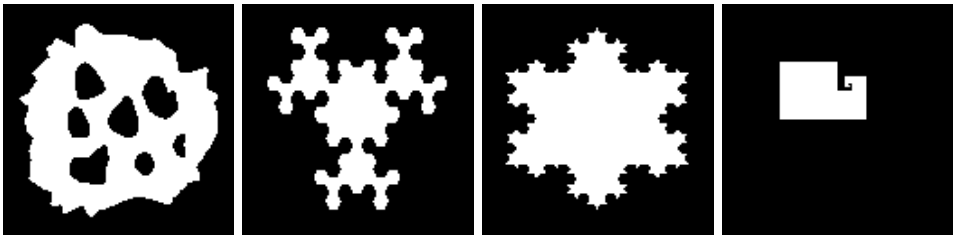


Figure 3.5: Phantom images used in the simulation experiments. (a) Phantom 1, (b) Phantom 2, (c) Phantom 3, (d) Phantom 4.

first start with 45 projections at angles ranging from 0 to π and subsequently reduce the number of projections. This setup is also known as sparse sampling, where the aim is to reduce the scan time by decreasing the number of angles.

Next, we consider a limited angle scenario. Such a situation usually arises in practice due to the limitations of the setup. For the test, we acquire projections in the range $[0, \theta_{\max}]$ for $\theta_{\max} \in \{5\pi/6, 2\pi/3, 7\pi/12, \pi/2\}$. Reconstruction of limited angle data is known to lead to so-called streak artifacts in the reconstructed image. Strategies to mitigate these streak artifacts include the use of regularization methods with some prior information. We will experiment how the discrete tomography can lead to the removal of these artifacts.

Finally, we test the performance of the proposed method in the presence of noise. We consider an additive Gaussian noise in these experiments. We measure the performance of our approach for tomographic data with a signal-to-noise ratio (SNR) of $\{10, 20, 30, 50\}$ dB.

To avoid *inverse crime* [107] in all the test scenarios, we generate data using strip kernel and use Joseph kernel for modeling.

3.4.3 Comparison with other reconstruction methods

There exist a vast amount of reconstruction methods for tomography. Here, consider the following three:

LSQR : Least squares QR method described in [144]. We perform a total of 1000 iterations with a tolerance of 10^{-6}. We segment the resulting reconstruction using Otsu's thresholding algorithm [143].

TV : The total-variation method leads to an optimization problem described below:

$$\min_{\mathbf{x}\in\mathbb{R}^N} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 + \lambda\|\mathbf{D}\mathbf{x}\|_1,$$

where λ is a corresponding regularization parameter, and matrix $\mathbf{D}$ captures the discrete gradient in both directions. We use the Chambolle-Pock method [49] to solve the above optimization problem with non-negativity constraints on the pixel values. For each case, we perform iterations till the relative duality gap reaches a tolerance value of 10^{-4}. To avoid slow convergence, we scale the

matrices $\mathbf{A}$ and $\mathbf{D}$ to have unit matrix norm. The regularization parameter λ is selected using Morozov's discrepancy principle using the noise level defined as $\|\mathbf{y} - \mathbf{A}\mathbf{x}_{\text{true}}\|$ [4]. Finally, we segment the TV-reconstructed image using Otsu's thresholding algorithm.

DART : We use the method described in [18] for DART on the above binary images. The gray values are taken to be the same as true grayvalues. We perform 20 Algebraic Reconstruction Method (ARM) iterations initially before performing 40 DART iterations. In each DART iteration, we do 3 algebraic reconstruction iterations. We use the segmented image as a result of the DART iterations to perform the further analysis of the method.

DP : We solve the dual formulation (3.14) with a smooth approximation of the ℓ_1-norm (as discussed in Section 3.2) using the L-BFGS method [123] with a maximum of 500 iterations.

3.4.4 Performance measures

In order to evaluate the performance of reconstruction methods, we use the following two criteria.

RMS The root-mean-square (RMS) error

$$\text{RMS} \triangleq \|\mathbf{A}\mathbf{x}^\star - \mathbf{y}\|,$$

measures how well the forward-projected reconstructed image matches the projection data. This measure is useful in practice, as it does not require knowledge of the ground truth. An RMS value close to the noise level of the data is considered as a good reconstruction.

JI The Jaccard index $\text{JI} \triangleq 1 - \sum_{i=1}^{N}(\alpha_i + \beta_i)/N$ measures the similarity between the reconstructed image ($\mathbf{x}^\star$) and the ground truth ($\mathbf{x}^{\text{true}}$) in a discrete sense. The parameters α and β represent missing and over-estimated pixels respectively and are given by

$$\alpha_i \triangleq (x_i^\star = u_0) \times \left(x_i^{\text{true}} = u_1\right),$$
$$\beta_i \triangleq (x_i^\star = u_1) \times \left(x_i^{\text{true}} = u_0\right).$$

The blue and red dots denote the missing and the overestimated pixels in Figure 3.7, 3.8, 3.9, 3.10, 3.11. If JI has high value (close to 1), the reconstruction is considered good. Although this measure is not readily applicable to real datasets, it is a handy measure to compare the various reconstruction methods on synthetic examples.

3.4.5 Experimental data setup

We use the experimental X-ray projection data of a carved cheese slice[37]. Figure 3.6 shows a high-resolution filtered back-projection reconstruction of the data. The cheese contains the letters C and T and the object is (approximately)

[4] Since the data is generated from a different modeling kernel, we compute noise level from the mismatch of the projection of true image with the data. The discrepancy principle then selects the parameter which fits data close to such noise level.

binary with two grey levels corresponding to calcium-containing organic compounds of cheese and air. The dataset consists of projection data with three different resolutions (128×128, 256×256, 512×512) and the corresponding projection matrix modeling the linear operation of X-ray transform.

We perform two sets of experiments: (1) Sparse sampling with 15 angles ranging from 0 to 2π and (2) limited-angle using 15 projections from 0 to $\pi/2$.

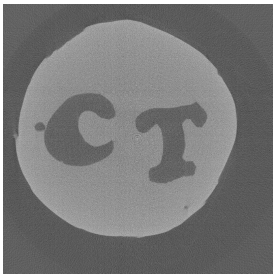


Figure 3.6: The high resolution (2000×2000 pixels) filtered back-projection reconstruction of the carved cheese from 360 projections from 0 to 2π.

3.4.6 Sparse projections test

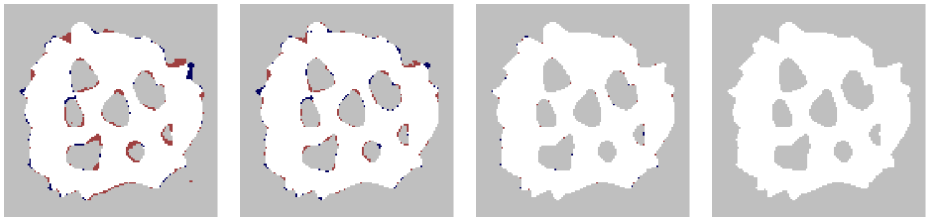
Figure 3.7 presents the reconstruction results from various methods for phantoms 1. The tomographic data are generated for ten equidistant projection angles from 0 to $\pi/2$. The reconstruction results show the difference between the reconstructed image and the ground truth. It is evident that, compared to the other methods, the proposed method (DP) reconstructions are very close to the ground truth. The results from LSQR are the worst as it does not incorporate any prior information about the model. The TV method also leads to artifacts as it includes partial information about the model. The DART and DP are very close to each other. For all the phantoms, we tabulate the data misfit (RMS) and Jaccard index (JI) in Table 3.2.

We also show reconstructions with the proposed approach for a varying number of projection angles in Figure 3.8. We note that the problem becomes harder to solve as the number of projections gets smaller. Hence, we may also expect the reconstruction to become poorer. We see that the proposed approach can reconstruct almost correctly with as few as ten projection angles.

3.4.7 Limited angle test

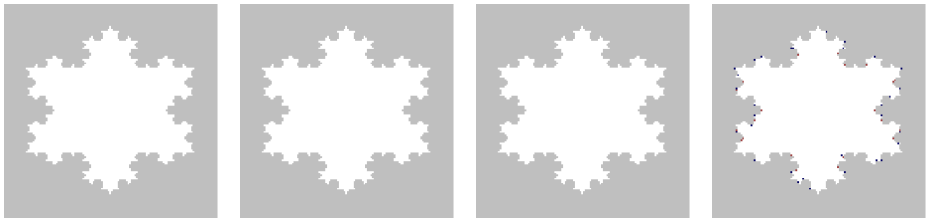
Figure 3.9 shows the results of phantom 2 with various reconstruction methods for limited angle tomography (10 equispaced angles in the range 0 to $\pi/2$). It is visible that the reconstructions from DART and the proposed method are very close to the true images of the phantoms. The values for data misfit and Jaccard index for all the tests with each of the synthetic phantoms are tabulated in Table 3.3.

We also look at how the reconstructions with the proposed method vary with limiting the angle (see Figure 3.10). As the angle gets limited, the reconstruction



(a) LSQR (b) TV (c) DART (d) DP

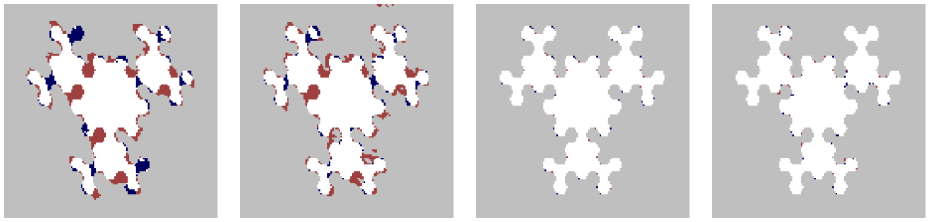
Figure 3.7: Limited projection test I for Phantom 1. Performance of various reconstruction methods with 10 projection angles from 0 to $\pi/2$.



(a) 45 (b) 20 (c) 10 (d) 5

Figure 3.8: Limited projection test II for Phantom 3. Performance of proposed method vs number of projections.

problem gets difficult. The proposed method can reconstruct almost correctly with angle limited to $\pi/2$.



(a) LSQR (b) TV (c) DART (d) DP

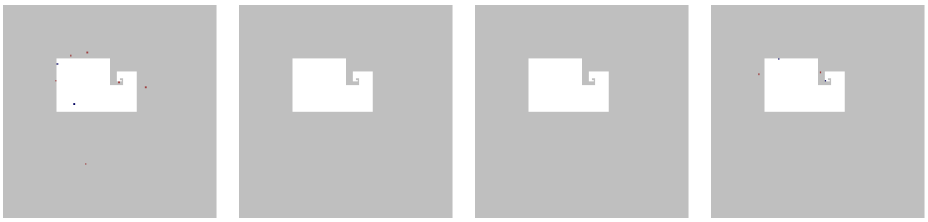
Figure 3.9: Limited angle test I for Phantom 2. Performance of various reconstruction methods with 10 projection angles from 0 to $\frac{\pi}{2}$.

3.4.8 Noisy projection test

This test aims to check the sensitivity of the proposed method to noise in the data. We perform four experiments with varying levels of Poisson noise in the data. In particular, we use expected incident photon counts $I_0 = \{10^6, 10^4, 10^3, 10^2\}$. Note

Table 3.2: Limited projection test performance measures

Test	Phantom	LSQR		TV		DART		DP	
		RMS	JI	RMS	JI	RMS	JI	RMS	JI
45	P1	31.4	99.7	19.2	99.9	31.5	99.7	17.6	100
	P2	26.2	99.6	12	100	36.4	99	12	100
	P3	48.1	99	24.5	99.8	52.4	98.8	16.4	100
	P4	11.6	99.8	5	100	13.6	99.6	5	100
20	P1	54.3	97.8	34.5	98.9	24.1	99.5	12	100
	P2	60.5	95.8	12.8	99.9	17	99.6	8.9	100
	P3	54.2	97.6	27	99.4	40.2	98.9	11.1	100
	P4	20.8	99.3	3.4	100	6.8	99.8	3.4	100
10	P1	122.3	93.1	62.4	96.2	22.4	99.1	8.3	99.9
	P2	254.4	76	81	91.1	17	99	11.2	99.7
	P3	73.8	94.9	38.6	98	28.7	99	7.4	100
	P4	42.7	97	5.4	99.9	5.9	99.7	2.4	00
5	P1	269.7	82.4	77.5	87	23.7	97.3	52.8	90.7
	P2	370.8	63.2	158.5	71.3	38.8	76.1	72.8	73.9
	P3	80	90.9	64.9	93.4	22.5	98.5	21.5	97.6
	P4	79.7	86.4	22.9	95.8	4.7	99.6	1.7	100



(a) $5\pi/6$ (b) $2\pi/3$ (c) $7\pi/12$ (d) $\pi/2$

Figure 3.10: Limited angle test II for Phantom 4. Performance of proposed method vs maximum angle.

that the lower incident photon counts lead to low SNR. Figure 3.11 shows the results on phantoms 1 and 2 for increasing noise level. We see that the reconstruction is stable against a moderate amount of noise and degrades gradually as the noise level increases.

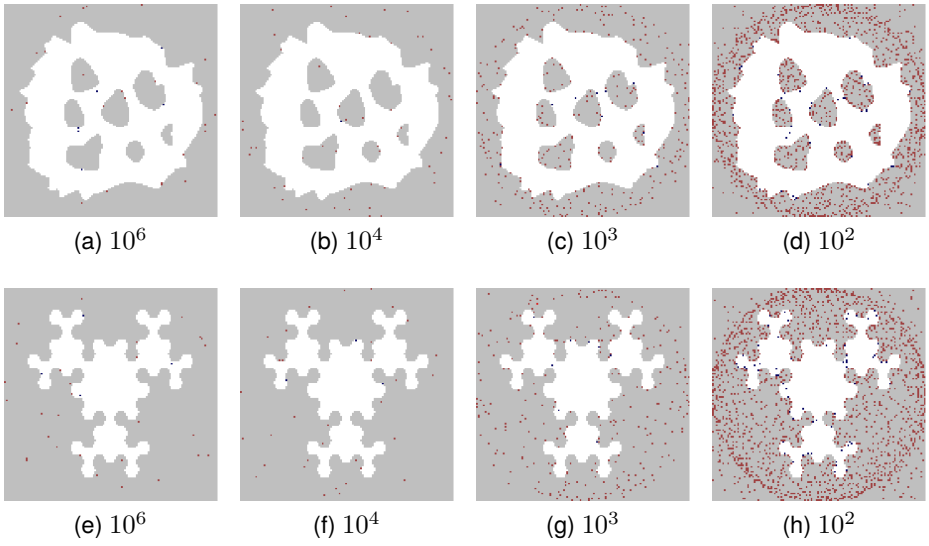
3.4.9 Real data test

We look at the results of reconstructions from the proposed method for two sets of experiments at various resolutions and compare them with the reconstructions from LSQR and TV. Since the ground truth image is not available, we compare these reconstructions visually.

In order to apply DP, we first need to estimate the grey values of the object. The

Table 3.3: Limited angle test performance measures

Test	Phantom	LSQR		TV		DART		DP	
		RMS	JI	RMS	JI	RMS	JI	RMS	JI
$5\pi/6$	P1	158.3	91.6	101.9	93.9	23.6	99.1	5.4	100
	P2	215.9	79.5	79.6	91.3	16.4	99.2	5	100
	P3	82.4	94.5	53.3	96.7	28.5	99	5.3	100
	P4	54.9	95.6	4.7	99.9	5.3	99.7	9.5	99.2
$2\pi/3$	P1	199.7	88.4	141.7	91.5	18.7	99.3	6.7	100
	P2	189.3	77.8	146.5	83	22	98	13.3	99.3
	P3	96.2	92.9	70.5	95	27.4	98.6	3.4	100
	P4	68.2	91.7	9.1	99.3	7.1	99.5	0.8	100
$7\pi/12$	P1	213.2	87	164.6	89.7	20.1	99.2	7.7	99.9
	P2	231.8	76.4	182.1	81.2	21.7	98	17	98.5
	P3	105.5	92.3	79.9	94.3	28.3	98.5	3.7	100
	P4	84.5	89.7	8.1	99.4	7.6	99.5	0.9	100
$\pi/2$	P1	258.6	84.9	293	85.2	19.1	99.3	18	99.2
	P2	205.3	75.3	192.5	80.9	22.7	98.2	18.2	98.5
	P3	127.1	90.5	99.2	93.5	31	98.3	4	100
	P4	128.2	82.9	27.1	96.7	7.8	0.99.5	7.2	99.5



(a) 10^6 (b) 10^4 (c) 10^3 (d) 10^2

(e) 10^6 (f) 10^4 (g) 10^3 (h) 10^2

Figure 3.11: Noisy Projection test on phantoms 1 and 2. Performance of proposed method vs expected total incident photon counts.

object, a thin slice of cheese, consists of two materials; the organic compound of the cheese, which is we assume to be homogeneous, and air. For air, the grey value

is zero. We estimate the grey value of the organic compound of cheese from the histogram of an FBP reconstruction provided with the data. Figure 3.12 represents the histogram. We obtain a value of 0.00696 for this compound.

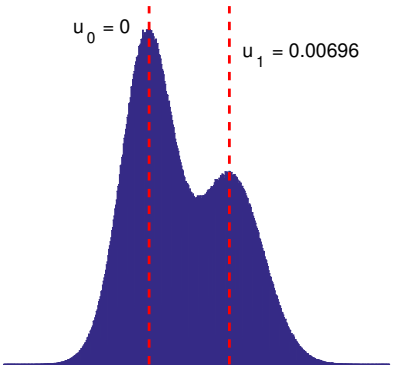


Figure 3.12: Histogram of filtered backprojection image of the carved cheese.

We first consider the reconstructions from sparse angular sampling (15 projections spanning from 0 to 2π). The tests are performed on two different resolutions: 128×128, and 512×512. Figure 3.13 presents the results of the reconstructions with LSQR, TV, and DP for these resolutions. The DP reconstruction is discrete and correctly identifies the letters C and T with also a little hole at the left side of C. Although LSQR reconstruction is poor for 128×128, it improves with the resolution. We still see the mild streak artifacts in these reconstructions. The TV reconstruction removes these streak artifacts but fails to identify the homogeneous cheese slice correctly.

In the second test, we limit the projection angles to $0 - \pi/2$. Figure 3.14 shows the results of the reconstructions from LSQR, TV, and DP for two different resolutions. We see that the reconstructions improve with increments in the resolution. LSQR reconstructions have severe streak artifacts, which are the characteristics of the limited data tomography. TV and DP reconstructions do not possess these artifacts. TV reconstruction can capture the shape of the cheese, but it blurs out the carved parts C and T. DP reconstructs the shape of cheese quite accurately and has C and T are also identified.

3.5 Conclusion

We presented a novel convex formulation for binary tomography. The problem is primarily a generalized LASSO problem that can be solved efficiently when the system matrix has full row rank or full column rank. Solving the dual problem is not guaranteed to give the optimal solution, but can at least be used to construct a feasible solution. In a complete enumeration of small binary test cases (images of $n \times n$ pixels for $n = 2, 3, 4$) we observed that if the problem has a unique solution, then the proposed dual approach finds it. In case the problem has multiple solutions, the dual approach finds the part that is common in all solutions. Based on these experiments we conjecture that this holds in the general case (beyond the

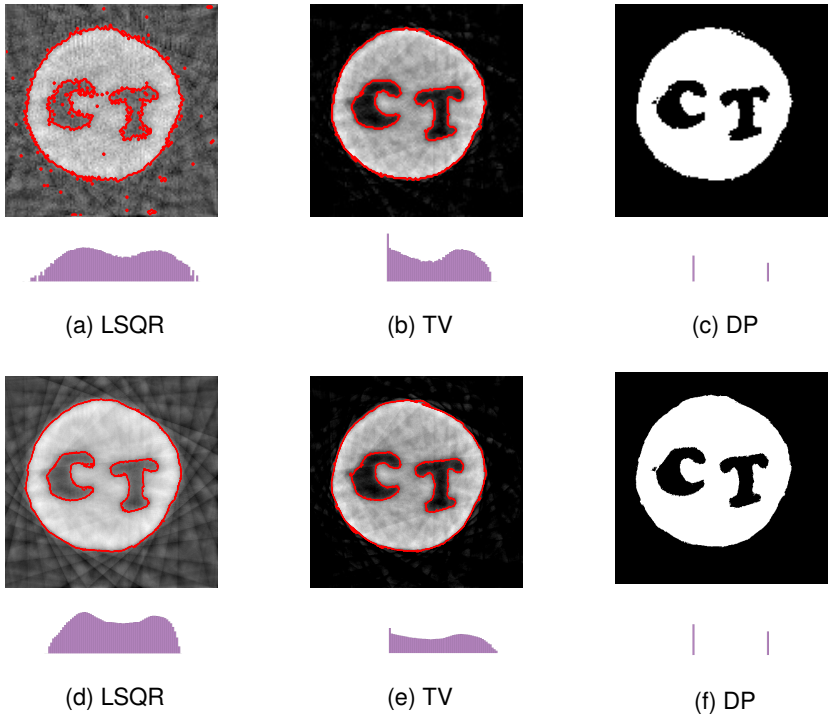
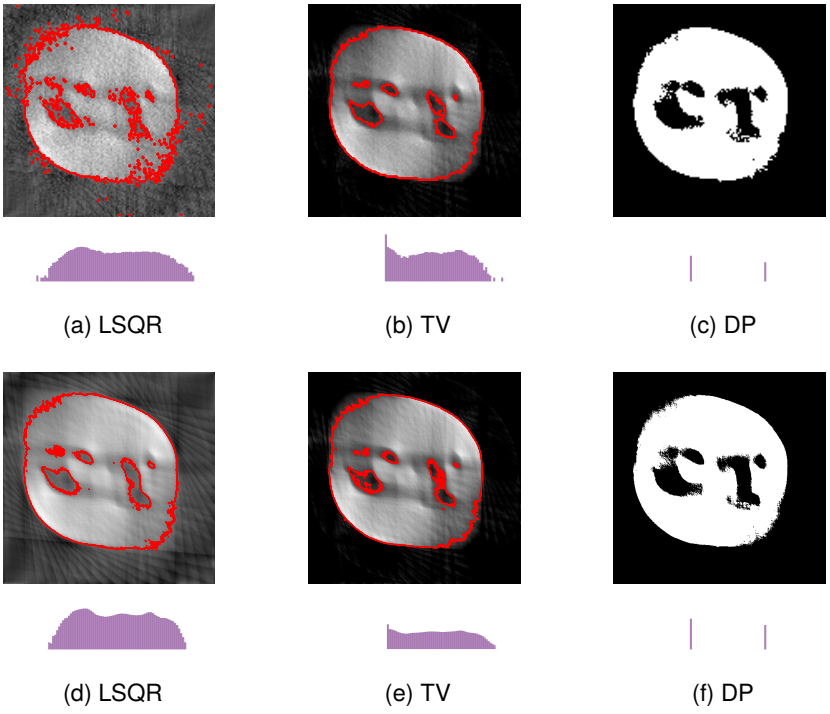


Figure 3.13: Real Data Test I - Sparse projection tomography. Performance of various methods with different resolutions. (a)-(c) correspond to 128×128 pixels. (d)-(f) correspond to 512×512 pixels. The figure below each image denotes the histogram. The red contours represent the thresholded image.

small test images). Of course, verifying beforehand if the problem has a unique solution may not be possible.

We test the proposed method on numerical phantoms and real data, showing that the method compares favorably to some of the state-of-the-art reconstruction techniques (Total Variation, DART). The proposed method is also reasonably stable against a moderate amount of noise.

We currently assume the grey levels are known apriori. Extension to multiple (i.e., more than 2) unknown grey levels is possible in the same framework but will be left for future work. To make the method more robust against noise additional regularization may be added.



(a) LSQR (b) TV (c) DP

(d) LSQR (e) TV (f) DP

Figure 3.14: Real Data Test II - Limited angle tomography. Performance of various methods with different resolutions. (a)-(c) correspond to 128×128 pixels. (d)-(f) correspond to 512×512 pixels. Figure below each image denote the histogram. The red contours represent the thresholded image.

3.6 Proofs

3.6.1 Proposition 3.2.1

Proof. In (3.4), the $g_1(\boldsymbol{\nu})$ has a closed-form expression for general $\mathbf{A}$. To see this, let us first denote

$$f(\mathbf{x}, \boldsymbol{\nu}) \triangleq \tfrac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 + \boldsymbol{\nu}^T\mathbf{x}. \tag{3.6.1}$$

We are interested in the infimum value of this function with respect to $\mathbf{x}$. To obtain this, we set the gradient of f with respect to $\mathbf{x}$ to zero

$$\nabla_{\mathbf{x}} f = \mathbf{A}^T(\mathbf{A}\mathbf{x} - \mathbf{y}) + \boldsymbol{\nu} = 0.$$

Since $\mathbf{A}$ is a general matrix, it may be rank-deficient. Hence, the optimal value $\mathbf{x}^\star$ only exists if $\boldsymbol{\nu}$ is in the range of $\mathbf{A}^T$ and it is given by

$$\mathbf{x}^\star = \left(\mathbf{A}^T\mathbf{A}\right)^\dagger \left(\mathbf{A}^T\mathbf{y} - \boldsymbol{\nu}\right).$$

Substituting this value in (3.6.1), we get the following:

$$\begin{aligned} g_1(\boldsymbol{\nu}) &= \inf_{\mathbf{x}} f(\mathbf{x}, \boldsymbol{\nu}) \\ &= \begin{cases} f(\mathbf{x}^\star, \boldsymbol{\nu}) & \boldsymbol{\nu} \in \mathcal{R}_{\mathbf{A}^T} \\ -\infty & \text{otherwise} \end{cases} \\ &= \begin{cases} -\frac{1}{2}\|\nu - \mathbf{A}^T\mathbf{y}\|_{(\mathbf{A}^T\mathbf{A})^\dagger} + \frac{1}{2}\mathbf{y}^T\mathbf{y} & \boldsymbol{\nu} \in \mathcal{R}_{\mathbf{A}^T} \\ -\infty & \text{otherwise,} \end{cases} \end{aligned}.$$

Now we return to the dual objective in equation (3.4). Substituting the explicit forms for $g_1(\boldsymbol{\nu})$ from above and $g_2(\boldsymbol{\nu})$ from equation (3.7), we get the expression for the dual objective:

$$g(\boldsymbol{\nu}) = \begin{cases} -\frac{1}{2}\|\boldsymbol{\nu} - \mathbf{A}^T\mathbf{y}\|^2_{(\mathbf{A}^T\mathbf{A})^\dagger} - \|\boldsymbol{\nu}\|_1 + \frac{1}{2}\mathbf{y}^T\mathbf{y} & \boldsymbol{\nu} \in \mathcal{R}_{\mathbf{A}^T}, \\ -\infty & \text{otherwise.} \end{cases}$$

The above dual objective leads to the following maximization problem with respect to the dual variable $\boldsymbol{\nu}$

$$\boldsymbol{\nu}^\star = \underset{\boldsymbol{\nu} \in \mathbb{R}^N}{\operatorname{argmax}} \quad g(\boldsymbol{\nu}).$$

As we are only interested in the maximum value of the dual objective, the space of $\boldsymbol{\nu}$ can be constrained to the range of $\mathbf{A}^T$. This is valid as the dual objective is $-\infty$ for the $\boldsymbol{\nu}$ outside the range of $\mathbf{A}^T$. Hence, the maximization problem reduced to the following minimization problem:

$$\boldsymbol{\nu}^\star = \underset{\boldsymbol{\nu} \in \mathcal{R}_{\mathbf{A}^T}}{\operatorname{argmin}} \quad \tfrac{1}{2}\|\boldsymbol{\nu} - \mathbf{A}^T\mathbf{y}\|^2_{(\mathbf{A}^T\mathbf{A})^\dagger} + \|\boldsymbol{\nu}\|_1.$$

□

3.6.2 Corollary 3.2.2

Proof. Since the search space for the dual variable $\boldsymbol{\nu}$ is constrained to the range of $\mathbf{A}^T$, we can express this variable as $\mathbf{A}^T\boldsymbol{\mu}$, where $\boldsymbol{\mu} \in \mathbb{R}^m$. Substituting $\boldsymbol{\nu} = \mathbf{A}^T\boldsymbol{\mu}$ in (3.12), we get

$$\boldsymbol{\mu}^\star = \underset{\boldsymbol{\mu} \in \mathbb{R}^m}{\operatorname{argmin}} \quad \tfrac{1}{2}\|\mathbf{A}^T(\boldsymbol{\mu} - \mathbf{y})\|^2_{(\mathbf{A}^T\mathbf{A})^\dagger} + \|\mathbf{A}^T\boldsymbol{\mu}\|_1. \tag{3.6.2}$$

Using the identities $(\mathbf{A}^T\mathbf{A})^\dagger\mathbf{A}^T = \mathbf{A}^\dagger$ and $\mathbf{A} = \mathbf{A}\mathbf{A}^\dagger\mathbf{A}$ [15], we can re-write the weighted norm $\|\mathbf{A}^T\mathbf{r}\|^2_{(\mathbf{A}^T\mathbf{A})^\dagger}$ as $\|AA^\dagger\mathbf{r}\|$. The dual problem (3.6.2) now reads

$$\boldsymbol{\mu}^\star = \underset{\boldsymbol{\mu} \in \mathbb{R}^m}{\operatorname{argmin}} \quad \tfrac{1}{2}\|\mathbf{A}\mathbf{A}^\dagger(\boldsymbol{\mu} - \mathbf{y})\|^2 + \|\mathbf{A}^T\boldsymbol{\mu}\|_1.$$

Correspondingly, the primal solution $\mathbf{x}^\star$ related to the dual optimal $\mu^\star$ is

$$\mathbf{x}^\star = \operatorname{sign}(\boldsymbol{\phi}^\star) = \operatorname{sign}(\boldsymbol{\nu}^\star) = \operatorname{sign}\left(\mathbf{A}^T\boldsymbol{\nu}^\star\right).$$

□

3.6.3 Proposition 3.2.2

Proof. The primal problem for binary tomography problem with levels $u_0 < u_1$ can be stated as:

$$\mathbf{x}^\star = \underset{\mathbf{x}}{\operatorname{argmin}} \left(\inf_{\boldsymbol{\phi}} \tfrac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 \right), \\ \text{subject to} \quad \mathbf{x} = u_0\mathbf{1} + (u_1 - u_0)\, H(\boldsymbol{\phi}),$$

where $H(\cdot)$ denotes the Heaviside function and $\boldsymbol{\phi}$ is an auxiliary variable. Such problem admits a Lagrangian

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\phi}, \boldsymbol{\nu}) = \tfrac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 + \boldsymbol{\nu}^T \left(\mathbf{x} - u_0\mathbf{1} - (u_1 - u_0)\, H(\boldsymbol{\phi})\right),$$

where $\boldsymbol{\nu} \in \mathbb{R}^N$ is a Lagrangian multiplier (also known as dual variable) corresponding to the equality constraint. This gives rise to a dual function

$$g(\boldsymbol{\nu}) = \underbrace{\inf_{\mathbf{x}} \left\{ \tfrac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 + \boldsymbol{\nu}^T\mathbf{x} \right\}}_{g_1(\boldsymbol{\nu})} + \underbrace{\inf_{\boldsymbol{\phi}} \left\{ -(u_1 - u_0)\, \boldsymbol{\nu}^T H(\boldsymbol{\phi}) \right\}}_{g_2(\boldsymbol{\nu})} - u_0\boldsymbol{\nu}^T\mathbf{1}.$$

Since we already know $g_1(\boldsymbol{\nu})$ (refer to equation (3.6)), we require the explicit form for $g_2(\boldsymbol{\nu})$. For its computation, we use the componentwise property of the Heaviside function to separate the infimum.

$$g_2(\boldsymbol{\nu}) = \sum_{i=1}^{N} \inf_{\phi_i} \left\{ -(u_1 - u_0)\, \nu_i H(\phi_i) \right\} = -\sum_{i=1}^{N} \sup_{\phi_i} \left\{ (u_1 - u_0)\, \nu_i H(\phi_i) \right\}$$

Since the range of Heaviside function is only two values, namely $\{0, 1\}$, we get the simple form for $g_2(\boldsymbol{\nu})$:

$$g_2(\boldsymbol{\nu}) = -\sum_{i=1}^{N} q(\nu_i) \quad \text{where} \quad q(\nu_i) = \begin{cases} (u_1 - u_0)\nu_i & \text{if } \nu_i > 0 \\ 0 & \text{otherwise} \end{cases} \\ = (u_1 - u_0)\max(\nu_i, 0).$$

This infimal value is attained at $\boldsymbol{\phi}^\star = H(\boldsymbol{\nu})$. Now the dual problem reads

$$\boldsymbol{\nu}^\star = \underset{\boldsymbol{\nu} \in \mathcal{R}_{\mathbf{A}^T}}{\operatorname{argmin}} \quad \left\{ \tfrac{1}{2}\|\boldsymbol{\nu} - \mathbf{A}^T\mathbf{y}\|^2_{(\mathbf{A}^T\mathbf{A})^\dagger} + \sum_i (u_1 - u_0)\max(\nu_i, 0) + u_0\boldsymbol{\nu}^T\mathbf{1} \right\}. \quad (3.6.3)$$

We note that the last two terms in the dual objective can be compactly represented by

$$p(\boldsymbol{\nu}) = \sum_i |u_0| \max(-\nu_i, 0) + |u_1| \max(\nu_i, 0),$$

where $p(\cdot)$ is known as an asymmetric one-norm. The optimal point of the problem (3.6.3) is denoted by $\boldsymbol{\nu}^\star$ and the corresponding primal optimal is retrieved using

$$\mathbf{x}^\star = u_0\mathbf{1} + (u_1 - u_0)H(\boldsymbol{\phi}^\star) = u_0\mathbf{1} + (u_1 - u_0)H(\boldsymbol{\nu}^\star).$$

□

3.6.4 Proposition 3.2.3

Proof. The minimization problem for the proximal operator of an asymmetric one-norm function $p(\cdot)$ reads

$$\min_{\mathbf{x}\in\mathbb{R}^N} \quad f(\mathbf{x}) = \tfrac{1}{2}\|\mathbf{x}-\mathbf{z}\|^2 + \lambda p(\mathbf{x}), \tag{3.6.4}$$

where $\lambda > 0$ is a parameter. Since the function is convex, we get the following from the first-order optimality condition [152]:

$$\mathbf{0} \in \partial f(\mathbf{x}^\star), \qquad \Longrightarrow \qquad \mathbf{0} \in \mathbf{x}^\star - \mathbf{z} + \lambda \partial p(\mathbf{x}^\star), \tag{3.6.5}$$

where $\mathbf{x}^\star$ is an optimal point of (3.6.4), and $\partial p(\mathbf{x})$ is a sub-differential of function $p(\cdot)$ at $\mathbf{x}$. This sub-differential is

$$\partial p(x_i) = \begin{cases} |u_1| & x_i > 0 \\ [-|u_0|, |u_1|] & x_i = 0 \\ -|u_0| & x_i < 0 \end{cases}.$$

Now coming back to the first-order optimality condition in (3.6.5), we get the explicit form for optimal solution $\mathbf{x}^\star$:

$$x_i^\star = \begin{cases} z_i - \lambda|u_1| & z_i \geq \lambda|u_1| \\ 0 & -\lambda|u_0| \leq z_i \leq \lambda|u_1| \\ z_i + \lambda|u_0| & z_i \leq -\lambda|u_0| \end{cases}.$$

We recognize this function as an asymmetric soft-thresholding function. □

CHAPTER 4

Partially discrete tomography

Abstract - *This chapter introduces a parametric level-set method for tomographic reconstruction of partially discrete images. Such images consist of a continuously varying background and an anomaly with a constant (known) grey-value. We represent the geometry of the anomaly using a level-set function, which we represent using radial basis functions. We pose the reconstruction problem as a bi-level optimization problem in terms of the background and coefficients for the level-set function. To constrain the background reconstruction, we impose smoothness through Tikhonov regularization. The bi-level optimization problem is solved in an alternating fashion; in each iteration we first reconstruct the background and consequently update the level-set function. We test our method on numerical phantoms and show that we can successfully reconstruct the geometry of the anomaly, even from limited data. On these phantoms, our method outperforms Total Variation reconstruction, DART and P-DART.*

This chapter is based on the following publications:

A Kadu, T van Leeuwen, and K J Batenburg. A parametric level-set method for partially discrete tomography. In *International Conference on Discrete Geometry for Computer Imagery*, pages 122–134. Springer, 2017.

" If people do not believe that mathematics is simple, it is only because they do not realize how complicated life is "
- John von Neumann

4.1 Introduction

The need to reconstruct (quantitative) images of an object from tomographic measurements appears in many applications. At the heart of many of these applications is a projection model based on the Radon transform. Characterizing the object under investigation by a function $x(\mathbf{r})$ with $\mathbf{r} \in \mathcal{D} = [0,1]^2$, tomographic measurements are modeled as

$$y_i = \int_{\mathcal{D}} u(\mathbf{r})\delta(s_i - \mathbf{n}(\theta_i)\cdot\mathbf{r})\,\mathrm{d}\mathbf{r}, \tag{4.1.1}$$

where $s_i \in [0,1]$ denotes the shift, $\theta_i \in [0,2\pi)$ denotes the angle and $\mathbf{n}(\theta) = (\cos\theta, \sin\theta)$. The goal is to retrieve x from a number, m, of such measurements for various shifts and directions.

If the shifts and angles are regularly and densely sampled, the transform can be inverted directly by Filtered back-projection or Fourier reconstruction [108]. A common approach for dealing with non-regularly sampled or missing data, is *algebraic reconstruction*. Here, we express x in terms of a basis

$$x(\mathbf{r}) = \sum_{j=1}^{n} u_j b(\mathbf{r} - \mathbf{r}_j),$$

where b are piece-wise polynomial basis functions and $\{\mathbf{r}_j\}_{j=1}^n$ is a regular (pixel) grid. This leads to a set of m linear equations in n unknowns

$$\mathbf{y} = A\mathbf{x},$$

with $a_{ij} = \int_{\mathcal{D}} b(\mathbf{r} - \mathbf{r}_j)\delta(s_i - \mathbf{n}(\theta_i)\cdot\mathbf{r})\,\mathrm{d}\mathbf{r}$. Due to noise in the data or errors in the projection model the system of equations is typically inconsistent, so a solution may not exist. Furthermore, there may be many solutions that fit the observations equally well when the system is underdetermined. A standard approach to mitigate these issues is to formulate a regularized least-squares problem

$$\min_{\mathbf{x}} \tfrac{1}{2}\|A\mathbf{x} - \mathbf{y}\|_2^2 + \tfrac{\lambda}{2}\|R\mathbf{x}\|_2^2,$$

where R is the regularization operator with parameter λ balancing the data-misfit and regularity of the solution. Such a formulation is popular mainly because very efficient algorithms exist for solving it. Depending on the choice of R, however, this formulation forces the solution to have certain properties which may not reflect the truth. For example, setting R to be the discrete Laplace operator will produce a smooth reconstruction, whereas setting R to be the identity matrix forces the individual coefficients x_i to be small. In many applications such quadratic regularization terms do not reflect the characteristics of the object we are reconstructing. For example, if we expect x to be piecewise constant, we could use a Total Variation regularization term $\|R\mathbf{x}\|_1$ where R is a discrete gradient operator

[173]. Recently, a lot of progress has been made in developing efficient algorithms for solving such non-smooth optimization problems [49]. If the object under investigation is known to consist of only two distinct materials, the regularization can be formulated in terms of a non-convex constraint $\mathbf{x} \in \{u_0, u_1\}^n$. The latter leads to a combinatorial optimization problem, solutions to which can be approximated using heuristic algorithms [18].

In this chapter, we consider tomographic reconstruction of *partially discrete* objects that consist of a region of constant density embedded in a continuously varying background. In this case, neither the quadratic, Total Variation nor non-convex constraints by themselves are suitable. We therefore propose the following parametrization

$$x(\mathbf{r}) = \begin{cases} u_1 & \text{if } \mathbf{r} \in \Omega, \\ u_0(\mathbf{r}) & \text{otherwise.} \end{cases}$$

The inverse problem now consists of finding $u_0(\mathbf{r})$, u_1 and the set Ω. We can subsequently apply suitable regularization to u_0 separately. To formulate a tractable optimization algorithm, we represent the set Ω using a level-set function $\phi(\mathbf{r})$ as follows:

$$\Omega = \{\mathbf{r} \,|\, \phi(\mathbf{r}) > 0\}.$$

In the following sections, we assume knowledge of u_1 and discuss how to formulate a variational problem to reconstruct Ω and u_0 based on a parametric level-set representation of Ω.

The outline of the chapter is as follows. In Section 4.2 we discuss the parametric level-set method and propose some practical heuristics for choosing various parameters that occur in the formulation. A joint background-anomaly reconstruction algorithm for partially discrete tomography is discussed in Section 4.3. The results on a few moderately complicated numerical phantoms are presented in Section 4.4. We provide some concluding remarks in Section 4.5.

4.2 Level-set methods

In terms of the level-set function, we can express u as

$$x(\mathbf{r}) = (\mathbf{1} - h(\phi(\mathbf{r})))u_0(\mathbf{r}) + h(\phi(\mathbf{r}))u_1,$$

where h is the Heaviside function.

Level-set methods have received much attention in geometric inverse problems, interface tracking, segmentation and shape optimization. In the classical level-set method, introduced by Sethian and Osher [142], the level-set is evolved according to the Hamilton-Jacobi equation

$$\frac{\partial \phi}{\partial t} + v|\nabla \phi| = 0,$$

where $\phi(\mathbf{r}, t)$ now denotes the level-set function as a time-dependent quantity for representing the shape and v denotes the normal velocity at the boundary of the

shape. In the inverse-problems setting, the velocity v is often derived from the gradient of the cost function with respect to the model parameter [40, 63]. There are various numerical issues associated with the numerical solution of level-set equation, e.g. reinitialization of the level-set. We refer the interested reader to a seminal paper in level-set methods [141] and its application to computational tomography [111].

Instead of taking this classical level-set approach, we employ a parametric level-set approach, first introduced by Aghasi et al [1]. In this method, the level-set function is parametrized using radial basis functions (RBF)

$$\phi(\mathbf{r}) = \sum_{j=1}^{n'} \alpha_j \Psi(\beta_j \|\mathbf{r} - \boldsymbol{\chi}_j\|_2),$$

where $\Psi(\cdot)$ is a radial basis function, $\{\alpha_j\}_{j=1}^{n'}$ and $\{\boldsymbol{\chi}_j\}_{j=1}^{n'}$ are the amplitudes and nodes respectively, and the parameters $\{\beta_j\}_{j=1}^{n'}$ control the widths. Introducing the kernel matrix $K(\boldsymbol{\chi}, \boldsymbol{\beta})$ with elements

$$k_{ij} = \Psi(\beta_j \|\mathbf{r}_i - \boldsymbol{\chi}_j\|_2),$$

we can now express $\mathbf{x}$ on the computational grid $\{\mathbf{r}_i\}_{i=1}^{n}$ as

$$\mathbf{x} = (\mathbf{1} - h(K(\boldsymbol{\chi}, \boldsymbol{\beta})\boldsymbol{\alpha})) \odot \mathbf{u}_0 + h(K(\boldsymbol{\chi}, \boldsymbol{\beta})\boldsymbol{\alpha})u_1, \tag{4.2.1}$$

where h is applied element-wise to the vector $K(\boldsymbol{\chi}, \boldsymbol{\beta})\boldsymbol{\alpha}$ and $\odot$ denotes the element-wise (Hadamard) product. By choosing the parameters $(\boldsymbol{\chi}, \boldsymbol{\beta}, \boldsymbol{\alpha})$ appropriately we can represent any (smooth) shape. To simplify matters and make the resulting optimization problem more tractable, we consider a fixed regular grid $\{\boldsymbol{\chi}_j\}_{j=1}^{n'}$ and a fixed width $\beta_j \equiv \beta$. In the following we choose β in accordance with the grid spacing $\Delta\chi$ as $\beta = 1/(\eta\Delta\chi)$, where η corresponds to the width of the RBF in grid points.

Example

To illustrate the representation of a shape using finitely many radial basis functions, we consider the green shape shown in Figure 4.1 (a). Starting from an initial guess (Figure 4.1 (a), red) we obtain the coefficients $\boldsymbol{\alpha}$ by solving a non-linear least-squares problem $\min_{\boldsymbol{\alpha}} \|h(K\boldsymbol{\alpha}) - \mathbf{y}\|_2^2$, where $\mathbf{y} \in \{0, 1\}^n$ indicates the true shape. This leads to the representation shown in Figure 4.1 (b). With $n' = 196$ RBFs, it is possible to reconstruct a smooth shape discretized on a grid with $n = 256 \times 256$ pixels.

Finally, the discretized reconstruction problem for determining the shape is now formulated as

$$\min_{\boldsymbol{\alpha}} \left\{ f(\boldsymbol{\alpha}) = \|A[(u_1\mathbf{1} - \mathbf{u}_0) \odot h_\epsilon(K\boldsymbol{\alpha})] - (\mathbf{y} - A\mathbf{u}_0)\|_2^2 \right\}, \tag{4.2.2}$$

where h_ϵ is a smooth approximation of the Heaviside function. The gradient and Gauss-Newton Hessian of $f(\boldsymbol{\alpha})$ are given by

$$\begin{aligned} \nabla f(\boldsymbol{\alpha}) &= K^T D_{\boldsymbol{\alpha}}^T A^T \mathbf{R}(\boldsymbol{\alpha}), \\ H_{GN}(f(\boldsymbol{\alpha})) &= K^T D_{\boldsymbol{\alpha}}^T A^T A D_{\boldsymbol{\alpha}} K, \end{aligned} \tag{4.2.3}$$

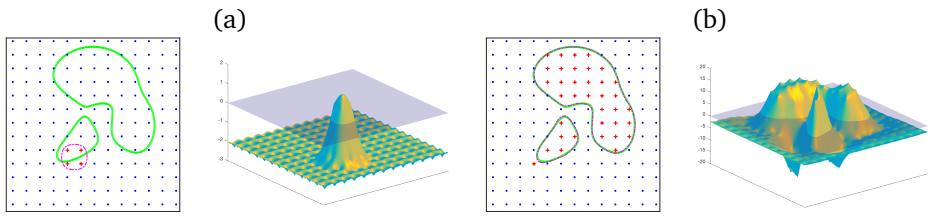


Figure 4.1: Any (sufficiently) smooth level-set can be reconstructed from radial basis functions. (a) The shape to be reconstructed is denoted in *green*. The initial shape (*dash-dotted* line) is generated by some positive RBF coefficients (denoted by *red pluses*) near the center and negative coefficients elsewhere (denoted by *blue dots*). Also shown is the corresponding initial level-set function. The reconstructed shape denoted with a *dash-dotted* line, the sign of the RBF-coefficients as well as the corresponding level-set function are shown in (b).

where the diagonal matrix and residual vectors are given by

$$D_{\boldsymbol{\alpha}} = \mathbf{diag}\left((u_1\mathbf{1} - \mathbf{u}_0) \odot h'_\epsilon(K\boldsymbol{\alpha})\right),$$
$$\mathbf{R}(\boldsymbol{\alpha}) = A[(u_1\mathbf{1} - \mathbf{u}_0) \odot h_\epsilon(K\boldsymbol{\alpha})] - (\mathbf{y} - A\mathbf{u}_0).$$

Using a Gauss-Newton method, the level-set parameters are updated as

$$\boldsymbol{\alpha}^{(k+1)} = \boldsymbol{\alpha}^{(k)} - \mu_k \left(H_{GN}(f(\boldsymbol{\alpha}^{(k)}))\right)^{-1} \nabla f(\boldsymbol{\alpha}^{(k)}), \tag{4.2.4}$$

where μ_k is a stepsize chosen to satisfy the weak Wolfe conditions [210] and $\boldsymbol{\alpha}^{(0)}$ is a given initial estimate of the shape. The weak Wolfe conditions consist of sufficient decrease and curvature conditions and ensure global convergence to a local minimum.

From Equation (4.2.3), it can be observed that the ability to update the level-set parameters depends on two main factors: 1) The difference between $\mathbf{u}_0$ and u_1, and 2) the derivative of the Heaviside function. Hence, the support and smoothness of h'_ϵ plays a crucial role in the sensitivity. More details on the choice of h_ϵ are discussed in Section 4.2.1.

Example

We demonstrate the parametric level-set method on a (binary) discrete tomography problem. We consider the model described in Figure 4.2(a). For a full-angle case ($0 \leq \theta \leq \pi$) with a large number of samples, Figure 4.2(c) shows that it is possible to accurately reconstruct a complex shape. The model is reconstructed by iteratively updating $\boldsymbol{\alpha}$ using Equation (4.2.4).

4.2.1 Approximation to Heaviside function

The update of the level-set function depends crucially on the choice of the Heaviside function. In equation (4.2.3) we see that h'_ϵ acts as a windowing function that controls which part of the level-set function is updated. The windowing

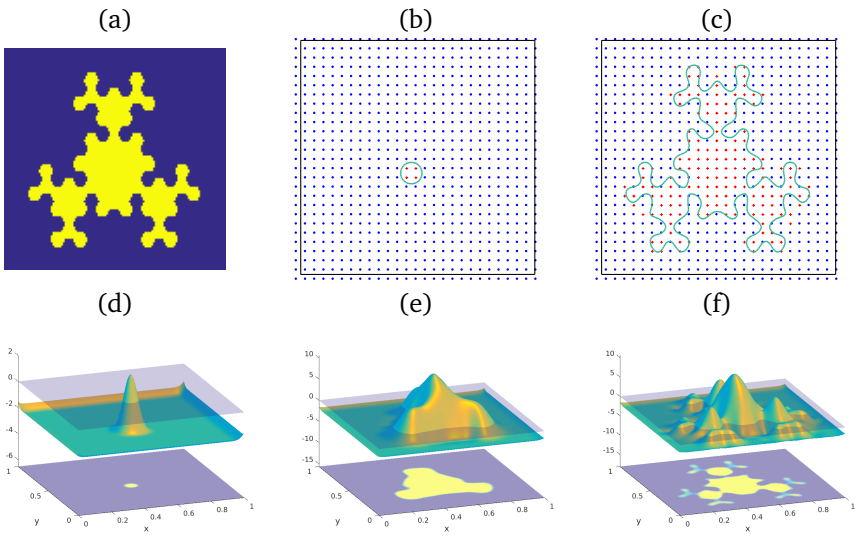


Figure 4.2: Parametric level-set method for Discrete tomography problem. (a) True model ($n = 256 \times 256$) (b) RBF grid ($n' = 27 \times 27$) with initial level-set denoted by *green line*, positive and negative RBFs are denoted by *red pluses* and *blue dots* respectively (c) Final level-set denoted by the *green line*, and the corresponding positive and negative RBFs (d) Initial level-set function (e) level-set function after 10 iterations (f) final level-set function after 25 iterations.

function should achieve the following: *(i)* limit the update to a small region around the boundary of the shape; *(ii)* have a uniform amplitude in the boundary region; and *(iii)* guarantee a minimum width of the boundary region. Failure to meet these requirements may result in poor updates for the level-set parameter $\boldsymbol{\alpha}$ and premature break-down of the algorithm.

Requirement (i) is easily fulfilled as any smooth approximation of the Heaviside will have a rapidly decaying derivative. To satisfy the second requirement we construct the Heaviside function by smoothing the piece-wise linear function $\frac{1}{2} + \frac{x}{2\epsilon}$ for $|x| \leq \epsilon$. This approximation is shown in Figure 4.3 alongside two common smooth approximations of the Heaviside. We now discuss how we can satisfy the third requirement, starting with a formal definition of the width of the level-set boundary layer as shown in Figure 4.3(c).

Definition 4.2.1. In accordance with the compact approximation of the Heaviside function with width ϵ, we define the minimum width of the level-set boundary layer as $\Delta = \min_{\mathbf{r}_0,\mathbf{r}_1} \|\mathbf{r}_0 - \mathbf{r}_1\|_2$ such that $\phi(\mathbf{r}_0) = 0$ and $|\phi(\mathbf{r}_1)| = \epsilon$.

Lemma 4.2.1. *For any smooth and compact approximation of the Heaviside function with finite width ϵ, the width of the level-set boundary layer, Δ, satisfies*

$$\Delta \geq \epsilon / \|\nabla \phi\|_\infty.$$

Proof. From a Taylor series expansion of $\phi(\mathbf{r})$ around $\mathbf{r}_0$ for which $\phi(\mathbf{r}_0) = 0$, we get

$$\phi(\mathbf{r}) = (\mathbf{r} - \mathbf{r}_0)^T \nabla\phi(\boldsymbol{\xi}),$$

with $\boldsymbol{\xi} = t\mathbf{r}_0 + (1-t)\mathbf{r}$ for some $t \in [0,1]$. This leads to

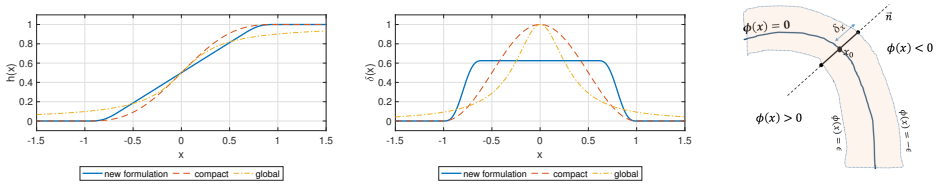
$$|\phi(\mathbf{r})| \leq \|\mathbf{r} - \mathbf{r}_0\|_2 \cdot \|\nabla\phi(\boldsymbol{\xi})\|_2 \leq \|\mathbf{r} - \mathbf{r}_0\|_2 \cdot \|\nabla\phi\|_\infty.$$

Choosing $\mathbf{r} = \mathbf{r}_1$ with $|\phi(\mathbf{r}_1)| = \epsilon$, we have $\|\mathbf{r}_1 - \mathbf{r}_0\|_2 \geq \epsilon / \|\nabla\phi\|_\infty$. Since this holds for all $\mathbf{r}_0, \mathbf{r}_1$ we obtain the desired result. □ □

To ensure a minimum width of the boundary layer, Lemma 4.2.1 suggest to choose ϵ proportional to $\|\nabla\phi\|_\infty$. For computational simplicity, we approximate this using upper and lower bounds [105] and set:

$$\epsilon = \kappa \left(\frac{\max(\phi(\mathbf{r})) - \min(\phi(\mathbf{r}))}{\Delta r} \right) = \kappa \left(\frac{\max(K\boldsymbol{\alpha}) - \min(K\boldsymbol{\alpha})}{\Delta r} \right), \tag{4.2.5}$$

where κ controls the width of level-set boundary in terms of the underlying computational grid. A small value of κ leads to a narrow boundary while big value leads a wide boundary.



(a) (b) (c)

Figure 4.3: New formulation for approximating the Heaviside function. The Heaviside functions (a) and corresponding Dirac-Delta functions (b) with $\epsilon = 1$. A global approximation is constructed from the inverse tangent function ($\frac{1}{2}(1 + \frac{2}{\pi}\arctan(\pi\frac{x}{\epsilon}))$), while compact one is composed of linear and sinusoid functions. (c) level-set boundary (*orange* region) around zero level-set denoted by *blue* line, n represents the normal direction at $\mathbf{x}_0$.

4.3 Joint reconstruction algorithm

Reconstructing both the shape and the background parameter can be cast as a bi-level optimization problem

$$\min_{\mathbf{u}_0, \boldsymbol{\alpha}} \left\{ f(\boldsymbol{\alpha}, \boldsymbol{u}_0) := \tfrac{1}{2}\|A[(\mathbf{1} - h(K\boldsymbol{\alpha}))\mathbf{u}_0 + h(K\boldsymbol{\alpha})u_1] - \mathbf{y}\|_2^2 + \tfrac{\lambda}{2}\|L\mathbf{u}_0\|_2^2 \right\}, \tag{4.3.1}$$

where L is of form $[L_x^T \quad L_y^T]^T$ where L_x and L_y is the second-order finite-difference operators in the x and y direction, respectively. This optimization problem is *separable*; it is quadratic in $\mathbf{u}_0$ and non-linear in $\boldsymbol{\alpha}$. In order to exploit

the fact that the problem has a closed-form solution in $\mathbf{u}_0$ for each $\boldsymbol{\alpha}$, we introduce a reduced objective

$$\overline{f}(\boldsymbol{\alpha}) = \min_{\mathbf{u}_0} f(\boldsymbol{\alpha}, \mathbf{u}_0).$$

The gradient and Hessian of this reduced objective are given by

$$\nabla \overline{f}(\boldsymbol{\alpha}) = \nabla_{\boldsymbol{\alpha}} f(\boldsymbol{\alpha}, \overline{\mathbf{u}}_0), \tag{4.3.2}$$

$$\nabla^2 \overline{f}(\boldsymbol{\alpha}) = \nabla^2_{\boldsymbol{\alpha}} f - \nabla^2_{\boldsymbol{\alpha},\mathbf{u}_0} f \left(\nabla^2_{\mathbf{u}_0} f\right)^{-1} \nabla^2_{\boldsymbol{\alpha},\mathbf{u}_0} f, \tag{4.3.3}$$

where $\overline{\mathbf{u}}_0 = \operatorname{argmin}_{\mathbf{u}_0} f(\boldsymbol{\alpha}, \mathbf{u}_0)$ [6].

Using a modified Gauss-Newton algorithm to find a minimizer of $\overline{f}$, it leads to the following alternating algorithm

$$\mathbf{u}_0^{(k+1)} = \arg\min_{\mathbf{u}_0} f(\boldsymbol{\alpha}^{(k)}, \mathbf{u}_0) \tag{4.3.4}$$

$$\boldsymbol{\alpha}^{(k+1)} = \boldsymbol{\alpha}^{(k)} - \mu_k \left(H_{GN}(f(\boldsymbol{\alpha}^{(k)}))\right)^{-1} \nabla_{\boldsymbol{\alpha}} f(\boldsymbol{\alpha}^{(k)}, \mathbf{u}_0^{(k+1)}), \tag{4.3.5}$$

where the expressions for the gradient and Gauss-Newton Hessian are given by (4.2.3). Convergence of this alternating approach to a local minimum of (4.3.1) is guaranteed as long as the step-length satisfies the strong Wolfe conditions [210].

The reconstruction algorithm based on this iterative scheme is presented in Algorithm 6. We use the LSQR method in step 3, with a pre-defined maximum number of iterations (typically 200) and a tolerance value. A trust-region method is applied to compute $\boldsymbol{\alpha}^{(k+1)}$ in step 4 restricting the conjugate gradient to *only* 10 iterations. We perform a total of $K = 50$ iterations to reconstruct the model.

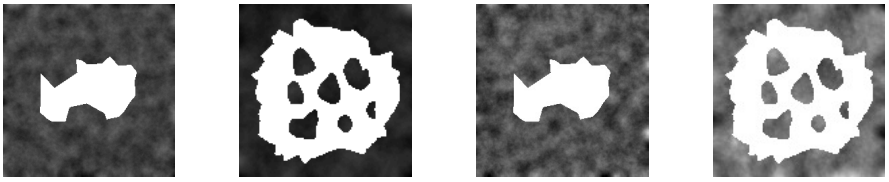
Algorithm 6 Joint Reconstruction Algorithm

Require: $\mathbf{p}$ - data, W - forward modeling operator, u_1 - anomaly property, A - RBF Kernel matrix, $\boldsymbol{\alpha}_0$ - initial RBF weights, κ - Heaviside parameters
Ensure: $\boldsymbol{\alpha}_K$ - final weights, $\mathbf{u}$ - corresponding model
1: **for** $k = 0$ to $K - 1$ **do**
2: compute Heaviside ϵ from Equation (4.2.5)
3: compute background parameter $\mathbf{u}_0^{(k+1)}$ by solving Equation (4.3.4)
4: compute level-set parameter $\boldsymbol{\alpha}^{(k+1)}$ from Equation (4.3.5)
5: **end for**
6: compute $\mathbf{u}$ from Equation (4.2.1).

4.4 Numerical experiments

The numerical experiments are performed on 4 phantoms shown in Figure 4.4. We scale the phantoms such that $u_1 = 1$. For the first two phantoms, the background varies from 0 to 0.5, while for the next two, it varies from 0 to 0.8. In order to

avoid the *inverse crime* [107], we use two different discretization schemes for Equation (4.1.1) (namely, the line kernel [27] for data generation, and the Joseph kernel [27] for forward modeling). We use the ASTRA toolbox to compute the forward and backward projections [27]. First, we show the results on the full-view data and later we compare various methods on a limited-angle case.



(a) Model A (b) Model B (c) Model C (d) Model D

Figure 4.4: Phantoms for Simulations. All the models have a resolution of 256×256 pixels.

For the parametric level-set method, we use Wendland compactly supported radial basis functions [105]. The RBF nodes are placed on a 5 times coarser grid than the model grid, with an extension of two points outside the model grid to avoid boundary effects. To constrain the initial level-set boundary to 4 grid-points, the Heaviside width parameter κ is set to be 0.01.

The level-set parameter $\boldsymbol{\alpha}$ is optimized using the *fminunc* package (trust-region algorithm) in MATLAB. A total of 50 iterations are performed for predicting the $\boldsymbol{\alpha}$, while 200 iterations are performed for predicting $\mathbf{u}_0(r)$ using LSQR at each step.

4.4.1 Regularization parameter selection

The reconstruction with the proposed algorithm is influenced by the regularization parameter λ (cf. (4.3.1)). In general, there are various strategies to choose this parameter, e.g., [189]. As our problem formulation is non-linear, many of these strategies do not apply. Instead we analyze the influence of the regularization parameter numerically as follows.

We define two measures (in the least-squares sense) to quantify the residuals: the data residual (DR), which determines the fit between the true data and reconstructed data, and the Jaccard index (JI), defined as a similarity coefficient between two sets, to capture the error in the reconstructed shape (on the scale of 0 to 100). In practice, one can only use the data residual to select the regularization parameter λ. It is evident from Figure 4.5 that there exists a sufficiently large region of λ for which the reconstructions are equally good. Moreover, this region is easily identifiable from the data residual plot for various λ values.

4.4.2 Full-view test

For the full-view case, the projection data is generated on a 256×256 grid with 256 detectors and 180 equidistant projections ($0 \leq \theta \leq \pi$). The Gaussian noise of 10 dB Signal-to-Noise ratio (SNR) is added to the data. The results on phantoms A, B, C and D with the full-view data are shown in Figure 4.6. The geometry of the

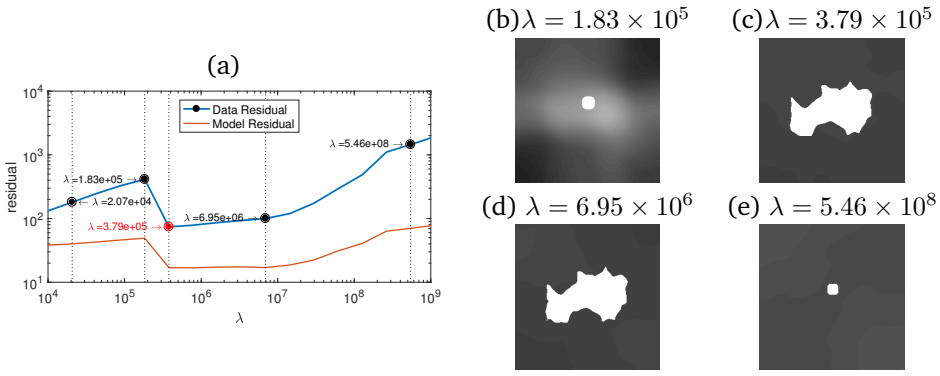


Figure 4.5: Variation of residuals with regularization parameter for Tikhonov. An appropriate region for choosing λ exists between 3.79×10^5 and 6.95×10^6. (a) behavior of DR and MR over λ for model A with noisy limited-angle data. (b),(c),(d),(e) show reconstructions for various λ values.

anomalies in all of these reconstructed models are very close to the ground truth, as is indicated by the Jaccard index shown above the figures. The background, though, has been smoothened out with the Tikhonov regularization.

Figure 4.6: Full-view test: Reconstructions with full-view data for the regularization parameter λ shown below it.

4.4.3 Limited-angle test

In this case, we generate synthetic data using *only* 5 projections, namely, $\theta = \{0, \pi/6, \pi/3, \pi/2, 2\pi/3\}$. We add Gaussian noise of 10 dB SNR to the synthetic data. To check the performance of the proposed method, we compare it to the Total Variation method [27], DART [18] and its modified version for partially discrete tomography, P-DART [153]. For Total Variation, we determine the shape from the final reconstruction via a simple segmentation step (tresholding). A total of 200 iterations were performed with the Total Variation method and the regularization parameter determined such that it optimally reconstructs the shape. In DART, the background is modeled using 20 discrete grey-values for model A and B, while 30 discrete grey-values for model C and D. The true model has been segmented per

mentioned grey-values to generate data for DART. 40 DART iterations were performed in each case. For P-DART, a total of 150 iterations were performed.

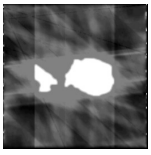
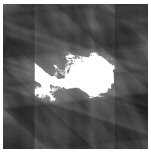
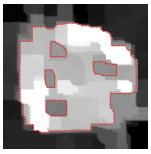
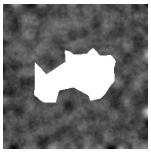
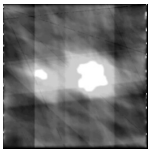
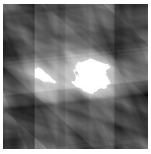
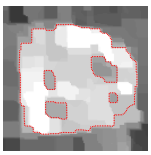
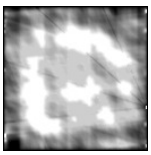
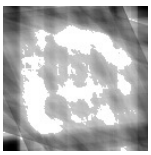
Phantom	Total Variation	DART	P-DART	Proposed Method
	$\lambda = 3.36$			$\lambda = 3.79 \times 10^5$
				
	DR = 59.91 JI = 92	DR = 100.36 JI = 49	**DR = 12.8** JI = 76	DR = 70.2 **JI = 96**
	$\lambda = 1.44$			$\lambda = 3.79 \times 10^5$
				
	DR = 39.39 JI = 78	DR = 32.8 JI = 40	**DR = 10.9** JI = 69	DR = 52.6 **JI = 91**
	$\lambda = 3.36$			$\lambda = 7.43 \times 10^5$
				
	DR = 70.35 JI = 92	DR = 115.28 JI = 19	**DR = 16.54** JI = 39	DR = 117.56 **JI = 95**
	$\lambda = 0.62$			$\lambda = 3.79 \times 10^5$
				
	DR = 21.52 JI = 80	DR = 78.30 JI = 25	**DR = 8.92** JI = 49	DR = 50.10 **JI = 87**

Figure 4.7: Reconstructions with noisy limited data. The first column shows the true models, while the last 4 columns show the reconstructions with various methods. Red dotted line shows the contour of the segmented model in Total Variation method. Different measures are also shown below each reconstructed model.

The results on limited-angle data are presented in Figure 4.7. The proposed method is able to capture most of the fine details (evident from the Jaccard Index) in the phantoms even with the very limited data with moderate noise. The P-DART method achieves the least amount of data residual in all the cases, but fails to capture the complete geometry of the anomaly. The Total variation method gives surprisingly good reconstructions of the shape. However, we obtained these results

by selecting the best over a large range of regularization parameters. The level-set method consistently gives the best reconstruction of the shape.

4.5 Conclusions and discussion

We discussed a parametric level-set method for partially discrete tomography. We model such objects as a constant-valued shape embedded in a continuously varying background. The shape is represented using a level-set function, which in turn is represented using radial basis functions. The reconstruction problem is posed as a bi-level optimization problem for the background and level-set parameters. This reconstruction problem can be efficiently solved using a variable projection approach, where the shape is iteratively updated. Each iteration requires a full reconstruction of the background. The algorithm includes some practical heuristics for choosing various parameters that are introduced as part of the parametric level-set method. Numerical experiments on a few numerical phantoms show that the proposed approach can outperform other popular methods for (partially) discrete tomography in terms of the reconstruction error. As the proposed algorithm requires repeated full reconstructions, it is currently an order of magnitude slower than the other methods. Future research is directed at making the method more efficient.

CHAPTER 5

Discrete waveform inversion

Abstract - *Seismic full-waveform inversion tries to estimate subsurface medium parameters from seismic data. Areas with subsurface salt bodies are of particular interest because they often have hydrocarbon reservoirs on their sides or underneath. Accurate reconstruction of their geometry is a challenge for current techniques. We present a parametric level-set method for the reconstruction of salt-bodies in seismic full-waveform inversion. We split the subsurface model in two parts: a background velocity model and the salt body with known velocity but undetermined shape. The salt geometry is represented by a level-set function that evolves during the inversion. We choose radial basis functions to represent the level-set function, leading to an optimization problem with a modest number of parameters. A common problem with level-set methods is to fine tune the width of the level-set boundary for optimal sensitivity. We propose a robust algorithm that dynamically adapts the width of the level-set boundary to ensure faster convergence. Tests on a suite of idealized salt geometries show that the proposed method is stable against a modest amount of noise. We also extend the method to joint inversion of both the background velocity model and the salt-geometry.*

This chapter is based on the following publications:

A Kadu, T van Leeuwen, and W A Mulder. Salt reconstruction in full-waveform inversion with a parametric level-set method. *IEEE Transactions on Computational Imaging*, 3(2):305–315, 2017.

" The shortest path between two truths in the real domain passes through the complex domain "
- Jacques Hadamard

5.1 Introduction

Seismic imaging attempts to obtain detailed images of the subsurface from seismic data. Such data are obtained by placing explosive or other types of sources on or near the surface and recording the response with a large array of receivers at the surface. By repeating the experiment many times for different source positions, enough data can be gathered to form useful images. The scale of these experiments varies from tens of meters, for instance, for near-surface imaging, to tens of kilometers in oil and gas exploration, up to the whole Earth in global seismology. For hydrocarbon exploration, depths typically extend to several kilometers.

One of the main challenges of seismic imaging is that the propagation velocity of the seismic waves traveling through the subsurface is unknown. Since this velocity can vary significantly, both laterally and with depth, it has to be estimated prior to applying a conventional imaging method. A wrong subsurface velocity can lead to severely distorted images. The conventional workflow for seismic imaging, therefore, is to first estimate the subsurface velocity and subsequently back-propagate the data to obtain an image. The success of this two-step approach relies crucially on the ability to isolate the reflection events in the data. This can only be done if the earth structure is relatively simple, e.g. smoothly varying with small perturbations. A particularly relevant setting in which the separation-of-scales argument fails to hold is in the presence of strong contrasts, such as salt diapirs, salt slabs, anhydrite or basalt layers. Salt geometries are of particular interest because they often have hydrocarbon reservoirs on their sides or underneath.

Full-waveform inversion (FWI) attempts to fit the data using a fairly precise numerical model for wave-propagation in heterogeneous media [77, 148, 204]. By posing the inverse problem as a nonlinear least-squares problem, the velocity structure in the subsurface can, in principle, be estimated quantitatively. However, this optimization problem is severely nonlinear and ill-posed, making it very difficult to obtain reasonable results without a good initial guess of the velocity parameters. Starting from an initial guess that is far away from the truth often leads to an incorrect subsurface model. While many approaches have been proposed to mitigate this problem, the issue remains unsolved[185, 39, 146, 200].

In this chapter, we aim to alleviate the ill-posedness of the problem to some extent by adding an appropriate regularization. In the particular setting of salt bodies, it is reasonable to assume that the subsurface can be described as one or more continuous bodies (salt) with known constant material parameters, surrounded by continuously varying parameters (sediment). Our approach is based on a level-set method, where we implicitly represent the shape of the salt body with a level-set function. By expanding the level-set function in some basis, we greatly reduce the effective number of parameters and obtain a nonlinear optimization problem that is better behaved than the original nonlinear least-squares problem.

The chapter is organized as follows. In Section 5.2, we review the details of the

classical FWI approach and give an overview of regularization approaches specifically aimed at the reconstruction of salt bodies. In Section 5.3, we present the basics of the parametric level-set method and discuss several practical issues in detail. Numerical results on a stylized seismic example are presented in Section 5.4. Finally, Section 5.5 concludes the chapter.

5.2 Full waveform inversion

Full waveform inversion (FWI) is nonlinear data-fitting scheme aiming to retrieve detailed estimates of subsurface properties from seismic data. The basic workflow of FWI is as follows: *(1)* predict the observed data by solving a wave-equation, given an initial guess of the subsurface velocity, *(2)* compute the difference between predicted and observed data, and *(3)* update the velocity in order to improve the data fit. This process is repeated in an iterative fashion until the residual drops below some tolerance. An excellent overview of various flavors of this basic scheme can be found in [204].

There are several ways to model seismic wave propagation, in either the frequency or time domain. We refer the reader to the reviews by [46] and [203]. For our purpose, it suffices to consider a two-dimensional scalar Helmholtz equation that models the acoustic pressure:

$$\omega^2 m(\mathbf{x}) u(\omega, \mathbf{x}) + \nabla^2 u(\omega, \mathbf{x}) = q(\omega, \mathbf{x}), \tag{5.2.1}$$

where $\mathbf{x} \in \mathbb{R}^2$ denotes the subsurface position, $m(\mathbf{x})$ is the squared slowness with units $\mathrm{s}^2/\mathrm{m}^2$, ω is the angular frequency, q the source term and u the pressure wavefield. We consider an unbounded domain and impose Sommerfeld radiation conditions The observed data are modeled by solving (5.2.1) for several sources $\{q_i\}_{i=1}^{n_s}$ and sampling the resulting wavefields u_i at locations $\{\mathbf{x}_j\}_{j=1}^{n_r}$ and frequencies $\{\omega_k\}_{k=1}^{n_f}$, i.e.,

$$d_{ijk} = u_i(\omega_k, \mathbf{x}_j).$$

The inverse problem is now to retrieve $m(\mathbf{x})$ from a set of observations d_{ijk}. This problem has been extensively studied and uniqueness results are available for a few specific cases, including layered earth-models, small perturbations around a known smooth reference medium and asymptotic versions of the problem. We refer to [183] for an extensive overview.

5.2.1 Discretization and optimization

A common approach to the inverse problem is to first discretize and solve the wave equation for forward modelling and to subsequently formulate a finite-dimensional data-fitting problem.

A finite-difference discretization of (5.2.1) on an N-point grid with absorbing boundary conditions leads to a sparse system of equations

$$A(\omega, \mathbf{m})\mathbf{u} = \mathbf{q},$$

where $A \in \mathbb{C}^{N \times N}$ is structurally symmetric and indefinite. Various dispersion-minimizing finite-difference stencils have been proposed in the literature

[172, 97, 187] and several absorbing boundary conditions are described in [52] and [54].

The system of equations can be solved by a decomposition of A such as the LU (lower and upper triangular) decomposition. The advantage of this direct approach is that, once the decomposition has been performed, the system can be solved efficiently for multiple sources by forward and backward substitution [125, 131]. The direct approach with nested dissection [78] is efficient for 2-D problems [97, 182, 96]. However, the time and memory complexities of LU factorization and its limited scalability on large-scale distributed memory platforms prevent its application to large-scale 3-D problems that may involve more than 10 million unknowns [139]. Helmholtz-specific factorization methods have been developed [205], but these are only suitable when the computational cost can be amortized over many sources.

Iterative solvers provide an alternative approach for solving the Helmholtz equation [151, 145, 66, 112]. These iterative solvers are usually implemented with Krylov subspace methods [158] that are preconditioned by solving a damped Helmholtz equation. The solution of the damped equation is computed efficiently with a multigrid method. The main advantage of the iterative approach is the low memory requirement, although the main drawback results from the difficulty to design an efficient preconditioner because A is indefinite [67]. Also, the iterative scheme has to be started again for each source.

Organizing the observations in a vector $\mathbf{d} \in \mathbb{C}^M$ with $M = n_s n_r n_f$, we introduce the forward operator

$$\mathbf{d} = F(\mathbf{m}).$$

Application of the forward operator on a given model $\mathbf{m}$ involves the solution of $n_f n_s$ Helmholtz equations, including the LU decompositions, and constitutes the main computational cost of FWI.

The conventional least-squares formulation of FWI can now be expressed as

$$\min_{\mathbf{m}} \{ f(\mathbf{m}) = \tfrac{1}{2} \| F(\mathbf{m}) - \mathbf{d} \|_2^2 \}. \tag{5.2.2}$$

This optimization problem is typically solved with a Newton-like algorithm [149]:

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \lambda_k H_k^{-1} \nabla f(\mathbf{m}_k),$$

where λ_k is the step size and H_k denotes (an approximation of) the Hessian of f at iteration k. The gradient of the objective is given by

$$\nabla f(\mathbf{m}) = J(\mathbf{m})^* (F(\mathbf{m}) - \mathbf{d}),$$

where $J(\mathbf{m})$ is the Jacobian of F and J^* denotes its adjoint or conjugate transpose. In practice, the Jacobian matrix is never formed explicitly but its action is computed using the *adjoint-state method* [84]. This entails solving $n_s n_f$ linear systems with A^* and the residual as right-hand sides. The action of the (Gauss-Newton) Hessian may be computed in a similar fashion at the cost of additional Helmholtz-solves [149]. This can be done cheaply in 2D if the LU factors are kept in memory while needed [133].

5.2.2 Ill-posedness and local minima

Unfortunately, waveform inversion is hampered by the presence of local minima in f [160]. In practice, this requires a good initial estimate of the parameters $\mathbf{m}$. To circumvent this problem, the data can be inverted in a multi-scale fashion, starting at some lowest frequency available in the data and using the inversion result to initialize a next pass at a higher frequency [39, 175]. The low frequencies are important in reconstructing the large-scale variations in the model, while high frequencies fill in the details [36]. All of these will fail, however, when the initial guess does not explain the observations well enough for the lowest available frequency.

Many alternative formulations have been proposed that depart from the usual data-fitting approach [185, 146, 200, 184, 23]. While these approaches can to some extent mitigate the non-linearity of the problem, they do not solve the inherent ill-posedness of the problem. This means that some features of $\mathbf{m}$ are simply not recoverable, regardless of the method we use to estimate the parameters. To address this problem, we need to add regularization.

5.2.3 Regularization

We distinguish two types of regularization: *implicit* regularization, where we add a penalty $\rho(\mathbf{m})$ to the objective in (5.2.2) to penalize unwanted features, and *explicit* regularization, where we expand $\mathbf{m}$ in an appropriate basis that contains only the features we desire. For example, when we expect the model to vary smoothly, we can penalize the second derivative by a penalty term

$$\rho(\mathbf{m}) = \|L\mathbf{m}\|_2^2,$$

with L the discrete Laplace operator. Alternatively, we can choose a representation of the form

$$\mathbf{m} = B\mathbf{a},$$

where B consists of smooth basis functions such as B-splines. This type of regularization works well when the scales of the model are separable, since we can invert for a smoothly varying velocity from low-frequency data [39].

In some geological settings, however, the scales do *not* separate, and we need to find an alternative form of regularization. If we expect our model to have strong discontinuities, a popular choice is a Total-Variation (TV) regularization with $\rho(\mathbf{m}) = \|D\mathbf{m}\|_1$, with D a discrete gradient operator [156, 122]. A natural basis is hard to define in this case. Alternatively, we can regularize the model by imposing the constraint $\|D\mathbf{m}\|_1 \leq \tau$ [68].

A disadvantage is that TV regularization acts globally and causes the model to be blocky everywhere. Nevertheless, some promising results have been obtained recently [68, 9, 5].

As an alternative to TV regularization, a level-set method for waveform inversion in the presence of salt-bodies has been considered by [121, 83]. For completeness sake and to set the notation, we first give a brief overview of the classical level-set approach and then introduce our approach.

5.3 Level-Set method

We represent $m(\mathbf{x})$ as being constant in a certain region and continuously varying elsewhere. We start again from the continuous formulation of the inverse problem and derive a finite-dimensional optimization problem analogous to (5.2.2). We represent m as

$$m(\mathbf{x}) = \begin{cases} m_1 & \text{if } \mathbf{x} \in \Omega, \\ m_0(\mathbf{x}) & \text{otherwise.} \end{cases}$$

Here, Ω indicates the salt-body, m_1 is the constant value of the model parameter inside the salt body and $m_0(\mathbf{x})$ denotes the spatially varying parameters in the sediment. Figure 5.1 sketches three different models, representing the smooth variation, blocky structure and a combination of both. Model 1 is a typical sediment structure, while model 2 represents the salt geometry. We generally expect a seismic velocity distribution similar to model 3, combining model 1 and 2.

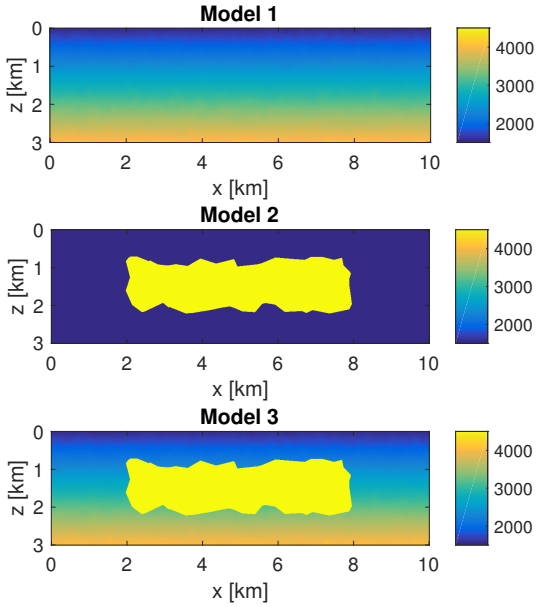


Figure 5.1: Model 1: smooth velocity variation (sediment). Model 2: random 2D body with higher velocity (salt). Model 3: combination of smooth variation and blocky model. All velocities in m/s.

We can represent the model formally as

$$m(\mathbf{x}) = [1 - \chi_\Omega(\mathbf{x})]m_0(\mathbf{x}) + \chi_\Omega(\mathbf{x})m_1,$$

where $\chi_\Omega(\mathbf{x})$ is the indicator function of Ω. The inverse problem now consists of finding the set Ω and the model parameters $m_0(\mathbf{x})$ and m_1. The basic idea behind the level-set method is to represent the domain Ω through a level-set function as

$\Omega = \{\mathbf{x} \,|\, \phi(\mathbf{x}) \geq 0\}$ [140, 132]. This then leads us to represent the indicator function as $\chi_\Omega(\mathbf{x}) = h(\phi(\mathbf{x}))$, where h is the Heaviside function $h(s) = (1 + \text{sign}(s))/2$.

To be able to compute sensitivities, one typically uses a smooth approximation of the Heaviside function. A common choice is

$$h_\epsilon(s) = \frac{1}{1 + e^{-s/\epsilon}},$$

where $h_\epsilon \to h$ as $\epsilon \to 0$. This function has the nice property that its derivative is non-zero everywhere. A disadvantage is that, in order to accurately represent the indicator function, the level-set function ϕ will need to tend to $\pm\infty$. This induces very steep gradients in ϕ around the boundary of the level-set. In turn, these steep gradients in ϕ require that we pick a proportionally large ϵ to remain sensitive to changes in the level-set. This suggest that we pick ϵ in accordance with the (maximum) gradient of ϕ. We will get back to this observation in Section 5.3.3.

To avoid some of these issues, we use a smooth Heaviside defined by

$$h_\epsilon(s) = \begin{cases} 0 & \text{if } s < -\epsilon, \\ \frac{1}{2}\left[1 + \frac{s}{\epsilon} + \frac{1}{\pi}\sin(\frac{\pi s}{\epsilon})\right] & \text{if } -\epsilon \leq s \leq \epsilon, \\ 1 & \text{if } s > \epsilon. \end{cases}$$

To avoid getting trapped in the region where $h'_\epsilon = 0$, we again have to pick ϵ in accordance with the (maximum) gradient of ϕ. A practical heuristic to pick ϵ and adapt it to the current ϕ will be discussed in Section 5.3.3.

The level-set method was originally introduced for tracking regions in fluid flow applications, providing a natural way to evolve the level-set by solving a Hamilton-Jacobi equation [142]. In applications like FWI, it is not obvious how to update ϕ away from the boundary of Ω, because $h'_\epsilon(x)$ quickly tends to zero. The problem of finding the level-set function is ill-posed.

5.3.1 A parametric level-set approach

We adopt a method proposed in [1] and represent the level-set function with a finite set of radial basis functions (RBFs):

$$\phi(\mathbf{x}) = \sum_{i=1}^{L} \alpha_i \Psi(\beta \|\mathbf{x} - \boldsymbol{\xi}_i\|_2),$$

where $\Psi(r)$ is a RBF, $\boldsymbol{\xi}_i$ are the nodes and β is a scaling parameter. The choice of RBF will be discussed in more detail in the next subsection.

Discretization on an N-point grid leads to the so-called RBF-kernel matrix $K \in \mathbb{R}^{N \times L}$ with elements $k_{ij} = \Psi(\beta \|\mathbf{x}_i - \boldsymbol{\xi}_j\|_2)$, allowing us to represent the parameters as

$$\mathbf{m}(\mathbf{m}_0, m_1, \boldsymbol{\alpha}) = \mathbf{m}_0 \odot (\mathbf{1} - h_\epsilon(K\boldsymbol{\alpha})) + m_1 h_\epsilon(K\boldsymbol{\alpha}), \tag{5.3.1}$$

where $\odot$ represents element-wise multiplication, also known as the Hadamard product. We can now define the corresponding optimization problem for the

Table 5.1: Global RBFs.

Name	$\Psi(r)$
Gaussian	$\exp(-r^2)$
Multiquadric	$\sqrt{1+r^2}$
Inverse multiquadric (IMQ)	$\frac{1}{\sqrt{1+r^2}}$
Inverse quadratic (IQ)	$\frac{1}{1+r^2}$

parametric level-set approach, for fixed $\mathbf{m}_0, m_1$, as minization over the cost functional

$$\widetilde{f}(\boldsymbol{\alpha}) = \tfrac{1}{2}\|F(\mathbf{m}(\boldsymbol{\alpha})) - \mathbf{d}\|_2^2. \tag{5.3.2}$$

The gradient of this objective is given by

$$\nabla \widetilde{f}(\boldsymbol{\alpha}) = \left(\frac{\partial \mathbf{m}}{\partial \boldsymbol{\alpha}}\right)^T \nabla f(\mathbf{m}(\boldsymbol{\alpha})), \text{ where } \frac{\partial \mathbf{m}}{\partial \boldsymbol{\alpha}} = \operatorname{diag}\left\{(m_1 \mathbf{1} - \mathbf{m_0}) \odot h_\epsilon'(K\boldsymbol{\alpha})\right\} K. \tag{5.3.3}$$

5.3.2 Radial basis functions

Radial basis functions are a means to approximate smooth multivariate functions. They have been extensively studied in the context of the interpolation of scattered data in high dimensions and for meshless methods [214, 38, 147]. RBFs are classified into two main types, global RBFs, which have infinite support, and compactly supported RBFs. Next, we discuss some relevant properties when we consider approximating a given smooth function ϕ using RBFs.

Global RBFs

Global RBFs have infinite support and hence the RBF kernel matrix K is dense. An overview of several common global RBFs is given in Table 5.1 and their radial behavior is shown in Figure 5.2. For the multiquadric, the kernel matrix is positive definite [129]. Among the advantages of global RBFs are (1) highly accurate and often exponentially convergent, (2) easily applicable to high-dimensional problems, (3) meshless in the approximation of multivariate scattered data and (4) numerical accuracy is easily improved by adding more points in regions with large gradients.

However, the corresponding interpolation matrix is dense and ill-conditioned and therefore sensitive to the shape parameter. As a result, the application of traditional RBF interpolation to large-scale problems is computationally expensive.

Compactly supported RBF

These result in a sparse, positive definite, and generally better conditioned kernel matrix [207]. However, the order of approximation is usually less than with global RBFs. Table 5.2 provides an overview of some common compactly supported RBFs. Figure 5.2 presents their radial behavior.

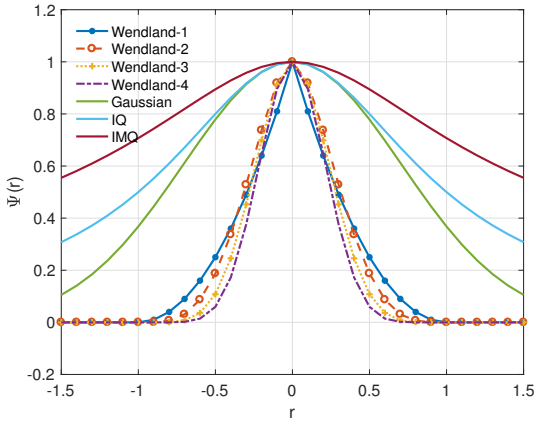


Figure 5.2: Various types of global and compact (Wendland) RBF.

Table 5.2: Wendland compactly-supported RBFs. Here, k denotes the order of smoothness (i.e., ϕ is k times continuously differentiable). $(x)_+$ operation gives value itself if $x \geq 0$, and set it to 0 if $x < 0$.

Name	$\Psi(r)$	k
Wendland-1	$(1-r)_+^2$	0
Wendland-2	$(1-r)_+^4(4r+1)$	2
Wendland-3	$(1-r)_+^6(35r^2+18r+3)$	4
Wendland-4	$(1-r)_+^8(32r^3+25r^2+8r+1)$	6

5.3.3 Shape representation

To determine which type of RBF is most suitable for our purpose, we study how well we can represent typical salt bodies with various RBFs. These salt models are discretized on a cartesian grid with grid spacing h. We choose the nodes of the RBFs on a cartesian grid with a larger grid spacing $h_r = 5h$ and normalize the scale parameter for the compactly supported RBFs with $\beta = \frac{1}{\gamma h_r}$.

We determine the coefficients $\boldsymbol{\alpha}$ by solving

$$\min_{\boldsymbol{\alpha}} \tfrac{1}{2}\|h_\epsilon(K_\gamma \boldsymbol{\alpha}) - \mathbf{m}\|_2^2,$$

with a L-BFGS method[136]. Results for the Wendland-4 RBF with $\gamma = 4$ and $\epsilon = 10^{-1}$ are shown in Figure 5.3(a–d). The lower-order Wendland RBFs gave a less good approximation. Results with the global RBFs are similar to those of the compact ones.

As noted earlier in Section 5.3, we need to pick ϵ in accordance with the gradient of the level-set function in order to have optimal sensitivity. We propose to choose ϵ adaptively based on the (fraction) of an upper bound on the gradient. It is

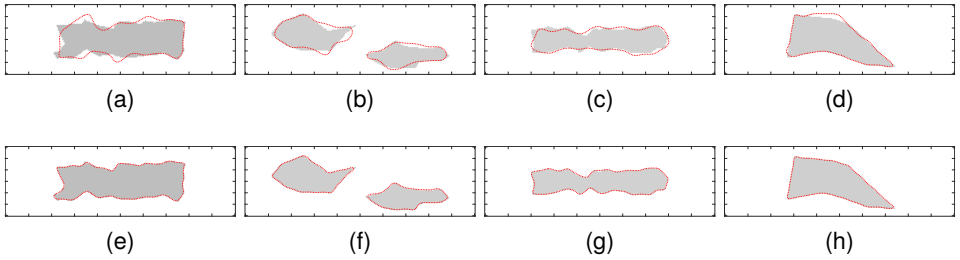


Figure 5.3: (a), (b), (c), (d) show level-set reconstructions of salt geometries with fixed $\epsilon = 0.1$, while (e), (f), (g), (h) reveal the improved reconstructions with new formulation of ϵ presented in equation (5.3.4) (red dotted line denotes the reconstructed zero contour of the level-set).

represented as

$$\epsilon = \kappa \left[\max(K_\gamma \boldsymbol{\alpha}) - \min(K_\gamma \boldsymbol{\alpha})\right]. \tag{5.3.4}$$

This choice of ϵ with $\kappa = 5 \times 10^{-2}$ produced the results in Figure 5.3(e–h).

5.3.4 Algorithm

Algorithm 7 summarizes parametric level-set full waveform inversion (PLS-FWI). This method introduces a small, $\mathcal{O}(N)$ flops overhead compared to the conventional FWI approach. This is negligible compared to the $\mathcal{O}(N^{(3/2)})$ complexity of solving the Helmholtz equation in 2D with nested dissection.

Algorithm 7 PLS-FWI Basic Algorithm

Require: Data $\mathbf{d} \in \mathbb{R}^{n_s \times n_r \times n_f}$, estimate of background parameter $\mathbf{m_0}$, salt parameter m_1, initial estimate of weights $\boldsymbol{\alpha}$, κ
Ensure: final weights $\boldsymbol{\alpha}$, model $\mathbf{m}$
1: **for** $j = 1$ to itermax **do**
2: compute Heaviside ϵ from equation (5.3.4).
3: compute misfit $\widetilde{f}(\boldsymbol{\alpha})$ from equation (5.3.2) and gradient $\nabla \widetilde{f}(\boldsymbol{\alpha})$ from equation (5.3.3). Form an estimate of the Hessian, $\widetilde{H}$ from the LBFGS procedure.
4: $\boldsymbol{\alpha} \leftarrow \boldsymbol{\alpha} + \lambda \widetilde{H}^{-1} \nabla \widetilde{f}(\boldsymbol{\alpha})$.
5: **end for**
6: compute $\mathbf{m}$ from equation (5.3.1).

Algorithm 8 outlines the multi-scale approach of PLS-FWI. We reduce κ for Heaviside ϵ after every frequency band to decrease the size of level-set boundary. The idea is to start optimization with large boundary (high κ) and small initial level-set to capture large sensitivity, allowing for large updates. Decreasing the level-set boundary then provides sharper images.

Algorithm 8 PLS-FWI Multi-scale Algorithm

Require: Data in frequency batches $\{\mathbf{d}_1, \mathbf{d}_2, \dots \mathbf{d}_n\}$, estimate of background parameter $\mathbf{m_0}$, salt parameter m_1, initial estimate of weights $\boldsymbol{\alpha}_0$
Ensure: model $\mathbf{m}$
1: $\kappa \leftarrow 0.1$
2: **for** $i = 1$ to n_{batches} **do**
3: $\quad \boldsymbol{\alpha}_i \leftarrow$ PLS-FWI$(\mathbf{d}_i, \mathbf{m_0}, m_1, \boldsymbol{\alpha}_{i-1}, \kappa)$
4: $\quad \kappa \leftarrow 0.8 \times \kappa$
5: **end for**
6: $\epsilon \leftarrow 0$
7: compute $\mathbf{m}$ from equation (5.3.1).

5.4 Results

To demonstrate the capabilities of the parametric level-set full waveform inversion, we perform simulations on four different velocity models. These models are shown in Figure 5.4. Each has a background velocity increasing linearly with depth z as $1500 + bz$ with $b = 0.8333$. The salt bodies in these model have a constant velocity of 4500 m/s. We choose a grid spacing of 50 m in each direction, providing a total of $N = 201 \times 61$ grid points.

The acquisition setup is shown in Figure 5.5, with 50 sources placed at the top of the model and 100 receivers placed at a depth of 50 m. Also shown is the part of the model to which the data are most sensitive. Features of the model in the lower left and right corners are hard to recover because of the limited aperture.

We use a frequency-domain finite-difference code [174] to generate the data for frequencies between 2.5 and 3.5 Hz with a spacing of 6.25×10^{-2} Hz. The data are weighted in frequency by a Ricker wavelet with a peak frequency of 15 Hz. For the inversion, we select four bands, $[2.5 - 2.75]$ Hz, $[2.75 - 3.0]$ Hz, $[3.0 - 3.25]$ Hz and $[3.25 - 3.5]$ Hz, each with 4 four frequencies.

For PLS-FWI, the RBF grid has a spacing of 250 m in both directions. As shown in Figure 5.6, two extra layers of RBF nodes are added outside the physical domain to provide flexibility in reconstructing the level-set near the boundary. We use a fourth-order Wendland RBF with $\gamma = 4$ and choose ϵ adaptively as described in Section 5.3.3. We found that ϵ usually stabilizes after the first few iterations.

We compare FWI and PLS-FWI on noise-free data and on data with white noise, with an SNR of 10 dB. Finally, we perform a joint reconstruction of both the salt geometry and the background model $\mathbf{m}_0 = 1500 + bz$, parametrized by the slope b.

For FWI, we use a projected Quasi-Newton method [166] to solve the resulting optimization problem with bound-constraints on $\mathbf{m}$ to ensure that the velocity stays within the feasible range of $[1500, 4500]$ m/s. To solve the resulting optimization problem in $\boldsymbol{\alpha}$ (PLS-FWI), we rely on the L-BFGS approximation of the Hessian[167]. We perform 150 iterations per frequency band while resetting the L-BFGS memory after handling each frequency band. For joint salt and background reconstruction with PLS-FWI, we use a bisection method to find the optimal b.

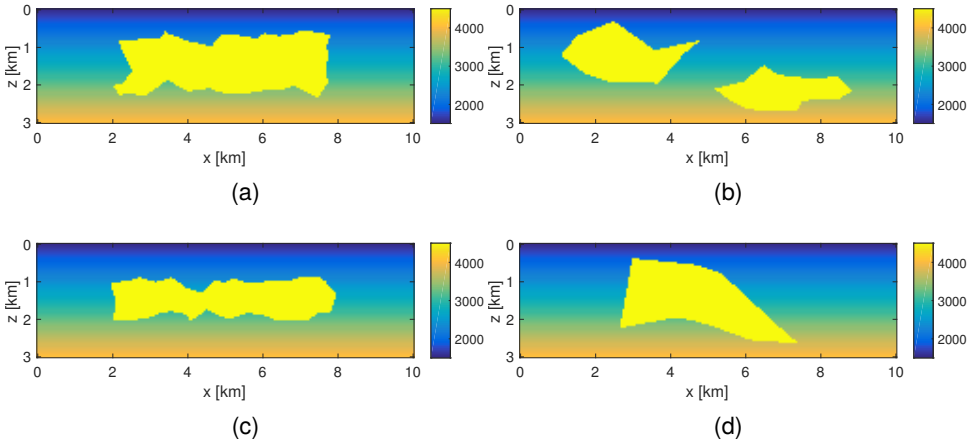


Figure 5.4: Velocity models (in m/s) created for the inversion test. They are referred to as model A (a), model B (b), model C (c) and model D (d).

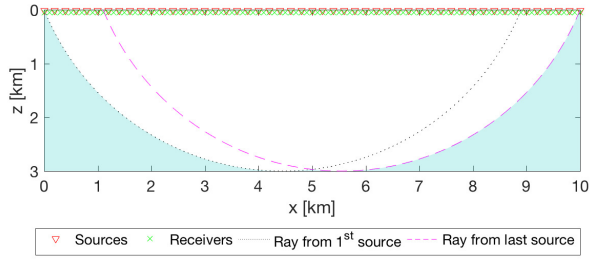


Figure 5.5: Source and receivers for the simulation. Sources are placed on top, receivers at a depth of 50 m. The two ray paths separate the region recoverable by inversion from the shaded non-recoverable part [115].

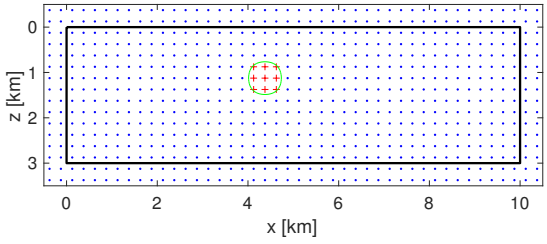


Figure 5.6: Placements of RBF on the computational grid (denoted by thick black line). To initialize the zero contour of the level set (green line) that defines the salt body, a few RBFs around the center have been allocated positive values (denoted by red plusses), others are negative (denoted by blue dots).

Table 5.3: Comparison of misfit and RRE values for classical FWI vs PLS-FWI

Model	ERF		RRE	
	FWI	PLS-FWI	FWI	PLS-FWI
A	0.0281	5.8197×10^{-4}	0.8213	0.0707
B	0.0221	2.9183×10^{-4}	0.8072	0.0856
C	0.0261	4.7127×10^{-4}	0.9102	0.0732
D	0.0339	9.0299×10^{-6}	0.8088	0.0434

5.4.1 Salt geometry determination

The initial model for conventional FWI and the background model for PLS-FWI are taken to be the same, linearly increasing velocity models as in the true ones. For PLS-FWI, we let m_1 correspond to the true velocity in the salt and initialize the level-set as shown in Figure 5.6. For model B, we initialize the level-set function with two positive regions near the top of the two salt bodies.

Noise-free data

Figure 5.7a, 5.7c, 5.7e and 5.7g show the reconstructed models using FWI. To a large extent, the results predict the top of the salt but fail to identify its proper shape. They also include some artefacts, as shown in the left bottom part of Figure 5.7g. On the other hand, PLS-FWI almost perfectly recovers the salt geometry in each of the models, as can be seen in Figure 5.7b, 5.7d, 5.7f and 5.7h. These models retain their sharpness even with low frequencies because we reduce ϵ to zero in step 6 of algorithm 8.

To compare the proposed method with conventional FWI, we define the error reduction factor (ERF),

$$\text{ERF}_{\text{recon}} = \frac{\|F(\mathbf{m}_{\text{recon}}) - \mathbf{d}\|_2}{\|F(\mathbf{m}_0) - \mathbf{d}\|_2}.$$

An ERF close to the best achievable ERF (0 in case of noise-free data) indicates a better performance of the reconstruction method in reducing the data misfit. Table 5.3 shows the improvement in the data misfit with PLS-FWI over classic FWI. The data misfit is reduced by a factor 10^{-4} on average with the use of PLS-FWI.

To compare the reconstructions for different methods, we define a measure called the Relative Reconstruction Error (RRE) as

$$\text{RRE}_{\text{recon}} = \frac{\|\mathbf{m}_{\text{recon}} - \mathbf{m}_{\text{true}}\|_2}{\|\mathbf{m}_0 - \mathbf{m}_{\text{true}}\|_2},$$

where $\mathbf{m}_{\text{recon}}$ is the model reconstructed by FWI or PLS-FWI. From Table 5.3, we observe that the RRE is reduced dramatically with PLS-FWI.

Noisy data

Figure 5.8a, 5.8c, 5.8e and 5.8g show the reconstructed models using FWI while Figure 5.8b, 5.8d, 5.8f, 5.8h show the models reconstructed by PLS-FWI. The results

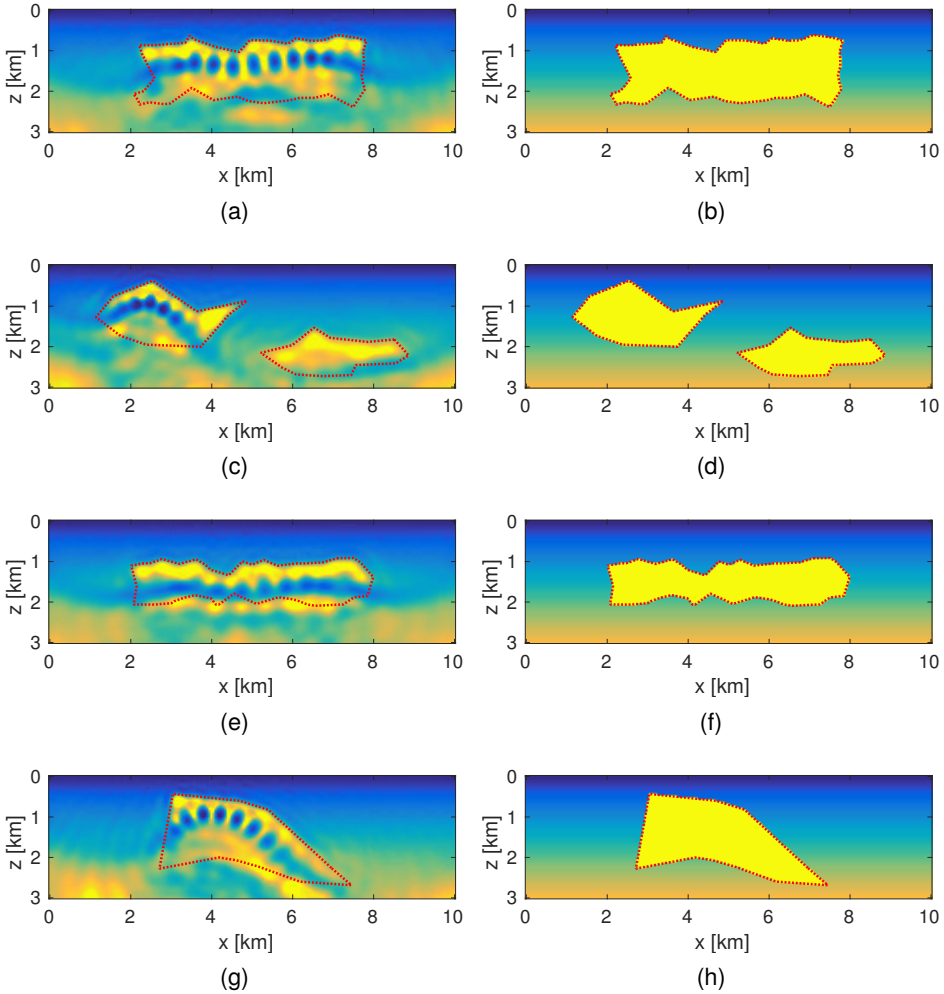


Figure 5.7: Salt reconstruction with classical FWI and parametric level-set full waveform inversion (PLS-FWI) from noise free data. (a), (c), (e) and (g) represent models A, B, C, and D, reconstructed with classical full waveform inversion. (b), (d), (f) and (h) represent reconstructed model A, B, C, D using PLS-FWI. The red dotted line shows the true geometry of salt.

are essentially the same as for the noise-free case, except for some artifacts outside the region of interest in the PLS-FWI results. Figure 5.9 exhibits the variation of the achieved level-set function for Model A. The gradient near the left bottom corner still is large, suggesting the method has difficulty in getting rid of the incorrect salt geometry.

Table 5.4 shows the improvement in the data misfit with PLS-FWI over classic

Table 5.4: Comparison of misfit and RRE values for classic FWI vs. PLS-FWI (noisy data).

Model	**ERF**			**RRE**	
	FWI	PLS-FWI	achievable	FWI	PLS-FWI
A	0.5425	0.5254	0.5246	0.8432	0.2437
B	0.6172	0.5956	0.5950	0.8332	0.2085
C	0.6090	0.5958	0.5951	0.9199	0.1817
D	0.5530	0.5325	0.5321	0.8239	0.1382

Table 5.5: Recovered values of b for different models (noisy data).

Model	A	B	C	D
b	0.8351	0.8351	0.8345	0.8333

FWI. For noisy data, we also look at the best achievable ERF, which is defined as

$$\text{ERF}_{\text{achievable}} = \frac{\|F(\mathbf{m}_{\text{true}}) - \mathbf{d}\|_2}{\|F(\mathbf{m}_0) - \mathbf{d}\|_2}.$$

This quantity denotes the smallest ERF achievable by any reconstruction method. The ERF for PLS-FWI is closer to the best achievable ERF than that of FWI. With the classic approach, the relative reconstruction error is slightly affected when noise is added to data. On the other hand, RRE changes by a large factor in the presence of noise with the proposed method. The false salt geometries mainly attribute to these large changes.

5.4.2 Simultaneous reconstruction of salt and background

Next, we jointly reconstruct the salt geometry and the background, parametrized by b. For each frequency band, we first estimate the optimal b, for fixed $\boldsymbol{\alpha}$ and setting $\epsilon = 0$, using a bisection method and subsequently estimate the salt geometry with PLS-FWI (Algorithm 7). The bisection method is initialized with the interval $[0, 1]$ for b.

Figure 5.10a, 5.10b, 5.10c and 5.10d show the reconstructed models for noisy data. The salt geometry is accurately predicted, but contains a few artifacts. Table 5.5 indicates the predicted values of b in each of the models, which are very close to true value of $b = 0.8333$.

5.5 Discussions and conclusion

Accurately determining the geometry of subsurface salt bodies from seismic data is a difficult problem. When casting the inverse problem into a nonlinear data-fitting problem, both the presence of local minima and the ill-posedness of the problem

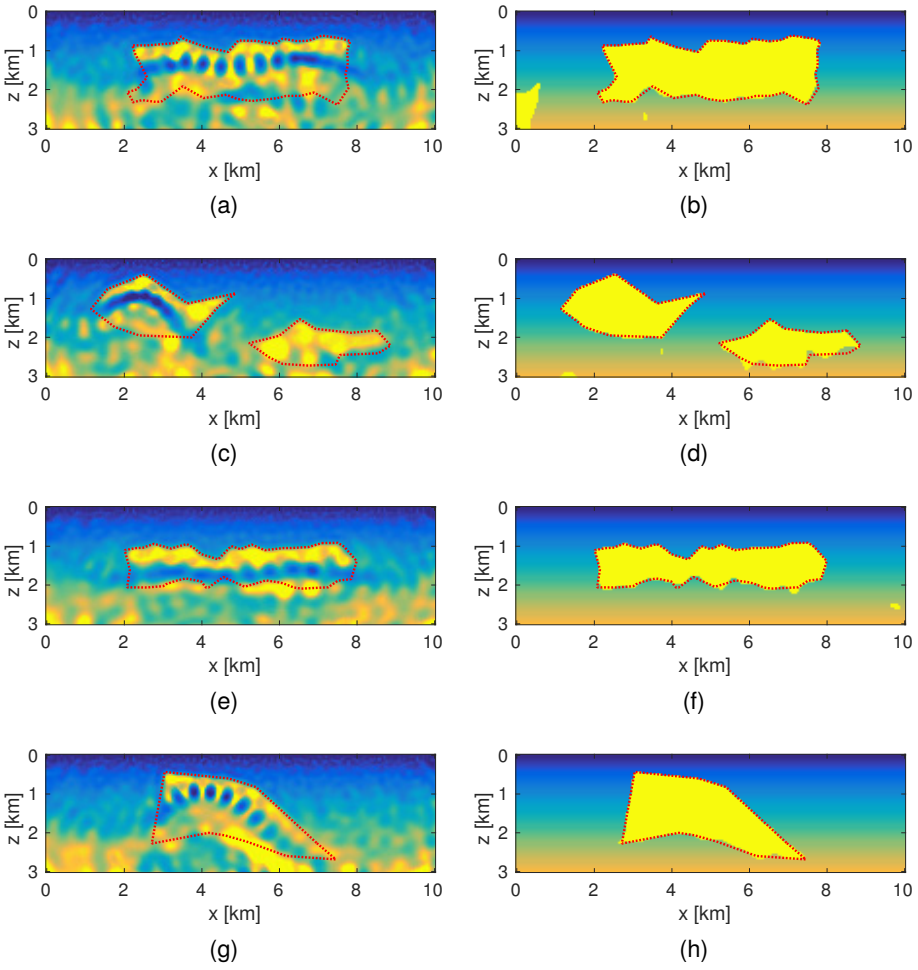


Figure 5.8: Salt reconstruction with classic FWI and Parametric level-set full waveform inversion (PLS-FWI) with noisy data of SNR 10 dB. (a), (c), (e), (g) represent reconstructed model A, B, C, D with classic full waveform inversion respectively. (b), (d), (f), (h) represent reconstructed model A, B, C, D with PLS-FWI respectively. Red dotted line shows the true geometry of salt.

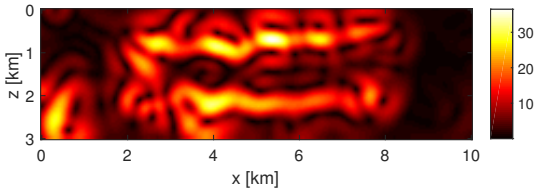


Figure 5.9: Gradient of the final level set function for model A with noisy data.

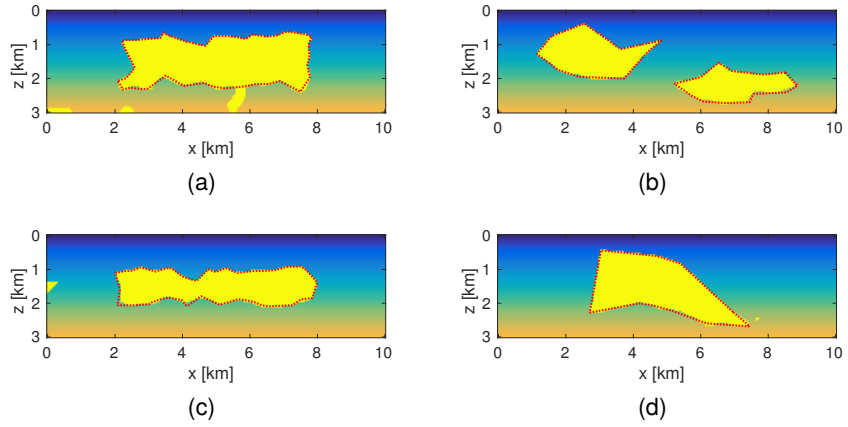


Figure 5.10: Simultaneous reconstruction of salt and background with Parametric level-set full waveform inversion (PLS-FWI) on noisy data with SNR $10\ dB$. (a), (b), (c), (d) represent reconstructed model A, B, C, D respectively. Red dotted line shows the true geometry of salt.

prevent accurate recovery of the salt-geometry. We have investigated the application of a parametric level-set method to address this problem. We represent the Earth model as a continuously varying background with an embedded salt body. The salt geometry is described by the zero contour of a level-set function, which in turn is represented with a relatively small number of radial basis functions. This formulation includes some additional parameters such as the width of the basis functions and the smoothness of the Heaviside function. The latter is of particular importance as it controls the sensitivity to changes in the salt geometry. We propose a robust algorithm that adaptively chooses the required smoothness parameter and tested the method on a suite of idealized Earth models with different salt geometries. For a fixed and accurate background model, the level-set method is shown to give superior estimates of the salt geometry and is stable against a moderate amount of noise. Additional results demonstrate that it is feasible to jointly estimate the background and the salt geometry.

To further develop the method as a viable alternative to conventional full-waveform inversion, tests on more realistic Earth models are needed. In particular, the joint estimation of the background model and salt geometry needs to be investigated further. Even when representing the level-set function with a finite basis, there are many level-set functions that result in the same salt geometry. To address this issue, additional regularization is needed. An often-used approach is to re-initialize the level-set function by solving an eikonal equation. In the parametric framework, this can be included by adding the discretized eikonal equation as a regularization term.

In this work, we fixed the RBF grid *a priori*. For very complex salt-geometries, this may no longer be feasible as it would require too many nodes for their accurate representation. An adaptive choice of the RBF grid may address this problem but it is not obvious how to refine the grid.

CHAPTER 6

Partially discrete waveform inversion

Abstract - *Full-waveform inversion attempts to estimate a high-resolution model of the Earth by inverting all the seismic data. This procedure fails if the Earth model contains high-contrast bodies such as salt and if sufficiently low frequencies are absent from the data. Salt bodies are important for hydrocarbon exploration because oil or gas reservoirs are often located on their sides or underneath. We represent the shape of the salt body with a level set, constructed from radial basis functions to keep its dimensionality low. We have shown earlier that the salt body can be completely recovered if the sediment structure is already known. We propose a strategy to simultaneously reconstruct the sediment and the salt. The sediment is implicitly represented by a bilinear interpolation kernel with a small number of variables. An alternating minimization technique solves the resulting optimization problem. The results on a synthetic model using Gauss-Newton approximation of the Hessian shows the feasibility of the approach.*

This chapter is based on the following publications:

A Kadu, T van Leeuwen, and W A Mulder. Parametric level-set full-waveform inversion in the presence of salt bodies. In *SEG Technical Program Expanded Abstracts 2017*, pages 1518–1522. Society of Exploration Geophysicists, 2017.

"The beauty of mathematics only shows
itself to more patient followers"
- Maryam Mirzakhani

6.1 Introduction

Full-waveform inversion (FWI) has become a popular technique to produce high-resolution maps of the subsurface velocity and density from the available seismic data. In its classic form, the method obtains a model by iteratively fitting modeled to observed data in a least-squares sense. Because the underlying minimization problem is highly nonlinear and seismic data typically lack sufficiently low frequencies, a good initial guess of the model is required to avoid convergence to the nearest local minimum, which generally does not represent the ground truth [204]. The problem becomes worse in the presence of salt bodies, which are of particular interest to the oil industry because hydrocarbon reservoir are often located nearby or underneath. Often, conventional FWI reconstructs the top of the salt but fails to obtain its interior, its bottom, and everything underneath.

A level set can help to regularize the problem. [161] introduced the method to geometric inverse problems. It implicitly defines the shape of the salt as the level set or zero contour of a smooth function. The conventional method deals with the level-set evolution through the Hamilton-Jacobi equation, also known as the level-set equation. This approach suffers from steepening and flattening of the level set during iterations. The classic solution, re-initialization of the level set, solves the issue but is computationally intensive. [57], [121] and [83] successfully applied the method in combination with full-waveform inversion.

Previously, we presented a robust implementation that dynamically adapts the width of the level-set boundary to obtain faster convergence [103]. This approach implicitly avoids flattening and steepening of the level-set function near the boundary by adjusting the width according to an approximate level-set gradient. We assumed, however, a known background sediment model. In the present chapter, we propose a joint-inversion approach and try to reconstruct the background and salt geometry simultaneously.

6.2 Theory

The FWI problem in its classic least-squares formulation [188] reads

$$\min_m \left\{ \tfrac{1}{2} \|F(m) - \mathbf{d}\|_2^2 \right\},$$

where F models the scalar Helmholtz equation, m defines the subsurface model, for instance, P-wave velocity or density or both, and $\mathbf{d}$ represents the data.

It is natural to separate the model into salt and sediment with constant and known salt velocity, m_1 and sediment model m_0:

$$m(\mathbf{x}) = \{1 - a(\mathbf{x})\} m_0(\mathbf{x}) + a(\mathbf{x}) m_1, \tag{6.2.1}$$

where a is an indicator function that separates the salt from the sediment. Solving for $a(\mathbf{x})$ is an NP-hard problem because of its discrete nature [61]. A level-set

function, $\phi : \mathbb{R}^2 \to \mathbb{R}$, will simplify the problem:

$$a(\mathbf{x}) = \begin{cases} 1, & \phi(\mathbf{x}) \geq 0; \\ 0, & \phi(\mathbf{x}) < 0. \end{cases}$$

Mathematically, we represent an indicator function with the Heaviside function, i.e., $a(\mathbf{x}) = h(\phi(\mathbf{x}))$. Now, the problem is to find a function $\phi(\mathbf{x})$ that represents the true salt geometry.

6.2.1 Parametric level-set method

[103] chose a representation of the level-set function by a linear combination of *finite* compactly-supported radial basis functions (RBFs):

$$\phi(\mathbf{x}) = \sum_{j=1}^{n_s} \alpha_j \psi(\|\mathbf{x} - \boldsymbol{\chi}_j\|_2).$$

Its discrete version is represented by a RBF Kernel Matrix, B, with entries $b_{ij} = [\psi(\|\mathbf{x}_i - \boldsymbol{\chi}_j\|_2)]$, where $\{\mathbf{x}_i\}_{i=1}^{n}$ represents the model grid points. The RBF nodes $\{\chi_j\}_{j=1}^{n_s}$ are placed on an equidistant model grid. The RBF amplitudes $\{\alpha_j\}_{j=1}^{n_s}$ control the level set. Typically, we use a *sufficiently* smooth Wendland RBF of the form

$$\psi(r) = (1-r)_+^8(32r^3 + 25r^2 + 8r + 1),$$

where x_+ is equal to x if $x > 0$ and 0 otherwise. The Parametric Level-Set Full-Waveform Inversion (PLS-FWI) for fixed m_0 and m_1 becomes

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^{n_s}} \left\{ f(\boldsymbol{\alpha}) = \tfrac{1}{2}\|F(\boldsymbol{\alpha}) - d\|_2^2 \right\}.$$

The gradient and Gauss-Newton Hessian for this problem are

$$\begin{aligned} g_{\boldsymbol{\alpha}} &= A^T D_{\boldsymbol{\alpha}}^T J^* \left(F(\boldsymbol{\alpha}) - \mathbf{d}\right), \\ H_{\boldsymbol{\alpha}} &= A^T D_{\boldsymbol{\alpha}}^T J^* J D_{\boldsymbol{\alpha}} A, \end{aligned}$$

where J is the Jacobian of the forward modeling operator F. The diagonal matrix $D_{\boldsymbol{\alpha}}$ denotes the element-wise multiplication of the difference between the salt and sediment velocity with the Dirac-Delta function:

$$D_{\boldsymbol{\alpha}} = \mathbf{diag}\,(m_1 - m_0) \odot h'_\epsilon(A\boldsymbol{\alpha}),$$

where $h_\epsilon(\cdot)$ is an approximation of the Heaviside function with width ϵ. It is important to note that the level-set sensitivities depend on two main factors: (1) the difference between the salt and sediment velocity and (2) the approximation of the Heaviside function.

We use a compact approximation of the Heaviside function, plotted in Figure 6.1. This formulation provides a constant region of sensitivity around the level-set boundary. This allows level-set parameters to take large steps as the FWI gradient

(i.e., $J^*(F(\boldsymbol{\alpha}) - \mathbf{d})$) is extrapolated by a constant factor. Due to its compactness, only neighboring RBFs are updated, providing less artifacts in the reconstruction.

The width of the Heaviside depends on the level-set boundary and the spatial gradient of the level-set function. Because this gradient is expensive to compute, we approximate it using the lower and upper bounds of the level-set function. Hence, the Heaviside width is given by:

$$\begin{aligned} \epsilon &= \kappa \Delta x \left(\frac{\max(\phi) - \min(\phi)}{\Delta x} \right) \\ &= \kappa \left[\max(A\boldsymbol{\alpha}) - \min(A\boldsymbol{\alpha}) \right] \end{aligned} \tag{6.2.2}$$

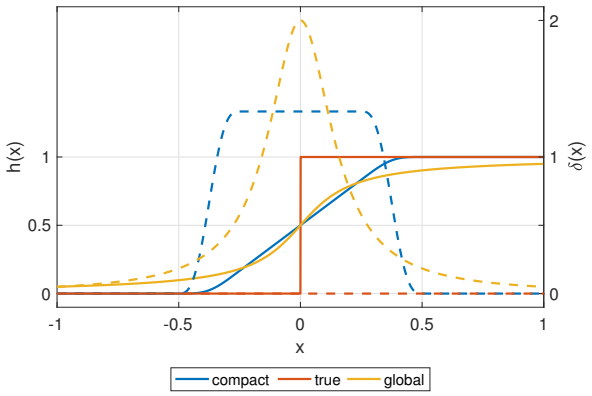


Figure 6.1: Heaviside and corresponding Dirac-Delta functions (dashed) with a width of 0.5.

We refer the reader to [105] for more details.

6.2.2 Bilinear Interpolation

We can impose smoothness constraints on the sediment velocity by means of bilinear interpolation from a model on a coarser mesh:

$$m_0(\mathbf{x}) = m_0^{\min} + \sum_{j=1}^{n_b} \beta_j l_j(\mathbf{x} - \widetilde{\boldsymbol{\chi}}_j).$$

Here, $m_0^{\min}$ denotes the water velocity, $l_j(\cdot)$ represents a piecewise linear basis function at node $\widetilde{\boldsymbol{\chi}}_j$ and β_j denotes its corresponding weight. The latter term can be captured using a bilinear interpolation kernel B when the model is discretized on a grid.

For fixed m_1 and $\phi(\mathbf{x})$ in equation (6.2.1), the optimization problem becomes

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^{n_b}} f(\boldsymbol{\beta}) = \tfrac{1}{2} \|F(\boldsymbol{\beta}) - \mathbf{d}\|^2.$$

The gradient and Gauss-Newton Hessian are given by

$$\begin{aligned} g_{\boldsymbol{\beta}} &= B^T J^* \left(F(\boldsymbol{\beta}) - \mathbf{d} \right), \\ H_{\boldsymbol{\beta}} &= B^T J^* J B. \end{aligned}$$

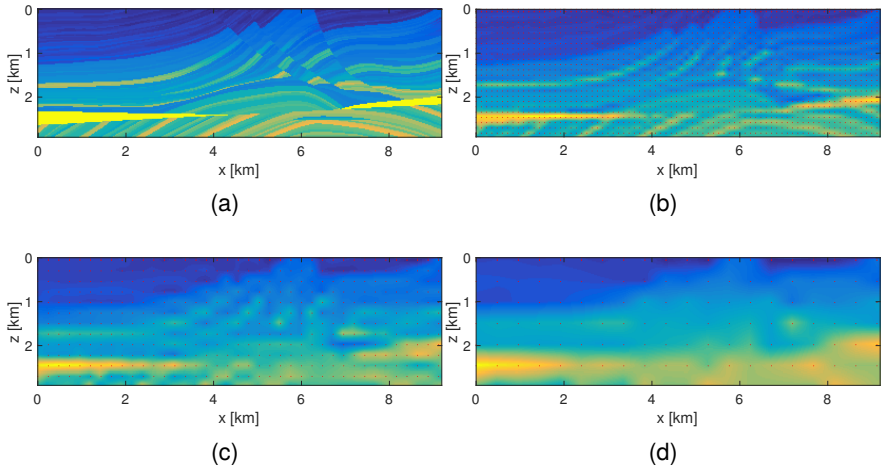


Figure 6.2: An example of bilinear interpolation. The less nodes, from (a) to (d), the smoother the representation of the model.

6.2.3 Joint reconstruction

The model is now represented in terms of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ as:

$$\mathbf{m} = \{1 - h(A\boldsymbol{\alpha})\} \odot (m_0^{\min} + B\boldsymbol{\beta}) + h(A\boldsymbol{\alpha})m_1. \tag{6.2.3}$$

With prior information about the salt velocity m_1 and water velocity $m_0^{\min}$, the minimization problem becomes

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left\{ f(\boldsymbol{\alpha},\boldsymbol{\beta}) = \tfrac{1}{2}\|F(\boldsymbol{\alpha},\boldsymbol{\beta}) - \mathbf{d}\|_2^2 \right\},$$
$$\text{such that } \boldsymbol{\beta}^{\min} \leq \boldsymbol{\beta} \leq \boldsymbol{\beta}^{\max}.$$

We use an alternating minimization strategy, splitting the optimization procedure in two parts, namely minimization over the level-set parameters $\boldsymbol{\alpha}$ and minimization over the background parameters $\boldsymbol{\beta}$. We alternately update the level set and the background in a multi-scale fashion. Algorithm 9 presents the basic algorithm, whereas Figure 6.3 outlines the multi-scale work-flow.

Algorithm 9 Basic Joint Reconstruction Algorithm

Require: $\mathbf{d}$ - data for frequency batch I_f, F - operator
$m_1, m_0^{\min}, \boldsymbol{\beta}^{\min}, \boldsymbol{\beta}^{\max}$ - model prior information
A, B - Kernel matrices, $\boldsymbol{\alpha}^0, \boldsymbol{\beta}^0$ - initial values
Ensure: $\{\boldsymbol{\alpha}^{\text{final}}, \boldsymbol{\beta}^{\text{final}}\}$, $\mathbf{m}$ - final model
for $i = 1$ to $\text{InnerIter}_{\max}$ **do**
  for $j = 1$ to J **do**
    $\boldsymbol{\beta}^{(j+1)} = \boldsymbol{\beta}^{(j)} + \left(H_{\boldsymbol{\beta}}\left(\boldsymbol{\alpha}^{(i-1)}, \boldsymbol{\beta}^{(j)}\right)\right)^{-1} g_{\boldsymbol{\beta}}\left(\boldsymbol{\alpha}^{(i-1)}, \boldsymbol{\beta}^{(j)}\right)$
  end for
  $\boldsymbol{\beta}^{(i)} = \boldsymbol{\beta}^{(J)}$
  for $k = 1$ to K **do**
    $\boldsymbol{\alpha}^{(k+1)} = \boldsymbol{\alpha}^{(k)} + \left(H_{\boldsymbol{\alpha}}\left(\boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(i)}\right)\right)^{-1} g_{\boldsymbol{\alpha}}\left(\boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(i)}\right)$
  end for
  $\boldsymbol{\alpha}^{(i)} = \boldsymbol{\alpha}^{(K)}$
end for
compute $\mathbf{m}$ from equation (6.2.3).

In step 5, we use an interior-point method that incorporates bounds on the parameters $\boldsymbol{\beta}$ and in step 9, a simple Newton method. In both these steps, the descent direction is calculated by a truncated conjugate-gradient method.

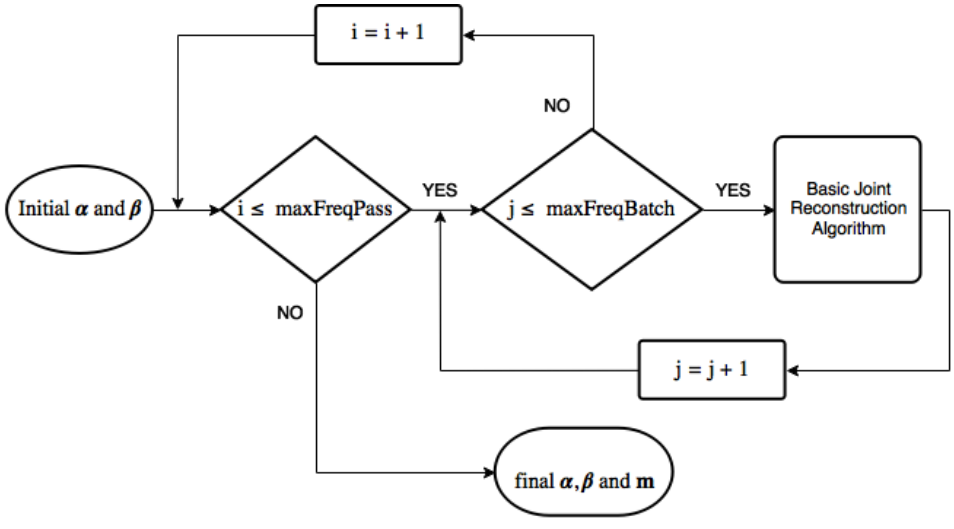


Figure 6.3: Multiscale Joint Reconstruction Algorithm

6.2.4 Bounds on sediment parameter

To avoid the delineation of salt in the sediment, it is important to put bounds on sediment velocity. Figure 6.4 shows an example. We allow for smooth updates of

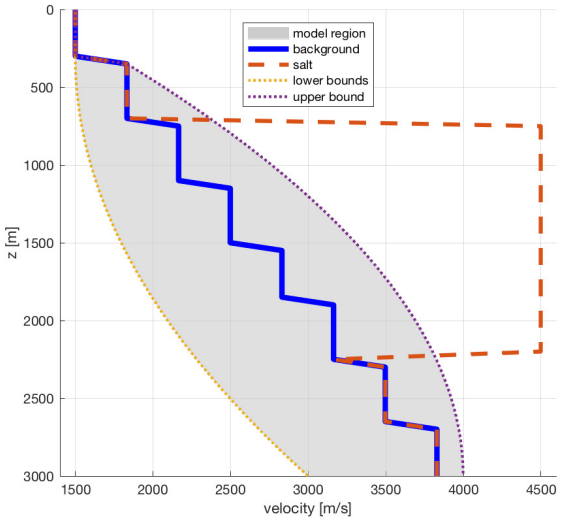


Figure 6.4: Bounds on the sediment velocity

the level set near the top region of the salt by constraining the sediment velocity to an upper bound close to the true velocity. As noticed in the FWI results, the region just below the top part demands very low velocities. This can also happen with the proposed approach and will slow down the reconstruction procedure. To accelerate the convergence, we impose an appropriate lower bound as a function of *depth*.

6.3 Examples

To demonstrate the capabilities of the proposed method, we present numerical experiments on a synthetic model with acoustic data. Figure 6.5(a) shows a model 10 km long and 3 km deep, discretized with a 50-m grid spacing. The sediment is a staircase in depth below a 300-m water layer with velocities between 1500 and 4000 m/s. The salt body, embedded in the sediment, has its top at 200 m below the water bottom with a constant velocity of 4500 m/s. Sources were placed 10 m deep and 200 m apart. The source is a Ricker wavelet with a 15-Hz peak frequency and zero time lag. The data were acquired with receivers placed 50 m apart starting at a smallest offset of 100 m up to a largest of 4 km. To avoid a full inverse crime, a different finite-difference code generated the data for a model discretized with a much finer grid spacing of 50/8 m. The amplitude of the source wavelet is estimated for each frequency at every step [148].

For the classic full-waveform inversion, we apply a spectral projected gradient method with bounds [166] on the velocity. For the initial model, we take a linear velocity profile with depth. The inversion is performed in a multi-scale fashion over the frequency range 2.5–4.5 Hz with 200 iterations for each frequency batch and a total of 3 passes [70] over the frequency range. Figure 6.5(b) displays the results. The top of the salt near the water bottom has been reconstructed reasonably well,

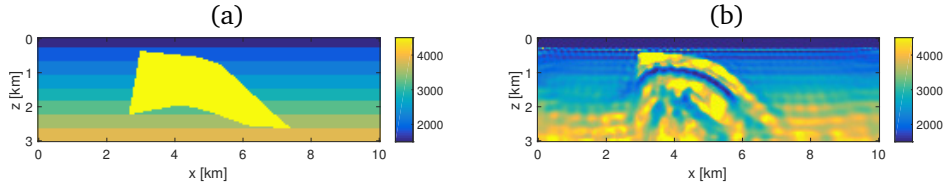


Figure 6.5: (a) True velocity (in m/s) of synthetic model. (b) Classic Full-Waveform Inversion results

but not the salt body below. The sediment structure at larger depths is also lost.

The parametric level-set parameters $\boldsymbol{\alpha}$ are initialized as shown in Figure 6.6(a). All negative RBFs have a value of -1 and all positive of $+1$. We place $25(z) \times 20(x)$ nodes over the model grid for the background. The background parameters $\boldsymbol{\beta}$ are initialized with a smooth linear trend in depth. We incorporate prior information about the water bottom, at $z = 350\,\text{m}$, in the initial model.

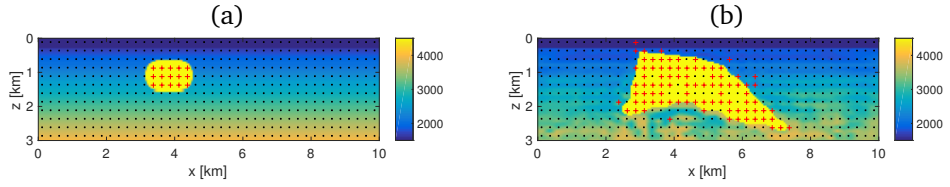


Figure 6.6: (a) Initial model with velocities in m/s. The level set is initialized around the top of the salt. Positive RBFs are denoted by *red plusses* and negative by *dots*. (b) Model obtained by the joint reconstruction approach. Corresponding positive (*pluses*) and negative (*dots*) RBFs.

The optimization over β is performed with the interior-point method (`fmincon` in MATLAB®). We restrict ourselves to $J = 10$ iterations in algorithm 9. We apply the `minFunc` [167] code in MATLAB® to optimize over α, limited to $K = 20$ iterations. In both these steps, the step direction is calculated by at most 10 iterations of the conjugate gradient method. A total of 4 passes are made over the frequency range, along with 2 inner iterations for each frequency batch. In total, we perform about 960 iterations. The Heaviside width parameter (κ) is initialized with 0.05 and reduced by 20% after each frequency pass to produce sharp boundaries for the salt.

Figure 6.6(b) shows the model obtained with the proposed method. The salt is reconstructed accurately at its top and sides. The sediment structure is reconstructed well down to a depth of 1.5 km as shown in Figure 6.8. Figure 6.7 illustrates the need for multiple passes over the frequency range. Figure 6.9 shows that the method manages to fit the data for the lower frequencies but not for the higher.

6.4 Conclusions

We have proposed a joint reconstruction approach for salt delineation in seismic full-waveform inversion. Our approach is based on the idea of separating the model into

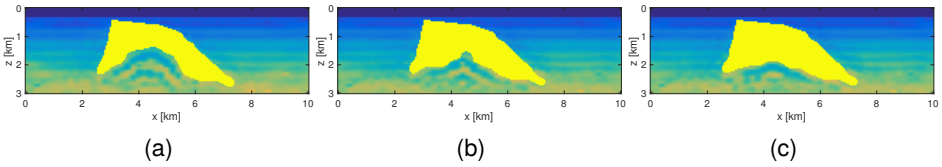


Figure 6.7: Reconstructed model after the 1st (a), 2nd (b) and 3rd (c) frequency pass.

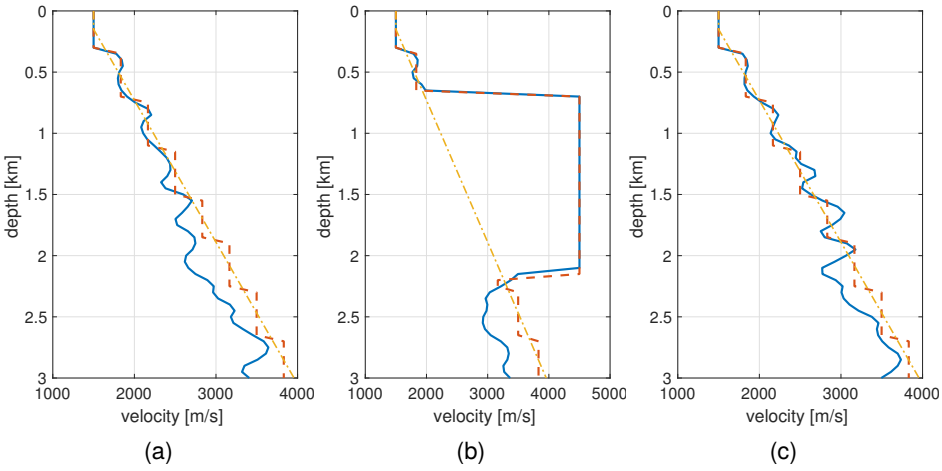


Figure 6.8: Velocity profile at $x = 2\,\text{km}$ (a), $x = 5\,\text{km}$ (b), and $x = 8\,\text{km}$ (c). The *blue* line denotes the reconstructed velocity, the *red dotted* line denotes the true velocity, while the *yellow dash-dotted* line represents the initial velocity.

salt and sediment. The salt geometry is defined with a level set represented by radial basis functions. The sediment, in turn, is represented by piecewise linear functions on a small number of nodes. This low-dimensional formulation of the model imposes smoothness on the sediment and on the salt geometries. The proposed compact approximation of the Heaviside function leads to faster convergence and produces no artifacts. We apply an alternating minimization strategy to optimize over the two different parameters. Results on a synthetic acoustic example demonstrate the method's capability. The proposed method accurately predicts the salt geometry where the conventional full-waveform inversion fails.

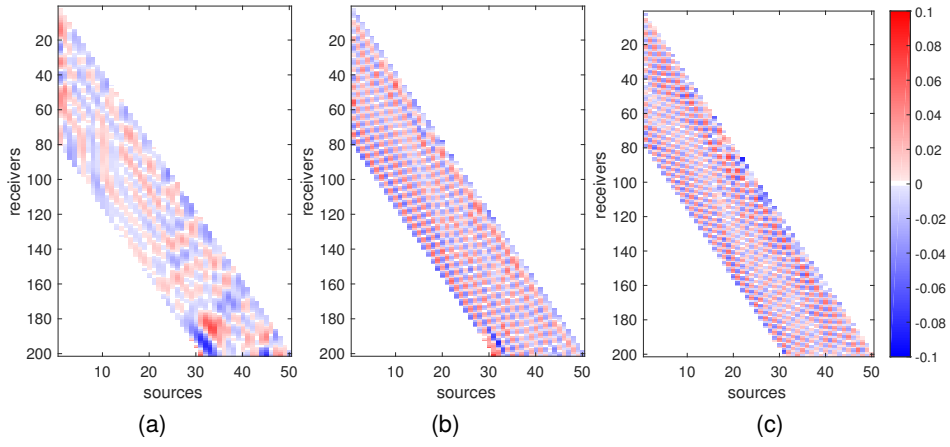


Figure 6.9: Normalized data misfit (only *real* part) for final reconstructed model at (a) 2.5 Hz, (b) 3.5 Hz, and (c) 4.5 Hz. We normalized the differences by the maximum value of the true data per frequency.

CHAPTER 7

High-contrast reflection tomography

Abstract - *Inverse scattering is the process of estimating the spatial distribution of the scattering potential of an object by measuring the scattered wavefields around it. We consider a limited-angle reflection tomography of high contrast objects that commonly occurs in ground-penetrating radar, exploration geophysics, terahertz imaging, ultrasound, and electron microscopy. Unlike conventional transmission tomography, the reflection regime is severely ill-posed since the measured wavefields contain far less spatial frequency information of the target object. We propose a constrained incremental frequency inversion framework that requires no side information from a background model of the object. Our framework solves a sequence of regularized least-squares subproblems that ensure consistency with the measured scattered wavefield while imposing total-variation and non-negativity constraints as a regularizer. We propose a proximal Quasi-Newton method to solve the resulting subproblem and devise an automatic parameter selection routine to determine the constraint in the regularizer of each subproblem. We validate the performance of our approach on synthetic low-resolution phantoms and with a non-inverse-crime test on a high-resolution phantom.*

This chapter is based on the following publications:

A Kadu, H Mansour, P T Boufounos, and D Liu. Reflection tomographic imaging of highly scattering objects using incremental frequency inversion. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7735–7739. IEEE, 2019.

" . . . in fact, the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity."
- R. Tyrrell Rockafellar

7.1 Introduction

Inverse scattering addresses the problem of reconstructing an image of the scattering potential of an object by probing it with electromagnetic or acoustic waves of finite bandwidth. An incident wavefield propagating inside the object induces multiple scattering of the waves that are generally measured on the boundary of the material. The scattered waves carry information about the spatial distribution of the scattering potential of the material, which has led to applications in numerous fields, such as, non-destructive testing [117], optical tomography [7], geophysical imaging [176, 204], ground penetrating radar [208], medical imaging [212, 88], and electron microscopy [95, 213, 16].

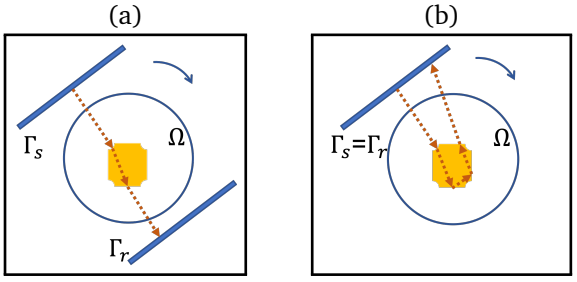


Figure 7.1: Two acquisition scenarios in inverse scattering, (a) Transmission, and (b) Reflection. Ω is the domain of interest, Γ_s is a source domain, and Γ_r is a receiver domain. A single experiment consists of - a set of sources in Γ_s sending a wavefield into Ω, and scattered wavefield being measured by a set of receivers in Γ_r. This experiment is repeated for various angles around Ω.

A scattering experiment consists of a transmission domain $\Gamma_s \subset \mathbb{R}^d$, an object domain $\Omega \subset \mathbb{R}^d$ and a receiver domain $\Gamma_r \subset \mathbb{R}^d$, where d is the dimension of the scene. A set of transmitters located in Γ_s sends waves into the scene that interact with an object in Ω. The interaction of the waves with the object leads to scattering of these waves. The scattered waves are then measured at the set of receivers located in Γ_r. Based on the location of transmission and receiver domain with respect to the object, we can classify the acquisition scheme into three different types: *(i) full-view*, where Γ_s and Γ_r surround the domain Ω; *(ii) transmission* mode, where Γ_s and Γ_r are located on opposite sides of the object; and *(iii) reflection* mode, where Γ_r and Γ_s co-located. Figure 7.1 illustrates these acquisition schemes. The *full-view* mode provides the most information about the spatial distribution of the object. The *transmission* mode offers less information than that of *full-view*, but it reduces the cost of the experiment due to the requirement of fewer transmitters and receivers. Tomographic imaging in this acquisition mode, known as *transmission tomography*, has found applications in many areas, for example, X-ray tomography in medicine

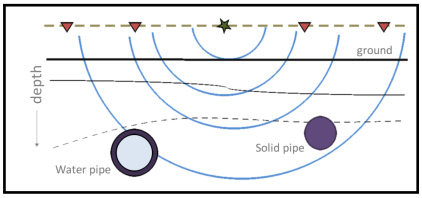


Figure 7.2: Acquisition scenario for underground imaging. A green star ($\star$) denotes the source and a triangle ($\bigtriangledown$) the receivers.

and non-destructive testing.

The *reflection* mode generally arises due to a limitation in the ability to access different sides of the material, as in the case of underground imaging, illustrated in Figure 7.2. We focus our presentation on the reflection tomography scenario where the problem is severely ill-posed. The ill-posedness arises due to restricted measurements and the limited availability of low spatial-frequency content in the measured wavefields (more discussion on this in Section 7.2.3). The underground imaging setup often appears in ground penetrating radar, seismic imaging, and ultrasound imaging.

The spatial scattering potential of a material can be described by its contrast level. The contrast indicates the power of interaction of a object material with a probing wave. A low-contrast material is semi-transparent, meaning that the interaction of the waves with it induces weak scattering. A high-contrast material strongly interacts with waves inducing *multiple* scattering events. In this chapter, we classify objects according to their contrast level, with a contrast below 1 being *low*, a contrast ranging from 1 to 10 being medium-contrast, and a contrast above 10 to be *high*. In general, the contrast varies with the frequency of the wave, but here we assume it to be independent of frequency.

7.1.1 Related Work

Numerous techniques have been proposed for solving the inverse scattering problem in the reflection regime. Earlier approaches dealt with iteratively linearizing the scattering model, using straight-ray theory, the Born approximation, and Rytov approximation [180, 31, 62, 20]. However, such linear models fail to account for the complex interaction between the wavefield and the material properties that result in multiple scattering. As a result, these methods require an accurate initial target model to enable the inversion and they generally suffer from poor reconstruction quality, especially when the material is inhomogeneous or contains highly scattering objects. Recently, the nonlinear interaction between the wavefield and the object has been incorporated into the inversion process using the wave equation (for example, [55, 11, 12, 29]). The inverse problem that deals with the wave-equation based scattering model is known as full-waveform inversion (FWI) [188, 148, 204]. FWI has been applied in multiple domains and across all modes of acquisition. A considerable amount of research has focused on full-view tomography and transmission tomography with FWI. Since we restrict ourself to

reflection tomography, we do not address the literature for other modes of tomography.

The research on the reflection tomography with FWI has mainly appeared in the geophysical community. Since the problem becomes nonlinear, the convergence of the inversion depends heavily on the initial model [134, 186]. Various new approaches have been proposed in the past decade to mitigate the effect of an initial model on the reconstructed solution [206, 200, 25, 201, 24]. These methods work well when the object does not have a high contrast. In the high-contrast regime, these methods are often complemented by the regularization [69, 71, 9]. These methods may require an additional set of parameters which may or may not be known apriori. Recently, the level-set approach, which incorporates quantitative information about the material in the object, is receiving attention [105, 121]. However, the level-set method requires a strong prior about the material, which may not be readily available.

Our work is inspired by the TV-regularization strategies proposed in [71]. Our contributions are different in the following three aspects: *(i)* We consider full-waveform inversion instead of its extended approach (the wavefield reconstruction inversion). Hence, we avoid estimating an extra hyperparameter. *(ii)* Instead of working with a batch approach for frequency, we keep the low frequencies in the cost function to avoid potential local minima. This makes our framework computationally more expensive. *(iii)* We work with a classical form of TV instead of its asymmetric version. Since the total variation parameter may be unknown (as also noted in [71]), we propose to estimate this value from the noise level in the data.

The sequential workflow has a long history in the geophysical imaging. It was introduced in [39] and named multiscale full-waveform inversion. We work with a regularized version of this multiscale approach. Since, we add one frequency at the time, as opposed to a frequency batch in [39], our approach is more robust against local minima (more discussions in Section 7.3.1).

7.1.2 Contributions and Outline

We develop an inversion framework for high-contrast limited-angle reflection tomography. Our contributions are three-fold:

- *Formulation*: We adopt a regularized sequential approach based on an incremental frequency. We keep the low frequencies in the cost function to avoid potential local minima. For a total of k frequencies in the data, we solve k constrained nonlinear least-squares problems sequentially.
- *Regularization and Optimization*: We introduce a combination of non-negative and total-variation regularization for the contrast function. Note that both regularizers are non-differentiable. To solve the regularized nonlinear least-squares problem, we propose a proximal Quasi-Newton (prox-QN) method. The proximal operations involved in the prox-QN are computed using primal-dual method.
- *Parameter Estimation*: We recommend a strategy to estimate the total-variation constraint parameter from the noise-level in the data.

We introduce the forward and the inverse scattering problem in Section 7.2. Here,

we also discuss the challenges of reflection tomography. In Section 7.3, we present the details of our sequential approach. We describe the regularization strategies and the optimization framework. We validate the proposed method on numerical phantoms and compare it with other methods in Section 7.5. We conclude the chapter in Section 7.6.

7.2 Inverse scattering problem

We begin by presenting the scattering model that describes the relationship between the wavefield and the contrast function. Next, we formulate the discrete inverse problem to reconstruct the contrast function from the set of the measured scattered wavefields. Finally, we discuss some challenges in estimating the contrast of an object in the reflection regime.

7.2.1 Forward problem

The *forward* scattering problem deals with constructing a map from a contrast function to the scattered waves measured at receivers. A wave-equation governs this map in the time-domain. For simplicity, we restrict ourself to scalar waves, but the map can naturally be constructed for other types of scattering problems with some modifications (see, for example, [31]).

Let us consider an object located in a bounded domain $\Omega \subset \mathbb{R}^d$, where $d = 2, 3$ denotes the dimension. The object has a spatial distribution of permittivity given by $\epsilon(\mathbf{r})$, where $\mathbf{r}$ denotes the spatial co-ordinate. The object lies in a homogeneous background (free space) of permittivity ϵ_b. We define the contrast function (or the relative permittivity) of the object as the difference of the permittivity of the object from the background, *i.e.*, $f(\mathbf{r}) = \epsilon(\mathbf{r}) - \epsilon_b$. We illuminate this object using the waves generated from a source function $q : \Gamma_t \to \mathbb{R}$. Subsequently, the scattered wavefield is measured inside the receiver domain Γ_r. Figure 7.1 illustrates the experimental setup.

The total wavefield $u : \Omega \to \mathbb{R}$ in the object domain Ω is related to contrast function f by an integral equation

$$u(\mathbf{r}) = u_{\text{in}}(\mathbf{r}) + k^2 \int_{\mathbf{r}' \in \Omega} g(\mathbf{r} - \mathbf{r}') u(\mathbf{r}') f(\mathbf{r}') \, \mathrm{d}\mathbf{r}' \qquad \forall \, \mathbf{r} \in \Omega, \tag{7.2.1}$$

where $g : \mathbb{R}^d \to \mathbb{R}$ is the Green function, $u_{\text{in}} : \mathbb{R}^d \to \mathbb{R}$ is an input wavefield, $k = 2\pi\omega/c$ represents the wave number in the vacuum, ω is the frequency and c denotes the speed of light in the vacuum. We assume that f is real, or in other words, the object is lossless. The equation (7.2.1) is known as the *Lippmann-Schwinger* equation. The input wavefield in equation (7.2.1) depends on the source function q as

$$u_{\text{in}}(\mathbf{r}) = k^2 \int_{\mathbf{r}' \in \Gamma_t} g(\mathbf{r} - \mathbf{r}') q(\mathbf{r}') \, \mathrm{d}\mathbf{r}' \qquad \forall \, \mathbf{r} \in \mathbb{R}^d. \tag{7.2.2}$$

Finally, the scattered wavefield measured in the receiver domain, $y : \Gamma_r \to \mathbb{R}$ is given by

$$y(\mathbf{r}) = \int_{\Omega} g(\mathbf{r} - \mathbf{r}') f(\mathbf{r}') u(\mathbf{r}') \, \mathrm{d}\mathbf{r}', \qquad \forall \, \mathbf{r} \in \Gamma_r. \tag{7.2.3}$$

We provide a detailed derivation of equations (7.2.1), (7.2.2), and (7.2.3) in Section 7.7.1. The forward problem finds the measurements y from the known source function q, the contrast function f, and the Green function g. In essence, it consists in first solving equation (7.2.2), then the *Lippmann-Schwinger* equation (7.2.1), and finally, the data equation (7.2.3). Generally, we pre-compute the input wavefield for each wavenumber k, since it is independent of the contrast function.

In the discrete setting, the scattering equation and data equation reduce to the following system of linear equations for each transmitter illumination and wave number:

$$\begin{aligned} \mathbf{u} &= \mathbf{v} + \mathbf{G}\,\mathbf{diag}\,(\mathbf{f})\,\mathbf{u}, \\ \mathbf{y} &= \mathbf{H}\,\mathbf{diag}(\mathbf{f})\mathbf{u}, \end{aligned} \tag{7.2.4}$$

where $\mathbf{u} \in \mathbb{C}^N$ and $\mathbf{v} \in \mathbf{C}^N$ are the total and input wavefields, respectively, N denotes the number of grid points used to discretize the domain Ω, $\mathbf{f} \in \mathbb{R}^N$ denotes the discretized contrast function, while $\mathbf{G} \in \mathbb{C}^{N\times N}$ and $\mathbf{H} \in \mathbb{C}^{n_r\times N}$ are the Green functions of the domain and receivers, respectively. Let n_r be the number of receivers that discretizes the receiver domain Γ, then $\mathbf{y} \in \mathbb{C}^{n_r}$ is the noise-free scattered wavefield measured at the receivers. The critical step in the forward problem involves estimating the wavefield $\mathbf{u}$ by inverting the matrix $\mathbf{A} := \mathbf{I} - \mathbf{G}\,\mathbf{diag}\,(\mathbf{f})$, where $\mathbf{I}$ denotes the identity operator. As the discretization dimension N increases, explicitly forming the matrix $\mathbf{A}$ and its inverse become prohibitively expensive. Therefore, a functional form of $\mathbf{A}$ along with the conjugate-gradient method (CG) are often used to perform the inversion. We note here that the convergence of CG depends on the conditioning of the operator $\mathbf{A}$, which becomes ill-conditioned for large wavenumber and high-contrast medium, *i.e.*, large value of $\|\mathbf{f}\|_\infty$.

Convolution operations (right-most term in the equation (7.2.1)) are an important step while computing the matrix-vector product $\mathbf{y} = \mathbf{A}(\mathbf{f})\mathbf{u}$. These operations are normally performed in the frequency domain, and the procedure is described as follows: *i)* We discretize the Green function on a grid with twice the grid points of the model. *ii)* We perform a circular shift to the Green function. *iii)* We obtain a contrast source from the pointwise multiplication of $\mathbf{u}$ and $\mathbf{f}$, and pad it with zeros from all sides to make it of the same size as the Green function. *iv)* We convolve the shifted Green's function with the contrast source by first taking its fast Fourier transform, and then taking an inverse Fourier transform. *v)* Finally, the top left part of the inverse Fourier transform gives the final output of the convolution.

7.2.2 Inverse problem

An *inverse scattering* problem is defined as the estimation of the contrast function given the measurement of scattered wavefield at n_r receivers for each input wavefield generated from n_t transmitters. Assuming that the measurements are contaminated by white Gaussian noise, we can formulate the discrete *inverse*

problem as follows:

$$\begin{aligned} \min_{\mathbf{f}} \quad & \sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}} \tfrac{1}{2}\|\mathbf{y}_{ij} - \mathbf{H}_j\,\mathbf{diag}(\mathbf{f})\mathbf{u}_{ij}\|^2, \\ \text{subject to} \quad & \left(\mathbf{I} - \mathbf{G}_j\,\mathbf{diag}\left(\mathbf{f}\right)\right)\mathbf{u}_{ij} = \mathbf{v}_{ij}\,\forall i \in \mathcal{I}, j \in \mathcal{J}, \end{aligned} \tag{7.2.5}$$

where $\mathcal{J} = \{1, \ldots, n_f\}$ and $\mathcal{I} = \{1, \ldots, n_r\}$ denote the index sets for frequencies and receivers respectively, and n_f represents the number of frequencies. We assume that the $\mathcal{J}$ is ordered according to the frequencies (in increasing order). For the rest of the chapter, $\|\cdot\|$ denotes the Euclidean norm (if there is no subscript). In general, problem (7.2.5) is ill-posed and admits multiple solutions. Therefore, spatial regularization in the form of a penalty function $\mathcal{R}(\mathbf{f})$ is often added to make the solution space smaller.

Let us introduce a data matrix $\mathbf{Y}_j \in \mathbb{R}^{n_r\times n_t}$, a wavefield matrix $\mathbf{U} \in \mathbb{C}^{N\times n_t}$ and the input wavefield matrix $\mathbf{V}_j \in \mathbb{C}^{N\times n_t}$ for $j \in \mathcal{J}$. Hence, for each frequency, the cost function and the constraint take the form

$$\begin{aligned} \mathcal{D}_j\left(\mathbf{f}, \mathbf{U}\right) &= \tfrac{1}{2}\|\mathbf{Y}_j - \mathbf{H}_j\,\mathbf{diag}(\mathbf{f})\mathbf{U}\|_2^2, \\ \mathbf{C}_j\left(\mathbf{f}, \mathbf{U}\right) &= \left(\mathbf{I} - \mathbf{G}_j\,\mathbf{diag}\left(\mathbf{f}\right)\right)\mathbf{U} - \mathbf{V}_j. \end{aligned}$$

It is possible to eliminate the wavefields $\mathbf{U}$ by satisfying the constraints, $\left(i.e., \mathbf{U}_j^\star = \left(\mathbf{I} - \mathbf{G}_j\,\mathbf{diag}(\mathbf{f})\right)^{-1}\mathbf{V}_j\right)$. Such reduced cost-function at frequency j is given by

$$\mathcal{F}_j(\mathbf{f}) \triangleq \left\{\mathcal{D}_j(\mathbf{f}, \mathbf{U}) \quad \text{subject to} \quad \mathbf{C}_j(\mathbf{f}, \mathbf{U}) = \mathbf{0}\right\}. \tag{7.2.6}$$

With the incorporation of this reduced form, the regularized version of (7.2.5) now reads

$$\min_{\mathbf{f}} \quad \sum_{j\in\mathcal{J}} \mathcal{F}_j\left(\mathbf{f}\right) + \mathcal{R}\left(\mathbf{f}\right). \tag{7.2.7}$$

This is a regularized least-squares optimization problem. The analytical solution is not available since the cost function, and the constraints are nonlinear. We resort to iterative methods to find a feasible solution to this problem.

7.2.3 Transmission vs reflection

A critical distinction between the transmission and reflection modes in inverse scattering manifests itself in the spatial frequency content that can be captured by the measured wavefields. In the transmission regime, the received measurements generally capture more of the lower spatial frequencies of the target distribution compared to the reflection regime. This is due to the fact that a probing pulse in the transmission mode is modulated by the complete object before reaching the receivers. On the other hand, the measured wavefields in the reflection mode are modulated by the discontinuities in the object permittivity that lead to reflections of the wavefields back to the receivers.

In order to illustrate the phenomenon described above, we simulate two sets of measurements $\{\mathbf{y}_T, \mathbf{y}_R\}$ of the scattered wavefield of the same object observed in the transmission and reflection modes through the measurement operators $\mathbf{H}_T$ and $\mathbf{H}_R$, respectively. The object is illuminated from its left side by a flat spectrum pulse containing 2, 3, and 5GHz frequency components. We want to identify the amount of spatial frequency content that is encoded in each of $\mathbf{y}_T$ and $\mathbf{y}_R$ without being affected by the nonconvexity of the inverse problem (7.2.5). Therefore, we provide the true scattered wavefields $\mathbf{U}_j^\star = (\mathbf{I} - \mathbf{G}_j \, \mathbf{diag}\,(\mathbf{f}^\star))^{-1} \mathbf{V}_j$ for each frequency, which reduces (7.2.5) to a convex linear inverse problem in $\mathbf{f}$. Consequently, we solve the convex form of (7.2.5) to compute $\mathbf{f}$ in each of the transmission and reflection modes and plot in Figure 7.3 the spatial frequency content (2D Fourier coefficients) of the reconstructed contrast $\mathbf{f}$ in each of the transmission and reflection modes. Notice how the recovered contrast in the reflection mode exhibits very little energy around the low spatial frequency subbands in the Fourier plane. This is in stark contrast to the transmission mode where a significant portion of spectral energy of the recovered contrast corresponds to the low spatial frequencies. The above illustration helps motivate the argument that the received measurements of the scattered wavefields in the reflection tomography mode encode very little spatial frequency information about the target object. Since the goal of tomographic imaging is to estimate the spatial distribution of the scattering potential of an object, the weak acquisition of spatial frequency information renders the problem severely ill-posed when compared to transmission tomography.

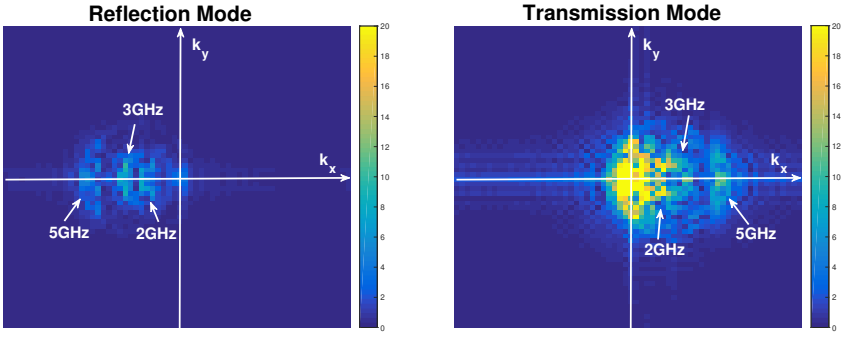


Figure 7.3: Comparison of the spatial frequency content of the received wavefields between the transmission mode and the reflection mode from a transmitted pulse containing 2GHz, 3GHz, and 5GHz frequency components.

7.3 Regularized multiscale approach

In this section, we present an incremental frequency inversion method that does not require a smooth initial model of the target image for successful recovery. We also discuss the regularization and the optimization strategy to solve the resulting problem.

7.3.1 Sequential workflow

The least-squares cost function in (7.2.7) provides a natural separation across frequencies. Moreover, the topology of the nonconvex cost function varies drastically between frequencies and can be leveraged to find a good local minimum. We illustrate this behavior using a simple cylindrical model for the target with a constant reflectivity c as shown in Figure 7.4(a). The true target has a reflectivity $c = 10$ and is illuminated with transmitters located at a y-position of -0.6m. We plot in Figure 7.4(b) the value of the data-fidelity cost function $\mathcal{F}_j(\mathbf{f})$ for various j values. Notice for the higher-frequency wavefields, the cost function starts to exhibit many local minima that are farther away from the global minimizer than for the low-frequency wavefields.

The most popular approach in the exploration geophysics community is to start with a low-frequency batch, and then slide linearly towards the high frequencies keeping the batch-size fixed. In Figure 7.4(c), we plot such cost function ($\sum_{j\in\mathcal{J}_\text{b}} \mathcal{F}_j(\mathbf{f})$ with $\mathcal{J}_\text{b} = \{j_0, j_0+1, \ldots, j_0+n_b-1\}$) for various frequency batches. We observe that the higher frequency batch has many local minima. The sliding approach works only when we get close to the global minimizer during a low-frequency batch inversion. A more robust approach would be to keep the low-frequencies as regularizer when inverting with high-frequency data. We plot the cost function ($\sum_{j=1}^{j_\text{max}} \mathcal{F}_j(\mathbf{f})$) in Figure 7.4(d). Notice that the cost functions are almost convex even when dealing with high-frequncy data.

The observations above led us to use an incremental frequency inversion framework where the model of the object's permittivity is sequentially updated as higher frequencies are included in the inversion. Given a measured wavefield containing n_f frequency components indexed in increasing order from 1 to n_f, our framework iteratively estimates the model from low to high-frequency while keeping the low-frequency cost function as a regularizer for high-frequency inversions.

$$\begin{aligned}&\text{for } k = 1, \ldots, n_f :\\ &\qquad \mathbf{f}^{(k)} \triangleq \underset{\mathbf{f}}{\operatorname{argmin}} \left\{ \sum_{j\in\mathcal{J}_k} \mathcal{F}_j(\mathbf{f}) + \mathcal{R}(\mathbf{f}) \right\}.\end{aligned} \tag{7.3.1}$$

Therefore, instead of solving a single nonconvex minimization problem in (7.2.7), we solve n_f subproblems sequentially according to (7.3.1), where the sequence of solutions moves us closer to the global minimizer of (7.2.7).

7.3.2 Regularization

Total-variation

The Total-Variation (TV) norm of a compactly supported function $u : \Omega \to \mathbb{R}$ is formally defined as

$$TV(u) \triangleq \sup \left\{ \int_\Omega u(\mathbf{x}) \, \text{div}\phi \, \mathrm{d}\mathbf{x} : \phi \in \mathcal{C}_c^1\left(\Omega, \mathbb{R}^d\right), \|\phi\|_\infty \leq 1 \right\},$$

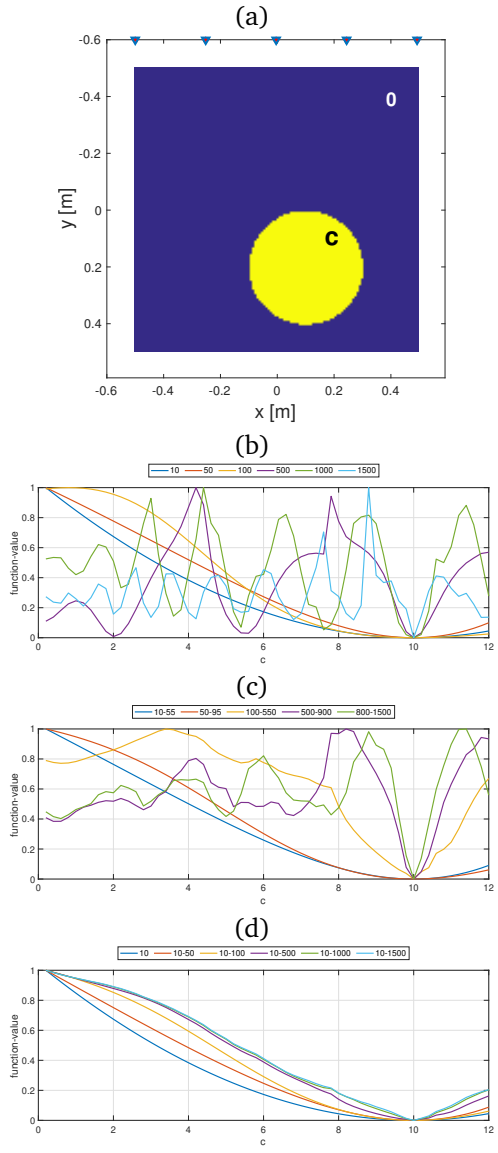


Figure 7.4: (a) Illustration of a cylindrical object with true reflectivity equal to $c^\star = 10$ measured by five co-located transmitters and receivers. Topology of the cost function per frequency (b), per frequency batch of size 10 (c), and incremental freqency batch (d) relative to the estimated reflectivity c.

where $\mathcal{C}_c^1\left(\Omega, \mathbb{R}^d\right)$ denotes the set of continuously differentiable functions of compact support contained in Ω. This norm measures the total change in the function over a finite domain [156]. If u is differentiable, then we can express the total-variation using an integral

$$TV(u) = \int_{\Omega} \|\nabla u(\mathbf{x})\|_1 \, \mathrm{d}\mathbf{x},$$

where $\|\cdot\|_1$ denotes the ℓ_1 norm or Manhattan norm. As a result, regularization with a TV norm promotes piecewise constant approximation of the true model [193]. In a discrete setting of two dimension, the TV-norm is represented as

$$TV(\mathbf{x}) = \|\mathbf{D}\mathbf{x}\|_1 \qquad \text{where} \quad \mathbf{D} = \begin{bmatrix} \mathbf{I}_x \otimes \mathbf{D}_y \\ \mathbf{D}_x \otimes \mathbf{I}_y \end{bmatrix}. \tag{7.3.2}$$

The $\mathbf{D}_x$ and $\mathbf{D}_y$ are the finite difference operators in x and y directions, and $\mathbf{I}_x, \mathbf{I}_y$ are the identity operators. We adopt the TV regularization in its constrained form [71], such that,

$$\begin{aligned} \mathcal{R}_{TV}(\mathbf{f}) &\triangleq \delta_{TV \leq \tau}(\mathbf{f}) \\ &\triangleq \delta_{\|\cdot\|_1 \leq \tau}(\mathbf{D}\mathbf{f}). \end{aligned} \tag{7.3.3}$$

where $\delta_{\mathcal{C}}(\cdot)$ is an indicator function to the set $\mathcal{C}$, and τ is a constraint parameter. The second line in (7.3.3) expresses the discretized version of the TV regularization function using the constrained ℓ_1-ball. We note here that the proximal for an indicator function to set $\mathcal{C}$ corresponds to the projection of a vector onto the set $\mathcal{C}$. Efficient algorithms exist for the projection onto the ℓ_1-norm ball (see, for example, [56]).

Non-negative Constraints

Since the contrast function is defined as the relative permittivity of an object (with respect to vacuum), it will always be non-negative. Hence, we impose this prior information (*i.e.*, the constraints) using a regularization defined as

$$\mathcal{R}_{\text{NN}}(\mathbf{f}) = \delta_{\geq 0}(\mathbf{f}),$$

where $\delta_{\geq 0}$ denotes the indicator to a non-negative orthant. This regularization function is convex, since the constraint set is convex. The proximal operator for this function corresponds to the projection of a vector onto a non-negative orthant. More specific, the projection operator is

$$\mathcal{P}_{\geq 0}(y) = \begin{cases} y & \text{if } y \geq 0 \\ 0 & \text{otherwise} \end{cases}.$$

Implementation

In order to impose the non-negative + TV constraints, we define the proximal operator:

$$\mathbf{prox}_{\gamma\mathcal{R}}(\mathbf{w}) \triangleq \underset{\mathbf{f}}{\operatorname{argmin}} \left\{ \frac{1}{2\gamma}\|\mathbf{f} - \mathbf{w}\|^2 + \mathcal{R}_{\text{TV}}(\mathbf{f}) + \mathcal{R}_{\text{NN}}(\mathbf{f}) \right\}, \tag{7.3.4}$$

with $\gamma > 0$. The proximal operator becomes a projection onto the intersection of the sets, the TV-norm ball and the non-negative orthant. Although there is no simple analytical expression for this proximal operator, it can be evaluated efficiently using various splitting methods, *e.g.*, the alternating direction method of multipliers (ADMM) [33] and/or primal-dual method [49]. Here, we work with the primal-dual method. We derive this method for the sum of three convex functions in Section 7.7.3. Algorithm 10 describes the primal-dual method to solve (7.3.4).

Algorithm 10 Proximal for Non-negative + Total-Variation

Input: $\mathbf{w} \in \mathbb{R}^n, \mathbf{D} \in \mathbb{R}^{m \times n}, \gamma > 0, \tau > 0, t_{\text{max}}$
Output: $\mathbf{f}^\star$

$\mathbf{f}^{(0)} = \mathbf{0}, \mathbf{u}^{(0)} = \mathbf{0}, \mathbf{v}^{(0)} = \mathbf{0}$
choose $\alpha \in \left(0, 1/\sqrt{\rho(\mathbf{D}^T\mathbf{D} + \mathbf{I})}\right)$
while $t < t_{\text{max}}$ **do**
  $\hat{\mathbf{f}} = \mathbf{f}^{(t)} - \alpha\left(\mathbf{D}^T\mathbf{u}^{(t)} + \mathbf{v}^{(t)}\right)$
  $\mathbf{f}^{(t+1)} = \left(\gamma\hat{\mathbf{f}} + \alpha\mathbf{w}\right) / (\alpha + \gamma)$
  $\hat{\mathbf{u}} = \mathbf{u}^{(t)} + \alpha\mathbf{D}\left(2\mathbf{f}^{(t+1)} - \mathbf{f}^{(t)}\right)$
  $\mathbf{u}^{(t+1)} = \mathbf{u}^{(t)} - \alpha\mathcal{P}_{\|\cdot\|_1 \leq \tau}\left(\hat{\mathbf{u}}/\alpha\right)$
  $\hat{\mathbf{v}} = \mathbf{v}^{(t)} + \alpha\left(2\mathbf{f}^{(t+1)} - \mathbf{f}^{(t)}\right)$
  $\mathbf{v}^{(t+1)} = \mathbf{v}^{(t)} - \alpha\mathcal{P}_{\geq 0}\left(\hat{\mathbf{v}}/\alpha\right)$
  $t = t + 1$
end while

7.3.3 Proximal quasi-Newton method

Let's consider a subproblem in (7.3.1):

$$\mathbf{f}^\star = \underset{\mathbf{f}}{\text{argmin}} \left\{ \sum_{j \in \mathcal{J}_k} \mathcal{F}_j\left(\mathbf{f}\right) + \mathcal{R}\left(\mathbf{f}\right) \right\}. \tag{7.3.5}$$

To solve the above problem, we propose a proximal Quasi-Newton (prox-QN) method. The method is derived in Section 7.7.4. Algorithm 11 enumerates the steps of prox-QN for problem (7.3.5).

Algorithm 11 Prox-QN method for solving (7.3.5)

Input: $\mathbf{f}^{(0)}, \tau > 0, \gamma \in (0,1)$
Output: $\mathbf{f}^{\star}$
1: **for** $i = 0$ to i_{max} **do**
2: compute the gradient $g_i = \sum_{i=1}^{k} \nabla \mathcal{F}_i \left(\mathbf{f}^{(i)}\right)$.
3: compute the approximate Hessian $\mathbf{H}_i$
4: solve $\mathbf{s}_i$ from equation (7.7.24)(a)
5: define $\hat{\mathbf{f}}(\alpha) = \mathbf{prox}_{\gamma \mathcal{R}} \left(\mathbf{f}^{(i)} + \alpha \mathbf{s}_i\right)$
6: $\alpha_i = \text{linesearch}_{\alpha} \left(\hat{\mathbf{f}}(\alpha)\right)$
7: $\mathbf{f}^{(i+1)} := \mathbf{prox}_{\gamma \mathcal{R}} \left(\mathbf{f}^{(i)} + \alpha_i \mathbf{s}_i\right)$
8: check optimality conditions
9: **end for**

The algorithm consists of an inner and an outer loop. The outer loop computes the next iterate based on the search direction and step length, while the inner loop finds the search direction. At every sequence of the outer loop, we first compute the gradient using an adjoint-state method, and form the approximate Hessian with the L-BFGS procedure. The procedure to compute the gradient is explained in Section 7.7.2. Once we have the gradient and approximate Hessian at the current iterate, we search for the descent direction using the primal-dual method (see (7.7.27)). Next, we search for the feasible step length using the backtracking linesearch. Finally, we compute the next iterate from the optimal search direction and the feasible step length.

7.4 Estimating the constraint parameter τ

Recall that for each subproblem (7.3.5) in our proposed framework, we are solving a TV-constrained nonlinear least-squares problem where the constraint parameter τ_k should bound the ℓ_1 ball of the total variation of the solution. Naturally, the choice of constraint parameter τ_k would significantly affect the reconstruction performance.

In order to estimate τ_k for each new subproblem, we develop a parameter estimation routine inspired by the approach in [197] for sparse optimization with linear least squares constraints. Suppose that we have an initial estimate of $\mathbf{f}^k$ obtained at the frequency corresponding to the kth subproblem for which the TV norm $\tau_k = TV(\mathbf{f}^k)$, specifically,

$$\mathbf{f}^k = \underset{\mathbf{f}}{\text{argmin}} \left\{ \sum_{j \in \mathcal{J}_k} \mathcal{F}_j \left(\mathbf{f}\right) \text{ s.t. } ||\mathbf{D}\mathbf{f}||_1 \leq \tau_k \right\}, \tag{7.4.1}$$

where $\mathcal{F}_j$ is as defined in (7.2.6), and the constraints $\mathbf{C}_j(\mathbf{f}^k, \mathbf{U}_j) = \mathbf{0}$ are satisfied for all $j \in \mathcal{J}_k$. At subproblem $k+1$, the cost $\mathcal{F}_{k+1}\left(\mathbf{f}\right)$ is added to the objective function, resulting in the potentially unsatisfied constraint

$$\mathbf{V}_{k+1} = \mathbf{A}_k \mathbf{U}_{k+1}, \tag{7.4.2}$$

where $\mathbf{A}_k \triangleq \mathbf{I} - \mathbf{G}\,\mathbf{diag}(\mathbf{f}_k)$. To overcome the nonconvexity of the objective function due to (7.4.2), we linearize the objective function around $\mathbf{f}^k$ by estimating $\mathbf{U}^\star_{k+1} = \mathbf{A}_k^{-1}\mathbf{V}_{k+1}$, thus reducing $\mathcal{F}_{k+1}(\mathbf{f})$ to a convex least squares cost function in $\mathbf{f}$, i.e.,

$$\mathcal{F}_{k+1}(\mathbf{f}) \approx \mathcal{D}_{k+1}\left(\mathbf{f}, \mathbf{U}^\star_{k+1}\right),$$

where $\mathcal{D}_{k+1}\left(\mathbf{f}, \mathbf{U}^\star_{k+1}\right)$ is the data mismatch cost function defined in (7.2.6). Consequently, we may now define a value function $\Phi(\tau)$ for the $k+1$ subproblem as

$$\begin{aligned}\Phi(\tau) &= \underset{\mathbf{f}}{\operatorname{argmin}} \left\{ \sum_{j\in\mathcal{J}_{k+1}} \mathcal{D}_j\left(\mathbf{f}, \mathbf{U}^\star_j\right) \text{ s.t. } ||\mathbf{D}\mathbf{f}||_1 \leq \tau \right\} \\ &= \underset{\lambda}{\operatorname{argmax}} \left\{ \frac{\sum_{j\in\mathcal{J}_{k+1}} \mathbf{r_j}^H \mathbf{Y}_j}{\|\mathbf{r}^{k+1}\|} - \tau\lambda \right. \\ &\qquad \left. \text{s.t. } \mathrm{TV}_{\text{polar}}\left(\frac{\sum_{j\in\mathcal{J}_{k+1}} \mathbf{diag}\left(\mathbf{U}^\star_j\right)\mathbf{H}_j^H \mathbf{r_j}}{\|\mathbf{r}^{k+1}\|} \right) \leq \lambda \right\}\end{aligned} \tag{7.4.3}$$

where $\mathbf{r}_j = \mathbf{Y}_j - \mathbf{H}_j\,\mathbf{diag}\left(\mathbf{U}^\star_j\right)\mathbf{f}_k$ is the data residual at the jth frequency, and $\mathbf{r}_{k+1}$ is the vector formed by concatenating all the vectors $\mathbf{r}_j$, such that, $\|\mathbf{r}^{k+1}\| = \sqrt{\sum_{j\in\mathcal{J}_{k+1}} ||\mathbf{r}_j||^2}$. The $\mathrm{TV}_{\text{polar}}$ function is defined as $\mathrm{TV}_{\text{polar}}(\mathbf{x}) = \|\mathbf{D}^{-T}\mathbf{x}\|_\infty$, with $\mathbf{D}^{-T}$ being the transposed pseudo-inverse of the finite difference operator $\mathbf{D}$ defined in (7.3.2). Note that (7.4.3) shows the primal and dual problems for computing the value function $\Phi(\tau)$.

The dual problem in (7.4.3) conveniently shows that the maximum is achieved when λ is at its minimum $\lambda^\star = \mathrm{TV}_{\text{polar}}\left(\frac{\sum_{j\in\mathcal{J}_{k+1}} \mathbf{diag}\left(\mathbf{U}^\star_j\right)\mathbf{H}_j^H \mathbf{r_j}}{\|\mathbf{r}\|} \right)$. Moreover, the gradient of $\Phi(\tau)$ with respect to τ is easily computed as $\nabla_\tau \Phi(\tau) = \lambda^\star$. Therefore, we can compute the update for τ using a Newton root finding step, such that,

$$\tau_{k+1} = \tau_k + \frac{\|\mathbf{r}^{k+1}\|\left(\|\mathbf{r}^{k+1}\| - \sigma_{k+1}\right)}{\mathrm{TV}_{\text{polar}}\left(\sum_{j\in\mathcal{J}_{k+1}} \mathbf{diag}\left(\mathbf{U}^\star_j\right)\mathbf{H}_j^H \mathbf{r_j} \right)}, \tag{7.4.4}$$

where σ_{k+1} is the upper bound on the ℓ_2 norm of the noise up to the $k+1$ frequency bin. Finally, we note that at the zeroth iteration, the parameter τ can be set to zero resulting in a homogeneous solution for $\mathbf{f}_0$.

7.5 Numerical experiment

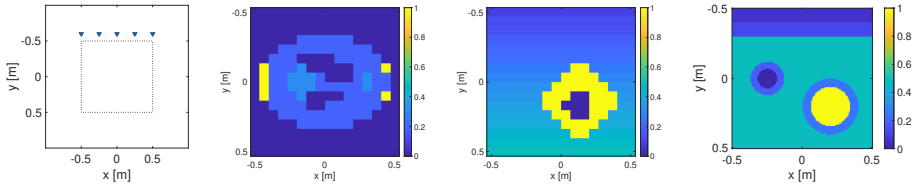
In this section, we describe the experimental setup for the reflection tomography. We evaluate our method on two numerical phantoms and compare it with two other approaches. We also experiment with a partially non-inverse-crime dataset towards the end.

7.5.1 Experimental details

We consider an experimental setup illustrated in Figure 7.5(a). The domain is 1 m $\times$ 1 m and extends in x-direction from $x = -0.5$ m to 0.5 m and in y-direction from $y = -0.5$ m to 0.5 m. There are total of five transmitters and receivers located on a line $y = -0.6$ m. Each transmitter illuminates a flat spectrum pulse occupying the frequency band $[10, 2000]$ MHz. All 5 receivers are activated for each transmitter. We consider three frequency bands: *i)* a low frequency band consisting of $\{10+5j\}$ MHz with $j = 0, \ldots, 17$, *ii)* a medium frequency band consisting of $\{100 + 50j\}$ MHz with $j = 0, \ldots, 17$, and *iii)* a high frequency band consisting of $\{1000 + 100j\}$ MHz with $j = 0, \ldots, 10$. Hence, in total, we consider 47 frequencies between 10 MHz and 2000 MHz.

We work with 3 phantoms shown in Figure 7.5(b)-(d). All phantoms have a length of 1 m in both x and y directions. Phantom1 is a Shepp-Logan phantom which resembles the brain. It is a well-known phantom in the image processing and tomography community. Here, we discretize it on a 16×16 grid. It has a total of 4 contrast values, namely $\{0, 0.2, 0.3, 1\}$. Phantom 2 resembles an underground scene. It has layer structure in the background whose contrast ranges from 0.1 to 0.5. A square-type hole (of contrast of 0) is embedded in a rhombus-type structure with a contrast of 1. This phantom also has a resolution of 16×16. We use these low resolution phantoms to compare our method with other exisiting methods and to check the robustness against the noise.

Phantom 3 is a high-resolution phantom depicting another underground scene. It has a resolution of 128×128. It contains 3 horizontal layers of contrast $\{0.05, 0.1235, 0.5\}$. The phantom consists of two circular pipes of outer diameter 0.4 m and 0.24 m with a thickness of 0.6 m and 0.5 m respectively. A large pipe has an inner region filled with a high contrast material of permittivity 1 and a small pipe has a vaccum inside. We use this phantom to perform a partially non-inverse-crime test as described in Section 7.5.5.



(a) Setup (b) Phantom 1 (c) Phantom 2 (d) Phantom 3

Figure 7.5: (a) Reflection tomography setup for all the numerical experiments. The dotted region denotes the object domain Ω. The transmitters and receivers are collocated at $y = -0.6$ m. (b), (c), (d) are the three numerical phantoms used for the experimentation.

7.5.2 Comparison with other methods

We restrict ourself to the two classical methods. For fair comparison we modify these methods to add the prescribed regularization. We do not compare with linearized methods like Born approximation and Rytov approximation as these methods have

shown to fail for high-contrast imaging [124].

CISOR : The CISOR algorithm aims to solve the equation described in (7.2.5) by taking all frequencies at once [124]. As opposed to penalization of TV-norm, we use the proposed regularization. We regularized it with total-variation and non-negative constraints. The problem is solved using a prox-QN method with a maximum of 5000 iterations or until the convergence is reached.

RL : Recursive linearization (RL) method was introduced in [51], and has been a standard while working with multi-frequency data. The method enjoys the computational benefit of solving a single constraint at the time, but might suffer in the high-contrast regime. It solves the sequence of problems

$$\mathbf{f}^{(j)} \triangleq \underset{\mathbf{f}}{\operatorname{argmin}} \Big\{ \mathcal{D}_j(\mathbf{f}, \mathbf{U}_j) \quad \text{subject to} \quad \mathbf{C}_j(\mathbf{f}, \mathbf{U}_j) = \mathbf{0} \Big\},$$

with an initial guess to each subproblem being the solution of the previous subproblem. We modify the cost function to include the regularization. Similar to the CISOR, we consider a non-negative + total-variation regularization. Each subproblem is solved using a prox-QN method with a maximum of 500 iterations or until the convergence is reached.

SF-τ : This method corresponds to the proposed sequential framework with known τ value. It solves the problem described in (7.3.1). We use a prox-QN method to solve each subproblem with a maximum of 500 iterations or until the convergence is reached.

SF-σ : This method corresponds to the proposed sequential framework with estimation of τ at each iteration. It solves the problem described in (7.3.1) with the τ estimation from (7.4.4). Here, we assume that the noise-level σ is known. We use a prox-QN method to solve each subproblem with a maximum of 500 iterations or until the convergence is reached.

For all the methods the initial model corresponds to a contrast of 0 everywhere.

7.5.3 Performance Measures

We use the following measures to test the performance of the proposed methods, and for comparison with other methods.

DR : The data residual (DR) measures the distance of the modeled data for the reconstructed model with the actual data in the Euclidean sense. For multi-frequency data the DR takes the following form

$$\text{DR} \triangleq 100 \times \frac{\sum_{j \in \mathcal{J}} \mathcal{F}_i(\mathbf{f}^\star)}{\sum_{j \in \mathcal{J}} \|\mathbf{Y}_j\|^2},$$

where $\mathbf{f}^\star$ is the reconstructed solution. Here, $\|\mathbf{Y}\|$ denotes the Frobenius norm for the matrix $\mathbf{Y}$. DR must be close to the noise-level for a method to be considered good.

SNR : The signal-to-noise ratio (SNR) for the reconstructed model $\mathbf{f}^\star$ with respect to the ground truth $\mathbf{f}^{\text{true}}$ is

$$\text{SNR} \triangleq -20 \log_{10} \left(\frac{\|\mathbf{f}^\star - \mathbf{f}^{\text{true}}\|}{\|\mathbf{f}^{\text{true}}\|} \right).$$

Again, this measure is only available if we know the ground truth. However, a reconstruction is considered good if it has high SNR.

7.5.4 Small-scale experiments

We mainly perform two kinds of experiments: *(i)* The noise-free test, and *(ii)* noisy test. In the noise-free test, we compare our methods with the other two methods (CISOR and RL). In the noisy test, we only look at the robustness of our methods against various levels of noise.

Noise-free test

We consider Phantom 1 and 2 for this test. We produce three types of phantoms by scaling these phantoms with a maximum contrast ($\mathbf{f}_{\text{max}}$) of $\{1, 10, 100\}$. The contrast of 1 is a low-contrast phantom. Medium contrast phantoms have a maximum contrast of 10. The phantom with maximum contrast 100 is a high-contrast phantom. We consider the reflection tomography setup illustrated in Figure 7.5(a) with the noiseless data. We check the performance of the methods SF-τ and SF-σ, and compare it with the CISOR and RL method. Figure 7.6 and Figure 7.7 show the reconstructions for various contrast levels for Phantom 1 and 2. We see that SF-τ consistently performs well except in the case of $\mathbf{f}_{\text{max}} = 100$ for Phantom 1, where all the methods fail. The reason for the failure is that Phantom 1 is ideal for the *transmission* or *full-view* tomography and not for *reflection* tomography. For an underground scene (depicted by Phantom 2) we see that the proposed methods performs well with the *reflection* tomography. We tabulate the values for the performance measures in Table 7.1 and 7.2. We conclude that the SF-τ and SF-σ performs superior to the existing methods (CISOR and RL).

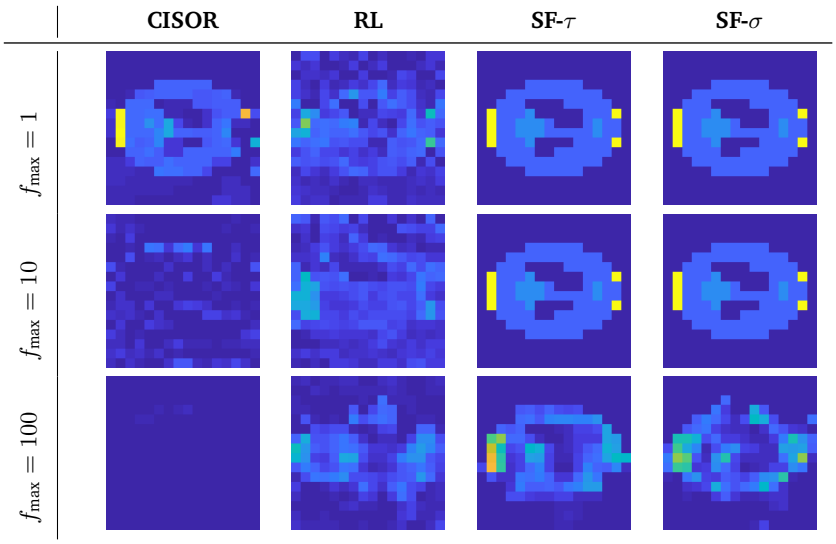


Figure 7.6: Phantom 1. Comparison of proposed methods with other methods

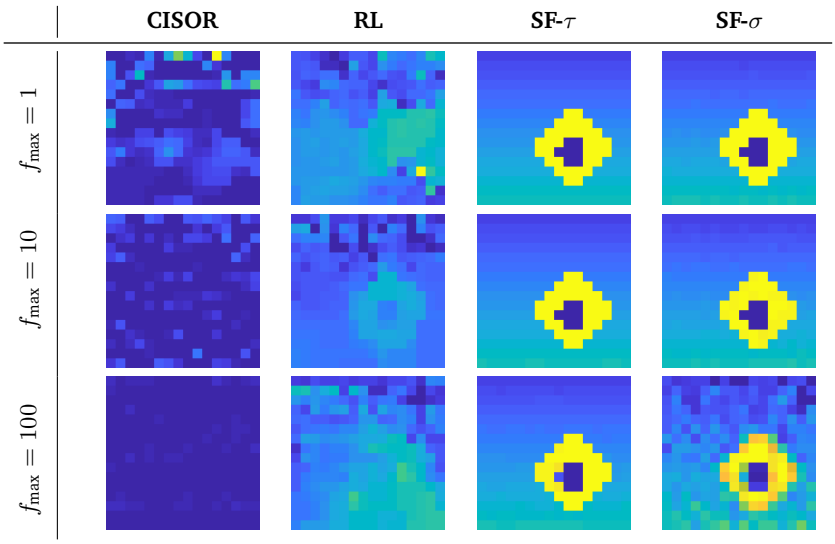


Figure 7.7: Phantom 2. Comparison of proposed methods with other methods

Table 7.1: Comparison of methods on Phantom 1

f_{max}	Measure	CISOR	RL	SF-τ	SF-σ
1	DR	28.24	91.59	0.06	0.03
	SNR	7.91	4.47	64.81	64.61
10	DR	58.68	151.20	0.36	1.04
	SNR	0.40	2.68	65.09	52.66
100	DR	456.04	85.92	7.95	19.84
	SNR	0.11	2.96	4.69	3.44

Table 7.2: Comparison of methods on Phantom 2

f_{max}	Measure	CISOR	RL	SF-τ	SF-σ
1	DR	442.35	131.80	0.48	1.54
	SNR	0.76	4.99	63.65	44.32
10	DR	334.22	160.74	0.59	4.13
	SNR	0.34	5.06	67.83	34.90
100	DR	449.01	120.20	0.59	14.13
	SNR	0.12	5.95	31.08	14.18

Noisy test

We consider Phantom 1 and 2, with the scaling $\{1, 10, 100\}$. We add a Gaussian noise of strengths 5%, 10%, and 20%. We check the performance of the methods SF-τ and SF-σ on these noise levels. Figure 7.8 and 7.9 show the reconstructions of these methods on various levels of noise and various levels of contrast values. The performance measures are tabulated in Table 7.3 and 7.4. We observe that SF-τ and SF-σ are robust against high noise in the low-contrast phantoms. SF-τ is also stable for moderate level of noise in the high-contrast regime.

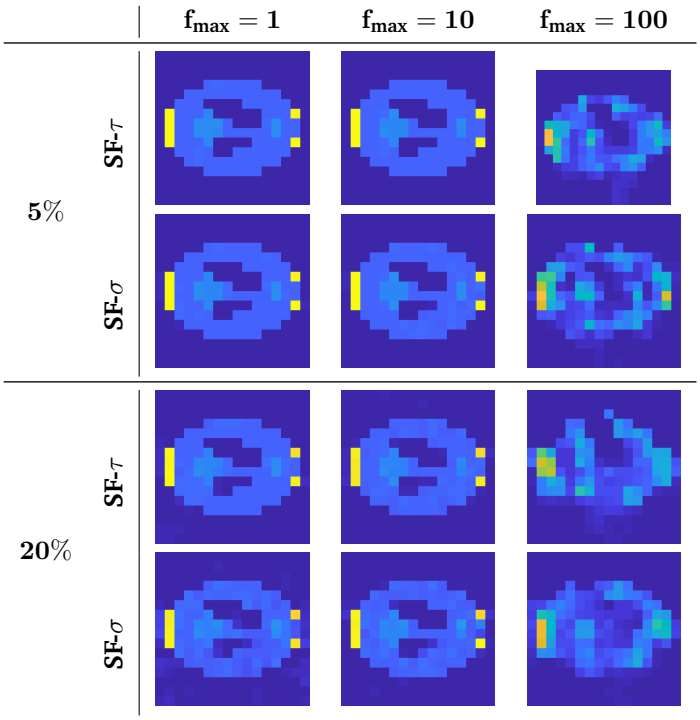


Figure 7.8: Phantom 1. Noise robustness of proposed methods

7.5.5 Partially Non-inverse-crime test

We consider Phantom 3 for this test that has a resolution of 128×128. To avoid an inverse crime, we discretize the model on a high-resolution grid of size 192×192. We use the nearest-neighbor algorithm for the rescaling to a high-resolution grid. We generate the data on the high-resolution grid. As a sanity check, we look at the difference between the data for high-resolution and low-resolution model, and found the relative difference of 20% for a few high frequencies. We test SF-τ and SF-σ methods with this high-resolution dataset. We assume a noise-level of 20% for SF-σ, while we set τ to be the TV-value of ground truth (low-resolution model) for SF-τ. The reconstruction results for SF-τ and SF-σ are presented in Figure 7.10. SF-τ has DR of 0.35, MR of 9.08 and SNR of 20.84, while SF-σ has a DR of 11.52,

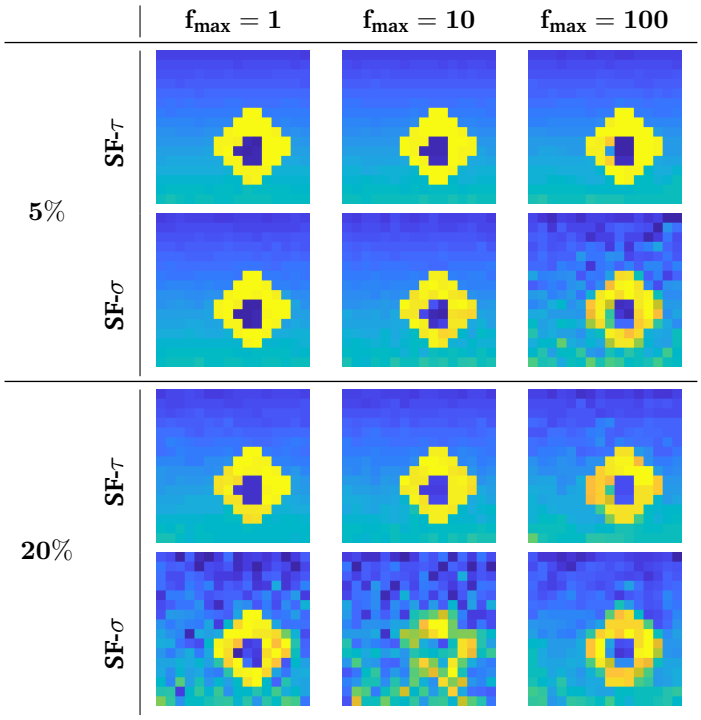


Figure 7.9: Phantom 2. Noise robustness of proposed methods

Table 7.3: Noise-Robustness of SF-τ and SF-σ on Phantom 1

f_{max}	Measure	SF-τ			SF-σ		
		5%	10%	20%	5%	10%	20%
1	DR	4.69	9.41	18.99	4.74	9.70	20.57
	SNR	37.02	30.98	25.11	33.15	24.28	18.35
10	DR	5.44	10.22	25.38	6.32	12.37	20.75
	SNR	35.16	30.93	23.41	31.30	25.17	21.54
100	DR	6.87	16.43	29.06	16.69	17.66	24.26
	SNR	4.35	4.02	3.22	4.68	4.93	4.76

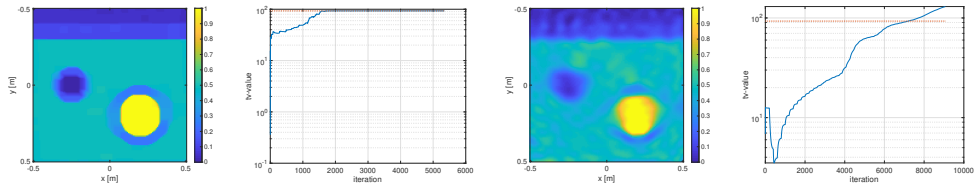
Table 7.4: Noise-Robustness of SF-τ and SF-σ on Phantom 2

f_{max}	Measure	SF-τ			SF-σ		
		5%	10%	20%	5%	10%	20%
1	DR	4.91	9.19	18.06	6.60	27.10	68.67
	SNR	34.13	28.85	25.58	32.41	19.50	13.86
10	DR	6.55	9.74	17.63	7.65	22.14	30.00
	SNR	33.61	29.62	24.52	24.51	15.79	7.87
100	DR	4.65	9.27	17.76	11.29	11.40	19.86
	SNR	24.75	19.65	16.54	14.14	13.82	13.10

MR of 18.60 and SNR of 14.61. We observe that the SF-τ is able to reconstruct the ground scene quite accurately. The top and the bottom regions of the pipes are retrieved to a good precision, while the method suffers from reconstructing the left and right sides of these pipes (see Figure 7.10(a)). SF-σ is able to locate the high-contrast and the low-contrast objects in the pipes but fails to get the boundary of the pipes accurately. Figure 7.10(b) and (d) are the plots for the TV-values of the models generated during the iterative process. Starting with a low TV-value model, SF-τ generates a sequence of models with the next model having the higher or the same TV-value as the previous model. This does not hold for the SF-σ procedure. We see that it generates a model with low TV-value for low-frequency batches. The SF-σ generated sequence also overshoots the true TV-value during high-frequency batches (corresponds to high iteration number).

7.6 Conclusions

We consider limited-angle reflection tomography of high-contrast objects. We show that the tomography problem is severely ill-posed due to the absence of low-frequency content and multiple scattering of waves. To find a feasible solution to this ill-posed problem, we developed a regularized multiscale approach. We pose



(a) SF-τ (b) TV (c) SF-σ (d) TV

Figure 7.10: (a) and (c) are the reconstruction results for SF-τ and SF-σ respectively. (b) and (d) are the evolution of TV-value of the generated sequence of iterates for the SF-τ and SF-σ. The red dotted line shows the TV-value of the ground truth.

the imaging problem as a nonlinear least-squares problem with constraints. The cost function includes the wave-based modeling that accounts for multiple scattering. The regularization, represented in the form of constraints, includes non-negativity and total variation. The total cost function is decomposed according to the frequency, and we observe that the low-frequencies bring smoothness while higher frequencies add details in the reconstruction. Hence, we solve a sequence of subproblems, where the k^{th} subproblem has a constrained cost function measured over the first k frequencies. We propose a proximal-Quasi-Newton method to solve the resulting constrained problem. The underlying proximal operations are performed using a primal-dual approach. We propose an automatic strategy to update the TV-constraint parameter based on the noise-level in the data. Through numerical experiments, we demonstrate that our methodologies outperform the existing methods and are robust against moderate noise. The proposed techniques can retrieve high-contrast object (contrast up to 100) for scenes similar to the underground.

7.7 Appendices

7.7.1 Scattering formalism

Consider a scattering setup illustrated in Figure 7.1. The scene (free-space with permittivity ϵ_b) has a dimension d. Transmitter domain $\Gamma_s \subset \mathbb{R}^d$ sends a pulse with a source function $q : \Gamma_s \to \mathbb{C}$. It generates an incident wavefield $u_{\text{in}} : \mathbb{R}^d \to \mathbb{C}$ everywhere. This incident wavefield interacts with an object in domain $\Omega \subset \mathbb{R}^d$ and generates a total wavefield $u : \mathbb{R}^d \to \mathbb{C}$. The scattered wavefield $u_{\text{sc}} : \mathbb{R}^d \to \mathbb{C}$ is then measured in the receiver domain $\Gamma_r \subset \mathbb{R}^d$.

The total wavefield is a superposition of an incident field $u_{\text{in}}(\mathbf{r})$ and a scattered field $u_{\text{sc}}(\mathbf{r})$,

$$u(\mathbf{r}) = u_{\text{in}}(\mathbf{r}) + u_{\text{sc}}(\mathbf{r}), \qquad \mathbf{r} \in \mathbb{R}^d. \tag{7.7.1}$$

The incident wavefield is the field in the absence of the scatterer, while the scattered field takes the presence of object into account. The incident wavefield satisfies the Helmholtz equation

$$\nabla^2 u_{\text{in}}(\mathbf{r}) - k^2 \epsilon_b u_{\text{in}}(\mathbf{r}) = -q(\mathbf{r}) \qquad \forall \mathbf{r} \in \mathbb{R}^d,$$

where k denotes the wavenumber. It is convenient to consider the above equation for inside and outside the object domain Ω:

$$\begin{aligned} \nabla^2 u_{\text{in}}(\mathbf{r}) - k^2 \epsilon_b u_{\text{in}}(\mathbf{r}) &= 0 \qquad \forall \mathbf{r} \in \Omega, \\ \nabla^2 u_{\text{in}}(\mathbf{r}) - k^2 \epsilon_b u_{\text{in}}(\mathbf{r}) &= -q(\mathbf{r}) \qquad \forall \mathbf{r} \notin \Omega, \end{aligned} \tag{7.7.2}$$

Similarly, the total wavefield satisfies the Helmholtz equation, and we can express it inside and outside the domain as follows,

$$\begin{aligned} \nabla^2 u(\mathbf{r}) - k^2 \epsilon(\mathbf{r}) u(\mathbf{r}) &= 0 \qquad \forall \mathbf{r} \in \Omega, \\ \nabla^2 u(\mathbf{r}) - k^2 \epsilon_b u(\mathbf{r}) &= -q(\mathbf{r}) \qquad \forall \mathbf{r} \notin \Omega, \end{aligned} \tag{7.7.3}$$

where $\epsilon(r)$ is the permittivity of the object. Now, from the equations (7.7.1), (7.7.2) and (7.7.3), the governing equation for the scattered wavefield reads

$$\begin{aligned} \nabla^2 u_{\text{sc}}(\mathbf{r}) - k^2 \epsilon_b u_{\text{sc}}(\mathbf{r}) &= -k^2 \left(\epsilon_b - \epsilon(\mathbf{r})\right) u(\mathbf{r}) \qquad \forall \mathbf{r} \in \Omega, \\ \nabla^2 u_{\text{sc}}(\mathbf{r}) - k^2 \epsilon_b u_{\text{sc}}(\mathbf{r}) &= 0 \qquad \forall \mathbf{r} \notin \Omega, \end{aligned}$$

These equations can be compactly written as

$$\nabla^2 u_{\text{sc}}(\mathbf{r}) - k^2 \epsilon_b u_{\text{sc}}(\mathbf{r}) = -k^2 f(\mathbf{r}) u(\mathbf{r}) \qquad \forall \mathbf{r} \in \mathbb{R}^d \tag{7.7.4}$$

where $f(\mathbf{r})$ is a *contrast function* that is equal to the difference between the permittivity, $\epsilon(r) - \epsilon_b$, inside the object domain Ω and 0 outside. Another interesting quantity to note is the contrast source $q_{\text{sc}}(\mathbf{r}) = f(\mathbf{r}) u(\mathbf{r})$. The contrast source acts as a modified source term contributing to the generation of scattered wavefield. We supplement the scattered wavefield equation (7.7.4) with the Sommerfeld radiation condition

$$\lim_{r \to \infty} r \left(\frac{\partial u_{\text{sc}}}{\partial r} - iku_{\text{sc}} \right) = 0$$

where $r = \|\mathbf{r}\|$. Equation (7.7.4) can be converted to an equivalent integral equation by introducing the free space Green function. The free space Green function $g : \mathbb{R}^d \to \mathbb{R}^d$ satisfies

$$\nabla^2 g(\mathbf{r}) + k^2 \epsilon_b g(\mathbf{r}) = -\delta(\mathbf{r}), \qquad \forall\, \mathbf{r} \in \mathbb{R}^d$$

together with the Sommerfeld radiation conditions. Here, δ is a dirac-delta function. The explicit representation for the Green function reads

$$g(\mathbf{r}) \triangleq \begin{cases} -\frac{i}{2k} e^{-ikr} & d = 1 \\ -\frac{i}{4} H_0^{(2)}(kr) & d = 2 \\ \frac{1}{4\pi r} e^{-ikr} & d = 3 \end{cases},$$

where $r = \|\mathbf{r}\|$, and $H_0^{(2)}$ is the zero-order Hankel function of second kind. Hence, the integral representation for the input wavefield is

$$u_{\text{in}}(\mathbf{r}) = k^2 \int_{\mathbf{r}' \in \Gamma_t} g(\mathbf{r} - \mathbf{r}') q(\mathbf{r}') \, \mathrm{d}\mathbf{r}' \qquad \forall\, \mathbf{r} \in \mathbb{R}^d,$$

and similarly, for the scattered wavefield is

$$u_{\text{sc}}(\mathbf{r}) = k^2 \int_{\mathbf{r}' \in \Omega} g(\mathbf{r} - \mathbf{r}') u(\mathbf{r}') f(\mathbf{r}') \, \mathrm{d}\mathbf{r}' \qquad \forall \, \mathbf{r} \in \mathbb{R}^d.$$

Noting that the scattered wavefield is the difference of the total wavefield and the input wavefield (refer (7.7.1)) and restricting our observations to the object domain Ω, we arrive at the well-known Lippmann-Schwinger equation

$$u(\mathbf{r}) = u_{\text{in}}(\mathbf{r}) + k^2 \int_{\mathbf{r}' \in \Omega} g(\mathbf{r} - \mathbf{r}') u(\mathbf{r}') f(\mathbf{r}') \, \mathrm{d}\mathbf{r}' \qquad \forall \, \mathbf{r} \in \Omega$$

The above equation describes the relation between the total-wavefield and the contrast function inside the object domain Ω. The scattered wavefield is then measured in the receiver domain Γ_r resulting in the following data equation:

$$y(\mathbf{x}) = \int_{\Omega} g(\mathbf{x} - \mathbf{r}) f(\mathbf{r}) u(\mathbf{r}) \, \mathrm{d}\mathbf{r}, \qquad \forall \, \mathbf{x} \in \Gamma_r.$$

7.7.2 Gradient computation

In this section, we derive a gradient for an equality constrained cost function

$$\mathcal{F}(\mathbf{f}) = \Big\{ h(\mathbf{f}, \mathbf{u}) \quad \text{subject to} \quad \mathbf{k}(\mathbf{f}, \mathbf{u}) = \mathbf{0} \Big\} \tag{7.7.5}$$

where $h : \mathbb{R}^n \times \mathbb{C}^n \to \mathbb{R}$ is a real-valued function and $\mathbf{k} : \mathbb{R}^n \times \mathbb{C}^n \to \mathbb{R}^n$ is a set valued function. We assume that both the functions h and $\mathbf{k}$ are smooth and hence, differentiable. For the constrained problem (7.7.5), the Lagrangian reads

$$\mathcal{L}(\mathbf{f}, \mathbf{u}, \boldsymbol{\lambda}) = h(\mathbf{f}, \mathbf{u}) + \boldsymbol{\lambda}^H \mathbf{k}(\mathbf{f}, \mathbf{u}), \tag{7.7.6}$$

where $\boldsymbol{\lambda} \in \mathbb{C}^n$ is a Lagrange multiplier corresponding to the constraints, and $\mathbf{X}^H$ represents the conjugate transpose of the matrix $\mathbf{X}$ with complex entries. The stationary point of the Lagrangian $\mathcal{L}$, $(\mathbf{f}, \mathbf{u}^\star, \boldsymbol{\lambda}^\star)$, satisfies

$$\frac{\partial \mathcal{L}}{\partial \mathbf{u}} = \mathbf{0}, \qquad \frac{\partial \mathcal{L}}{\partial \boldsymbol{\lambda}} = \mathbf{0}.$$

The first condition gives rise to an adjoint equation

$$\frac{\partial h}{\partial \mathbf{u}}(\mathbf{f}, \mathbf{u}^\star) + \left(\frac{\partial \mathbf{k}}{\partial \mathbf{u}}(\mathbf{f}, \mathbf{u}^\star) \right)^H \boldsymbol{\lambda}^\star = 0, \tag{7.7.7}$$

while the second condition is the states equation

$$\mathbf{k}(\mathbf{f}, \mathbf{u}^\star) = \mathbf{0}. \tag{7.7.8}$$

The states equation generates a wavefield $\mathbf{u}^\star$ for a given parameter value $\mathbf{f}$. The adjoint equation calculates the Lagrange multiplier (also called adjoint wavefield) corresponding to wavefield $\mathbf{u}^\star$ for given $\mathbf{f}$. The gradient of $\mathcal{F}$ is now retrieved from the partial derivative of the Lagrangian with respect to $\mathbf{f}$,

$$\nabla \mathcal{F}(\mathbf{f}) = \frac{\partial \mathcal{L}}{\partial \mathbf{u}} = \frac{\partial h}{\partial \mathbf{f}}(\mathbf{f}, \mathbf{u}^\star) + \left(\frac{\partial \mathbf{k}}{\partial \mathbf{f}}(\mathbf{f}, \mathbf{u}^\star) \right)^H \boldsymbol{\lambda}^\star. \tag{7.7.9}$$

This method is popularly known as the *adjoint-state* method [145].

Inverse scattering Example

For an inverse scattering problem, h represents the misfit function between the simulated and the measured wavefields and $\mathbf{k} = \mathbf{0}$ is a Lippmann-Schwinger equation,

$$h(\mathbf{f},\mathbf{u}) \triangleq \tfrac{1}{2}\|\mathbf{y} - \mathbf{H}\,\mathbf{diag}(\mathbf{u})\mathbf{f}\|^2, \qquad \text{and} \qquad \mathbf{k}(\mathbf{f},\mathbf{u}) \triangleq (\mathbf{I} - \mathbf{G}\,\mathbf{diag}(\mathbf{f}))\,\mathbf{u} - \mathbf{v}.$$

At a given value of $\mathbf{f}$, the adjoint system for the Lippmann-Schwinger equation is

$$\left(\mathbf{I} - \mathbf{G}^H\,\mathbf{diag}(\mathbf{f})\right)\boldsymbol{\lambda}^\star = \mathbf{diag}(\mathbf{f})\mathbf{H}^H\left(\mathbf{y} - \mathbf{H}\,\mathbf{diag}(\mathbf{f})\mathbf{u}^\star\right). \tag{7.7.10}$$

Here, $\boldsymbol{\lambda}^\star$ is the adjoint wavefield and the $\mathbf{u}^\star$ is obtained satisfying the constraints at given value of $\mathbf{f}$:

$$(\mathbf{I} - \mathbf{G}\,\mathbf{diag}(\mathbf{f}))\,\mathbf{u}^\star = \mathbf{v}.$$

Once the forward wavefield $\mathbf{u}^\star$ and the adjoint wavefield $\boldsymbol{\lambda}^\star$ are computed, the gradient is

$$\nabla\mathcal{F}(\mathbf{f}) = \mathbf{diag}(\mathbf{u}^\star)^H\mathbf{H}^H\left(\mathbf{H}\,\mathbf{diag}(\mathbf{u}^\star)\mathbf{f} - \mathbf{y}\right) - \mathbf{diag}(\mathbf{u}^\star)^H\mathbf{G}^H\boldsymbol{\lambda}^\star. \tag{7.7.11}$$

It requires only one solve of each, the forward system ($\mathbf{k}(\mathbf{u},\mathbf{f}) = \mathbf{0}$), and the adjoint system (7.7.10), to compute the gradient.

7.7.3 Primal-Dual method

We consider a class of optimization problem

$$\min_{\mathbf{x}} \quad h(\mathbf{x}) + g(\mathbf{L}\mathbf{x}) + k(\mathbf{x}), \tag{7.7.12}$$

where $h : \mathbb{R}^n \to \mathbb{R}$ is a differentiable closed convex function. $g : \mathbb{R}^m \to \mathbb{R}$ and $k : \mathbb{R}^n \to \mathbb{R}$ are closed non-differentiable convex functions. We assume that the proximal operators for the functions h, g and k are inexpensive. $\mathbf{L} \in \mathbb{R}^{m\times n}$ denotes a structured matrix. For example, in TV regularization, $\mathbf{L}$ represents a discrete gradient operator. We assume that the matrix $\mathbf{L}$ may be potentially non-invertible, such as is the case in TV regularization. Here, we derive a primal-dual algorithm to find an optimal solution to problem (7.7.12).

Let us look at the optimality conditions for (7.7.12). It states that a zero-vector must be in the subdifferential of the cost function, *i.e.*,

$$\mathbf{0} \in \nabla h(\mathbf{x}) + \mathbf{L}^T\partial g(\mathbf{L}\mathbf{x}) + \partial k(\mathbf{x}), \tag{7.7.13}$$

where $\partial g : \mathbb{R}^m \to \mathbb{R}^m$ and $\partial k : \mathbb{R}^n \to \mathbb{R}^n$ are the respective subdifferentials of functions g and k. Let's consider variables $\mathbf{u} \in \mathbb{R}^m$ in the subdifferential of g and $\mathbf{v} \in \mathbb{R}^n$ in the subdifferential of k,

$$\mathbf{u} \in \partial g(\mathbf{L}\mathbf{x}), \qquad \mathbf{v} \in \partial k(\mathbf{x}). \tag{7.7.14}$$

The equations in (7.7.14) can be restated as follows.

$$\mathbf{0} \in \partial g^\star(\mathbf{u}) - \mathbf{L}\mathbf{x}, \qquad \mathbf{0} \in \partial k^\star(\mathbf{v}) - \mathbf{x} \tag{7.7.15}$$

where $g^\star$ and $k^\star$ are the convex conjugate of the functions g and k respectively. From equations (7.7.13) and (7.7.15), we can write the optimality conditions in the form of a following system

$$\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix} \in \underbrace{\begin{bmatrix}\nabla h & \mathbf{L}^T & \mathbf{I}_n\\ -\mathbf{L} & \partial g^\star & \mathbf{0}\\ -\mathbf{I}_n & \mathbf{0} & \partial k^\star\end{bmatrix}}_{\mathcal{A}} \underbrace{\begin{bmatrix}\mathbf{x}\\ \mathbf{u}\\ \mathbf{v}\end{bmatrix}}_{\mathbf{z}}, \tag{7.7.16}$$

where $\mathbf{I}_n \in \mathbb{R}^{n\times n}$ is an Identity matrix. Consider a preconditioner operator

$$\mathcal{P} = \begin{bmatrix}\frac{1}{\gamma}\mathcal{I} & -\mathbf{L}^T & -\mathbf{I}_n\\ -\mathbf{L} & \frac{1}{\gamma}\mathcal{I} & \mathbf{0}\\ -\mathbf{I}_n & \mathbf{0} & \frac{1}{\gamma}\mathcal{I}\end{bmatrix},$$

with $\gamma > 0$, the preconditioned fixed-point iteration scheme

$$\mathbf{z}^{(k+1)} = \left(\mathcal{I} + \mathcal{P}^{-1}\mathcal{A}\right)\mathbf{z}^{(k)}$$

results in the following primal-dual algorithm:

$$\begin{aligned}
\mathbf{x}^{(t+1)} &= (\mathcal{I} + \gamma\nabla h)^{-1}\left(\mathbf{x}^t - \gamma\mathbf{L}^T\mathbf{u}^{(t)} - \gamma\mathbf{v}^{(t)}\right)\\
\mathbf{u}^{(t+1)} &= (\mathcal{I} + \gamma\partial g^\star)^{-1}\left(\mathbf{u}^t - \gamma\mathbf{L}\left(\mathbf{x}^{(t)} - 2\mathbf{x}^{(t+1)}\right)\right)\\
\mathbf{v}^{(t+1)} &= (\mathcal{I} + \gamma\partial k^\star)^{-1}\left(\mathbf{v}^t - \gamma\left(\mathbf{x}^{(t)} - 2\mathbf{x}^{(t+1)}\right)\right)
\end{aligned} \tag{7.7.17}$$

7.7.4 Proximal quasi-Newton method

Here, we discuss the Quasi-Newton (QN) method and its proximal version (prox-QN). The QN method falls in the category of gradient-based methods. Let's assume the cost function f is twice differentiable. QN aims to solve the problem

$$\mathbf{x}^\star = \underset{\mathbf{x}}{\operatorname{argmin}} \quad f(\mathbf{x}) \tag{7.7.18}$$

by generating a sequence based on the quadratic approximation to the fuction f at every iterate of the sequence. The procedure is as follows:

$$\mathbf{s}^{(k)} = -\mathbf{H}_k^{-1}\nabla f\left(\mathbf{x}^{(k)}\right), \tag{7.7.19}$$

$$\alpha_k = \text{linesearch}\left(f(\mathbf{x}^k + \alpha\mathbf{s}^{(k)})\right), \tag{7.7.20}$$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k\mathbf{s}^{(k)}. \tag{7.7.21}$$

Here, $\mathbf{H}_k$ is a L-BFGS approximation of the Hessian of function f at $\mathbf{x}^{(k)}$. This method differs from the Newton method, as the former relies on an approximation, while the latter on the exact Hessian. If f is a convex function, the QN method converges to a global minimum. If f is non-convex, the QN can only guarantee the convergence to a local optimum.

We are interested in adapting the QN method for solving the problem of form

$$\mathbf{x}^\star = \underset{\mathbf{x}}{\operatorname{argmin}} \Big\{ f(\mathbf{x}) \mid g(\mathbf{L}\mathbf{x}) \leq \tau, \mathbf{x} \geq 0 \Big\}. \tag{7.7.22}$$

Here, $f : \mathbb{R}^n \to \mathbb{R}$ is a twice-differentiable function, and $g : \mathbb{R}^n \to \mathbb{R}$ is a convex but potentially non-differentiable function. For convenience, we rewrite the problem (7.7.22) as

$$\mathbf{x}^\star = \underset{\mathbf{x}}{\operatorname{argmin}} \Big\{ f(\mathbf{x}) + \delta_g(\mathbf{L}\mathbf{x}) + \delta_k(\mathbf{x}) \Big\}, \tag{7.7.23}$$

where, δ_g is an indicator to the set $\{\mathbf{x} : g(\mathbf{x}) \leq \tau\}$, and δ_k is an indicator to the set $\{\mathbf{x} : \mathbf{x} > \mathbf{0}\}$. We propose the following modification to the Quasi-Newton method, and name it as Proximal Quasi-Newton (Prox-QN) method:

$$\begin{aligned}
&(a) \quad \mathbf{s}^{(k)} = \underset{\mathbf{s}}{\operatorname{argmin}} \left\{ \mathbf{s}^T \nabla f\left(\mathbf{x}^{(k)}\right) + \frac{1}{2}\mathbf{s}^T \mathbf{H}_k \mathbf{s} + \delta_g\left(\mathbf{L}(\mathbf{x}^{(k)} + \mathbf{s})\right) + \delta_k\left(\mathbf{x}^{(k)} + \mathbf{s}\right) \right\} \\
&(b) \quad \text{define} \quad \hat{\mathbf{x}}(\alpha) = \mathbf{prox}_{\alpha\delta_g}\left(\mathbf{x}^k + \alpha \mathbf{s}^{(k)}\right) \\
&(c) \quad \alpha_k = \underset{\alpha}{\operatorname{argmin}} \left\{ f\left(\hat{\mathbf{x}}(\alpha)\right) \right\} \\
&(d) \quad \mathbf{x}^{(k+1)} = \hat{\mathbf{x}}\left(\alpha_k\right)
\end{aligned} \tag{7.7.24}$$

The steps in (7.7.24) can be summarized as follows: The step (a) finds a search direction $\mathbf{s}^k$. It involves the minimization of a quadratic approximation of f at $\mathbf{x}^{(k)}$, ensuring that it satisfies the constraints. In step (b), we define a function $\hat{\mathbf{x}} : \mathbb{R} \to \mathbb{R}$ which is a proximal of the iterate $\mathbf{x}^k + \alpha \mathbf{s}^{(k)}$ with respect to g. The function $\hat{\mathbf{x}}$ ensures that the step length, α, must satisfy the constraints. Step (c) does a linesearch with respect to the feasible α. Once we obtain the correct α, we update our variable of interest $\mathbf{x}$ in step (d).

The minimization problem in step (a) of (7.7.24), is a convex minimization problem. The cost function is the sum of three functions. The function

$$h(\mathbf{s}) = \mathbf{s}^T \nabla f\left(\mathbf{x}^{(k)}\right) + \frac{1}{2}\mathbf{s}^T \mathbf{H}_k \mathbf{s}$$

is a convex quadratic function, while the remaining two, δ_g and δ_k, are non-differentiable convex functions. To solve this minimization problem, we use the first-order primal-dual method described in Section 7.7.3. The iterates for $t = 0, \ldots, T$ are

$$\mathbf{s}^{(t+1)} = \mathbf{prox}_{\gamma h}\left(\mathbf{s}^{(t)} - \gamma \mathbf{L}^T \mathbf{u}^{(t)} - \gamma \mathbf{v}^{(t)}\right), \tag{7.7.25}$$

$$\mathbf{u}^{(t+1)} = \mathbf{prox}_{\gamma \delta_g^\star}\left(\mathbf{u}^{(t)} + \gamma \mathbf{L}\left(\mathbf{x}^{(k)} + 2\mathbf{s}^{(t+1)} - \mathbf{s}^{(t)}\right)\right), \tag{7.7.26}$$

$$\mathbf{v}^{(t+1)} = \mathbf{prox}_{\gamma \delta_k^\star}\left(\mathbf{v}^{(t)} + \gamma\left(\mathbf{x}^{(k)} + 2\mathbf{s}^{(t+1)} - \mathbf{s}^{(t)}\right)\right), \tag{7.7.27}$$

with $\gamma > 0$ controlling the speed of convergence. The proximal operations for h, δ_g and δ_k are expressed as follows:

$$\begin{aligned}
\mathbf{prox}_{\gamma h}(\mathbf{y}) &= (\mathbf{I} + \gamma \mathbf{H}_k)^{-1} \left(\mathbf{y} - \gamma \nabla f \left(\mathbf{x}^{(k)} \right) \right) \\
\mathbf{prox}_{\gamma \delta_g}(\mathbf{y}) &= \mathbf{proj}_{\|\cdot\|_1 \leq \tau}(\mathbf{y}) \\
\mathbf{prox}_{\gamma \delta_k}(y) &= \begin{cases} y & y > 0 \\ 0 & y \leq 0 \end{cases}
\end{aligned}$$

The proposed method (prox-QN) differs from [120] in two aspects: *(i)* The function g can be potentially be more than ℓ_1 type penalty. For example, we can work with total-variation-type regularization. *(ii)* The linesearch ensures that the chosen α is strictly feasible.

CHAPTER 8

Conclusions and outlook

" I was born not knowing and have had only a little time
to change that here and there. "
- Richard Feynman

In this thesis, we have developed efficient algorithms to retrieve discrete and partially discrete objects from their tomographic projections. Our key contributions are:

- **Convex formulation**
  We proposed a novel convex program to recover binary images from their linear tomographic projections. The convex program turns out to be an ℓ_1-regularized least-squares problem (known as LASSO). We conjecture that the convex program gives a solution if it is unique and an intersection of the solutions in the multiple solutions case (see chapter 3). Next, we consider ray tomography for partially discrete objects. We developed a bi-level approach by separating the homogeneous medium from the heterogeneous. In this bi-level approach, the inner-level (corresponding to the heterogeneous medium) is a convex problem, while the outer-level (corresponding to the homogeneous medium) remains non-convex (see chapter 4).
- **Parametric approach**
  We developed a level-set based approach to recover the shape of salt bodies in the earth from reflection measurements. We use the full-waveform inversion framework that models the interaction of the waves with the object in a nonlinear fashion. We represent the geometry of the salt using the zero level-set of the characteristic function composed of a few radial basis functions (RBF) spread over the domain. First, with the assumption of a known background medium, we show that salt geometry can be retrieved by obtaining the right combination of the RBF coefficients (see chapter 5). Later, we also parametrize the background medium with the RBFs and propose an alternating minimization strategy to retrieve both the background and the salt (see chapter 6).
- **Total-variation regularized multiscale approach**
  We extended the multiscale approach to image high-contrast objects from their scattered measurements. We mainly work in the reflection regime and use an integral representation for the scattered wavefield. We developed a combination of total-variation and non-negative constraint regularization to steer the image towards the correct solution appropriately. We propose an automated strategy to select the regularization parameter. We numerically show that the proposed algorithm correctly identifies high-contrast objects with a moderate amount of noise in the data (see chapter 7).

Based on the work in this thesis, we identify a couple of future directions. These directions can be classified into i) convex programs, and ii) parametric models. Convex programs deal with finding a suitable convex formulation of the non-convex optimization problems, while parametric models deal with dimensionality reduction of the shape reconstruction problem.

8.1 Convex programs

Convex programs are compelling for their ability to solve the problem using fast and accurate optimization techniques. The programs are reliable since they provide a global solution with minimal computational efforts. Since we have already developed a convex program for binary tomography, we see the following future opportunities that rely on the developed formulation.

8.1.1 Binary to discrete

A discrete image is an extension of a binary image to more than two grey levels. Presently, there are few heuristic techniques to obtain discrete images from their tomographic projections [19, 215]. It is possible to extend the convex program developed for the binary tomography to the discrete tomography case. A discrete image of k distinct grey levels can be thought of as a convex combination of k binary images. Hence, we can formulate the discrete tomography problem as a binary tomography problem with an added constraint.

8.1.2 Vector-valued greylevels

An attractive extension of binary images is an image composed of two heterogeneous regions. That is, an image with two function $u_0 : \Omega \to \mathbb{R}$ and $u_1 : \Omega \to \mathbb{R}$ is

$$x\left(\mathbf{r}\right) = \left(1 - a\left(\mathbf{r}\right)\right) u_0(\mathbf{r}) + a(\mathbf{r})\, u_1(\mathbf{r})\,,$$

where $a : \Omega \to \{0, 1\}$ is an indicator function. Since u_0 and u_1 are no more spatially invariant, we call this image a binary image with vector-valued grey levels. A partially discrete image is a particular case of this image where u_1 is constant. Although not trivial, a convex program can be formulated for linear tomographic imaging of these images.

These images often occur in the image segmentation literature, where one is interested in finding various regions in the image (depending on the structures). Methods like Mumford-Shah segmentation are popular (implicit) regularization techniques to obtain such images [135, 91].

8.1.3 Waveform inversion

Convex programs are easy to derive for linear operators due to the availability of explicit expressions for their Fenchel conjugates. This does not hold for nonlinear operators. Hence, to derive a convex program for waveform inversion, it is crucial to separate variables such that it becomes a linear operator for at least one variable. This can be achieved either by posing the waveform inversion problem in an ADMM framework or using the wavefield reconstruction inversion approach [201].

8.2 Parametric models

Parametric models are a useful tool in reducing the dimensionality of the problem. For example, in the parametric level-set approach, a level set function was

constructed from a set of radial basis functions (RBF). Hence, instead of finding a level-set function (in a discrete setting, its n point evaluations) that fits the shape, we need only $m \ll n$ coefficients of RBF for the parametric approach. Following are some of the directions in this area.

8.2.1 Representation

It is vital to find a relationship between the shape (or its contour) and the parametric model. That is, we need to answer the following questions: *i)* Can we represent a given shape with a zero level-set of the function constructed from the parametric model? *ii)* What are the conditions on the parametric model? *iii)* What is the resolution limit of the shape with a given parametric model? These kinds of question come under the umbrella of 'representation' which is important for understanding shape recovery with parametric models.

8.2.2 Computational efficiency

Our iterative algorithms rely on the computation of the gradient and the Hessian (or its approximation) at every iterate. With the parametric models, the size of the gradient and the Hessian decreases (for example, by representing the function evaluated at n discrete points with an m-dimensional parametric model ($m \ll n$), the gradient is an m-dimensional vector and the Hessian is a $m \times m$ matrix). Hence, parametric models reduce the memory of the algorithm, but do they also decrease the number of computations? Can we use randomization techniques to compute these small dimensional quantities quickly?

8.2.3 Vector-images

In wavefield imaging, the interaction of the object with waves depends on multiple characteristics of the object. For example, in geophysical application, the interaction is characterized by wave-speed, density, attenuation coefficients, Lamé parameters. In electromagnetics, this interaction is governed by the permittivity, permeability, conductivity, and the magnetic loss of the material. Hence, the imaging problem finds a map of these parameters, *i.e.*, a vector-valued image, from the scattered wavefield measurements. In our discrete tomography framework, the question arises if it is possible to represent vector-images with a single parametric model.

Bibliography

[1] Alireza Aghasi, Misha Kilmer, and Eric L Miller. Parametric level set methods for inverse problems. *SIAM Journal on Imaging Sciences*, 4(2):618–650, 2011.

[2] MS Albert, GD Cates, B Driehuys, W Happer, B Saam, CS Springer Jr, and A Wishnia. Biological magnetic resonance imaging using laser-polarized 129xe. *Nature*, 370(6486):199, 1994.

[3] Grégoire Allaire and Antoine Henrot. On some recent advances in shape optimization. *Comptes Rendus de l'Académie des Sciences-Series IIB-Mechanics*, 329(5):383–396, 2001.

[4] Grégoire Allaire, François Jouve, and Anca-Maria Toader. A level-set method for shape optimization. *Comptes Rendus Mathematique*, 334(12):1125–1130, 2002.

[5] A. Y. Anagaw and M. D. Sacchi. Edge-preserving seismic imaging using the total variation method. *Journal of Geophysics and Engineering*, 9(2):138, 2012.

[6] Aleksandr Y Aravkin and Tristan van Leeuwen. Estimating nuisance parameters in inverse problems. *Inverse Problems*, 28(11):115016, 2012.

[7] Simon R Arridge. Optical tomography in medical imaging. *Inverse problems*, 15(2):R41, 1999.

[8] Kenneth J Arrow, Leonid Hurwicz, and Hirofumi Uzawa. Studies in linear and non-linear programming. 1958.

[9] A. Asnaashari, R. Brossier, S. Garambois, F. Audebert, P. Thore, and J. Virieux. Regularized seismic full waveform inversion with prior model information. *Geophysics*, 78(2):R25–R36, 2013.

[10] Márton Balaskó, Attila Kuba, Antal Nagy, Zoltán Kiss, Lajos Rodek, and László Ruskó. Neutron-, gamma- and x-ray three-dimensional computed tomography at the budapest research reactor site. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 542(1-3):22–27, 2005.

[11] Gang Bao and Peijun Li. Inverse medium scattering problems for electromagnetic waves. *SIAM Journal on Applied Mathematics*, 65(6):2049–2066, 2005.

[12] Gang Bao, Peijun Li, Junshan Lin, and Faouzi Triki. Inverse scattering problems with multi-frequencies. *Inverse Problems*, 31(9):093001, 2015.

[13] Elena Barcucci, Sara Brunetti, Alberto Del Lungo, and Maurice Nivat. Reconstruction of lattice sets from their horizontal, vertical and diagonal X-rays. *Discrete Mathematics*, 241(1-3):65–78, 2001.

[14] Elena Barcucci, Alberto Del Lungo, Maurice Nivat, and Renzo Pinzani. Reconstructing convex polyominoes from horizontal and vertical projections. *Theoretical Computer Science*, 155(2):321–347, 1996.

[15] Stephen Barnett. *Matrices: methods and applications*. Clarendon Press, 1990.

[16] John J Barton. Photoelectron holography. *Physical review letters*, 61(12):1356, 1988.

[17] Kees Joost Batenburg, Sara Bals, J Sijbers, C Kübel, PA Midgley, JC Hernandez, U Kaiser, ER Encina, EA Coronado, and G Van Tendeloo. 3D imaging of nanomaterials by discrete tomography. *Ultramicroscopy*, 109(6):730–740, 2009.

[18] Kees Joost Batenburg and Jan Sijbers. DART: a practical reconstruction algorithm for discrete tomography. *IEEE Transactions on Image Processing*, 20(9):2542–2553, 2011.

[19] K.J. Batenburg and J. Sijbers. DART: A fast heuristic algebraic reconstruction algorithm for discrete tomography. In *2007 IEEE International Conference on Image Processing*, pages IV – 133–IV – 136. IEEE, 2007.

[20] Edip Baysal, Dan D Kosloff, and John WC Sherwood. Reverse time migration. *Geophysics*, 48(11):1514–1524, 1983.

[21] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[22] Martin P Bendsøe and Ole Sigmund. *Optimization of structural topology, shape, and material*, volume 414. Springer, 1995.

[23] P. Bharadwaj, W. A. Mulder, and G. G. Drijkoningen. Full waveform inversion with an auxiliary bump functional. *Geophysical Journal International*, 206(2):1076–1092, 2016.

[24] Pawan Bharadwaj, Wim Mulder, and Guy Drijkoningen. Full waveform inversion with an auxiliary bump functional. *Geophysical Journal International*, 206(2):1076–1092, 2016.

[25] Biondo Biondi and Ali Almomin. Tomographic full waveform inversion (TFWI) by combining full waveform inversion with wave-equation migration velocity analysis. In *SEG Technical Program Expanded Abstracts 2012*, pages 1–5. Society of Exploration Geophysicists, 2012.

[26] Folkert Bleichrodt, Frank Tabak, and Kees Joost Batenburg. SDART: An algorithm for discrete tomography from noisy projections. *Computer Vision and Image Understanding*, 129:63–74, 2014.

[27] Folkert Bleichrodt, Tristan van Leeuwen, Willem Jan Palenstijn, Wim van Aarle, Jan Sijbers, and K Joost Batenburg. Easy implementation of advanced tomography algorithms using the ASTRA toolbox with spot operators. *Numerical algorithms*, 71(3):673–697, 2016.

[28] Felix Bloch and Isidor Isaac Rabi. Atoms in variable magnetic fields. *Reviews of Modern Physics*, 17(2-3):237, 1945.

[29] Carlos Borges, Adrianna Gillman, and Leslie Greengard. High resolution

inverse scattering in two dimensions using recursive linearization. *SIAM Journal on Imaging Sciences*, 10(2):641–664, 2017.

[30] Max Born. Quantenmechanik der stoßvorgänge. *Zeitschrift für Physik*, 38(11-12):803–827, 1926.

[31] Max Born and Emil Wolf. *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Elsevier, 2013.

[32] A Bose, A Kadu, H Mansour, P Wang, P Boufounos, PV Orlik, and M Soltanalian. Thz multi-layer imaging via nonlinear inverse scattering. In *2019 44th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*, pages 1–2. IEEE, 2019.

[33] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine learning*, 3(1):1–122, 2011.

[34] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

[35] Haim Brezis. *Operateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, volume 5. Elsevier, 1973.

[36] Romain Brossier, Stéphane Operto, and Jean Virieux. Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion. *Geophysics*, 74(6):WCC105–WCC118, 2009.

[37] Tatiana A Bubba, Markus Juvonen, Jonatan Lehtonen, Maximilian März, Alexander Meaney, Zenith Purisha, and Samuli Siltanen. Tomographic x-ray data of carved cheese. *arXiv preprint arXiv:1705.05732*, 2017.

[38] M. D. Buhmann. Radial basis functions: theory and implementations. *Cambridge Monographs on Applied and Computational Mathematics*, 12:147–165, 2003.

[39] C. Bunks, F. M. Saleck, S. Zaleski, and G. Chavent. Multiscale seismic waveform inversion. *Geophysics*, 60(5):1457–1473, 1995.

[40] Martin Burger. A level set method for inverse problems. *Inverse problems*, 17(5):1327, 2001.

[41] Martin Burger and Stanley J Osher. A survey on level set methods for inverse problems and optimal design. *European Journal of Applied Mathematics*, 16(2):263–301, 2005.

[42] Hans Busch. Berechnung der Bahn von Kathodenstrahlen im axialsymmetrischen elektromagnetischen Felde. *Annalen der Physik*, 386(25):974–993, 1926.

[43] EJ Candes, J Romberg, and T Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.

[44] TD Capricelli and PL Combettes. A convex programming algorithm for noisy discrete tomography. In *Advances in Discrete Tomography and Its Applications*, pages 207–226. Springer, 2007.

[45] José-Maria Carazo, CO Sorzano, Eicke Rietzel, R Schröder, and Roberto Marabini. Discrete tomography in electron microscopy. In *Discrete Tomography*, pages 405–416. Springer, 1999.

[46] J. Carcione, G. C. Herman, and A. P. E. ten Kroode. Seismic modeling. *Geophysics*, 67(4):1304–1325, 2002.

[47] Bruno M Carvalho, Gabor T Herman, Samuel Matej, Claudia Salzberg, and Eilat Vardi. Binary tomography for triplane cardiography. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 29–41. Springer, 1999.

[48] Yair Censor and Samuel Matej. Binary steering of nonbinary iterative algorithms. In *Discrete tomography*, pages 285–296. Springer, 1999.

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

[50] Michael T Chan, Gabor T Herman, and Emanuel Levitan. Probabilistic modeling of discrete images. In *Discrete Tomography*, pages 213–235. Springer, 1999.

[51] Yu Chen. Inverse scattering via Heisenberg's uncertainty principle. *Inverse problems*, 13(2):253, 1997.

[52] W. C. Chew and Q. H. Liu. Perfectly matched layers for elastodynamics: a new absorbing boundary condition. *Journal of Computational Acoustics*, 4(4):341–359, 1996.

[53] Jon F Claerbout. *Imaging the earth's interior*, volume 1. Blackwell scientific publications Oxford, 1985.

[54] R. W. Clayton and B. Engquist. Absorbing boundary conditions for wave-equation migration. *Geophysics*, 45(5):895–904, 1980.

[55] David Colton and Rainer Kress. *Inverse acoustic and electromagnetic scattering theory*, volume 93. Springer Science & Business Media, 2012.

[56] Laurent Condat. Fast projection onto the simplex and the $\boldsymbol{l_1}$ ball. *Mathematical Programming*, 158(1-2):575–585, 2016.

[57] Taylor Dahlke, Biondo Biondi, and Robert Clapp. Domain decomposition of level set updates for salt segmentation. In *SEG Technical Program Expanded Abstracts 2015*, pages 1366–1371. Society of Exploration Geophysicists, 2015.

[58] Ingrid Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE transactions on information theory*, 36(5):961–1005, 1990.

[59] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 57(11):1413–1457, 2004.

[60] Stanley R Deans. *The Radon transform and some of its applications*. Courier Corporation, 2007.

[61] Alberto Del Lungo and Maurice Nivat. Reconstruction of connected sets from two projections. In *Discrete tomography*, pages 163–188. Springer, 1999.

[62] AJ Devaney. Inverse-scattering theory within the Rytov approximation. *Optics letters*, 6(8):374–376, 1981.

[63] Oliver Dorn and Dominique Lesselier. Level set methods for inverse scattering. *Inverse Problems*, 22(4):R67, 2006.

[64] Jacques Dubochet, Marc Adrian, Jiin-Ju Chang, Jean-Claude Homo, Jean Lepault, Alasdair W McDowall, and Patrick Schultz. Cryo-electron microscopy of vitrified specimens. *Quarterly reviews of biophysics*, 21(2):129–228, 1988.

[65] Thomas Lee Duvall, S D'silva, SM Jefferies, JW Harvey, and J Schou. Downflows under sunspots detected by helioseismic tomography. *Nature*, 379(6562):235, 1996.

[66] Y. A. Erlangga and F. J. Herrmann. An iterative multilevel method for computing wavefields in frequency-domain seismic inversion. In *SEG Technical Program Expanded Abstracts*. Society of Exploration Geophysicists, 2008.

[67] O. G. Ernst and M. J. Gander. Why it is difficult to solve Helmholtz problems with classical iterative methods. In *Numerical Analysis of Multiscale Problems*, pages 325–363. 2012.

[68] E. Esser, L. Guasch, T. van Leeuwen amd A. Y. Aravkin, and F. J. Herrmann. Automatic salt delineation — Wavefield Reconstruction Inversion with convex constraints. In *SEG Technical Program Expanded Abstracts 2015*, pages 1337–1343. Society of Exploration Geophysicists, 2015.

[69] Ernie Esser, Lluís Guasch, Felix J Herrmann, and Mike Warner. Constrained waveform inversion for automatic salt flooding. *The Leading Edge*, 35(3):235–239, 2016.

[70] Ernie Esser, Lluís Guasch, Tristan van Leeuwen, Aleksandr Y. Aravkin, and Felix J. Herrmann. Automatic salt delineation – wavefield reconstruction inversion with convex constraints. In *SEG Technical Program Expanded Abstracts 2015*, pages 1337–1343. Society of Exploration Geophysicists, 2015.

[71] Ernie Esser, Lluis Guasch, Tristan van Leeuwen, Aleksandr Y Aravkin, and Felix J Herrmann. Total variation regularization strategies in full-waveform inversion. *SIAM Journal on Imaging Sciences*, 11(1):376–406, 2018.

[72] Jeffrey A Fessler. Penalized weighted least-squares image reconstruction for positron emission tomography. *IEEE Transactions on Medical Imaging*, 13(2):290–300, 1994.

[73] Roger Fletcher. *Practical methods of optimization*. John Wiley & Sons, 2013.

[74] Thomas Frese, Charles A Bouman, and Ken Sauer. Multiscale bayesian methods for discrete tomography. In *Discrete Tomography*, pages 237–264. Springer, 1999.

[75] James G Fujimoto, Mark E Brezinski, Guillermo J Tearney, Stephen A Boppart, Brett Bouma, Michael R Hee, James F Southern, and Eric A Swanson. Optical biopsy and imaging using optical coherence tomography. *Nature Medicine*, 1(9):970–972, 1995.

[76] R J Gardner, P Gritzmann, and D Prangenberg. On the computational complexity of reconstructing lattice sets from their X-rays. *Discrete Mathematics*, 202(1–3):45–71, 1999.

[77] Odile Gauthier, Jean Virieux, and Albert Tarantola. Two-dimensional nonlinear inversion of seismic waveforms: Numerical results. *Geophysics*, 51(7):1387–1403, 1986.

[78] A. George and J. W.-H. Liu. *Computer Solution of Large Sparse Positive Definite*

Systems. Prentice-Hall, Englewood Cliffs, NJ, 1981.

[79] Tom Goldstein and Stanley Osher. The split Bregman method for L1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

[80] Richard Gordon, Robert Bender, and Gabor T Herman. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography. *Journal of theoretical Biology*, 29(3):471–481, 1970.

[81] Michael Grant, Stephen Boyd, and Yinyu Ye. Cvx: Matlab software for disciplined convex programming, 2008.

[82] Renliang Gu and Aleksandar Dogandžić. Projected Nesterov's proximal-gradient algorithm for sparse signal recovery. *IEEE Transactions on Signal Processing*, 65(13):3510–3525, 2017.

[83] Zhaohui Guo and Maarten V de Hoop. Shape optimization and level set method in full waveform inversion with 3D body reconstruction. In *SEG Technical Program Expanded Abstracts 2013*, pages 1079–1083. Society of Exploration Geophysicists, 2013.

[84] Eldad Haber, Uri M Ascher, and Doug Oldenburg. On optimization techniques for solving nonlinear inverse problems. *Inverse Problems*, 16(5):1263–1280, 2000.

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

[86] Lajos Hajdu and R Tijdeman. Algebraic aspects of discrete tomography. *Journal fur die Reine und Angewandte Mathematik*, 534:119–128, 2001.

[87] Jaroslav Haslinger and Raino AE Mäkinen. *Introduction to shape optimization: theory, approximation, and computation*. SIAM, 2003.

[88] Mark Haynes and Mahta Moghaddam. Large-domain, low-contrast acoustic inverse scattering for ultrasound breast imaging. *IEEE Transactions on Biomedical Engineering*, 57(11):2712–2722, 2010.

[89] Gabor T Herman and Attila Kuba. *Discrete Tomography: Foundations, Algorithms, and Applications*. Springer Science & Business Media, 1999.

[90] Gabor T Herman and Attila Kuba. *Advances in discrete tomography and its applications*. Springer Science & Business Media, 2008.

[91] Kilian Hohm, Martin Storath, and Andreas Weinmann. An algorithmic framework for mumford–shah regularization of inverse problems in imaging. *Inverse Problems*, 31(11):115011, 2015.

[92] Martin F Hohmann-Marriott, Alioscka A Sousa, Afrouz A Azari, Svetlana Glushakova, Guofeng Zhang, Joshua Zimmerberg, and Richard D Leapman. Nanoscale 3D cellular imaging by axial scanning transmission electron tomography. *Nature Methods*, 6(10):729, 2009.

[93] Godfrey N Hounsfield. Computerized transverse axial scanning (tomography): Part 1. description of system. *The British Journal of Radiology*, 46(552):1016–1022, 1973.

[94] JA Hudson and JR Heritage. The use of the Born approximation in seismic scattering problems. *Geophysical Journal International*, 66(1):221–240, 1981.

[95] MJ Humphry, B Kraus, AC Hurst, AM Maiden, and JM Rodenburg.

Ptychographic electron microscopy using high-angle dark-field scattering for sub-nanometre resolution imaging. *Nature Communications*, 3:730, 2012.

[96] B. Hustedt, S. Operto, and J. Virieux. Mixed-grid and staggered-grid finite-difference methods for frequency-domain acoustic wave modelling. *Geophysical Journal International*, 157(3):1269–1296, 2004.

[97] Churl-Hyun Jo, Changsoo Shin, and Jung Hee Suh. An optimal 9-point, finite-difference, frequency-space, 2-D scalar wave extrapolator. *Geophysics*, 61(2):529–537, 1996.

[98] A Kadu and R Kumar. Decentralized full-waveform inversion. In *80th EAGE Conference and Exhibition 2018*, 2018.

[99] A Kadu, R Kumar, and T van Leeuwen. Full-waveform inversion with mumford-shah regularization. In *SEG Technical Program Expanded Abstracts 2018*, pages 1258–1262. Society of Exploration Geophysicists, 2018.

[100] A Kadu, H Mansour, P T Boufounos, and D Liu. Reflection tomographic imaging of highly scattering objects using incremental frequency inversion. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7735–7739. IEEE, 2019.

[101] A Kadu and T van Leeuwen. A convex formulation for binary tomography. *IEEE Transactions on Computational Imaging*, pages 1–1, 2019.

[102] A Kadu, T van Leeuwen, and K J Batenburg. A parametric level-set method for partially discrete tomography. In *International Conference on Discrete Geometry for Computer Imagery*, pages 122–134. Springer, 2017.

[103] A Kadu, T van Leeuwen, and W. A. Mulder. A parametric level-set approach for seismic full-waveform inversion. In *SEG Technical Program Expanded Abstracts 2016*, pages 1146–1150. Society of Exploration Geophysicists, 2016.

[104] A Kadu, T van Leeuwen, and W A Mulder. Parametric level-set full-waveform inversion in the presence of salt bodies. In *SEG Technical Program Expanded Abstracts 2017*, pages 1518–1522. Society of Exploration Geophysicists, 2017.

[105] A Kadu, T van Leeuwen, and W A Mulder. Salt reconstruction in full-waveform inversion with a parametric level-set method. *IEEE Transactions on Computational Imaging*, 3(2):305–315, 2017.

[106] A Kadu, T van Leeuwen, and W A Mulder. Discrete seismic tomography. *SIAM Online News*, 2019.

[107] Jari Kaipio and Erkki Somersalo. Statistical inverse problems: discretization, model reduction and inverse crimes. *Journal of Computational and Applied Mathematics*, 198(2):493–504, 2007.

[108] Avinash C Kak and Malcolm Slaney. *Principles of computerized tomographic imaging*. SIAM, 2001.

[109] S Karczmarz. Angenaherte Auflosung von Systemen linearer Glei-Chungen. *Bull. Int. Acad. Pol. Sic. Let., Cl. Sci. Math. Nat.*, pages 355–357, 1937.

[110] Richard M Karp. Reducibility among combinatorial problems. In *Complexity of computer computations*, pages 85–103. Springer, 1972.

[111] Esther Klann, Ronny Ramlau, and Wolfgang Ring. A Mumford-Shah level-set approach for the inversion and segmentation of SPECT/CT data. *Inverse*

Probl. Imaging, 5(1):137–166, 2011.

[112] H. Knibbe, W. A. Mulder, C. W. Oosterlee, and C. Vuik. Closing the performance gap between an iterative frequency-domain solver and an explicit time-domain scheme for 3D migration on parallel architectures. *Geophysics*, 79(2):S47–S61, 2014.

[113] Max Knoll and Ernst Ruska. Das Elektronenmikroskop. *Zeitschrift für Physik*, 78(5-6):318–339, 1932.

[114] Jan Kuske, Paul Swoboda, and Stefania Petra. A novel convex relaxation for non-binary discrete tomography. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 235–246. Springer, 2017.

[115] B. N. Kuvshinov and W. A. Mulder. The exact solution of the time-harmonic wave equation for a linear velocity profile. *Geophysical Journal International*, 167(2):659–662, 2006.

[116] Marc Laruelle, Anissa Abi-Dargham, Christopher H Van Dyck, Roberto Gil, Cyril D D'Souza, Joseph Erdos, Elinore McCance, William Rosenblatt, Christine Fingado, Sami S Zoghbi, et al. Single photon emission computerized tomography imaging of amphetamine-induced dopamine release in drug-free schizophrenic subjects. *Proceedings of the National Academy of Sciences*, 93(17):9235–9240, 1996.

[117] S Laurens, JP Balayssac, J Rhazi, G Klysz, and G Arliguie. Non-destructive evaluation of concrete moisture by GPR: experimental study and direct modeling. *Materials and structures*, 38(9):827–832, 2005.

[118] PC LAUTERBUR. Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature*, 242(5394):190, 1973.

[119] Chau Le, Tyler Bruns, and Daniel Tortorelli. A gradient-based, parameter-free approach to shape optimization. *Computer Methods in Applied Mechanics and Engineering*, 200(9-12):985–996, 2011.

[120] Jason D Lee, Yuekai Sun, and Michael A Saunders. Proximal Newton-type methods for minimizing composite functions. *SIAM Journal on Optimization*, 24(3):1420–1443, 2014.

[121] Winston Lewis, Bill Starr, and Denes Vigh. A level set approach to salt geometry inversion in full-waveform inversion. In *SEG Technical Program Expanded Abstracts 2012*, pages 1–5. Society of Exploration Geophysicists, 2012.

[122] Y. Lin and L. Huang. Acoustic- and elastic-waveform inversion using a modified total-variation regularization scheme. *Geophysical Journal International*, 200(1):489–502, 2014.

[123] Dong C Liu and Jorge Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(1-3):503–528, 1989.

[124] Yanting Ma, Hassan Mansour, Dehong Liu, Petros T Boufounos, and Ulugbek S Kamilov. Accelerated image reconstruction for nonlinear diffractive imaging. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6473–6477. IEEE, 2018.

[125] Kurt J Marfurt. Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations. *Geophysics*, 49(5):533–549, 1984.

[126] Samuel Matej, Gabor T Herman, and Avi Vardi. Binary tomography on the hexagonal grid using Gibbs priors. *International Journal of Imaging Systems and Technology*, 9(2-3):126–131, 1998.

[127] Samuel Matej, Avi Vardi, Gabor T Herman, and Eilat Vardi. Binary tomography using Gibbs priors. In *Discrete Tomography*, pages 191–212. Springer, 1999.

[128] Russell Manning Mersereau and Alan V Oppenheim. Digital reconstruction of multidimensional signals from their projections. *Proceedings of the IEEE*, 62(10):1319–1338, 1974.

[129] Charles A Micchelli. Interpolation of scattered data: distance matrices and conditionally positive definite functions. In *Approximation theory and spline functions*, pages 143–145. Springer, 1984.

[130] Paul A Midgley and Rafal E Dunin-Borkowski. Electron tomography and holography in materials science. *Nature Materials*, 8(4):271, 2009.

[131] Dong-Joo Min, Changsoo Shin, R Gerhard Pratt, and Hai Soo Yoo. Weighted-averaging finite-element method for 2D elastic wave equations in the frequency domain. *Bulletin of the Seismological Society of America*, 93(2):904–921, 2003.

[132] W. A. Mulder, S. Osher, and J. A. Sethian. Computing interface motion in compressible gas dynamics. *Journal of Computational Physics*, 100(2):209–228, 1992.

[133] W. A. Mulder and R.-E. Plessix. Time-versus frequency-domain modelling of seismic wave propagation. In *64th EAGE Conference & Exhibition*, 2002.

[134] WA Mulder and R-E Plessix. Exploring some issues in acoustic full waveform inversion. *Geophysical Prospecting*, 56(6):827–841, 2008.

[135] David Mumford and Jayant Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989.

[136] Jorge Nocedal. Updating quasi-Newton matrices with limited storage. *Mathematics of Computation*, 35(151):773–782, 1980.

[137] Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

[138] Seiji Ogawa, Tso-Ming Lee, Alan R Kay, and David W Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *proceedings of the National Academy of Sciences*, 87(24):9868–9872, 1990.

[139] Stéphane Operto, Jean Virieux, Patrick Amestoy, Jean-Yves L'Excellent, Luc Giraud, and Hafedh Ben Hadj Ali. 3D finite-difference frequency-domain modeling of visco-acoustic wave propagation using a massively parallel direct solver: A feasibility study. *Geophysics*, 72(5):SM195–SM211, 2007.

[140] S. Osher and R. P. Fedkiw. Level set methods: an overview and some recent results. *Journal of Computational Physics*, 169(2):463–502, 2001.

[141] Stanley Osher and Ronald Fedkiw. *Level set methods and dynamic implicit surfaces*, volume 153. Springer Science & Business Media, 2006.

[142] Stanley Osher and James A Sethian. Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79(1):12–49, 1988.

[143] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

[144] Christopher C Paige and Michael A Saunders. Lsqr: An algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software (TOMS)*, 8(1):43–71, 1982.

[145] R.-E. Plessix. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167(2):495–503, 2006.

[146] R.-E. Plessix, G. Chavent, and Y.-H. De Roeck. Waveform inversion of reflection seismic data for kinematic parameters by local inversion. *SIAM Journal on Scientific Computing*, 20(3):1033–1052, 1999.

[147] M. J. D. Powell. Radial basis functions for multivariable interpolation: a review. In *Algorithms for approximation*, pages 143–167. Clarendon Press, 1987.

[148] R Gerhard Pratt. Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model. *Geophysics*, 64(3):888–901, 1999.

[149] R Gerhard Pratt, Changsoo Shin, and GJ Hick. Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion. *Geophysical Journal International*, 133(2):341–362, 1998.

[150] J Radon. On the determination of functions from their integrals along certain manifolds. *Ber. Verh, Sachs Akad Wiss.*, 69:262–277, 1917.

[151] C. D. Riyanti, Y. A. Erlangga, R.-E. Plessix, W. A. Mulder, C. Vuik, and C. Oosterlee. A new iterative solver for the time-harmonic wave equation. *Geophysics*, 71(5):E57–E63, 2006.

[152] RT Rockafellar. Convex analysis. 1970.

[153] Tom Roelandts, KJ Batenburg, E Biermans, C Kübel, S Bals, and J Sijbers. Accurate segmentation of dense nanoparticles by partially discrete electron tomography. *Ultramicroscopy*, 114:96–105, 2012.

[154] Justin Romberg. Imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):14–20, 2008.

[155] Wilhelm Conrad Röntgen. On a new kind of rays. *Science*, 3(59):227–231, 1896.

[156] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

[157] Ernest K Ryu and Stephen Boyd. Primer on monotone operator methods. *Appl. Comput. Math*, 15(1):3–43, 2016.

[158] Y. Saad. *Iterative Methods for Sparse Linear Systems*. Society for Industrial and Applied Mathematics, second edition, 2003.

[159] M Carmen San Martin, N Patrick J Stamford, Nada Dammerova, Nicholas E Dixon, and José M Carazo. A structural model for the Escherichia coli DnaB helicase based on electron microscopy data. *Journal of structural biology*, 114(3):167–176, 1995.

[160] F. Santosa and W. W. Symes. *An analysis of least-squares velocity inversion*. Geophysical monograph series. Society of exploration geophysicists, 1989.

[161] Fadil Santosa. A level-set approach for inverse problems involving obstacles. *ESAIM: Control, Optimisation and Calculus of Variations*, 1:17–33, 1996.

[162] Fadil Santosa and William W Symes. Linear inversion of band-limited reflection seismograms. *SIAM Journal on Scientific and Statistical Computing*, 7(4):1307–1330, 1986.

[163] Jorge LC Sanz, Eric B Hinkle, and Anil K Jain. Radon and projection transform-based computer vision: algorithms, a pipeline architecture, and industrial applications. 1988.

[164] Ken Sauer and Charles Bouman. A local update strategy for iterative reconstruction from projections. *IEEE Transactions on Signal Processing*, 41(2):534–548, 1993.

[165] William C Scarfe, Allan G Farman, Predag Sukovic, et al. Clinical applications of cone-beam computed tomography in dental practice. *Journal-Canadian Dental Association*, 72(1):75, 2006.

[166] M. W. Schmidt, E. Van Den Berg, M. P. Friedlander, and K. P. Murphy. Optimizing costly functions with simple constraints: A limited-memory projected quasi-Newton algorithm. In *AISTATS*, volume 5, page 2009, 2009.

[167] Mark Schmidt. minfunc: unconstrained differentiable multivariate optimization in matlab. 2012.

[168] P Schuetz, D Guggisberg, I Jerjen, MT Fröhlich-Wyder, J Hofmann, D Wechsler, A Flisch, W Bisig, U Sennhauser, and H-P Bachmann. Quantitative comparison of the eye formation in cheese using radiography and computed tomography data. *International Dairy Journal*, 31(2):150–155, 2013.

[169] Thomas Schüle, Christoph Schnörr, Stefan Weber, and Joachim Hornegger. Discrete tomography by convex–concave regularization and DC programming. *Discrete Applied Mathematics*, 151(1-3):229–243, 2005.

[170] Myriam Servieres, JP Guédon, and Nicolas Normand. A discrete tomography approach to PET reconstruction. In *Fully 3D Reconstruction In Radiology and Nuclear Medicine*, number 1, pages 1–4, 2003.

[171] Behzad Sharif and Behnam Sharif. Discrete tomography in discrete deconvolution: Deconvolution of binary images using Ryser's algorithm. *Electronic Notes in Discrete Mathematics*, 20:555–571, 2005.

[172] C. Shin and H. Sohn. A frequency-space 2-D scalar wave extrapolator using extended 25-point finite-difference operator. *Geophysics*, 63(1):289–296, 1998.

[173] Emil Y Sidky and Xiaochuan Pan. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in medicine and biology*, 53(17):4777, 2008.

[174] C. Da Silva and F. J. Herrmann. *A unified 2D/3D software environment for large-scale time-harmonic full-waveform inversion*, pages 1169–1173. 2016.

[175] L. Sirgue and R. G. Pratt. Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics*, 69(1):231–248, 2004.

[176] Laurent Sirgue, OI Barkved, J Dellinger, J Etgen, U Albertin, and JH Kommedal. Thematic set: Full waveform inversion: The next leap forward

in imaging at Valhall. *First Break*, 28(4):65–70, 2010.
[177] Morton Slater. Lagrange multipliers revisited. Technical report, Cowles Foundation for Research in Economics, Yale University, 1959.
[178] Jan Sokolowski and Antoni Zochowski. Topological derivative in shape optimization. *Encyclopedia of Optimization*, pages 3908–3918, 2009.
[179] Jan Sokolowski and Jean-Paul Zolésio. Introduction to shape optimization. In *Introduction to Shape Optimization*, pages 5–12. Springer, 1992.
[180] GH Spencer and MVRK Murty. General ray-tracing procedure. *JOSA*, 52(6):672–678, 1962.
[181] Jean-Luc Starck, Emmanuel J Candès, and David L Donoho. The curvelet transform for image denoising. *IEEE Transactions on Image Processing*, 11(6):670–684, 2002.
[182] I Štekl and R Gerhard Pratt. Accurate viscoelastic modeling by frequency-domain finite differences using rotated operators. *Geophysics*, 63(5):1779–1794, 1998.
[183] W. W. Symes. The seismic reflection inverse problem. *Inverse Problems*, 25(12):123008, 2009.
[184] W. W. Symes. Seismic inverse poblems: recent developments in theory and practice. In A.K. Louis, S. Arridge, and B. Rundell, editors, *Proceedings of the Inverse Problems from Theory to Applications Conference*, pages 2–6. IOP Publishing, 2014.
[185] W. W. Symes and J. J. Carazzone. Velocity inversion by differential semblance optimization. *Geophysics*, 56(5):654–663, 1991.
[186] William W Symes. Migration velocity analysis and waveform inversion. *Geophysical prospecting*, 56(6):765–790, 2008.
[187] Christopher KW Tam and Jay C Webb. Dispersion-relation-preserving finite difference schemes for computational acoustics. *Journal of computational physics*, 107(2):262–281, 1993.
[188] Albert Tarantola. Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8):1259–1266, 1984.
[189] Alan M Thompson, John C Brown, Jim W Kay, and D Michael Titterington. A study of methods of choosing the smoothing parameter in image restoration by regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4):326–339, 1991.
[190] Robert Tibshirani. Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.
[191] Andrei Nikolaevich Tikhonov. On the solution of ill-posed problems and the method of regularization. In *Doklady Akademii Nauk*, volume 151, pages 501–504. Russian Academy of Sciences, 1963.
[192] Ahmet Tuysuzoglu, Yuehaw Khoo, and W Clem Karl. Fast and robust discrete computational imaging. *Electronic Imaging*, 2017(17):49–54, 2017.
[193] Michael Unser, Julien Fageot, and John Paul Ward. Splines are universal solutions of linear inverse problems with generalized TV regularization. *SIAM Review*, 59(4):769–793, 2017.
[194] Wim van Aarle, Kees Joost Batenburg, Gert Van Gompel, Elke Van de

Casteele, and Jan Sijbers. Super-resolution for computed tomography based on discrete tomography. *IEEE Transactions on image processing*, 23(3):1181–1193, 2014.

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

[196] Sandra Van Aert, Kees J Batenburg, Marta D Rossell, Rolf Erni, and Gustaaf Van Tendeloo. Three-dimensional atomic imaging of crystalline nanoparticles. *Nature*, 470(7334):374, 2011.

[197] Ewout van den Berg and Michael P. Friedlander. Sparse optimization with least-squares constraints. *SIAM J. Optimization*, 21(4):1201–1229, 2011.

[198] Nico P van Dijk, Kurt Maute, Matthijs Langelaar, and Fred Van Keulen. Level-set methods for structural topology optimization: a review. *Structural and Multidisciplinary Optimization*, 48(3):437–472, 2013.

[199] T van Leeuwen, A Kadu, and W A Mulder. Geometric imaging for subsurface salt bodies. ERCIM, 2017.

[200] Tristan Van Leeuwen and Felix J Herrmann. Mitigating local minima in full-waveform inversion by expanding the search space. *Geophysical Journal International*, 195(1):661–667, 2013.

[201] Tristan van Leeuwen and Felix J Herrmann. A penalty method for PDE-constrained optimization in inverse problems. *Inverse Problems*, 32(1):015007, 2015.

[202] Yehuda Vardi and Cun-Hui Zhang. Reconstruction of binary images via the EM algorithm. In *Discrete Tomography*, pages 297–316. Springer, 1999.

[203] J. Virieux, V. Etienne, V. Cruz-Atienza, R. Brossier, E. Chaljub, O. Coutant, S. Garambois, D. Mercerat, V. Prieux, S. Operto, A. Ribodetti, and J. Tago. Modelling seismic wave propagation for geophysical imaging. In Masaki Kanao, editor, *Seismic Waves – Research and Analysis*. InTech, 2012.

[204] Jean Virieux and Stéphane Operto. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26, 2009.

[205] S. Wang, M. V. de Hoop, and J. Xia. On 3D modeling of seismic wave propagation via a structured parallel multifrontal direct Helmholtz solver. *Geophysical Prospecting*, 59(5):857–873, 2011.

[206] Michael Warner and Lluís Guasch. Adaptive waveform inversion: Theory. *Geophysics*, 81(6):R429–R445, 2016.

[207] H. Wendland. Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in Computational Mathematics*, 4(1):389–396, 1995.

[208] Alan J Witten, John E Molyneux, and Jonathan E Nyquist. Ground penetrating radar tomography: Algorithms and case studies. *IEEE Transactions on geoscience and remote sensing*, 32(2):461–467, 1994.

[209] Dean F Wong, Henry N Wagner, Larry E Tune, Robert F Dannals, Godfrey D Pearlson, Jonathan M Links, Carol A Tamminga, Emmanuel P Broussolle, Hayden T Ravert, Alan A Wilson, et al. Positron emission tomography reveals elevated D2 dopamine receptors in drug-naive schizophrenics. *Science*,

234(4783):1558–1563, 1986.

[210] Stephen Wright and Jorge Nocedal. Numerical optimization. *Springer Science*, 35:67–68, 1999.

[211] Andrew E Yagle. An algebraic solution for discrete tomography. In *Discrete Tomography*, pages 265–284. Springer, 1999.

[212] Zhen Yuan and Huabei Jiang. Three-dimensional finite-element-based photoacoustic tomography: Reconstruction algorithm and simulations. *Medical physics*, 34(2):538–546, 2007.

[213] Ahmed H Zewail. Four-dimensional electron microscopy. *Science*, 328(5975):187–193, 2010.

[214] X. Zhang, K. Z. Song, M. W. Lu, and X. Liu. Meshless methods based on collocation with radial basis functions. *Computational Mechanics*, 26(4):333–343, 2000.

[215] Xiaodong Zhuge, Willem Jan Palenstijn, and Kees Joost Batenburg. TVR-DART: a more robust algorithm for discrete tomography from limited projection data with automated gray value estimation. *IEEE Transactions on Image Processing*, 25(1):455–468, 2015.

Summary

Discrete Seismic Tomography

Tomographic imaging reveals the interior of an object. Similar to a camera, it images an object by illuminating it with electromagnetic or acoustic waves (or a beam of photons). Due to its non-invasive nature, it has found applications in various fields of sciences and engineering, including the medical sciences, material sciences, and geophysical exploration. Recent challenges include fast and accurate reconstructions of high contrasts in the object from a limited number of tomographic measurements. To tackle this, prior information about the object under inspection is beneficial. One particular example is a discrete prior where an object is made up of only a few homogeneous materials of known grey levels. The tomography that deals with this prior is known as discrete tomography. The discrete assumption holds approximately for many objects. In general, the low-contrast surrounding is not spatially invariant. This leads to a class of objects called partially discrete objects, where high-contrast materials lie in the non-homogeneous low-contrast background. This thesis develops algorithms based on the theory of convex optimization, regularization, and the level-set method to reconstruct discrete and partially discrete objects from limited measurements.

In X-ray tomography, the challenge to reconstruct an object from limited measurements stems from the high levels of radiation dose from X-rays. Although the tomographic problem is linear, the discrete prior makes it non-convex. Many heuristic algorithms exist to image discrete objects from a few of their ray projections. These algorithms often require manual tuning of parameters and may suffer in the noisy scenario. We develop a convex program to recover the binary objects (*i.e.*, objects composed of only two materials) that relies on the Lagrangian duality. The resulting problem is an ℓ_1-regularized least-squares problem (LASSO) that can be solved quickly. Based on small-scale experiments, we conjecture that if the binary tomography problem admits a unique solution, it can be recovered using the proposed formulation. In the case of multiple solutions, the convex program gives the intersection of the solutions. The proposed algorithm compares favorably to existing state-of-the-art algorithms like total variation and DART.

In wavefield imaging, the access to the measurements on one side of the object and the nonlinear interaction of waves with the object leads to challenges in the recovery of high contrast objects. The famous example of high contrast is the

salt-bodies in the earth, which are good indicators of hydrocarbon reservoirs. In the first instance, we assume that the grey level of the high contrast material is known and it is homogeneous. We develop a level-set approach that explicitly separates the high-contrast object from the low-contrast surrounding. We make use of a parametrization technique to represent the level-set function and make the problem lower-dimensional and well-behaved. On synthetic phantoms, we show that the high-contrast objects are recoverable under one-sided limited measurements. The parametric level-set approach is extended to partially discrete objects as well. In the second instance, we assume that the data consists of noise, and the noise level is known. We developed a total variation regularized multiscale framework that solves a series of least-squares problems. The regularization parameter is estimated automatically based on the noise level in the data. The proposed algorithm outperforms the classical methodologies on the numerical phantoms. Since the algorithm does not account for discreteness of the object, it is also applicable for the imaging of partially discrete objects.

Samenvatting

Discrete seismische tomografie

Met tomografische beeldvorming wordt een beeld gemaakt van de binnenkant van een object via non-destructieve metingen. Op soortgelijke wijze als bij een camera wordt een beeld gevormd door het object te belichten met elektromagnetische of akoestische golven (of een straal van fotonen). Door de non-invasieve aard heeft het veel toepassingen in de wetenschap, waaronder de medische wetenschap, materiaalwetenschap en geofysische exploratie. Hedendaagse uitdagingen in het onderzoeksgebied zijn het ontwikkelen van snelle en accurate reconstructies uit een beperkt aantal tomografische metingen. Dit wordt veelal gedaan door voorkennis over het te meten object mee te nemen in de reconstructie. Een specifiek voorbeeld is het gebruik van zogeheten discrete voorkennis waarbij het te meten object bestaat uit slechts een aantal bekende grijstinten. De tomografie die zich bezighoudt met dit type voorkennis heet discrete tomografie. De aanname dat het object bestaat uit discrete grijstinten is grotendeels geldig voor een groot aantal objecten. Over het algemeen heeft de omgeving een laag contrast en is het niet invariant in de ruimte. Dit geeft aanleiding tot zogeheten partieel discrete objecten, waar een object met hoog contrast omgeven wordt door een inhomogene achtergrond met laag contrast. In deze scriptie worden algoritmen ontwikkeld gebaseerd op theorie uit convexe optimalisatie, regularizatie en de niveau-lijn methode om de discrete en partieel discrete objecten te reconstrueren uit beperkte metingen.

In röntgentomografie kan men maar beperkt metingen doen door de destructieve aard van de röntgenstralen. Hierdoor is er doorgaans te weinig informatie om het object te reconstrueren zonder extra voorkennis. Het tomografische probleem is linear, maar de voorkennis dat het object discreet is maakt het probleem non-convex. Er bestaan veel heuristieke algoritmen die een beeld vormen van discrete objecten uit een beperkt aantal projecties van de stralen. Deze algoritmen hebben vaak parameters die handmatig afgesteld moeten worden en zijn gevoelig voor ruis in de data. Wij ontwikkelen een convex programma dat de discrete objecten reconstrueert gebaseerd op Lagrange dualiteit. Het daaruitvolgende probleem is een ℓ_1-geregulariseerd kleinste kwadraten probleem (LASSO) dat snel opgelost kan worden. Gebaseerd op kleinschalige experimenten, vermoeden we dat het algorithme het object reconstrueert als het binaire tomografische probleem een

unieke oplossing heeft. Als er meerdere oplossingen zijn geeft het algoritme de doorsnede van deze oplossingen. Het voorgestelde algoritme vergelijkt gunstig ten opzichte van bestaande moderne algoritmen zoals totale variatie en DART.

In beeldvorming met golfvelden wordt het object maar van een kant gemeten en is de interactie van de golfvelden met het object niet-lineair. Dit leidt tot problemen in het reconstrueren van objecten met een hoog contrast. Een bekend voorbeeld van de reconstructie van een object met hoog contrast is het reconstrueren van een zoutlichaam in de ondergrond. De aanwezigheid van een zoutlichaam geeft een goede indicatie geven van de aanwezigheid van koolwaterstoffen. In eerste instantie nemen we aan dat de grijstinten van het materiaal bekend zijn en dat het object homogeen is. We ontwikkelen een niveau-lijn methode die expliciet het object met hoog contrast onderscheidt van de achtergrond met laag contrast. We gebruiken een parametrisatie techniek om de niveau-lijn te representeren, wat het probleem laag-dimensionaal en goedgesteld maakt. Op een synthetisch voorbeeld laten we zien dat het object gereconstrueerd kan worden met beperkte metingen van één kant. De parametrische niveau-lijn methode wordt uitgebreid naar een gedeeltelijk discreet object. In tweede instantie nemen we aan dat er ruis op de data zit en dat het ruisgehalte bekend is. We ontwikkelen een door totale variatie geregulariseerd multischaal raamwerk dat een reeks kleinste kwadraten problemen oplost. De regularizatieparameter wordt automatisch geschat op basis van het ruisgehalte. Het voorgestelde algoritme werkt beter dan de klassieke methoden op onze voorbeelden. Aangezien het algoritme niet expliciet de discrete aard van het object in beschouwing neemt is de methode ook toepasbaar voor het reconstrueren van gedeeltelijk discrete objecten.

Acknowledgments

It is my pleasure to express gratitude to the people who have been an essential part of my life for the last four years.

First of all, I'm greatly indebted to my Ph.D. supervisor, **Tristan**. To be honest, Ph.D. was never on my career plan a few years back. I was a bit reluctant at the start of my Ph.D., but Tristan made the best efforts to accommodate me in the Dutch research community. In the initial period, he offered me an ample amount of time to discuss random ideas. He eventually taught me to structure them and how to break them down into small pieces. His *mantra* of 'start with the simplest possible example' has become an essential part of my research. I'm delighted that he provided me numerous opportunities during my Ph.D. This includes the formation of a student-chapter, various scientific conferences, an internship in the USA, long-term visits to different research groups in the world. Thank you very much, Tristan!

My Ph.D. promoters, **Wim**, and **Joost**, have played a significant role in my research life. Both were superbly friendly and had a vibrant research environment. Wim has been very critical and helped me to notice even the smallest mistake. I enjoyed the discussions with him regarding Shell and the ups-and-downs of the Energy industry. I did not get much chance to work with Joost, but I look forward to working with him for the next few months on the 'Lego Problem.' I would also like to thank **Jason** for being a promoter from Utrecht. Without much involvement, he provided an external view of the project. I would also like to thank my Masters' supervisors, **R. P. Shimpi** and **N. Hemachandra**, who motivated me to pursue Ph.D.

I express my sincere thanks and indebtedness to **Felix Herrmann** for hosting me numerous times in his group. It is because of the SLIM group that I got a head start in my Ph.D. I would also like to thank **Rajiv**, **Bas**, **Phillipe**, **Shashin**, and **Gabrio** for friendly discussions and exciting collaborations. In particular, Rajiv was my academic mentor throughout my Ph.D. Thank you, Rajiv, for helping me out with geophysics literature. I would like to thank **Hassan**, **Petros**, **Dehong**, and **Perry** from MERL. Hassan has provided me an excellent opportunity to work on a cutting-edge problem and a chance to live in the most beautiful city in America. Thanks, **Oscar**, **Arindam**, and **Suhas**, for an exhilarating summer in Boston. I loved hanging out with you and especially watching the skyline from Arindam's rooftop.

Thanks, **Sasha** and **Peng**, for making my visit to Seattle worthwhile. I have a lot to learn from Sasha, and I look forward to working with him in the future.

My teammates, **Nick** and **Sarah**, have been of invaluable technical help and motivation to me. The coffee breaks, lunches, and small seminar meetings made me feel at ease with the small but an excellent team that we made together. Nick, in particular, became a very caring friend and an invaluable critique of my work. Thanks, **Andreas** and **Izzat**, for helpful discussions. I would like to thank the CCSS team for providing office space during last year of my Ph.D. **Pauline**, many thanks for those stimulating conversations during long coffee-breaks.

I am thankful to all the current and previous members of the Mathematical Institute for providing me all the comfort and stimulus environment to carry out my research. In particular, I would like to thank **Felix**, **Ties**, **Bart**, **Sehar**, **Jan Willem**, **Harry**, **Davide**, and **Shan**. I enjoyed all the discussions over lunch and nice coffee breaks we have had. I would also like to thank **Rob**, **Paul**, **Wilberd** and **Svetlana**, whom I occasionally met during coffee breaks.

Kudos to **Sakshi** and **Aridaman** for creating a mini-India environment in Utrecht. Your initiatives of gathering people at festivities made me not missthe home at all. Thank you, Sakshi, for providing moral support during low moments in life, and Aridaman, for motivating me to pursue an academic career. **Siddharth**, what can I say, I could not have asked for a better flatmate than you. You were like a big brother to me. Thanks, **Divyae** (aka Keto), for joining our house. Our conversations were always turned into heated debates due to our differences, but nonetheless, I enjoyed those moments. Thanks, **Abhinandan**, for being a gym-buddy and a very caring friend. Sid, Keto, you and I make a great team. I enjoyed all the activities with you guys and look forward to more in my remaining time in the Netherlands. Thanks, **Priyanka** and **Shamika**, for short trips and evening hangouts. Thanks, **Akshay**, for providing me an opportunity to work in your startup. I have always admired your passion and would love to learn entrepreneurship skills from you. Thanks, **Tanmay**, **Aditya**, **Sandeep**, **Ayush**, **Deepak**, **Amol**, **Rama**, **Somil**, and **Harini**, for nice hangouts and conversations. I would also like to thank all my IIT-B friends (including Saket, Pawan, Uttam, Manas, Akash, Majnu, Lohit, Arjun, Vikrant, Ketan, Chaitanya, Sayantam) for making an undergrad life exciting and supporting my Ph.D. decision. I would like to thank my school friends **Sagar**, **Taresh**, **Hitesh** and **Ajinkya Pachagade** for maintaining the friendship and the our yearly trips to forests.

This thesis would not have been possible without constant support from my parents. They have always encouraged me to pursue my ambitions and provided financial assistance when needed. Thank you, **Aai** and **Baba**, for your endless love and showing confidence in my decisions. **Anand**, you are a great brother. Your hard work has taught me many lessons. Thanks, **Ananta**, **Nikita**, **Isha**, **Sai**, **Ishaan**, and all my uncles and aunts (including Sandeep kaka, Vaishali kaku, Mamaji, Anju and Manju Aatya, Jaya kaku and Nitin kaka) for providing emotional support to me.

Publications & Dissemination

Journal/Conference Articles:

- A Bose, **A Kadu**, H Mansour, P Wang, P Boufounos, PV Orlik, and M Soltanalian. Thz multi-layer imaging via nonlinear inverse scattering. In *2019 44th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*, pages 1–2. IEEE, 2019.

- **A Kadu**, H Mansour, P T Boufounos, and D Liu. Reflection tomographic imaging of highly scattering objects using incremental frequency inversion. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7735–7739. IEEE, 2019.

- **A Kadu**, and T van Leeuwen. A convex formulation for binary tomography. *IEEE Transactions on Computational Imaging*, pages 1–1, 2019.

- **A Kadu** and R Kumar. Decentralized full-waveform inversion. In *80th EAGE Conference and Exhibition 2018*, 2018.

- **A Kadu**, R Kumar, and T van Leeuwen. Full-waveform inversion with mumford-shah regularization. In *SEG Technical Program Expanded Abstracts 2018*, pages 1258–1262. Society of Exploration Geophysicists, 2018.

- **A Kadu**, T van Leeuwen, and W A Mulder. Salt reconstruction in full-waveform inversion with a parametric level-set method. *IEEE Transactions on Computational Imaging*, 3(2):305–315, 2017.

- **A Kadu**, T van Leeuwen, and K J Batenburg. A parametric level-set method for partially discrete tomography. In *International Conference on Discrete Geometry for Computer Imagery*, pages 122–134. Springer, 2017.

- **A Kadu**, T van Leeuwen, and W A Mulder. Parametric level-set full-waveform inversion in the presence of salt bodies. In *SEG Technical Program Expanded Abstracts 2017*, pages 1518–1522. Society of Exploration Geophysicists, 2017.

- **A Kadu**, T van Leeuwen, and W. A. Mulder. A parametric level-set approach for seismic full-waveform inversion. In *SEG Technical Program Expanded Abstracts 2016*, pages 1146–1150. Society of Exploration Geophysicists, 2016.

Oral Presentations:

- **Oberwolfach** Meeting on Tomographic Inverse Problems, Germany Jan 2019
- CASA seminar, **TU Eindhoven**, The Netherlands Nov 2018
- Imaging and Computing group, **MIT**, Boston, USA Oct 2018
- Center for Signal and Information Processing Seminar, **GaTech**, USA Oct 2018
- SEG Annual Meeting, **Anaheim**, California, USA Oct 2018
- SIAM Conference on Imaging Science, **Bologna**, Italy June 2018
- SEG Annual Meeting, **Houston**, TX, USA Sep 2017
- Intl Conf on Discrete Geometry for Computer Imagery, **Vienna**, Austria Sep 2017
- SIAM Conf on Math and Comp Issues in Geosciences, **Erlangen**, Germany Sep 2017
- SIAM Annual Meeting, **Pittsburgh**, PA, USA July 2017
- NDNS+ PhD Days, **Lunteren**, Netherlands May 2017
- Utrecht SIAM Student Chapter talk, Utrecht, Netherlands April 2017
- Computational Imaging Group, **CWI Amsterdam**, Netherlands Jan 2017
- SEG Annual Meeting, **Dallas**, Texas Oct 2016
- **CSER** Future Energy Conference, Utrecht, Netherlands Oct 2016
- Conference on Multiscale Inverse Problem, **Loka Brunn**, Sweden Aug 2016
- Mini-Symposium on Seismic Imaging, **Delft**, Netherlands May 2016
- SLIM Seismic Seminar, **Vancouver**, Canada April 2016

Posters:

- **IMA** workshop on Computational Imaging, Minneapolis, USA October 2019
- Woudschoten Conference, **Zeist**, Netherlands Oct 2019
- **ICERM** Computational Imaging, Providence, USA March 2019
- EAGE Annual Meeting, **Copenhagen**, Denmark June 2018
- Houston Imaging Sciences Symposium, **Houston**, TX, USA Oct 2017
- INdAM MATHTECH Workshop on Biomedical Imaging, **Rome**, Italy Feb 2017
- SIGMA Conference, CIRM **Marseille**, France Nov 2016
- Woudschoten Conference, **Zeist**, Netherlands Oct 2016
- NDNS+ Workshop, **Enschede**, Netherlands June 2016
- Woudschoten Conference, **Zeist**, Netherlands Oct 2015

News Articles:

- **A Kadu**, T van Leeuwen, and W A Mulder. Discrete seismic tomography. *SIAM Online News*, 2019.
- T van Leeuwen, **A Kadu**, and W A Mulder. Geometric imaging for subsurface salt bodies. ERCIM, 2017.

Curriculum Vitae

Ajinkya Kadu was born on April 20, 1993 in Bhilai, India. After finishing the high-school in Hyderabad, he joined the **Indian Institute of Technology, Bombay** (IIT-B) in Mumbai, India in the Aerospace Engineering Department. During his studies at IIT-B, he interned at **Honeywell Technology Labs** in Bengaluru (Summer 2013) and **Rediff Labs** in Mumbai (Summer 2014). He graduated with a Bachelor and Master of Technology (2015) in Aerospace Engineering. He performed his thesis on 'Risk aversion in supply chains related to the aviation market' under the supervision of Prof. N Hemachandra and Prof. R.P. Shimpi. At IIT-B, he was involved in organizing events of the **Aerospace Engineering Association** and part of the controls team of **Pratham** - a student satellite project. He worked as an academic mentor for two years and headed the **Department Mentorship program** in the year 2014-15.

He started a Ph.D. project in July 2015 at **Utrecht University** in the Netherlands under the supervision of Dr. Tristan van Leeuwen, Prof. Wim A. Mulder, and Prof. K. Joost Batenburg. His research was part of the project 'Computational Sciences for Energy Research,' a public-private research program of the Netherlands Organisation for Scientific Research with Shell Global Solutions. During his Ph.D., he spent Winter 2016 at the **University of British Columbia** in Vancouver, Canada, and Summer 2018 at **Mitsubishi Electric Research Labs** in Cambridge, MA, USA. He also visited the Seismic Imaging group at **Georgia Institute of Technology** in Atlanta (October 2018 and October 2019) and Applied Modeling and Optimization group at the **University of Washington** in Seattle (March 2019). He also co-founded the SIAM student chapter to foster interactions among fellow applied mathematicians.

After his Ph.D., he will be a research associate at Utrecht University for eight months working on the project '*Motifs in three-dimensional tomographic imaging*,' funded through the NWO-XS grant.

Please visit `https://ajinkyakadu125.github.io/` for more info.