

# High-Contrast Reflection Tomography with Total-Variation Constraints

Ajinkya Kadu, Hassan Mansour, Petros T. Boufounos

**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. In this paper, we consider 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. We propose a proximal Quasi-Newton method to solve the resulting subproblem and devise an automatic parameter selection routine to determine the constraint of each subproblem. We validate the performance of our approach on synthetic low-resolution phantoms and with a mismatched forward model test on a high-resolution phantom.

**Index Terms**—Computational imaging, inverse scattering, total variation regularization, reflection tomography, limited data

## I. 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 [2], optical tomography [3], geophysical imaging [4], [5], ground-penetrating radar [6], medical imaging [7], [8], and electron microscopy [9–11].

A scattering experiment consists of a transmission domain  $\Gamma_t \subset \mathbb{R}^d$ , an object domain  $\Omega \subset \mathbb{R}^d$  and a receiver domain  $\Gamma_r \subset \mathbb{R}^d$ , where  $d(= 2, 3)$  is the dimension of the scene. A set of transmitters located in  $\Gamma_t$  sends incident waves into the scene that interact with an object in  $\Omega$ . This interaction leads to scattering of the incident waves. The scattered waves are then measured at the set of receivers located in  $\Gamma_r$ . Based on the location of transmission and receiver domain with respect to the object, we can classify the acquisition scheme into three

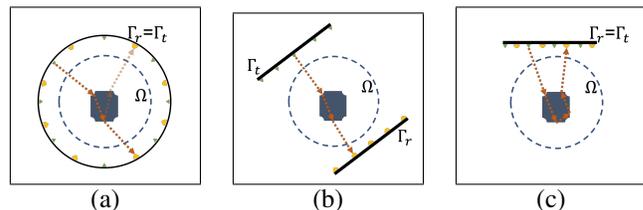


Fig. 1: Three acquisition scenarios in inverse scattering, (a) full-view, (b) Transmission, and (c) Reflection.  $\Omega$  is the domain of interest,  $\Gamma_t$  is a transmission domain, and  $\Gamma_r$  is a receiver domain. A single experiment consists of - a set of transmitters in  $\Gamma_t$  sending a wavefield into  $\Omega$ , and scattered wavefield being measured by set of receivers in  $\Gamma_r$ .

different types: (i) *full-view*, where  $\Gamma_t$  and  $\Gamma_r$  surround the domain  $\Omega$ ; (ii) *transmission* mode, where  $\Gamma_t$  and  $\Gamma_r$  are located on opposite sides of the object; and (iii) *reflection* mode, where  $\Gamma_r$  and  $\Gamma_t$  are co-located. Figure 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 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. 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. We discuss this further in Section II-C. 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 an 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 paper, 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. We also restrict the

A preliminary version of this work appeared in [1].

A. Kadu is with the Mathematical Institute of Utrecht University in The Netherlands. The work was partially completed while A. Kadu was with MERL.

H. Mansour and P. T. Boufounos are with Mitsubishi Electric Research Laboratories (MERL), 201 Broadway, Cambridge, MA, 02139.

study to lossless objects. Extension of the analysis to lossy objects would require that the scattering potential be complex-valued with a non-negative real component. However, such an extension is out of the scope of this paper.

### A. Related Work

Numerous techniques have been proposed for solving the inverse scattering problem in the reflection regime. Earlier approaches linearize the scattering model iteratively using straight-ray theory, the Born approximation, and the Rytov approximation [12–15]. 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 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, [16–20]). The inverse problem that deals with the wave-equation based scattering model is known as full-waveform inversion (FWI) [5], [21], [22]. 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 are mainly interested in reflection tomography, we do not address the literature for other modes of tomography and nonlinear inverse problems, see for example [23], [24]. We note, however, that contrary to other modes of tomography, measurements in reflection tomography are dominated by the high spatial frequencies of the contrast map. Moreover, contrary to other nonlinear inverse problems, the reflection tomography problem suffers from the fact that for every new frequency included in the measurements, a significantly larger number of unknowns are added to the estimation problem.

Reflection tomography with FWI has been heavily investigated in the geophysical community. Since the problem is nonlinear and nonconvex, the convergence of the inversion depends heavily on the initial model [25], [26]. Various approaches have been proposed to mitigate the effect of an initial model on the reconstructed solution [27–31]. While these methods work well in the low contrast regime, they require regularization and additional constraints in the high contrast regime to deliver good reconstruction [32–35]. Our work is also inspired by the TV-regularization strategies proposed in [33]. However, since the total variation parameter may be unknown [33], we develop a framework to estimate this parameter from the noise level in the data.

The sequential workflow has a long history in the geophysical imaging literature. It was introduced in [36] under the name 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 [36], our approach is more robust against local minima (more discussions in Section III-A).

### B. 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 incremental frequency inclusion. 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 the regularizers are non-differentiable. To solve the regularized nonlinear least-squares problem, we propose a proximal Quasi-Newton (prox-QN) method that is computed using a primal-dual method.
- *Parameter Estimation:* We develop a strategy for estimating the total-variation constraint parameter from the noise-level in the data.

We introduce the forward and the inverse scattering problem in Section II. Here, we also discuss the challenges of reflection tomography. In Section III, we present the details of our sequential approach and describe the optimization framework as well as the regularization strategies. We validate the proposed method on numerical phantoms and compare it with other methods in Section V, and conclude the paper in Section VI.

## II. 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 measured scattered wavefields. Finally, we discuss some challenges in estimating the contrast of an object in the reflection regime.

### A. Forward problem

The *forward* scattering problem constructs a mapping from a contrast function (determined by materials in the object) to the scattered waves measured at receivers. A wave-equation governs this mapping in the frequency-domain. For simplicity, we restrict our discussion to scalar waves, but the map can naturally be constructed for the other types of scattering problems with some modifications (see, for example, [13]).

Consider the setup shown in Figure 1 where an object is 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  $\epsilon_b$ . 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.*,  $f(\mathbf{r}) = \epsilon(\mathbf{r}) - \epsilon_b$ . In this section, we consider the single frequency setting and drop the frequency index from the equations. The variables are expressed as scalar functions of the position  $\mathbf{r}$ . We will later use vector notation to represent the variables over the entire domain  $\Omega$ .

We illuminate the target object using the waves generated from a source function  $q : \Gamma_t \mapsto \mathbb{C}$ . Subsequently, the scattered wavefield is measured inside the receiver domain  $\Gamma_r$ . The total

wavefield  $u : \Omega \mapsto \mathbb{C}$  in the object domain  $\Omega$  is related to the contrast function  $f$  by the *Lippmann-Schwinger* integral equation

$$u(\mathbf{r}) = u_{\text{in}}(\mathbf{r}) + k^2 \int_{\mathbf{r}' \in \Omega} g(\mathbf{r} - \mathbf{r}') u(\mathbf{r}') f(\mathbf{r}') d\mathbf{r}' \quad \forall \mathbf{r} \in \Omega, \quad (1)$$

where  $g : \mathbb{C}^d \mapsto \mathbb{C}$  is the Green function,  $u_{\text{in}} : \mathbb{C}^d \mapsto \mathbb{C}$  is an input wavefield,  $k = 2\pi\omega/c$  represents the wavenumber in vacuum,  $\omega$  is the frequency and  $c$  denotes the speed of light in vacuum. We assume that  $f$  is real, or in other words, the object is lossless. The input wavefield in (1) depends on the source function  $q$  as

$$u_{\text{in}}(\mathbf{r}) = k^2 \int_{\mathbf{r}' \in \Gamma_t} g(\mathbf{r} - \mathbf{r}') q(\mathbf{r}') d\mathbf{r}' \quad \forall \mathbf{r} \in \mathbb{R}^d. \quad (2)$$

Finally, the scattered wavefield measured in the receiver domain,  $y : \Gamma_r \mapsto \mathbb{R}$  is given by

$$y(\mathbf{r}) = \int_{\Omega} g(\mathbf{r} - \mathbf{r}') f(\mathbf{r}') u(\mathbf{r}') d\mathbf{r}', \quad \forall \mathbf{r} \in \Gamma_r. \quad (3)$$

We provide detailed derivation of equations (1), (2), and (3) in Appendix A. 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 of solving equation (2), the *Lippmann-Schwinger* equation (1), and finally, the data equation (3). Generally, we pre-compute the input wavefield  $u_{\text{in}}$  for each wavenumber  $k$ , since it is independent of the contrast function.

In the discrete setting, the scattering equation (1) and data equation (3) reduce to the following system of linear equations for each transmitter illumination and the wavenumber:

$$\begin{aligned} \mathbf{u} &= \mathbf{v} + \mathbf{G} \text{diag}(\mathbf{f}) \mathbf{u}, \\ \mathbf{y} &= \mathbf{H} \text{diag}(\mathbf{f}) \mathbf{u}, \end{aligned} \quad (4)$$

where  $\mathbf{u} \in \mathbb{C}^N$  and  $\mathbf{v} \in \mathbb{C}^N$  are the total and input wavefields, respectively,  $N$  denotes the number of grid points used to discretize the domain  $\Omega$ ,  $\mathbf{f} \in \mathbb{R}^N$  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  $\Gamma$ , 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  $\mathbf{A} := \mathbf{I} - \mathbf{G} \text{diag}(\mathbf{f})$ , where  $\mathbf{I}$  denotes the identity operator. As the discretization dimension  $N$  increases, explicitly forming the matrix  $\mathbf{A}$  and computing 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 media, *i.e.*, for large values of  $\|\mathbf{f}\|_{\infty}$ .

### B. Inverse problem

An *inverse scattering* problem is defined as the estimation of the contrast function given the measurement of the scattered wavefield at  $n_r$  receivers for each input wavefield generated from  $n_t$  transmitters. We use  $\mathcal{J} = \{1, \dots, n_f\}$  and  $\mathcal{I} =$

$\{1, \dots, n_t\}$  to denote the index sets for frequencies and transmitters respectively,  $n_f$  to represent the number of frequencies, and  $n_t$  to represent the number of transmitters. Let  $\mathbf{y}_{ijl}$  be the measured signal at frequency  $j$  and receiver  $l \in \{1, \dots, n_r\}$  and illuminated by transmitter  $i$ . Also,  $\mathbf{H}_{jl}$  denotes the forward operator mapping associated with frequency  $j$  and receiver  $l$ . Assuming that the measurements are contaminated by white Gaussian noise, we can formulate the discrete *inverse* problem as a constrained least-squares problem:

$$\min_{\mathbf{f}} \sum_{l=1}^{n_r} \sum_{j \in \mathcal{J}, i \in \mathcal{I}} \frac{1}{2} \|\mathbf{y}_{ijl} - \mathbf{H}_{jl} \text{diag}(\mathbf{f}) \mathbf{u}_{ij}\|^2, \quad (5)$$

$$\text{subject to } (\mathbf{I} - \mathbf{G}_j \text{diag}(\mathbf{f})) \mathbf{u}_{ij} = \mathbf{v}_{ij} \quad \forall i, j$$

We assume that the  $\mathcal{J}$  is ordered according to the frequencies (in an increasing order). For the rest of this paper,  $\|\cdot\|$  denotes the Euclidean norm (if there is no subscript). In general, problem (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, for each frequency  $j \in \mathcal{J}$ , a data matrix  $\mathbf{Y}_j \in \mathbb{C}^{n_r \times n_t}$ , a wavefield matrix  $\mathbf{U}_j \in \mathbb{C}^{N \times n_t}$  and the input wavefield matrix  $\mathbf{V}_j \in \mathbb{C}^{N \times n_t}$ . Hence, the cost function and the constraint for each frequency takes the form

$$\mathcal{D}_j(\mathbf{f}, \mathbf{U}_j) = \frac{1}{2} \|\mathbf{Y}_j - \mathbf{H}_j \text{diag}(\mathbf{f}) \mathbf{U}_j\|_F^2,$$

$$\mathbf{C}_j(\mathbf{f}, \mathbf{U}_j) = (\mathbf{I} - \mathbf{G}_j \text{diag}(\mathbf{f})) \mathbf{U}_j - \mathbf{V}_j,$$

where  $\|\cdot\|_F$  denotes the Frobenius norm. It is possible to eliminate the wavefields  $\mathbf{U}$  by satisfying the constraints, *i.e.*,  $\mathbf{U}_j^* = (\mathbf{I} - \mathbf{G}_j \text{diag}(\mathbf{f}))^{-1} \mathbf{V}_j$ . Such reduced cost-function at frequency  $j$  is given by

$$\mathcal{F}_j(\mathbf{f}) \triangleq \left\{ \mathcal{D}_j(\mathbf{f}, \mathbf{U}_j) \quad \text{subject to} \quad \mathbf{C}_j(\mathbf{f}, \mathbf{U}_j) = \mathbf{0} \right\}. \quad (6)$$

With the incorporation of this reduced form, the regularized version of (5) now reads

$$\min_{\mathbf{f}} \sum_{j \in \mathcal{J}} \mathcal{F}_j(\mathbf{f}) + \mathcal{R}(\mathbf{f}). \quad (7)$$

Since both the cost function and constraints are nonlinear, we resort to iterative methods to find a feasible solution to the regularized least-squares optimization problem shown above.

### C. 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 large amount 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.

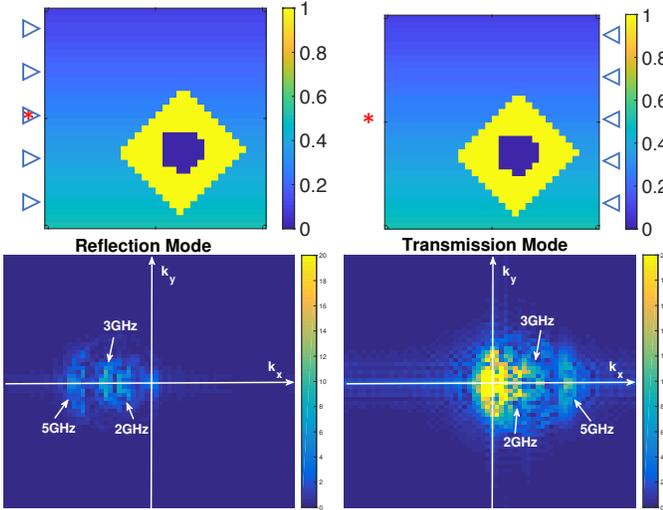


Fig. 2: Comparison of the spatial frequency content of the received wavefields between the transmission mode and the reflection mode from a transmitted pulse containing 2GHz, 3GHz, and 5GHz frequency components.

In order to illustrate this phenomenon, we simulate two sets of measurements  $\{y_T, y_R\}$  of the scattered wavefield from the same object, observed in the transmission and reflection modes through the measurement operators  $\mathbf{H}_T$  and  $\mathbf{H}_R$ , respectively. Figure 2 illustrates the imaging setup where the object is illuminated from its left side by a transmitter, denoted by the red asterisk, with a flat spectrum pulse containing 2, 3, and 5GHz frequency components. Five receivers, denoted by blue triangles, are used to measure the scattered wavefield in both the reflection and transmission regimes. We want to identify the amount of spatial frequency content that is encoded in each of  $y_T$  and  $y_R$  without being affected by the nonconvexity of the inverse problem (5). Therefore, we provide the true scattered wavefields  $\mathbf{U}_j^* = (\mathbf{I} - \mathbf{G}_j \text{diag}(\mathbf{f}^*))^{-1} \mathbf{V}_j$  for each frequency, which reduces (5) to a convex linear inverse problem in  $\mathbf{f}$ . Consequently, we solve the convex form of (5) to compute  $\mathbf{f}$  in each of the transmission and reflection modes and plot in Figure 2 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 spectral energy of the recovered contrast corresponds to the low spatial frequencies. The illustration above 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 the scattering potential of an object, the weak acquisition of spatial frequency information renders the problem severely ill-posed when compared to transmission tomography.

### III. 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.

#### A. Sequential Workflow

The least-squares cost function in (7) provides a natural separation across frequencies. Moreover, the topology of the non-convex 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 3(a). The true target has a reflectivity  $c = 10$  and is illuminated with five transmitters located at a  $y$ -position of  $-0.6\text{m}$ . The transmitters and receivers are collocated and are equidistantly placed between  $x = -0.5\text{m}$  and  $x = 0.5\text{m}$ . We plot in Figure 3(b) the value of the data-fidelity cost function  $\mathcal{F}_j(\mathbf{f})$  for various  $j$  values. Notice that for higher-frequency wavefields, the cost function exhibits many local minima that are farther away from the global minimizer than for the low-frequency wavefields.

A popular approach in the exploration geophysics community is to solve a sequence of inverse problems starting with a low-frequency batch, and then sliding linearly towards the high frequencies keeping the batch-size fixed. In Figure 3(c), we plot such cost function  $\left(\sum_{j \in \mathcal{J}_b} \mathcal{F}_j(\mathbf{f})\right)$  with  $\mathcal{J}_b = \{j_0, j_0 + 1, \dots, j_0 + n_b - 1\}$  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 the low-frequency batch inversions. A more robust approach would be to keep the low-frequencies as regularizer when inverting with high-frequency data. We plot the cost function  $\left(\sum_{j=1}^{j_{\max}} \mathcal{F}_j(\mathbf{f})\right)$  in Figure 3(d). Notice that the cost functions are almost convex even when dealing with high-frequency data. However, the functions are well-behaved primarily due to the very simplistic setting of this example, where everything about the target is known except for the permittivity  $c$ . The function behavior will be significantly more erratic when the structure of the target and its surrounding medium are unknown.

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.

for  $k = 1, \dots, n_f$ :

$$\mathbf{f}^{(k)} \triangleq \underset{\mathbf{f}}{\text{argmin}} \left\{ \sum_{j \in \mathcal{J}_k} \mathcal{F}_j(\mathbf{f}) + \mathcal{R}(\mathbf{f}) \right\}. \quad (8)$$

Therefore, instead of solving a single nonconvex minimization problem in (7), we solve  $n_f$  subproblems sequentially according to (8), where the sequence of solutions moves us closer to the global minimizer of (7).

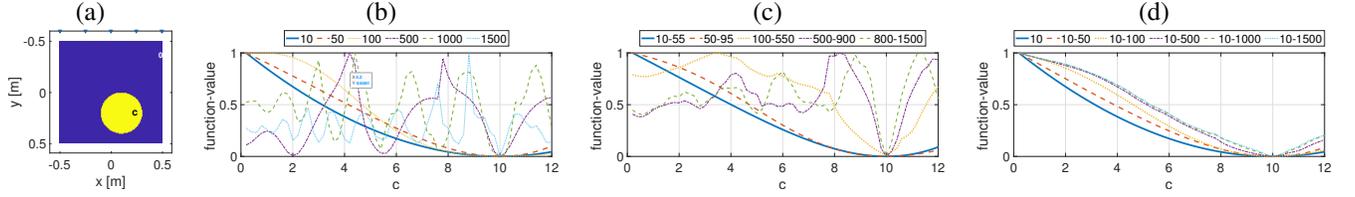


Fig. 3: (a) Illustration of a cylindrical object with true reflectivity equal to  $c^* = 10$  measured by five co-located transmitters and receivers. Topology of the cost function per frequency (b), per frequency batch of size 10 (c), and incremental frequency batch (d) relative to the estimated reflectivity  $c$ . All the frequencies are in MHz.

### B. Regularization

In this section we provide details on the total variation norm and the non-negativity constraints we use to regularize the problem, as well as their implementation through a proximal operator.

1) *Total-variation*: The Total-Variation (TV) norm of a compactly supported function  $u : \Omega \mapsto \mathbb{R}$  is formally defined as

$$TV(u) \triangleq \sup \left\{ \int_{\Omega} u(\mathbf{x}) \operatorname{div} \phi \, d\mathbf{x} : \|\phi\|_{\infty} \leq 1 \right\},$$

where  $\phi \in \mathcal{C}_c^1(\Omega, \mathbb{R}^d)$  denotes the set of continuously differentiable functions of compact support contained in  $\Omega$ . This norm measures the total change in the function over a finite domain [37]. If  $u$  is differentiable, then we can express the total-variation using an integral

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

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 [38]. In a discrete two dimensional setting, the TV-norm is represented as

$$TV(\mathbf{f}) = \|\mathbf{D}\mathbf{f}\|_1 \quad \text{where} \quad \mathbf{D} = \begin{bmatrix} \mathbf{I}_x \otimes \mathbf{D}_y \\ \mathbf{D}_x \otimes \mathbf{I}_y \end{bmatrix}. \quad (9)$$

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 [33], such that,

$$\mathcal{R}_{TV}(\mathbf{f}) \triangleq \delta_{TV \leq \tau}(\mathbf{f}) = \delta_{\|\cdot\|_1 \leq \tau}(\mathbf{D}\mathbf{f}). \quad (10)$$

where  $\delta_{\mathcal{C}}(\cdot)$  is an indicator function to the set  $\mathcal{C}$ , and  $\tau$  is a constraint parameter. The second line in (10) expresses the discretized version of the TV regularization function using the constrained  $\ell_1$ -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  $\ell_1$ -norm ball (see, for example, [39]).

2) *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 using a regularization

$$\mathcal{R}_{NN}(\mathbf{f}) = \delta_{\geq 0}(\mathbf{f}),$$

where  $\delta_{\geq 0}$  denotes the indicator to a non-negative orthant. The proximal operator for this function corresponds the projection of a vector onto a non-negative orthant. In specific, the

projection operator is

$$\mathcal{P}_{\geq 0}(y) = \begin{cases} y & \text{if } y \geq 0 \\ 0 & \text{otherwise} \end{cases}.$$

3) *Implementation*: In order to impose the non-negative + TV constraints, we define the proximal operator:

$$\operatorname{prox}_{\gamma\mathcal{R}}(\mathbf{w}) \triangleq \operatorname{argmin}_{\mathbf{f}} \left\{ \frac{1}{2\gamma} \|\mathbf{f} - \mathbf{w}\|^2 + \mathcal{R}_{TV}(\mathbf{f}) + \mathcal{R}_{NN}(\mathbf{f}) \right\}, \quad (11)$$

with  $\gamma > 0$ . The proximal operator becomes a projection onto the intersection of the sets: the TV-norm ball set and the non-negative orthant set. 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) [40] and/or primal-dual method [41]. Here, we use the primal-dual method, which we derive for the sum of three convex functions in Appendix C. Algorithm 1 describes the primal-dual method to solve (11).

### C. Proximal Quasi-Newton Method

To solve each subproblem in (8), *i.e.*,

$$\mathbf{f}^* = \operatorname{argmin}_{\mathbf{f}} \left\{ \sum_{j \in \mathcal{J}_k} \mathcal{F}_j(\mathbf{f}) + \mathcal{R}(\mathbf{f}) \right\}. \quad (12)$$

we propose a proximal Quasi-Newton (prox-QN) method. For simplicity of illustration, we enumerate the steps in Algorithm 2, but provide a complete derivation in Appendix D.

The algorithm consists of two loops. The inner loop, implicit in step 4 and described in Appendix D, finds the search direction, while the outer loop computes the next iterate based on the computed search direction and step length. At every

---

#### Algorithm 1 Proximal for Non-negative + Total-Variation

---

**Input:**  $\mathbf{w} \in \mathbb{R}^n, \mathbf{D} \in \mathbb{R}^{m \times n}, \gamma > 0, \tau > 0, t_{\max}$

**Output:**  $\mathbf{f}^* \approx \mathbf{f}_{t_{\max}}$

- 1:  $\mathbf{f}_0 = \mathbf{0}, \mathbf{u}_0 = \mathbf{0}, \mathbf{v}_0 = \mathbf{0}$
  - 2: choose  $\alpha \in \left(0, 1/\sqrt{\|\mathbf{D}^T \mathbf{D} + \mathbf{I}\|}\right)$
  - 3: **while**  $t < t_{\max}$  **do**
  - 4:  $\hat{\mathbf{f}} = \mathbf{f}_t - \alpha (\mathbf{D}^T \mathbf{u}_t + \mathbf{v}_t)$
  - 5:  $\mathbf{f}_{t+1} = (\gamma \hat{\mathbf{f}} + \alpha \mathbf{w}) / (\alpha + \gamma)$
  - 6:  $\mathbf{u}_{t+1} = \mathbf{u}_t - \alpha \mathcal{P}_{\|\cdot\|_1 \leq \tau}(\mathbf{u}_t / \alpha + \mathbf{D} (2\mathbf{f}_{t+1} - \mathbf{f}_t))$
  - 7:  $\mathbf{v}_{t+1} = \mathbf{v}_t - \alpha \mathcal{P}_{\geq 0}(\mathbf{v}_t / \alpha + (2\mathbf{f}_{t+1} - \mathbf{f}_t))$
  - 8:  $t = t + 1$
  - 9: **end while**
-

---

**Algorithm 2** Prox-QN method for solving (12)

---

**Input:**  $\mathbf{f}^{(0)}, \tau > 0, \gamma \in (0, 1)$

**Output:**  $\mathbf{f}^*$

- 1: **for**  $i = 0$  to  $i_{\max}$  **do**
  - 2:   compute the gradient  $g_i = \sum_{j=1}^k \nabla \mathcal{F}_j(\mathbf{f}^{(i)})$ .
  - 3:   compute the approximate Hessian  $\mathbf{H}_i$
  - 4:   compute  $\mathbf{s}_i$  from equation (D.5)(a).
  - 5:   define  $\hat{\mathbf{f}}(\alpha) = \mathbf{prox}_{\gamma \mathcal{R}}(\mathbf{f}^{(i)} + \alpha \mathbf{s}_i)$
  - 6:    $\alpha_i = \text{linesearch}_{\alpha}(\hat{\mathbf{f}}(\alpha))$ .
  - 7:    $\mathbf{f}^{(i+1)} := \mathbf{prox}_{\gamma \mathcal{R}}(\mathbf{f}^{(i)} + \alpha_i \mathbf{s}_i)$  using Algorithm 1.
  - 8:   check optimality conditions
  - 9: **end for**
- 

sequence of the outer loop, we 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 Appendix B. Once we have the gradient and approximate Hessian at the current iterate, we compute the search direction using the primal-dual method (see (D.6)). Next, we search for the feasible step length using the backtracking linesearch. Finally, we compute the next iterate using Algorithm 1.

The computational complexity of Prox-QN method relies on step 2 of Algorithm 2. The gradient computation involves solving an adjoint of Lippmann-Schwinger equation. We use GMRES method which has complexity of  $\mathcal{O}(nt^2)$ , where  $n$  is the size of image, and  $t$  is the number of iterations. For  $n_t$  transmitters and  $n_f$  frequencies, the step 2 involves  $\mathcal{O}(nt^2 n_t n_f)$  floating-point operations. Steps 3 to 8 have lower complexity order than that of step 2. Hence, Prox-QN has  $\mathcal{O}(nt^2 n_t n_f i_{\max})$  computational complexity. For a single frequency of 100 Hz ( $n_f = 1$ ) with 5 transmitters on a  $32 \times 32$  image, it took approximately 2 minutes to run 100 iterations of prox-QN on a 2.7 GHz Intel Core i5 processor with 8 GB RAM.

#### IV. ESTIMATING THE CONSTRAINT PARAMETER $\tau$

Recall that for each subproblem (12) in our proposed framework, we are solving a TV-constrained nonlinear least-squares problem where the constraint parameter  $\tau_k$  should bound the total variation of the solution. Naturally, the choice of constraint parameter  $\tau_k$  would significantly affect the reconstruction performance.

In order to estimate  $\tau_k$  for each new subproblem, we develop a parameter estimation routine inspired by the approach in [42] 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  $k^{\text{th}}$  subproblem for which the TV norm  $\tau_k = TV(\mathbf{f}^k)$ , specifically,

$$\mathbf{f}^k = \underset{\mathbf{f}}{\operatorname{argmin}} \left\{ \sum_{j \in \mathcal{J}_k} \mathcal{F}_j(\mathbf{f}) \text{ s.t. } \|\mathbf{Df}\|_1 \leq \tau_k \right\}, \quad (13)$$

where  $\mathcal{F}_j$  is as defined in (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  $\mathcal{F}_{k+1}(\mathbf{f})$  is added to the objective function, resulting in the potentially unsatisfied constraint

$$\mathbf{V}_{k+1} = \mathbf{A}_k \mathbf{U}_{k+1}, \quad (14)$$

where  $\mathbf{A}_k \triangleq \mathbf{I} - \mathbf{G} \operatorname{diag}(\mathbf{f}^k)$ . To overcome the nonconvexity of the objective function due to (14), we linearize the objective function around  $\mathbf{f}^k$  by estimating  $\mathbf{U}_{k+1}^* = \mathbf{A}_k^{-1} \mathbf{V}_{k+1}$ , thus reducing  $\mathcal{F}_{k+1}(\mathbf{f})$  to a convex least squares cost function in  $\mathbf{f}$ , i.e.,

$$\mathcal{F}_{k+1}(\mathbf{f}) \approx \mathcal{D}_{k+1}(\mathbf{f}, \mathbf{U}_{k+1}^*),$$

where  $\mathcal{D}_{k+1}(\mathbf{f}, \mathbf{U}_{k+1}^*)$  is the data mismatch cost function defined in (6). Consequently, we may now define a value function  $\Phi(\tau)$  for the  $(k + 1)^{\text{th}}$  subproblem as

$$\begin{aligned} \Phi(\tau) &= \underset{\mathbf{f}}{\operatorname{argmin}} \left\{ \sum_{j \in \mathcal{J}_{k+1}} \mathcal{D}_j(\mathbf{f}, \mathbf{U}_j^*) \text{ s.t. } \|\mathbf{Df}\|_1 \leq \tau \right\} \\ &= \underset{\lambda}{\operatorname{argmax}} \left\{ \sum_{j \in \mathcal{J}_{k+1}} \mathbf{r}_j^H \mathbf{Y}_j / \|\mathbf{r}^{k+1}\| - \tau \lambda \right. \\ &\quad \left. \text{s.t. } \operatorname{TV}_{\text{polar}} \left( \frac{\sum_{j \in \mathcal{J}_{k+1}} \operatorname{diag}(\mathbf{U}_j^*) \mathbf{H}_j^H \mathbf{r}_j}{\|\mathbf{r}^{k+1}\|} \right) \leq \lambda \right\} \end{aligned} \quad (15)$$

where  $\mathbf{r}_j = \mathbf{Y}_j - \mathbf{H}_j \operatorname{diag}(\mathbf{U}_j^*) \mathbf{f}_k$  is the data residual at the  $j^{\text{th}}$  frequency, and  $\mathbf{r}_{k+1}$  is the vector formed by concatenating all the vectors  $\mathbf{r}_j$ , such that,  $\|\mathbf{r}^{k+1}\|^2 = \sum_{j \in \mathcal{J}_{k+1}} \|\mathbf{r}_j\|^2$ . The  $\operatorname{TV}_{\text{polar}}$  function is defined as  $\operatorname{TV}_{\text{polar}}(\mathbf{x}) = \|\mathbf{D}^{-T} \mathbf{x}\|_{\infty}$ , with  $\mathbf{D}^{-T}$  being the transposed pseudo-inverse of the finite difference operator  $\mathbf{D}$  defined in (9). Note that (15) shows the primal and dual problems for computing the value function  $\Phi(\tau)$ .

The dual problem in (15) conveniently shows that the maximum is achieved when  $\lambda$  is at its minimum  $\lambda^* = \operatorname{TV}_{\text{polar}} \left( \sum_{j \in \mathcal{J}_{k+1}} \operatorname{diag}(\mathbf{U}_j^*) \mathbf{H}_j^H \mathbf{r}_j / \|\mathbf{r}\| \right)$ . Moreover, the gradient of  $\Phi(\tau)$  with respect to  $\tau$  is easily computed as  $\nabla_{\tau} \Phi(\tau) = \lambda^*$ . Therefore, we can compute the update for  $\tau$  using a Newton root finding step, such that,

$$\tau_{k+1} = \tau_k + \frac{\|\mathbf{r}^{k+1}\| (\|\mathbf{r}^{k+1}\| - \sigma_{k+1})}{\operatorname{TV}_{\text{polar}} \left( \sum_{j \in \mathcal{J}_{k+1}} \operatorname{diag}(\mathbf{U}_j^*) \mathbf{H}_j^H \mathbf{r}_j \right)}, \quad (16)$$

where  $\sigma_{k+1}$  is the upper bound on the  $\ell_2$  norm of the noise up to the  $k + 1$  frequency bin. Finally, we note that at the zeroth iteration, the parameter  $\tau$  can be set to zero, resulting in a homogeneous solution for  $\mathbf{f}_0$ .

#### V. 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 in Section V-E.

##### A. Experimental details

We consider an experimental setup illustrated in Figure 4(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

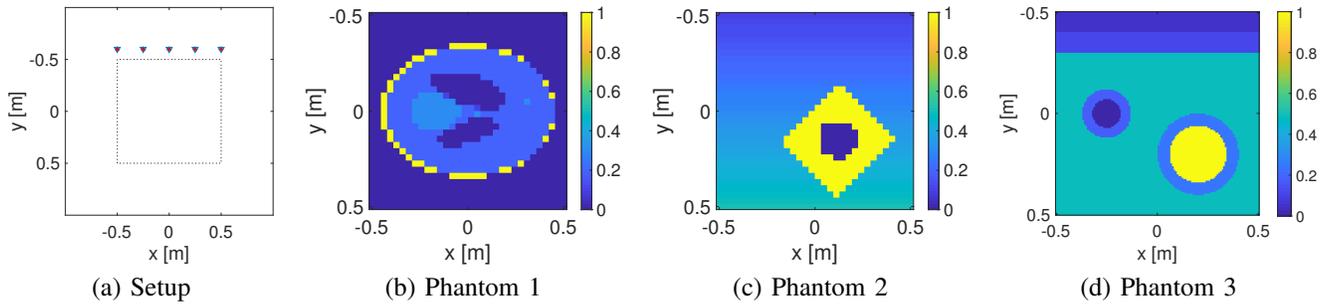


Fig. 4: (a) Tomography setup for all the numerical experiments. The dotted region denotes the object domain  $\Omega$ . The transmitters and receivers are collocated at  $y = -0.6$  m. (b), (c), (d) are the three numerical phantoms used for the experimentation.

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  $j = 0, \dots, 17$ , *ii*) a medium frequency band consisting of  $\{100 + 50j\}$  MHz with  $j = 0, \dots, 17$ , and *iii*) a high frequency band consisting of  $\{1000 + 100j\}$  MHz with  $j = 0, \dots, 10$ . Hence, in total, we consider 47 frequencies between 10 MHz and 2000 MHz.

We work with 3 phantoms shown in Figure 4(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  $32 \times 32$  grid. It has 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  $32 \times 32$ . We use these low resolution phantoms to compare our method with other existing 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  $128 \times 128$ . 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 vacuum inside. We use this phantom to perform a partially non-inverse-crime test as described in Section V-E.

### B. 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 [43].

**CISOR:** The CISOR algorithm aims to solve (5) by taking all frequencies at once [43]. As opposed to TV-norm penalization, we use the proposed regularization, *i.e.*, we regularize it with non-negative and total-variation constraints with known  $\tau$  value. The TV constraint

parameter is set to the total-variation of the true model. The problem is solved using a prox-QN method with a maximum of 5000 iterations or until convergence (norm of the gradient below  $10^{-6}$ ).

**RL:** Recursive linearization (RL) method was introduced in [44], and has been a standard while working with multi-frequency data. The method enjoys the computational benefit of solving a single constraint (*i.e.*, solving a single linear system of equations) at a time, but might suffer in the high-contrast regime. It solves the sequence of problems

$$\mathbf{f}^{(j)} \triangleq \underset{\mathbf{f}}{\operatorname{argmin}} \left\{ \mathcal{D}_j(\mathbf{f}, \mathbf{U}_j) \quad \text{s.t.} \quad \mathbf{C}_j(\mathbf{f}, \mathbf{U}_j) = \mathbf{0} \right\},$$

with an initial guess to each subproblem being the solution of the previous subproblem. We modify the cost function to include the regularization. Similarly to the CISOR, we consider non-negative and TV regularization with known  $\tau$  value. Each subproblem is solved using a prox-QN method with a maximum of 500 iterations.

**SF- $\tau$ :** This method corresponds to the proposed sequential framework with known  $\tau$  value. It solves the problem described in (8). We use a prox-QN method to solve each subproblem with a maximum of 500 iterations or until converge.

**SF- $\sigma$ :** This method corresponds to the proposed sequential framework with estimation of  $\tau$  at each iteration. It solves the problem described in (8) with the  $\tau$  estimation from (16). Here, we assume that the noise-level  $\sigma$  is known. We use a prox-QN method to solve each subproblem with a maximum of 500 iterations or until converge.

For all the methods the initial model corresponds to a contrast of 0 everywhere.

### C. Performance Measures

We use the following measures to evaluate the performance of the proposed methods and to compare with other methods.

**DR:** The data residual (DR) measures the distance of the modeled data for reconstructed model with the actual

data in the euclidean sense. For multi-frequency data the DR takes the following form

$$\text{DR} \triangleq 100 \times \frac{\sum_{j \in \mathcal{J}} \mathcal{F}_j(\mathbf{f}^*)}{\sum_{j \in \mathcal{J}} \|\mathbf{Y}_j\|^2},$$

where  $\mathbf{f}^*$  is the reconstructed solution. Here,  $\|\mathbf{Y}\|$  denotes the Frobenius norm for the matrix  $\mathbf{Y}$ . DR

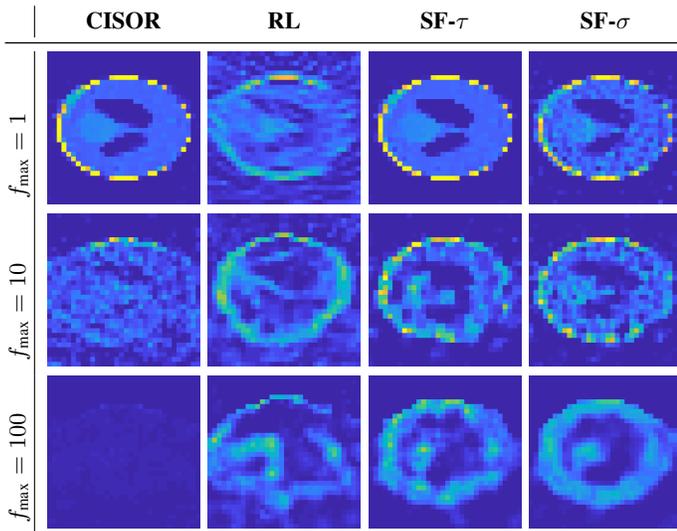


Fig. 5: Comparison of methods on Phantom 1.

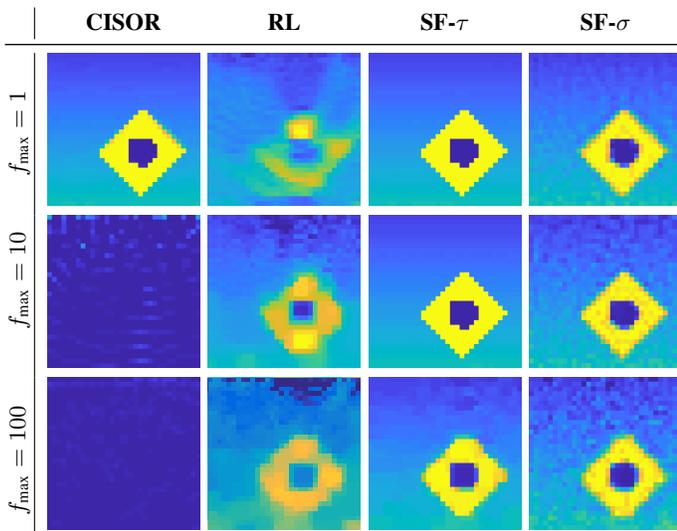


Fig. 6: Comparison of methods on Phantom 2.

TABLE I: Comparison of methods on Phantom 1 and Phantom 2

$f_{\max}$		Phantom 1				Phantom 2			
		CISOR	RL	SF- $\tau$	SF- $\sigma$	CISOR	RL	SF- $\tau$	SF- $\sigma$
1	DR	0.87	45.89	0.74	2.36	0.32	29.16	0.05	0.06
	SNR	14.73	3.87	15.12	9.19	27.63	8.84	42.79	18.22
10	DR	28.08	75.27	8.78	11.13	945.16	260.43	3.77	24.75
	SNR	2.17	1.94	3.83	4.47	0.16	11.08	47.07	18.00
100	DR	295.41	97.12	2.72	4.95	344.59	69.02	5.76	1.52
	SNR	0.27	1.15	2.60	3.08	0.11	10.72	17.18	14.42

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  $\mathbf{f}^*$  with respect to the ground truth  $\mathbf{f}^{\text{true}}$  is

$$\text{SNR} \triangleq -20 \log_{10} \left( \frac{\|\mathbf{f}^* - \mathbf{f}^{\text{true}}\|}{\|\mathbf{f}^{\text{true}}\|} \right).$$

A reconstruction is considered good if it has high SNR. This measure is only available if we know the ground truth.

#### D. Exact-model experiments

To evaluate our approach, we perform both noise-free and noisy experiments, in which the exact model is known. In the noise-free experiments, we compare our methods with the other two methods (CISOR and RL). In noisy ones, we only examine the robustness of our methods against various levels of noise.

1) *Noise-free experiment:* We consider Phantom 1 and 2 for this experiment. We produce three types of phantoms by scaling these phantoms with a maximum contrast ( $f_{\max}$ ) of  $\{1, 10, 100\}$ , which we consider to be low, medium, and

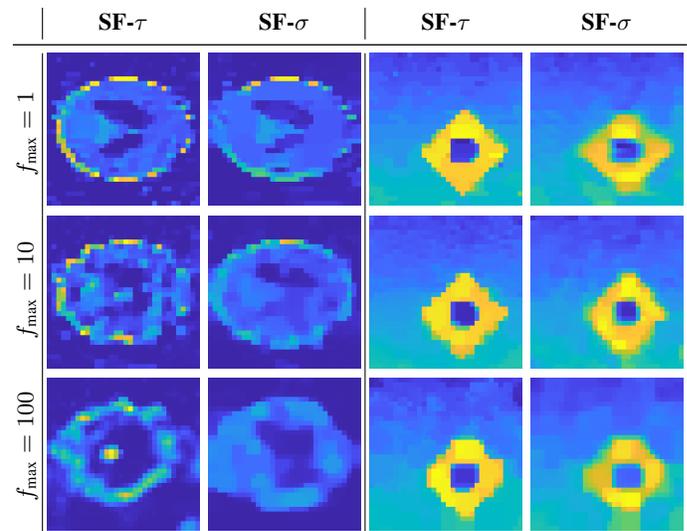


Fig. 7: Noise robustness of proposed methods on Phantom 1 and Phantom 2 with 20% noise. Left and right columns show reconstructions of Phantom 1 and 2 respectively.

TABLE II: Noise-Robustness of SF- $\tau$  and SF- $\sigma$

$f_{\max}$		Phantom 1				Phantom 2			
		SF- $\tau$		SF- $\sigma$		SF- $\tau$		SF- $\sigma$	
		10%	20%	10%	20%	10%	20%	10%	20%
1	DR	7.00	15.10	13.20	25.79	9.53	19.77	11.50	22.73
	SNR	10.59	8.64	6.97	5.35	19.74	15.77	14.06	12.04
10	DR	19.91	35.38	21.77	51.59	41.49	52.53	42.24	62.78
	SNR	3.23	3.23	4.37	3.43	18.98	15.15	14.68	12.34
100	DR	42.42	74.96	49.12	87.96	27.85	49.16	35.49	73.38
	SNR	2.10	1.27	3.15	2.54	13.69	14.02	12.43	11.46

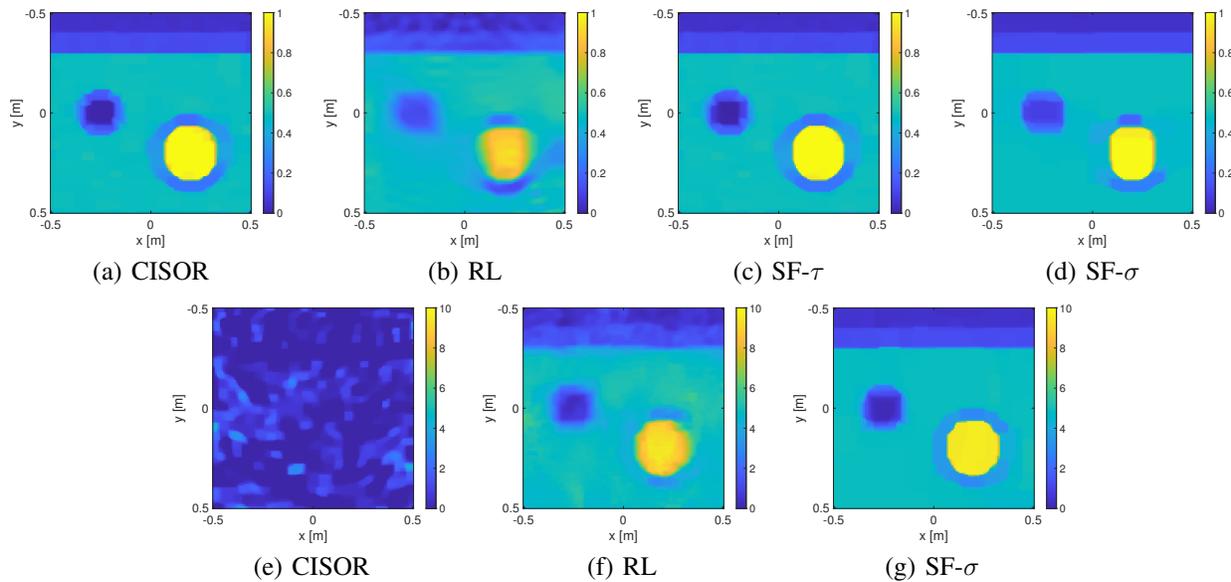


Fig. 8: Inexact model experiments on Phantom 3. The top row shows the reconstructions on low-contrast phantom, while the bottom row shows the reconstructions on medium-contrast phantom.

high-contrast phantoms, respectively. For the simulations we use the reflection tomography setup illustrated in Figure 4(a) with noiseless data. We examine the performance of the methods SF- $\tau$  and SF- $\sigma$ , and compare it with the CISOR and RL method. Figures 5 and 6 show the reconstructions for various contrast levels for Phantom 1 and 2, respectively. We see that SF- $\tau$  consistently performs well except in the case of  $f_{\max} = 100$  for Phantom 1, where all the methods fail. The reason for the failure is that Phantom 1 is ideal for 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 I. We conclude that the SF- $\tau$  and SF- $\sigma$  perform superior to the existing methods (CISOR and RL).

2) *Noisy experiment:* We consider Phantom 1 and 2, with the scaling  $\{1, 10, 100\}$ . We add a Gaussian noise of relative energy 10% (20dB measurement SNR), and 20% (14dB measurement SNR), and examine the performance of SF- $\tau$  and SF- $\sigma$  on these noise levels. Figure 7 shows the reconstructions using these methods for 20% relative noise energy and various levels of contrast values. The performance measures are tabulated in Table II. We observe that SF- $\tau$  and SF- $\sigma$  are robust against high noise in the low-contrast phantoms. SF- $\tau$  is also stable for moderate level of noise in high-contrast regime.

#### E. Inexact-model experiment

To verify the robustness of our approach, we consider Phantom 3 for this test that has a resolution of  $128 \times 128$ , using an inexact model for the reconstruction. To avoid *inverse-crime*, we generate the measurements with a high-resolution modeling grid with additional Gaussian noise, and use a low-resolution grid as a forward solver [45]. In particular, we first discretize the model on a high-resolution grid of size

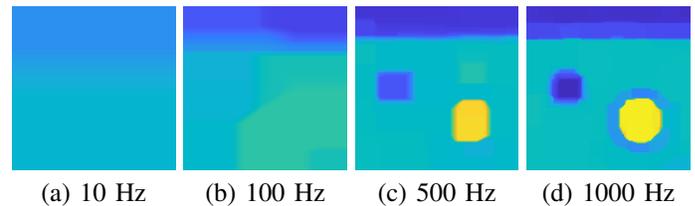


Fig. 9: Intermediate reflectivity maps from SF- $\sigma$  for the inexact model for Phantom 3 on medium-contrast phantom highlighting the low-to-high spatial resolution of the reconstruction.

$192 \times 192$ . We use the nearest-neighbor algorithm for the rescaling to a high-resolution grid. We generate the data on the high-resolution grid, and add 10% (20dB measurement SNR) white Gaussian noise relative to the signal power. As a sanity check, we look at the difference between the data for high-resolution and low-resolution model, and found the relative difference is less than 20%. We test CISOR, RL, SF- $\tau$  and SF- $\sigma$  on low-contrast phantom ( $f_{\max} = 1$ ) with this high-resolution dataset. We assume a noise level of 20% for SF- $\sigma$ , while we set  $\tau$  to be the TV-value of the ground truth (low-resolution model) for CISOR, RL and SF- $\tau$ . The reconstruction results for these methods are presented in Figure 8(a)-(d). CISOR has DR of 1.47 and SNR of 20.32dB, while RL has DR of 12.52 and SNR of 14.23dB. Similarly, SF- $\tau$  has DR of 1.46, and SNR of 19.95dB, while SF- $\sigma$  has a DR of 4.74, and SNR of 16.71dB. We observe that CISOR and SF- $\tau$  are able to reconstruct the ground scene accurately: the top and the bottom regions of the pipes are retrieved to high precision. SF- $\sigma$  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. Moreover, RL struggles to predict the accurate geometries of pipes as well as layers. As we have seen in Figures 5 and 6, we conclude that the low-contrast phantom can be reconstructed with CISOR, SF- $\sigma$  and SF- $\tau$ .

Next, we run a similar test for medium-contrast phantom,

*i.e.*, Phantom 3 scaled to the maximum contrast of 10. For this test, we only show the reconstructions from SF- $\sigma$  with CISOR and RL, since SF- $\tau$  consistently performs better. Figure 8(e)-(g) provide reconstruction results for these methods. The CISOR has DR of 342.82, SNR of 0.64dB, while RL achieves DR of 24.24 and SNR of 15.19dB. Compared to these methods, SF- $\sigma$  has a DR of 13.78 and SNR of 18.73dB. Hence, CISOR fails on medium-contrast phantoms, while RL still struggles to provide an accurate picture of target image. On the other hand, the reconstruction from SF- $\sigma$  quite accurately recovers the layers and the pipes.

Finally, we plot in Figure 9 the intermediate solutions from SF- $\sigma$  on medium-contrast phantoms. In particular, we plot the solutions for batches 10 Hz, 10-100 Hz, 10-500 Hz, and 10-1000 Hz. We observe that the solution at 10 Hz obtains an almost constant image due to the low TV constraint. As we move towards higher frequencies, the reconstruction method starts to fill in the details in the image by allowing for higher values of TV and higher frequencies in the measurements. This behavior is reminiscent of “multi-grid methods” while avoiding their complex bookkeeping requirements.

## VI. CONCLUSIONS

We consider limited-angle reflection tomography of high-contrast objects and 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 develop a regularized multiscale approach. We pose the imaging problem as a nonlinear least-squares problem with constraints. The cost function includes the wave-based modeling that accounts for multiple scattering and a regularization term that includes non-negativity and total variation constraints. The total cost function is decomposed according to the frequency, and we observe that the low-frequencies promote smoothness while higher frequencies add details in the reconstruction. Hence, we solve a sequence of subproblems, where the  $k^{\text{th}}$  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 also 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 is robust against moderate noise. The proposed techniques can retrieve high-contrast object (contrast up to 100) for scenes similar to the underground.

## APPENDIX A SCATTERING FORMALISM

Consider a scattering setup illustrated in Figure 1. The scene (free-space with permittivity  $\epsilon_b$ ) has a dimension  $d$ . The transmitter domain  $\Gamma_t \subset \mathbb{R}^d$  emits a source function  $q : \Gamma_t \mapsto \mathbb{C}$ , which generates an incident wavefield  $u_{\text{in}} : \mathbb{R}^d \mapsto \mathbb{C}$  everywhere. This incident wavefield interacts with an object in domain  $\Omega \subset \mathbb{R}^d$  and generates a total wavefield  $u : \mathbb{R}^d \mapsto \mathbb{C}$ .

The scattered wavefield  $u_{\text{sc}} := u - u_{\text{in}}$  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  $u_{\text{sc}}(\mathbf{r})$ ,

$$u(\mathbf{r}) = u_{\text{in}}(\mathbf{r}) + u_{\text{sc}}(\mathbf{r}), \quad \mathbf{r} \in \mathbb{R}^d. \quad (\text{A.1})$$

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

$$\nabla^2 u_{\text{in}}(\mathbf{r}) - k^2 \epsilon_b u_{\text{in}}(\mathbf{r}) = -q(\mathbf{r}) \quad \forall \mathbf{r} \in \mathbb{R}^d,$$

where  $k$  denotes the wavenumber. It is convenient to consider the above equation for inside and outside the object domain  $\Omega$ :

$$\begin{aligned} \nabla^2 u_{\text{in}}(\mathbf{r}) - k^2 \epsilon_b u_{\text{in}}(\mathbf{r}) &= 0 & \forall \mathbf{r} \in \Omega, \\ \nabla^2 u_{\text{in}}(\mathbf{r}) - k^2 \epsilon_b u_{\text{in}}(\mathbf{r}) &= -q(\mathbf{r}) & \forall \mathbf{r} \notin \Omega, \end{aligned} \quad (\text{A.2})$$

Similarly, the total wavefield satisfies the Helmholtz equation, and we can express it inside and outside the domain as follows,

$$\begin{aligned} \nabla^2 u(\mathbf{r}) - k^2 \epsilon(\mathbf{r}) u(\mathbf{r}) &= 0 & \forall \mathbf{r} \in \Omega, \\ \nabla^2 u(\mathbf{r}) - k^2 \epsilon_b u(\mathbf{r}) &= -q(\mathbf{r}) & \forall \mathbf{r} \notin \Omega, \end{aligned} \quad (\text{A.3})$$

where  $\epsilon(r)$  is the permittivity of the object. Now, from the equations (A.1), (A.2) and (A.3), the governing equation for the scattered wavefield reads

$$\begin{aligned} \nabla^2 u_{\text{sc}}(\mathbf{r}) - k^2 \epsilon_b u_{\text{sc}}(\mathbf{r}) &= -k^2 (\epsilon_b - \epsilon(\mathbf{r})) u(\mathbf{r}) & \forall \mathbf{r} \in \Omega, \\ \nabla^2 u_{\text{sc}}(\mathbf{r}) - k^2 \epsilon_b u_{\text{sc}}(\mathbf{r}) &= 0 & \forall \mathbf{r} \notin \Omega, \end{aligned}$$

These equations can be compactly written as

$$\nabla^2 u_{\text{sc}}(\mathbf{r}) - k^2 \epsilon_b u_{\text{sc}}(\mathbf{r}) = -k^2 f(\mathbf{r}) u(\mathbf{r}) \quad \forall \mathbf{r} \in \mathbb{R}^d \quad (\text{A.4})$$

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  $\Omega$  and 0 outside. We supplement the scattered wavefield equation (A.4) with the Sommerfeld radiation condition

$$\lim_{r \rightarrow \infty} r \left( \frac{\partial u_{\text{sc}}}{\partial r} - i k u_{\text{sc}} \right) = 0$$

where  $r = \|\mathbf{r}\|$ . Equation (A.4) can be converted to an equivalent integral equation by introducing the free space Green function. The free space Green function  $g : \mathbb{R}^d \mapsto \mathbb{R}^d$  satisfies

$$\nabla^2 g(\mathbf{r}) + k^2 \epsilon_b g(\mathbf{r}) = -\delta(\mathbf{r}), \quad \forall \mathbf{r} \in \mathbb{R}^d$$

together with the Sommerfeld radiation conditions. Here,  $\delta$  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

$$u_{\text{in}}(\mathbf{r}) = k^2 \int_{\mathbf{r}' \in \Gamma_t} g(\mathbf{r} - \mathbf{r}') q(\mathbf{r}') d\mathbf{r}' \quad \forall \mathbf{r} \in \mathbb{R}^d,$$

and similarly, for the scattered wavefield is

$$u_{\text{sc}}(\mathbf{r}) = k^2 \int_{\mathbf{r}' \in \Omega} g(\mathbf{r} - \mathbf{r}') u(\mathbf{r}') f(\mathbf{r}') d\mathbf{r}' \quad \forall \mathbf{r} \in \mathbb{R}^d.$$

Noting that the scattered wavefield is the difference of the total wavefield and the input wavefield (see (A.1)) and restricting our observations to the object domain  $\Omega$ , we arrive at the well-known Lippmann-Schwinger equation

$$u(\mathbf{r}) = u_{\text{in}}(\mathbf{r}) + k^2 \int_{\mathbf{r}' \in \Omega} g(\mathbf{r} - \mathbf{r}') u(\mathbf{r}') f(\mathbf{r}') d\mathbf{r}' \quad \forall \mathbf{r} \in \Omega$$

The equation above describes the relation between the total-wavefield and the contrast function inside the object domain  $\Omega$ . The scattered wavefield is then measured in the receiver domain  $\Gamma_r$  resulting in the following data equation:

$$y(\mathbf{x}) = \int_{\Omega} g(\mathbf{x} - \mathbf{r}) f(\mathbf{r}) u(\mathbf{r}) d\mathbf{r}, \quad \forall \mathbf{x} \in \Gamma_r.$$

## APPENDIX B GRADIENT COMPUTATION

In this section, we derive a gradient for an equality constrained cost function

$$\mathcal{F}(\mathbf{f}) = \left\{ h(\mathbf{f}, \mathbf{u}) \quad \text{subject to} \quad \mathbf{k}(\mathbf{f}, \mathbf{u}) = \mathbf{0} \right\} \quad (\text{B.1})$$

where  $h : \mathbb{R}^n \times \mathbb{C}^n \mapsto \mathbb{R}$  is a real-valued function and  $\mathbf{k} : \mathbb{R}^n \times \mathbb{C}^n \mapsto \mathbb{C}^n$  is a set valued function. We assume that both the functions  $h$  and  $\mathbf{k}$  are smooth and hence, differentiable. For the constrained problem (B.1), the Lagrangian reads

$$\mathcal{L}(\mathbf{f}, \mathbf{u}, \boldsymbol{\lambda}) = h(\mathbf{f}, \mathbf{u}) + \boldsymbol{\lambda}^H \mathbf{k}(\mathbf{f}, \mathbf{u}), \quad (\text{B.2})$$

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 vector  $\mathbf{x}$  with complex entries. The stationary point of the Lagrangian  $\mathcal{L}$ , denoted by  $(\mathbf{f}, \mathbf{u}^*, \boldsymbol{\lambda}^*)$ , satisfies

$$\frac{\partial \mathcal{L}}{\partial \mathbf{u}} = \mathbf{0}, \quad \frac{\partial \mathcal{L}}{\partial \boldsymbol{\lambda}} = \mathbf{0}.$$

The first condition gives rise to an adjoint equation

$$\frac{\partial h}{\partial \mathbf{u}}(\mathbf{f}, \mathbf{u}^*) + \left( \frac{\partial \mathbf{k}}{\partial \mathbf{u}}(\mathbf{f}, \mathbf{u}^*) \right)^H \boldsymbol{\lambda}^* = \mathbf{0}, \quad (\text{B.3})$$

while the second condition is the states equation

$$\mathbf{k}(\mathbf{f}, \mathbf{u}^*) = \mathbf{0}. \quad (\text{B.4})$$

The states equation generates a wavefield  $\mathbf{u}^*$  for a given parameter value  $\mathbf{f}$ . The adjoint equation calculates the Lagrange multiplier (also called adjoint wavefield) corresponding to wavefield  $\mathbf{u}^*$  for given  $\mathbf{f}$ . The gradient of  $\mathcal{F}$  is now retrieved from the partial derivative of the Lagrangian with respect to  $\mathbf{f}$ ,

$$\nabla \mathcal{F}(\mathbf{f}) = \frac{\partial \mathcal{L}}{\partial \mathbf{f}} = \frac{\partial h}{\partial \mathbf{f}}(\mathbf{f}, \mathbf{u}^*) + \left( \frac{\partial \mathbf{k}}{\partial \mathbf{f}}(\mathbf{f}, \mathbf{u}^*) \right)^H \boldsymbol{\lambda}^*. \quad (\text{B.5})$$

This method is known as the *adjoint-state* method [46].

## 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,

$$h \triangleq \frac{1}{2} \|\mathbf{y} - \mathbf{H} \text{diag}(\mathbf{u}) \mathbf{f}\|^2, \quad \mathbf{k} \triangleq (\mathbf{I} - \mathbf{G} \text{diag}(\mathbf{f})) \mathbf{u} - \mathbf{v}.$$

At a given value of  $\mathbf{f}$ , the adjoint system for the Lippmann-Schwinger equation is

$$(\mathbf{I} - \mathbf{G}^H \text{diag}(\mathbf{f})) \boldsymbol{\lambda}^* = \text{diag}(\mathbf{f}) \mathbf{H}^H (\mathbf{y} - \mathbf{H} \text{diag}(\mathbf{f}) \mathbf{u}^*). \quad (\text{B.6})$$

Here,  $\boldsymbol{\lambda}^*$  is the adjoint wavefield and the  $\mathbf{u}^*$  is obtained satisfying the constraints at given value of  $\mathbf{f}$ :

$$(\mathbf{I} - \mathbf{G} \text{diag}(\mathbf{f})) \mathbf{u}^* = \mathbf{v}. \quad (\text{B.7})$$

Once the forward wavefield  $\mathbf{u}^*$  and the adjoint wavefield  $\boldsymbol{\lambda}^*$  are computed, the gradient is

$$\begin{aligned} \nabla \mathcal{F}(\mathbf{f}) &= \text{diag}(\mathbf{u}^*)^H \mathbf{H}^H (\mathbf{H} \text{diag}(\mathbf{u}^*) \mathbf{f} - \mathbf{y}) \\ &\quad - \text{diag}(\mathbf{u}^*)^H \mathbf{G}^H \boldsymbol{\lambda}^*. \end{aligned} \quad (\text{B.8})$$

Computing the gradient requires solving the forward (B.7) and the adjoint (B.6) systems only once each.

## APPENDIX C PRIMAL-DUAL METHOD

We consider a class of optimization problems

$$\min_{\mathbf{x}} h(\mathbf{x}) + g(\mathbf{L}\mathbf{x}) + k(\mathbf{x}), \quad (\text{C.1})$$

where  $h : \mathbb{R}^n \mapsto \mathbb{R}$  is a differentiable closed convex function.  $g : \mathbb{R}^m \mapsto \mathbb{R}$  and  $k : \mathbb{R}^n \mapsto \mathbb{R}$  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 is the case in TV regularization. In this section, we derive a primal-dual algorithm to find an optimal solution to problem (C.1).

### A. Preliminaries

A set-valued operator  $\mathcal{H} : \mathbb{R}^n \mapsto \mathbb{R}^n$ , that maps a point  $\mathbf{x} \in \mathbb{R}^n$  to sets  $\mathcal{H}(\mathbf{x}) \in \mathbb{R}^n$  is *monotone* if

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

The operator  $(\mathcal{I} + \gamma \mathcal{H})^{-1}$ , with  $\gamma > 0$  is called as the resolvent of the operator  $\mathcal{H}$ , where  $\mathcal{I}$  is an identity operator. The value  $\mathbf{x} = (\mathcal{I} + \gamma \mathcal{H})^{-1}(\mathbf{y})$  of the resolvent is the unique solution of the monotone inclusion  $\mathbf{y} \in \mathbf{x} + \gamma \mathcal{H}(\mathbf{x})$ . A resolvent of a monotone operator is a non-expansive operator. An operator  $\mathcal{H}$  is non-expansive if  $\|\mathcal{H}(\mathbf{x}) - \mathcal{H}(\mathbf{y})\| \leq \|\mathbf{x} - \mathbf{y}\|, \forall \mathbf{x}, \mathbf{y}$ .

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

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

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

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

A (convex) conjugate of a general function  $h : \mathbb{R}^n \mapsto \mathbb{R}$  is

$$h^*(\mathbf{y}) = \sup_{\mathbf{x} \in \mathbb{R}^n} \left\{ \mathbf{y}^T \mathbf{x} - h(\mathbf{x}) \right\}.$$

A conjugate of a function is always *convex*. The prox-operator of a function and its conjugate is related by the *Moreau identity*,

$$\operatorname{prox}_{\alpha h^*}(\mathbf{x}) + \alpha \operatorname{prox}_{h/\alpha} \left( \frac{\mathbf{x}}{\alpha} \right) = \mathbf{x}. \quad (\text{C.2})$$

### B. Fixed point method

A fixed point of an operator  $\mathcal{T} : \mathbb{R}^n \mapsto \mathbb{R}^n$  is defined as the set of points  $\mathbf{x} \in \mathbb{R}^n$  such that  $\mathcal{T}(\mathbf{x}) = \mathbf{x}$ . A fixed point method finds one such point by generating a sequence of iterates  $\mathbf{x}^{(k)}$  with  $k = 1, \dots, n$  of form

$$\mathbf{x}^{(k+1)} = \mathcal{T} \left( \mathbf{x}^{(k)} \right)$$

for a given initial point  $\mathbf{x}^{(0)}$ . The iterates converge to one of the fixed point if  $\mathcal{T}$  is a non-expansive operator.

Now recall that the resolvent of a monotone operator  $\mathcal{H}$  is a non-expansive operator, *i.e.*,  $\mathcal{T} = (\mathcal{I} + \alpha \mathcal{H})^{-1}$ . Also, it can be easily seen that the zeros of the monotone operator  $\mathcal{H}$  are the fixed points of its resolvent. Hence, the fixed point iterations takes the following form to find the zeros of a monotone operator  $\mathcal{H}$ :

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

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

$$\mathbf{x}^{(k+1)} = (\mathcal{I} + \mathcal{P}^{-1} \mathcal{H})^{-1} \mathbf{x}^{(k)},$$

with  $\mathcal{P}$  as a symmetric positive-definite linear operator. This sequence can be simplified to

$$(\mathcal{P} + \mathcal{H}) \mathbf{x}^{(k+1)} = \mathcal{P} \mathbf{x}^{(k)} \quad (\text{C.3})$$

### C. Primal-Dual algorithm

To compute monotone operator for (C.1), we look at its first-order optimality condition. It states that a zero-vector must be in the subdifferential of the cost function, *i.e.*,

$$\mathbf{0} \in \nabla h(\mathbf{x}) + \mathbf{L}^T \partial g(\mathbf{L}\mathbf{x}) + \partial k(\mathbf{x}), \quad (\text{C.4})$$

where  $\partial g : \mathbb{R}^m \mapsto \mathbb{R}^m$  and  $\partial k : \mathbb{R}^n \mapsto \mathbb{R}^n$  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  $k$ ,

$$\mathbf{u} \in \partial g(\mathbf{L}\mathbf{x}), \quad \mathbf{v} \in \partial k(\mathbf{x}). \quad (\text{C.5})$$

The equations in (C.5) can be restated as follows.

$$\mathbf{0} \in \partial g^*(\mathbf{u}) - \mathbf{L}\mathbf{x}, \quad \mathbf{0} \in \partial k^*(\mathbf{v}) - \mathbf{x} \quad (\text{C.6})$$

where  $g^*$  and  $k^*$  are the convex conjugate of the functions  $g$  and  $k$  respectively. From equations (C.4) and (C.6), we can write the optimality conditions in the form of the following system

$$\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \in \underbrace{\begin{bmatrix} \nabla h & \mathbf{L}^T & \mathbf{I}_n \\ -\mathbf{L} & \partial g^* & \mathbf{0} \\ -\mathbf{I}_n & \mathbf{0} & \partial k^* \end{bmatrix}}_{\mathcal{H}} \underbrace{\begin{bmatrix} \mathbf{x} \\ \mathbf{u} \\ \mathbf{v} \end{bmatrix}}_{\mathbf{z}}, \quad (\text{C.7})$$

where  $\mathbf{I}_n \in \mathbb{R}^{n \times n}$  is the identity matrix. It is easy to show that the operator  $\mathcal{H}$  in (C.7) is a monotone operator. 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 in (C.3) results in the following primal-dual algorithm:

$$\begin{aligned} \mathbf{x}^{(t+1)} &= (\mathcal{I} + \gamma \nabla h)^{-1} \left( \mathbf{x}^{(t)} - \gamma \mathbf{L}^T \mathbf{u}^{(t)} - \gamma \mathbf{v}^{(t)} \right) \\ \mathbf{u}^{(t+1)} &= (\mathcal{I} + \gamma \partial g^*)^{-1} \left( \mathbf{u}^{(t)} - \gamma \mathbf{L} \left( \mathbf{x}^{(t)} - 2\mathbf{x}^{(t+1)} \right) \right) \\ \mathbf{v}^{(t+1)} &= (\mathcal{I} + \gamma \partial k^*)^{-1} \left( \mathbf{v}^{(t)} - \gamma \left( \mathbf{x}^{(t)} - 2\mathbf{x}^{(t+1)} \right) \right) \end{aligned}$$

If the proximal operators of functions  $h, g$  and  $k$  are simple, then the each iteration can be computed efficiently.

## APPENDIX D PROXIMAL QUASI-NEWTON METHOD

In this section, we discuss the Quasi-Newton (QN) method and its proximal version (prox-QN). Assuming the cost function  $f$  is twice differentiable, QN aims to solve

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} f(\mathbf{x}) \quad (\text{D.1})$$

by generating a sequence based on the quadratic approximation to the function  $f$  at every iterate of the sequence. The procedure is as follows:

$$\begin{aligned} \mathbf{s}^{(k)} &= -\mathbf{H}_k^{-1} \nabla f \left( \mathbf{x}^{(k)} \right), \\ \alpha_k &= \text{linesearch} \left( f \left( \mathbf{x}^{(k)} + \alpha \mathbf{s}^{(k)} \right) \right), \\ \mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} + \alpha_k \mathbf{s}^{(k)}. \end{aligned} \quad (\text{D.2})$$

Here,  $\mathbf{H}_k$  is (an 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 computes 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 to solve problems of form

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ f(\mathbf{x}) \text{ subject to } g(\mathbf{L}\mathbf{x}) \leq \tau, \mathbf{x} \geq \mathbf{0} \right\}. \quad (\text{D.3})$$

Here,  $f : \mathbb{R}^n \mapsto \mathbb{R}$  is a twice-differentiable function, and  $g : \mathbb{R}^n \mapsto \mathbb{R}$  is a convex but potentially non-differentiable

function. For convenience, we rewrite the problem (D.3) as

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ f(\mathbf{x}) + \delta_g(\mathbf{L}\mathbf{x}) + \delta_k(\mathbf{x}) \right\}, \quad (\text{D.4})$$

where,  $\delta_g$  is an indicator to the set  $\{\mathbf{x} : g(\mathbf{x}) \leq \tau\}$ , and  $\delta_k$  is an indicator to the set  $\{\mathbf{x} : \mathbf{x} > \mathbf{0}\}$ . We propose a following modification to the Quasi-Newton method, and call it Proximal Quasi-Newton (prox-QN) method:

$$\begin{aligned} (a) \quad & \mathbf{s}^{(k)} = \underset{\mathbf{s}}{\operatorname{argmin}} \left\{ \mathbf{s}^T \nabla f(\mathbf{x}^{(k)}) + \frac{1}{2} \mathbf{s}^T \mathbf{H}_k \mathbf{s} \right. \\ & \left. + \delta_g(\mathbf{L}(\mathbf{x}^{(k)} + \mathbf{s})) + \delta_k(\mathbf{x}^{(k)} + \mathbf{s}) \right\} \\ (b) \quad & \text{define } \hat{\mathbf{x}}(\alpha) = \operatorname{prox}_{\alpha(\delta_g + \delta_k)}(\mathbf{x}^k + \alpha \mathbf{s}^{(k)}) \\ (c) \quad & \alpha_k = \underset{\alpha}{\operatorname{argmin}} \{f(\hat{\mathbf{x}}(\alpha))\} \\ (d) \quad & \mathbf{x}^{(k+1)} = \hat{\mathbf{x}}(\alpha_k) \end{aligned} \quad (\text{D.5})$$

The steps in (D.5) can be summarized as follows: Step (a) finds a search direction  $\mathbf{s}^k$ . It minimizes the quadratic approximation of  $f$  at  $\mathbf{x}^{(k)}$ , ensuring that it satisfies the constraints. In step (b), we define a function  $\hat{\mathbf{x}} : \mathbb{R} \mapsto \mathbb{R}^n$  which is a proximal of the iterate  $\mathbf{x}^k + \alpha \mathbf{s}^{(k)}$  with respect to indicators to functions  $g$  and  $k$ . The function  $\hat{\mathbf{x}}$  ensures that the step length,  $\alpha$ , must satisfy the constraints. Step (c) does a linesearch with respect to the feasible  $\alpha$ . Once we obtain the correct  $\alpha$ , we update our variable of interest  $\mathbf{x}$  in step (d).

The minimization problem in step (a) of (D.5), is a convex minimization problem. The cost function is the sum of three functions  $h$ ,  $\delta_g$ ,  $\delta_k$ . The function

$$h(\mathbf{s}) = \mathbf{s}^T \nabla f(\mathbf{x}^{(k)}) + \frac{1}{2} \mathbf{s}^T \mathbf{H}_k \mathbf{s}$$

is a convex quadratic function, while the remaining two,  $\delta_g$  and  $\delta_k$ , are non-differentiable convex functions. To solve this minimization problem, we use first-order primal-dual method described in Appendix C. The iterates for  $t = 0, \dots, T$  are

$$\begin{aligned} \mathbf{s}^{(t+1)} &= \operatorname{prox}_{\gamma h}(\mathbf{s}^{(t)} - \gamma \mathbf{L}^T \mathbf{u}^{(t)} - \gamma \mathbf{v}^{(t)}), \\ \mathbf{u}^{(t+1)} &= \operatorname{prox}_{\gamma \delta_g^*}(\mathbf{u}^{(t)} + \gamma \mathbf{L}(2\mathbf{s}^{(t+1)} - \mathbf{s}^{(t)})), \\ \mathbf{v}^{(t+1)} &= \operatorname{prox}_{\gamma \delta_k^*}(\mathbf{v}^{(t)} + \gamma(2\mathbf{s}^{(t+1)} - \mathbf{s}^{(t)})), \end{aligned} \quad (\text{D.6})$$

with  $\gamma > 0$  controlling the speed of convergence. The proximal operations for  $h$ ,  $\delta_g$  and  $\delta_k$  are expressed as follows:

$$\begin{aligned} \operatorname{prox}_{\gamma h}(\mathbf{y}) &= (\mathbf{I} + \gamma \mathbf{H}_k)^{-1} (\mathbf{y} - \gamma \nabla f(\mathbf{x}^{(k)})) \\ \operatorname{prox}_{\gamma \delta_g}(\mathbf{y}) &= \operatorname{proj}_{\|\cdot\|_1 \leq \tau}(\mathbf{y}) \\ \operatorname{prox}_{\gamma \delta_k}(\mathbf{y}) &= \begin{cases} \mathbf{y} & \mathbf{y} > \mathbf{0} \\ \mathbf{0} & \mathbf{y} \leq \mathbf{0} \end{cases} \end{aligned}$$

The proposed method (prox-QN) differs from [47] in two aspects: (i) The function  $g$  can be potentially be more than  $\ell_1$  type penalty. For example, we can work with total-variation-type regularization. (ii) The linesearch ensures that the chosen  $\alpha$  is strictly feasible.

## REFERENCES

- [1] 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)*, May 2019, pp. 7735–7739.
- [2] S. Laurens, J. Balayssac, J. Rhazi, G. Klysz, and G. Arliguie, "Non-destructive evaluation of concrete moisture by gpr: experimental study and direct modeling," *Materials and structures*, vol. 38, no. 9, pp. 827–832, 2005.
- [3] S. R. Arridge, "Optical tomography in medical imaging," *Inverse problems*, vol. 15, no. 2, p. R41, 1999.
- [4] L. Sirgue, O. Barkved, J. Dellinger, J. Etgen, U. Albertin, and J. Kommedal, "Thematic set: Full waveform inversion: The next leap forward in imaging at valhall," *First Break*, vol. 28, no. 4, pp. 65–70, 2010.
- [5] J. Virieux and S. Operto, "An overview of full-waveform inversion in exploration geophysics," *Geophysics*, vol. 74, no. 6, pp. WCC1–WCC26, 2009.
- [6] A. J. Witten, J. E. Molyneux, and J. E. Nyquist, "Ground penetrating radar tomography: Algorithms and case studies," *IEEE Transactions on geoscience and remote sensing*, vol. 32, no. 2, pp. 461–467, 1994.
- [7] Z. Yuan and H. Jiang, "Three-dimensional finite-element-based photoacoustic tomography: Reconstruction algorithm and simulations," *Medical physics*, vol. 34, no. 2, pp. 538–546, 2007.
- [8] M. Haynes and M. Moghaddam, "Large-domain, low-contrast acoustic inverse scattering for ultrasound breast imaging," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 11, pp. 2712–2722, 2010.
- [9] M. Humphry, B. Kraus, A. Hurst, A. Maiden, and J. Rodenburg, "Ptychographic electron microscopy using high-angle dark-field scattering for sub-nanometre resolution imaging," *Nature communications*, vol. 3, p. 730, 2012.
- [10] A. H. Zewail, "Four-dimensional electron microscopy," *Science*, vol. 328, no. 5975, pp. 187–193, 2010.
- [11] J. J. Barton, "Photoelectron holography," *Physical review letters*, vol. 61, no. 12, p. 1356, 1988.
- [12] G. Spencer and M. Murty, "General ray-tracing procedure," *JOSA*, vol. 52, no. 6, pp. 672–678, 1962.
- [13] M. Born and E. Wolf, *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Elsevier, 2013.
- [14] A. Devaney, "Inverse-scattering theory within the rytov approximation," *Optics letters*, vol. 6, no. 8, pp. 374–376, 1981.
- [15] E. Baysal, D. D. Kosloff, and J. W. Sherwood, "Reverse time migration," *Geophysics*, vol. 48, no. 11, pp. 1514–1524, 1983.
- [16] D. Colton and R. Kress, *Inverse acoustic and electromagnetic scattering theory*. Springer Science & Business Media, 2012, vol. 93.
- [17] G. Bao and P. Li, "Inverse medium scattering problems for electromagnetic waves," *SIAM Journal on Applied Mathematics*, vol. 65, no. 6, pp. 2049–2066, 2005.
- [18] G. Bao, P. Li, J. Lin, and F. Triki, "Inverse scattering problems with multi-frequencies," *Inverse Problems*, vol. 31, no. 9, p. 093001, 2015.
- [19] C. Borges, A. Gillman, and L. Greengard, "High resolution inverse scattering in two dimensions using recursive linearization," *SIAM Journal on Imaging Sciences*, vol. 10, no. 2, pp. 641–664, 2017.
- [20] C. Estatico, A. Fedeli, M. Pastorino, and A. Randazzo, "A Multifrequency Inexact-Newton Method in Lp Banach Spaces for Buried Objects Detection," *IEEE Transactions on Antennas and Propagation*, vol. 63, no. 9, pp. 4198–4204, 2015.
- [21] A. Tarantola, "Inversion of seismic reflection data in the acoustic approximation," *Geophysics*, vol. 49, no. 8, pp. 1259–1266, 1984.
- [22] R. G. Pratt, "Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model," *Geophysics*, vol. 64, no. 3, pp. 888–901, 1999.
- [23] T. Isernia, F. Soldovieri, G. Leone, and R. Pierri, "On the local minima in phase reconstruction algorithms," *Radio Science*, vol. 31, no. 6, pp. 1887–1899, 1996. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/96RS02154>
- [24] T. Isernia, V. Pascazio, and R. Pierri, "On the local minima in a tomographic imaging technique," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1596–1607, 2001.
- [25] W. Mulder and R.-E. Plessix, "Exploring some issues in acoustic full waveform inversion," *Geophysical Prospecting*, vol. 56, no. 6, pp. 827–841, 2008.
- [26] W. W. Symes, "Migration velocity analysis and waveform inversion," *Geophysical prospecting*, vol. 56, no. 6, pp. 765–790, 2008.
- [27] M. Warner and L. Guasch, "Adaptive waveform inversion: Theory," *Geophysics*, vol. 81, no. 6, pp. R429–R445, 2016.

- [28] T. Van Leeuwen and F. J. Herrmann, "Mitigating local minima in full-waveform inversion by expanding the search space," *Geophysical Journal International*, vol. 195, no. 1, pp. 661–667, 2013.
- [29] B. Biondi and A. Almomin, "Tomographic full waveform inversion (tfwi) by combining full waveform inversion with wave-equation migration velocity analysis," in *SEG Technical Program Expanded Abstracts 2012*. Society of Exploration Geophysicists, 2012, pp. 1–5.
- [30] T. van Leeuwen and F. J. Herrmann, "A penalty method for pde-constrained optimization in inverse problems," *Inverse Problems*, vol. 32, no. 1, p. 015007, 2015.
- [31] P. Bharadwaj, W. Mulder, and G. Drijkoningen, "Full waveform inversion with an auxiliary bump functional," *Geophysical Journal International*, vol. 206, no. 2, pp. 1076–1092, 2016.
- [32] E. Esser, L. Guasch, F. J. Herrmann, and M. Warner, "Constrained waveform inversion for automatic salt flooding," *The Leading Edge*, vol. 35, no. 3, pp. 235–239, 2016.
- [33] E. Esser, L. Guasch, T. van Leeuwen, A. Y. Aravkin, and F. J. Herrmann, "Total variation regularization strategies in full-waveform inversion," *SIAM Journal on Imaging Sciences*, vol. 11, no. 1, pp. 376–406, 2018.
- [34] A. Asnaashari, R. Brossier, S. Garambois, F. Audebert, P. Thore, and J. Virieux, "Regularized seismic full waveform inversion with prior model information," *Geophysics*, vol. 78, no. 2, pp. R25–R36, 2013.
- [35] M. Ambrosanio and V. Pascazio, "A compressive-sensing-based approach for the detection and characterization of buried objects," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 7, pp. 3386–3395, 2015.
- [36] C. Bunks, F. M. Saleck, S. Zaleski, and G. Chavent, "Multiscale seismic waveform inversion," *Geophysics*, vol. 60, no. 5, pp. 1457–1473, 1995. [Online]. Available: <https://doi.org/10.1190/1.1443880>
- [37] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.
- [38] M. Unser, J. Fageot, and J. P. Ward, "Splines are universal solutions of linear inverse problems with generalized tv regularization," *SIAM Review*, vol. 59, no. 4, pp. 769–793, 2017.
- [39] L. Condat, "Fast projection onto the simplex and the  $\mathbb{L}_1$  ball," *Mathematical Programming*, vol. 158, no. 1-2, pp. 575–585, 2016.
- [40] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [41] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of mathematical imaging and vision*, vol. 40, no. 1, pp. 120–145, 2011.
- [42] E. van den Berg and M. P. Friedlander, "Sparse optimization with least-squares constraints," *SIAM J. Optimization*, vol. 21, no. 4, pp. 1201–1229, 2011.
- [43] Y. Ma, H. Mansour, D. Liu, P. T. Boufounos, and U. S. Kamilov, "Accelerated image reconstruction for nonlinear diffractive imaging," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 6473–6477.
- [44] Y. Chen, "Inverse scattering via heisenberg's uncertainty principle," *Inverse problems*, vol. 13, no. 2, p. 253, 1997.
- [45] J. Kaipio and E. Somersalo, "Statistical inverse problems: discretization, model reduction and inverse crimes," *Journal of computational and applied mathematics*, vol. 198, no. 2, pp. 493–504, 2007.
- [46] R.-E. Plessix, "A review of the adjoint-state method for computing the gradient of a functional with geophysical applications," *Geophysical Journal International*, vol. 167, no. 2, pp. 495–503, 2006.
- [47] J. D. Lee, Y. Sun, and M. A. Saunders, "Proximal newton-type methods for minimizing composite functions," *SIAM Journal on Optimization*, vol. 24, no. 3, pp. 1420–1443, 2014.