SEAGLE: Sparsity-Driven Image Reconstruction under Multiple Scattering

Hsiou-Yuan Liu, *Student Member, IEEE*, Dehong Liu, *Senior Member, IEEE*, Hassan Mansour, *Member, IEEE*, Petros T. Boufounos, *Senior Member, IEEE*, Laura Waller, and Ulugbek S. Kamilov, *Member, IEEE*

***Abstract*—Multiple scattering of an electromagnetic wave as it passes through an object is a fundamental problem that limits the performance of current imaging systems. In this paper, we describe a new technique—called Series Expansion with Accelerated Gradient Descent on Lippmann-Schwinger Equation (SEAGLE)—for robust imaging under multiple scattering based on a combination of an iterative forward model and a total variation (TV) regularizer. The proposed method can account for multiple scattering, which makes it advantageous in applications where single scattering approximations are inaccurate. Specifically, the method relies on a series expansion of the scattered wave with an accelerated-gradient method. This expansion guarantees the convergence of the forward model even for strongly scattering objects. One of our key insights is that it is possible to obtain an explicit formula for computing the gradient of an iterative forward model with respect to the unknown object, thus enabling fast image reconstruction with the state-of-the-art fast iterative shrinkage/thresholding algorithm (FISTA). The proposed method is validated on diffraction tomography where complex electric field is captured at different illumination angles.**

***Index Terms*—Diffraction tomography, nonconvex optimization, sparse optimization, total variation, computational imaging.**

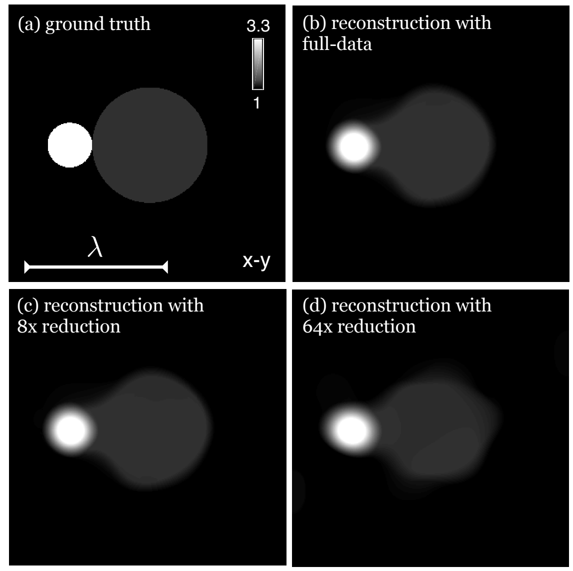

Fig. 1: SEAGLE can be used to reconstruct the spatial distribution of dielectric permittivity from measurements of complex scattered waves at different illumination angles. Illustration using experimental data at 3 GHz: (a) ground truth; (b) using full data; (c) $8\times$ data reduction; (d) $64\times$ data reduction.

I. Introduction

Reconstruction of the spatial permittivity distribution of an unknown object from the measurements of the scattered waves at different illumination angles is common in numerous applications. Traditional formulations of the problem are based on linearizing the relationship between the permittivity and the measured wave. For example, if one assumes a straight-ray propagation of waves, the phase of the transmitted wave can be interpreted as a line integral of the permittivity along the propagation direction. This approximation leads to an efficient reconstruction with the filtered back-projection algorithm [1]. Diffraction tomography uses a more refined linear scattering model based on the first Born or Rytov approximations [2]–[4]. It establishes a Fourier transform-based relationship between the measured wave and the permittivity, and thus enables the reconstruction of the latter with a direct numerical application of the inverse Fourier transform.

Recent research in compressive sensing and sparse signal processing has established that sparse regularization can dramatically improve the quality of reconstructed images, even when the amount of measured data is severely limited [5], [6]. This has popularized optimization-based inverse scattering approaches that combine linear forward models with regularizers that mitigate ill-posedness by promoting solutions that are sparse in a suitable transform domain. One class of such regularizers is total variation (TV) [7], which substantially reduces undesired artifacts due to missing data [8]–[10].

The main advantage of imaging with linear forward models is that the reconstruction can be reduced to a convex optimization problem that is relatively simple and efficient to solve [11]–[14]. However, multiple scattering of waves limit the validity of Born and Rytov approximations when the objects are relatively large or have high permittivity contrasts compared to the background [15]. Multiple scattering is a fundamental problem in diffraction tomography and its complete resolution would enable imaging through strongly scattering objects such as human tissue [16]. As multiple scattering leads to nonlinear forward models, the challenge is in finding computationally tractable methods that can account for the nonlinearity while also making the image reconstruction tractable. To that end, we propose a new method that efficiently

H.-Y. Liu and L. Waller are with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720, USA. This work was completed while H.-Y. Liu was with Mitsubishi Electric Research Laboratories (MERL).

D. Liu, H. Mansour, P. T. Boufounos, and U. S. Kamilov (email: kamilov@merl.com) are with Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139, USA.

combines a nonlinear forward model with the TV regularizer, thus enabling high-quality imaging from a limited number of measurements. Figure 1 provides an example, illustrating the quality of reconstruction for a strongly scattering object from a public dataset [17]. In particular, the 320×320 image in Figure 1(d) was obtained from only 16 experimentally collected measurements at different illumination angles.

A. Contributions

Our work builds upon prior work on inverse scattering that has been applied to a variety of practical problems in optical, microwave, and radar imaging. The proposed method—called Series Expansion with Accelerated Gradient Descent on Lippmann-Schwinger Equation (SEAGLE)—further extends this work by considering an iterative forward model that still enables efficient sparsity-driven inversion. The performance of SEAGLE is robust to large permittivity contrasts, data reduction, and measurement noise.

The key contributions of this paper is a novel image reconstruction strategy based on the explicit evaluation of the gradient of an iterative forward model with respect to the unknown parameters that correspond to the permittivity of the object. Specifically, we rely on the Nesterov's accelerated-gradient method (AGM) [18] to iteratively approximate the scattered waves. The key benefit is the guaranteed convergence of the latter even for objects with large permittivity contrasts, even for general nonlinear functions of the scattered wave. Albeit, the solution to AGM may not be unique if the functions are not strongly convex. We additionally present extensive validation of our approach on analytical, simulated, and experimental data. The experimental data used in our evaluations comes from a public dataset [17], which enables easy comparisons with several other related approaches.

B. Related Work

Imaging systems, such as optical projection tomography (OPT), diffraction tomography, optical coherence tomography (OCT), digital holography, and subsurface radar, rely on the linearization of object-wave interaction [8]–[10], [19]–[31]. Early work in microwave imaging has shown the promise of accounting for the nonlinear nature of scattering [32]–[35]. These have been extended by a large number of techniques incorporating the nonlinear nature of scattering. Several recent publications have reviewed these methods [36]–[38], which include conjugate gradient method (CGM) [39], [40], contrast source inversion method (CSIM) [41], hybrid method (HM) [37], and recursive Born method (RBM) [42]. Some recent work has explored the idea of statistical modeling of multiple scattering for imaging through diffusive or turbid media [43]–[45]. Other work has explored the combination of nonlinear scattering with sparse regularization [45]–[48]. This paper extends our preliminary work [49] by including key mathematical derivations, as well as more extensive validation on experimentally collected data.

Recently, the beam propagation method (BPM) was proposed for performing nonlinear inverse scattering in transmission [48], [50]–[53]. BPM-based methods numerically propagate the field slice-by-slice through the object. The Jacobian matrix of BPM can be efficiently computed with error backpropagation algorithm, which enables fast image reconstruction. The nonlinear model presented here is based on the Lippmann-Schwinger equation [54]. The main advantage of the proposed formulation is that it accounts for both transmitted and reflected waves. This makes the proposed method more suitable when reflections are important.

The Lippmann-Schwinger equation, also known as the Foldy-Lax multiple scattering model, has been extensively used in the inverse scattering literature for imaging under wave scattering [55], [56], diffuse optical tomography [57], [58], impedance tomography [59], and for elastic wave scattering problems [60]. The common theme involves first estimating the contrast source reflectivity of the target by exploiting the joint sparsity across multiple illuminations. Then, the Lippmann-Schwinger equation is used to estimate the total field, which in turn is used to separate the target permittivity from the estimated contrast source reflectivity. Our proposed method differs from these works in that we jointly estimate the total field as well as the target permittivity in a closed loop framework that allows us to exploit the spatial structure of the target through regularization.

The problem setting in this paper is also related to that in full-waveform inversion (FWI) [61]–[63] used in geophysical applications. We propose an alternative approach for solving such problems, namely modeling the forward scattering process using AGM and estimating the material parameters. FWI-based methods often use a differential form of Helmholtz equation, while we rely on integral-domain formulation given by the Lippmann-Schwinger equation. Moreover, established frameworks in FWI-based methods utilize Krylov based solvers along with the adjoint state method [64] to estimate the gradient of the forward wave propagation model with respect to the permittivity. However, these methods rely on the linearity of the wave equation as a function the scattered field. Otherwise, for nonlinear functions of the scattered field, these methods require iterative linearization which can become slow and lacks convergence guarantees. Our AGM-based forward model on the other hand is guaranteed to converge for general nonlinear convex functions of the scattered field, and has the same asymptotic convergence rate as Krylov-based methods for strictly convex functions. Moreover, a key difference of our method is in the combination of the AGM-based forward model and sparsity-driven image reconstruction using the fast iterative shrinkage/thresholding algorithm (FISTA) [65]. Our experiments show that our formulation is promising, as it enables fast, stable, and reliable convergence when working with a limited amount of data.

The experimental data used in this paper comes from a public dataset of complex wave-field measurements of several objects at various illumination angles and frequencies [17]. Several other methods have been tested on this dataset [41], [66]–[69]. This enables qualitative evaluation of the performance of the proposed technique against other algorithms. While many of the methods tested on the data use multiple frequencies, the results here rely on a single frequency. However, the method uses the latest techniques in large-scale

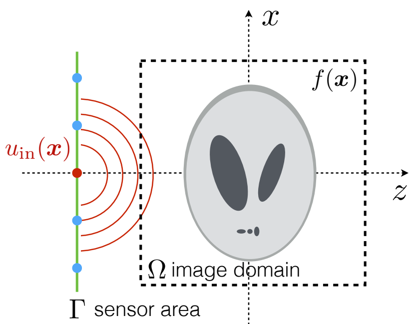

Fig. 2: Schematic representation of the scattering experiment. An object with a real scattering potential $f(\boldsymbol{x}), \boldsymbol{x} \in \Omega$, is illuminated with an input wave u_{in}, which interacts with the object and leads to the scattered wave u_{sc} measured at the sensing region Γ represented with a green line.

optimization with sparse regularization, which enables sub-sampling and leads to improvements in imaging performance. We expect the performance of the proposed method to improve further if multi-frequency measurements are incorporated.

II. Forward model

The forward problem computes the scattered field given a distribution of inhomogeneous permittivity, while the inverse problem reconstructs this distribution. The model we propose here can be interpreted as a series expansion based on the iterates of the gradient method. This expansion can be made arbitrarily accurate and is stable for high permittivity objects. Additionally, it enables efficient computation of the gradient of the cost function, which is essential for fast image reconstruction. In this section we focus on introducing the forward model, leaving the gradient evaluation and inversion to the next section. Our derivations are for the scenario of a single illumination, but the generalization to an arbitrary number of illuminations is straightforward.

A. Problem formulation

Consider the scattering problem in Figure 2, where an object of the permittivity distribution $\epsilon(\boldsymbol{x})$ in the bounded domain $\Omega \subseteq \mathbb{R}^D$, with $D \in \{2,3\}$, is immersed into a background medium of permittivity ϵ_b, and illuminated with the incident electric field $u_{\text{in}}(\boldsymbol{x})$. We assume that the incident field is monochromatic and coherent, and it is known inside Ω and at the locations of the sensors Γ. The result of object-wave interaction is measured at the location of the sensors as a scattered field $u_{\text{sc}}(\boldsymbol{x})$. The scattering of light can be accurately described by the Lippmann-Schwinger equation inside the image domain [54]

$$u(\boldsymbol{x}) = u_{\text{in}}(\boldsymbol{x}) + \int_\Omega g(\boldsymbol{x}-\boldsymbol{x}')\, f(\boldsymbol{x}')\, u(\boldsymbol{x}')\, \mathrm{d}\boldsymbol{x}', \quad (\boldsymbol{x} \in \Omega) \quad (1)$$

where $u(\boldsymbol{x}) = u_{\text{in}}(\boldsymbol{x}) + u_{\text{sc}}(\boldsymbol{x})$ is the total electric field, $f(\boldsymbol{x}) \triangleq k^2(\epsilon(\boldsymbol{x}) - \epsilon_b)$ is the scattering potential, which is assumed to be real, and $k = 2\pi/\lambda$ is the wavenumber in vacuum. The function $g(\boldsymbol{x})$ is the Green's function defined as

$$g(\boldsymbol{x}) \triangleq \begin{cases} \dfrac{\mathrm{j}}{4} H_0^{(1)}(k_b \|\boldsymbol{x}\|_{\ell_2}) & \text{in 2D} \\ \dfrac{\mathrm{e}^{\mathrm{j}k_b\|\boldsymbol{x}\|_{\ell_2}}}{4\pi\|\boldsymbol{x}\|_{\ell_2}} & \text{in 3D,} \end{cases} \quad (2)$$

where $k_b \triangleq k\sqrt{\epsilon_b}$ is the wavenumber of the background medium and $H_0^{(1)}$ is the zero-order Hankel function of the first kind. Note that the Green's function satisfies the Helmholtz equation

$$\left(\nabla^2 + k_b^2 \mathrm{I}\right) g(\boldsymbol{x}) = -\delta(\boldsymbol{x}),$$

as well as the outgoing-wave boundary condition, and a time-dependence convention under which the physical electric field equals to $\mathrm{Re}\left\{u(\boldsymbol{x})\mathrm{e}^{-\mathrm{j}\omega t}\right\}$. The knowledge of the total-field u inside the image domain Ω enables the prediction of the scattered field at the sensor area

$$u_{\text{sc}}(\boldsymbol{x}) = \int_\Omega g(\boldsymbol{x}-\boldsymbol{x}')\, f(\boldsymbol{x}')\, u(\boldsymbol{x}')\, \mathrm{d}\boldsymbol{x}'. \quad (\boldsymbol{x} \in \Gamma) \quad (3)$$

The computation of the scattered wave is equivalent to solving (1) for $u(\boldsymbol{x})$ inside the image and evaluating (3) for $u_{\text{sc}}(\boldsymbol{x})$ at the sensor locations. Note that u is present on both sides of (1) and that the relation between u and the scattering potential f is nonlinear. The first Born and the Rytov approximations [2]–[4] are linear models that replace u in (3) with a suitable approximation that decouples the nonlinear dependence of u on f. However, such linearization imposes a strong assumption that the object is weakly scattering, which makes the corresponding reconstruction methods not applicable to a large variety of imaging problems [15]. Our forward model described next is a fast nonlinear method for solving (1) that overcomes the weakly scattering object assumptions.

The results presented in this paper rely on the scalar theory of diffraction, which yields accurate results when two conditions are met: (*i*) objects are sufficiently large compared with the wavelength; (*ii*) the measurements are taken sufficiently far from the object. Experimental demonstration of this was provided by Silver [70] and the applicability of scalar theory to instrumentation was extensively discussion in the classical literature in electromagnetics and optics [71], [72]. Note also that both of these conditions are generally met in optical imaging, which is an important 3D application area for the method in this paper.

B. Algorithmic Expansion of the Scattered Waves

We separate the computation of the electric field into two parts: the total field $u(\boldsymbol{x})$ in the image domain and the scattered field $u_{\text{sc}}(\boldsymbol{x})$ at the sensors. The discretization and combination of (1) and (3) leads to the following matrix-vector description of the forward problem

$$\mathbf{y} = \mathbf{H}(\mathbf{u} \bullet \mathbf{f}) + \mathbf{e} \quad (4a)$$
$$\mathbf{u} = \mathbf{u}_{\text{in}} + \mathbf{G}(\mathbf{u} \bullet \mathbf{f}), \quad (4b)$$

where $\mathbf{f} \in \mathbb{R}^N$ is the discretized scattering potential f, $\mathbf{y} \in \mathbb{C}^M$ is the measured scattered field u_{sc} at Γ, $\mathbf{u}_{\text{in}} \in \mathbb{C}^N$ is the

Algorithm 1 Forward model computation

intput: Image $\mathbf{f} \in \mathbb{R}^N$, maximum number of iterations K, tolerance δ_{tol}, and initialization $\mathbf{u}_{\text{init}} = \mathbf{u}_{\text{in}}$.
set: $\mathbf{u}^{-1} \leftarrow \mathbf{u}_{\text{init}},\ \mathbf{u}^0 \leftarrow \mathbf{u}_{\text{init}},\ t_0 \leftarrow 0$
for $k \leftarrow 1$ to K **do**
  $t_k \leftarrow (1 + \sqrt{1 + 4t_{k-1}^2})/2,$
  $\mu_k \leftarrow (1 - t_{k-1})/t_k$
  $\mathbf{s}^k \leftarrow (1 - \mu_k)\mathbf{u}^{k-1} + \mu_k \mathbf{u}^{k-2}$
  $\mathbf{g} \leftarrow \mathbf{A}^{\mathsf{H}}(\mathbf{A}\mathbf{s}^k - \mathbf{u}_{\text{in}})$ ▷ gradient at $\mathbf{s}^k$
  if $\|\mathbf{g}\|_2 < \delta_{\text{tol}}$ **then** $K \leftarrow k$, break the loop
  $\gamma_k \leftarrow \|\mathbf{g}\|_2^2 / \|\mathbf{A}\mathbf{g}\|_2^2$
  $\mathbf{u}^k \leftarrow \mathbf{s}^k - \gamma_k \mathbf{g}$
$\widehat{\mathbf{u}} \leftarrow \mathbf{u}^K$
$\mathbf{z} \leftarrow \mathbf{H}(\widehat{\mathbf{u}} \bullet \mathbf{f})$
return: predicted scattered wave $\mathbf{z}$, as well as $\widehat{\mathbf{u}}$, $\{\mathbf{s}^k\}$, $\{\gamma_k\}$, and $\{\mu_k\}$.

input field u_{in} inside Ω, $\mathbf{H} \in \mathbb{C}^{M \times N}$ is the discretization of the Green's function at Γ, $\mathbf{G} \in \mathbb{C}^{N \times N}$ is the discretization of the Green's function inside Ω, the symbol $\bullet$ denotes a component-wise multiplication between two vectors, and $\mathbf{e} \in \mathbb{C}^M$ models the random noise at the measurements. Using the shorthand notation $\mathbf{A} \triangleq \mathbf{I} - \mathbf{G}\mathrm{diag}(\mathbf{f})$, where $\mathbf{I} \in \mathbb{R}^{N \times N}$ is the identity matrix and $\mathrm{diag}(\cdot)$ is an operator that forms a diagonal matrix from its argument, we can represent the forward scattering in (4b) as a minimization problem

$$\widehat{\mathbf{u}}(\mathbf{f}) \triangleq \underset{\mathbf{u} \in \mathbb{C}^N}{\arg\min} \{\mathcal{S}(\mathbf{u})\} \quad \text{with} \quad \mathcal{S}(\mathbf{u}) \triangleq \frac{1}{2}\|\mathbf{A}\mathbf{u} - \mathbf{u}_{\text{in}}\|_{\ell_2}^2, \tag{5}$$

where the matrix $\mathbf{A}$ is a function of $\mathbf{f}$. The gradient of $\mathcal{S}$ can be computed as

$$\nabla \mathcal{S}(\mathbf{u}) = \mathbf{A}^{\mathsf{H}}(\mathbf{A}\mathbf{u} - \mathbf{u}_{\text{in}}). \tag{6}$$

Since (5) corresponds to the optimization of a differentiable function, it is possible to compute the total field $\widehat{\mathbf{u}}$ iteratively using Nesterov's AGM [18]

$$t_k \leftarrow \frac{1}{2}\left(1 + \sqrt{1 + t_{k-1}^2}\right) \tag{7a}$$
$$\mathbf{s}^k \leftarrow \mathbf{u}^{k-1} + ((t_{k-1} - 1)/t_k)(\mathbf{u}^{k-1} - \mathbf{u}^{k-2}) \tag{7b}$$
$$\mathbf{u}^k \leftarrow \mathbf{s}^k - \nu \mathbf{A}^{\mathsf{H}}(\mathbf{A}\mathbf{s}^k - \mathbf{u}_{\text{in}}), \tag{7c}$$

for $k = 1, 2, \dots, K$, where $\mathbf{u}^0 = \mathbf{u}^{-1} = \mathbf{u}_{\text{in}}$, $q_0 = 1$, and $\nu > 0$ is the step-size. At any moment, the predicted scattered field can be set to $\mathbf{z}^k = \mathbf{H}(\mathbf{u}^k \bullet \mathbf{f})$ with $\mathbf{u}^k$ given by (7c). Note that the resulting set of fields $\{\mathbf{u}^k\}_{k \in [1 \dots K]}$ and $\{\mathbf{z}^k\}_{k \in [1 \dots K]}$ can be interpreted as a K-term series expansion of the wave-fields inside the object and at the sensor locations, respectively. The full procedure for forward computation with a convenient adaptive step-size is summarized in Algorithm 1.

There are strong parallels between the AGM-based formulation of scattering and the popular Born series expansion [42], [54]. Both approaches produce a sequence of wave-field vectors $\{\mathbf{u}_k\}_{k \in [1, \dots, K]}$, starting from the initial $\mathbf{u} = \mathbf{u}_{\text{in}}$. The final field $\widehat{\mathbf{u}}$ is the linear combination of all the intermediate field vectors, which indicates that this is an expansion of the field $\widehat{\mathbf{u}}$ with respect to K-terms, where each term brings additional information about scattering. The traditional Born series and SEAGLE are thus identical for $K = 0$, but yield different intermediate fields for any other K. Crucially, while Born series is known to diverge for strong scatterers, AGM is guaranteed to converge for sufficiently large K, as (5) is a smooth and convex optimization problem [18], [65], [73]

Algorithm 2 Image formation with FISTA

input: scattered field $\mathbf{y}$, initial guess $\mathbf{f}^0$, initial step $\gamma_0 > 0$, step reduction rate $0 < \eta < 1$, and regularization parameter $\tau > 0$.
set: $t \leftarrow 1$, $\tilde{\mathbf{f}}^0 \leftarrow \mathbf{f}^0$, $q_0 \leftarrow 1$
repeat
  $\gamma_t \leftarrow \gamma_{t-1}/\eta$
  repeat ▷ line search
    $\gamma_t \leftarrow \eta\gamma_t$
    $\mathbf{f}^t \leftarrow \mathrm{prox}_{\gamma_t \tau \mathcal{R}}(\tilde{\mathbf{f}}^{t-1} - \gamma_t \nabla \mathcal{D}(\tilde{\mathbf{f}}^{t-1}))$
  until $\mathcal{D}(\mathbf{f}^t) + \tau\mathcal{R}(\mathbf{f}^t) \leq Q_{\gamma_t}(\mathbf{f}^t, \tilde{\mathbf{f}}^{t-1})$
  $q_t \leftarrow \frac{1}{2}\left(1 + \sqrt{1 + 4q_{t-1}^2}\right)$
  $\tilde{\mathbf{f}}^t \leftarrow \mathbf{f}^t + ((q_{t-1} - 1)/q_t)(\mathbf{f}^t - \mathbf{f}^{t-1})$
  $t \leftarrow t + 1$
until stopping criterion
return: estimate of the scattering potential $\mathbf{f}^t$.

Algorithm 3 Error backpropagation for $\nabla\mathcal{D}(\mathbf{f})$

intput: Image $\mathbf{f} \in \mathbb{R}^N$, measurements $\mathbf{y} \in \mathbb{C}^M$, input wave field $\mathbf{u}_{\text{in}} \in \mathbb{C}^N$.
$(\mathbf{z}, \widehat{\mathbf{u}}, \{\mathbf{s}^k\}, \{\gamma_k\}, \{\mu_k\}) \leftarrow$ run Algorithm 1
$\mathbf{q}^{K+1} \leftarrow \mathbf{0}$
$\mathbf{q}^K \leftarrow \mathrm{diag}(\mathbf{f})^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}(\mathbf{z} - \mathbf{y})$
$\mathbf{r}^K \leftarrow \mathrm{diag}(\widehat{\mathbf{u}})^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}(\mathbf{z} - \mathbf{y})$
for $k \leftarrow K$ to 1 **do**
  $\mathbf{S}^k \triangleq \mathbf{I} - \gamma_k \mathbf{A}^{\mathsf{H}}\mathbf{A}$
  $\mathbf{T}^k \triangleq \mathrm{diag}(\mathbf{G}^{\mathsf{H}}(\mathbf{A}\mathbf{s}^k - \mathbf{u}_{\text{in}}))^{\mathsf{H}} + \mathrm{diag}(\mathbf{s}^k)^{\mathsf{H}}\mathbf{G}^{\mathsf{H}}\mathbf{A}$
  $\mathbf{q}^{k-1} \leftarrow (1 - \mu_k)\mathbf{S}^k\mathbf{q}^k + \mu_{k+1}\mathbf{S}^{k+1}\mathbf{q}^{k+1}$
  $\mathbf{r}^{k-1} \leftarrow \mathbf{r}^k + \gamma_k \mathbf{T}^k \mathbf{q}^k$
return: $\nabla\mathcal{D}(\mathbf{f}) = \mathrm{Re}\{\mathbf{r}^0\}$ the gradient in (10).

III. Inverse Problem

We now present the overall image reconstruction algorithm, based on the state-of-the-art fast iterative shrinkage/thresholding algorithm (FISTA) [65]. The application of FISTA to nonlinear inverse scattering is, however, nontrivial due to the requirement of the gradient of the scattered field with respect to the object. We solve this by providing an explicit formula, based on error backpropagation [74].

A. Image Reconstruction

We formulate image reconstruction as the following optimization problem

$$\widehat{\mathbf{f}} = \underset{\mathbf{f} \in \mathbb{R}^N}{\arg\min} \{\mathcal{D}(\mathbf{f}) + \tau\mathcal{R}(\mathbf{f})\}, \tag{8a}$$

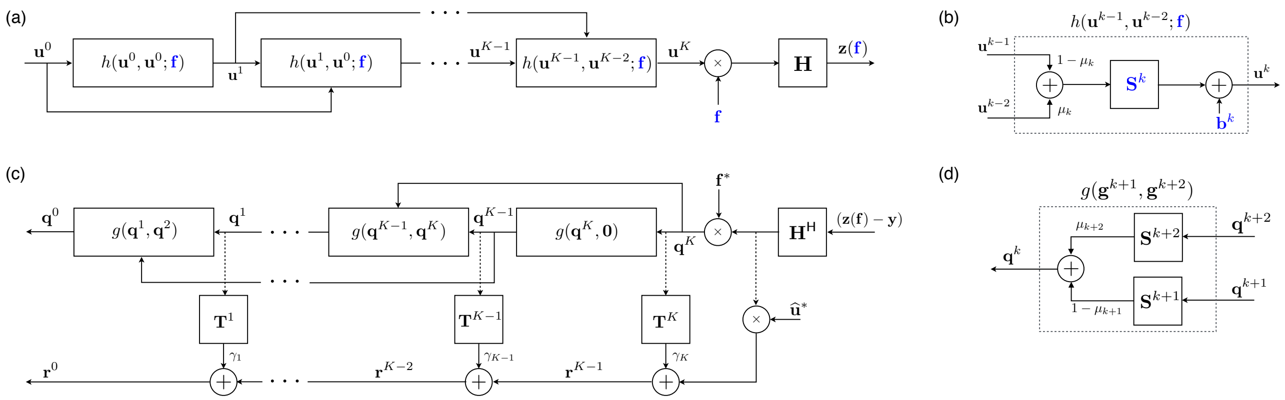

Fig. 3: A schematic representation of the method with adaptable parameters $\mathbf{S}^k \triangleq \mathbf{I} - \gamma_k \mathbf{A}^\mathsf{H}\mathbf{A}$ and $\mathbf{b}^k \triangleq \gamma_k \mathbf{A}^\mathsf{H}\mathbf{u}^0$ that depend on the scattering potential $\mathbf{f}$. (a) Forward model computation corresponding to Algorithm 1 for K forward iterations. Adaptable parameters are marked in blue. (b) The schematic view of a single forward iteration. (c) Error backpropagation corresponding to Algorithm 3 for K iterations. (d) The schematic view of a single backward iteration. The notation $\mathbf{v}^*$ indicates the elementwise complex conjugation of the vector $\mathbf{v}$. Note that the algorithm does not require physical storage of matrices, as they can be efficiently implemented as convolutions using FFT.

where

$$\mathcal{D}(\mathbf{f}) \triangleq \frac{1}{2}\|\mathbf{y} - \mathbf{z}(\mathbf{f})\|_{\ell_2}^2 \quad \text{and} \tag{8b}$$

$$\mathcal{R}(\mathbf{f}) \triangleq \sum_{n=1}^{N} \|[\mathbf{D}\mathbf{f}]_n\|_{\ell_2} = \sum_{n=1}^{N} \sqrt{\sum_{d=1}^{D} |[\mathbf{D}_d\mathbf{f}]_n|^2}. \tag{8c}$$

The data-fidelity term $\mathcal{D}$ measures the discrepancy between the actual measurements $\mathbf{y}$ and the ones predicted by our scattering model $\mathbf{z}$. The function $\mathcal{R}$ is the isotropic TV regularizer and the parameter $\tau > 0$ controls the strength of the regularization, where $\mathbf{D} : \mathbb{R}^N \to \mathbb{R}^{N \times D}$ is the discrete gradient operator with matrix $\mathbf{D}_d$ denoting the finite difference operation along dimension d.

The image can then be formed iteratively using a first order method such as ISTA [75]–[77]

$$\mathbf{f}^t \leftarrow \mathrm{prox}_{\gamma\tau\mathcal{R}}\left(\mathbf{f}^{t-1} - \gamma\nabla\mathcal{D}(\mathbf{f}^{t-1})\right), \tag{9}$$

for $t = 1, 2, 3, \ldots$ or its accelerated variant FISTA [65] summarized in Algorithm 2. Note that the algorithm relies on the definition of the quadratic upper bound

$$Q_\gamma(x, y) \triangleq \mathcal{D}(y) + \nabla\mathcal{D}(y)^\mathrm{T}(x - y) + \frac{1}{2\gamma}\|x - y\|_2^2 + \tau\mathcal{R}(x)$$

for setting the step-size parameter $\gamma > 0$ using the line search. The operator $\mathrm{prox}_{\gamma\tau\mathcal{R}}$ denotes the proximity operator, and for isotropic TV it can be efficiently evaluated [14], [78]. Finally, an efficient implementation of the imaging algorithm requires the gradient of the data-fidelity term

$$\nabla\mathcal{D}(\mathbf{f}) = \mathrm{Re}\left\{\left[\frac{\partial\,\mathbf{z}(\mathbf{f})}{\partial\mathbf{f}}\right]^\mathsf{H}(\mathbf{z}(\mathbf{f}) - \mathbf{y})\right\}, \tag{10}$$

which can be evaluated explicitly using Algorithm 3.

The mathematical derivation of Algorithm 3 is given in Appendix I. It is similar to the derivation of the standard backpropagation used in deep learning [74], [79]. Figure 3 visually illustrates the steps required for the forward model and backpropagation computations in Algorithm 1 and Algorithm 3, respectively. In particular, Figure 3(a) and Figure 3(b) illustrate the way intermediate iterates are combined during the K forward iterations and the schematic of a single iteration, respectively. Similarly, Figure 3(c) and Figure 3(d) illustrate the computation of intermediate iterates in K backward iterations and the schematic of a single such iteration, respectively. Note that the algorithm does not need to explicitly store the matrices, as they can be implemented as convolutions using the fast Fourier transform (FFT) algorithm. Thus the memory required for the algorithm only includes the storage of iterate vectors in Algorithm 3. The overall per iteration complexity of SEAGLE for each illumination is $\mathcal{O}(KN\log N)$ where K is the number of AGM terms and N is the dimension of the imaging domain.

One of the key benefits of SEAGLE is that it offers high levels of parallelism making it well suited for GPU implementations. In particular, computations can be treated independently in parallel for each illumination, which greatly reduces the computational cost of image formation. It is possible to further reduce the cost of each imaging iteration, by considering the incremental variant of Algorithm 2 that processes only a subset of illuminations at each iteration [80].

While the theoretical convergence of FISTA is difficult to analyze for nonconvex functions, it is often used as a faster alternative to the standard gradient-based methods in the context deep learning and broader machine learning [81]–[83]. In fact, we observed that our method reliably converges and achieves excellent results on a wide array of problems, as reported in Section IV.

IV. Experimental Evaluation

We now present the results of validating our method on analytically obtained scattering data for simple scenarios, scattering data obtained with a high-fidelity simulator, and experimentally collected data from the public dataset [84].

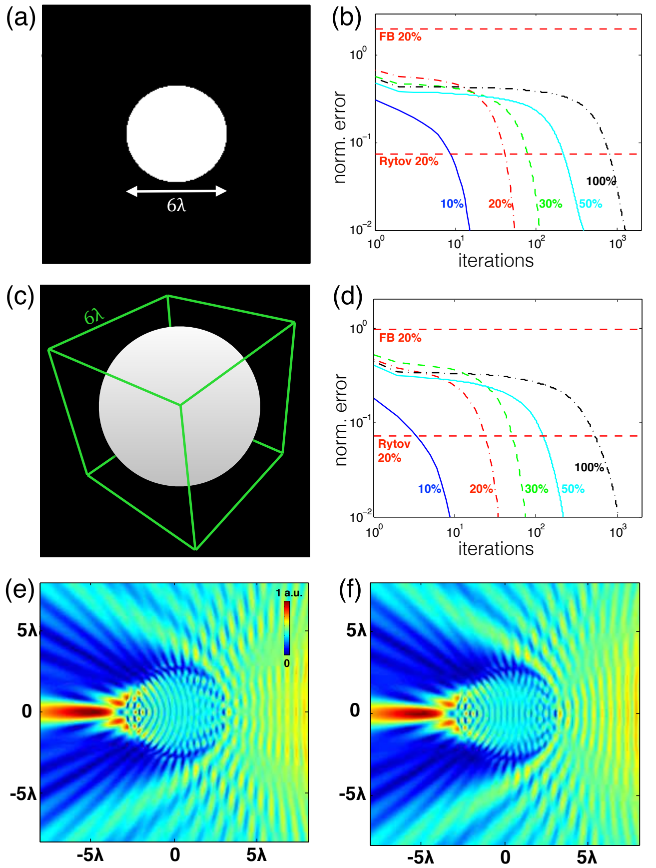

Fig. 4: Analytical validation of the forward model: (a) cylinder of diameter 6λ; (b) normalized errors in scattering for different contrast levels; (c) sphere of diameter 6λ; (d) normalized errors in scattering for different contrast levels; (e) analytic field for a cylinder at a contrast level of 100%; (f) corresponding field computed by our forward model.

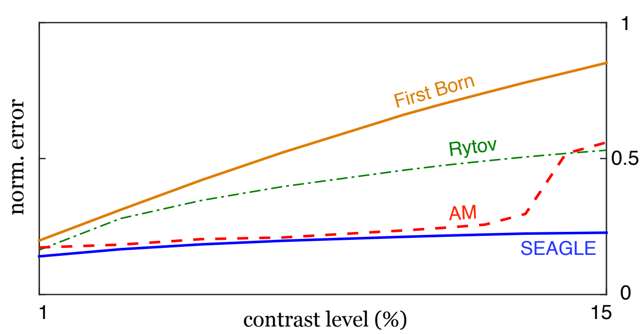

Fig. 5: Quantitative evaluation of normalized reconstruction error against the contrast level for four methods: first-Born, Rytov, AM, and SEAGLE. All the results were obtained by using TV regularization.

Note that all the image reconstruction results reported in this section rely on TV regularization.

A. Validation on analytic data

In the first set of experiments, we validated our forward model for two simple objects where analytic expressions

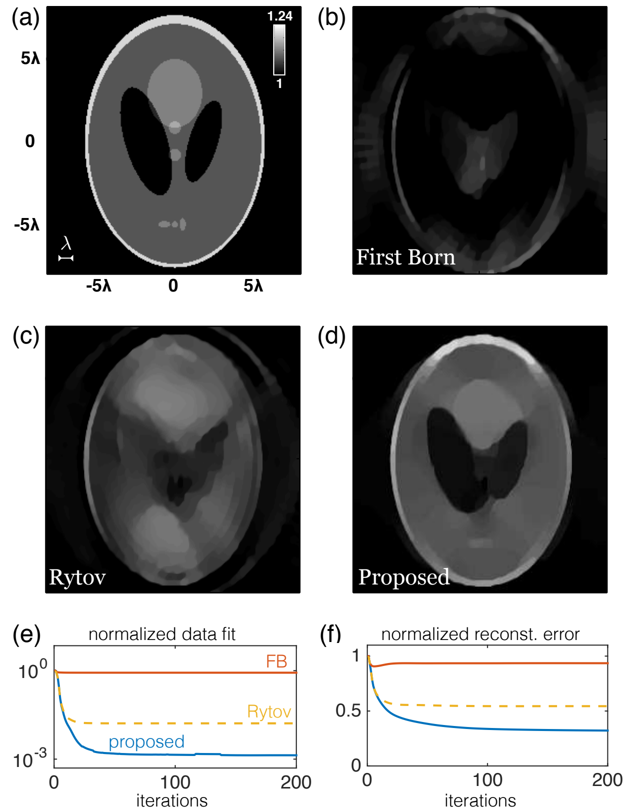

Fig. 6: Comparison of the proposed approach with baseline methods on simulated data. (a) Shepp-Logan at 20% contrast; the reconstructed results with (b) the first Born approximation; (c) the Rytov approximation, and (d) our method; (e) per-iteration data fit; (f) per-iteration error.

of the scattered wave exist: a two-dimensional point source scattered by a cylinder, and a three-dimensional point source scattered by a sphere. The expressions are derived following the mathematical formalism in [85], which we review in Appendix II for completeness. As illustrated in Figures 4(a) and (b), in both cases, the objects have diameters equal to 6 wavelengths. The wavelength is set to 74.9 mm, the source is placed 1 m away from the center of the objects, the grid size is set to 4.8 mm (6 mm), and there are 250 points (128 points) along each axis in 2D (3D). The contrast of an object is defined as $\max(|\mathbf{f}|)/k_b^2$. In Figures 4(b) and (d), we quantitatively evaluate the performance of our forward model with the normalized error defined as

$$\text{normalized error} \triangleq \frac{\|\widehat{\mathbf{u}} - \mathbf{u}_{\text{true}}\|_{\ell_2}^2}{\|\mathbf{u}_{\text{true}}\|_{\ell_2}^2}, \tag{11}$$

where $\widehat{\mathbf{u}}$ is the solution of (5) and $\mathbf{u}_{\text{true}}$ is the analytic expression. For comparison, we additionally provide the performance of the first-Born (FB) and Rytov approximations at 20% contrast. In Fig. 4(e) and 4(f), we demonstrate a visual comparison between the analytic expression and the result of our model. In 2D, the forward computations took 0.03, 0.18, 0.62, 1.3, 4.4, and 14.4 seconds for 5%, 10%, 20%, 30%, 50%, and 100% contrast levels, respectively. Similarly, in 3D, the computations took 2.1, 7.7, 32, 200, and 981.7 seconds for the same contrast levels. Overall, we observed that, by allowing for a large enough value of K, our forward model can match the analytically obtained field with arbitrarily high precision. The actual value of K depends on the severity of multiple scattering and must be adapted on the basis of the application of interest. For example, we observed that for objects closely resembling biological samples, one generally requires $10 \leq K \leq 30$.

B. Validation on simulated data

We next validated the proposed technique for reconstructing the Shepp-Logan phantom in an ill-posed, strongly scattering, and compressive regime ($M = 25 \times 338$ and $N = 250 \times 250$). Specifically, we consider the setup in Fig. 2 where the scattered wave measurements are generated with an FDTD simulator. The object is of size 84.9 cm $\times$ 113 cm. We place two linear detectors on either side of the phantom at a distance of 95.9 cm from the center of the object. Each detector has 169 sensors placed with a spacing of 3.84 cm. The transmitters are positioned on a line 48.0 cm left to the left detector. They are spaced uniformly in azimuth with respect to the center of the phantom (every $5°$ within $\pm 60°$). We set up a 120 cm $\times$ 120 cm square area for reconstructing the object, with pixel size 0.479 cm. The wavelength of the illuminating light is 7.49 cm.

We compare results of our approach against three alternative methods. We regularize the solution of all the methods with TV. As the first reference method, we consider the solution of the linearized model based on FB, which is known to be valid only for weakly scattering objects. Additionally, we consider an inverse scattering approach that is based on the Rytov approximation, which is known to be more robust to moderate levels of scattering. Finally, we consider a popular optimization scheme extensively used in optical imaging and FWI, denoted AM for *alternating minimization*, for strongly scattering objects that iteratively alternates between updating the contrast function for a fixed field and updating the field for a fixed contrast function [33], [39], [40]. All three methods minimize the same error functional; however, each method relies on a distinct forward model. Image reconstruction in all the approaches was done using FISTA with TV regularizer that was empirically set for the best performance. The order of SEAGLE's forward model is set to $K = 120$, but Algorithm 1 may terminate earlier when the objective function (5) is below $\delta_{\text{tol}} = 5 \times 10^{-7} \|\mathbf{u}_{\text{in}}\|_{\ell_2}^2$.

Figure 5 summarizes the normalized error performance $\|\widehat{\mathbf{f}} - \mathbf{f}\|_{\ell_2}^2 / \|\mathbf{f}\|_{\ell_2}^2$, where $\mathbf{f}$ and $\widehat{\mathbf{f}}$ denote the true and estimated object, of all the methods for various contrast levels between 1% and 15%. The results confirm that while all the methods yield good performance at low contrasts, the performance of linearized methods, FB and Rytov, degenerate as contrast levels increase. One can also observe that the performance of AM is similar to SEAGLE for low to moderate contrast levels, but SEAGLE outperforms AM for higher contrasts. In fact, we generally observed that the performance of AM rapidly degenerates for very high contrast levels, while the performance of SEAGLE is relatively stable.

Figure 6 summarizes the performance of the same methods for the contrast level of 20%. We have omitted the result of AM as the method was not able to reconstruct the image at this high contrast level. The figure additionally provides quantitative performance evaluation in terms of data fit

$$\text{normalized data fit} \triangleq \frac{\mathcal{D}(\widehat{\mathbf{f}})}{\mathcal{D}(\mathbf{0})} = \frac{\|\mathbf{z}(\widehat{\mathbf{f}}) - \mathbf{y}\|_{\ell_2}^2}{\|\mathbf{y}\|_{\ell_2}^2}. \qquad (12)$$

Simulation results corroborate the benefit of using the proposed method for strongly scattering objects. It can be seen that, due to the ill-posed nature of the measurements, the reconstructed images suffer from missing frequency artifacts [86]. However, the proposed method is still able to accurately capture most features of the object while the linear methods fail to do so. Note also, that our method was initialized with the background value of the dielectric permittivity, $\epsilon_{\text{b}} = 1$, and that it takes fewer than 50 FISTA iterations for converging to a stationary point (see convergence plot in Figure 6(f)). It took SEAGLE about 1.5 seconds on average to process an illumination at each FISTA iteration.

C. Validation on experimental data

We apply our method to three objects from the public dataset provided by the Fresnel institute [17]: *FoamDielExtTM*, *FoamDielIntTM*, and *FoamTwinDielTM*. These objects are placed in a region of size 15 cm $\times$ 15 cm at the center of a circular rim of radius 1.67 m and measured using 360 detectors and 8 transmitters evenly distributed on the rim. The number of transmitters is increased to 18 for *FoamTwinDielTM* and are also uniformly spaced. In all cases, only one transmitter is turned on at a time, while 241 detectors are used for each transmitter by excluding 119 detectors that are closest to the transmitter. While the full data contains multiple frequency measurements, we only use the data corresponding to the 3 GHz. As before, we compare the result of our method with the first-Born and Rytov approximations, as well as AM, all regularized with TV. We set the highest order of SEAGLE terms to $K = 200$ and the TV regularization to $\tau = 0.25 \times 10^{-8} \|\mathbf{y}\|_{\ell_2}^2$, and run the image formation algorithm for 40 FISTA iterations. The reconstruction was initialized with the background value of the dielectric permittivity, which in this case corresponds to $\epsilon_{\text{b}} = 1$.

Figure 7 summarizes the imaging results on the experimental data. The quantitative evaluation is performed using the same metrics as before. The results show that our method successfully captures the shape of the objects, as well as the value of the permittivity. Both first Born and Rytov approximations underestimate the permittivity. One can also see that the data-fit error for both of the linear forward models remain high as iterations progress. On the other hand, the object reconstructed by the proposed method closely agrees with the measured data (see rightmost column in Figure 7). It took the proposed method about 6.59 seconds on average to process an illumination at each FISTA iteration.

Figure 8 illustrates the performance of our method when using a limited number of measured data. In particular, we consider the reconstruction of the same three objects, but

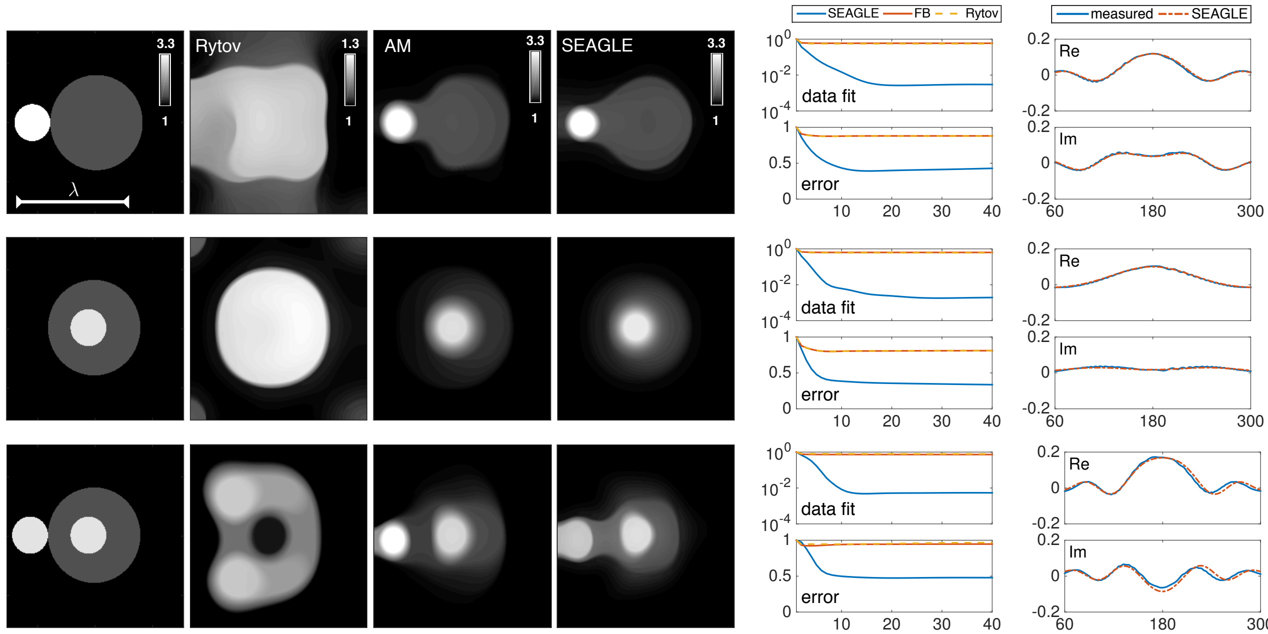

Fig. 7: Reconstruction from an experimentally measured objects at 3 GHz, from top to bottom: *FoamDielExtTM*, *FoamDielIntTM*, and *FoamTwinDielTM*. From left to right: ground truth; reconstruction using the Rytov approximation; reconstruction based on alternating minimization; reconstruction with SEAGLE; evolution of normalized data-fit (top) and the normalized reconstruction error (bottom); and the true and predicted measurements for the transmission angle zero.

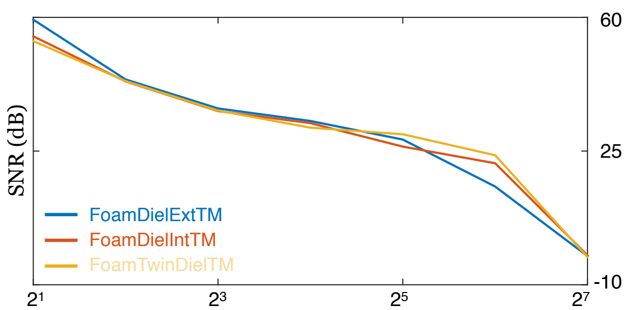

Fig. 8: Reconstruction quality (see text) of the proposed method at various values of the data-reduction factor.

reduce the number of measurements for each transmission using regular downsampling by factors of $2, 4, 8, 16, 32, 64$, and 128. The full dataset consists of 8 transmissions with 241 measurements each; a factor of 128 downsampling reduces to 8 transmissions with 2 measurements each. The size of the reconstructed image is set to 320×320 pixels. The reconstruction performance is quantified as

$$\text{SNR (dB)} \triangleq 10 \log_{10} \left(\frac{\|\mathbf{f}_{\text{ref}}\|_{\ell_2}^2}{\|\widehat{\mathbf{f}} - \mathbf{f}_{\text{ref}}\|_{\ell_2}^2} \right), \qquad (13)$$

where $\mathbf{f}_{\text{ref}}$ is the reconstructed image with all the measured data (see Figure 7). The visual illustration is provided for *FoamDielExtTM* in Figure 1. This result highlights the stability of the proposed method to subsampling and experimental noise, even at highly nonlinear scattering scenarios.

Note that several other methods have been tested on this dataset [41], [66]–[69]. Qualitative comparison of our results in Figure 7 with the results of those methods indicates that our approach achieves comparable performance using only a fraction of data (*i.e.*, a single frequency with possible subsampling). Additionally, we observe a reliably stable and fast convergence starting from the initialization to the background permittivity, which is desirable in strongly scattering regimes.

V. Conclusion

In conclusion, we have demonstrated a nonconvex optimization technique for solving nonlinear inverse scattering problems. We have applied the technique to simulated and experimentally measured data in microwave frequencies. The scattering was modeled iteratively as a series expansion with Nesterov's accelerated-gradient method. By structuring the expansion as a recursive feedforward network, we derived a backpropagation formula for evaluating the gradient that can be used for fast iterative image reconstruction. The algorithm yields images of better quality than methods using linear forward models and is competitive with state-of-the-art inverse scattering approaches, tested on the same dataset. While the optimization problem is not convex, we have observed that the algorithm converges reliably within 100 iterations from a constant initialization of the permittivity. Our approach provides a promising framework for active correction of scattering in various applications and has the potential of significantly increasing the resolution and robustness when imaging strongly scattering objects.

Appendix I
Derivation of Error Backpropagation

In this appendix, we provide the derivation of error back-propagation applied to our method. The method essentially computes the gradient of the data-fidelity term. This gradient is a key step of updating the scattering potential in solving the inverse problem. We now present the mathematical derivation of the gradient computation and relate it to Algorithm 3.

The inputs of the error-back propagation are the data mismatch and the intermediate variables ($\{\mathbf{s}^k\}$, $\{\gamma_k\}$, $\{\mu_k\}$, $\widehat{\mathbf{u}} = \mathbf{u}^K$) of the forward model computation, and the output is the gradient. Here we follow the differentiation conventions for vectors: $(\frac{\partial \mathbf{u}}{\partial \mathbf{f}})_{ij} = \frac{\partial \mathbf{u}_i}{\partial \mathbf{f}_j}$ and $(\nabla_{\mathbf{f}} \mathbf{u})_{ij} = [(\frac{\partial \mathbf{u}}{\partial \mathbf{f}})^{\mathsf{H}}]_{ij} = \frac{\partial \mathbf{u}_j^*}{\partial \mathbf{f}_i}$. All boldface lower-case variables are column vectors.

Let us begin with the gradient of $\mathcal{D} = \frac{1}{2}\|\mathbf{z}-\mathbf{y}\|_2^2$.

$$\begin{aligned}\nabla_{\mathbf{f}}\mathcal{D} =& \frac{1}{2}\nabla_{\mathbf{f}}[(\mathbf{z}-\mathbf{y})^{\mathsf{H}}(\mathbf{z}-\mathbf{y})] \\ =& \frac{1}{2}[(\nabla_{\mathbf{f}}\mathbf{z})(\mathbf{z}-\mathbf{y}) + ((\mathbf{z}-\mathbf{y})^{\mathsf{H}}(\nabla_{\mathbf{f}}\mathbf{z})^{\mathsf{H}})^{\mathsf{T}}] \\ =& \mathrm{Re}\left\{(\nabla_{\mathbf{f}}\mathbf{z})(\mathbf{z}-\mathbf{y})\right\}. \end{aligned} \tag{14}$$

This can be evaluated by applying the chain rule to $\nabla_{\mathbf{f}}\mathbf{z}$ and all the variables composing $\mathbf{z}$. The equations leading from the initialization all the way to $\mathbf{z}$ are listed below:

$$\begin{aligned}\mathbf{z} &= \mathbf{u}_{\text{in}} + \mathbf{H}\mathrm{diag}(\mathbf{f})\mathbf{u}^K \\ \mathbf{s}^k &= (1-\mu_k)\mathbf{u}^{k-1} + \mu_k\mathbf{u}^{k-2} \\ \mathbf{u}^k &= \mathbf{s}^k - \gamma_k\mathbf{A}^{\mathsf{H}}(\mathbf{A}\mathbf{s}^k - \mathbf{u}_{\text{in}}),\end{aligned}$$

for $k = 1, \dots, K$, where $\mathbf{A} \triangleq \mathbf{I} - \mathbf{G}\mathrm{diag}(\mathbf{f})$, and $\mathbf{u}^{-1} = \mathbf{u}^0$. It is worth noting that, while the step-size γ_k also depends on $\mathbf{f}$, we ignore this dependency to simplify the computation. The rationale for this simplification is that the step-size can be replaced by a fixed one. Furthermore, in practice, γ_k attains a stationary value for large enough k, which indicates that this simplification has a negligible effect on backpropagation.

A. Initialization of backpropagation

The initialization in Algorithm 3 is obtained by differentiating the first of the above equations with respect to $\mathbf{f}$. With $\mathrm{diag}(\mathbf{f})\mathbf{u} = \mathrm{diag}(\mathbf{u})\mathbf{f}$, we have

$$\begin{aligned}\nabla_{\mathbf{f}}\mathbf{z} &= \left[\mathbf{H}\frac{\partial \mathbf{f}}{\partial \mathbf{f}}\mathrm{diag}(\mathbf{u}^K) + \mathbf{H}\mathrm{diag}(\mathbf{f})\frac{\partial \mathbf{u}^K}{\partial \mathbf{f}}\right]^{\mathsf{H}} \\ &= \mathrm{diag}(\mathbf{u}^K)^{\mathsf{H}}\mathbf{H}^{\mathsf{H}} + (\nabla_{\mathbf{f}}\mathbf{u}^K)\mathrm{diag}(\mathbf{f})^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\end{aligned} \tag{15}$$

The first term gives the remainder that contributes to the final result while the second term gives the vector that multiplies with $\nabla_{\mathbf{f}}\mathbf{u}^K$. For convenience, we define two sets of vectors:

- $\mathbf{q}^k$: the vector that multiplies with $\nabla_{\mathbf{f}}\mathbf{u}^k$
- $\mathbf{r}^k$: the remainder before computing $(\nabla_{\mathbf{f}}\mathbf{u}^k)\mathbf{q}^k$.

In addition, due to the acceleration step in the forward computation, we expect subsequent $\mathbf{q}^{k-1}$ to have a contribution from $(\nabla_{\mathbf{f}}\mathbf{u}^{k+1})\mathbf{q}^{k+1}$ in addition to the contribution from its direct neighbor $(\nabla_{\mathbf{f}}\mathbf{u}^k)\mathbf{q}^k$. This leads to the third set of vectors:

- $\mathbf{p}^k$: the explicit contribution of $(\nabla_{\mathbf{f}}\mathbf{u}^{k+1})\mathbf{q}^{k+1}$ to $\mathbf{q}^{k-1}$.

Finally, multiplying (15) with $(\mathbf{z}-\mathbf{y})$ we identify

$$\mathbf{r}^K = \mathrm{diag}(\mathbf{u}^K)^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}(\mathbf{z}-\mathbf{y}) \tag{16}$$

$$\mathbf{q}^K = \mathrm{diag}(\mathbf{f})^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}(\mathbf{z}-\mathbf{y}). \tag{17}$$

Since there is no term multiplying with $\nabla_{\mathbf{f}}\mathbf{u}^{K-1}$ explicitly (hence nothing to pass to $\mathbf{q}^{K-1}$), we have

$$\mathbf{p}^K = \mathbf{0}. \tag{18}$$

B. Recursive updates for $\mathbf{u}^k$

The computation of $(\nabla_{\mathbf{f}}\mathbf{u}^k)\mathbf{q}^k$ is the key step in error-back propagation. We evaluate the gradient $\nabla_{\mathbf{f}}\mathbf{u}^k$ by taking the Hermitian of the derivative, and the multiplication with $\mathbf{q}^k$ follows. The result should be passed onto another gradient with smaller k. Before we start, let us write out the gradient of $\mathbf{s}^k$ which is straightforward from its definition,

$$\nabla_{\mathbf{f}}\mathbf{s}^k = (1-\mu_k)\nabla_{\mathbf{f}}\mathbf{u}^{k-1} + \mu_k\nabla_{\mathbf{f}}\mathbf{u}^{k-2}. \tag{19}$$

The derivative of $\mathbf{u}^k$ is

$$\begin{aligned}\frac{\partial \mathbf{u}^k}{\partial \mathbf{f}} = &-\gamma_k\frac{\partial \mathbf{A}^{\mathsf{H}}}{\partial \mathbf{f}}(\mathbf{A}\mathbf{s}^k - \mathbf{u}_{\text{in}}) - \gamma_k\mathbf{A}^{\mathsf{H}}\frac{\partial \mathbf{A}}{\partial \mathbf{f}}\mathbf{s}^k \\ &+ (\mathbf{I} - \gamma_k\mathbf{A}^{\mathsf{H}}\mathbf{A})\frac{\partial \mathbf{s}^k}{\partial \mathbf{f}}.\end{aligned} \tag{20}$$

The first term becomes

$$\begin{aligned}-\gamma_k\frac{\partial \mathbf{A}^{\mathsf{H}}}{\partial \mathbf{f}}(\mathbf{A}\mathbf{s}^k - \mathbf{u}_{\text{in}}) &= \gamma_k\left(\frac{\partial}{\partial \mathbf{f}}\mathrm{diag}(\mathbf{f})\right)\mathbf{G}^{\mathsf{H}}(\mathbf{A}\mathbf{s}^k - \mathbf{u}_{\text{in}}) \\ &= \gamma_k\mathrm{diag}(\mathbf{G}^{\mathsf{H}}(\mathbf{A}\mathbf{s}^k - \mathbf{u}_{\text{in}})),\end{aligned}$$

and the second term becomes

$$\begin{aligned}-\gamma_k\mathbf{A}^{\mathsf{H}}\frac{\partial \mathbf{A}}{\partial \mathbf{f}}\mathbf{s}^k &= \gamma_k\mathbf{A}^{\mathsf{H}}\mathbf{G}\left(\frac{\partial}{\partial \mathbf{f}}\mathrm{diag}(\mathbf{f})\right)\mathbf{s}^k \\ &= \gamma_k\mathbf{A}^{\mathsf{H}}\mathbf{G}\mathrm{diag}(\mathbf{s}^k).\end{aligned}$$

By taking Hermitian transpose of (20), we have

$$\nabla_{\mathbf{f}}\mathbf{u}^k = \gamma_k\mathbf{T}^k + (\nabla_{\mathbf{f}}\mathbf{s}^k)\mathbf{S}^k \tag{21}$$

where

$$\mathbf{T}^k = \mathrm{diag}(\mathbf{G}^{\mathsf{H}}(\mathbf{A}\mathbf{s}^k - \mathbf{u}_{\text{in}}))^{\mathsf{H}} + \mathrm{diag}(\mathbf{s}^k)^{\mathsf{H}}\mathbf{G}^{\mathsf{H}}\mathbf{A} \tag{22}$$

$$\mathbf{S}^k = (\mathbf{I} - \gamma_k\mathbf{A}^{\mathsf{H}}\mathbf{A})^{\mathsf{H}} = \mathbf{I} - \gamma_k\mathbf{A}^{\mathsf{H}}\mathbf{A}. \tag{23}$$

By multiplying (21) with $\mathbf{q}^k$ and substituting with (19), we obtain the expression for $(\nabla_{\mathbf{f}}\mathbf{u}^k)\mathbf{q}^k$,

$$\begin{aligned}(\nabla_{\mathbf{f}}\mathbf{u}^k)\mathbf{q}^k = &\gamma_k\mathbf{T}^k\mathbf{q}^k \\ &+ (\nabla_{\mathbf{f}}\mathbf{u}^{k-1})\left[(1-\mu_k)\mathbf{S}^k\mathbf{q}^k\right] \\ &+ (\nabla_{\mathbf{f}}\mathbf{u}^{k-2})\left[\mu_k\mathbf{S}^k\mathbf{q}^k\right].\end{aligned} \tag{24}$$

Note that because we set $\mathbf{u}^{-1} = \mathbf{u}^0$,

$$(\nabla_{\mathbf{f}}\mathbf{u}^1)\mathbf{q}^1 = \gamma_1\mathbf{T}^1\mathbf{q}^1 + (\nabla_{\mathbf{f}}\mathbf{u}^0)\mathbf{S}^1\mathbf{q}^1. \tag{25}$$

C. Error backpropagation equations

From equations (15) to (18), we have

$$(\nabla_{\mathbf{f}}\mathbf{z})(\mathbf{z}-\mathbf{y}) = \mathbf{r}^K + (\nabla_{\mathbf{f}}\mathbf{u}^K)\mathbf{q}^K + (\nabla_{\mathbf{f}}\mathbf{u}^{K-1})\mathbf{p}^K \quad (26)$$

Substituting (24) into (26), we have the following expressions

$$\begin{aligned}(\nabla_{\mathbf{f}}\mathbf{z})(\mathbf{z}-\mathbf{y}) &= \mathbf{r}^K + (\nabla_{\mathbf{f}}\mathbf{u}^K)\mathbf{q}^K + (\nabla_{\mathbf{f}}\mathbf{u}^{K-1})\mathbf{p}^K \\ &= \mathbf{r}^{K-1} + (\nabla_{\mathbf{f}}\mathbf{u}^{K-1})\mathbf{q}^{K-1} + (\nabla_{\mathbf{f}}\mathbf{u}^{K-2})\mathbf{p}^{K-1} \\ &= \dots \\ &= \mathbf{r}^1 + (\nabla_{\mathbf{f}}\mathbf{u}^1)\mathbf{q}^1 + (\nabla_{\mathbf{f}}\mathbf{u}^0)\mathbf{p}^1 \end{aligned} \quad (27)$$

and the recursion relations for $k = 2\dots K$

$$\mathbf{r}^{k-1} = \mathbf{r}^k + \gamma_k \mathbf{T}^k\mathbf{q}^k \quad (28)$$

$$\mathbf{q}^{k-1} = \mathbf{p}^k + (1-\mu_k)\mathbf{S}^k\mathbf{q}^k \quad (29)$$

$$\mathbf{p}_{k-1} = \mu_k \mathbf{S}^k\mathbf{q}^k. \quad (30)$$

For the case $k = 1$, or namely $\mathbf{r}_0$ and $\mathbf{q}_0$ (note that $\mathbf{p}_0$ does not exist due to $\mathbf{u}^{-1} = \mathbf{u}^0$), we plug (25) into (27),

$$(\nabla_{\mathbf{f}}\mathbf{z})(\mathbf{z}-\mathbf{y}) = \mathbf{r}^1 + \gamma_1\mathbf{T}^1\mathbf{q}^1 + (\nabla_{\mathbf{f}}\mathbf{u}^0)\left[\mathbf{S}^1\mathbf{q}^1 + \mathbf{p}^1\right].$$

Hence we have

$$\mathbf{r}^0 = \mathbf{r}^1 + \gamma_1\mathbf{T}^1\mathbf{q}^1 \quad (31)$$

$$\mathbf{q}^0 = \mathbf{p}^1 + \mathbf{S}^1\mathbf{q}^1. \quad (32)$$

In the initialization of our forward model, $\mathbf{u}^0$ is the incident field and does not depend on $\mathbf{f}$. Therefore $\nabla_{\mathbf{f}}\mathbf{u}^0 = \mathbf{0}$ and the gradient of data-fidelity is

$$\nabla_{\mathbf{f}}\mathcal{D} = \mathrm{Re}\left\{(\nabla_{\mathbf{f}}\mathbf{z})(\mathbf{z}-\mathbf{y})\right\} = \mathrm{Re}\left\{\mathbf{r}^0\right\}. \quad (33)$$

We summarize these recursion relations of error backpropagation in Algorithm 3 of the main text.

APPENDIX II
ANALYTIC SOLUTIONS TO SCATTERING

In this section, our aim is to present the analytic expressions for scalar electric fields resulting from a point source outside a dielectric sphere in 2D and 3D (strictly speaking, the 2D case should be understood as an infinitely long line source illuminating a cylinder parallel to it and looking at the cross-section). A sketch of the derivation is provided after the actual expressions. A more complete description can be found in a number of standard textbooks such as [85].

A. Expressions

Consider a sphere of a radius r_{sph} and a refractive index $n = \sqrt{\epsilon}$. The source is located r_{s} distance away from the center of the sphere and the wavenumber of the source outside the sphere is k_b.

2D case:
We consider the polar coordinates

$$\mathbf{r} = (r\cos\theta, r\sin\theta),$$

and, without loss of generality, assume that the source is at $\theta_{\text{s}} = 0$. The field can be expressed as

$$E(\mathbf{r}; r_{\text{s}}) = \sum_{m=-\infty}^{\infty} R_m(r, r_{\text{s}})\frac{e^{\mathrm{j}m\theta}}{2\pi} \quad (34)$$

where $\rho = k_b r$, $\rho_{\text{sph}} = k_b r_{\text{sph}}$ and $\rho_{\text{s}} = k_b r_{\text{s}}$,

$$R_m(r, r_{\text{s}}) = \begin{cases} a_m J_m(n\rho)H_m^{(1)}(\rho_{\text{s}}), & r < r_{\text{sph}} \\ (b_m J_m(\rho) + c_m Y_m(\rho))H_m^{(1)}(\rho_{\text{s}}), & r_{\text{sph}} \le r < r_{\text{s}} \\ (b_m J_m(\rho_{\text{s}}) + c_m Y_m(\rho_{\text{s}}))H_m^{(1)}(\rho), & r_{\text{s}} \le r \end{cases} \quad (35)$$

$$a_m = \frac{-1}{\rho_{\text{sph}}\Delta_m} \quad (36)$$

$$b_m = \frac{-\pi}{2\Delta_m}\begin{vmatrix} J_m(n\rho_{\text{sph}}) & nJ_{m-1}(n\rho_{\text{sph}}) \\ Y_m(\rho_{\text{sph}}) & Y_{m-1}(\rho_{\text{sph}}) \end{vmatrix} \quad (37)$$

$$c_m = \frac{\pi}{2\Delta_m}\begin{vmatrix} J_m(n\rho_{\text{sph}}) & nJ_{m-1}(n\rho_{\text{sph}}) \\ J_m(\rho_{\text{sph}}) & J_{m-1}(\rho_{\text{sph}}) \end{vmatrix} \quad (38)$$

$$\Delta_m = \begin{vmatrix} J_m(n\rho_{\text{sph}}) & nJ_{m-1}(n\rho_{\text{sph}}) \\ H_m^{(1)}(\rho_{\text{sph}}) & H_{m-1}^{(1)}(\rho_{\text{sph}}) \end{vmatrix} \quad (39)$$

and J_m and Y_m are the m'th order Bessel functions of the first kind and the second kind, and $H_m^{(1)} = J_m + \mathrm{j}Y_m$ is the m'th order Hankel's function of the first kind.

3D case:
We consider the spherical coordinates

$$\mathbf{r} = (r\sin\theta\cos\phi, r\sin\theta\sin\phi, r\cos\theta),$$

and, without loss of generality, assume that the source has zenith angle $\theta_{\text{s}} = 0$ and azimuthal angle $\phi_{\text{s}} = 0$. The field then reads

$$E(\mathbf{r}; r_{\text{s}}) = \sum_{l=0}^{\infty} R_l(r, r_{\text{s}})\left(\frac{2l+1}{4\pi}\right)P_l(\cos\theta) \quad (40)$$

where, with $\rho = k_b r$, $\rho_{\text{sph}} = k_b r_{\text{sph}}$ and $\rho_{\text{s}} = k_b r_{\text{s}}$,

$$R_l(r, r_{\text{s}}) = \begin{cases} A_l j_l(n\rho)h_l^{(1)}(\rho_{\text{s}}), & r < r_{\text{sph}} \\ (B_l j_l(\rho) + C_l n_l(\rho))h_l^{(1)}(\rho_{\text{s}}), & r_{\text{sph}} \le r < r_{\text{s}} \\ (B_l j_l(\rho_{\text{s}}) + C_l n_l(\rho_{\text{s}}))h_l^{(1)}(\rho), & r_{\text{s}} \le r \end{cases} \quad (41)$$

$$A_m = \frac{k_b}{\rho_{\text{sph}}^2 D_m} \quad (42)$$

$$B_m = \frac{-k_b}{D_m}\begin{vmatrix} j_l(n\rho_{\text{sph}}) & nj_{l+1}(n\rho_{\text{sph}}) \\ n_l(\rho_{\text{sph}}) & n_{l+1}(\rho_{\text{sph}}) \end{vmatrix} \quad (43)$$

$$C_m = \frac{k_b}{D_m}\begin{vmatrix} j_l(n\rho_{\text{sph}}) & nj_{l+1}(n\rho_{\text{sph}}) \\ j_l(\rho_{\text{sph}}) & j_{l+1}(\rho_{\text{sph}}) \end{vmatrix} \quad (44)$$

$$D_m = \begin{vmatrix} j_l(n\rho_{\text{sph}}) & nj_{l+1}(n\rho_{\text{sph}}) \\ h_l^{(1)}(\rho_{\text{sph}}) & h_{l+1}^{(1)}(\rho_{\text{sph}}) \end{vmatrix} \quad (45)$$

and j_l and n_l are the l'th order spherical Bessel function of the first kind and the second kind, $h_l^{(1)} = j_l + \mathrm{j}n_l$ is the l'th order of spherical Hankel function of the first kind, and $P_l(x)$ is the Legendre polynomial defined as

$$P_l(x) = \frac{1}{2^l l!}\frac{d^l}{dx^l}(x^2+1)^l. \quad (46)$$

B. Helmholtz equation

The Helmholtz equation for a point source is

$$\nabla^2_{\mathbf{r}} E(\mathbf{r}, \mathbf{r}_{\text{s}}) + k^2(\mathbf{r}) E(\mathbf{r}, \mathbf{r}_{\text{s}}) = -\delta(\mathbf{r} - \mathbf{r}_{\text{s}}) \tag{47}$$

where E is the complex electric field at position $\mathbf{r}$ when the point source is at $\mathbf{r}_{\text{s}}$ and $k^2(\mathbf{r})$ is defined as

$$k^2(\mathbf{r}) = k^2(\|\mathbf{r}\|_2) = \begin{cases} n^2 k_b^2, & \text{for } \|\mathbf{r}\|_2 < r_{\text{sph}} \\ k_b^2, & \text{for } \|\mathbf{r}\|_2 > r_{\text{sph}} \end{cases}. \tag{48}$$

Note that both 2D and 3D cases follow the same form. Their difference is that $\mathbf{r}$ and $\mathbf{r}_{\text{s}}$ have 2 or 3 coordinates.

C. Derivation for 2D case

We consider the polar coordinates, assume $\theta_{\text{s}} = 0$ without loss of generality, and use an ansatz for the electric field in (34). With the Laplacian in the polar coordinate and the following expansion of a 2D delta function [85]

$$\delta(\mathbf{r} - \mathbf{r}_{\text{s}}) = \frac{1}{r} \delta(r - r_{\text{s}}) \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} e^{jm(\theta - \theta_{\text{s}})}, \tag{49}$$

eq. (47) becomes a sequence of equations on $R_m(r, r_{\text{s}})$

$$\begin{aligned} &\frac{\partial}{\partial r}\left(r \frac{\partial}{\partial r} R_m(r, r_{\text{s}}) \right) + \left(r k^2(r) - \frac{m^2}{r} \right) R_m(r, r_{\text{s}}) \\ &= -\,\delta(r - r_{\text{s}}) \end{aligned} \tag{50}$$

for each m. Each equations is a Bessel differential equation so the solution can be composed of Bessel functions of order m. The boundary conditions for R_m are as follows

1) finite value at $r = 0$
2) only outgoing component at $r = \infty$
3) continuous and first-derivative-continuous at $r = r_{\text{sph}}$
4) continuous at $r = r_{\text{s}}$
5) $\frac{\partial R_m}{\partial r}\big|_{r_{\text{s}}^+} - \frac{\partial R_m}{\partial r}\big|_{r_{\text{s}}^-} = -\frac{1}{r_{\text{s}}}$ (integrate (50) around r_{s})

The above condition and equations lead to Eqs. (34)-(39).

D. Derivation for 3D case

We consider the spherical coordinates and assume that the source lies on the zenith axis. The ansatz for the electric field is (40), the expansion of a 3D delta function is

$$\begin{aligned} \delta(\mathbf{r} - \mathbf{r}_{\text{s}}) &= \frac{1}{r^2} \delta(r - r_{\text{s}}) \sum_{l=0}^{\infty} \sum_{m=-l}^{l} Y_l^m(\theta, \phi) Y_l^m(0, 0) \\ &= \frac{1}{r^2} \delta(r - r_{\text{s}}) \sum_{l=0}^{\infty} \left(\frac{2l+1}{4\pi} \right) P_l(\cos\theta) \end{aligned} \tag{51}$$

and eq. (47) becomes

$$\begin{aligned} &\frac{\partial}{\partial r}\left(r^2 \frac{\partial}{\partial r} R_l(r, r_{\text{s}}) \right) + \left(k^2(r) r^2 - l(l+1) \right) R_l(r, r_{\text{s}}) \\ &= -\,\delta(r - r_{\text{s}}) \end{aligned} \tag{52}$$

for each l. These equations are spherical Bessel equations and there are corresponding spherical Bessel functions to compose the solution. The boundary conditions for the solution are the same as listed above except the last one becoming

$$\left.\frac{\partial R_l}{\partial r}\right|_{r_{\text{s}}^+} - \left.\frac{\partial R_l}{\partial r}\right|_{r_{\text{s}}^-} = -\frac{1}{r_{\text{s}}^2}. \text{ (integrate (52) around } r_{\text{s}})$$

References

[1] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*. IEEE, 1988.

[2] E. Wolf, “Three-dimensional structure determination of semi-transparent objects from holographic data,” *Opt. Commun.*, vol. 1, no. 4, pp. 153–156, September/October 1969.

[3] A. J. Devaney, “Inverse-scattering theory within the Rytov approximation,” *Opt. Lett.*, vol. 6, no. 8, pp. 374–376, August 1981.

[4] M. Born and E. Wolf, *Principles of Optics*, 7th ed. Cambridge Univ. Press, 1999.

[5] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, February 2006.

[6] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[7] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D*, vol. 60, no. 1–4, pp. 259–268, November 1992.

[8] M. M. Bronstein, A. M. Bronstein, M. Zibulevsky, and H. Azhari, “Reconstruction in diffraction ultrasound tomography using nonuniform FFT,” *IEEE Trans. Med. Imag.*, vol. 21, no. 11, pp. 1395–1401, November 2002.

[9] J. W. Lim, K. R. Lee, K. H. Jin, S. Shin, S. E. Lee, Y. K. Park, and J. C. Ye, “Comparative study of iterative reconstruction algorithms for missing cone problems in optical diffraction tomography,” *Opt. Express*, vol. 23, no. 13, pp. 16 933–16 948, June 2015.

[10] Y. Sung and R. R. Dasari, “Deterministic regularization of three-dimensional optical diffraction tomography,” *J. Opt. Soc. Am. A*, vol. 28, no. 8, pp. 1554–1561, August 2011.

[11] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge Univ. Press, 2004.

[12] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. Springer, 2006.

[13] J. M. Bioucas-Dias and M. A. T. Figueiredo, “A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration,” *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2992–3004, December 2007.

[14] A. Beck and M. Teboulle, “Fast gradient-based algorithm for constrained total variation image denoising and deblurring problems,” *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, November 2009.

[15] B. Chen and J. J. Stamnes, “Validity of diffraction tomography based on the first born and the first rytov approximations,” *Appl. Opt.*, vol. 37, no. 14, pp. 2996–3006, May 1998.

[16] V. Ntziachristos, “Going deeper than microscopy: the optical imaging frontier in biology,” *Nat. Methods*, vol. 7, no. 8, pp. 603–614, August 2010.

[17] J.-M. Geffrin, P. Sabouroux, and C. Eyraud, “Free space experimental scattering database continuation: experimental set-up and measurement precision,” *Inv. Probl.*, vol. 21, no. 6, pp. S117–S130, 2005.

[18] Y. E. Nesterov, “A method for solving the convex programming problem with convergence rate $O(1/k^2)$,” *Dokl. Akad. Nauk SSSR*, vol. 269, pp. 543–547, 1983, (in Russian).

[19] J. Sharpe, U. Ahlgren, P. Perry, B. Hill, A. Ross, J. Hecksher-Sørensen, R. Baldock, and D. Davidson, “Optical projection tomography as a tool for 3D microscopy and gene expression studies,” *Science*, vol. 296, no. 5567, pp. 541–545, April 2002.

[20] W. Choi, C. Fang-Yen, K. Badizadegan, S. Oh, N. Lue, R. R. Dasari, and M. S. Feld, “Tomographic phase microscopy,” *Nat. Methods*, vol. 4, no. 9, pp. 717–719, September 2007.

[21] V. Lauer, “New approach to optical diffraction tomography yielding a vector equation of diffraction tomography and a novel tomographic microscope,” *J. Microsc.*, vol. 205, no. 2, pp. 165–176, 2002.

[22] Y. Sung, W. Choi, C. Fang-Yen, K. Badizadegan, R. R. Dasari, and M. S. Feld, “Optical diffraction tomography for high resolution live cell imaging,” *Opt. Express*, vol. 17, no. 1, pp. 266–277, December 2009.

[23] T. Kim, R. Zhou, M. Mir, S. Babacan, P. Carney, L. Goddard, and G. Popescu, “White-light diffraction tomography of unlabelled live cells,” *Nat. Photonics*, vol. 8, pp. 256–263, March 2014.

[24] T. S. Ralston, D. L. Marks, P. S. Carney, and S. A. Boppart, “Inverse scattering for optical coherence tomography,” *J. Opt. Soc. Am. A*, vol. 23, no. 5, pp. 1027–1037, May 2006.

[25] B. J. Davis, S. C. Schlachter, D. L. Marks, T. S. Ralston, S. A. Boppart, and P. S. Carney, “Nonparaxial vector-field modeling of optical coherence tomography and interferometric synthetic aperture microscopy,” *J. Opt. Soc. Am. A*, vol. 24, no. 9, pp. 2527–2542, September 2007.

[26] D. J. Brady, K. Choi, D. L. Marks, R. Horisaki, and S. Lim, "Compressive holography," *Opt. Express*, vol. 17, no. 15, pp. 13 040–13 049, 2009.

[27] L. Tian, N. Loomis, J. A. Dominguez-Caballero, and G. Barbastathis, "Quantitative measurement of size and three-dimensional position of fast-moving bubbles in airwater mixture flows using digital holography," *Appl. Opt.*, vol. 49, no. 9, pp. 1549–1554, March 2010.

[28] W. Chen, L. Tian, S. Rehman, Z. Zhang, H. P. Lee, and G. Barbastathis, "Empirical concentration bounds for compressive holographic bubble imaging based on a Mie scattering model," *Opt. E*, vol. 23, no. 4, p. February, 2015.

[29] H. M. Jol, Ed., *Ground Penetrating Radar: Theory and Applications*. Amsterdam: Elsevier, 2009.

[30] M. Leigsnering, F. Ahmad, M. Amin, and A. Zoubir, "Multipath exploitation in through-the-wall radar imaging using sparse reconstruction," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 2, pp. 920–939, April 2014.

[31] D. Liu, U. S. Kamilov, and P. T. Boufounos, "Compressive tomographic radar imaging with total variation regularization," in *Proc. IEEE 4th International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar, and Remote Sensing (CoSeRa 2016)*, Aachen, Germany, September 19-22, 2016, pp. 120–123.

[32] A. G. Tijhuis, "Born-type reconstruction of material parameters of an inhomogeneous, lossy dielectric slab from reflected-field data," *Wave Motion*, vol. 11, no. 2, pp. 151–173, May 1989.

[33] Y. M. Wang and W. Chew, "An iterative solution of the two-dimensional electromagnetic inverse scattering problem," *Int. J. Imag. Syst Tech.*, vol. 1, pp. 100–108, 1989.

[34] W. C. Chew and Y. M. Wang, "Reconstruction of two-dimensional permittivity distribution using the distorted Born iterative method," *IEEE Trans. Med. Imag.*, vol. 9, no. 2, pp. 218–225, June 1990.

[35] R. E. Kleinman, G. F. Roach, and P. M. van den Berg, "Convergent Born series for large refractive indices," *J. Opt. Soc. Am. A*, vol. 7, no. 5, pp. 890–897, May 1990.

[36] O. Haeberlé, K. Belkebir, H. Giovaninni, and A. Sentenac, "Tomographic diffractive microscopy: basic, techniques, and perspectives," *J. Mod. Opt.*, vol. 57, no. 9, pp. 686–699, May 2010.

[37] E. Mudry, P. C. Chaumet, K. Belkebir, and A. Sentenac, "Electromagnetic wave imaging of three-dimensional targets using a hybrid iterative inversion method," *Inv. Probl.*, vol. 28, no. 6, p. 065007, April 2012.

[38] D. Jin, R. Zhou, Z. Yaqoob, and P. So, "Tomographic phase microscopy: Principles and applications in bioimaging," *J. Opt. Soc. Am. B*, vol. 34, no. 5, pp. B64–B77, 2017.

[39] P. C. Chaumet and K. Belkebir, "Three-dimensional reconstruction from real data using a conjugate gradient-coupled dipole method," *Inv. Probl.*, vol. 25, no. 2, p. 024003, 2009.

[40] K. Belkebir, P. C. Chaumet, and A. Sentenac, "Superresolution in total internal reflection tomography," *J. Opt. Soc. Am. A*, vol. 22, no. 9, pp. 1889–1897, September 2005.

[41] A. Dubois, K. Belkebir, and M. Saillard, "Retrieval of inhomogeneous targets from experimental frequency diversity data," *Inv. Probl.*, vol. 21, no. 6, pp. S65–S81, 2005.

[42] U. S. Kamilov, D. Liu, H. Mansour, and P. T. Boufounos, "A recursive Born approach to nonlinear inverse scattering," *IEEE Signal Process. Lett.*, vol. 23, no. 8, pp. 1052–1056, August 2016.

[43] M. Charnotskii, "Extended Huygens-Fresnel principle and optical waves propagation in turbulence: discussion," *J. Opt. Soc. Am. A*, vol. 32, no. 7, pp. 1357–1365, July 2015.

[44] A. K. Singh, D. N. Naik, G. Pedrini, M. Takeda, and W. Osten, "Exploiting scattering media for exploring 3D objects," *Light Sci. Appl.*, vol. 6, no. 2, pp. 1–7, February 2017.

[45] H.-Y. Liu, E. Jonas, L. Tian, J. Zhong, B. Recht, and L. Waller, "3D imaging in volumetric scattering media using phase-space measurements," *Opt. Express*, vol. 23, no. 11, pp. 14 461–14 471, June 2015.

[46] T. Zhang, C. Godavarthi, P. C. Chaumet, G. Maire, H. Giovannini, A. Talneau, M. Allain, K. Belkebir, and A. Sentenac, "Far-field diffraction microscopy at $\lambda/10$ resolution," *Optica*, vol. 3, no. 6, pp. 609–612, June 2016.

[47] F. Bürgel, K. S. Kazimierski, and A. Lechleiter, "A sparsity regularization and total variation based computational framework for the inverse medium problem in scattering," *J. Comput. Physics*, vol. 339, pp. 1–30, 2017.

[48] U. S. Kamilov, I. N. Papadopoulos, M. H. Shoreh, A. Goy, C. Vonesch, M. Unser, and D. Psaltis, "Optical tomographic image reconstruction based on beam propagation and sparse regularization," *IEEE Trans. Comp. Imag.*, vol. 2, no. 1, pp. 59–70,, March 2016.

[49] H.-Y. Liu, U. S. Kamilov, D. Liu, H. Mansour, and P. T. Boufounos, "Compressive imaging with iterative forward models," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP 2017)*, New Orleans, LA, USA, March 5-9, 2017, pp. 6025–6029.

[50] A. I. Kirkland, W. O. Saxton, and G. Chand, "Multi beam tilt microscopy for super resolved imaging," *J. Electron Microsc. (Tokyo)*, vol. 46, no. 11, pp. 11–22, January 1997.

[51] A. M. Maiden, M. J. Humphry, and J. M. Rodenburg, "Ptychographic transmission microscopy in three dimensions using a multi-slice approach," *J. Opt. Soc. Am. A.*, vol. 29, no. 8, pp. 1606–1614, August 2012.

[52] L. Tian and L. Waller, "3D intensity and phase imaging from light field measurements in an LED array microscope," *Optica*, vol. 2, pp. 104–111, 2015.

[53] U. S. Kamilov, I. N. Papadopoulos, M. H. Shoreh, A. Goy, C. Vonesch, M. Unser, and D. Psaltis, "Learning approach to optical tomography," *Optica*, vol. 2, no. 6, pp. 517–522, June 2015.

[54] M. Born and E. Wolf, *Principles of Optics*, 7th ed. Cambridge Univ. Press, 2003, ch. Scattering from inhomogeneous media, pp. 695–734.

[55] T. C. Ye and S. Y. Lee, "Non-iterative exact inverse scattering using simultaneous orthogonal matching pursuit (s-omp)," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2008, pp. 2457–2460.

[56] A. C. Fannjiang, "Compressive inverse scattering: I. high-frequency simo/miso and mimo measurements," *Inverse Problems*, vol. 26, no. 3, p. 035008, 2010. [Online]. Available: http://stacks.iop.org/0266-5611/26/i=3/a=035008

[57] O. Lee, J. M. Kim, Y. Bresler, and J. C. Ye, "Compressive diffuse optical tomography: Noniterative exact reconstruction using joint sparsity," *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1129–1142, May 2011.

[58] O. Lee and J. C. Ye, "Joint sparsity-driven non-iterative simultaneous reconstruction of absorption and scattering in diffuse optical tomography," *Opt. Express*, vol. 21, no. 22, pp. 26 589–26 604, Nov 2013. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-21-22-26589

[59] O. K. Lee, H. Kang, J. C. Ye, and M. Lim, "A non-iterative method for the electrical impedance tomography based on joint sparse recovery," *Inverse Problems*, vol. 31, no. 7, p. 075002, 2015. [Online]. Available: http://stacks.iop.org/0266-5611/31/i=7/a=075002

[60] J. Yoo, Y. Jung, M. Lim, J. C. Ye, and A. Wahab, "A joint sparse recovery framework for accurate reconstruction of inclusions in elastic media," *SIAM Journal on Imaging Sciences*, vol. 10, no. 3, pp. 1104–1138, 2017. [Online]. Available: https://doi.org/10.1137/16M110318X

[61] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, 2005.

[62] J. Virieux and S. Operto, "An overview of full-waveform inversion in exploration geophysics," *Geophyscs*, vol. 74, no. 6, pp. WCC127–WCC152, November-December 2009.

[63] F. J. Herrmann and T. van Leeuwen, "Mitigating local minima in full-waveform inversion by expanding the search space," *Geophys. J. Int.*, vol. 195, no. 1, pp. 661–667, 2013.

[64] N. K. Nikolova, J. Zhu, D. Li, M. H. Bakr, and J. W. Bandler, "Sensitivity analysis of network parameters with electromagnetic frequency-domain simulators," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 2, pp. 670–681, Feb 2006.

[65] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[66] A. Abubakar, P. M. van den Berg, and T. M. Habashy, "Application of the multiplicative regularized contrast source inversion method tm- and te-polarized experimental fresnel data," *Inv. Probl.*, vol. 21, no. 6, pp. S5–S14, 2005.

[67] A. Baussard, "Inversion of multi-frequency experimental data using an adaptive multiscale approach," *Inv. Probl.*, vol. 21, no. 6, pp. S15–S32, 2005.

[68] L. Crocco, M. D'Urso, and T. Isernia, "Testing the contrast source extended born inversion method against real data: the tm case," *Inv*, vol. 21, no. 6, pp. S33–S51, 2005.

[69] O. Féron, B. Duchêne, and A. Mohammad-Djafari, "Microwave imaging of inhomogeneous objects made of a finite number of dielectric and conductive materials from experimental data," *Inv. Probl.*, vol. 21, no. 6, pp. S95–S117, 2005.

[70] S. Silver, "Microwave aperture antennas and diffraction theory," *J. Opt. Soc. Am.*, vol. 52, no. 2, pp. 131–139, February 1962.

[71] M. Born and E. Wolf, *Principles of Optics*, 7th ed. Cambridge Univ. Press, 2003, ch. Transition to a Scalar Theory, pp. 430–436.

[72] J. W. Goodman, *Introduction to Fourier Optics*. McGraw-Hill, 1996, ch. Foundations of Scalar Diffraction Theory, pp. 32–62.
[73] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers, 2004.
[74] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, 1995.
[75] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 906–916, August 2003.
[76] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, November 2004.
[77] J. Bect, L. Blanc-Feraud, G. Aubert, and A. Chambolle, "A ℓ_1-unified variational framework for image restoration," in *Proc. ECCV*, Springer, Ed., vol. 3024, New York, 2004, pp. 1–13.
[78] U. S. Kamilov, "A parallel proximal algorithm for anisotropic total variation minimization," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 539–548, February 2017.
[79] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 28, 2015.
[80] D. P. Bertsekas, "Incremental proximal methods for large scale convex optimization," *Math. Program. Ser. B*, vol. 129, pp. 163–195, 2011.
[81] L. Bottou, *Neural Networks: Tricks of the Trade*, 2nd ed. Springer, September 2012, ch. Stochastic Gradient Descent Tricks, pp. 421–437.
[82] J. Mairal, F. Bach, and J. Ponce, "Sparse modeling for image and vision processing," *Foundations and Trends in Machine Learning*, vol. 8, no. 2-3, pp. 1–199, 2014.
[83] L. Bottou, F. E. Curtis, and J. Nocedal, "Optimization methods for large-scale machine learning," June 2016, arXiv:1606.04838 [stat.ML].
[84] J.-M. Geffrin and P. Sabouroux, "Continuing with the Fresnel database: experimental setup and improvements in 3D scattering measurements," *Inv. Probl.*, vol. 25, no. 2, p. 024001, 2009.
[85] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. Wiley, 1999.
[86] C. Sheppard and C. J. Cogswell, "Three-dimensional image formation in confocal microscopy," *J. Microsc.*, vol. 159, no. 2, pp. 179–194, August 1990.