MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

# Accelerated Image Reconstruction for Nonlinear Diffractive Imaging

Ma, Y.; Mansour, H.; Liu, D.; Boufounos, P.T.; Kamilov, U.

TR2018-008 April 15, 2018

**Abstract**

The radar autofocus problem arises in situations where radar measurements are acquired of a scene using antennas that suffer from position ambiguity. Current techniques model the antenna ambiguity as a global phase error affecting the received radar measurement at every antenna. However, the phase error signal model is only valid in the far field regime where the position error can be approximated by a one dimensional shift in the down-range direction. We propose in this paper an alternate formulation where the antenna position error is modeled using a two-dimensional shift operator in the imagedomain. The radar autofocus problem then becomes a multichannel two-dimensional blind deconvolution problem where the static radar image is convolved with a two dimensional shift kernel for each antenna measurement. We develop an alternating minimization framework that leverages the sparsity and piece-wise smoothness of the radar scene, as well as the one-sparse property of the two dimensional shift kernels.

*IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*

# ACCELERATED IMAGE RECONSTRUCTION FOR NONLINEAR DIFFRACTIVE IMAGING

*Yanting Ma*[⋆], *Hassan Mansour*[†], *Dehong Liu*[†], *Petros T. Boufounos*[†], *and Ulugbek S. Kamilov*[♭]

[⋆] Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27606
[†] Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139, USA
[♭] Department of Computer Science and Engineering, Washington University in St. Louis, St. Louis, MO 63130

**ABSTRACT**

The problem of reconstructing an object from the measurements of the light it scatters is common in numerous imaging applications. While the most popular formulations of the problem are based on linearizing the object-light relationship, there is an increased interest in considering nonlinear formulations that can account for multiple light scattering. In this paper, we propose an image reconstruction method, called CISOR, for nonlinear diffractive imaging, based on our new variant of fast iterative shrinkage/thresholding algorithm (FISTA) and total variation (TV) regularization. We prove that CISOR reliably converges for our nonconvex optimization problem, and systematically compare our method with other state-of-the-art methods on simulated as well as experimentally measured data.

***Index Terms*—** Diffraction tomography, proximal gradient method, total variation regularization, nonconvex optimization

## 1. INTRODUCTION

Estimation of the spatial permittivity distribution of an object from the scattered wave measurements is ubiquitous in numerous applications. Although the classical linear scattering models such as the first Born approximation [1] and the Rytov approximation [2] can be solved by comparatively simple inverse algorithms, such models are highly inaccurate when the physical size of the object is large or the permittivity contrast of the object compared to the background is high [3]. In order to be able to reconstruct strongly scattering objects, nonlinear formulations that can model multiple scattering need to be considered. Recent work has been trying to integrate the nonlinearity and design new inverse algorithms to reconstruct the object. Examples of nonlinear methods include iterative linearization [4, 5], contrast source inversion [6–8], hybrid methods [9–11], and optimization with error backpropagation [12–16].

A standard way for solving inverse problems is via optimization. The cost function usually consists of a smooth data-fidelity term and a non-smooth regularization term whose proximal mapping is easily computed. For such cost functions, the proximal gradient method ISTA [17–19] or its accelerated variant FISTA [20] can be applied. Theoretical convergence analysis of FISTA is well-understood for convex problems, whereas no convergence guarantee is known in nonconvex cases. A variant of FISTA has been proposed in [21] for nonconvex optimization with convergence guarantees. This algorithm computes two estimates from ISTA and FISTA, respectively, at each iteration, and selects the one with lower cost function value as the final estimate at the current iteration. Therefore, both the gradient and the cost function need to be evaluated at two different points at each iteration. While such extra computation may be insignificant in some applications, it can be prohibitive in the inverse scattering problem, where additional evaluations of the gradient and the cost function require the computation of the entire forward model.

In this work, we propose a new image reconstruction method called *Convergent Inverse Scattering using Optimization and Regularization (CISOR)*. CISOR is based on our novel nonconvex optimization formulation that can account for multiple scattering, while enabling fast computation of the gradient of the cost functional. Additionally, CISOR relies on our new relaxed variant of FISTA for nonconvex optimization problems with convergence guarantees that we establish here. Our new FISTA variant may be of interest on its own as a general nonconvex solver. While we were concluding this manuscript, we became aware of very recent related work in [22], which considered a similar nonconvex formulation as in this paper. However, the work in [22] uses FISTA to solve the nonconvex problem and does not have theoretical convergence analysis.

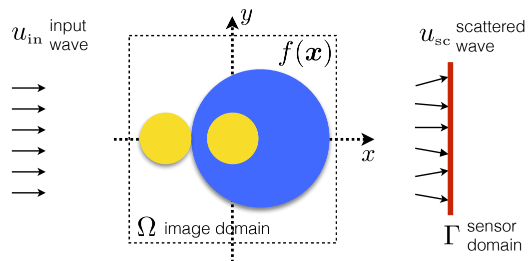


**Fig. 1**: The measurement scenario considered in this paper.

## 2. PROBLEM FORMULATION

The problem of inverse scattering is described as follows and illustrated in Figure 1. Suppose that an object is placed within a bounded domain $\Omega \subset \mathbb{R}^2$. The object is illuminated by an incident wave $u_{\text{in}}$, and the scattered wave $u_{\text{sc}}$ is measured by the sensors placed in a sensing region $\Gamma \subset \mathbb{R}^2$. Let $u$ denote the total field, which satisfies $u(\boldsymbol{x}) = u_{\text{in}}(\boldsymbol{x}) + u_{\text{sc}}(\boldsymbol{x}), \forall \boldsymbol{x} \in \mathbb{R}^2$. The scalar Lippmann-Schwinger equation [1] establishes the fundamental object-wave relationship

$$u(\boldsymbol{x}) = u_{\text{in}}(\boldsymbol{x}) + \int_{\Omega} g(\boldsymbol{x} - \boldsymbol{x}')u(\boldsymbol{x}')f(\boldsymbol{x}')d\boldsymbol{x}', \quad \forall \boldsymbol{x} \in \mathbb{R}^2.$$

In the above, $f(\boldsymbol{x}) = k^2(\epsilon(\boldsymbol{x}) - \epsilon_b)$ is the scattering potential, where $\epsilon(\boldsymbol{x})$ is the permittivity of the object, $\epsilon_b$ is the permittivity of the

This work was completed while Y. Ma and U. S. Kamilov (email: kamilov@wustl.edu) were with MERL.

background, and $k=2\pi/\lambda$ is the wavenumber in vacuum. The free-space Green's function in 2D is defined as $g(\boldsymbol{x})=-\frac{j}{4}H_0^{(1)}(k_b\|\boldsymbol{x}\|)$, where $H_0^{(1)}$ is the Hankel function of first kind, $k_b=k\sqrt{\epsilon_b}$ is the wavenumber of the background medium, and $\|\cdot\|$ denotes the $\ell_2$ norm. The discrete system is then

$$\begin{aligned}\mathbf{y}&=\mathbf{H}\mathrm{diag}(\mathbf{f})\mathbf{u}+\mathbf{e}\\ \mathbf{u}&=\mathbf{u}_{\text{in}}+\mathbf{G}\mathrm{diag}(\mathbf{f})\mathbf{u},\end{aligned} \tag{1}$$

where $\mathbf{f}\in\mathbb{R}^N$, $\mathbf{u}\in\mathbb{C}^N$, $\mathbf{u}_{\text{in}}\in\mathbb{C}^N$ are $N$ uniformly spaced samples of $f(\boldsymbol{x})$, $u(\boldsymbol{x})$, and $u_{\text{in}}(\boldsymbol{x})$ on $\Omega$, respectively, $\mathrm{diag}(\mathbf{f})$ is a diagonal matrix with $\mathbf{f}$ on its diagonal, and $\mathbf{y}\in\mathbb{C}^M$ is the measured scattered wave at the sensors with measurement error $\mathbf{e}\in\mathbb{C}^M$. The matrix $\mathbf{H}\in\mathbb{C}^{M\times N}$ is the discretization of Green's function $g(\boldsymbol{x}-\boldsymbol{x}')$ with $\boldsymbol{x}\in\Gamma$ and $\boldsymbol{x}'\in\Omega$, whereas $\mathbf{G}\in\mathbb{C}^{N\times N}$ is the discretization of Green's function with $\boldsymbol{x},\boldsymbol{x}'\in\Omega$. The nonlinear inverse scattering problem is then to estimate $\mathbf{f}$ in (1) given $\mathbf{y}$, $\mathbf{H}$, $\mathbf{G}$, and $\mathbf{u}_{\text{in}}$.

## 3. PROPOSED METHOD

Our proposed method is based on a nonconvex optimization formulation with total variation regularization. Let $\mathbf{A}:=\mathbf{I}-\mathbf{G}\mathrm{diag}(\mathbf{f})$ and $\mathcal{Z}(\mathbf{f}):=\mathbf{H}\mathrm{diag}(\mathbf{f})\mathbf{u}$. Moreover, let $\mathcal{C}\subset\mathbb{R}^N$ be a set that contains all possible values that $\mathbf{f}$ can take, and we assume there exists a constant $M>0$ such that $\|\mathbf{f}\|\leq M, \forall\mathbf{f}\in\mathcal{C}$. We estimate $\mathbf{f}$ from (1) by solving the following optimization problem:

$$\widehat{\mathbf{f}}=\arg\min_{\mathbf{f}\in\mathbb{R}^N}\mathcal{F}(\mathbf{f}):=\mathcal{D}(\mathbf{f})+\mathcal{R}(\mathbf{f}), \tag{2}$$

with

$$\mathcal{D}(\mathbf{f})=\frac{1}{2}\|\mathbf{y}-\mathcal{Z}(\mathbf{f})\|^2, \tag{3}$$

$$\mathcal{R}(\mathbf{f})=\tau\sum_{n=1}^{N}\sqrt{\sum_{d=1}^{2}|[\mathbf{D}_d\mathbf{f}]_n|^2}+\chi_{\mathcal{C}}(\mathbf{f}), \tag{4}$$

where $\mathbf{D}_d$ is the discrete gradient operator in the $d$th dimension, hence the first term in (4) is the total variation (TV) regularizer. The function $\chi_{\mathcal{C}}(\mathbf{f})=0$ if $\mathbf{f}\in\mathcal{C}$ and $\chi_{\mathcal{C}}(\mathbf{f})=\infty$ otherwise.

### 3.1. Relaxed FISTA

We now propose a new variant of FISTA to solve (2) and provide its theoretical convergence guarantee. Starting with some initialization $\mathbf{f}_0\in\mathbb{R}^N$ and setting $\mathbf{s}_1=\mathbf{f}_0$, $t_0=1$, $\alpha\in[0,1)$, for $k\geq 1$, the proposed algorithm proceeds as follows:

$$\mathbf{f}_k=\mathrm{prox}_{\gamma\mathcal{R}}\left(\mathbf{s}_k-\gamma\nabla\mathcal{D}(\mathbf{s}_k)\right) \tag{5}$$

$$t_{k+1}=\frac{\sqrt{4t_k^2+1}+1}{2} \tag{6}$$

$$\mathbf{s}_{k+1}=\mathbf{f}_k+\alpha\left(\frac{t_k-1}{t_{k+1}}\right)(\mathbf{f}_k-\mathbf{f}_{k-1}), \tag{7}$$

where the choice of the step-size $\gamma$ to ensure convergence will be discussed in Section 3.2. Notice that the algorithm (5)-(7) is equivalent to ISTA when $\alpha=0$ and is equivalent to FISTA when $\alpha=1$. For this reason, we call it relaxed FISTA. Figure 2 shows that the empirical convergence speed of relaxed FISTA improves as $\alpha$ increases from 0 to 1. The plot was obtained by using the experimentally measured scattered microwave data collected by the Fresnel Institute [23]. Our

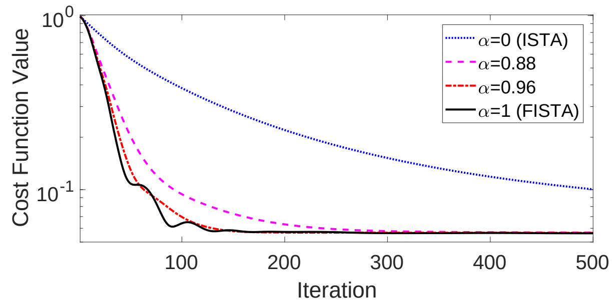


**Fig. 2**: Empirical convergence speed for relaxed FISTA with various $\alpha$ values tested on experimentally measured data.

theoretical analysis of relaxed FISTA in Section 3.2 establishes convergence for any $\alpha\in[0,1)$.

The two main elements of relaxed FISTA are the computation of the gradient $\nabla\mathcal{D}$ and of the proximal mapping $\mathrm{prox}_{\gamma\mathcal{R}}$. Given $\nabla\mathcal{D}(\mathbf{s}_k)$, the proximal mapping (5) can be efficiently solved [24,25]. The following proposition provides an explicit formula for $\nabla\mathcal{D}$.

**Proposition 1.** Define $\mathbf{r}:=\mathbf{H}\mathrm{diag}(\mathbf{f})\mathbf{u}-\mathbf{y}$. Then we have

$$\nabla\mathcal{D}(\mathbf{f})=\mathsf{Re}\left\{\mathrm{diag}(\mathbf{u})^{\mathsf{H}}\left(\mathbf{H}^{\mathsf{H}}\mathbf{r}+\mathbf{G}^{\mathsf{H}}\mathbf{v}\right)\right\}, \tag{8}$$

where $\mathbf{u}$ and $\mathbf{v}$ are obtained from the linear systems

$$\mathbf{A}\mathbf{u}=\mathbf{u}_{\text{in}},\quad \text{and}\quad \mathbf{A}^{\mathsf{H}}\mathbf{v}=\mathrm{diag}(\mathbf{f})\mathbf{H}^{\mathsf{H}}\mathbf{r}. \tag{9}$$

*Proof.* See Appendix 5.1. □

Note that in the above, $\mathbf{u}$ and $\mathbf{v}$ can be efficiently solved by conjugate gradient. In our implementation, $\mathbf{A}$ is an operator rather than an explicit matrix, and the convolution with the Green's function is computed using the fast Fourier transform (FFT) algorithm.

### 3.2. Convergence Analysis

The following proposition shows that the data-fidelity term (3) has Lipschitz gradient on a bounded domain, which is essential to prove the convergence of relaxed FISTA.

**Proposition 2.** Suppose that $\mathcal{U}\subset\mathbb{R}^N$ is bounded. Assume that $\|\mathbf{u}_{\text{in}}\|<\infty$ and the matrix $\mathbf{A}=\mathbf{I}-\mathbf{G}\mathrm{diag}(\mathbf{s})$ is non-singular for all $\mathbf{s}\in\mathcal{U}$. Then $\mathcal{D}(\mathbf{s})$ has Lipschitz gradient on $\mathcal{U}$. That is, there exists an $L\in(0,\infty)$ such that

$$\|\nabla\mathcal{D}(\mathbf{s}_1)-\nabla\mathcal{D}(\mathbf{s}_2)\|\leq L\|\mathbf{s}_1-\mathbf{s}_2\|,\quad\forall\mathbf{s}_1,\mathbf{s}_2\in\mathcal{U}. \tag{10}$$

*Proof.* See Appendix 5.2. □

Notice that all $\mathbf{f}_k$ obtained from (5) are within a bounded set $\mathcal{C}$, and each $\mathbf{s}_{k+1}$ obtained from (7) is a linear combination of $\mathbf{f}_k$ and $\mathbf{f}_{k-1}$, where the weight $\alpha\left(\frac{t_k-1}{t_{k+1}}\right)\in[0,1)$ since $\alpha\in[0,1)$ and $\frac{t_k-1}{t_{k+1}}\leq 1$ by (6). Hence, the set that covers all possible values for $\{\mathbf{f}_k\}_{k\geq 0}$ and $\{\mathbf{s}_k\}_{k\geq 1}$ is bounded. Using this fact, we have the following convergence guarantee for relaxed FISTA.

**Proposition 3.** Let $\mathcal{U}$ in Proposition 2 be the set that covers all possible values for $\{\mathbf{f}_k\}_{k\geq 0}$ and $\{\mathbf{s}_k\}_{k\geq 1}$ obtained from (5) and (7), $L$ be the corresponding Lipschitz constant defined in (10). Choose $\gamma\leq\frac{1-\alpha^2}{2L}$ for any fixed $\alpha\in[0,1)$. Define the gradient mapping as

$$\mathcal{G}_\gamma(\mathbf{s}):=\frac{\mathbf{s}-\mathrm{prox}_{\gamma\mathcal{R}}\left(\mathbf{s}-\gamma\nabla\mathcal{D}(\mathbf{s})\right)}{\gamma}. \tag{11}$$

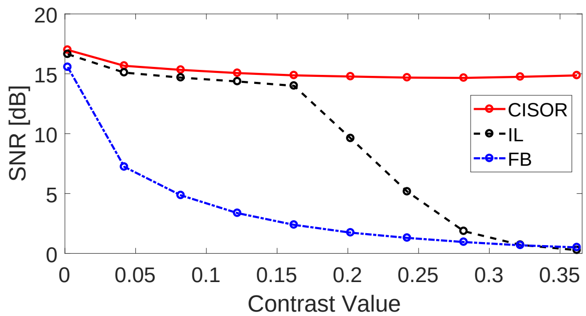


**Fig. 3**: Comparison of different reconstruction methods for various contrast levels tested on simulated data.

Then, relaxed FISTA converges to a stationary point in the sense that the gradient mapping norm satisfies

$$\lim_{k\to\infty} \|\mathcal{G}_\gamma(\mathbf{s}_k)\| = 0. \tag{12}$$

*Proof.* See Appendix 5.3. □

Note that in practice, one can use backtracking line search to determine the value of $\gamma$ when $L$ is not known explicitly. Define $\widehat{\mathbf{s}} := \lim_{k\to\infty} \mathbf{s}_k$. We notice that $\mathcal{G}_\gamma(\widehat{\mathbf{s}}) = 0$ implies $0 \in \partial\mathcal{F}(\widehat{\mathbf{s}})$, where $\partial\mathcal{F}$ denotes the limiting subdifferential of $\mathcal{F}$ [26]. Hence, $\widehat{\mathbf{s}}$ is a stationary point of $\mathcal{F}$. Moreover, by (5) and (11), we have $\mathcal{G}_\gamma(\mathbf{s}_k) = \frac{1}{\gamma}(\mathbf{s}_k - \mathbf{f}_k)$. Therefore, (12) implies that $\lim_{k\to\infty} \|\mathbf{s}_k - \mathbf{f}_k\| = 0$, thus $\lim_{k\to\infty} \mathbf{f}_k = \lim_{k\to\infty} \mathbf{s}_k = \widehat{\mathbf{s}}$. This establishes that the sequence $\{\mathbf{f}_k\}_{k\geq 0}$ generated by relaxed FISTA converges to a stationary point of the nonconvex problem (2).

## 4. EXPERIMENTAL RESULTS

We compare CISOR with state-of-the-art methods, iterative linearization (IL) [4, 5], contrast sourse inversion (CSI) [6–8], and SEAGLE [15], as well as a linear method, the first Born approximation (FB) [1]. The proximal operator of TV in (5) is implemented following [24]. CISOR uses the relaxed FISTA defined in Section3.1 with $\alpha = 0.96$ and fixed step-size $\gamma$, which is manually tuned. The other methods use the standard FISTA, also with manually tuned $\gamma$.

**Comparison on simulated data.** Figure 3 shows the performance of three algorithms on the simulated data using objects with various contrast values. The contrast of an object $\mathbf{f}$ is defined as $\max(|\mathbf{f}|)/k_b^2$. We consider the Shepp-Logan phantom and change its contrast to the desired value to obtain the ground-truth $\mathbf{f}_{\text{true}}$. We then solve the Lippmann-Schwinger equation to generate the scattered waves that are then used as measurements. The center of the image is the origin and the physical size of the image is set to 120 cm $\times$ 120 cm. Two linear detectors are placed on two opposite sides of the image at a distance of 95.9 cm from the origin. Each detector has 169 sensors with a spacing of 3.84 cm. The transmitters are placed on a line 48.0 cm left to the left detector, and they are spaced uniformly in azimuth with respect to the origin within a range of $[-60^\circ, 60^\circ]$ at every $5^\circ$. The wavelength of the incident wave is 7.49 cm and the pixel size is 0.94 cm. The reconstructed SNR, which is defined as $20\log_{10}(\|\mathbf{f}_{\text{true}}\|/\|\hat{\mathbf{f}} - \mathbf{f}_{\text{true}}\|)$, is used as the comparison criterion. For each contrast value and each algorithm, we run the algorithm with five different regularization parameter values and select the result that yields the highest reconstructed SNR. Figure 3 shows that as the contrast increases, the reconstructed SNR of FB and IL decreases, whereas that of CISOR is more stable.

**Comparison on experimental data.** We use two objects from the public dataset provided by the Fresnel Institute [23]: *FoamDielExtTM* and *FoamDielintTM*. The objects are placed within a 15 cm $\times$ 15 cm square region centered at the origin of the coordinate system. The number of transmitters is 8 and the number of receivers is 360 for all objects. The transmitters and the receivers are placed on a circle centered at the origin with radius 1.67 m and are spaced uniformly in azimuth. Only one transmitter is turn on at a time and only 241 receivers are active for each transmitter. That is, the 119 receivers that are closest to a transmitter are inactive for that transmitter. While the dataset contains multiple frequency measurements, we only use the ones corresponding to 3 GHz, hence the wavelength of the incident wave is 9.99 cm. The pixel size of the reconstructed images is 0.12 cm.

Figure 4 provides a visual comparison of the reconstructed images obtained by different algorithms. For each object and each algorithm, we run the algorithm with five different regularization parameter values and select the result that has the best visual quality. Figure 4 shows that all nonlinear methods CISOR, SEAGLE, IL, and CSI obtained reasonable reconstruction results in terms of both the contrast value and the shape of the object, whereas the linear method FB significantly underestimated the contrast value and failed to capture the shape. These results demonstrate that the proposed method is competitive with several state-of-the-art methods. Two key advantages of CISOR over other methods are in its memory efficiency and convergence guarantees.

## 5. APPENDIX

### 5.1. Proof for Proposition 1

The gradient of $\mathcal{D}(\cdot)$ is $\nabla\mathcal{D}(\mathbf{f}) = \text{Re}\{\mathbf{J}_{\mathcal{Z}}^{\mathsf{H}}\mathbf{r}\}$, where $\mathbf{J}_{\mathcal{Z}}$ is the Jocobian matrix of $\mathcal{Z}(\mathbf{f}) := \mathbf{H}\text{diag}(\mathbf{f})\mathbf{u}$. Recall that $\mathbf{A} = (\mathbf{I} - \mathbf{G}\text{diag}(\mathbf{f}))$ and $\mathbf{u} = \mathbf{A}^{-1}\mathbf{u}_{\text{in}}$, hence both $\mathbf{A}$ and $\mathbf{u}$ are functions of $\mathbf{f}$ and we write $\mathbf{u}(\mathbf{f})$ and $\mathbf{A}(\mathbf{f})$ to emphasize the dependencies. Following the chain rule of differentiation, we have

$$\frac{\partial \mathcal{Z}_m}{\partial f_n} = H_{m,n} u_n(\mathbf{f}) + \sum_{i=1}^{N} \left[\frac{\partial u_i(\mathbf{f})}{\partial f_n}\right] H_{m,i} f_i.$$

Using the definition $\mathbf{r} = \mathcal{Z}(\mathbf{f}) - \mathbf{y}$ and summing over $m = 1, \ldots, M$,

$$\begin{aligned}
[\nabla\mathcal{D}(\mathbf{f})]_n &= \sum_{m=1}^{M} \left[\frac{\partial \mathcal{Z}_m}{\partial f_n}\right]^* r_m \\
&= u_n^*(\mathbf{f}) \sum_{m=1}^{M} H_{m,n}^* r_m + \sum_{i=1}^{N} \left[\frac{\partial u_i(\mathbf{f})}{\partial f_n}\right]^* f_i \sum_{m=1}^{M} H_{m,i}^* r_m \\
&= u_n^*(\mathbf{f}) \left[\mathbf{H}^{\mathsf{H}}\mathbf{r}\right]_n + \sum_{i=1}^{N} \left[\frac{\partial u_i(\mathbf{f})}{\partial f_n}\right]^* f_i \left[\mathbf{H}^{\mathsf{H}}\mathbf{r}\right]_i,
\end{aligned} \tag{13}$$

where $a^*$ denotes the complex conjugate of $a \in \mathbb{C}$. Label the two terms in (13) as $T_1$ and $T_2$, then

$$T_1 = [\text{diag}(\mathbf{u}^*(\mathbf{f}))\mathbf{H}\mathbf{r}]_n, \tag{14}$$

$$\begin{aligned}
T_2 &\overset{(a)}{=} \mathbf{u}_{\text{in}}^{\mathsf{H}} \left[\frac{\partial \mathbf{A}^{-1}(\mathbf{f})}{\partial f_n}\right]^{\mathsf{H}} \text{diag}(\mathbf{f})\mathbf{H}^{\mathsf{H}}\mathbf{r} \\
&\overset{(b)}{=} -\mathbf{u}_{\text{in}}^{\mathsf{H}}\mathbf{A}^{-\mathsf{H}}(\mathbf{f}) \left[\frac{\partial \mathbf{A}(\mathbf{f})}{\partial f_n}\right]^{\mathsf{H}} \mathbf{A}^{-\mathsf{H}}(\mathbf{f})\text{diag}(\mathbf{f})\mathbf{H}^{\mathsf{H}}\mathbf{r} \\
&\overset{(c)}{=} -\mathbf{u}^{\mathsf{H}}(\mathbf{f}) \left[\frac{\partial \mathbf{A}(\mathbf{f})}{\partial f_n}\right]^{\mathsf{H}} \mathbf{v}(\mathbf{f}) \overset{(d)}{=} [\text{diag}(\mathbf{u}^*(\mathbf{f}))\mathbf{G}^{\mathsf{H}}\mathbf{v}(\mathbf{f})]_n.
\end{aligned} \tag{15}$$

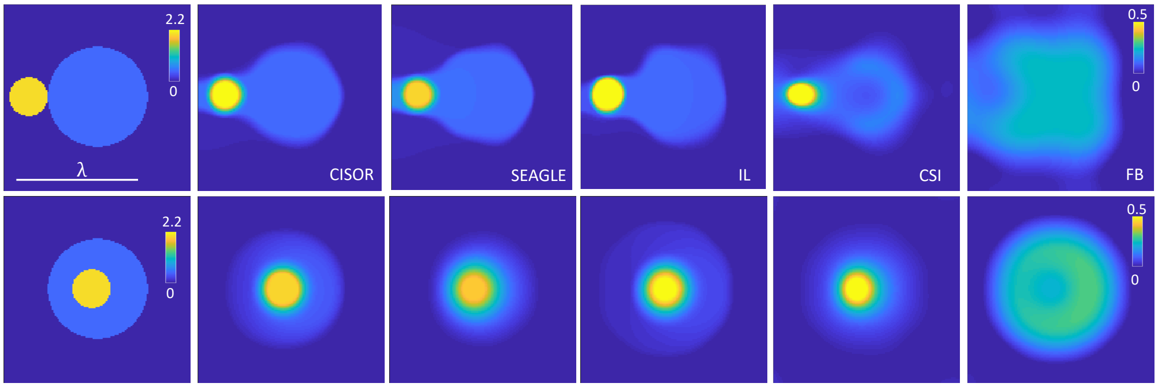


**Fig. 4**: Reconstructed images obtained by different algorithms from experimentally measured data. The first row uses *FoamDielExtTM*, and the second row uses *FoamDielIntTM*. From left to right: ground truth, reconstructed images by CISOR, SEAGLE, IL, CSI and FB. Note that the color-map used for FB is different from the rest, because FB significantly underestimated the contrast values.

In the above, step $(a)$ holds by plugging in $u_i(\mathbf{f}) = \left[\mathbf{A}^{-1}(\mathbf{f})\mathbf{u}_{\text{in}}\right]_i$. Step $(b)$ uses the identity

$$\frac{\partial \mathbf{A}^{-1}(\mathbf{f})}{\partial f_n} = -\mathbf{A}^{-1}(\mathbf{f})\frac{\partial \mathbf{A}(\mathbf{f})}{\partial f_n}\mathbf{A}^{-1}(\mathbf{f})$$

which follows by differentiating both sides of $\mathbf{A}(\mathbf{f})\mathbf{A}^{-1}(\mathbf{f}) = \mathbf{I}$,

$$\frac{\partial \mathbf{A}(\mathbf{f})}{\partial f_n}\mathbf{A}^{-1}(\mathbf{f}) + \mathbf{A}(\mathbf{f})\frac{\partial \mathbf{A}^{-1}(\mathbf{f})}{\partial f_n} = 0.$$

From step $(b)$ to step $(c)$, we used the fact that $\mathbf{u}(\mathbf{f}) = \mathbf{A}^{-1}(\mathbf{f})\mathbf{u}_{\text{in}}$ and defined $\mathbf{v}(\mathbf{f}) := \mathbf{A}^{-\mathsf{H}}(\mathbf{f})\text{diag}(\mathbf{f})\mathbf{H}^{\mathsf{H}}\mathbf{r}$, which matches (9). Finally, step $(d)$ follows by plugging in $\mathbf{A}(\mathbf{f}) = \mathbf{I} - \mathbf{G}\text{diag}(\mathbf{f})$. Combining (13), (14), and (15), we have obtained the expression in (8).

### 5.2. Proof for Proposition 2

Let $\mathbf{A}_i = \mathbf{I} - \mathbf{G}\text{diag}(\mathbf{s}_i)$, $\mathbf{u}_i = \mathbf{A}_i^{-1}\mathbf{u}_{\text{in}}$, $\mathbf{z}_i = \mathcal{Z}(\mathbf{s}_i)$, $\mathbf{r}_i = \mathbf{z}_i - \mathbf{y}$, and $\mathbf{v}_i = \mathbf{A}_i^{-\mathsf{H}}\mathbf{H}\text{diag}(\mathbf{s}_i)\mathbf{u}_i$ for $i = 1, 2$. Then,

$$\begin{aligned}\|\nabla\mathcal{D}(\mathbf{s}_1) - \nabla\mathcal{D}(\mathbf{s}_2)\| &\le \|\text{diag}(\mathbf{u}_1)^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\mathbf{r}_1 - \text{diag}(\mathbf{u}_2)^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\mathbf{r}_2\| \\ &+ \|\text{diag}(\mathbf{u}_1)^{\mathsf{H}}\mathbf{G}^{\mathsf{H}}\mathbf{v}_1 - \text{diag}(\mathbf{u}_2)^{\mathsf{H}}\mathbf{G}^{\mathsf{H}}\mathbf{v}_2\|.\end{aligned}$$

Label the two terms on the RHS as $T_1$ and $T_2$. We will prove $T_1 \le L_1\|\mathbf{s}_1 - \mathbf{s}_2\|$ for some constant $L_1 > 0$, then the proof for $T_2 \le L_2\|\mathbf{s}_1 - \mathbf{s}_2\|$ follows similarly.

$$\begin{aligned}T_1 &\le \|\text{diag}(\mathbf{u}_1)^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\mathbf{r}_1 - \text{diag}(\mathbf{u}_2)^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\mathbf{r}_1\| \\ &+ \|\text{diag}(\mathbf{u}_2)^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\mathbf{r}_1 - \text{diag}(\mathbf{u}_2)^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\mathbf{r}_2\| \\ &\le \|\mathbf{u}_1 - \mathbf{u}_2\|\|\mathbf{H}\|_{\text{op}}\|\mathbf{r}_1\| + \|\mathbf{A}_2^{-1}\|_{\text{op}}\|\mathbf{u}_{\text{in}}\|\|\mathbf{H}\|_{\text{op}}\|\mathbf{z}_1 - \mathbf{z}_2\|,\end{aligned}$$

where $\|\cdot\|_{\text{op}}$ denotes the operator norm and the last inequality uses the fact that for a diagonal matrix $\text{diag}(\mathbf{d})$, $\|\text{diag}(\mathbf{d})\|_{\text{op}} = \max_{n\in[N]}|d_n| \le \|\mathbf{d}\|$.

$$\begin{aligned}\|\mathbf{u}_1 - \mathbf{u}_2\| &\le \|\mathbf{A}_1^{-1}(\mathbf{A}_2 - \mathbf{A}_1)\mathbf{A}_2^{-1}\|\|\mathbf{u}_{\text{in}}\| \\ &\le \|\mathbf{A}_1^{-1}\|_{\text{op}}\|\mathbf{G}\|_{\text{op}}\|\mathbf{s}_1 - \mathbf{s}_2\|\|\mathbf{A}_2^{-1}\|_{\text{op}}\|\mathbf{u}_{\text{in}}\|, \\ \|\mathbf{z}_1 - \mathbf{z}_2\| &\le \|\mathbf{H}\text{diag}(\mathbf{s}_1)\mathbf{u}_1 - \mathbf{H}\text{diag}(\mathbf{s}_1)\mathbf{u}_2\| \\ &+ \|\mathbf{H}\text{diag}(\mathbf{s}_1)\mathbf{u}_2 - \mathbf{H}\text{diag}(\mathbf{s}_2)\mathbf{u}_2\| \\ &\le \|\mathbf{H}\|_{\text{op}}\|\mathbf{s}_1\|\|\mathbf{u}_1 - \mathbf{u}_2\| + \|\mathbf{H}\|_{\text{op}}\|\mathbf{s}_1 - \mathbf{s}_2\|\|\mathbf{A}_2^{-1}\|_{\text{op}}\|\mathbf{u}_{\text{in}}\|.\end{aligned}$$

Then the result $T_1 \le L_1\|\mathbf{s}_1 - \mathbf{s}_2\|$ follows by noticing that $\|\mathbf{s}_1\|$, $\|\mathbf{u}_{\text{in}}\|$, $\|\mathbf{G}\|_{\text{op}}$, $\|\mathbf{H}\|_{\text{op}}$, and $\|\mathbf{A}_i^{-1}\|_{\text{op}}$ for $i = 1, 2$ are bounded, and the fact that $\|\mathbf{r}_1\| \le \|\mathbf{y}\| + \|\mathbf{H}\|_{\text{op}}\|\mathbf{s}_1\|\|\mathbf{A}_1^{-1}\|_{\text{op}}\|\mathbf{u}_{\text{in}}\| < \infty$.

### 5.3. Proof for Proposition 3

By (10), we have that for all $\mathbf{x}, \mathbf{y} \in \mathcal{U}$,

$$|\mathcal{D}(\mathbf{x}) - \mathcal{D}(\mathbf{y}) - \langle\nabla\mathcal{D}(\mathbf{y}), \mathbf{x} - \mathbf{y}\rangle| \le \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|^2. \tag{16}$$

By (5), we have that for all $\mathbf{x} \in \mathcal{U}$, $t \ge 0$,

$$\mathcal{R}(\mathbf{x}) \ge \mathcal{R}(\mathbf{f}_t) + \langle\frac{\mathbf{s}_t - \mathbf{f}_t}{\gamma} - \nabla\mathcal{D}(\mathbf{s}_t), \mathbf{x} - \mathbf{f}_t\rangle. \tag{17}$$

Let $\mathbf{x} = \mathbf{f}_k$, $\mathbf{y} = \mathbf{f}_{k+1}$ in (16) and $\mathbf{x} = \mathbf{f}_k$, $t = k+1$ in (17). Then, adding the two inequalities, we have

$$\begin{aligned}&\mathcal{F}(\mathbf{f}_{k+1}) - \mathcal{F}(\mathbf{f}_k) \le \langle\nabla\mathcal{D}(\mathbf{f}_{k+1}) - \nabla\mathcal{D}(\mathbf{s}_{k+1}), \mathbf{f}_{k+1} - \mathbf{f}_k\rangle \\ &+ \frac{1}{\gamma}\langle\mathbf{s}_{k+1} - \mathbf{f}_{k+1}, \mathbf{f}_{k+1} - \mathbf{f}_k\rangle + \frac{L}{2}\|\mathbf{f}_{k+1} - \mathbf{f}_k\|^2 \qquad (18) \\ &\overset{(a)}{\le} \frac{L}{2}\|\mathbf{s}_{k+1} - \mathbf{f}_{k+1}\|^2 + \frac{L}{2}\|\mathbf{f}_{k+1} - \mathbf{f}_k\|^2 + \frac{1}{2\gamma}\|\mathbf{s}_{k+1} - \mathbf{f}_k\|^2 \\ &- \frac{1}{2\gamma}\|\mathbf{s}_{k+1} - \mathbf{f}_{k+1}\|^2 - \frac{1}{2\gamma}\|\mathbf{f}_{k+1} - \mathbf{f}_k\|^2 + \frac{L}{2}\|\mathbf{f}_{k+1} - \mathbf{f}_k\|^2 \\ &\overset{(b)}{\le} \left(\frac{1}{2\gamma} - L\right)\left(\|\mathbf{f}_k - \mathbf{f}_{k-1}\|^2 - \|\mathbf{f}_{k+1} - \mathbf{f}_k\|^2\right) - \left(\frac{1}{2\gamma} - \frac{L}{2}\right)\|\mathbf{s}_{k+1} - \mathbf{f}_{k+1}\|^2.\end{aligned}$$

In the above, step $(a)$ uses Cauchy-Schwarz, Proposition 2, as well as the fact that $2ab \le a^2 + b^2$ and $2\langle\mathbf{a} - \mathbf{b}, \mathbf{b} - \mathbf{c}\rangle = \|\mathbf{a} - \mathbf{c}\|^2 - \|\mathbf{a} - \mathbf{c}\|^2 - \|\mathbf{b} - \mathbf{c}\|^2$. Step $(b)$ uses the condition in the proposition statement that $\gamma \le \frac{1-\alpha^2}{2L}$ and (7), which implies $\|\mathbf{s}_{k+1} - \mathbf{f}_k\| \le \alpha\frac{t_k - 1}{t_{k+1}}\|\mathbf{f}_k - \mathbf{f}_{k-1}\|$, where we notice that $\frac{t_k - 1}{t_{k+1}} \le 1$ by (6), and $\alpha < 1$ by our assumption. Summing both sides from $k = 0$ to $K$:

$$\begin{aligned}&\left(\frac{1}{2\gamma} - \frac{L}{2}\right)\sum_{k=0}^{K-1}\|\mathbf{s}_{k+1} - \mathbf{f}_{k+1}\|^2 \le \mathcal{F}(\mathbf{f}_0) - \mathcal{F}(\mathbf{f}_K) \\ &+ \left(\frac{1}{2\gamma} - L\right)\left(\|\mathbf{f}_0 - \mathbf{f}_{-1}\|^2 - \|\mathbf{f}_K - \mathbf{f}_{K-1}\|^2\right) \le \mathcal{F}(\mathbf{f}_0) - \mathcal{F}^*,\end{aligned}$$

where $\mathcal{F}^*$ is the global minimum. The last step follows by letting $\mathbf{f}_{-1} = \mathbf{f}_0$, which satisfies (7) for the initialization $\mathbf{s}_1 = \mathbf{f}_0$, and the fact that $\mathcal{F}^* \le \mathcal{F}(\mathbf{f}_K)$. Since $\mathcal{G}_\gamma(\mathbf{s}_k) = \frac{\mathbf{s}_k - \mathbf{f}_k}{\gamma}$, we have

$$\lim_{K\to\infty}\sum_{k=1}^{K}\|\mathcal{G}_\gamma(\mathbf{s}_k)\|^2 \le \frac{2L\left(\mathcal{F}(\mathbf{f}_0) - \mathcal{F}^*\right)}{\gamma L(1 - \gamma L)} < \infty.$$

Therefore, $\lim_{k\to\infty}\|\mathcal{G}_\gamma(\mathbf{s}_k)\| = 0$.

## 6. REFERENCES


[1] M. Born and E. Wolf, *Principles of Optics*, chapter Scattering from inhomogeneous media, pp. 695–734, Cambridge Univ. Press, 7 edition, 2003.

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

[3] 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.

[4] 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.

[5] 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, pp. 024003, 2009.

[6] P. M. van den Berg and R. E. Kleinman, “A contrast source inversion method,” *Inv. Probl.*, vol. 13, no. 6, pp. 1607–1620, December 1997.

[7] 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.

[8] M. T. Bevacquad, L. Crocco, L. Di Donato, and T. Isernia, “Non-linear inverse scattering via sparsity regularized contrast source inversion,” *IEEE Trans. Comp. Imag.*, 2017.

[9] Kamal Belkebir and Anne Sentenac, “High-resolution optical diffraction microscopy,” *J. Opt. Soc. Am. A*, vol. 20, no. 7, pp. 1223–1229, July 2003.

[10] 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, pp. 065007, April 2012.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] H.-Y. Liu, D. Liu, H. Mansour, P. T. Boufounos, L. Waller, and U. S. Kamilov, “SEAGLE: Sparsity-driven image reconstruction under multiple scattering,” May 2017, arXiv:1705.04281 [cs.CV].

[17] 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.

[18] I. Daubechies, M. Defrise, and C. De 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.

[19] J. Bect, L. Blanc-Feraud, G. Aubert, and A. Chambolle, “A $\ell_1$-unified variational framework for image restoration,” in *Proc. ECCV*, Springer, Ed., New York, 2004, vol. 3024, pp. 1–13.

[20] 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.

[21] H. Li and Z. Lin, “Accelerated proximal gradient methods for nonconvex programming,” in *Proc. Advances in Neural Information Processing Systems 28*, Montreal, Canada, December 7-12 2015.

[22] Emmanuel Soubies, Thanh-An Pham, and Michael Unser, “Efficient inversion of multiple-scattering model for optical diffraction tomography,” *Opt. Express*, vol. 25, no. 18, pp. 21786–21800, Sep 2017.

[23] 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.

[24] 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.

[25] 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.

[26] R. Tyrrell Rockafellar and Roger J-B Wets, *Variational Analysis*, Springer Science & Business Media, 2009.