

Accelerated Image Reconstruction for Nonlinear Diffractive Imaging

Yanting Ma^{*}, Hassan Mansour[†], Dehong Liu[†],
Petros T. Boufounos[†], and Ulugbek S. Kamilov[†]

August 4, 2017

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.

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 object, nonlinear formulations that can model multiple scattering need to be considered. More recent work has been trying to integrate the nonlinearity in the forward model and design new nonlinear inverse algorithms to reconstruct the object. Examples of existing nonlinear methods include iterative linearization [4, 5], contrast source inversion [6–8], hybrid methods [9–11], and optimization with error backpropagation [12–16].

One way for solving the inverse problem is via optimization. The objective function usually consists of a smooth data fidelity term and a non-smooth regularization term whose proximal mapping is easily computed. For such objective 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 for FISTA applied to nonconvex cases. A variant of FISTA has been proposed in [21] for nonconvex optimization with convergence guarantee. This algorithm computes two estimates from ISTA and FISTA, respectively, at each iteration, and selects the one with lower objective function value as the final estimate at the current iteration. Therefore, both the gradient and the objective function need to be evaluated at two different points at each iteration. While such extra computation cost may be insignificant in some applications, it can be prohibitive in the inverse scattering problem that we consider here, since additional evaluations of the gradient and the objective function require computation of the entire forward model.

In this work, we propose a new image reconstruction method called *Convergent Inverse Scattering based on Optimization and Regularization (CISOR)*. CISOR is based on our novel nonconvex optimization formulation that

^{*}Y. Ma (email: yma7@ncsu.edu) is with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27606. This work was completed while H.-Y. Liu was with Mitsubishi Electric Research Laboratories (MERL).

[†]H. Mansour (email: mansour@merl.com), D. Liu (email: liudh@merl.com), P. T. Boufounos (email: petrosb@merl.com), and U. S. Kamilov (email: kamilov@merl.com) are with Mitsubishi Electric Research Laboratories (MERL), 201 Broadway, Cambridge, MA 02139, USA.

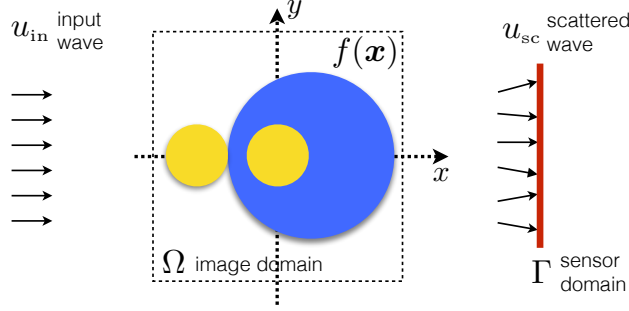


Figure 1: Visual representation of the measurement scenario considered in this paper. An object with a real scattering potential $f(x)$ is illuminated with an input wave $u_{\text{in}}(x)$, which interacts with the object and results in the scattered wave u_{sc} at the sensor domain $\Gamma \subset \mathbb{R}^2$. The complex scattered wave is captured at the sensor and the algorithm proposed here is used for estimating the potential f .

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. The convergence analysis of our new FISTA variant may be of interest on its own as a general nonconvex solver.

This paper is organized as follows. In Section II, we present the formulation of the scattering problem. In Section III, we present CISOR and analyze its convergence. In Section IV, we present experimental results illustrating the performance of our algorithm on simulated as well as experimental data. The technical details of our proofs are presented in the appendix.

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_{in} , and the scattered wave u_{sc} is measured by the sensors placed in a sensing region $\Gamma \subset \mathbb{R}^2$. Let u denote the total field, which satisfies $u(x) = u_{\text{in}}(x) + u_{\text{sc}}(x)$, $\forall x \in \mathbb{R}^2$. The Lippmann-Schwinger equation [1] establishes the fundamental object-wave relationship

$$u_{\text{sc}}(x) = \int_{\Omega} g(x-x')u(x')f(x')dx', \forall x \in \Gamma$$

$$u(x) = u_{\text{in}}(x) + \int_{\Omega} g(x-x')u(x')f(x')dx', \forall x \in \Omega.$$

In the above, $f(x) = k^2(\epsilon(x) - \epsilon_b)$ is the scattering potential, where $\epsilon(x)$ is the permittivity of the object, ϵ_b is the permittivity of the background, and $k = 2\pi/\lambda$ is the wavenumber in vacuum. The free-space Green's function in 2D is defined as $g(x) = -\frac{j}{4}H_0^{(1)}(k_b\|x\|)$, where $H_0^{(1)}$ is the Hankel function of first kind and $k_b = k\sqrt{\epsilon_b}$ is the wavenumber of the background medium. The corresponding discrete system is then

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

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 $f(x)$, $u(x)$, and $u_{\text{in}}(x)$ on Ω , respectively, 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 the Green's function $g(x-x')$ with $x \in \Gamma$ and $x' \in \Omega$, whereas $\mathbf{G} \in \mathbb{C}^{N \times N}$ is the discretization of the Green's function with $x, x' \in \Omega$. The inverse scattering problem is then to estimate \mathbf{f} given \mathbf{y} , \mathbf{H} , \mathbf{G} , and \mathbf{u}_{in} . Note that this is a nonlinear inverse problem, because \mathbf{u} depends on \mathbf{f} through $\mathbf{u} = (\mathbf{I} - \mathbf{G}\text{diag}(\mathbf{f}))^{-1}\mathbf{u}_{\text{in}}$.

3 Proposed Method

Our proposed method is based on a nonconvex optimization formulation with total variation regularization. Let $\mathbf{A} := \mathbf{I} - \mathbf{G}\text{diag}(\mathbf{f})$ and $\mathcal{Z}(\mathbf{f}) := \mathbf{H}\text{diag}(\mathbf{f})\mathbf{u} = \mathbf{H}\text{diag}(\mathbf{f})\mathbf{A}^{-1}\mathbf{u}_{\text{in}}$. 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:

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

with

$$\mathcal{D}(\mathbf{f}) = \frac{1}{2} \|\mathbf{y} - \mathcal{Z}(\mathbf{f})\|_2^2 \quad (3)$$

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

where \mathbf{D}_d is the discrete gradient operator in the d th dimension, and $\chi_{\mathcal{C}}(\cdot)$ is defined as

$$\chi_{\mathcal{C}}(\mathbf{f}) := \begin{cases} 0, & \text{if } \mathbf{f} \in \mathcal{C} \\ \infty, & \text{if } \mathbf{f} \notin \mathcal{C} \end{cases}.$$

Note that $\mathcal{D}(\cdot)$ is differentiable whenever \mathbf{A} is non-singular, $\mathcal{R}(\cdot)$ is proper convex closed (lower semi-continuous) whenever \mathcal{C} is convex and closed, and $\mathcal{F}(\cdot)$ is nonconvex since $\mathcal{D}(\cdot)$ is nonconvex.

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:

$$\mathbf{f}_k = \text{prox}_{\gamma \mathcal{R}}(\mathbf{s}_k - \gamma \nabla \mathcal{D}(\mathbf{s}_k)) \quad (5)$$

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

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

where the choice of the step-size γ 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 illustrates the convergence rate of relaxed FISTA for values of α between 0 and 1. The plot was obtained by using the experimentally measured scattered microwave data collected by the Fresnel institute [22]. We can see from Figure 2 that the empirical convergence speed increases as α increases from 0 to 1. Our theoretical convergence analysis of relaxed FISTA is presented in Section 3.2 and 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 $\text{prox}_{\gamma \mathcal{R}}$. Given $\nabla \mathcal{D}(\mathbf{s}_k)$, the proximal mapping (5) can be efficiently solved by constrained TV-FISTA [23]. The following proposition provides an explicit formula for $\nabla \mathcal{D}$.

Proposition 1. Let $\mathcal{Z}(\mathbf{f}) = \mathbf{H}\text{diag}(\mathbf{f})\mathbf{u}$, $\mathbf{r} = \mathbf{y} - \mathcal{Z}(\mathbf{f})$, and $\mathbf{J}_{\mathcal{Z}}$ be the Jacobian matrix of \mathcal{Z} with Hermitian $\mathbf{J}_{\mathcal{Z}}^H$. Then we have

$$\nabla \mathcal{D}(\mathbf{f}) = \text{Re} \{ \mathbf{J}_{\mathcal{Z}}^H \mathbf{r} \} = \text{Re} \{ \text{diag}(\mathbf{u})^H (\mathbf{H}^H \mathbf{r} + \mathbf{G}^H \mathbf{v}) \}, \quad (8)$$

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

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

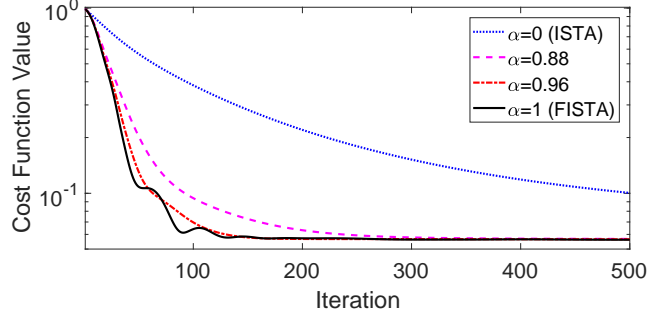


Figure 2: Empirical convergence speed for relaxed FISTA with various α values tested on experimentally measured data.

Proof. See Appendix ??.

□

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 Performance Guarantee

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 $\mathbf{A} = \mathbf{I} - \mathbf{G}\text{diag}(\mathbf{s})$ 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

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

Proof. See Appendix ??.

□

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

$$\mathcal{G}_\gamma(\mathbf{s}) := \frac{\mathbf{s} - \text{prox}_{\gamma \mathcal{D}}(\mathbf{s} - \gamma \nabla \mathcal{D}(\mathbf{s}))}{\gamma}.$$

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

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

Proof. See Appendix ??.

□

Note that when $\mathcal{G}_\gamma(\hat{\mathbf{s}}) = 0$, this implies that $0 \in \partial \mathcal{F}(\hat{\mathbf{s}})$, where $\partial \mathcal{F}$ denotes the limiting subdifferentiation of \mathcal{F} [24]. Hence, $\hat{\mathbf{s}}$ is a stationary point of \mathcal{F} . This establishes that relaxed FISTA converges to the stationary point of the problem (2).

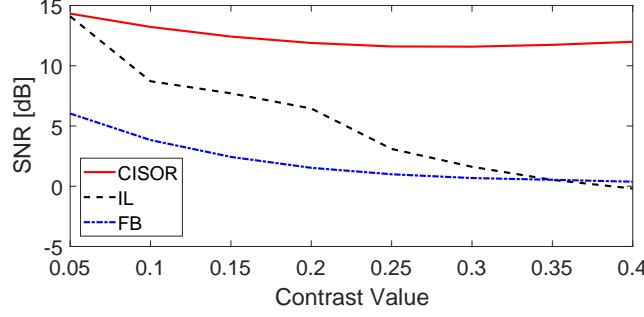


Figure 3: Comparison of different reconstruction methods for various contrast levels tested on simulated data.

4 Experimental Results

We now compare our method CISOR with the state-of-the-art method SEAGLE [15], as well as several conventional methods, including the first Born approximation (FB) [1], iterative linearization (IL) [4, 5], and contrast source inversion (CSI) [6–8].

Comparison on simulated data. We first use simulated data to demonstrate the reconstruction quality of different methods for objects with various contrast values, where the contrast of an object \mathbf{f} is defined as $\max(|\mathbf{f}|)/k_b^2$. In the experiment, we generate a Shepp-Logan image in Matlab and change its contrast to the desired value to obtain the ground-truth \mathbf{f}_{true} . We then run our forward model to generate the scattered wave, which is then used as measurements. Let the center of the image be the origin, and the physical size of the image is assumed to be $120 \text{ cm} \times 120 \text{ cm}$. 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° . 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 5 different regularization parameter values: $\tau = 2^a \times 2.5 \times 10^{-5} \|\mathbf{y}\|^2$ with $a = -2, -1, 0, 1, 2$, 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 the proposed method CISOR is not affected by the increasing contrast value.

Comparison on experimental data. We use three objects from the public dataset provided by the Fresnel institute [22]: *FoamDielExtTM*, *FoamDielIntTM*, and *FoamTwinDielTM*. The objects are placed within a $15 \text{ cm} \times 15 \text{ cm}$ square region centered at the origin of the coordinate system. The number of transmitters is 8 for *FoamDielExtTM*, *FoamDielIntTM*, and 18 for *FoamTwinDielTM*. 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 demonstrates the relative reconstructed SNR, which is defined as $20 \log_{10}(\|\mathbf{f}_{\text{ref}}\|/\|\hat{\mathbf{f}} - \mathbf{f}_{\text{ref}}\|)$, obtained by CISOR with different measurement reduction factors, where \mathbf{f}_{ref} is the reconstructed image using all measurements, namely, 8 transmitter with 241 measurements each. A reduction factor of 64 reduces the number of measurements for each transmitter to 7. The regularization parameter τ is set to be $2.5 \times 10^{-5} \|\mathbf{y}_{\mathcal{S}}\|^2$, where $\mathbf{y}_{\mathcal{S}}$ is the subset of the full measurements that is used for reconstruction.

Figure 5 provides a visual comparison of the reconstructed images obtained by different algorithms. For each object and each algorithm, we run the algorithm with 5 different regularization parameter values: $\tau = 2^a \times 2.5 \times 10^{-5} \|\mathbf{y}\|^2$ with $a = -2, -1, 0, 1, 2$, and select the result that has the best visual quality. Figure 5 shows that all nonlinear methods CISOR, SEAGLE, CSI, and IL obtained reasonable 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. Among the nonlinear methods, CISOR and SEAGLE obtained images with the

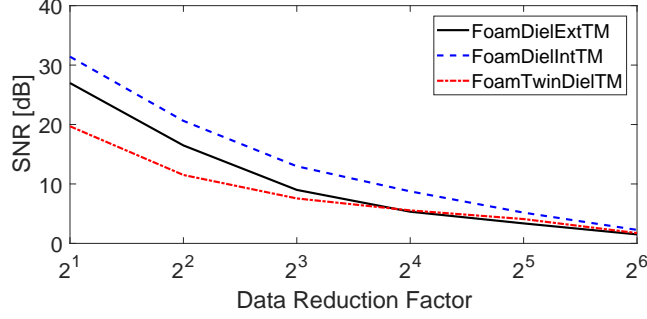


Figure 4: Reconstruction quality CISOR with various data reduction factors tested on experimentally measured data.

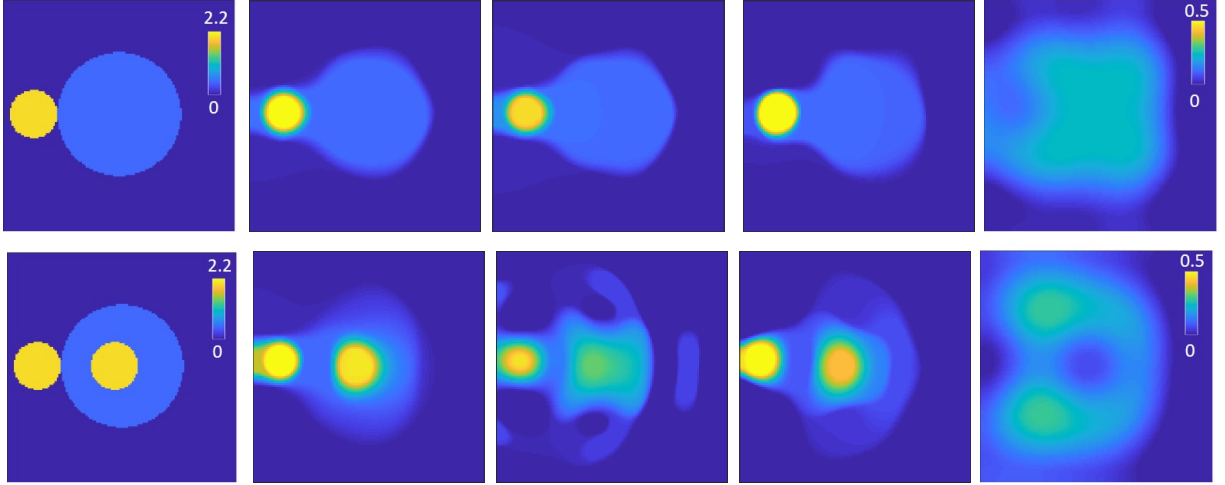


Figure 5: Reconstructed images obtained by different algorithms from experimentally measured data. The first row uses *FoamDielExtTM*, and the second row uses *FoamTwinDielTM*. From left to right: ground truth, reconstructed images by CISOR, SEAGLE, IL, and FB. Note that the color-map used for FB is different from the rest, because FB significantly underestimated the contrast values.

best visual quality.

Note that one advantage of CISOR over SEAGLE is that CISOR computes the gradient $\mathcal{D}(\cdot)$ (8) on the fly at each FISTA iteration, whereas SEAGLE needs to store iterates from the forward model for computing the gradient. Therefore, in order to obtain accurate gradient, CISOR can set the stopping criterion for the conjugate gradient algorithm to be sufficiently small when computing \mathbf{u} and \mathbf{v} in (9), whereas SEAGLE needs to consider the availability of memory to store the iterates. Moreover, the relaxed FISTA applied in CISOR has theoretical convergence guarantee for the nonconvex problem (2), whereas the standard FISTA applied in SEAGLE is not guaranteed to convergence for nonconvex problems.

5 Conclusion

In this paper, we proposed a nonconvex formulation for nonlinear diffractive imaging. The nonconvex optimization problem was solved by our new variant of FISTA. We provided an explicit formula for fast computation of the gradient at each FISTA iteration and proved that the algorithm converges for our nonconvex problem. Numerical results demonstrated that the proposed method is competitive with several state-of-the-art methods. Two key advantages of CISOR over other methods are in its (i) memory efficiency and (ii) convergence guarantees.

6 Appendix

6.1 Proof for Proposition 2

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

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

Label the two terms on the RHS as T_1 and T_2 . We will prove $T_1 \leq L_1\|\mathbf{s}_1 - \mathbf{s}_2\|$ for some $L_1 \in (0, \infty)$, and $T_2 \leq L_2\|\mathbf{s}_1 - \mathbf{s}_2\|$ can be proved in a similar way for some $L_2 \in (0, \infty)$. The result (10) is then obtained by letting $L = L_1 + L_2$.

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

where the last inequality uses the fact that for a diagonal matrix, $\|\text{diag}(\mathbf{d})\| = \max_{n \in [N]} |d_n| \leq \|\mathbf{d}\|$. Notice that

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

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

6.2 Proof for Proposition 3

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

$$|\mathcal{D}(\mathbf{x}) - \mathcal{D}(\mathbf{y}) - \langle \nabla\mathcal{D}(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle| \leq \frac{L}{2} \|\mathbf{x} - \mathbf{y}\|^2. \quad (11)$$

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

$$\mathcal{R}(\mathbf{x}) \geq \mathcal{R}(\mathbf{f}_t) + \left\langle \frac{\mathbf{s}_t - \mathbf{f}_t}{\gamma} - \nabla\mathcal{D}(\mathbf{s}_t), \mathbf{x} - \mathbf{f}_t \right\rangle. \quad (12)$$

Let $\mathbf{x} = \mathbf{f}_k$, $\mathbf{y} = \mathbf{f}_{k+1}$ in (11) and $\mathbf{x} = \mathbf{f}_k$, $t = k + 1$ in (12). Then, adding the two inequalities, we have

$$\begin{aligned} \mathcal{R}(\mathbf{f}_{k+1}) - \mathcal{R}(\mathbf{f}_k) &\leq \langle \nabla\mathcal{D}(\mathbf{f}_{k+1}) - \nabla\mathcal{D}(\mathbf{s}_{k+1}), \mathbf{f}_{k+1} - \mathbf{f}_k \rangle \\ &\quad + \frac{1}{\gamma} \langle \mathbf{s}_{k+1} - \mathbf{f}_{k+1}, \mathbf{f}_{k+1} - \mathbf{f}_k \rangle + \frac{L}{2} \|\mathbf{f}_{k+1} - \mathbf{f}_k\|^2 \\ &\stackrel{(a)}{\leq} \frac{L}{2} \|\mathbf{s}_{k+1} - \mathbf{f}_{k+1}\|^2 + \frac{L}{2} \|\mathbf{f}_{k+1} - \mathbf{f}_k\|^2 + \frac{1}{2\gamma} \|\mathbf{s}_{k+1} - \mathbf{f}_k\|^2 \\ &\quad - \frac{1}{2\gamma} \|\mathbf{s}_{k+1} - \mathbf{f}_{k+1}\|^2 - \frac{1}{2\gamma} \|\mathbf{f}_{k+1} - \mathbf{f}_k\|^2 + \frac{L}{2} \|\mathbf{f}_{k+1} - \mathbf{f}_k\|^2 \\ &\stackrel{(b)}{\leq} \left(\frac{1}{2\gamma} - L \right) (\|\mathbf{f}_k - \mathbf{f}_{k-1}\|^2 - \|\mathbf{f}_{k+1} - \mathbf{f}_k\|^2) \\ &\quad - \left(\frac{1}{2\gamma} - \frac{L}{2} \right) \|\mathbf{s}_{k+1} - \mathbf{f}_{k+1}\|^2. \end{aligned} \quad (13)$$

In the above, step (a) uses Cauchy-Swartz and Proposition 2, as well as the fact that $2ab \leq a^2 + b^2$, for the first term in (13), and uses the formula $2\langle \mathbf{a}-\mathbf{b}, \mathbf{b}-\mathbf{c} \rangle = \|\mathbf{a}-\mathbf{c}\|^2 - \|\mathbf{a}-\mathbf{b}\|^2 - \|\mathbf{b}-\mathbf{c}\|^2$ for the second term in (13). Step (b) uses the condition in the proposition statement that $\gamma \leq \frac{1-\alpha^2}{2L}$ and (7), which implies $\|\mathbf{s}_{k+1}-\mathbf{f}_k\| \leq \alpha \frac{t_k-1}{t_{k+1}} \|\mathbf{f}_k-\mathbf{f}_{k-1}\|$, where we notice that $\frac{t_k-1}{t_{k+1}} \leq 1$ by (6), and $\alpha < 1$ by our assumption. Summing both sides from $k = 0$ to K :

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

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

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

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

References

- [1] M. Born and E. Wolf, *Principles of Optics*, 7th ed. Cambridge Univ. Press, 2003, ch. Scattering from inhomogeneous media, pp. 695–734.
- [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, p. 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] K. Belkebir and A. 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, p. 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] —, “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. 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.
- [19] 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.
- [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] 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.
- [23] 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.
- [24] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*. Springer Science & Business Media, 2009.