SIMBA: Scalable Inversion in Optical Tomography using Deep Denoising Priors

Zihui Wu*, Yu Sun*, Alex Matlock†, Jiaming Liu‡,
Lei Tian†, and Ulugbek S. Kamilov*,‡

Abstract

Two features desired in a three-dimensional (3D) optical tomographic image reconstruction algorithm are the ability to reduce imaging artifacts and to do fast processing of large data volumes. Traditional iterative inversion algorithms are impractical in this context due to their heavy computational and memory requirements. We propose and experimentally validate a novel *scalable iterative minibatch algorithm (SIMBA)* for fast and high-quality optical tomographic imaging. SIMBA enables high-quality imaging by combining two complementary information sources: the physics of the imaging system characterized by its forward model and the imaging prior characterized by a denoising deep neural net. SIMBA easily scales to very large 3D tomographic datasets by processing only a small subset of measurements at each iteration. We establish the theoretical fixed-point convergence of SIMBA under nonexpansive denoisers for convex data-fidelity terms. We validate SIMBA on both simulated and experimentally collected intensity diffraction tomography (IDT) datasets. Our results show that SIMBA can significantly reduce the computational burden of 3D image formation without sacrificing the imaging quality.

1 Introduction

Optical tomographic imaging seeks to recover the three-dimensional (3D) distribution of the refractive index of an object from its light measurements. In a standard setup (see Figure 1 for an example), the sample is illuminated multiple times from different angles and the scattered light-field is recorded with a camera. In the interferometry-based microscopy, one measures both the amplitude and the phase of the scattered field [1–3], while in the intensity-only setups one measures only the amplitude of the light-field [4–6]. A tomographic reconstruction algorithm is then used to computationally reconstruct the 3D distribution of the sample's refractive index. The quantitative characterization of the refractive index is important in biomedical imaging since it allows to visualize the internal structure of a tissue, as well as characterize physical changes within biological samples.

The reconstruction of the refractive index is often formulated as an inverse problem. In this context, the *forward model* characterizes the physics of data-acquisition and can be used to ensure the consistency of the final estimate with respect to the measurements. However, the need for processing large-scale tomographic data limits the utility of traditional iterative methods in 3D optical tomography. Traditional *batch* algorithms process the whole tomographic dataset at every iteration. On the other hand, *online* algorithms can effectively scale to large datasets by processing only a small subset of data per iteration.

Using imaging priors is a standard strategy for mitigating the ill-posed nature of many tomographic imaging problems. Popular imaging priors include Tikhonov [7] and total variation (TV) [8] regularizers. Recently,

*Department of Computer Science & Engineering, Washington University in St. Louis, St. Louis, MO 63130.
†Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215, USA.
‡Department of Electrical & Systems Engineering, Washington University in St. Louis, St. Louis, MO 63130.

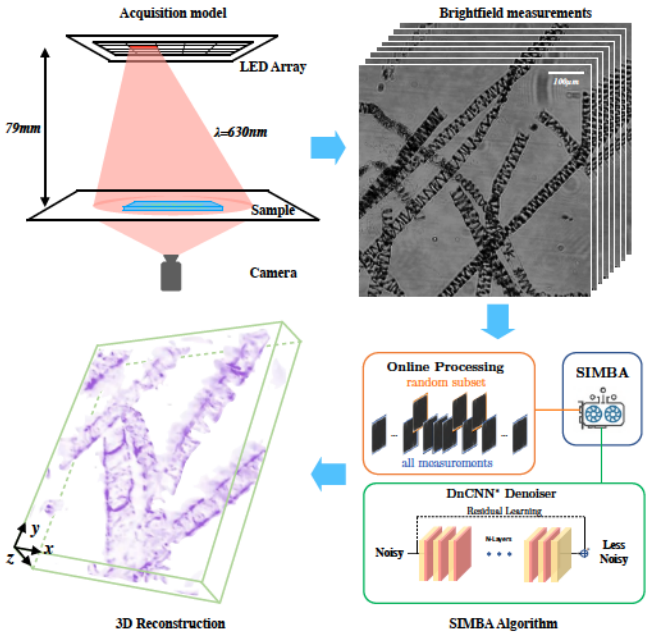

Figure 1: The conceptual illustration of the proposed inversion algorithm for optical tomographic imaging. The brightfield measurements of the scattered light-field are collected with a standard computational microscope platform. An *online* reconstruction algorithm, SIMBA, facilitated by a *convolutional neural network (CNN)* denoiser is then used to form a 3D phase image. Unlike traditional batch algorithms, SIMBA processes a subset of measurements at a time, making it scalable for processing very large tomographic datasets.

a new class of methods, called *plug-and-play priors (PnP)* [9], have popularized the idea of using general image denoisers as imaging priors within iterative inversion. By leveraging advanced image denoisers, such as BM3D [10] and DnCNN [11], PnP methods have achieved the state-of-the-art performance in various imaging applications [12–21]. An alternative framework for using image denoisers is the *regularization by denoising (RED)* [22], where the denoiser is used to formulate an explicit regularizer that has a simple gradient. The work [23] has clarified the existence of RED regularizers for certain class of denoisers, and the excellent performance of the framework has been demonstrated in phase retrieval [24] and image super-resolution [25] using DnCNN and the deep image prior, respectively. In short, using advanced denoisers has proven to be effective for improving the reconstruction quality in various imaging contexts. Note that the concept of regularization described in this paper is distinct from the concept of regularization for training deep neural nets by injecting noise [26].

In this paper, we present a new *scalable iterative minibatch algorithm (SIMBA)* for the regularized inversion in optical tomography. SIMBA is an online extension of the traditional RED framework. It can thus leverage powerful convolutional neural network (CNN) denoisers as imaging priors, while also taking advantage of the physical information available through the forward model. However, unlike traditional RED algorithms, SIMBA is scalable to datasets that are too large for batch processing since it only uses a subset of measurements at a time. We prove that SIMBA converges in expectation to the same set of fixed points as its batch counterparts under a set of transparent assumptions. Thus, SIMBA benefits from the excellent imaging quality offered by RED, but does so in a computationally tractable way for optical tomographic imaging.

We validate SIMBA in the context of intensity-only microscopy called *intensity diffraction tomography (IDT)*. IDT microscopes are relatively cheap and easy to implement since they do not collect the phase of the light. We adopt the IDT forward model in [27] that establishes a linear relationship between the desired object and the intensity measurements by neglecting the terms corresponding to higher order light scattering. We show that SIMBA can efficiently reconstruct a high-resolution ($1024 \times 1024 \times 25$ pixels) IDT

image while also offering improvements in the 3D sectioning capability. The preliminary version of this work was presented in [28]. The current paper significantly extends [28] by including the IDT model, providing additional simulations, and validating the method on an experimentally collected 3D IDT dataset.

This paper is organized as follows. In Section 2, we introduce the IDT forward model and the RED framework. In Section 3, we present the algorithmic details of SIMBA. In Section 4, we analyze the fixed-point convergence under a set of assumptions. In Section 5, we provide simulations and experiments that illustrates the efficiency and effectiveness of SIMBA. Section 6 concludes the paper.

2 Background

In this section, we provide the background on IDT and image-denoising priors. We start by describing the IDT forward model, then formulate the corresponding inverse problem, and finally introduce the RED framework as a strategy to leverage image denoisers as priors.

2.1 Linearized IDT

Consider a 3D object with the permittivity distribution $\epsilon(\mathbf{r})$ in a bounded sample domain $\Omega \subset \mathbb{R}^3$, immersed into the background medium of permittivity ϵ_b. We use $\Delta\epsilon = \Delta\epsilon_{\mathrm{Re}} + i\Delta\epsilon_{\mathrm{Im}} = \epsilon - \epsilon_b$ to denote the permittivity contrast between the object and the background medium. The real part $\Delta\epsilon_{\mathrm{Re}}$ corresponds to the phase effect, and the imaginary part $\Delta\epsilon_{\mathrm{Im}}$ accounts for the absorption. The object is illuminated by an angled incident light field $u_{\mathrm{in}}(\mathbf{r})$. The incident field u_{in} is assumed to be known inside Ω as well as at the camera domain $\Gamma \subset \mathbb{R}^2$. The total light-field $u(\mathbf{r})$ is measured only through its intensity at the camera. Here, $\mathbf{r} = (x, y, z)$ denotes the 3D spatial coordinates. Under the first Born approximation [29], the light-sample interaction is described by the following equation

$$u(\mathbf{r}) = u_{\mathrm{in}}(\mathbf{r}) + \int_{\Omega} g(\mathbf{r} - \boldsymbol{r}')\, v(\mathbf{r}')\, u_{\mathrm{in}}(\mathbf{r}')\, \mathrm{d}\mathbf{r}', \quad \mathbf{r} \in \Omega \tag{1}$$

where $u(\mathbf{r}) = u_{\mathrm{in}}(\mathbf{r}) + u_{\mathrm{sc}}(\mathbf{r})$ is the total light field, $v(\mathbf{r}) = \frac{1}{4\pi} k^2 \Delta\epsilon$ is the scattering potential, $k = 2\pi/\lambda$ is wave number in free space, and λ is the wavelength of the illumination. In the 3D space, the Green's function at the camera plane Γ is given by

$$g(\mathbf{r}) = \frac{\mathbf{e}^{ik_b \|\mathbf{r}\|_2}}{\|\mathbf{r}\|_2},$$

where $k_b = \sqrt{\epsilon_b} k$ is the wavenumber of the background medium, and $\|\cdot\|_2$ denotes the ℓ_2-norm. For a single illumination, the intensity of the light field after propagating through the sample is given by

$$\mathbf{I} = |u(\mathbf{r}) * p|^2, \tag{2}$$

where p is the point spread function of the microscope, and the operator $*$ denotes the 2D convolution. Eq. (2) can be expanded into the summation of four components

$$\mathbf{I} = \mathbf{I}^{ii} + \mathbf{I}^{ss} + \mathbf{I}^{is} + \mathbf{I}^{si}, \tag{3}$$

where $\mathbf{I}^{ii}$ is the constant background intensity, $\mathbf{I}^{ss}$ is the squared modulus of the scattered field, and $\mathbf{I}^{is} = (\mathbf{I}^{si})^*$ are the cross terms that relate the unscattered and scattered field. Here, $(\cdot)^*$ denotes the complex conjugate. Due to the first Born approximation, $\mathbf{I}^{ss}$ can be assumed to be small and thus neglected. By modeling the 3D object as a series of slices along the axial dimension z, one can represent the spectrum of the total scattered field as the summation of the sub-scattered fields produced by each slice [27]

$$\widetilde{\mathbf{I}} = \widetilde{\mathbf{I}}^{ii} + \int \left[H_{\mathrm{Re}}(z) \widetilde{\Delta\epsilon}_{\mathrm{Re}}(z) + H_{\mathrm{Im}}(z) \widetilde{\Delta\epsilon}_{\mathrm{Im}}(z) \right] \mathrm{d}z \tag{4}$$

where $\widetilde{\cdot}$ denotes 2D Fourier transform, and $\widetilde{\mathbf{I}}^{ii}$ is the background intensity spectrum measured at Γ. In (4), H_{Re} and H_{Im} are the angle-dependent phase and absorption transfer functions (TF) for each sample slice at depth z, respectively. These TFs linearly map the Fourier transform of the permittivity contrast to the intensity spectrum of the scattered field. We refer the reader to [27] for the full details of the TF for IDT.

By discretizing (4) and explicitly including the Fourier transform into the equation, we obtain the following linear model in the spatial domain for the i^{th} illumination

$$\boldsymbol{I}_i = \boldsymbol{I}_i^{ii} + \Re\left\{ \sum_{j=0}^{J} \boldsymbol{A}_{ij}\boldsymbol{x}_j \right\}, \quad \text{with} \quad \boldsymbol{A}_{ij} = \boldsymbol{F}^{\mathsf{H}}\boldsymbol{H}_{ij}\boldsymbol{F}, \tag{5}$$

where $j = 0, \ldots, J$ discretely indexes the axial direction z, $\boldsymbol{x}_j \in \mathbb{C}^N$ is the discretized complex permittivity contrast of the j^{th} slice, $\boldsymbol{I}_i$ is the measured intensity of the total field, $\boldsymbol{I}_i^{ii}$ is the discretized intensity of the background, and $\boldsymbol{H}_{ij}$ is the discretized TF accounting for both phase and absorption at z_j. We use $\boldsymbol{F}$ and $\boldsymbol{F}^{\mathsf{H}}$ to denote the 2D discrete Fourier transform and its inverse, respectively. By re-arranging the terms, we can obtain the following linear forward model

$$\boldsymbol{y}_i = \boldsymbol{A}_i\boldsymbol{x} + \boldsymbol{e}, \quad \text{with} \quad \boldsymbol{A}_i^{\mathsf{H}} = \begin{bmatrix} \boldsymbol{A}_{i0}^{\mathsf{H}} \\ \vdots \\ \boldsymbol{A}_{iJ}^{\mathsf{H}} \end{bmatrix}, \; \boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_0 \\ \vdots \\ \boldsymbol{x}_J \end{bmatrix} \tag{6}$$

where the operator $(\cdot)^{\mathsf{H}}$ denotes the conjugate transpose, $\boldsymbol{y}_i \coloneqq \boldsymbol{I}_i - \boldsymbol{I}_i^{ii} \in \mathbb{R}^N$ is the measured intensity with the removal of the background intensity for the i^{th} illumination, and $\boldsymbol{e} \in \mathbb{R}^N$ is the error term. Note that, as was discussed in [27], the IDT forward model does not contain any information on the DC component of the phase.

2.2 Inverse Problem

Since image reconstruction in optical tomography is often ill-posed, it is typically formulated as the regularized inversion problem

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}\in\mathbb{C}^N}{\mathsf{arg\,min}} \left\{ g(\boldsymbol{x}) + h(\boldsymbol{x}) \right\}, \tag{7}$$

where g is the data-fidelity term that ensures the consistency with the measured data, and h is the regularization term that imposes the prior knowledge on the desired image. For example, the Tikhonov regularization [30] assumes a Gaussian prior on the unknown image. It has been previously used in IDT for deriving a closed form solution [27]. More recent regularizers, such as the sparsity-promoting ℓ_1-norm penalty [31] and the edge-preserving total variation (TV) [32], are nonsmooth and do not have closed-form solutions, thus requiring iterative algorithms for image formation. In particular, the family of proximal methods—such as proximal gradient method (PGM) [33–36] and alternating direction method of multipliers (ADMM) [37–40]—avoid the need to differentiate the regularizer by using the proximal map [41].

Recently, deep learning has gained popularity in imaging inverse problems [42–50]. Traditional strategy trains the convolutional neural network (CNN) to learn the direct mapping from the measurements to some ground-truth image. Despite their excellent performance in some image reconstruction problems, this strategy does not leverage the known physics of the imaging system and does not insure consistency with the measured data. In this paper, we propose SIMBA to reconcile the model-based and learning-based approaches by infusing deep denoising priors into online iterative algorithms.

2.3 Regularization by Denoising

RED [22] is a recently introduced framework to leverage powerful image denoisers. It has been successfully applied in many regularized imaging tasks, including image deblurring [22], super-resolution [25], and phase

Algorithm 1 SIMBA

input: $\boldsymbol{x}^0 \in \mathbb{R}^n$, $\tau > 0$, $\sigma > 0$, and $B \geq 1$
for $k = 1, 2, \ldots$ **do**
  $\widehat{\nabla} g(\boldsymbol{x}^{k-1}) \leftarrow \mathsf{minibatchGradient}(\boldsymbol{x}^{k-1}, B)$
  $\widehat{\mathsf{G}}(\boldsymbol{x}^{k-1}) \leftarrow \widehat{\nabla} g(\boldsymbol{x}^{k-1}) + \tau(\boldsymbol{x}^{k-1} - \mathsf{D}_\sigma(\boldsymbol{x}^{k-1}))$
  $\boldsymbol{x}^k \leftarrow \boldsymbol{x}^{k-1} - \gamma \widehat{\mathsf{G}}(\boldsymbol{x}^{k-1})$

retrieval [24]. The framework aims to find a fixed point $\boldsymbol{x}^*$ that satisfies

$$\mathsf{G}(\boldsymbol{x}^*) = \nabla g(\boldsymbol{x}^*) + \tau(\boldsymbol{x}^* - \mathsf{D}_\sigma(\boldsymbol{x}^*)) = 0, \tag{8}$$

where ∇g denotes the gradient of g, D_σ is the image denoiser, and $\tau > 0$ adjusts the tradeoff between the data-fidelity and the prior. RED algorithms seek a vector $\boldsymbol{x}^*$ that lies in the zero set of $\mathsf{G} : \mathbb{R}^n \to \mathbb{R}^n$

$$\boldsymbol{x}^* \in \mathsf{zer}(\mathsf{G}) \coloneqq \{\boldsymbol{x} \in \mathbb{R}^n : \mathsf{G}(\boldsymbol{x}) = 0\}. \tag{9}$$

For example, the gradient-method variant of RED (denoted as GM-RED) can be implemented as

$$\begin{aligned} \boldsymbol{x}^k &\leftarrow \boldsymbol{x}^{k-1} - \gamma(\nabla g(\boldsymbol{x}^{k-1}) + \mathsf{H}(\boldsymbol{x}^{k-1})) \\ \text{where} \quad \mathsf{H}(\boldsymbol{x}) &\coloneqq \tau(\boldsymbol{x} - \mathsf{D}_\sigma(\boldsymbol{x})). \end{aligned} \tag{10}$$

Here, the parameter $\gamma > 0$ is the step-size. When the denoiser D_σ is locally homogeneous and has a symmetric Jacobian [22, 23], the operator H corresponds to the gradient of the following regularizer

$$h(\boldsymbol{x}) = \frac{\tau}{2} \boldsymbol{x}^\mathsf{T}(\boldsymbol{x} - \mathsf{D}_\sigma(\boldsymbol{x})). \tag{11}$$

By having a closed-form objective function, one can use the classical optimization theory to analyze the convergence of RED algorithms [22]. On the other hand, fixed-point convergence has also been established without having an explicit objective function [20, 23]. Reehorst *et al.* [23] have shown that RED proximal gradient methods (RED-PG) converges to a fixed point by utilizing the monotone operator theory. Sun *et al.* [51] have established the explicit convergence rate for the block coordinate variant of RED (BC-RED) under a nonexpansive D_σ. In this paper, we extend these prior analyses to the randomized processing of the measurements instead of image blocks, which opens up applications to tomographic imaging with a large number of projections.

3 Proposed Method

We now introduce SIMBA that combines the iterative usage of the forward model with a deep denoising prior. At each iteration, SIMBA updates $\boldsymbol{x}$ by combining a stochastic gradient for increasing data-consistency with a CNN denoiser for artifact reduction. SIMBA is ideal for data-intensive biomedical imaging applications where the object features are difficult to characterize using traditional regularizers.

3.1 Iterative Online Procedure

In IDT, the data-fidelity term can be written as an average over a set of distinct components functions

$$g(\boldsymbol{x}) = \frac{1}{I} \sum_{i=1}^{I} g_i(\boldsymbol{x}), \tag{12}$$

where each component function g_i is evaluated only on the subset $\mathbf{y}_i$ of the full measurements $\mathbf{y}$

$$g_i(\boldsymbol{x}) = \mathcal{L}(\boldsymbol{y}_i, \boldsymbol{A}_i\boldsymbol{x}) \;, \tag{13}$$

where $\mathcal{L}$ is a smooth loss function quantifying the discrepancy between the predicted measurement $\boldsymbol{A}_i\boldsymbol{x}$ and the actual measurements $\boldsymbol{y}_i$. For example, all the results in this paper were obtained using

$$g_i(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{y}_i - \boldsymbol{A}_i\boldsymbol{x}\|_2^2 \quad \Rightarrow \quad \nabla g_i(\boldsymbol{x}) = \boldsymbol{A}_i^{\mathsf{H}}(\boldsymbol{A}_i\boldsymbol{x} - \boldsymbol{y}_i).$$

The computation of the gradient of g

$$\nabla g(\boldsymbol{x}) = \frac{1}{I}\sum_{i=1}^{I} \nabla g_i(\boldsymbol{x}), \tag{14}$$

is proportional to the total number of illuminations I. A large I effectively precludes the applicability of the batch RED algorithms due to the computational cost of evaluating ∇g.

SIMBA, summarized in Algorithm 1, improves scalability through partial randomized processing of gradient components ∇g_i via the following *minibatch* approximation of the gradient

$$\widehat{\nabla} g(\boldsymbol{x}) = \frac{1}{B}\sum_{b=1}^{B} \nabla g_{i_b}(\boldsymbol{x}), \tag{15}$$

where $i_1, \ldots, i_B$ are independent random indices that are distributed uniformly over $\{1, \ldots, I\}$. Due to its ability to control the minibatch size $1 \leq B \leq I$, SIMBA benefits from considerable *flexibility* for trading off different practical considerations compared to the batch RED algorithms. For example, one can consider using small minibatches ($B \ll I$) for problems where ∇g is computationally expensive and D_σ is relatively efficient. On the other hand, one can consider larger minibatches for problems where D_σ is relatively slow compared to ∇g. In this paper we focus on image denoisers corresponding to deep neural nets, thus obtaining D_σ that is both fast and effective for many practical imaging problems.

3.2 CNN-based Denoiser

In recent years, CNNs have been shown to achieve the state-of-the-art performance on image denoising [11,52]. We propose a simple denoising network DnCNN* as the deep learning module in SIMBA. The architecture of the neural network, illustrated in Figure 1, is adapted from the popular DnCNN. In general, DnCNN* consists of two parts. The first part contains $N_\ell - 1$ sequential composite convolutional layers, each of which has one convolutional layer followed by a rectified linear unit (ReLU) layer. The second part is a single convolutional layer that outputs the final denoised image, resulting the total number of layers in DnCNN* to be N_ℓ. All the convolution filters are implemented with size 3×3, and every feature map has 64 channels. In SIMBA, we apply this 2D image denoising network to the 3D sample by performing the layer-by-layer denoising along the axial direction z.

We generated the training dataset by adding AWGN to the natural images from BSD400 and applying standard data augmentation strategies including flipping, rotating, and rescaling. Note that our training dataset does not include any biomedical image. We employed the residual learning technique [53] in DnCNN* so that the network is forced to learn the noise residual in the noisy input. DnCNN* was trained to minimize the following loss

$$\mathcal{L}_\theta = \frac{1}{p}\sum_{i=1}^{p} \left\{ \|f_\theta(\boldsymbol{x}_i) - \boldsymbol{y}_i\|_2^2 + \rho\|f_\theta(\boldsymbol{x}_i) - \boldsymbol{y}_i\|_1 \right\}, \tag{16}$$

where $\boldsymbol{x}_i$ is the noisy input, $\boldsymbol{y}_i$ is the noise, and $f_\theta(\boldsymbol{x})$ represents the noise predicted by the neural network. Eq. (16) penalizes both the mean squared error (MSE) and the mean absolute error (MAE) between the

Figure 2: Eight test images used in the experiments. Top row from left to right: *Aircraft*, *Boat*, *Cameraman*, *Foreman*. Bottom row from left to right: *House*, *Monarch*, *Parrot*, *Pirate*.

estimated noise and the ground truth. A *loss parameter* $\rho > 0$ is thus introduced to adjust the tradeoff between the two errors for the best training performance. Our results show that our simple DnCNN* is competitive with traditional denoisers in terms of the imaging quality.

4 Convergence Analysis

Our analysis relies on the fixed-point convergence of averaged operators, which is well known as the Krasnosel'skii-Mann theorem [54]. Here, we extend the result to the iterative online algorithms under the RED formulation and show the worst-case convergence rates. Note that our analysis does not assume that the denoiser corresponds to any explicit RED regularizer. We first introduce the assumptions necessary for our analysis and then present the main results.

Assumption 1. *We make the following assumptions on the data-fidelity term g:*

(a) The component functions g_i are all convex and differentiable with the same Lipschitz constant $L > 0$.
(b) At every iteration, the gradient estimate is unbiased and has a bounded variance:

$$\mathbb{E}\left[\widehat{\nabla} g(\boldsymbol{x})\right] = \nabla g(\boldsymbol{x}), \;\; \mathbb{E}\left[\left\|\nabla g(\boldsymbol{x}) - \widehat{\nabla} g(\boldsymbol{x})\right\|_2^2\right] \leq \frac{\nu^2}{B},$$

for some constant $\nu > 0$.

Assumption 1 (a) implies that the overall data-fidelity g is also convex and has Lipschitz continuous gradient with constant L. Assumption 1 (b) assumes that the minibatch gradient is an unbiased estimate of the full gradient. The bounded variance assumption is a standard assumption used in the analysis of online and stochastic algorithms [55–57]

Assumption 2. *The operator G is such that $\mathsf{zer}(\mathsf{G}) \neq \varnothing$.*

Assimption 2 is a mild assumption that simply asserts the existence of a solution to (8). It is related to the existence of stationary points in traditional smooth optimization [58].

Assumption 3. *Given $\sigma > 0$, the denoiser D_σ is a nonexpansive operator such that*

$$\|\mathsf{D}_\sigma(\boldsymbol{x}) - \mathsf{D}_\sigma(\boldsymbol{y})\|_2 \leq \|\boldsymbol{x} - \boldsymbol{y}\|_2 \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n,$$

Nonexpansive variants of several widely used denoisers, such as NLM and DnCNN, have been developed in [12, 14, 51, 59]. Under the above assumptions, we can establish the following for SIMBA.

Table 1: List of parameters of the experimental setup

Experimental parameters		Simulations (5.2)	Experiments (5.3)
λ	wavelength of LED light	630 nm	630 nm
ϵ_b	background medium index	1.33	1.33
z_{LED}	axial position of LEDs	-70 mm	-79 mm
z	axial position of the sample	0 µm	$(-20, 100)$ µm
MO	microscope objectives	$40\times$	$10\times$
NA	numerical aperture	0.65	0.25

Table 2: List of algorithmic hyperparameters

Hyperparameters		Simulations (5.2)	Experiments (5.3)
$\boldsymbol{x}^0$	initial point of reconstructions	$\mathbf{0}$	$\mathbf{0}$
B	minibatch size	20	10
I	batch size	60	89
γ	step size	$\frac{1}{L+2\tau}$	$\frac{1}{L+2\tau}$
σ	input noise level for DnCNN*	10	5
ρ	loss function parameter	0	1
N_ℓ	number of layers in DnCNN*	7	10
τ	level of regularization in GM-RED and SIMBA	optimized for each image	optimized for the dataset

Theorem 1. *Run SIMBA for $t \geq 1$ iterations under Assumptions 1-3 using a fixed step-size $0 < \gamma \leq 1/(L+2\tau)$ and a fixed minibatch size $B \geq 1$. Then, we have*

$$\mathbb{E}\left[\frac{1}{t}\sum_{k=1}^{t}\|\mathsf{G}(\boldsymbol{x}^{k-1})\|_2^2\right] \leq (L+2\tau)\left[\frac{\|\boldsymbol{x}^0-\boldsymbol{x}^*\|_2^2}{\gamma t}+\frac{\gamma\nu^2}{B}\right].$$

Proof. See Appendix B. □

Thereom 1 is analogous to the convergence of the *minibatch stochastic gradient descent (SGD)* [60] in the sense that the convergence can be established up to an error term that depends on γ and B. Similarly, the accuracy of the expected convergence of SIMBA to $\mathsf{zer}(\mathsf{G})$ improves with smaller γ and larger B. For example, by setting $\gamma = 1/[(L+2\tau)\sqrt{t}]$, we get

$$\mathbb{E}\left[\min_{k\in\{1,\dots,t\}}\|\mathsf{G}(\boldsymbol{x}^{k-1})\|_2^2\right] \leq \mathbb{E}\left[\frac{1}{t}\sum_{k=1}^{t}\|\mathsf{G}(\boldsymbol{x}^{k-1})\|_2^2\right] \leq \frac{C}{\sqrt{t}},$$

where C is some positive constant.

Finally, note that the analysis in Theorem 1 only provides *sufficient conditions* for the convergence of SIMBA. As corroborated by our numerical studies in Section 5, the actual convergence of SIMBA is more general and often holds beyond nonexpansive denoisers (such as BM4D). One plausible explanation for this is that such denoisers are *locally nonexpansive* over the set of input vectors used in testing (see also discussion in [59]). On the other hand, the recent techniques for spectral-normalization of deep neural nets [61–63] provide a convenient tool for building *globally nonexpansive* neural denoisers that result in provable convergence of SIMBA.

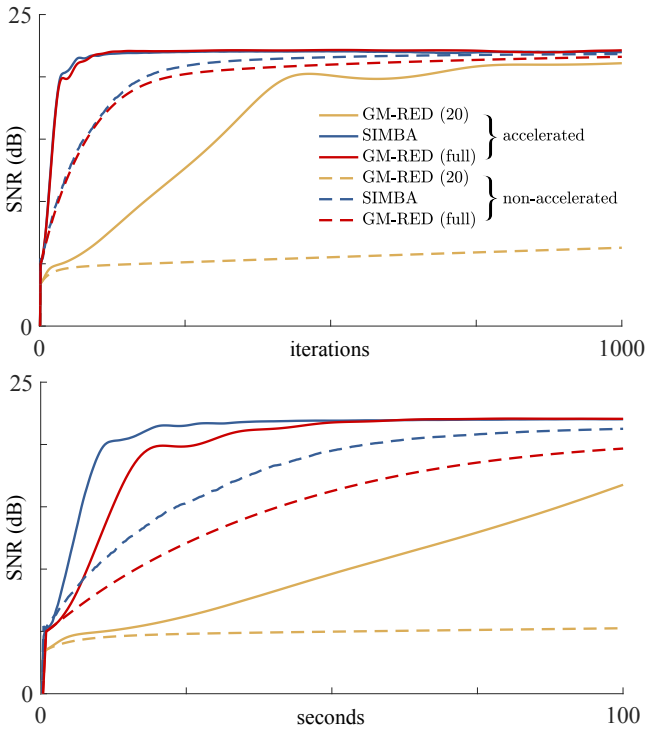

Figure 3: Illustration of convergence in SNR of SIMBA with minibatch size $B = 20$ under the DnCNN* denoiser. The top and bottom figures plot the SNR values against the number of itartions and running time, respectively. Two batch algorithms, GM-RED (20) and GM-RED (full), are plotted for comparison. Under the same per-iteration complexity, SIMBA converges to significantly higher SNR than GM-RED (20) due to its actual usage of the full data. Moreover, online processing makes SIMBA converge significantly faster than GM-RED (full). The acceleration is due to the lower computational cost of processing a small random subset of the full data. The same trend is observed for both accelerated and normal versions of the algorithms.

5 Experimental Validation

In this section, we validate SIMBA on both simulated and experimental data. We first numerically demonstrate the efficiency and practical convergence of SIMBA in simulations. Next, we apply SIMBA to reconstruct a 3D model from a set of real intensity-only measurements. Our results highlight the applicability and effectiveness of SIMBA for the iterative inversion in optical tomography.

5.1 Setup

In simulations, we reconstruct eight grayscale natural images, representing the phase component of the complex permittivity contrast, displayed in Figure 2. They are assumed to be on the focal plane $z = 0\,\mu\text{m}$ with LEDs located at $z_{\text{LED}} = -70\,\text{mm}$. We generate $I = 60$ simulated intensity measurements with $40\times$ microscope objectives (MO) and 0.65 numerical aperture (NA). All simulated measurements are corrupted by AWGN corresponding to 20 dB of *input signal-to-noise ratio (SNR)*. As a quantitative metric for measuring the quality of reconstructions, we use the SNR defined as follows

$$\text{SNR}(\hat{\boldsymbol{y}}, \boldsymbol{y}) \triangleq \max_{a,b\in\mathbb{R}} \left\{ 20\log_{10}\left(\frac{\|\boldsymbol{y}\|_{\ell_2}}{\|\boldsymbol{y} - a\hat{\boldsymbol{y}} + b\|_{\ell_2}} \right) \right\}$$

Table 3: Optimized SNR for each test image in dB

Algorithms	GM (20)	SGM	GM (full)	GM-RED (20)		SIMBA		GM-RED (full)	
Denoisers	—	—	—	BM3D	DnCNN*	BM3D	DnCNN*	BM3D	DnCNN*
Aircraft	17.44	18.00	18.01	18.82	19.85	19.65	20.48	19.47	20.44
Boat	18.09	18.78	18.82	19.96	20.40	21.23	21.29	21.12	21.55

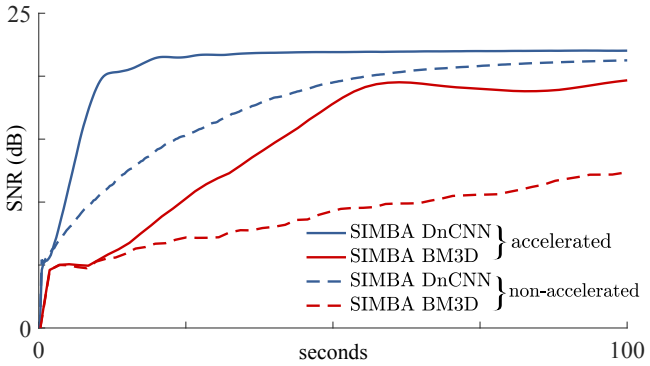

Figure 4: Illustration of the convergence speed of SIMBA with minibatch size $B = 20$ under the DnCNN* and BM3D denoisers. The figure plots the SNR values against the running time in seconds. SIMBA DnCNN is significantly faster due to the fast GPU implementation of the denoiser.

where $\hat{\boldsymbol{y}}$ represents the noisy vector and $\boldsymbol{y}$ denotes the ground truth. In experiments, we recover a 3D algae sample from real IDT measurements. The 3D sample is located over the range $(-20, 100)$ µm and $z_{\text{LED}} = -79\,\text{mm}$. We set the slice spacing as $5\,\mu\text{m}$, so each slice represents the average over the sample thickness. We take $I = 89$ measurements with $10\times$ MO and 0.25 NA for reconstruction. We refer to Table 1 for the detailed summary of the experimental parameters. All experiments in this paper were performed on a machine equipped with an Intel Xeon E5-2620 v4 Processor that has 4 cores of 2.1 GHz and 256 GBs of DDR memory. We trained all neural nets using NVIDIA RTX 2080 GPUs.

The algorithmic hyperparameters are summarized in Table 2. All algorithms start from $\boldsymbol{x}^0 = \boldsymbol{0}$. We trained DnCNN* for the removal of AWGN at four noise levels corresponding to $\sigma \in \{5, 10, 15, 20\}$. The same set of σ is used for BM3D. All algorithmic parameters are optimized for the best performance.

5.2 Simulated Data

In this section, we numerically illustrate the advantages of SIMBA in tomographic imaging over the batch GM-RED. The advantages are: (1) better SNR under a limited memory budget; (2) better time efficiency when all the measurements are used.

Figure 3 (top) plots the average SNR over test images against the iteration number for SIMBA and GM-RED (20), both using DnCNN* as the denoiser. GM-RED (20) uses a fixed set of 20 (out of 60) measurements, while SIMBA selects a random subset of 20 at every iteration. Under the same computational complexity, SIMBA achieves a SNR boost of about 1 dB over GM-RED (20) because the former has access to all the measurements. Visual examples are presented in Figure 5. As a reference, we also plot the SNR for

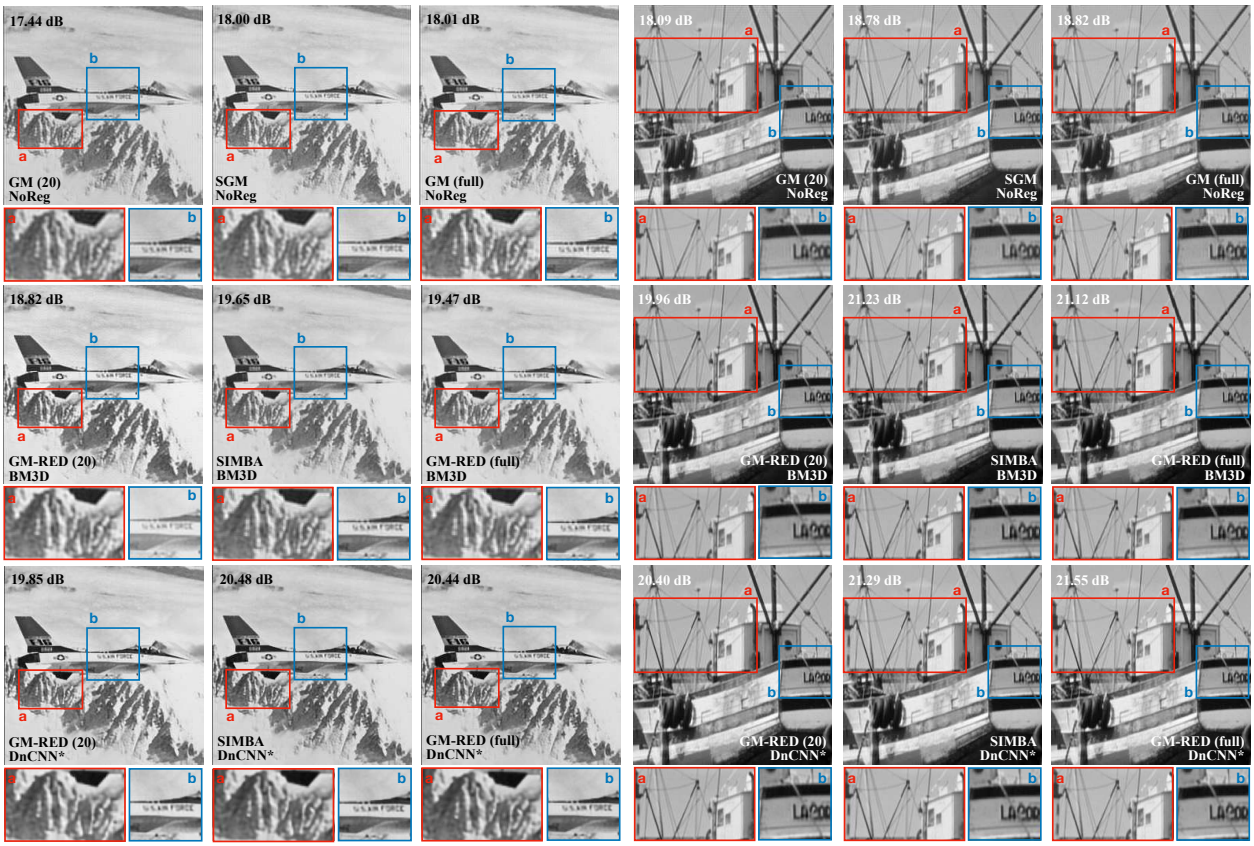

Figure 5: Visual examples of reconstructed *Aircrafts* (left) and *Boat* (right) images by different algorithms. Three columns correspond to algorithms using fixed 20, random 20 out of 60, and full 60 measurements, respectively. The first row presents the unregularized results and the second and third row show the results given by a well-known BM3D denoiser and a state-of-the-art deep learning prior, respectively. Differences are zoomed in using boxes inside the images. Each image is labeled by its SNR (dB) with respect to the original image. Note that our proposed algorithm SIMBA recovers the details lost by the batch algorithm with the same computational cost and achieves the same level of SNR and visual quality as the full batch algorithm.

GM-RED using all 60 measurements, denoted as GM-RED (full).

Figure 3 (bottom) highlights the faster time convergence of SIMBA compared to GM-RED (full) to the same level of SNR. Figure 5 highlights that the SNR values and the visual quality obtained by SIMBA and GM-RED (full) are nearly identical. SIMBA, however, significantly reduces the reconstruction time by processing one third of all measurements at each iteration. Specifically, the average per-iteration times of GM-RED (20), SIMBA, and GM-RED (full) are 0.30 second, 0.31 second, and 0.52 second, respectively. We also note that by processing only a subset of measurements, SIMBA leads to a more favorable tradeoff between computational cost and memory compared to GM-RED (full). This makes SIMBA beneficial for processing datasets containing a large number of tomographic measurements. The convergence speed of SIMBA can be significantly improved by using deep learning denoisers implemented on GPUs. This is highlighted in Figure 4, where SIMBA DnCNN* is compared against SIMBA BM3D.

Table 3 shows final SNRs of all reconstructions we performed. We run all simulations using the accelerated versions of these algorithms, which are analogous to the accelerated gradient method by Nesterov [58]. Empirically, they converge to the same solution as the non-accelerated counterparts. For reference, we show the evolution of SNR for non-accelerated versions by the dotted lines in Figure 3. Table 3 shows that our DnCNN denoiser has higher average SNR than BM3D. The compatibility of SIMBA with DnCNN*, which is

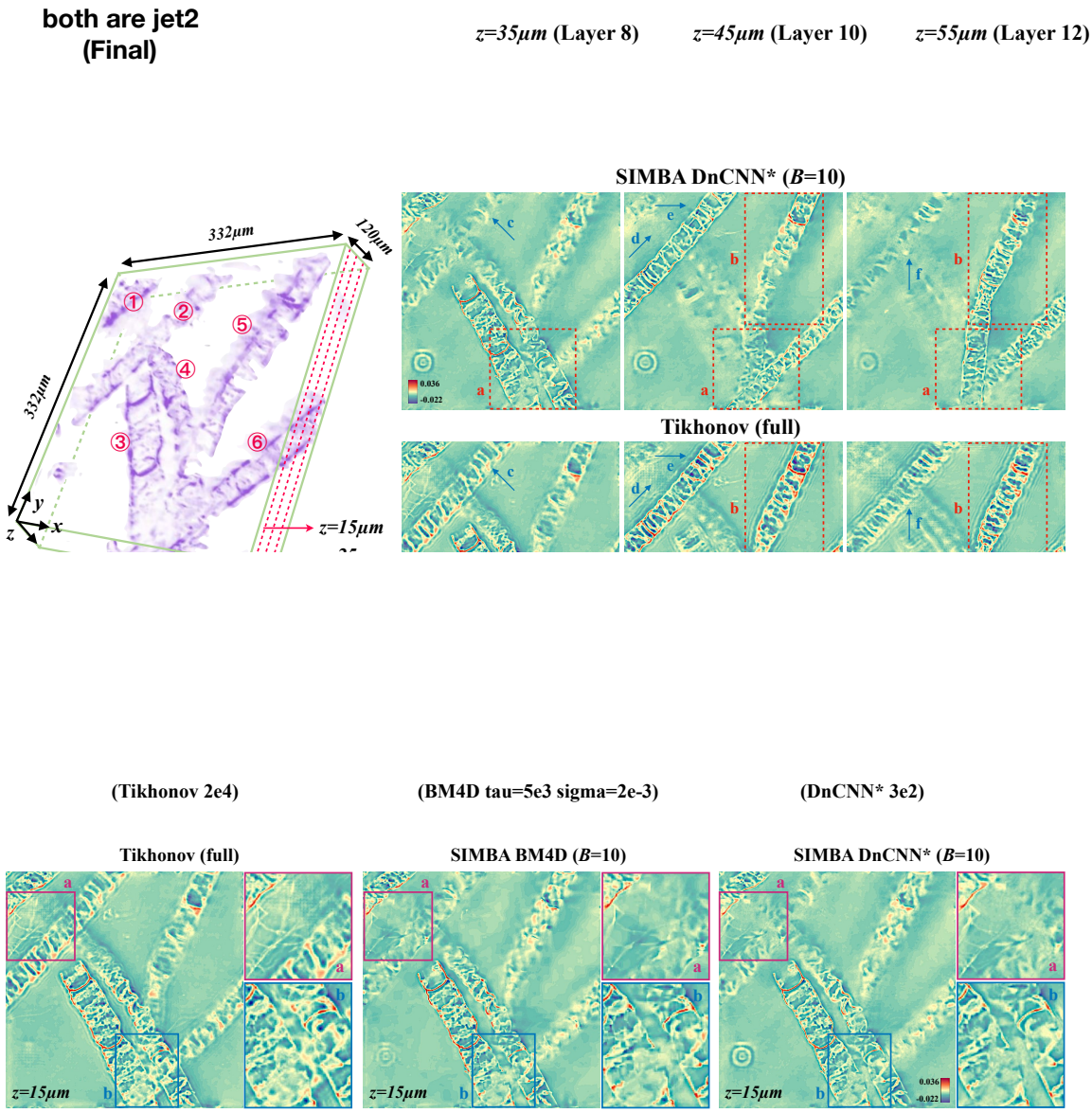

Figure 7: Comparison of SIMBA under BM4D and DnCNN* against Tikhonov. Regions (a) and (b) are zoomed in to highlight visual differences. Tikhonov reconstructed image contains grid-shape artifacts and interfering contents from other slices, while BM4D generates blocks and nonsmoothness. SIMBA under DnCNN* produces the most real recovery with the clearest shape of the algae.

a low-complexity denoiser, increases the potential of applying SIMBA to large scale image reconstructions.

5.3 Experimental IDT Dataset

In this section, we use SIMBA to reconstruct a 3D algae sample of $1024 \times 1024 \times 25$ pixels from 89 high-resolution measurements. The large sample volume dramatically increase the memory usage and computational cost, and prohibits the applicability of the full batch algorithms. Experimental results show that SIMBA successfully overcomes these difficulties by processing a small subset of all measurements ($B = 10$) at every iteration and leads to significant performance improvements compared to the method reported in [27].

Figure 6 provides a 3D visualization of the phase component of the image recovered by SIMBA, with different algae labeled by circled numbers (there are 6 of them). Figure 7 compares three slices of our SIMBA results and the Tikhonov (full) results obtained by algorithm in [27], which uses all 89 measurements. As discussed in [27], the DC component of the phase is lost in the IDT forward model, we thus set the mean of

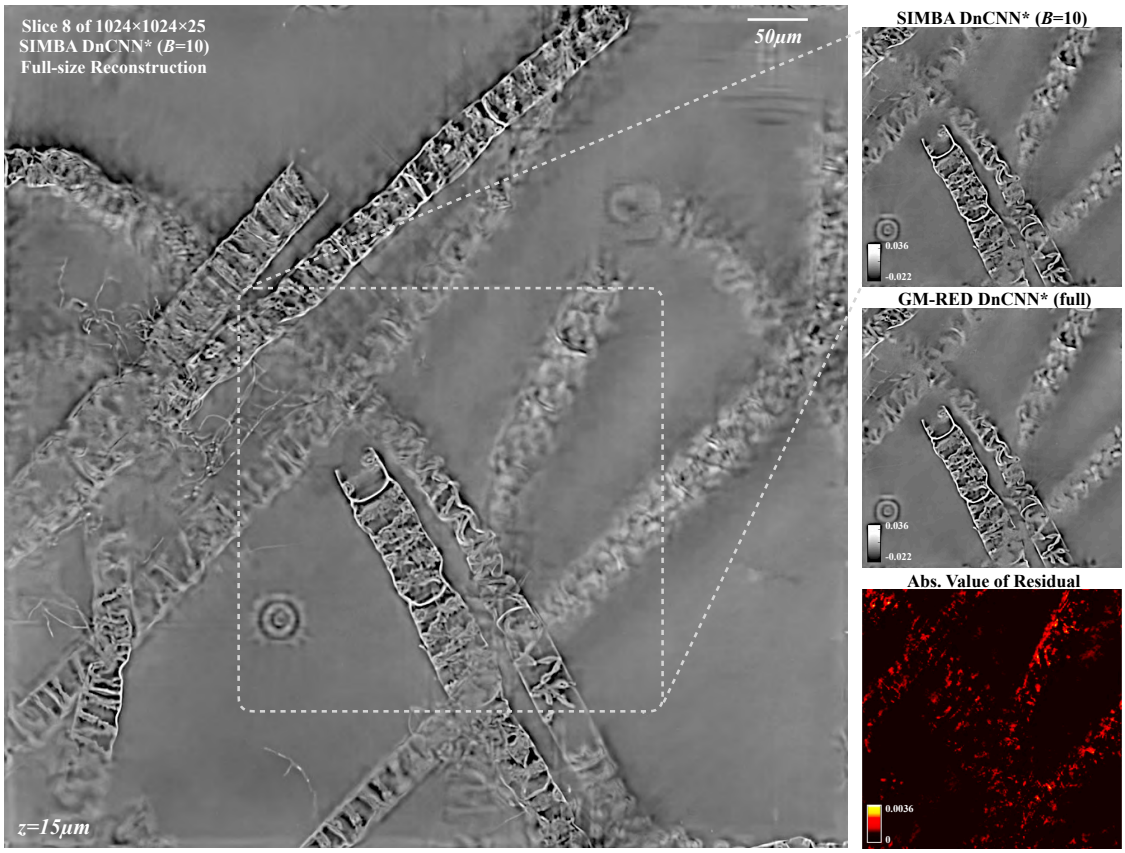

Figure 8: A slice from the full $1024 \times 1024 \times 25$ reconstruction by SIMBA under DnCNN*. On the right is a comparison between SIMBA and the full batch results for the dotted region. The two reconstructions are visually indistinguishable, and the absolute value of the residual between them highlights the numerical proximity of SIMBA to the full batch reconstruction. Note the small numerical scale of the residual compared to that of the two reconstructions.

all the results to the one of the Tikhonov reconstruction for a more uniform comparison. We evaluate the quality of different reconstructions by comparing their axial sectioning effect and the ability to eliminates artifacts. In the 3D tomographic model with strong sectioning effect, a pattern emerges only in the slice it belongs to and fades away as we go axially to different depths. Sectioning enables us to better predict the axial location of the patterns within a 3D object and thus better understand its internal structure, which is crucial for biomedical imaging applications. While Tikhonov regularization is attractive from computational perspective, it is known that to lead to excessive smoothing. This complicates the understanding of the axial structure of the sample. On the other hand, by leveraging the DnCNN* prior, SIMBA improves the performance, while also mitigating the computational complexity with online processing. Our results show that SIMBA with DnCNN* enables better sectioning of the object compared to the Tikhonov prior. For example, maintaining the clarity and sharpness of algae ② in slice $z = 25\,\mu\text{m}$, SIMBA successfully reduces the artifacts generated by the content of adjacent slices, which exist in the region (a) of Tikhonov. In the other two slices, algae ② fades away and does not generate strong shadowy artifacts as indicated by arrows (c) and (f). By horizontally comparing the two rows, the algae cluster in region (a) is visually better resolved by SIMBA than Tikhonov. Moreover, in SIMBA reconstructions, the top half of algae ⑤ in region (b) looks sharp in slice $z = 25\,\mu\text{m}$ and the bottom half appears clear in slice $z = 35\,\mu\text{m}$. This inter-slice information implies that algae ⑤ penetrate through $z = 25\,\mu\text{m}$ and $z = 35\,\mu\text{m}$. However, the whole structure of algae ⑤ is present in both slices of Tikhonov reconstructions, which fails in illustrating the axial position. Note that SIMBA also better eliminates artifacts pointed out by arrows (d) and (e). To further analyze the performance

Table 4: Per-iteration memory usage specification for processing our experimental data

Algorithms			SIMBA (10)		GM-RED (full)	
Variables		Data type	Shape	Size	Shape	Size
$\{\boldsymbol{A}_i\}$	phase	complex64	$1024\times1024\times25\times10$	3.91 Gb	$1024\times1024\times25\times89$	34.77 Gb
	absorption	complex64	$1024\times1024\times25\times10$	3.91 Gb	$1024\times1024\times25\times89$	34.77 Gb
$\{\boldsymbol{y}_i\}$		complex128	$1024\times1024\times10$	0.31 Gb	$1024\times1024\times89$	2.78 Gb
others combined		—	—	3.13 Gb	—	3.13 Gb
Total		—	—	**11.26 Gb**	—	75.45 Gb

of the priors, we bring BM4D, the 3D version of the well-known denoiser BM3D, into comparison. In zoom-in region (a) of Figure 7, Tikhonov reconstruction contains grid-shape artifacts. BM4D generates small blocks due to its block-matching mechanism. DnCNN* provides a more real and sharper result than the other two. In region (b), Tikhonov reconstruction is of satisfactory visual quality but the shadow of algae ⑤ and ⑥ in the background interferes with the actual content in this slice. BM4D erases the shadow in the background but it again generates blocky artifacts which makes its reconstruction not as real as DnCNN* result.

Finally, we present one slice of the full $1024\times1024\times25$ reconstruction by SIMBA under DnCNN* in Figure 8. For comparison, we run GM-RED (full) under DnCNN* but only for the dotted region because of the high computational cost of the full batch reconstruction. The result is juxtaposed with our SIMBA result. These two algorithms are run with the same τ value until convergence. Visually, they look almost identical and we present the absolute value of the residual between the two for reference. The residual is negligible compared to the numerical scale of the two results. Quantitatively, if we assume the result of the full batch algorithm to be the "ground truth", the SNR of SIMBA is 47.03 dB. This substantiates that SIMBA sufficiently matches the full batch algorithms in terms of the final reconstruction quality. Specifically, the average per-iteration running time of SIMBA for reconstructing the dotted region is 22 seconds, while that of GM-RED is 192 seconds, which corresponds to a 9× speed-up. SIMBA also requires less memory at every iteration by processing only about one ninths of full measurements. The reduced running time and memory usage in processing such an intensive amount of data highlighted the efficiency improvement of SIMBA compared to the traditional batch GM-RED.

We would also like to note that the memory considerations in image reconstruction must take into account the size of all the variables related to the image volume $\boldsymbol{x}$, the measured data $\{\boldsymbol{y}_i\}$, and the measurement operators $\{\boldsymbol{A}_i\}$. The goal of this paper is to address the problems where the bottleneck is in the storage and processing of the measurements and measurement operators. Table 4 records the total memory (Gb) used by SIMBA and GM-RED (full) in each iteration. Note that our implementation stores each $\boldsymbol{A}_i$ as two separate matrices for phase and absorption. Additionally, each matrix is stored in the Fourier space to reduce computational complexity of computing convolutions. This results in the storage of complex valued arrays for each, consisting of pairs of double precision floats for every element. While GM-RED (full) requires 75.45 Gb of memory due to its processing of all measurements in every iteration, SIMBA uses only 11.26 Gb, about one-seventh of the full volume, which makes the algorithm particularly well suited to tomographic applications where one needs to process a very large number of views.

6 Conclusion

We proposed an extension of RED for solving imaging inverse problems in optical tomography. Our method is scalable to large measurements and uses a deep denoising prior to improve the final estimate. We proved the fixed-point convergence of the method without assuming an explicit objective function, which complements the current theoretical analysis of RED for large-scale image reconstruction. We validated the method on both simulated and experimental IDT data. Especially, the 3D reconstruction of a large algae sample

fully elucidates the benefits of our method in data-intensive imaging problems. Future work includes the application of SIMBA in other advanced IDT modalities with coded illumination patterns [64] and accelerated data acquisition [65].

Acknowledgements

This work was partially supported by grants NSF CCF-1813910.

References

[1] 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.

[2] 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.

[3] 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.

[4] G. Gbur and E. Wolf, "Diffraction tomography without phase information," *Opt. Lett.*, vol. 27, no. 21, pp. 1890–1892, Nov 2002.

[5] L. Tian, J. Wang, and L. Waller, "3D differential phase-contrast microscopy with computational illumination using an led array," *Opt. Lett.*, vol. 39, no. 5, pp. 1326–1329, Mar 2014.

[6] 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.

[7] A. Ribés and F. Schmitt, "Linear inverse problems in imaging," *IEEE Signal Process. Mag.*, vol. 25, no. 4, pp. 84–99, July 2008.

[8] 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.

[9] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proc. IEEE Global Conf. Signal Process. and Inf. Process. (GlobalSIP)*, Austin, TX, USA, December 3-5, 2013, pp. 945–948.

[10] A. Danielyan, "Block-based collaborative 3-D transform domain modeing in inverse imaging," Publication 1145, Tampere University of Technology, May 2013.

[11] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, July 2017.

[12] S. Sreehari, S. V. Venkatakrishnan, B. Wohlberg, G. T. Buzzard, L. F. Drummy, J. P. Simmons, and C. A. Bouman, "Plug-and-play priors for bright field electron tomography and sparse interpolation," *IEEE Trans. Comput. Imaging*, vol. 2, no. 4, pp. 408–423, December 2016.

[13] S. H. Chan, X. Wang, and O. A. Elgendy, "Plug-and-play ADMM for image restoration: Fixed-point convergence and applications," *IEEE Trans. Comp. Imag.*, vol. 3, no. 1, pp. 84–98, March 2017.

[14] A. M. Teodoro, J. M. Bioucas-Dias, and M. Figueiredo, "A convergent image fusion algorithm using scene-adapted Gaussian-mixture-based denoising," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 451–463, Jan. 2019.

[15] A. Brifman, Y. Romano, and M. Elad, "Turning a denoiser into a super-resolver using plug and play priors," in *Proc. IEEE Int. Conf. Image Proc. (ICIP)*, Phoenix, AZ, USA, September 25-28, 2016, pp. 1404–1408.

[16] A. M. Teodoro, J. M. Biocas-Dias, and M. A. T. Figueiredo, "Image restoration and reconstruction using variable splitting and class-adapted image priors," in *Proc. IEEE Int. Conf. Image Proc. (ICIP)*, Phoenix, AZ, USA, September 25-28, 2016, pp. 3518–3522.

[17] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Honolulu, USA, July 21-26, 2017, pp. 3929–3938.

[18] T. Meinhardt, M. Moeller, C. Hazirbas, and D. Cremers, "Learning proximal operators: Using denoising networks for regularizing inverse imaging problems," in *Proc. IEEE Int. Conf. Comp. Vis. (ICCV)*, Venice, Italy, October 22-29, 2017, pp. 1799–1808.

[19] U. S. Kamilov, H. Mansour, and B. Wohlberg, "A plug-and-play priors approach for solving nonlinear imaging inverse problems," *IEEE Signal. Proc. Let.*, vol. 24, no. 12, pp. 1872–1876, December 2017.

[20] Y. Sun, B. Wohlberg, and U. S. Kamilov, "An online plug-and-play algorithm for regularized image reconstruction," *IEEE Trans. Comput. Imaging*, 2019.

[21] Y. Sun, S. Xu, Y. Li, L. Tian, B. Wohlberg, and U. S. Kamilov, "Regularized Fourier ptychography using an online plug-and-play algorithm," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP)*, Brighton, UK, May 12-17, 2019, pp. 7665–7669.

[22] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (RED)," *SIAM J. Imaging Sci.*, vol. 10, no. 4, pp. 1804–1844, 2017.

[23] E. T. Reehorst and P. Schniter, "Regularization by denoising: Clarifications and new interpretations," *IEEE Trans. Comput. Imag.*, vol. 5, no. 1, pp. 52–67, Mar. 2019.

[24] C. Metzler, P. Schniter, A. Veeraraghavan, and R. Baraniuk, "prDeep: Robust phase retrieval with a flexible deep network," in *Proc. 35th Int. Conf. Machine Learning (ICML)*, Stockholmsmässan, Stockholm Sweden, 10–15 Jul 2018, pp. 3501–3510.

[25] G. Mataev, P. Milanfar, and M. Elad, "DeepRED: Deep image prior powered by RED," in *The IEEE International Conference on Computer Vision (ICCV) Workshops*, Oct 2019.

[26] A.-K. Seghouane, Y. Moudden, and G. Fleury, "Regularizing the effect of input noise injection in feedforward neural networks training," *Neural Comput. Applic.*, vol. 13, no. 3, pp. 248–254, 2004.

[27] R. Ling, W. Tahir, H.-Y. Lin, H. Lee, and L. Tian, "High-throughput intensity diffraction tomography with a computational microscope," *Biomed. Opt. Express*, vol. 9, no. 5, pp. 2130–2141, May 2018.

[28] Z. Wu, Y. Sun, J. Liu, and U. Kamilov, "Online regularization by denoising with applications to phase retrieval," in *The IEEE International Conference on Computer Vision (ICCV) Workshops*, Oct 2019.

[29] 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.

[30] A. N. Tikhonov and V. Y. Arsening, *Solution of Ill-Posed Problems*. Winston-Wiley, 1977.

[31] M. A. T. Figueiredo and R. D. Nowak, "Wavelet-based image estimation: An empirical Bayes approach using Jeffreys' noninformative prior," *IEEE Trans. Image Process.*, vol. 10, no. 9, pp. 1322–1331, September 2001.

[32] 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.

[33] 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.

[34] 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.

[35] J. Bect, L. Blanc-Feraud, G. Aubert, and A. Chambolle, "A ℓ_1-unified variational framework for image restoration," in *Proc. Euro. Conf. Comp. Vis. (ECCV)*, vol. 3024, New York, 2004, pp. 1–13.

[36] 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.

[37] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, pp. 293–318, 1992.

[38] M. V. Afonso, J. M.Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345–2356, September 2010.

[39] M. K. Ng, P. Weiss, and X. Yuan, "Solving constrained total-variation image restoration and reconstruction problems via alternating direction methods," *SIAM J. Sci. Comput.*, vol. 32, no. 5, pp. 2710–2736, August 2010.

[40] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[41] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 123–231, 2014.

[42] A. Mousavi, A. B. Patel, and R. G. Baraniuk, "A deep learning approach to structured signal recovery," in *Proc. Allerton Conf. Communication, Control, and Computing*, Allerton Park, IL, USA, September 30-October 2, 2015, pp. 1336–1343.

[43] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, Sep. 2017.

[44] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction," *Medical Physics*, vol. 44, no. 10, pp. e360–e375, 2017.

[45] J. C. Ye, Y. Han, and E. Cha, "Deep convolutional framelets: A general deep learning framework for inverse problems," *SIAM Journal on Imaging Sciences*, vol. 11, no. 2, pp. 991–1048, 2018.

[46] Y. Sun, Z. Xia, and U. S. Kamilov, "Efficient and accurate inversion of multiple scattering with deep learning," *Opt. Express*, vol. 26, no. 11, pp. 14 678–14 688, May 2018.

[47] Y. Li, Y. Xue, and L. Tian, "Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media," *Optica*, vol. 5, no. 10, pp. 1181–1190, Oct 2018.

[48] J. Yoo, S. Sabir, D. Heo, K. H. Kim, A. Wahab, Y. Choi, S. Lee, E. Y. Chae, H. H. Kim, Y. M. Bae, Y. Choi, S. Cho, and J. C. Ye, "Deep learning diffuse optical tomography," *IEEE Transactions on Medical Imaging*, pp. 1–1, 2019.

[49] A. Goy, G. Rughoobur, S. Li, K. Arthur, A. I. Akinwande, and G. Barbastathis, "High-resolution limited-angle phase tomography of dense layered objects using deep neural networks," *Proceedings of the National Academy of Sciences*, vol. 116, no. 40, pp. 19 848–19 856, 2019.

[50] E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, "Deep-STORM: super-resolution single-molecule microscopy by deep learning," *Optica*, vol. 5, no. 4, pp. 458–464, Apr 2018.

[51] Y. Sun, J. Liu, and U. Kamilov, "Block coordinate regularization by denoising," in *Advances in Neural Information Processing Systems (NeurIPS)*, Vancouver, BC, Canada, December 8-14, 2019, pp. 380–390.

[52] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.

[53] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 770–778.

[54] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed. Springer, 2017.

[55] S. Ghadimi and G. Lan, "Accelerated gradient methods for nonconvex nonlinear and stochastic programming," *Math. Program. Ser. A*, vol. 156, no. 1, pp. 59–99, March 2016.

[56] J. Bernstein, Y.-X. Wang, K. Azizzadenesheli, and A. Anandkumar, "signSGD: Compressed optimization for non-convex problems," in *Proc. 35th Int. Conf. Machine Learning (ICML)*, vol. 80, Stockholm, Sweden, Jul. 2018, pp. 560–569.

[57] X. Xu and U. S. Kamilov, "Signprox: One-bit proximal algorithm for nonconvex stochastic optimization," in *IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP)*, Brighton, UK, May 2019, pp. 7800–7804.

[58] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers, 2004.

[59] E. K. Ryu, J. Liu, S. Wang, X. Chen, Z. Wang, and W. Yin, "Plug-and-play methods provably converge with properly trained denoisers," in *Proc. 36th Int. Conf. Machine Learning (ICML)*, 2019, pp. 5546–5557.

[60] H. Robbins and S. Monro, "A stochastic approximation method," *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400–407, September 1951.

[61] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," in *International Conference on Learning Representations (ICLR)*, Vancouver, Canada, Apr. 2018.

[62] H. Sedghi, V. Gupta, and P. M. Long, "The singular values of convolutional layers," in *International Conference on Learning Representations (ICLR)*, New Orleans, LA, USA, May 2019.

[63] H. Gouk, E. Frank, B. Pfahringer, and M. Cree, "Regularisation of neural networks by enforcing Lipschitz continuity," 2018, arXiv:1804.04368.

[64] A. Matlock and L. Tian, "High-throughput, volumetric quantitative phase imaging with multiplexed intensity diffraction tomography," *Biomed. Opt. Express*, vol. 10, no. 12, pp. 6432–6448, Dec 2019.

[65] J. Li, A. Matlock, Y. Li, Q. Chen, C. Zuo, and L. Tian, "High-speed in vitro intensity diffraction tomography," *arXiv: 1904.06004 [physics.optics]*, 2019.

[66] R. T. Rockafellar and R. Wets, *Variational Analysis*. Springer, 1998.

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

A Background Material

The results in this section are well-known in the optimization literature and can be found in different forms in standard textbooks [54, 58, 66, 67]. For completeness, we summarize the key results useful for our analysis.

Definition 1. *An operator* T *is Lipschitz continuous with constant* $\lambda > 0$ *if*

$$\|\mathsf{T}\boldsymbol{x} - \mathsf{T}\boldsymbol{y}\| \leq \lambda\|\boldsymbol{x} - \boldsymbol{y}\|, \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n.$$

When $\lambda = 1$*, we say that* T *is nonexpansive.*

Definition 2. T *is cocoercive with constant* $\beta > 0$ *if*

$$(\mathsf{T}\boldsymbol{x} - \mathsf{T}\boldsymbol{y})^{\mathsf{T}}(\boldsymbol{x} - \boldsymbol{y}) \geq \beta\|\mathsf{T}\boldsymbol{x} - \mathsf{T}\boldsymbol{y}\|^2, \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n.$$

When $\beta = 1$*, we say that* T *is firmly nonexpansive.*

The following results are derived from the definition above.

Proposition 1. *Let* $\mathsf{T}_i : \mathbb{R}^n \to \mathbb{R}^n$ *for* $i \in I$ *be a set of nonexpansive operators. Then, their convex combination*

$$\mathsf{T} \coloneqq \sum_{i \in I} \theta_i \mathsf{T}_i, \quad \textit{with} \quad \theta_i > 0 \textit{ and } \sum_{i \in I} \theta_i = 1,$$

is nonexpansive.

Proof. By using the triangular inequality and the definition of nonexpansiveness, we obtain

$$\begin{aligned}\|\mathsf{T}\boldsymbol{x} - \mathsf{T}\boldsymbol{y}\| &\leq \sum_{i \in I} \theta_i \|\mathsf{T}_i\boldsymbol{x} - \mathsf{T}_i\boldsymbol{y}\| \\ &\leq \left(\sum_{i \in I} \theta_i\right) \|\boldsymbol{x} - \boldsymbol{y}\| = \|\boldsymbol{x} - \boldsymbol{y}\|,\end{aligned}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. □

Proposition 2. *Consider* $\mathsf{R} = \mathsf{I} - \mathsf{T}$ *where* $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^n$*.*

$$\mathsf{T} \textit{ is nonexpansive} \;\Leftrightarrow\; \mathsf{R} \textit{ is } (1/2)\textit{-cocoercive.}$$

Proof. First suppose that R is $1/2$ cocoercive. Let $\boldsymbol{h} \coloneqq \boldsymbol{x} - \boldsymbol{y}$ for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. We then have

$$\frac{1}{2}\|\mathsf{R}\boldsymbol{x} - \mathsf{R}\boldsymbol{y}\|^2 \leq (\mathsf{R}\boldsymbol{x} - \mathsf{R}\boldsymbol{y})^{\mathsf{T}}\boldsymbol{h} = \|\boldsymbol{h}\|^2 - (\mathsf{T}\boldsymbol{x} - \mathsf{T}\boldsymbol{y})^{\mathsf{T}}\boldsymbol{h}.$$

We also have that

$$\frac{1}{2}\|\mathsf{R}\boldsymbol{x} - \mathsf{R}\boldsymbol{y}\|^2 = \frac{1}{2}\|\boldsymbol{h}\|^2 - (\mathsf{T}\boldsymbol{x} - \mathsf{T}\boldsymbol{y})^{\mathsf{T}}\boldsymbol{h} + \frac{1}{2}\|\mathsf{T}\boldsymbol{x} - \mathsf{T}\boldsymbol{y}\|^2.$$

By combining these two and simplifying the expression

$$\|\mathsf{T}\boldsymbol{x} - \mathsf{T}\boldsymbol{y}\| \leq \|\boldsymbol{h}\|.$$

The converse can be proved by following this logic in reverse. □

Definition 3. *For a constant* $\alpha \in (0, 1)$*, we say that* T *is* α*-averaged, if there exists a nonexpansive operator* N *such that* $\mathsf{T} = (1 - \alpha)\mathsf{I} + \alpha\mathsf{N}$*.*

The following characterization is often convenient.

Proposition 3. *For a nonexpansive operator* T*, a constant* $\alpha \in (0,1)$*, and the operator* $\mathsf{R} \coloneqq \mathsf{I} - \mathsf{T}$*, the following are equivalent*

(a) T *is* α*-averaged*

(b) $(1-1/\alpha)\mathsf{I} + (1/\alpha)\mathsf{T}$ *is nonexpansive*

(c) $\|\mathsf{T}\boldsymbol{x} - \mathsf{T}\boldsymbol{y}\|^2 \leq \|\boldsymbol{x}-\boldsymbol{y}\|^2 - \left(\frac{1-\alpha}{\alpha}\right)\|\mathsf{R}\boldsymbol{x} - \mathsf{R}\boldsymbol{y}\|^2$, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$.

Proof. See Proposition 4.35 in [54]. □

Proposition 4. *Consider* $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^n$ *and* $\beta > 0$*. Then, the following are equivalent*

(a) T *is* β*-cocoercive*

(b) $\beta\mathsf{T}$ *is firmly nonexpansive*

(c) $\mathsf{I} - \beta\mathsf{T}$ *is firmly nonexpansive.*

(d) $\beta\mathsf{T}$ *is* $(1/2)$*-averaged.*

(e) $\mathsf{I} - 2\beta\mathsf{T}$ *is nonexpansive.*

Proof. For any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, let $\boldsymbol{h} \coloneqq \boldsymbol{x} - \boldsymbol{y}$. The equivalence between (a) and (b) is readily observed by defining $\mathsf{P} \coloneqq \beta\mathsf{T}$ and noting that

$$\begin{aligned}(\mathsf{P}\boldsymbol{x} - \mathsf{P}\boldsymbol{y})^\mathsf{T}\boldsymbol{h} &= \beta(\mathsf{T}\boldsymbol{x} - \mathsf{T}\boldsymbol{y})^\mathsf{T}\boldsymbol{h} \\ \text{and} \quad \|\mathsf{P}\boldsymbol{x} - \mathsf{P}\boldsymbol{y}\|^2 &= \beta^2\|\mathsf{T}\boldsymbol{x} - \mathsf{T}\boldsymbol{y}\|. \end{aligned} \tag{17}$$

Define $\mathsf{R} \coloneqq \mathsf{I} - \mathsf{P}$ and suppose (b) is true, then

$$\begin{aligned}&(\mathsf{R}\boldsymbol{x} - \mathsf{R}\boldsymbol{y})^\mathsf{T}\boldsymbol{h} = \|\boldsymbol{h}\|^2 - (\mathsf{P}\boldsymbol{x} - \mathsf{P}\boldsymbol{y})^\mathsf{T}\boldsymbol{h} \\ &= \|\mathsf{R}\boldsymbol{x} - \mathsf{R}\boldsymbol{y}\|^2 + (\mathsf{P}\boldsymbol{x} - \mathsf{P}\boldsymbol{y})^\mathsf{T}\boldsymbol{h} - \|\mathsf{P}\boldsymbol{x} - \mathsf{P}\boldsymbol{y}\|^2 \\ &\geq \|\mathsf{R}\boldsymbol{x} - \mathsf{R}\boldsymbol{y}\|^2.\end{aligned}$$

By repeating the same argument for $\mathsf{P} = \mathsf{I} - \mathsf{R}$, we establish the full equivalence between (b) and (c).

The equivalence of (b) and (d) can be seen by noting that

$$\begin{aligned}&2\|\mathsf{P}\boldsymbol{x} - \mathsf{P}\boldsymbol{y}\|^2 \leq 2(\mathsf{P}\boldsymbol{x} - \mathsf{P}\boldsymbol{y})^\mathsf{T}\boldsymbol{h} \\ &\Leftrightarrow \quad \|\mathsf{P}\boldsymbol{x} - \mathsf{P}\boldsymbol{y}\|^2 \leq 2(\mathsf{P}\boldsymbol{x} - \mathsf{P}\boldsymbol{y})^\mathsf{T}\boldsymbol{h} - \|\mathsf{P}\boldsymbol{x} - \mathsf{P}\boldsymbol{y}\|^2 \\ &= \|\boldsymbol{h}\|^2 - (\|\boldsymbol{h}\|^2 - 2(\mathsf{P}\boldsymbol{x} - \mathsf{P}\boldsymbol{y})^\mathsf{T}\boldsymbol{h} + \|\mathsf{P}\boldsymbol{x} - \mathsf{P}\boldsymbol{y}\|^2) \\ &= \|\boldsymbol{h}\|^2 - \|\mathsf{R}\boldsymbol{x} - \mathsf{R}\boldsymbol{y}\|^2.\end{aligned}$$

To show the equivalence with (e), first suppose that $\mathsf{N} \coloneqq \mathsf{I} - 2\mathsf{P}$ is nonexpansive, then $\mathsf{P} = \frac{1}{2}(\mathsf{I} + (-\mathsf{N}))$ is $1/2$-averaged, which means that it is firmly nonexpansive. On the other hand, if P is firmly nonexpansive, then it is $1/2$-averaged, which means that from Proposition 3(b) we have that $(1-2)\mathsf{I} + 2\mathsf{P} = 2\mathsf{P} - \mathsf{I} = -\mathsf{N}$ is nonexpansive. This directly means that N is nonexpansive. □

B Proof of Theorem 1

We consider the following operators

$$\mathsf{G} \coloneqq \nabla g + \mathsf{H} \quad \text{and} \quad \widehat{\mathsf{G}} \coloneqq \widehat{\nabla} g + \mathsf{H} \quad \text{with} \quad \mathsf{H} \coloneqq \tau(\mathsf{I} - \mathsf{D}_\sigma),$$

where $\widehat{\mathsf{G}}$ is the minibatch approximation of G. The direct application of Assumption 1 implies that for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$

$$\mathbb{E}[\widehat{\mathsf{G}}(\boldsymbol{x})] = \mathbb{E}[\widehat{\nabla} g(\boldsymbol{x})] + \mathsf{H}(\boldsymbol{x}) = \mathsf{G}(\boldsymbol{x}) \tag{18a}$$

$$\mathbb{E}[\|\mathsf{G}(\boldsymbol{x}) - \widehat{\mathsf{G}}(\boldsymbol{x})\|_2^2] = \mathbb{E}[\|\nabla g(\boldsymbol{x}) - \widehat{\nabla} g(\boldsymbol{x})\|_2^2] \leq \frac{\nu^2}{B} \tag{18b}$$

Now we prove Theorem 1 in several steps.

(a) Since ∇g is L-Lipschitz continuous, we know that it is $(1/L)$-cocoercive (see Theorem 2.1.5 in Section 2.1 of [58]). Then from Proposition 4, we know that the operator $(\mathsf{I} - (2/L)\nabla g)$ is nonexpansive.

(b) From the definition of H and the fact that D is nonexpansive, we know that $(\mathsf{I} - (1/\tau)\mathsf{H}) = \mathsf{D}$ is nonexpansive.

(c) From Proposition 1, we know that a convex combination of nonexpansive operators is also nonexpansive, hence

$$\begin{aligned}\mathsf{I} - \frac{2}{L+2\tau}\mathsf{G} = &\left(\frac{2}{L+2\tau} \cdot \frac{L}{2}\right)\left[\mathsf{I} - \frac{2}{L}\nabla g\right] \\ &+ \left(\frac{2}{L+2\tau} \cdot \frac{2\tau}{2}\right)\left[\mathsf{I} - \frac{1}{\tau}\mathsf{H}\right],\end{aligned}$$

is nonexpansive. Then from Proposition 4, we know that G is $1/(L+2\tau)$-cocoercive.

(d) Consider any $\boldsymbol{x}^* \in \mathsf{zer}(\mathsf{G})$ and $\boldsymbol{x} \in \mathbb{R}^n$. We then have

$$\begin{aligned}&\|\boldsymbol{x} - \boldsymbol{x}^* - \gamma\mathsf{G}\boldsymbol{x}\|^2 \\ &= \|\boldsymbol{x} - \boldsymbol{x}^*\|^2 - 2\gamma(\mathsf{G}\boldsymbol{x} - \mathsf{G}\boldsymbol{x}^*)^\mathsf{T}(\boldsymbol{x} - \boldsymbol{x}^*) + \gamma^2\|\mathsf{G}\boldsymbol{x}\|^2 \\ &\leq \|\boldsymbol{x} - \boldsymbol{x}^*\|^2 - \frac{2\gamma - (L+2\tau)\gamma^2}{L+2\tau}\|\mathsf{G}\boldsymbol{x}\|^2 \\ &\leq \|\boldsymbol{x} - \boldsymbol{x}^*\|^2 - \frac{\gamma}{L+2\tau}\|\mathsf{G}\boldsymbol{x}\|^2,\end{aligned} \tag{19}$$

where we used $\mathsf{G}\boldsymbol{x}^* = \boldsymbol{0}$, the cocoercivity of G, and the fact that $0 < \gamma \leq 1/(L+2\tau)$.

(e) For a single iteration of SIMBA $\boldsymbol{x}^+ = \boldsymbol{x} - \gamma\widehat{\mathsf{G}}\boldsymbol{x}$, we have

$$\begin{aligned}&\|\boldsymbol{x}^+ - \boldsymbol{x}^*\|^2 = \|\boldsymbol{x} - \boldsymbol{x}^* - \gamma\widehat{\mathsf{G}}\boldsymbol{x}\|^2 \\ &= \|\boldsymbol{x} - \boldsymbol{x}^* - \gamma\mathsf{G}\boldsymbol{x} + \gamma(\mathsf{G}\boldsymbol{x} - \widehat{\mathsf{G}}\boldsymbol{x})\|^2 \\ &= \|\boldsymbol{x} - \boldsymbol{x}^* - \gamma\mathsf{G}\boldsymbol{x}\|^2 + \gamma^2\|\mathsf{G}\boldsymbol{x} - \widehat{\mathsf{G}}\boldsymbol{x}\|^2 \\ &+ 2\gamma(\mathsf{G}\boldsymbol{x} - \widehat{\mathsf{G}}\boldsymbol{x})^\mathsf{T}(\boldsymbol{x} - \boldsymbol{x}^* - \gamma\mathsf{G}\boldsymbol{x}).\end{aligned}$$

By taking the conditional expectation with respect to the previous iterate $\boldsymbol{x}$, using (18), and applying the bound (19), we obtain

$$\begin{aligned}\mathbb{E}\left[\|\boldsymbol{x}^+ - \boldsymbol{x}^*\|^2 \mid \boldsymbol{x}\right] &\leq \|\boldsymbol{x} - \boldsymbol{x}^* - \gamma\mathsf{G}\boldsymbol{x}\|^2 + \frac{\gamma^2\nu^2}{B} \\ &\leq \|\boldsymbol{x} - \boldsymbol{x}^*\|^2 - \frac{\gamma}{L+2\tau}\|\mathsf{G}\boldsymbol{x}\|^2 + \frac{\gamma^2\nu^2}{B}.\end{aligned} \tag{20}$$

(f) By simply rearranging the terms, we obtain

$$\frac{\gamma}{L+2\tau}\|\mathsf{G}\boldsymbol{x}\|^2 \leq \mathbb{E}\left[\|\boldsymbol{x}-\boldsymbol{x}^*\|^2 - \|\boldsymbol{x}^+-\boldsymbol{x}^*\|^2|\boldsymbol{x}\right] + \frac{\gamma^2\nu^2}{B}$$

(g) Hence, by averaging over $t \geq 1$ iterations and taking the total expectation, we obtain our main result

$$\mathbb{E}\left[\frac{1}{t}\sum_{k=1}^{t}\|\mathsf{G}\boldsymbol{x}^{k-1}\|^2\right] \leq \frac{(L+2\tau)}{\gamma}\left[\frac{\|\boldsymbol{x}^0-\boldsymbol{x}^*\|^2}{t} + \frac{\gamma^2\nu^2}{B}\right].$$