Plug-and-Play Priors for Bright Field Electron Tomography and Sparse Interpolation

Suhas Sreehari, *Student Member, IEEE*, S. V. Venkatakrishnan, Brendt Wohlberg, *Senior Member, IEEE*, Gregery T. Buzzard, Lawrence F. Drummy, Jeffrey P. Simmons, *Member, IEEE*, and Charles A. Bouman, *Fellow, IEEE*

***Abstract*—Many material and biological samples in scientific imaging are characterized by nonlocal repeating structures. These are studied using scanning electron microscopy and electron tomography. Sparse sampling of individual pixels in a two-dimensional image acquisition geometry, or sparse sampling of projection images with large tilt increments in a tomography experiment, can enable high speed data acquisition and minimize sample damage caused by the electron beam. In this paper, we present an algorithm for electron tomographic reconstruction and sparse image interpolation that exploits the nonlocal redundancy in images. We adapt a framework, termed plug-and-play priors, to solve these imaging problems in a regularized inversion setting. The power of the plug-and-play approach is that it allows a wide array of modern denoising algorithms to be used as a "prior model" for tomography and image interpolation. We also present sufficient mathematical conditions that ensure convergence of the plug-and-play approach, and we use these insights to design a new nonlocal means denoising algorithm. Finally, we demonstrate that the algorithm produces higher quality reconstructions on both simulated and real electron microscope data, along with improved convergence properties compared to other methods.**

***Index Terms*—Plug-and-play, prior modeling, bright field electron tomography, sparse interpolation, non-local means, doubly stochastic gradient non-local means, BM3D.**

I. Introduction

TRANSMISSION electron microscopes are widely used for characterization of material and biological samples at the nano-meter scale [1]–[3]. In many cases, these electron microscopy samples contain many repeating structures that are similar or identical to each other. High quality reconstruction of these samples from tomographic projections is possible by exploiting the redundancy caused by repeating structures. As an important example, cryo-electron microscope (EM) tomography involves single particle reconstructions using several views of the same particle [1]. However, in the more general area of 3D transmission electron microscopy (TEM) tomography, no solution currently exists to fully exploit the redundancy in images constituted by many similar or identical particles.

Another important imaging problem is that raster scanning an electron beam across a large field of view is time consuming and can damage the sample. For this reason, there is growing interest in reconstructing full resolution images from sparsely sampled pixels [4], [5]. The redundancy in material and biological samples suggests that it is possible to reconstruct such images with sufficient fidelity by acquiring only a few random samples in the image and using an advanced image reconstruction algorithm that exploits non-local redundancies.

Conventionally, model-based iterative reconstruction (MBIR) solves a single optimization problem that tightly couples the log likelihood term (based on the data) and the log of the prior probability [6]–[14]. MBIR can, in principle, exploit redundancy in microscope images for tomographic reconstruction, but this requires selection of the appropriate log prior probability, which is very challenging in practice.

Patch-based denoising algorithms such as non-local means (NLM) [15]–[17] and BM3D [18] have been very successful in exploiting non-local redundancy in images. Inspired by the success of NLM, several researchers have proposed methods for non-locally regularized inversion. Chen *et al.* [19] proposed an MRF-style prior, but with non-local spatial dependencies, to perform Bayesian tomographic reconstruction. They adopted a two-step optimization involving a non-local weight update, followed by the image update. However, since the cost function changes with every iteration, there is no single fixed cost function that is minimized. Chun *et al.* [20] proposed non-local regularizers for emission tomography based on alternating direction method of multipliers (ADMM) [21]–[24], using the Fair potential [25] as the non-local regularizer. This model is restricted to convex potential functions, which in practice is a very strong constraint, and severely limits how expressive the model can be. Yang and Jacob proposed a unifying energy minimization framework for non-local regularization [26], resulting in a model that captures the intrinsically non-convex behavior required for

Manuscript received November 19, 2015; revised April 25, 2016 and July 2, 2016; accepted July 8, 2016. Date of publication August 11, 2016; date of current version November 4, 2016. This work was supported in part by an AFOSR/MURI Grant #FA9550-12-1-0458, by UES, Inc., under the Broad Spectrum Engineered Materials contract, and in part by the Electronic Imaging component of the ICMD Program of the Materials and Manufacturing Directorate of the Air Force Research Laboratory, Andrew Rosenberger, Program Manager. The work of Wohlberg was supported by the U.S. Department of Energy through the LANL/LDRD Program. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Alessandro Foi.

S. Sreehari and C. A. Bouman are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: ssreehar@purdue.edu; bouman@purdue.edu).

S. V. Venkatakrishnan is with Lawrence Berkeley National Laboratory, Berkeley, CA 94720 USA (e-mail: svvenkatakrishnan@lbl.gov).

B. Wohlberg is with the Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545 USA (e-mail: brendt@ieee.org).

G. T. Buzzard is with the Department of Mathematics, Purdue University, West Lafayette, IN 47907 USA (e-mail: buzzard@math.purdue.edu).

L. F. Drummy and J. P Simmons are with the Air Force Research Laboratory, Dayton, OH 45433 USA (e-mail: lawrence.drummy.1@us.af.mil; jeff.simmons.3@us.af.mil).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCI.2016.2599778

modeling distant particles with similar structure. However, it is not clear under what conditions their method converges. Non-local regularizers using PDE-like evolutions and total variation on the image patch graph have been proposed to solve inverse problems [27], [28]. However, since patch-based denoisers are typically not formulated as cost functions, it is unclear how to use them as prior models in Bayesian iterative reconstruction frameworks like MBIR.

Image interpolation or inpainting is also a widely researched problem [29]. Existing approaches can be broadly classified into two categories-those based on local regularization and those on non-local regularization. In local approaches, the missing pixels are reconstructed from an immediate neighborhood surrounding the unknown values to encourage similarity between spatially neighboring pixels [30].

An important subclass of these problems is interpolation from a sparsely sampled set of points. We refer to this type of problem as "sparse interpolation" to distinguish it from the interpolation or inpainting of relatively large contiguous regions from densely sampled points on the boundaries of those regions. Spurred by the success of non-local means, there have been several efforts to solve the sparse interpolation problem using global patch based dictionary models [5], [31]–[34]. Li and Zeng [35] adapted a two stage approach similar to [36] and used the BM3D denoising algorithm for the problem of sparse image reconstruction. However, this approach is not immediately applicable to the incorporation of a wider variety of non-local denoising algorithms such as are described by Milanfar in [37]–[39]. The simplicity and success of NLM and BM3D has also led to the question of how these algorithms can be used to solve other inverse problems. In fact, Danielyan *et al.* [36] have adapted BM3D for image deblurring through the optimization of two cost functions balanced by the generalized Nash equilibrium.

Venkatakrishnan *et al.* [40] developed a semi-empirical framework termed plug-and-play priors, but limited results were presented and the convergence of the algorithm was not discussed. Rond *et al.* [41] have used the plug-and-play framework to propose a Gaussian denoising algorithm for Poisson noisy inverse problems. Chan *et al.* have adapted plug-and-play for a variety of image restoration problems [42], while Teodoro *et al.* have used a class-adapted Gaussian mixture model as a prior within the plug-and-play framework for image deblurring and sparse reconstruction [43]. Several techniques have proposed to recover degraded and incomplete images by repeatedly applying denoising filters [44]–[46]. Heide *et al.* [47] used ADMM and the primal-dual algorithm to apply any Gaussian denoiser as an image prior in various blocks of the camera imaging pipeline, including demosaicing and interlaced high dynamic range (HDR) imaging.

In this paper, we extend our plug-and-play framework of [40], [48] to present a robust algorithm for tomographic reconstruction and sparse image interpolation that exploits the non-local redundancies in microscope images. The plug-and-play framework is based on ADMM [22], [23] which decouples the forward model and the prior model terms in the optimization procedure. This results in an algorithm that involves repeated application of two steps: an inversion step only dependent on the forward model, and a denoising step only dependent on the image prior model. The plug-and-play takes ADMM one step further by replacing the prior model optimization by a denoising operator.

However, while it is convenient to be able to use any denoising operator as a prior model, this new plug-and-play framework also raises the question as to whether plug-and-play necessarily inherits the convergence properties of ADMM. We answer this important question by presenting a theorem that outlines the sufficiency conditions to be satisfied by the denoising operator in order to guarantee convergence of the plug-and-play algorithm. We also present a proof for this convergence theorem partly based on the ideas presented by Moreau [49] and Williamson *et al.* [50]. Using this result, we then modify NLM to satisfy these sufficiency conditions, naming the resultant variant doubly stochastic gradient NLM (DSG-NLM).

We then apply DSG-NLM as a prior model to the tomographic reconstruction and sparse interpolation problems. This new DSG-NLM algorithm is based on symmetrizing the filter corresponding to the traditional NLM algorithm. Interestingly, Milanfar [51] has also discussed the benefit of symmetrizing the denoising operator, albeit in the context of improving the performance of denoising algorithms.

The plug-and-play electron tomography solution presented in this paper builds on the existing MBIR framework for bright field electron tomography [14], which models Bragg scatter and anomaly detection. We demonstrate that our proposed algorithm produces high quality tomographic reconstructions and interpolation on both simulated and real electron microscope images. Additionally our method has improved convergence properties compared to using the standard NLM or the BM3D algorithm as a regularizer for the reconstruction. Due to the generality of the plug-and-play technique, this work results in an MBIR framework that is compatible with any denoising algorithm as a prior model[1], and thus opens up a huge opportunity to adopt a wide variety of spatial constraints to solve a wide variety of inverse problems.

II. Plug-and-Play Framework

In this section, we outline the plug-and-play framework for model-based iterative reconstructions [40]. Let $x \in \mathbb{R}^N$ be an unknown image with a prior distribution given by $p_\beta(x)$ where β parameterizes a family of distributions, and let $y \in \mathbb{R}^M$ be the associated measurements of the image with conditional distribution given by $p(y|x)$. We will refer to $p(y|x)$ as the forward model for the measurement system. Then the maximum *a posteriori* (MAP) estimate of the image x is given by

$$\hat{x}_{\text{MAP}} = \underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \{l(x) + \beta s(x)\}, \tag{1}$$

where $l(x) = -\log p(y|x)$ and $\beta s(x) = -\log p_\beta(x)$. Notice that in this case β is a positive scalar used to control the level of regularization in the MAP reconstruction. In order to allow for the possibility of convex constraints, we will allow both $l(x)$

[1]However convergence of plug-and-play is not guaranteed unless the denoising operator meets the sufficiency conditions that we outline in Theorem III.1.

and $s(x)$ to take values on the extended real line, $\mathbb{R} \cup \{+\infty\}$. Using this convention, we can, for example, enforce positivity by setting $l(x) = +\infty$ for $x \leq 0$.

The solution to equation (1) may be computed using a variety of optimization methods, such as the alternating direction method of multipliers (ADMM), the fast iterative shrinkage thresholding algorithm (FISTA) [52], the two-step iterative shrinkage thresholding (TwIST) [53], or sparse reconstruction by separable approximation (SpaRSA) [54]. We choose the ADMM algorithm because it provides an effective framework for incorporating denoising operators as prior models by decoupling the forward and prior models. There are other methods, including those listed above, that also allow such a decoupling, but we focus on ADMM due to its competitive convergence properties [55] and do not consider the corresponding alternative forms of the plug-and-play framework.

The first step in applying ADMM for solving equation (1) is to split the variable x, resulting in an equivalent expression for the MAP estimate given by

$$(\hat{x}, \hat{v}) = \arg \min_{\substack{x,v \in \mathbb{R}^N \\ x=v}} \{l(x) + \beta s(v)\} \, . \tag{2}$$

As in [24, Section 3.1.1], this constrained optimization problem has an associated scaled-form augmented Lagrangian given by

$$L_\lambda(x, v; u) = l(x) + \beta s(v) + \frac{1}{2\sigma_\lambda^2} \|x - v + u\|_2^2 - \frac{\|u\|_2^2}{2\sigma_\lambda^2}, \tag{3}$$

where $\sigma_\lambda > 0$ is the augmented Lagrangian parameter.[2]

For our problem, the ADMM algorithm consists of iteration over the following steps:

$$\hat{x} \leftarrow \arg \min_{x \in \mathbb{R}^N} L_\lambda(x, \hat{v}; u) \tag{4}$$

$$\hat{v} \leftarrow \arg \min_{v \in \mathbb{R}^N} L_\lambda(\hat{x}, v; u) \tag{5}$$

$$u \leftarrow u + (\hat{x} - \hat{v}) \, , \tag{6}$$

where $\hat{v}$ is initialized to some value and u is typically initialized as zero.

In fact, if $l(x)$ and $s(x)$ are both proper, closed, and convex functions, and a saddle point solution exists, then the ADMM converges to the global minimum [22]–[24].

We can express the ADMM iterations more compactly by defining two operators. The first is an inversion operator F defined by

$$F(\tilde{x}; \sigma_\lambda) = \operatorname*{argmin}_{x \in \mathbb{R}^N} \left\{ l(x) + \frac{\|x - \tilde{x}\|_2^2}{2\sigma_\lambda^2} \right\} \, , \tag{7}$$

and the second is a denoising operator H given by

$$H(\tilde{v}; \sigma_n) = \operatorname*{argmin}_{v \in \mathbb{R}^N} \left\{ \frac{\|\tilde{v} - v\|_2^2}{2\sigma_n^2} + s(v) \right\} \, , \tag{8}$$

where $\sigma_n = \sqrt{\beta}\sigma_\lambda$ can be interpreted as the assumed noise standard deviation in the denoising operator. We say that H is a Moreau proximity operator, i.e., H is the proximal mapping for the proper, closed, and convex function $s : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$. Further, if $l(x)$ is convex, then the inversion operator, F, is also a Moreau proximity operator.

Using these two operators, we can easily derive the plug-and-play algorithm shown in Algorithm 1 as a special case of the ADMM iterations. This formulation has a number of practical and theoretical advantages. First, in this form we can now "plug in" denoising operators that are not in the explicit form of the optimization of equation (8). So for example, we will later see that popular and effective denoising operators such as non-local means (NLM) [56] or BM3D [18], which are not easily represented in an optimization framework can be used in the plug-and-play iterations. Second, this framework allows for decomposition of the problem into separate software systems for the implementation of the inversion operator, F, and the denoising operator, H. As software systems for large inversion problems become more complex, the ability to decompose them into separate modules, while retaining the global optimality of the solution, can be extremely valuable.

One important advantage of the plug-and-play method is that it can simplify the selection of the regularization parameter. In order to understand this, notice that the plug-and-play algorithm only requires the selection of two parameters, β and σ_λ, with the remaining parameter being set to $\sigma_n = \sqrt{\beta}\sigma_\lambda$. We can see from equation (1) that only β controls the amount of regularization, and therefore the value of σ_λ does not affect the final reconstructed image. However from equation (3) we see that σ_λ controls the weight of the constraint term in the augmented Lagrangian, and therefore the value of σ_λ must be chosen appropriately in order to achieve rapid convergence.

Importantly, the regularization parameter β is unitless. The fact that β is unitless makes its selection much easier in practice. This is in contrast to conventional regularization parameters that can be tricky to set because they have units related to the quantity being reconstructed [57]. In fact, it turns out that 1 if often a good starting point for the value of β. The value of β can then be increased or decreased for more or less regularization, respectively. Intuitively, denoising algorithms such as NLM only requires the specification of the noise variance, but do not require knowledge about the scale of the image being restored. Consequently, the plug-and-play algorithm inherits the scale invariance property of denoising algorithms such as NLM, and the regularization of the resulting plug-and-play reconstruction algorithm can be therefore set with a unitless parameter β.

In theory, the value of σ_λ does not affect the reconstruction for a convex optimization problem, but in practice, a well-chosen value of σ_λ can substantially speed up ADMM convergence [24], [58], [59]; so the careful choice of σ_λ is important. Our approach is to choose the value of σ_λ to be approximately equal to the amount of variation in the reconstruction. Formally stated, we choose

Formally stated, we choose

$$\sigma_\lambda^2 \approx \frac{1}{N} \sum_{i=0}^{N-1} \operatorname{var}[x_i | y] \, . \tag{9}$$

[2]The augmented Lagrangian parameter, σ_λ, is related to the ADMM penalty parameter, λ, through the simple expression $\sigma_\lambda = \frac{1}{\sqrt{\lambda}}$.

Algorithm 1: Plug-and-play algorithm for implementation of a general inversion operator $F(\tilde{x};\sigma_\lambda)$, and a prior model specified by the denoising operator in $H(\tilde{v};\sigma_n)$.

initialize $\hat{v}$
$u \leftarrow 0$
while not converged **do**
  $\tilde{x} \leftarrow \hat{v} - u$
  $\hat{x} \leftarrow F(\tilde{x};\sigma_\lambda)$
  $\tilde{v} \leftarrow \hat{x} + u$
  $\hat{v} \leftarrow H(\tilde{v};\sigma_n)$
  $u \leftarrow u + (\hat{x} - \hat{v})$
end while

to be the average variance of the pixels in x.

This choice for the value of σ_λ^2 is motivated by its role as the inverse regularizer in equation (7). In practice, this can be done by first computing an approximate reconstruction using some baseline algorithm, and then computing the mean squared pixel variation in the approximate reconstruction.

Of course, for an arbitrary denoising algorithm, the question remains of whether the plug-and-play algorithm converges. The following section provides practical conditions for the denoising operator to meet that ensure convergence of the iterations.

III. Convergence of the Plug-and-Play Algorithm

It is well known that the ADMM algorithm is guaranteed to converge under appropriate technical conditions. For example, if the optimization problem is convex and a saddle point solution exists then the iterations of ADMM converge [22]–[24]. However, in our plug-and-play approach, we will be using general denoising algorithms to implement the operator $H(\tilde{v};\sigma_n)$, and therefore, the function $s(x)$ is not available for analysis. This raises the question of what conditions $H(\tilde{v};\sigma_n)$ and $l(x)$ must satisfy in order to ensure that the plug-and-play algorithm converges.

In the following theorem, we give conditions on both the log likelihood function, $l(x)$, and the denoising operator, $H(x)$, that are sufficient to guarantee convergence of the plug-and-play algorithm to the global minimum of some implicitly defined MAP cost function. This is interesting because it does not ever require that one know or explicitly specify the function $s(x)$. Instead, $s(x)$ is implicitly defined through the choice of $H(x)$.

Theorem III.1: Let the negative log likelihood function $l : \mathbb{R}^N \rightarrow \mathbb{R} \cup \{+\infty\}$ and the denoising operator $H : \mathbb{R}^N \rightarrow \mathbb{R}^N$ meet the following conditions:

1) H is a continuously differentiable function on $\mathbb{R}^N$;
2) $\forall x \in \mathbb{R}^N, \nabla H(x)$ is a symmetric matrix with eigenvalues in the interval $[0, 1]$;
3) There exists a y in the range of H such that $l(y) < \infty$;
4) There exists at least one fixed point of $H : H(x_0) = x_0$;
5) l is a proper closed convex function which is lower bounded by a function $f(\|x\|)$ such that $f(r)$ is monotone increasing with

$$\lim_{r\rightarrow\infty} f(r) = \infty\ .$$

Then the following results hold:

1) H is a proximal mapping for some proper closed convex function s on $\mathbb{R}^N$;
2) There exists an element which attains the infimum, which is the MAP estimate, $\hat{x}_{\mathrm{MAP}}$, such that

$$p^* = \inf_{x\in\mathbb{R}^N} \{l(x) + \beta s(x)\} = l(\hat{x}_{\mathrm{MAP}}) + \beta s(\hat{x}_{\mathrm{MAP}})\ ;$$

3) The plug-and-play algorithm converges in the following sense,

$$\lim_{k\rightarrow\infty} \{\hat{x}^{(k)} - \hat{v}^{(k)}\} = 0;$$
$$\lim_{k\rightarrow\infty} \{l(\hat{x}^{(k)}) + \beta s(\hat{v}^{(k)})\} = p^*\ ,$$

where $\hat{x}^{(k)}$ and $\hat{v}^{(k)}$ denote the result of the k^{th} iteration.

The proof of this theorem, which is presented in Appendix A, depends on a powerful theorem proved by Moreau in 1965 [49]. This theorem states that H is a proximal mapping if and only if it is non-expansive and the sub-gradient of a convex function on $\mathbb{R}^N$. Intuitively, once we can show that the denoising operator, H, is a proximal mapping, then we know that it is effectively implementing an update step of the form of the ADMM algorithm of equation (8).

The first and second conditions of the theorem ensure that the conditions of Moreau's theorem are met. This is because the structure of $H(x)$ with eigenvalues in the interval $[0, 1]$ ensures that H is the gradient of some function ϕ, that ϕ is convex, and that H is non-expansive.

The additional three conditions of Theorem III.1 ensure that the infimum can be attained and the MAP estimate actually exists for the problem. Importantly, this is done without explicit reference to the prior function $s(x)$. More specifically, the third condition ensures that the set of feasible solutions is not empty, the fourth condition will be seen to imply that the function $s(x)$ in the conclusion takes on its global minimum value, i.e., that the minimum is not achieved toward infinity, and the fifth condition ensures that the MAP cost function grows towards infinity, forcing the sum $l(x) + \beta s(x)$ to take on its global minimum value. Since H is a denoising operator, the assumption of a fixed point is quite natural: a constant image is a fixed point of most denoising operators.

In general, verifying the sufficiency conditions of Theorem III.1 may be more difficult for some denoising algorithms than it is for others. There could exist some denoising operators for which these sufficiency conditions may not hold. But importantly, in the next section, we will show that some real denoising operators can be modified to meet the conditions of this theorem. In particular, the symmetrized non-local means filters investigated by Milanfar [51], [60] are designed to create a symmetric gradient, and we build on this approach to introduce a NLM filter that is symmetric and doubly stochastic with eigenvalues in the interval $[0, 1]$.

IV. Non-Local Means Denoising With Symmetric, Doubly Stochastic Gradient

In order to satisfy the conditions for convergence, the gradient of the denoising operator must be a symmetric matrix with eigenvalues in the interval $[0, 1]$. However, the standard NLM denoising algorithm does not satisfy this condition. In addition, to ensure that the NLM filter has unit gain, the NLM weight matrix must be stochastic. Due to the symmetry requirement, the unit-gain NLM filter must therefore have a doubly stochastic weight matrix.

In this section, we introduce a simple modification of the NLM approach, which we refer to as the doubly stochastic gradient NLM (DSG-NLM), that satisfies the required convergence conditions. Interestingly, the symmetrized non-local means filters investigated by Milanfar [51] also achieve a symmetric and stochastic gradient, but require the use of a more complex iterative algorithm to symmetrize the operator.

The NLM algorithm is known to produce much higher quality results than traditional local smoothing-based denoising methods [56]. It works by estimating each pixel/voxel[3] as a weighted mean of all pixels in the image[4]. In this section, $\tilde{v}$ will denote a noisy image with pixel values $\tilde{v}_s$ at locations $s \in S$. Generally, S is a discrete lattice, so for 2D images $S = \mathbb{Z}^2$ and for 3D volumes $S = \mathbb{Z}^3$.

Using this notation, the NLM denoising method can be represented as

$$\hat{v}_s = \sum_{r \in \Omega_s} w_{s,r} \tilde{v}_r \ , \tag{10}$$

where $\hat{v}_s$ is the denoised result, the coefficients $w_{s,r}$ are the NLM weights, and Ω_s is the NLM search window defined by

$$\Omega_s = \{r \in S : \|r - s\|_\infty \leq N_s\} \ .$$

Note that the integer N_s controls the size of the NLM search window; we assume $N_s \geq 1$ to avoid the trivial case of a window with a single point. In general, larger values of N_s can yield better results but with higher computational cost.

Using this notation, the plug-and-play denoising operator is given by

$$H(\tilde{v}; \sigma_n) = W\tilde{v} \ ,$$

where the matrix

$$W_{s,r} = \begin{cases} w_{s,r} & \text{if } r \in \Omega_s \\ 0 & \text{otherwise} \end{cases} \ .$$

Now if we fix the weights, then it is clear that

$$\nabla H(v; \sigma_n) = W \ .$$

Condition 2 of Theorem III.1 requires that W be a symmetric matrix with eigenvalues in the interval $[0, 1]$. In order to enforce this, we propose the following five step procedure for computing the DSG-NLM filter weights,

$$w_{s,r} \leftarrow \exp\left\{\frac{-\|P_r - P_s\|_2^2}{2N_p^2\sigma_n^2}\right\} \Lambda\left(\frac{s-r}{N_s+1}\right) \tag{11}$$

$$w_{s,r} \leftarrow \frac{w_{s,r}}{\sqrt{\left(\sum_{r \in \Omega_s} w_{s,r}\right)\left(\sum_{s \in \Omega_r} w_{r,s}\right)}} \tag{12}$$

$$\alpha \leftarrow \frac{1}{\max_s \sum_{r \in \Omega_s} w_{s,r}} \tag{13}$$

$$w_{s,r} \leftarrow \alpha w_{s,r} \tag{14}$$

$$w_{s,s} \leftarrow w_{s,s} + \left(1 - \sum_{r \in \Omega_s} w_{s,r}\right) \ , \tag{15}$$

where $P_s \in \mathbb{R}^{N_p^2}$ denotes a patch of size $N_p \times N_p$ centered at position s, and the function

$$\Lambda(s) = \prod_{i=1}^{\dim(s)} (1 - |s_i|)_+$$

is the separable extension of the triangle function.

Notice that equation (12) renormalizes that weights so that the rows and columns sum approximately to 1. Therefore, in our experiments we observed that α from equation (13) typically takes on a value of approximately 0.95. Equations (14) and (15) then rescale the weights and add a positive quantity to the diagonal that ensure that rows and columns of the matrix must sum to 1. Further, we stop adapting the DSG-NLM weights as the plug-and-play algorithm progresses. This is a critical step to ensure that we indeed have a symmetric gradient with eigenvalues in the interval $[0, 1]$. Such a gradient fulfills condition 2 of Theorem III.1 and therefore meets the conditions that guarantee convergence of the plug-and-play algorithm.

The following theorem ensures that the DSG-NLM weight matrix W generated by equations (11) through (15) meets condition 2 of Theorem III.1 that is required for convergence of the plug-and-play algorithm.

Theorem IV.1: The DSG-NLM weight matrix W generated by equations (11) through (15) is a symmetric, doubly stochastic matrix with eigenvalues in the interval $[0, 1]$.

Proof: See Appendix B.

V. 3D Bright Field EM Forward Model

In this section, we formulate the explicit form of the inversion operator, $F(\tilde{x}, \sigma_\lambda)$, for the application of 3D bright field EM tomography. For this problem, we adopted both the forward model and optimization algorithms described in [14]. More specifically, the negative log likelihood function is given by

$$l(x, d, \sigma) = \frac{1}{2}\sum_{k=1}^{K}\sum_{i=1}^{M} \beta_{T,\delta}\left((y_{k,i} - A_{k,i,*}x - d_k)\frac{\sqrt{\Lambda_{k,\text{ii}}}}{\sigma}\right) + MK\log(\sigma) + C \ ,$$

[3]From this point forward, we will use the term "pixel" for both the 2 and 3-D cases.

[4]In practice, we only compute the weighted mean of pixels/voxels in a search window, instead of the whole image.

where K is the number of tilts, $\lambda_{k,i}$ is the electron counts corresponding to the ith measurement at the kth tilt, $y_{k,i} = -\log \lambda_{k,i}$, $\lambda_{D,k}$ is the blank scan value at the kth tilt, $d_k = -\log \lambda_{D,k}$, A_k is the $M \times N$ tomographic forward projection matrix associated with the kth tilt, $A_{k,i,*}$ is the ith row of A_k, σ^2 is a proportionality constant, Λ_k is a diagonal matrix whose entries are set such that $\frac{\sigma^2}{\Lambda_{k,\mathrm{ii}}}$ is the variance of $y_{k,i}$, $d = [d_1, \ldots, d_K]$ is the offset parameter vector, C is a constant, and $\beta_{T,\delta}(\cdot)$ is the generalized Huber function defined as,

$$\beta_{T,\delta}(x) = \begin{cases} x^2 & \text{if } |x| < T \\ 2\delta T|x| + T^2(1-2\delta) & \text{if } |x| \geq T \,. \end{cases} \tag{16}$$

The generalized Huber function is used to reject measurements with large errors. This is useful because measurement may vary from the assumed model for many practical reasons. For example, in bright field EM, Bragg scatter can cause highly attenuated measurements that otherwise would cause visible streaks on the reconstruction [61].

To compute the inversion operator F of equation (7), we minimize the cost function below with respect to x, d, and σ.

$$\begin{aligned} &c(x, d, \sigma; \tilde{x}, \sigma_\lambda) \\ &= \frac{1}{2} \sum_{k=1}^{K} \sum_{i=1}^{M} \beta_{T,\delta} \left((y_{k,i} - A_{k,i,*}x - d_k) \frac{\sqrt{\Lambda_{k,\mathrm{ii}}}}{\sigma} \right) \\ &\quad + \; MK \log(\sigma) + \frac{\|x - \tilde{x}\|_2^2}{2\sigma_\lambda^2}. \end{aligned} \tag{17}$$

So the inversion operator is computed as

$$F(\tilde{x}; \sigma_\lambda) = \arg \min_{x \geq 0, d, \sigma} c(x, d, \sigma; \tilde{x}, \sigma_\lambda). \tag{18}$$

As in the case of sparse interpolation, we set $c(x, d, \sigma; \tilde{x}, \sigma_\lambda) = +\infty$ for $x < 0$ in order to enforce positivity.

The details of the optimization algorithm required for equation (18) are described in [14]. The optimization algorithm is based on alternating minimization with respect to the three quantities and uses a majorization based on a surrogate function to handle the minimization of the generalized Huber function [62].

For this complex problem, we note some practical deviations from the theory. First, the negative log likelihood function, $l(x)$, is not convex in this case, so the assumptions of the plug-and-play convergence do not hold. With such a non-convex optimization, it is not possible to guarantee convergence to a global minimum, but in practice most optimization algorithms generate very good results. Also, the global optimization of equation (18) is approximated by three iterations of alternating minimization with respect to x, d, and σ. Nonetheless, in our experimental results section, we will illustrate our empirical observation that the plug-and-play algorithm consistently converges even with these approximations to the ideal case.

VI. Sparse Interpolation Forward Model

In this section, we formulate the explicit form of the inversion operator, $F(\tilde{x}, \sigma_\lambda)$, for the application of sparse interpolation (the noise-free limit of approximation by sparse measurements). More specifically, our objective will be to recover an image $x \in \mathbb{R}^N$ from a noisy and sparsely subsampled version denoted by $y \in \mathbb{R}^M$ where $M \ll N$. More formally, the forward model for this problem is given by

$$y = Ax + \epsilon \,, \tag{19}$$

where $A \in \mathbb{R}^M \times \mathbb{R}^N$. The sampling matrix A contains a subset of the rows of an identity matrix. We also define $I(j) = \sum_i A_{i,j}$ so that $I(j) = 1$ when the j^{th} pixel is sampled, and $I(j) = 0$, if it is not. Furthermore, ϵ is an M-dimensional vector of i.i.d. Gaussian random variables with mean zero and variance σ_w^2.

For such a sparse sampling system, we can write the negative log likelihood function as

$$l(x) = \frac{1}{2\sigma_w^2} \|y - Ax\|_2^2 + C \,, \tag{20}$$

where C is a constant. In order to enforce positivity, we also modify the negative likelihood function by setting $l(x) = +\infty$ if any component of x is negative. We include positivity in $l(x)$ rather than in the denoising operator so that H remains continuously differentiable. Also, A has a nontrivial null space, so $l(x)$ does not satisfy the growth condition in Theorem III.1 (5). However, setting $l(x)$ to be $+\infty$ outside some large ball (depending on maximum possible pixel value) produces a function that does satisfy the growth condition and that does not affect the final result. In practice, this modification to $l(x)$ is not needed since the denoising step is nonexpansive.

Using equation (7), the interpolation inversion operator is given by

$$F(\tilde{x}; \sigma_\lambda) = \underset{x \geq 0}{\operatorname{argmin}} \left\{ \frac{1}{2\sigma_w^2} \|y - Ax\|_2^2 + \frac{1}{2\sigma_\lambda^2} \|x - \tilde{x}\|_2^2 \right\}.$$

Due to the simple structure of the matrix A, we can also calculate an explicit pixel-wise expression for F. In the limit as σ_w^2 tends to 0, F reduces to the following form

$$F_i(\tilde{x}; \sigma_\lambda) = \begin{cases} [y_i]_+ & \text{if } I(i) = 1 \\ [\tilde{x}_i]_+ & \text{if } I(i) = 0. \end{cases} \tag{21}$$

where $[\cdot]_+$ represents zeroing of any negative argument. In this noise-free limit, the minimizer is forced to take on the measured

TABLE I
RMSE of The Reconstructed Al Spheres Image Compared to The Ground Truth (After 200 Plug-and-Play Iterations)

FBP	qGGMRF	NLM	DSG-NLM
14.608 $\times 10^{-4}$ nm^{-1}	4.581 $\times 10^{-4}$ nm^{-1}	2.531 $\times 10^{-4}$ nm^{-1}	2.488 $\times 10^{-4}$ nm^{-1}

TABLE II
Plug-and-Play Parameters for Tomographic Reconstructions

Plug-and-play parameters	Al spheres (simulated)	Al spheres (real)	Silicon dioxide (real)
σ_λ (nm^{-1})	8.66×10^{-4}	8.66×10^{-4}	8.66×10^{-4}
β	3.68	4.77	4.30

TABLE III
NORMALIZED PRIMAL RESIDUAL CONVERGENCE ERROR FOR TOMOGRAPHY EXPERIMENTS (AFTER 200 PLUG-AND-PLAY ITERATIONS)

Primal residual convergence error	qGGMRF	NLM	DSG-NLM
Al spheres (simulated)	3.46×10^{-12}	2.12×10^{-3}	1.02×10^{-10}
Al spheres (real)	7.06×10^{-11}	3.66×10^{-4}	2.77×10^{-9}
Silicon dioxide (real)	4.99×10^{-12}	8.12×10^{-3}	8.34×10^{-10}

TABLE IV
NORMALIZED DUAL RESIDUAL CONVERGENCE ERROR FOR TOMOGRAPHY EXPERIMENTS (AFTER 200 PLUG-AND-PLAY ITERATIONS)

Dual residual convergence error	qGGMRF	NLM	DSG-NLM
Al spheres (simulated)	1.55×10^{-10}	7.22×10^{-3}	1.04×10^{-9}
Al spheres (real)	2.61×10^{-10}	1.12×10^{-3}	5.73×10^{-9}
Silicon dioxide (real)	9.06×10^{-11}	5.49×10^{-2}	3.55×10^{-9}

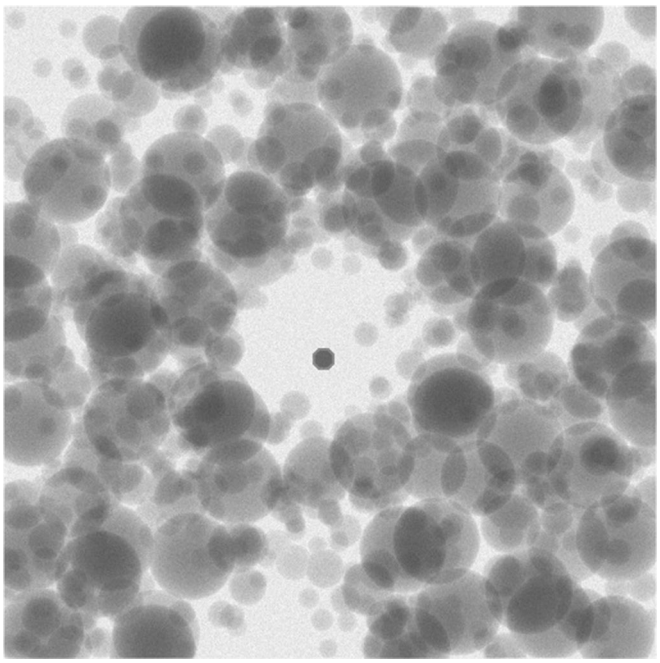


Fig. 1. 0° tilt of the aluminum spheres (simulated) dataset.

values at the sample points, hence motivating the name sparse interpolation.

VII. RESULTS AND DISCUSSION

In this section, we present experimental results on both real and simulated data for the applications of bright-field EM tomography and sparse interpolation. For all experiments, we present convergence plots that compare both primal and dual residuals using different priors. The normalized primal and dual

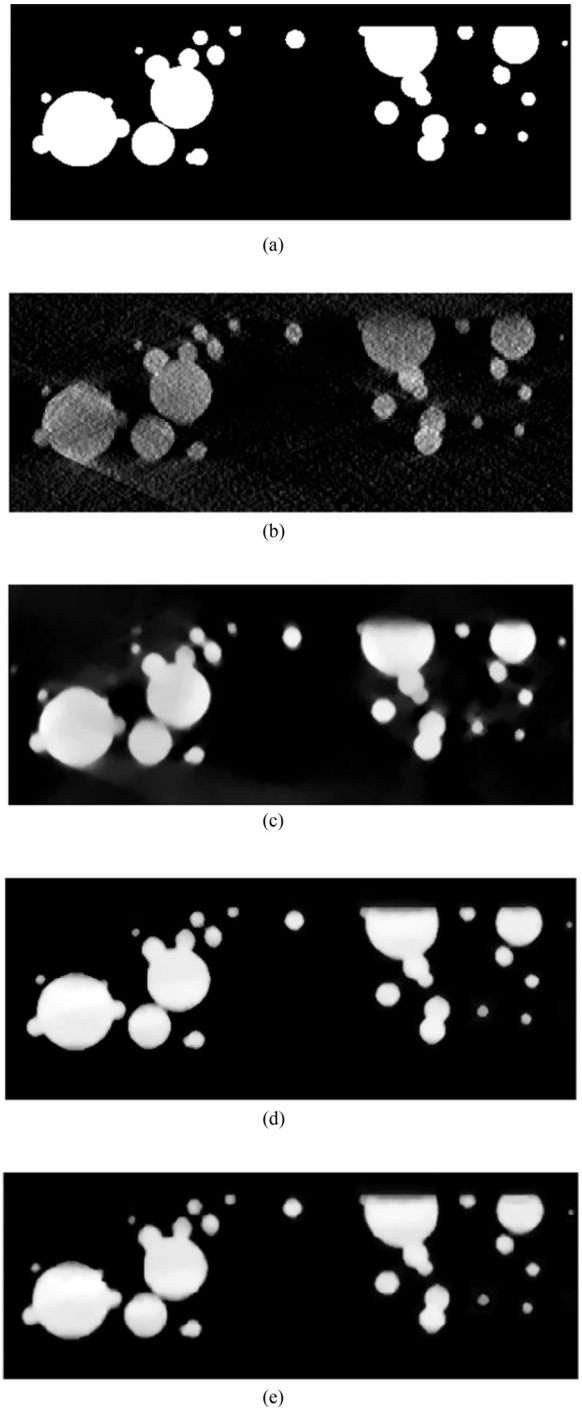


Fig. 2. Tomographic reconstruction of the simulated aluminum spheres dataset. NLM and DSG-NLM reconstructions are clearer and relatively artifact-free (a) The aluminum spheres phantom (ground truth); (b) Filtered Backprojection; (c) qGGMRF (T = 3; = 0:5); (d) 3D NLM using plug-and-play; (e) 3D DSG-NLM using plug-and-play.

residues [24, p. 18], $r^{(k)}$ and $s^{(k)}$ respectively, at the kth iteration of the plug-and-play algorithm are given by

$$r^{(k)} = \frac{\|\hat{x}^{(k)} - \hat{v}^{(k)}\|_2}{\|\hat{x}^{(\infty)}\|_2}; \tag{22}$$

$$s^{(k)} = \frac{\|\hat{v}^{(k)} - \hat{v}^{(k-1)}\|_2}{\|u^{(k)}\|_2}, \tag{23}$$

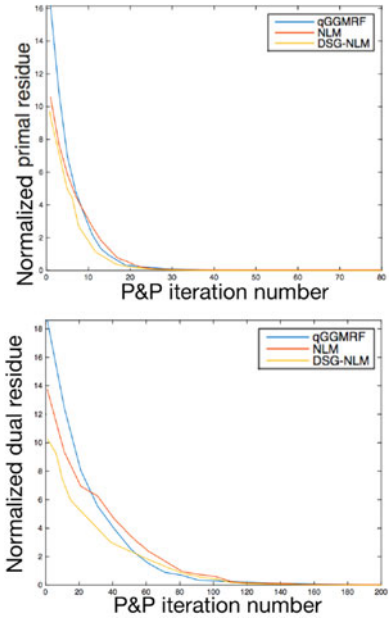


Fig. 3. Plug-and-play primal and dual residual convergence for tomographic reconstruction of (simulated) aluminum spheres. DSG-NLM achieves complete convergence.

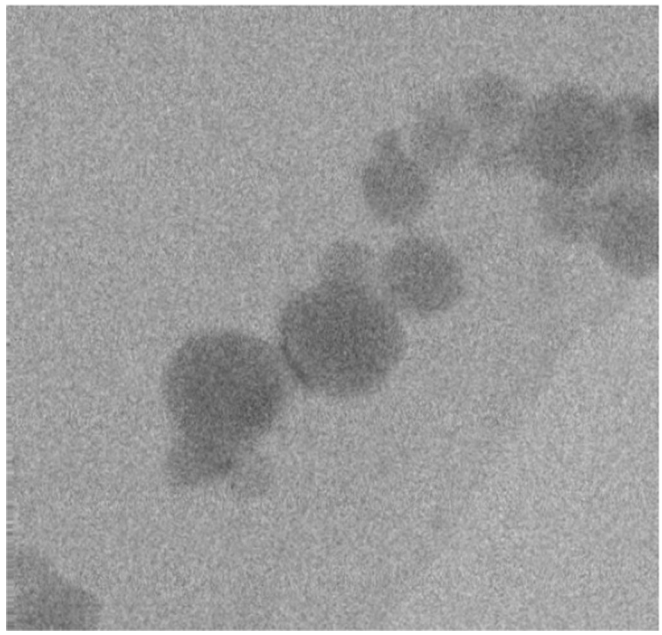


Fig. 4. 0° tilt of the very noisy aluminum spheres (real) dataset.

where $\hat{x}^{(k)}$, $\hat{v}^{(k)}$, and $u^{(k)}$ are the values of $\hat{x}$, $\hat{v}$, and u respectively after the kth iteration of the plug-and-play algorithm, respectively, and $\hat{x}^{(\infty)}$ is the final value of the reconstruction, $\hat{x}$. The primal and dual residues are normalized to ensure that they represent relative feasibility measures rather than absolute feasibility measures. This property makes plug-and-play residues scale invariant [63].

We stopped adapting the DSG-NLM weights after 20 iterations in the tomography experiments and after 12 iterations in the sparse interpolation experiments. This step was critical in achieving a doubly stochastic gradient for DSG-NLM, thereby satisfying the sufficiency conditions of our convergence theorem. Finally, using an interpolation experiment, we discuss the case where we do not stop adapting the weights of DSG-NLM, and show that plug-and-play may not converge in this case.

The plug-and-play parameters have been chosen to produce the best possible reconstruction for each prior, based on mean squared error or visual inspection when the ground truth was unavailable.

A. Bright Field EM Tomography

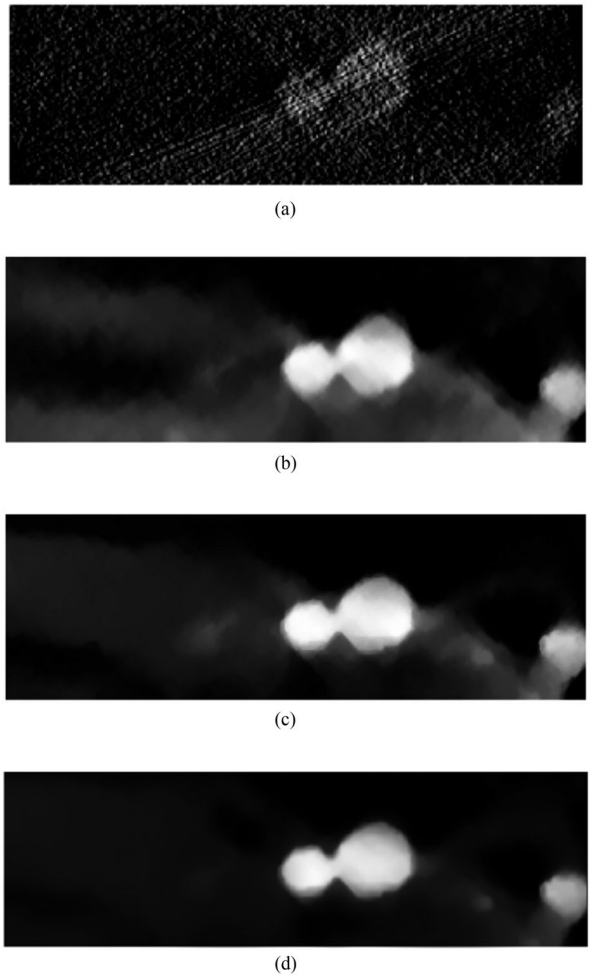
In this section, we present the results of bright field tomographic reconstruction of (1) a simulated dataset of aluminum spheres of varying radii, (2) a real dataset of aluminum spheres, and (3) a real dataset of silicon dioxide. We compare four reconstruction methods–filtered backprojection, MBIR with qGGMRF prior [64], plug-and-play reconstructions with 3D NLM and 3D DSG-NLM as prior models. We used qGGMRF, 3D NLM and 3D DSG-NLM as prior models within the plug-and-play framework. Filtered backprojection was used as the initialization for all MBIR-based reconstructions. All the reconstruction results shown below are x-z slices (i.e., slices parallel to the electron beam). The qGGMRF parameters used for all reconstructions are as follows: $q = 1$, $p = 1.2$, and $c = 0.001$. The NLM and DSG-NLM patch size used for all reconstructions is $5 \times 5 \times 5$. We stopped adapting the DSG-NLM weights at 20 iterations of the plug-and-play algorithm. The plug-and-play parameters used are given in Table II.

In all the experiments, we observe from Tables III and IV Figs. 3, 6, and 9 that the DSG-NLM ensures that the plug-and-play algorithm converges fully, while NLM achieves convergence to within a fraction of a percent. We note that the qGGMRF operator has the form of equation (8) with $s(v)$ a convex function that attains its minimum. In this case, Prop 7.d of [49] implies that this operator is continuously differentiable, Theorem A.1 (originally of the same paper) implies that condition 2 of Theorem III.1 is also satisfied, and Lemma IX.4 implies that condition 4 is satisfied. The remaining conditions are satisfied by l, so plug-and-play is guaranteed to converge fully, which we observe in the plots.

1) Aluminum Spheres (Simulated) Dataset: The aluminum spheres simulated dataset contains 47 equally-spaced tilts about the y-axis, spanning $[-70^\circ, +70^\circ]$. The attenuation co-efficient of the spheres are assumed to be 7.45×10^{-3} nm. The noise model is Gaussian, with variance set equal to the mean. The phantom also contains effects that resemble Bragg scatter. The dimensions of the phantom are 256 nm, 512 nm, and 512 nm along z, x, and y axes, respectively.

Fig. 1 shows a 0° tilt projection of the simulated TEM data. Since this is a bright-field image, the aluminum spheres appear dark against a bright background. Fig. 2 shows the ground truth along with three reconstructions of slice 280 along the x-z plane. The NLM and DSG-NLM reconstructions have no shadow artifacts, and also have low RMSE values (see Table I). The edges are also sharper in the NLM and DSG-NLM reconstructions.

2) Aluminum Spheres (Real) Dataset: The aluminum spheres dataset (see Fig. 4) has 67 equally-spaced tilts about



(a) (b) (c) (d)

Fig. 5. Tomographic reconstruction of the real aluminum spheres dataset. NLM and DSG-NLM reconstructions are clearer and have fewer smear and missing-wedge artifacts (a) Filtered Backprojection; (b) qGGMRF (T = 3; = 0:5); (c) 3D NLM using plug-and-play; (d) 3D DSG-NLM using plug-and-play.

TABLE V
PLUG-AND-PLAY PARAMETER, σ_λ, FOR THE 10% SAMPLING CASE

Plug-and-play parameter, σ_λ	NLM	DSG-NLM	BM3D
Super ellipses	7.41	7.41	7.41
Zinc oxide nano-rods	9.16	9.16	9.16

the y-axis, spanning $[-65°, +65°]$. Fig. 4 shows a $0°$ tilt projection of the real aluminum spheres TEM data. Fig. 5 shows three reconstructions along the x-z plane. The NLM-based reconstruction has fewer smear artifacts than the qGGMRF reconstruction, and more clarity than the filtered backprojection reconstruction. Also, the NLM and DSG-NLM reconstructions have visibly suppressed missing-wedge artifact.

3) Silicon Dioxide (Real) Dataset: The silicon dioxide dataset (see Fig. 7) has 31 tilts about the y-axis, spanning $[-65°, +65°]$.

Fig. 7 shows a $0°$ tilt projection of the real silicon dioxide TEM data. Fig. 8 shows three reconstructions along the x-z plane. The NLM and DSG-NLM reconstructions have less

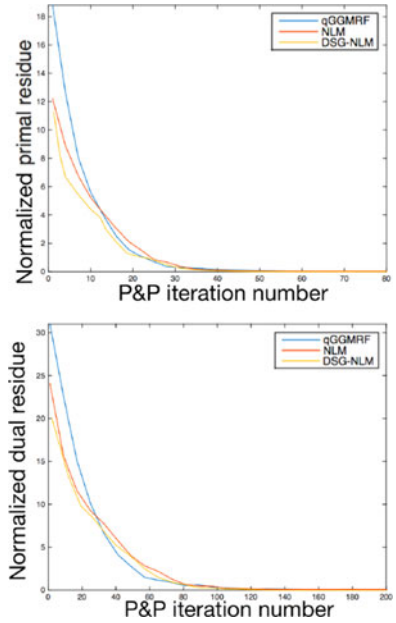


Fig. 6. Plug-and-play primal and dual residual convergence for tomographic reconstruction of (real) aluminum spheres. DSG-NLM achieves complete convergence.

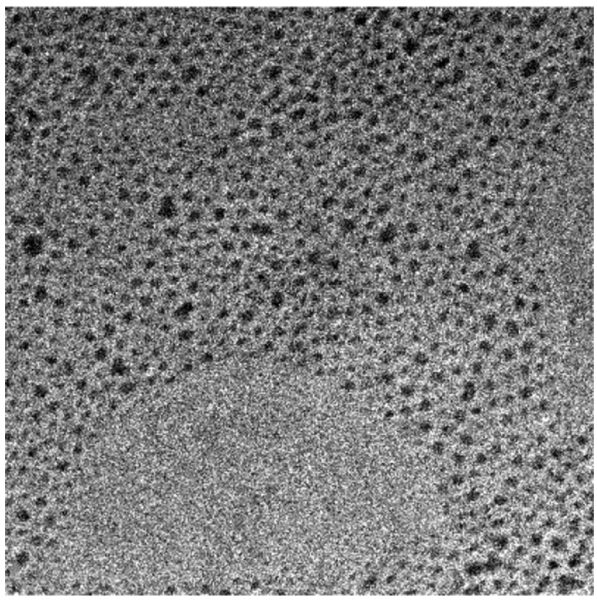


Fig. 7. Contrast-adjusted version of the $0°$ tilt of the silicon dioxide (real) dataset.

smear artifacts than the qGGMRF reconstruction, and far more clarity than the filtered backprojection reconstruction.

B. Sparse Interpolation

In this section, we present sparse interpolation results on both simulated and real microscope images (see Fig. 10 and 11). We show that a variety of denoising algorithms like NLM, DSG-NLM, and BM3D can be plugged in as prior models to reconstruct images from sparse samples. In all the sparse interpolation experiments, we stopped adapting the weights of the DSG-NLM after 12 iterations of the plug-and-play algorithm.

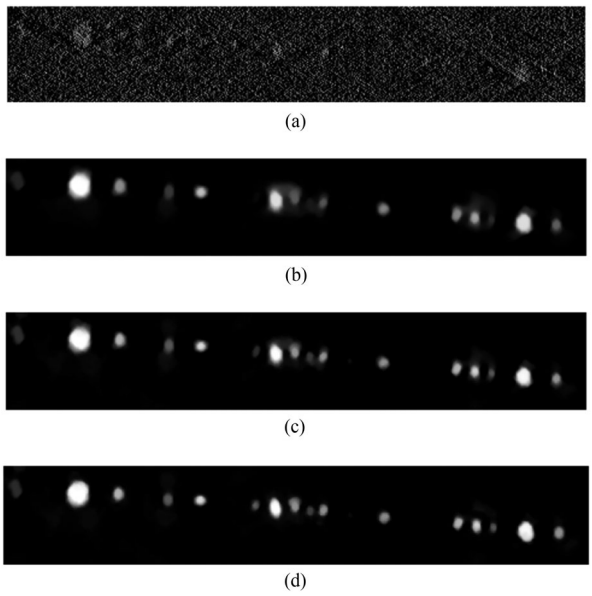


Fig. 8. Tomographic reconstruction of the silicon dioxide dataset. NLM reconstruction is clearer and has less smear artifacts. DSG-NLM reconstruction improves upon the NLM result through clear reconstruction of the structure on the left (a) Filtered Backprojection; (b) qGGMRF (T = 3; = 0:5); (c) 3D NLM using plug-and-play; (d) 3D DSG-NLM using plug-and-play.

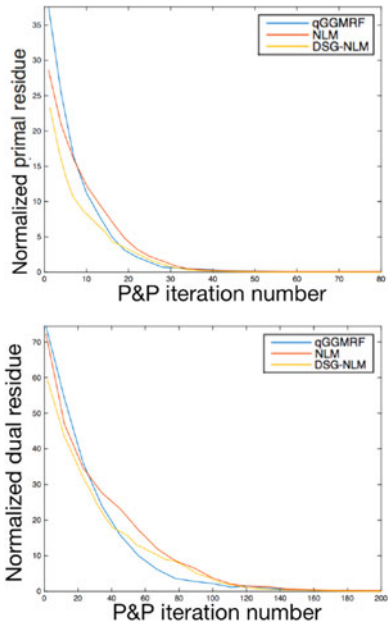


Fig. 9. Plug-and-play primal and dual residual convergence for tomographic reconstruction of (real) silicon dioxide. DSG-NLM achieves complete convergence.

TABLE VI
PLUG-AND-PLAY PARAMETER, β, FOR THE 10% SAMPLING CASE

Plug-and-play parameter, β	NLM	DSG-NLM	BM3D
Super ellipses	0.9	0.85	0.55
Zinc oxide nano-rods	0.81	0.79	0.49

TABLE VII
NORMALIZED PRIMAL RESIDUAL CONVERGENCE ERROR FOR THE 10% SAMPLING CASE (AFTER 150 PLUG-AND-PLAY ITERATIONS)

Primal residual convergence error	NLM	DSG-NLM	BM3D
Super ellipses	1.31×10^{-3}	2.89×10^{-8}	1.20×10^{-3}
Zinc oxide nano-rods	2.02×10^{-3}	1.33×10^{-9}	1.14×10^{-3}

TABLE VIII
NORMALIZED DUAL RESIDUAL CONVERGENCE ERROR FOR THE 10% SAMPLING CASE (AFTER 150 PLUG-AND-PLAY ITERATIONS)

Dual residual convergence error	NLM	DSG-NLM	BM3D
Super ellipses	9.10×10^{-3}	9.93×10^{-8}	8.71×10^{-3}
Zinc oxide nano-rods	1.14×10^{-2}	7.68×10^{-9}	3.23×10^{-2}

TABLE IX
INTERPOLATION ERROR (AFTER 150 PLUG-AND-PLAY ITERATIONS): NORMALIZED RMSE OF THE INTERPOLATED IMAGE COMPARED TO THE GROUND TRUTH

Interpolation error	Method	5% random sampling	10% random sampling
Super ellipses	Shepard	10.61%	8.99%
	NLM	8.51%	7.12%
	DSG-NLM	**8.10%**	**6.79%**
	BM3D	9.75%	7.46%
Zinc oxide nano-rods	Shepard	6.01%	5.49%
	NLM	4.35%	3.67%
	DSG-NLM	**4.32%**	**3.37%**
	BM3D	4.72%	3.80%

The plug-and-play parameters used are given in Tables V and Table VI.

Our first dataset is a set of simulated super ellipses that mimic the shapes of several material grains like Ni-Cr-Al alloy [65]. The next dataset is a real microscope image of zinc oxide nano-rods [66]. All the images are scaled to the range $[0, 255]$. Furthermore, no noise was added to either of the images, thus reducing the sparse interpolation inversion operator to the simple form of equation (21).

In all experiments, the plug-and-play sparse interpolation results are clearer than Shepard interpolation results. We observe fromTable IX that DSG-NLM typically results in the least RMS interpolation error, though for some images we have observed BM3D can produce lower RMSE. The RMSE values are normalized as $\frac{\|x - \hat{x}\|_2}{\|x\|_2}$, where $\hat{x}$ is the interpolated image and x is the ground truth image. Furthermore, we can see from Tables VII and VIII that DSG-NLM makes plug-and-play converge fully.

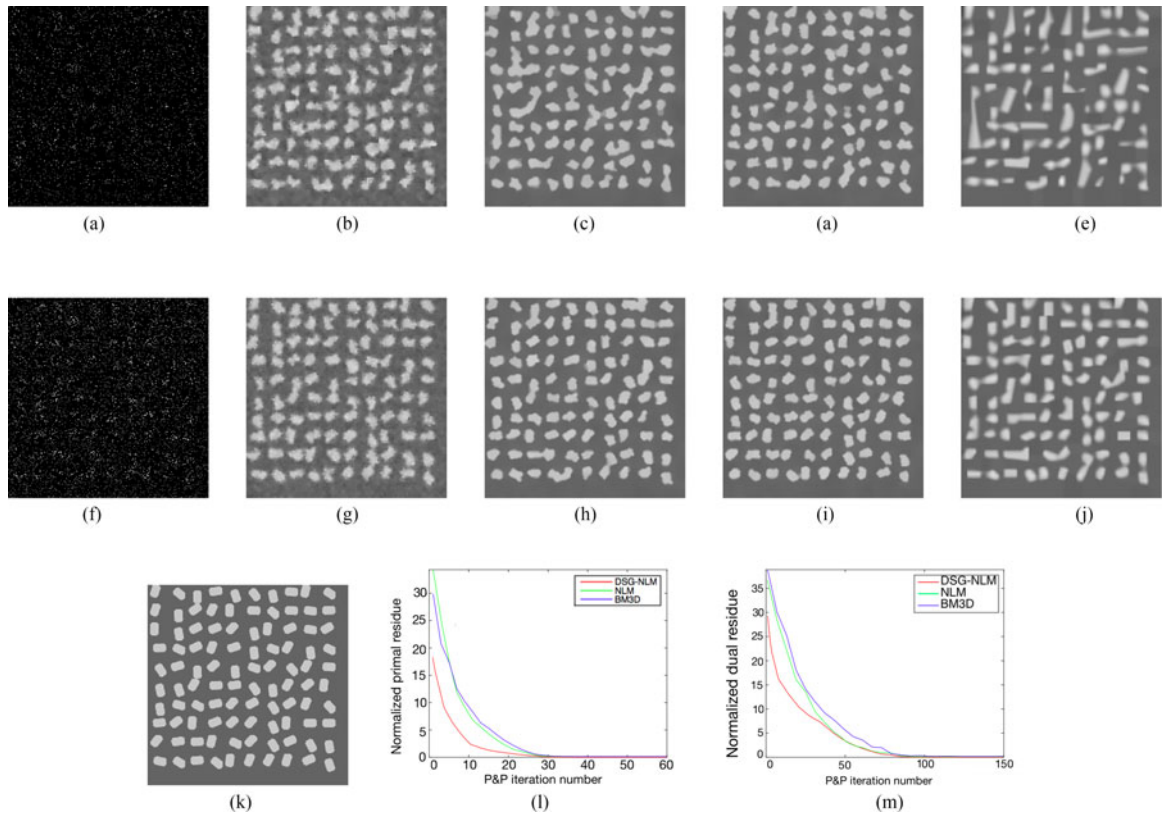


Fig. 10. Interpolation of a 256×256 grayscale image of a set of super ellipses (a) 5% sampling; (b) Shepard 5% random sampling; (c) NLM 5% random sampling; (d) DSG-NLM 5% random sampling; (e) BM3D 5% random sampling; (f) 10% sampling; (g) Shepard 10%; (h) NLM 10%; (i) DSG-NLM 10%; (j) BM3D 10%; (k) Ground truth; (l) Primal residual convergence for 10% sampling; (m) Dual residual convergence for 10% sampling.

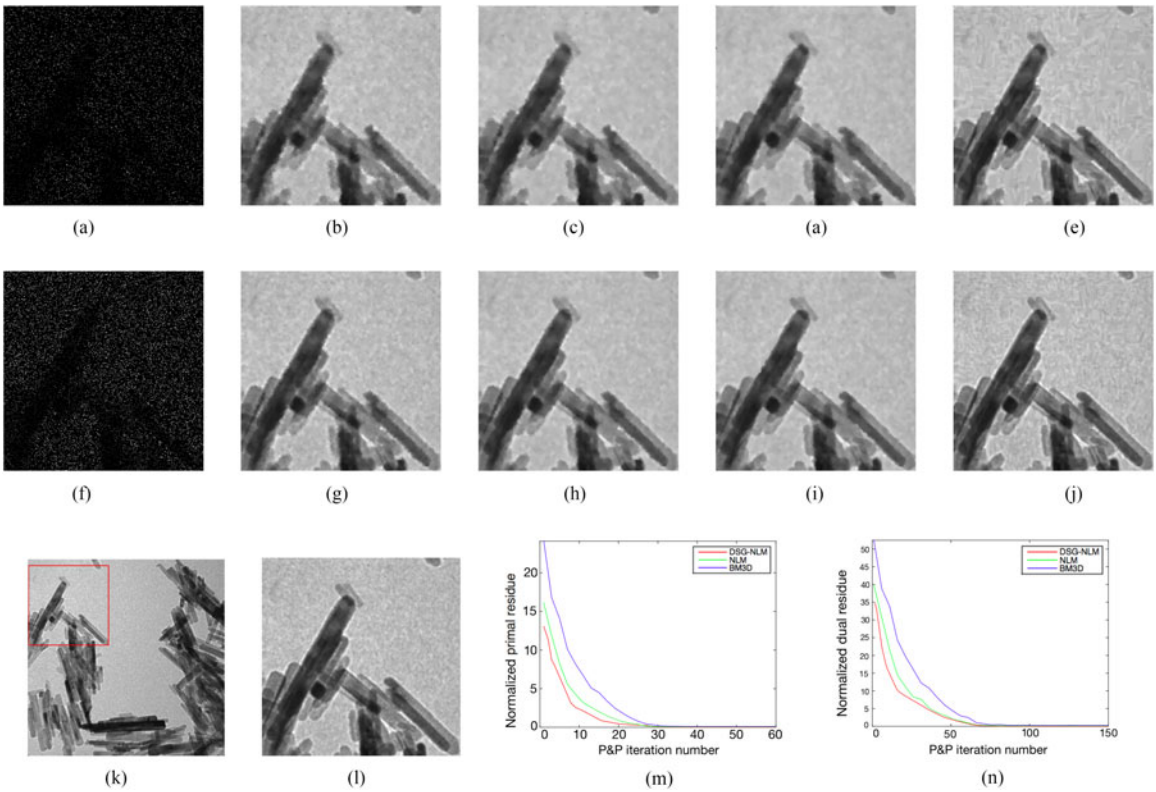


Fig. 11. Interpolation of a 414×414 grayscale image of zinc oxide nano-rods (a) 5% sampling; (b) Shepard 5%; (c) NLM 5%; (d) DSG-NLM 5%; (e) BM3D 5%; (f) 10% sampling; (g) Shepard 10%; (h) NLM 10%; (i) DSG-NLM 10%; (j) BM3D 10%; (k) Ground truthfull view; (l) Ground truthzoomed into the red box; (m) Primal residual convergence for 10% sampling; (n) Dual residual convergence for 10% sampling.

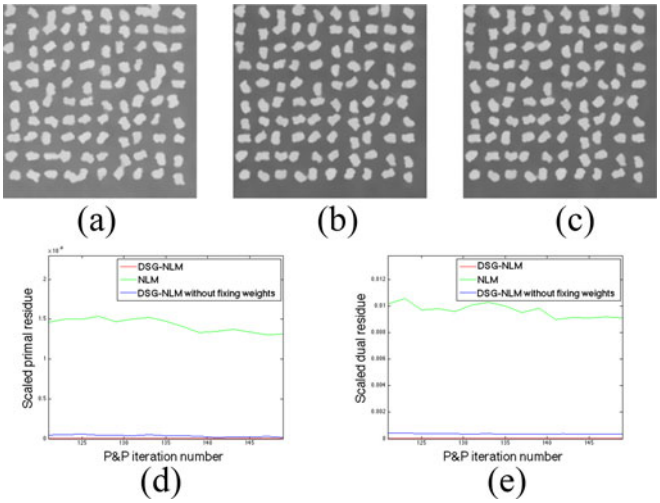


Fig. 12. Interpolation of simulated image of super ellipses for the convergence analysis of DSG-NLM (a) NLM 10%; (b) DSG-NLM (adaptive weights); (c) DSG-NLM (weights fixed after 12 iterations); (d) Zoomed-in primal residual convergence; (e) Zoomed-in dual residual convergence.

TABLE X
CONVERGENCE AND INTERPOLATIONS ERRORS FOR 10% SPARSE INTERPOLATION OF SIMULATED SUPER ELLIPSES (AFTER 150 PLUG-AND-PLAY ITERATIONS)

	NLM	DSG-NLM (adaptive weights)	DSG-NLM (fixed weights)
Primary residual convergence error	1.31×10^{-3}	1.75×10^{-5}	2.89×10^{-8}
Dual residual convergence error	9.10×10^{-3}	3.59×10^{-4}	9.93×10^{-8}
Interpolation error (normalized RMSE)	7.12%	6.72%	6.79%

C. Importance of Having a Fixed Doubly Stochastic Gradient for Plug-and-Play Convergence

In this section, in order to illustrate the importance of our denoising operator having a fixed doubly stochastic gradient for plug-and-play convergence, we present a sparse interpolation experiment with the same super ellipses simulated dataset as above. In addition to the NLM, and DSG-NLM, we use another version of DSG-NLM where we do not stop adapting the filter weights. We see from Fig. 12 and Table X that when the DSG-NLM weights are adapted every iteration, the primal and dual residuals enter a limit cycle and do not seem to fully converge.

Also, there is a very small improvement in RMSE when the DSG-NLM weights are adapted with every iteration. However, this slight reduction in RMSE comes at the price of increased computation.

Finally, we have empirically observed that even if the (standard) NLM weights are fixed after a few plug-and-play iterations, plug-and-play does not converge fully.

VIII. CONCLUSION

Microscope images of material and biological samples contain several repeating structures at distant locations. High quality reconstruction of these samples is possible by exploiting non-local repetitive structures. Though model-based iterative reconstruction (MBIR) could in principle exploit these repetitions, practically choosing the appropriate log probability term is very challenging. To solve this problem, we presented the "plug-and-play" framework which is based on ADMM. ADMM is a popular method to decouple the log likelihood and the log prior probability terms in the MBIR cost function. Plug-and-play takes ADMM one step further by replacing the optimization step related to the prior model by a denoising operation. This approach has two major advantages: First, it allows the use of a variety of modern denoising operators as implicit prior models; and second, it allows for more modular implementation of software systems for the solution of complex inverse problems.

We next presented and proved theoretical conditions for convergence of the plug-and-play algorithm which depend on the gradient of the denoising operator being a doubly stochastic matrix. We also re-designed the non-local means (NLM) denoising algorithm to have a doubly stochastic gradient, thereby ensuring plug-and-play convergence.

In order to demonstrate the value of our method, we applied the plug-and-play algorithm to two important problems: bright field electron tomography and sparse image interpolation. The results indicate that the plug-and-play algorithm when used with the NLM and DSG-NLM priors were able to reduce artifacts, improve clarity, and reduce RMSE (for the simulated dataset) as compared to the filtered back-projection and qGGMRF reconstructions. Then we performed sparse interpolation on simulated and real microscope images with as little as 5% of the pixels sampled–using three denoising operators: NLM, doubly stochastic gradient NLM (DSG-NLM), and BM3D. We then compared the results against Shepard's interpolation as the baseline. In all experiments, DSG-NLM resulted in the least RMSE and also complete convergence of the plug-and-play algorithm, as predicted by theory.

APPENDIX A
PROOF OF PLUG AND PLAY CONVERGENCE THEOREM

This appendix provides a proof of Theorem III.1. We start by defining a proximal mapping as any function $H : \mathbb{R}^N \to \mathbb{R}^N$ which can be expressed in the form

$$H(x) = \arg\min_{v \in \mathbb{R}^N} \left\{ \frac{\|x - v\|^2}{2} + s(v) \right\}, \tag{24}$$

where $s : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is a proper closed convex function on $\mathbb{R}^N$. With this definition, we can formally state the theorem proved by Moreau in 1965 [49] which gives necessary and sufficient conditions for when H is a proximal mapping.

Theorem A.1 (Moreau 1965 [49]): A function $H : \mathbb{R}^N \to \mathbb{R}^N$ is a proximal mapping if and only if

(1) H is non-expansive and,

(2) H is the sub-gradient of a convex function $\phi : \mathbb{R}^N \to \mathbb{R}$.

In fact, if there exists a function $\phi : \mathbb{R}^N \to \mathbb{R}$ such that $\forall x \in \mathbb{R}^N$

$$H(x) = \nabla\phi(x) \ ,$$

then we say that H is a conservative function or vector field. The concept of conservative functions is widely used on electromagnetics, for example. The next well known theorem (see for example [50, Theorem 2.6, p. 527]). gives necessary and sufficient conditions for a continuously differentiable function to be conservative on $\mathbb{R}^N$.

Theorem A.2: Let $H : \mathbb{R}^N \to \mathbb{R}^N$ be a continuously differentiable function. Then $H(x)$ is conservative if and only if $\forall x \in \mathbb{R}^N$, $\nabla H(x) = [\nabla H(x)]^t$.

In general, the sum of two proper closed convex functions, $h = f + g$, is not necessarily proper. This is because the intersection of the two sets $A = \{x \in \mathbb{R}^N : f(x) < \infty\}$ and $B = \{x \in \mathbb{R}^N : g(x) < \infty\}$ might be empty. Therefore, the following lemma will be needed in order to handle the addition of proper closed convex functions.

Lemma A.3: Let f and g both be proper closed convex functions and let $h = f + g$ be proper. Then h is proper, closed, and convex.

Proof: A proper convex function is closed if and only if it is lower semi-continuous. Therefore, both f and g must be lower semi-continuous. This implies that h is also lower semi-continuous. Since h is formed by the sum of two convex functions, it must be convex. Putting this together, h is proper, convex, and lower semi-continuous, and therfore it must be closed. Therefore, h is a proper, closed, and convex function on $\mathbb{R}^N$. ■

Finally, we give a general result relating the existence of a fixed point of a proximal map and the minimum of the associated function s.

Lemma A.4: Let H be a proximal mapping for a proper, closed, convex function s on $\mathbb{R}^N$. Then $H(x_0) = x_0$ if and only if $s(x_0)$ is the global minimum of s.

Proof: First suppose $s(x_0) \leq s(v)$ for all $v \in \mathbb{R}^N$. Then the unique minimizer of $M(v) = \|x_0 - v\|^2/2 + s(v)$ is x_0, so $H(x_0) = x_0$ is a fixed point. Next suppose there exists x_1 with $s(x_1) < s(x_0)$, and let $u = x_1 - x_0$ and $\Delta = s(x_1) - s(x_0) < 0$. Since s is convex, for all $r \in [0, 1]$ we have $s(x_0 + ru) \leq s(x_0) + r\Delta$, and hence

$$\frac{\|x_0 - (x_0 + ru)\|^2}{2} + s(x_0 + ru) \leq \frac{\|u\|^2}{2} r^2 + s(x_0) + r\Delta.$$

Since $\Delta < 0$, there exists $r_0 > 0$ so that $r_0(\Delta + r_0\|u\|^2/2) < 0$. Take $v_0 = x_0 + r_0 u$ to get $M(v_0) < M(x_0)$, and so $H(x_0) \neq x_0$. ■

As an immediate corollary, if H has a fixed point, then s attains its finite minimum, hence is bounded below by some constant $c > -\infty$.

Using these results, we next provide a proof of Theorem III.1. Without loss of generality, we will assume $\beta = 1$ and $\sigma_n^2 = 1$ in order to simplify the notation of the proof.

Proof of Theorem III.1 . To show that H is a proximal mapping for some proper, closed, and convex function $s(x)$, we use Moreau's result stated above in Theorem A.1. Note first that $\nabla H(x)$ is assumed to be symmetric, so Theorem A.2 implies that $H(x)$ is conservative and hence there exists a function ϕ so that

$$H(x) = \nabla\phi(x) \ .$$

Furthermore, by condition 2, $\nabla H(x)$ has real eigenvalues in the range $[0, 1]$. Since the eigenvalues are non-negative, ϕ must be convex. Furthermore, since the eigenvalues are ≤ 1, H must also be non-expansive. Since both conditions are satisfied, Moreau's Theorem implies that H is a proximal mapping of some proper, closed, and convex function $s(x)$; i.e., there exists a proper, closed, and convex function, $s(x)$, on $\mathbb{R}^N$ such that H can be expressed as

$$H(x) = \arg\min_{v \in \mathbb{R}^N} \left\{ \frac{\|x - v\|^2}{2} + s(v) \right\} \ . \tag{25}$$

We next show result 2 of the theorem, that a MAP estimate exists. This is equivalent to saying that the function $h(x) = l(x) + s(x)$ takes on its global minimum value for some $x = \hat{x}_{\text{MAP}}$.

By condition 3 of Theorem III.1 there exists an x and y such that $y = H(x)$ and $l(y) < \infty$. Since, $y = H(x)$ we also know that $s(y) < \infty$. Therefore, $h(y) < \infty$ and h is proper. By Lemma A.3, h must also be proper, closed, and convex.

Now to show that $h(x)$ takes on its global minimum, we need only show that there exists a threshold $\alpha \in \mathbb{R}$ such that the sublevel set of h is a non-empty compact set; that is,

$$A_\alpha = \{x \in \mathbb{R}^N : h(x) \leq \alpha\}$$

is a non-empty, compact subset of $\mathbb{R}^N$. Since h is a closed function, A_α must be a closed set. Therefore, it is necessary only to show that A_α is nonempty and bounded.

Define

$$p^* = \inf_{x \in \mathbb{R}^N} h(x) \ .$$

Since $h(x)$ is proper, we know that $p^* < \infty$. Select any $\alpha > p^*$, in which case A_α is nonempty.

To show that A_α is bounded, note that condition 4 of Theorem III.1 (that H has a fixed point) with Lemma A.4 implies that s attains its minimum. Hence there is a constant c so that $s(x) \geq c$ for all $x \in \mathbb{R}^N$. By condition 5 of Theorem III.1 (lower bound on $l(x)$), there exists $r > 0$ so that

$$f(r) > |c| + \alpha \ .$$

In this case, for all $\|x\| > r$, we have that

$$\begin{aligned} h(x) &= l(x) + s(x) \\ &\geq f(r) - |c| \\ &> \alpha. \end{aligned}$$

Therefore, if $x \in A_\alpha$, then $\|x\| \leq r$, and hence A_α is a nonempty bounded and therefore compact set. Consequently, h must take on its global minimum value for some value $\hat{x}_{\text{MAP}}$ in the compact set A_α.

Finally, we show result 3 of the theorem, that the plug-and-play algorithm converges. Since the plug-and-play algorithm is just an application of the ADMM algorithm, we can use standard ADMM convergence theorems. We use the standard theorem as

stated in [24, p. 16]. This depends on two assumptions. The first assumption is that $l(x)$ and $s(x)$ must be proper, closed, and convex functions, which we have already shown. The second assumption is that the standard (un-augmented) Lagrangian must have a saddle point.

The standard Lagrangian for this problem is given by

$$L(x, v; \lambda) = l(x) + s(v) + \lambda^t (x - v) \,, \tag{26}$$

and the associated dual function is denoted by

$$g(\lambda) = \inf_{x,v \in \mathbb{R}^N} L(x, v; \lambda) \,.$$

We say that $x^* \in \mathbb{R}^N$, $v^* \in \mathbb{R}^N$, $\lambda^* \in \mathbb{R}^K$ are a saddle point if for all x, v, and λ,

$$L(x, v; \lambda^*) \geq L(x^*, y^*; \lambda^*) \geq L(x^*, y^*; \lambda) \,.$$

Now we have already proved that a solution to our optimization problem exists and is given by $x^* = v^* = \hat{x}_{\mathrm{MAP}}$. So we know that the primal problem has a solution given by

$$\begin{aligned} p^* &= \inf_{\substack{x,v \in \mathbb{R}^N \\ v = x}} \{l(x) + s(v)\} \\ &= l(x^*) + s(v^*) \,. \end{aligned}$$

Now the pair (x^*, v^*) is a strictly feasible solution to the constrained optimization problem because x^* and v^* meet the constraint and they both fall within the open set $\mathbb{R}^N$. This means Slater's conditions hold, and by Slater's theorem, strong duality must also hold for some λ^* [67], [68]. More specifically, we know that there must exist a $\lambda^* \in \mathbb{R}^N$ such that

$$p^* = g(\lambda^*) \,.$$

Using this result, we have that for any x and v,

$$\begin{aligned} L(x^*, v^*; \lambda^*) &= l(x^*) + s(v^*) + [\lambda^*]^t (x^* - v^*) \\ &= l(x^*) + s(v^*) \\ &= p^* = g(\lambda^*) \\ &\leq L(x, v; \lambda^*). \end{aligned}$$

So we have that $L(x, v;\ \lambda^*) \geq L(x^*, v^*;\ \lambda^*)$. Furthermore since $x^* = v^*$, we know that $L(x^*, v^*; \lambda^*) = L(x^*, v^*; \lambda)$ for all λ. Putting together these two results, we have that $L(x, v; \lambda^*) \geq L(x^*, y^*; \lambda^*) \geq L(x^*, y^*; \lambda)$, thus proving the existence of a saddle point of the un-augmented Lagrangian, $L(x, v; \lambda)$.

Adapting the theorem of [24, p. 16], we then have the stated convergence results of equation (10).

$$\lim_{k \to \infty} \{x^{(k)} - v^{(k)}\} = 0;$$

$$\lim_{k \to \infty} \{l(x^{(k)}) + s(v^{(k)})\} = p^*. \qquad \blacksquare$$

APPENDIX B
PROOF OF THEOREM IV.1

To prove that the DSG-NLM matrix W generated by equations (11) through (15) is a symmetric doubly stochastic matrix with eigenvalues in the interval $[0, 1]$, we start by first noticing that the final resulting matrix is symmetric because all five steps of equations (11) through (15) are symmetric in s and r, so they produce symmetric weights with the property that $w_{s,r} = w_{r,s}$.

Next notice that equation (14) scales $w_{s,r}$ so that the sums of rows and columns are ≤ 1. Equation (15) then adds the non-negative value $1 - \sum_{r \in \Omega_s} w_{s,r}$ to the diagonal of the matrix to produce a final matrix with non-negative entries and rows and columns that sum to 1. Putting this together, we have shown that W is a symmetric, doubly stochastic matrix.

Finally, we show that the eigenvalues of W fall in the interval $[0, 1]$. Since $N_s \geq 1$, any pixel can be reached from any other by a path with each step restricted to a neighboring pixel. This implies that W is irreducible, and the fact that each diagonal entry is positive implies that W is primitive, so the Perron-Frobenius theorem [69], [70] implies that its eigenvalues fall in the interval $[-1, 1]$.

Furthermore, W must also be positive definite. To see this, note that Bochner's theorem [71] implies that $\exp\left\{-\|P\|_2^2\right\}$ and $\Lambda(s)$ are positive definite functions on their respective domains (since each is the Fourier transform of a non-negative function). From this and the Schur Product Theorem, the matrix W resulting from equation (11) is non-negative definite. Furthermore, the result of equation (12) is also non-negative definite because it results from multiplication on the left and right by the same positive diagonal matrix. Finally, equation (14) scales the matrix by a positive number and equation (15) adds a non-negative value to the diagonal of the matrix, so the final matrix must be non-negative definite with eigenvalues that lie in the interval $[0, 1]$. ■

ACKNOWLEDGEMENTS

The authors would like to thank M. DeGraef, professor of material science at Carnegie Mellon University, for providing simulated aluminum spheres tomography datasets.

REFERENCES

[1] K. Grünewald *et al.*, "Three-dimensional structure of herpes simplex virus from cryo-electron tomography," *Science*, vol. 302, no. 5649, pp. 1396–1398, 2003.
[2] M. Bárcena and A. J. Koster, "Electron tomography in life science," *Semin. Cell Develop. Biol.*, vol. 20, no. 8, pp. 920–930, 2009.
[3] P. A. Midgley and R. E. Dunin-Borkowski, "Electron tomography and holography in materials science," *Nature Mater.*, vol. 8, no. 4, pp. 271–280, 2009.
[4] H. S. Anderson, J. Ilic-Helms, B. Rohrer, J. Wheeler, and K. Larson, "Sparse imaging for fast electron microscopy," *Proc. SPIE*, vol. 8657, 2013, Art. ID. 86570C.
[5] A. Stevens, H. Yang, L. Carin, I. Arslan, and N. D. Browning, "The potential for Bayesian compressive sensing to significantly reduce electron dose in high-resolution STEM images," *Microscopy*, vol. 63, no. 1, pp. 41–51, 2014.
[6] A. Mohammad-Djafari, "Joint estimation of parameters and hyperparameters in a bayesian approach of solving inverse problems," in *Proc. Int. Conf. Image Process.*, Sep. 1996, vol. 1, pp. 473–476.
[7] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Process.*, vol. 2, no. 3, pp. 296–310, Jul.1993.
[8] C. A. Bouman, *Model Based Image and Signal Processing*, 2013.
[9] S. Z. Sullivan *et al.*, "High frame-rate multichannel beam-scanning microscopy based on lissajous trajectories," *Opt. Express*, vol. 22, pp. 24224–24234, 2014.
[10] S. Oh, A. Milstein, C. Bouman, and K. Webb, "A general framework for nonlinear multigrid inversion," *IEEE Trans. Image Process.*, vol. 14, no. 1, pp. 125–140, Jan. 2005.

[11] E. Mumcuoglu, R. Leahy, S. Cherry, and Z. Zhou, "Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 687–701, Dec. 1994.

[12] D. B. Husarik *et al.*, "Radiation dose reduction in abdominal computed tomography during the late hepatic arterial phase using a model-based iterative reconstruction algorithm: How low can we go?" *Investigative Radiol.*, vol. 47, no. 8, pp. 468–474, 2012.

[13] S. Venkatakrishnan, L. Drummy, M. A. Jackson, M. De Graef, J. Simmons, and C. Bouman, "A model based iterative reconstruction algorithm for high angle annular dark field-scanning transmission electron microscope (HAADF-STEM) tomography," *IEEE Trans. Image Process.*, vol. 22, no. 11, pp. 4532–4544, Nov. 2013.

[14] S. Venkatakrishnan, L. Drummy, M. Jackson, M. D. Graef, J. Simmons, and C. Bouman, "Model-based iterative reconstruction for bright-field electron tomography," *IEEE Trans. Comput. Imag.*, vol. 1, no. 1, pp. 1–15, Mar. 2015.

[15] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, 2005.

[16] A. Wong and J. Orchard, "A nonlocal-means approach to exemplar-based inpainting," in *Proc. 15th IEEE Intl. Conf. Image Process.*, 2008, pp. 2600–2603.

[17] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. IEEE 12th Int. Conf. Comput. Vision*, 2009, pp. 2272–2279.

[18] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[19] Y. Chen, J. Ma, Q. Feng, L. Luo, P. Shi, and W. Chen, "Nonlocal prior Bayesian tomographic reconstruction," *J. Math. Imag. Vision*, vol. 30, no. 2, pp. 133–146, 2008.

[20] S. Chun, Y. Dewaraja, and J. Fessler, "Alternating direction method of multiplier for emission tomography with non-local regularizers," in *Proc. Fully 3D Image Reconstruction Radiol. Nucl. Med.*, 2013, pp. 62–75.

[21] R. Glowinski and A. Marroco, "Sur l'approximation, par elements finis d'ordre un, et la resolution, par penalisation-dualite d'une classe de problemes de dirichlet non lineaires," *ESAIM: Math. Model. Numer. Anal.-Modél Math Anal Numérique*, vol. 9, no. R2, pp. 41–76, 1975.

[22] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Comput. Math. Appl.*, vol. 2, no. 1, pp. 17–40, 1976.

[23] J. Eckstein and D. P. Bertsekas, "On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, nos. 1–3, pp. 293–318, 1992.

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

[25] R. C. Fair, "On the robust estimation of econometric models," in *Annals of Economic and Social Measurement*, vol. 3, no. 4. Cambridge, MA, USA: Nat. Bureau Econ. Res., 1974, pp. 117–128.

[26] Z. Yang and M. Jacob, "Nonlocal regularization of inverse problems: A unified variational framework," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3192–3203, Sep. 2013.

[27] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," *Multiscale Model. Simul.*, vol. 7, no. 3, pp. 1005–1028, 2008.

[28] G. Peyré, S. Bougleux, and L. Cohen, "Non-local regularization of inverse problems," in *Proc. 10th Eur. Conf. Comput. Vision*, 2008, pp. 57–68.

[29] C. Guillemot and O. Le Meur, "Image inpainting: Overview and recent advances," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 127–144, Jan. 2014.

[30] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proc. 27th Annu. Conf. Comput. Graph. Interactive Tech.*, 2000, pp. 417–424.

[31] M. Elad, J.-L. Starck, P. Querre, and D. Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *Appl. Comput. Harmon. Anal.*, vol. 19, no. 3, pp. 340–358, 2005.

[32] W. Dong, L. Zhang, R. Lukac, and G. Shi, "Sparse representation based image interpolation with nonlocal autoregressive modeling," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1382–1394, Apr. 2013.

[33] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, Mar. 2010.

[34] J. Yang *et al.*, "Compressive sensing by learning a Gaussian mixture model from measurements," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 106–119, Jan. 2015.

[35] F. Li and T. Zeng, "A universal variational framework for sparsity-based image inpainting," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4242–4254, Oct. 2014.

[36] A. Danielyan, V. Katkovnik, and K. Egiazarian, "BM3D frames and variational image deblurring," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1715–1728, Apr. 2012.

[37] P. Milanfar, "A tour of modern image filtering: New insights and methods, both practical and theoretical," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 106–128, Sep. 2013.

[38] P. Chatterjee and P. Milanfar, "Patch-based near-optimal image denoising," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1635–1649, Apr. 2012.

[39] H. Talebi and P. Milanfar, "Global image denoising," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 755–768, Feb. 2014.

[40] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proc. IEEE Global Conf. Signal Inf. Process.*, 2013, pp. 945–948.

[41] A. Rond, R. Giryes, and M. Elad, "Poisson inverse problems by the plug-and-play scheme," 2015, *arXiv:1511.02500*.

[42] S. H. Chan, X. Wang, and O. A. Elgendy, "Plug-and-play ADMM for image restoration: Fixed point convergence and applications," 2016, *arXiv:1605.01710*.

[43] A. M. Teodoro, J. M. Bioucas-Dias, and M. A. Figueiredo, "Image restoration and reconstruction using variable splitting and class-adapted image priors," 2016, *arXiv:1602.04052*.

[44] S. Oymak, B. Recht, and M. Soltanolkotabi, "Sharp time–data tradeoffs for linear inverse problems," 2015, *arXiv:1507.04793*.

[45] K. Egiazarian, A. Foi, and V. Katkovnik, "Compressed sensing image reconstruction via recursive spatially adaptive filtering," in *Proc. IEEE Int. Conf. Image Process.*, 2007, vol. 1, pp. 549–552.

[46] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image restoration by sparse 3d transform-domain collaborative filtering," in *Proc. Int. Soc. Opt. Photon. Electron. Imaging*, 2008, pp. 681 207–681 207.

[47] F. Heide *et al.*, "Flexisp: A flexible camera image processing framework," *ACM Trans. Graph.*, vol. 33, no. 6, 2014, Art. no. 231.

[48] S. Sreehari, S. Venkatakrishnan, L. F. Drummy, J. P. Simmons, and C. A. Bouman, "Advanced prior modeling for 3D bright field electron tomography," *Proc. SPIE*, 2015, Art. ID. 940108.

[49] J.-J. Moreau, "Proximité et dualité dans un espace hilbertien," *Bull. Soc. Math. France*, vol. 93, pp. 273–299, 1965.

[50] R. E. Williamson, R. H. Crowell, and H. F. Trotter, *Calculus of Vector Functions*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1972.

[51] P. Milanfar, "Symmetrizing smoothing filters," *SIAM J. Imag. Sci.*, vol. 6, no. 1, pp. 263–284, 2013.

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

[53] J. M. Bioucas-Dias and M. A. Figueiredo, "A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2992–3004, Jan. 2008.

[54] S. J. Wright, R. D. Nowak, and M. A. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2479–2493, Jul.2009.

[55] M. Afonso, J. 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, Sep. 2010.

[56] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, 2005, vol. 2, pp. 60–65.

[57] R. Autieri, G. Ferraiuolo, and V. Pascazio, "Bayesian regularization in nonlinear imaging: Reconstructions from experimental data in nonlinearized microwave tomography," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 801–813, Feb. 2011.

[58] B. Wahlberg, S. Boyd, M. Annergren, and Y. Wang, "An ADMM algorithm for a class of total variation regularized estimation problems," *IFAC Proceedings Volumes*, vol. 45, no. 16, pp. 83–88, 2012.

[59] E. Ghadimi, A. Teixeira, I. Shames, and M. Johansson, "Optimal parameter selection for the alternating direction method of multipliers (ADMM): Quadratic problems," *IEEE Trans. Autom. Control*, vol. 60, no. 3, pp. 644–658, Feb. 2015.

[60] A. Kheradmand and P. Milanfar, "Motion deblurring with graph Laplacian regularization," *Proc. SPIE*, vol. 9404, 2015, Art. ID. 94040C.

[61] M. De Graef, *Introduction to Conventional Transmission Electron Microscopy*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[62] R. Stevenson and E. Delp, "Fitting curves with discontinuities," in *Proc. 1st Int. Workshop Robust Comput. Vision*, 1990, pp. 127–136.

[63] B. Wohlberg, "Adaptive penalty parameter selection for ADMM algorithms," 2016, submitted for publication.
[64] J.-B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multislice helical CT," *Med. Phys.*, vol. 34, no. 11, pp. 4526–4544, 2007.
[65] J. MacSleyne, M. Uchic, J. Simmons, and M. De Graef, "Three-dimensional analysis of secondary γ' precipitates in René-88 DT and UMF-20 superalloys," *Acta Materialia*, vol. 57, no. 20, pp. 6251–6267, 2009.
[66] H. Koerner *et al.*, "ZnO nanorod-thermoplastic polyurethane nanocomposites: Morphology and shape memory performance," *Macromolecules*, vol. 42, no. 22, pp. 8933–8942, 2009.
[67] M. Slater, "Lagrange multipliers revisited," Cowles Found. Res. Econ., Yale Univ., New Haven, CT, USA, *Tech. Rep. 80*, Nov. 1950.
[68] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2009.
[69] O. Perron, "Zur theorie der matrices," *Mathematische Annalen*, vol. 64, no. 2, pp. 248–263, 1907.
[70] G. Frobenius, "Über matrizen aus positiven elementen," S. B. Preuss. Akad. Wiss. Berlin, Germany, 1908.
[71] Y. Katznelson, *An Introduction to Harmonic Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

Suhas Sreehari (S'15) received the B.E. degree in telecommunications engineering from Businayana Mukundadas Sreenivasaiah College of Engineering, Visveswaraiah Technological University, Belgaum, India, in 2010, and the M.A.Sc. degree in electrical and computer engineering from the University of Windsor, Windsor, ON, Canada, in 2012. He is currently working toward the Ph.D. degree in the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA. His research interests include inverse imaging problems, computational imaging, and mathematical modeling.

S. V. Venkatakrishnan received the B.Tech. degree in electronics and communication engineering from the National Institute of Technology Tiruchirappalli, Tiruchirappalli, India, in 2007, and the M.S. and Ph.D. degrees in electrical and computer engineering from the Purdue University, West Lafayette, IN, USA, in 2009 and 2014, respectively. He is currently a Postdoctoral Fellow at the Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. His research interests include statistical information processing, inverse problems, computational imaging, and machine learning.

Brendt Wohlberg (SM'13) received the B.Sc. (Hons.) degree in applied mathematics, the M.Sc. degree in applied science, and the Ph.D. degree in electrical engineering from the University of Cape Town, Cape Town, South Africa, in 1990, 1993, and 1996, respectively. He is currently a Staff Scientist in the Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, USA. His research interest includes signal and image processing inverse problems, with an emphasis on sparse representations and exemplar-based methods. He was an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING from 2010 to 2014, and is currently the Chair of the Computational Imaging Special Interest Group of the IEEE Signal Processing Society and an Associate Editor of the IEEE TRANSACTIONS ON COMPUTATIONAL IMAGING.

Gregery T. Buzzard received the degrees in violin performance, computer science, and mathematics from Michigan State University, East Lansing, MI, USA, and the Ph.D. degree in mathematics from the University of Michigan, Ann Arbor, MI, in 1995. He held postdoctoral positions at Indiana University and Cornell University before joining the mathematics faculty, in 2002, at the Purdue University, West Lafayette, IN, USA, where he currently serves as the Head of the Department of Mathematics. His research interests include dynamical systems, experiment design, and uncertainty quantification.

Lawrence F. Drummy received the B.S. degree in physics from Rensselaer Polytechnic Institute, Troy, NY, USA, in 1998, while developing image processing tools for scanning tunneling microscopy, and the Ph.D. degree from the Department of Materials Science and Engineering, University of Michigan, Ann Arbor, MI, USA, in 2003, performing research on flexible electronic materials. He is a Materials Engineer in the Soft Matter Materials Branch, Functional Materials Division, Materials and Manufacturing Directorate, Air Force Research Laboratory, Dayton, OH, USA . His research interests include tomography and inverse problems, signal and image processing, high-resolution transmission electron microscopy, nanocomposites, organic–inorganic interfaces, structural proteins, metamaterials, and organic electronics.

Jeffrey P. Simmons (M'96) received the B.S. degree in metallurgical engineering from New Mexico Institute of Mining and Technology, Socorro, NM, USA, in 1983, the M.S. degree in metallurgical engineering and materials science, and the Ph.D. degree in materials science and engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 1985 and 1992, respectively. He is a Research Scientist in the Structural Materials Division, Materials and Manufacturing Directorate, Air Force Research Laboratory, Dayton, OH, USA. He is currently working with developing mathematical algorithms for data analysis of emerging large digital datasets produced by advances in microscope characterization capabilities.

Charles A. Bouman (S'86–M'89–SM'97–F'01) received the B.S.E.E. degree from the University of Pennsylvania, Philadelphia, PA, USA, in 1981, the M.S. degree from the University of California, Berkeley, CA, USA, in 1982, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, USA, in 1989. From 1982 to 1985, he was a Full Staff Member at the MIT Lincoln Laboratory. In 1989, he joined the faculty of Purdue University, West Lafayette, IN, USA, where he is currently the Showalter Professor of electrical and computer engineering and biomedical engineering. His research interests include the use of statistical image models, multiscale techniques, fast algorithms in applications including tomographic reconstruction, medical imaging, and document rendering and acquisition. He is a Fellow of the American Institute for Medical and Biological Engineering, a Fellow of the Society for Imaging Science and Technology, a Fellow of the SPIE Professional Society, and a Fellow of the National Academy of Inventors.