Abstract—In a particle beam microscope, a raster-scanned focused beam of particles interacts with a sample to generate a secondary electron (SE) signal pixel by pixel. Conventionally formed micrographs are noisy because of limitations on acquisition time and dose. The raw SE data can be modeled with a compound Poisson (Neyman Type A) likelihood, which implies variance that is signal-dependent and greater than the variation in the underlying particle–sample interaction. These statistical properties make methods that assume additive white Gaussian noise ineffective. This paper introduces methods for micrograph denoising that exploit approximations of the data likelihood that vary in accuracy and computational complexity. These are used within the plug-and-play ADMM framework in combination with denoising by total variation regularization, BM3D, and DuCNN. Methods are provided for both conventional and time-resolved measurements. In simulations representative of helium ion microscopy and scanning electron microscopy, significant improvements in mean-squared error (MSE), structural similarity index measure (SSIM), and qualitative appearance are obtained. Reductions in MSE are by factors ranging from 5.4 to 11.3.

Index Terms—electron microscopy, focused ion beam, helium ion microscopy, Neyman Type A distribution, Poisson processes

I. INTRODUCTION

In this paper, we apply the plug-and-play (PnP) alternating direction method of multipliers (ADMM) framework [12] to create several denoising algorithms for particle beam micrographs. PnP ADMM algorithms alternate between an inversion step that minimizes a data fidelity term and a denoising step. Their distinguishing feature is that the denoising step can be the application of an arbitrary off-the-shelf denoiser, which may not be inspired by any explicit prior for the signal of interest. PnP ADMM algorithms have achieved state-of-the-art performance in many applications [13]–[16] and are gaining in popularity. The Neyman Type A measurement distribution makes naïve use of the negative log-likelihood as the data fidelity term computationally unattractive. Here we introduce and compare alternative data fidelity terms with varying accuracy and computational complexities that operate on both conventional data and data that is not conventionally acquired.

Though commercial PBMs more commonly use scintillators and photomultiplier tubes, direct SE detection offers a higher signal-to-noise ratio [17] and is easier to model. Therefore we concentrate on direct SE detection, where the readout is the number of detected SEs. Furthermore, we recently introduced the concept of time-resolved (TR) measurement in particle beam microscopy in [18], [19]. With TR measurement, each pixel’s dwell time is divided into disjoint subacquisitions. The number of SEs is measured separately for each subacquisition, forming a vector of measurements at each pixel. In our previous work, we demonstrated that these TR measurements are more informative than a single measurement with the same total dose. This was exhibited through Fisher information and through the lower mean-squared error (MSE) of estimates computed from TR measurements. It was shown in [20], [21] that the TR techniques in [18], [19] are also robust to unknown and variable beam current, preventing stripe artifacts. More
recently, we established that TR measurements can be used to explicitly estimate unknown beam current from the same SE count data used to form a micrograph, without the use of a calibrated sample [22], [23].

TR estimation techniques introduced in [18], [19] operate pixelwise. The only previous regularized reconstructions from TR measurements are restricted to total variation (TV) regularization [24], and they do not emphasize efficient implementations [22], [25]. The TV regularizer is nonsmooth, necessitating the use of iterative algorithms to minimize the resulting cost function. Proximal approaches [26], such as the proximal gradient method (PGM) [27]–[29], and ADMM [30]–[32] have been adopted to avoid differentiating the regularizer. In this work, we use ADMM due to its modular structure and demonstrated success in a variety of applications [33], [34]. In addition to denoising by TV regularization, we plug in the extraordinarily successful BM3D image denoiser [35] and a pre-trained deep neural network denoiser DnCNN [36].

A. Main Contributions

- Introduction of data fidelity terms to make PnP ADMM efficiently applicable to denoising of particle beam micrographs. Five data fidelity terms are presented: one for conventional (non-TR) measurements, one to study performance assuming an oracle provides the number of incident particles, and three for TR measurements.

- Experimental evaluations of PnP ADMM in emulations of HIM and SEM. We combined the five data fidelity terms with three denoisers: TV, BM3D, and DnCNN. The improvement from regularization is greater for HIM than for SEM and for conventional data than for TR data. MSE reduction factors range from 5.4 for SEM with TR data to 11.3 for HIM with conventional data.

B. Outline

In Section II, we summarize the abstraction, measurement models and pixelwise estimators in PBM, and in Section III, we review the PnP ADMM framework. The key novelties of this paper are in Section IV, where we introduce data fidelity terms that allow the application of PnP ADMM to PBM denoising. The data fidelity terms have varying levels of complexity and accuracy to the negative log-likelihood function of the physical data generation process. Section V describes the collection of PnP ADMM algorithms we obtain by combining the data fidelity terms with three different denoisers. We present experimental results in both simulated HIM and SEM settings in Section VI. These show that time-resolved measurements and PnP ADMM provide significant improvements in estimation accuracy.

II. MEASUREMENT MODELS AND PIXELWISE ESTIMATORS

In this section, we introduce our measurement model for direct SE detection in PBM and several pixelwise estimators; see [18], [19] for additional details. The central goal of this paper is to develop regularized estimators analogous to these pixelwise estimators. While the incident particles may be electrons or ions, we refer to them as ions for simplicity.

A. Abstract Model

A sample is raster scanned with a focused beam of ions. During a fixed dwell time t at any one raster scan location, the number of incident ions M is well modeled as a Poisson random variable with mean $\lambda = \Lambda t$, where $\Lambda$ represents the known rate of incident ions per unit time. In particular, $\lambda$ is the number of incident ions per unit area and per unit time. Assuming that this differs substantially from Poisson-distributed data, notice the dependence of both the mean and variance on $\eta$ and that this differs substantially from Poisson-distributed data, assuming $\eta$ is not too small.

\begin{align}
\Pr(Y_k; \eta, \lambda) = & e^{-\lambda \eta} \sum_{m=0}^{\infty} \frac{(\lambda e^{-\eta})^m}{m!}, \\
& y = 0, 1, \ldots, \\
\text{mean} & \quad \mathbb{E}[Y] = \lambda \eta, \\
\text{variance} & \quad \text{Var}(Y) = \lambda \eta (\eta + 1).
\end{align}

Notice the dependence of both the mean and variance on $\eta$ and the dependence of the variance on $\eta + 1$. Notice that this differs substantially from Poisson-distributed data, assuming $\eta$ is not too small.

B. Conventional Measurement

The conventional measurement $y^C \in \mathbb{R}^d$ gathers measurements across all d pixels, with each entry drawn from the distribution in (1). Because the entries of $y^C$ are independent, its joint PMF is given by

\begin{align}
\Pr(y^C; \eta, \lambda) = & \prod_{k=1}^{d} \Pr(y_k; \eta_k, \lambda), \\
\end{align}

where $\eta_k$ is the SE yield at the kth pixel and $\lambda$ is the per-pixel dose.

C. Time-Resolved Measurement

With TR measurement, the per-pixel dwell time $t$ is split into $n$ sub-acquisitions, each of length $t/n$, yielding an $n$-length vector of measurements at each pixel. The vector $y \in \mathbb{R}^{dn}$ gathers these TR measurements across all pixels with the measurement vector at the kth pixel given by $y_k = [y_k^{(1)}, y_k^{(2)}, \ldots, y_k^{(n)}]$. Each observation is sampled from the distribution in (1) with $\lambda$ being replaced by $\lambda/n$. We note that at a given pixel, the conventional measurement may be obtained by summing the vector of time-resolved measurements:

\begin{align}
y^C_k = \sum_{i=1}^{n} y_k^{(i)}. \\
\end{align}

Although certain operating conditions may cause the beam current to stray from the desired setting due to contamination [37], we have shown that the unknown beam current can be estimated at each pixel using time-resolved measurements [22], [23].
The entries in $y$ are independent so its joint PMF is given by:

$$P_Y(y; \eta, \lambda) = \prod_{k=1}^{d} \prod_{i=1}^{n} P_Y(y_{k}^{(i)}; \eta_k, \lambda/n).$$  \hfill (6)

D. Pixelwise Estimators

We now review estimation methods that operate at a single pixel without regularization. We include an oracle estimator that relies on knowledge of the number of incident ions $M$ (which is not available in practice) and estimators that can be applied with conventional or TR data.

1) Conventional Estimator: From (2), scaling the observed SE counts at the $k$th pixel by $\lambda$ yields an unbiased estimator:

$$\hat{\eta}_k^{\text{conv}} = \frac{y_k^C}{\lambda}.$$  \hfill (7)

From (3), its mean-squared error (MSE) is $\eta_k(\eta_k + 1)/\lambda$. The $\eta_k + 1$ factor in the MSE stems from the excess variance in (3) compared with the variance of a Poisson random variable with the same mean. This excess variance can be attributed to source shot noise—the randomness of the number of incident ions.

2) Oracle Estimator: If one were able to know the number of incident ions $M_k$ at the $k$th pixel, dividing $y_k^C$ by $M_k$ will produce a superior estimator:

$$\hat{\eta}_k^{\text{oracle}} = \frac{y_k^C}{M_k},$$  \hfill (8)

which has the MSE

$$\text{MSE}(\hat{\eta}_k^{\text{oracle}}) = e^{-\lambda}(1 - e^{-\lambda})(\eta_k - \eta_0)^2 + \sum_{m=1}^{\infty} \frac{\eta_k \lambda^m}{m!} e^{-\lambda},$$  \hfill (9)

where $\eta_0$ is an arbitrarily assigned number when $M_k = 0$ \cite{18}. For large enough $\lambda$, the choice of $\eta_0$ has little effect on the MSE. Hence, for large $\lambda$, the MSE satisfies

$$\text{MSE}(\hat{\eta}_k^{\text{oracle}}) \approx \frac{\eta_k}{\lambda},$$  \hfill (10)

which eliminates the excess MSE due to the randomness of incident ions. We emphasize that such an estimator is unimplementable since $M_k$ cannot be known exactly from only observing $y_k^C$.

3) Quotient Mode Estimator: To approach the performance of the oracle estimator, we may naturally seek a proxy for $M_k$ that is computable from observed quantities. When $n$ is large enough, the dose for each subacquisition $\lambda/n$ becomes so small that the probability that more than one ion will arrive during one subacquisition is negligible. Assuming that $\eta$ is large enough, most ions will produce at least one SE. In this case, the number of subacquisitions at the $k$th pixel measuring a positive number of SEs,

$$L_k = \sum_{i=1}^{n} I_{\{y_k^{(i)} > 0\}},$$  \hfill (11)

is a good approximation for the number of incident ions $M_k$, where $I_{\{y_k^{(i)} > 0\}}$ is equal to 1 when $\{y_k^{(i)} > 0\}$ and is equal to 0 otherwise. Analogous to the oracle estimator in (8), the quotient mode (QM) estimator is defined as

$$\hat{\eta}_k^{\text{QM}} = \frac{y_k^C}{L_k} = \frac{y_k^{(1)} + y_k^{(2)} + \ldots + y_k^{(n)}}{\sum_{i=1}^{n} I_{\{y_k^{(i)}>0\}}}.$$  \hfill (12)

The QM name is taken from a similar concept proposed by John Notte in [38], where counting of the analog-domain pulses produced by SE bursts are adopted as the denominator of an estimator similar to (12). A closed-form expression for the MSE of $\hat{\eta}_k^{\text{QM}}$ is given in [19]. The MSE of $\hat{\eta}_k^{\text{QM}}$ is significantly lower than that of $\hat{\eta}_k^{\text{conv}}$ except when $\eta$ is small.

4) Lambert Quotient Mode Estimator: When $\eta$ is small, $L_k = \sum_{i=1}^{n} I_{\{y_k^{(i)}>0\}}$ significantly underestimates $M_k$ because the probability of an ion generating zero detected SEs cannot be neglected. In this case, the bias in $\hat{\eta}_k^{\text{QM}}$ caused by the underestimation can be reduced by replacing $L_k$ with $(1 - e^{-\eta K})^{-1} L_k$. Since the probability of an incident ion resulting in at least 1 detected SE is $1 - e^{-\eta K}$, the adjusted $(1 - e^{-\eta K})^{-1} \sum_{i=1}^{\infty} I_{\{y_k^{(i)}>0\}}$ is a more accurate estimate of $M_k$. Given that $\eta_k$ is unknown, this expression results in a transcendental equation, which has the solution

$$\hat{\eta}_k^{\text{LQM}} = W \left( -\hat{\eta}_k^{\text{QM}} e^{-\hat{\eta}_k^{\text{QM}}} \right) + \hat{\eta}_k^{\text{QM}},$$  \hfill (13)

where $W(\cdot)$ represents the Lambert W function \cite{39}. Hence, we name $\hat{\eta}_k^{\text{LQM}}$ the Lambert quotient mode (LQM) estimator. In MSE, $\hat{\eta}_k^{\text{LQM}}$ significantly improves upon $\hat{\eta}_k^{\text{QM}}$ when $\eta$ is small and is nearly indistinguishable from $\hat{\eta}_k^{\text{QM}}$ otherwise.

5) Time-Resolved Maximum Likelihood Estimator: With TR data at the $k$th pixel, $[y_k^{(1)}, y_k^{(2)}, \ldots, y_k^{(n)}]$, the value of $\eta$ that maximizes the joint likelihood is the time-resolved maximum likelihood (TRML) estimator:

$$\hat{\eta}_k^{\text{TRML}} = \arg\max_{\eta \in [0, \infty]} \prod_{i=1}^{n} P_Y(y_k^{(i)}; \eta, \lambda/n),$$  \hfill (14)

where $P_Y(\cdot; \cdot, \cdot)$ is given by (1). Since the decision variable is a scalar, solving the optimization via grid search is not impractical. According to [19, Fig. 5(a)], $\hat{\eta}_k^{\text{TRML}}$ has lower MSE than $\hat{\eta}_k^{\text{conv}}, \hat{\eta}_k^{\text{QM}}$, and $\hat{\eta}_k^{\text{LQM}}$ across all $\eta$ values. However, since that domination does not hold for continuous-time observations (see [19, Fig. 3(a)]), in the discrete-time setting of the present work, the relative performances of estimators may depend on the choices of $\lambda$ and $n$.

III. PLUG-AND-PLAY ADMM

In this section, we review the concepts of ADMM and PlnP-ADMM for image reconstruction incorporating image priors.

A. Problem Formulation

Given a vector of conventional measurements $y^C \in \mathbb{R}^d$ or TR measurements $y \in \mathbb{R}^{dn}$, we seek to reconstruct the underlying SE yield image $\eta \in \mathbb{R}^d$. This image reconstruction task can often be written as an optimization problem of the form

$$\hat{\eta} = \arg\min_{\eta} f(\eta) + \beta g(\eta),$$  \hfill (15)
where \(f\) is a data-fidelity term that encourages consistency with \(y^C\) or \(y\), \(g\) is a regularizer that promotes solutions with desirable properties, and \(\beta\) is a tuning parameter that controls the regularization strength.

From a Bayesian perspective, (15) can arise as the maximum a posteriori (MAP) estimator when

\[
f(\eta) = -\log p(y | \eta),
\]

(16)

the negative log-likelihood of the observation \(y\) (or similarly with \(y^C\)); and

\[
\beta g(\eta) = -\log p(\eta),
\]

(17)

the negative log prior of \(\eta\).

B. ADMM

ADMM converts the unconstrained problem in (15) into a constrained one:

\[
(\hat{\eta}, \hat{v}) = \arg\min_{\eta, v} f(\eta) + \beta g(v), \quad \text{subject to} \quad \eta = v.
\]

(18)

This constrained problem can be solved by minimizing its augmented Lagrangian function

\[
\mathcal{L}(\eta, v) = f(\eta) + \beta g(v) + u^T(\eta - v) + \frac{\rho}{2} \|\eta - v\|^2,
\]

(19)

where \(\rho\) is a penalty parameter and \(u\) is the Lagrangian multiplier. The solution can be obtained by iteratively solving the following subproblems:

\[
\eta^{(t+1)} = \arg\min_{\eta \in \mathbb{R}^d} f(\eta) + \frac{\rho}{2} \|\eta - (u^{(t)} - u^{(t)})\|^2,
\]

(20a)

\[
u^{(t+1)} = \arg\min_{v \in \mathbb{R}^d} g(v) + \frac{1}{2\sigma^2} \|v - (\eta^{(t+1)} + u^{(t)})\|^2,
\]

(20b)

\[
u^{(t+1)} = u^{(t)} + (\eta^{(t+1)} - u^{(t+1)}),
\]

(20c)

where \(\sigma = \sqrt{\beta/\rho}\).

C. Plug-in Denoiser

The data fidelity term and prior term are decoupled into two separate subproblems with similar structures. Specifically, (20a) can be viewed as an inversion step since \(f(\eta)\) is determined by the forward image measurement model, and (20b) can be recognized as a denoising step since \(g(\eta)\) represents an image prior. The latter subproblem is equivalent to Gaussian denoising on \(\eta^{(t+1)} + u^{(t)}\) with noise level \(\sigma = \sqrt{\beta/\rho}\). Based on this intuition, Venkatakrishnan et al. [12] proposed the PnP ADMM algorithm, which does not specify \(g\) explicitly. Instead, (20b) can be replaced with any off-the-shelf denoiser, denoted as \(D_\sigma\), to yield

\[
u^{(t+1)} = D_\sigma(\eta^{(t+1)} + u^{(t)}).
\]

(21)

Notice that when the regularizer is

\[
g(v) = \|v\|_{TV} = \sum_{i=1}^{d} \|[Dv]_i\|_2,
\]

where \(D\) denotes the discrete image gradient, (21) is a standard total variation denoising problem. The prior \(g(\cdot)\) can also be implicitly specified by a denoiser, such as the popular BM3D method [35], or by deep neural networks that enable us to leverage the powerful modeling capacity and flexibility of network architectures.

IV. DATA FIDELITY TERMS

In many applications of PnP ADMM, the data fidelity term \(f(\eta)\) is derived from a measurement process that involves linear mixing and signal-independent additive white Gaussian noise (AWGN); (16) then results in \(f(\eta) \propto \|y - Ay\|^2\) for some matrix \(A\), which is quite convenient for computations. In this work, we wish to apply PnP ADMM to particle beam micrograph denoising, where the challenge is not rooted in linear mixing. Instead, difficulties arise from the measurement likelihood function even though it is separable. To directly apply (16) with the measurement likelihood function (4) or (6) is highly inconvenient because of the form of the Neyman Type A PMF (1). This is an obstacle to regularized estimation of any form, whether or not one employs PnP ADMM.

This section introduces several data fidelity terms \(f(\eta)\) that vary in their closeness to (16) and their computational complexity. For conventional measurement data, we have a data fidelity term based on a spatially adapted Gaussian approximation. For oracle or TR data, the data fidelity terms have correspondences with the oracle, QM, LQM, and TRML estimators of Section II. In one case, (20a) has very low computational complexity because there is a simple closed-form solution; for the others, we derive the derivatives of \(f(\eta)\) with respect to the entries of \(\eta\). We will use these derivatives later to solve the optimization problems of Section V.

A. Gaussian

We may approximate the entries of \(y^C\) as independent Gaussian random variables, each with mean and variance given in (2) and (3):

\[
y^C_k \sim N(\lambda \eta_k, \lambda \eta_k(\eta_k + 1)).
\]

(22)

Simple point evaluation of the Gaussian probability density function gives a reasonable approximation of the Neyman Type A probability mass function provided that \(\lambda \eta_k\) is not small, and there is pointwise convergence of the moment generating function for \(\lambda \to \infty\) [40]. Omitting a constant term, the corresponding negative log-likelihood function is

\[
f_{conv}(\eta) = \sum_{k=1}^{d} \left( \frac{1}{2} \log \eta_k + \frac{1}{2} \log(\eta_k + 1) + \frac{(y^C_k - \lambda \eta_k)^2}{2 \lambda \eta_k(\eta_k + 1)} \right),
\]

(23)

with derivatives

\[
\frac{\partial f_{conv}(\eta)}{\partial \eta_k} = \frac{1}{2 \eta_k} + \frac{1}{2(\eta_k + 1)}
\]

\[
+ \frac{(\lambda \eta_k - y^C_k)(\lambda \eta_k + 2 \eta_k y^C_k + y^C_k)}{2 \lambda \eta_k^2(\eta_k + 1)^2}.
\]

(24)

To be able to replace a gradient method with a closed-form solution when we compute the inverse step (20a), we can further approximate (22) as

\[
y^C_k \sim N(\lambda \eta_k, \lambda \eta_k^{prev}(\eta_k^{prev} + 1)),
\]

(25)
where the variance is set to be independent of \( \eta_k \) by using \( \eta_k^{prev} \), an estimated value of \( \eta_k \) at the previous iteration. In this case, the observation \( y^C \) becomes a Gaussian random vector with constant, diagonal covariance matrix, and the negative log-likelihood function is

\[
f_{conv}(\eta) = \sum_{k=1}^{d} \frac{(y^C_k - \eta_k)(y^C_k - \eta_k)^T}{2\eta_k^{prev} + 1} \tag{26}
\]

after dropping terms that do not depend on \( \eta \). With this \( f \), the ADMM inversion step (20a) is separable over the \( d \) components, with closed-form solution

\[
\eta_k^{(t+1)} = \left( \eta_k^{prev} + \rho \right)^{-1} \left( \eta_k^{prev} + \rho \right) \eta_k^{(t)} + \rho v_k^{(t+1)} - u_k^{(t)} \tag{27}
\]

In the model (25), we are positing a variance for pixel \( k \) that gives spatially varying strength to the regularizer \( g \). When the detected \( y^C \) is unluckily small relative to a moderate or larger underlying true value \( \eta_k \), estimation will perform poorly if too much confidence is ascribed to the observed \( y^C \). To avoid this, performance is improved by allowing the local variance \( \lambda \eta_k^{prev} + 1 \) to depend on a neighborhood of pixel \( k \) rather than pixel \( k \) alone. Specifically, when computing \( \eta_k^{(t+1)} \), we choose \( \eta_k^{prev} \) to be the average of \( \eta_k^{(t)} \) across a neighborhood of pixel \( k \):

\[
\eta_k^{prev} = \frac{1}{9} \sum_{i \in N(k)} \eta_i^{(t)}, \tag{28}
\]

where \( N(k) \) is a \( 3 \times 3 \) patch centered at pixel \( k \).

**B. Oracle**

When the number of incident ions is exactly known, the number of SEs observed at the \( k \)th pixel may be modeled as a Poisson random variable with parameter \( M_k \eta_k \):

\[
y^C_k \sim \text{Poisson}(M_k \eta_k). \tag{29}
\]

The negative log-likelihood function is

\[
f_{oracle}(\eta) = \sum_{k=1}^{d} (M_k \eta_k - y^C_k \log \eta_k) \tag{30}
\]

after dropping terms that do not depend on \( \eta \). The derivative of \( f_{oracle}(\eta) \) with respect to an entry of \( \eta \) is

\[
\frac{\partial f_{oracle}(\eta)}{\partial \eta_k} = M_k - \frac{y^C_k}{\eta_k}. \tag{31}
\]

Although instruments are not capable of measuring \( M_k \), the oracle data fidelity term serves as an interesting benchmark.

**C. Quotient Mode**

As in (12), our QM data fidelity term uses the number of subacquisitions where more that one SE was observed as a proxy for \( M_k \). Here we have

\[
y^C_k \sim \text{Poisson}(L_k \eta_k), \tag{32}
\]

with \( L_k \) defined in (11). The negative log-likelihood function is

\[
f_{QM}(\eta) = \sum_{k=1}^{d} (L_k \eta_k - y^C_k \log \eta_k) \tag{33}
\]

after dropping terms that do not depend on \( \eta \). Its derivatives are

\[
\frac{\partial f_{QM}(\eta)}{\partial \eta_k} = L_k - \frac{y^C_k}{\eta_k}. \tag{34}
\]

Note that although the left hand side of (32) is the conventional measurement \( y^C_k \), \( L_k \) is computed using TR measurements.

**D. Lambert Quotient Mode**

Using the same adjustment as in (13) to compensate for the underestimate of \( M_k \) in \( L_k \), our QM data fidelity term is based upon the model

\[
y^C_k \sim \text{Poisson}((1 - e^{-\eta_k})^{-1} L_k \eta_k). \tag{35}
\]

The negative log-likelihood function after dropping terms that do not depend on \( \eta \) is

\[
f_{LQM}(\eta) = \sum_{k=1}^{d} \left( \frac{\eta_k}{1 - e^{-\eta_k}} L_k + y^C_k \log(1 - e^{-\eta_k}) \right), \tag{36}
\]

with derivatives

\[
\frac{\partial f_{LQM}(\eta)}{\partial \eta_k} = \frac{e^{\eta_k} \left( e^{\eta_k} - \eta_k - 1 \right) L_k}{(e^{\eta_k} - 1)^2} + \frac{y^C_k}{(e^{\eta_k} - 1) \eta_k}. \tag{37}
\]

For faster computation, (35) may be approximated as

\[
y^C_k \sim \text{Poisson}((1 - e^{-\eta_k})^{-1} L_k \eta_k), \tag{38}
\]

where \( \eta_k^{prev} \) is based on the previous iteration as in Section IV-A. Then

\[
f_{LQM}(\eta) \approx \frac{L_k \eta_k}{1 - e^{-\eta_k}} - y^C_k \log \eta_k \tag{39}
\]

and

\[
\frac{\partial f_{LQM}(\eta)}{\partial \eta_k} \approx \frac{L_k}{1 - e^{-\eta_k}} - \frac{y^C_k}{\eta_k}. \tag{40}
\]

**E. Time-Resolved Maximum Likelihood**

In our TRML data fidelity term, we use the full likelihood in (6). Here we have

\[
f_{TRML}(\eta) = -\log P_y(y \mid \eta, \lambda). \tag{41}
\]

To use derivatives of (41) directly with the substitution of (1) and (6) is computationally expensive and delicate. However, as derived in [22, App. D], the derivatives of (41) with respect to the entries in \( \eta \) are approximately

\[
\frac{\partial f_{TRML}(\eta)}{\partial \eta_k} \approx \frac{\lambda}{n} e^{-\eta_k} \left( y^C_k \right) + \sum_{i \in S} \frac{n + (2 \eta_k^{prev} - 1)}{n} \lambda e^{-\eta_k}, \tag{42}
\]
where $S = \{i : y_k^{(i)} > 0\}$. The approximation is accurate when $\lambda/n$ is small. The experimental results in Section VI are for $\lambda/n = 0.1$, which is small enough for accurate approximation and for most of the gains from TR measurement to be realized [19].

V. PROPOSED METHODS

In this work, we test the five data fidelity terms proposed in Section IV within the PnP ADMM framework using three different denoisers: TV-regularized least squares, BM3D [35], and a deep neural network. This section explains the overall algorithm design and different types of denoising.

A. Algorithm Overview

Algorithm 1 outlines the key steps of PnP ADMM, as detailed in Section III, adapted to our setting. Line 2, the inversion step, incorporates the data fidelity term. For the conventional data fidelity term (Section IV-A), it is computed with (27); in the other cases, the minimization is solved using gradient descent. Line 3 is the denoising step of (21). Line 4 updates the Lagrangian multiplier. These three steps are repeated within a loop until convergence is achieved. Here, convergence is declared when $\Delta t_{t+1} \leq \alpha$, where $\alpha$ is a threshold and

$$\Delta t_{t+1} := \frac{1}{\sqrt{d}} \left( \|\eta(t+1) - \eta(t)\|_2 + \|\nu(t+1) - \nu(t)\|_2 \right)$$

B. Denoisers

We employ three denoising methods.

1) Total Variation: Under this formulation, the denoising step is regularized least-squares estimation (20b) with regularizer $g(\eta) = \|\eta\|_{TV}$. The TV cost $\|\eta\|_{TV}$ is given by

$$\|\eta\|_{TV} = \sum_{i=1}^{d} \|D\eta_i\|_2,$$

where $D$ denotes the discrete image gradient. This cost term is designed to promote estimates that are piecewise smooth while maintaining edge features. The corresponding denoising step is solved iteratively as in [27].

2) BM3D: One branch of denoising algorithms exploits non-local similarity of image patches to recover a clean image from the noisy observation. BM3D [35] is one of the best among these methods. It groups image patches based on similarity, then applies collaborative filtering and recombines to yield a reconstructed image.

3) Deep Neural Network: In recent years, deep learning-based methods have achieved great success in image denoising tasks. Most existing deep neural networks use a large number of clean–noisy image pairs as training samples, which is one key factor contributing to performance improvements. However, collecting a large dataset of clean–noisy image pairs can be expensive and challenging, especially in PBM. Without access to such a dataset, [25] used synthetic data, generated using accurate knowledge of the PBM forward model, to train a deep neural network. The network’s weights were optimized to minimize the $L_2$ difference between reconstructed and ground truth images. This method did not explicitly leverage the model information during inversion but instead relied on the PBM model to generate a noisy counterpart to each clean training image. In contrast, we use the PnP ADMM framework to combine knowledge of the PBM model with the assistance of a deep neural network to extract meaningful features, resulting in improved performance. In the denoising step (Line 3 of Algorithm 1), we adopt a deep neural network called DnCNN [36]. It combines residual learning [41] and batch normalization [42], and it has been demonstrated to be effective in removing AWGN.

In the formulation of ADMM in Section III-B, the Gaussian denoising problem (20b) has a noise standard deviation $\sigma = \sqrt{\beta/\rho}$ derived directly from the regularization parameter $\beta$ and variable-splitting penalty $\rho$. This dependence on $\sigma$ can be troublesome in PnP ADMM because the denoiser may require training that depends on $\sigma$ or it may have no analogous parameter.

Denoiser scaling [43] is a method for introducing tunable regularization without assuming that a denoiser matched to a specific $\sigma$ value is available. Suppose only that some denoiser $D$ is given. Then a scaled denoiser is defined as

$$D_\mu(\eta) = (1/\mu)D(\mu\eta),$$

where $\mu > 0$ is the denoiser scaling parameter. This approach makes it possible to use a single pre-trained network across a variety of noise levels.

VI. EXPERIMENTAL RESULTS

In this section, we compare the performances of the proposed methods using simulated HIM and SEM micrograph images. While previous works [18], [19] have evaluated the QM, LQM, and TRML estimators of Section II-D, here we aim to compare how these and the benchmark conventional and oracle estimators perform in combination with the three denoisers in Section V-B, within the PnP ADMM framework. The five estimators are included in PnP ADMM through the data fidelity terms detailed in Section IV.

Experiment details: We take images from the ThermoFisher Scientific website as the ground truth and scale [3]https://www.fei.com/image-gallery/
them to $\eta \in [2, 8]$ to emulate HIM [44] and to $\eta \in [1, 2]$ for SEM. Given the ground truth image, we apply (6) to generate noisy TR measurements pseudorandomly using total dose $\lambda = 20$ split over $n = 200$ sub-acquisitions for each pixel for the HIM sample and total dose $\lambda = 50$ split over $n = 500$ sub-acquisitions per pixel for the SEM sample. The conventional measurement is obtained by summing over the $n$ subacquisitions at each pixel. All PnP ADMM experiments use a stopping threshold of $\alpha = 5 \times 10^{-4}$ as defined in Section V-A. We apply the scaled denoiser for estimations with DnCNN. Parameters $\beta$, $\mu$, and $\rho$ are tuned to minimize the MSE.

We pre-train the DnCNN on the Berkeley Segmentation Data Set and Benchmarks 500 (BSDS500) [45]. We corrupt each image with Gaussian noise with standard deviation of 25 for images on the scale of $\{0, 1, \ldots, 255\}$. The network is trained on clean–noisy image pairs to output reconstructed images to minimize their $\ell_2$-norm distances with their clean counterparts. After pre-training, the network’s weights are fixed and the network is used as a denoiser to reduce noise of an input image. For DnCNN-related methods, $\mu$ typically lies between 0.1 and 1.5 to give the best result. In addition, DnCNN-related methods are GPU-accelerated for faster computation.

Along with the combinations of data fidelity terms and denoisers yielding fifteen PnP ADMM methods (bottom three rows and right five columns in Figures 1 and 2), we compute eight additional estimates for comparison. To validate the use of regularization in general, we show the performance of pixelwise application of each of the five estimation methods (right five columns of the top row in Figures 1 and 2). To demonstrate the virtue of accurate modeling of the acquisition process, we also compare with naïve estimators assuming $y^C/\lambda$ is Gaussian with mean $\eta$ and constant variance (bottom three rows of the left column in Figures 1 and 2). These are computed by solving

$$\hat{\eta}^G = \arg\min_{\eta \in [0, \infty)^d} \|\eta - y^C/\lambda\|^2_2 + \beta g(\eta)$$

for each of the three regularizers $g$ given explicitly or implicitly in Section V-B. Absolute error images for the sub-regions marked by the green boxes are included for each micrograph as well.

**Pixelwise estimators (top rows):** Without regularization, the best TR methods significantly outperform the conventional method, as demonstrated in [19] as well. The MSE reduction of the best of these – TRML – is by a factor of 4.1 for HIM and 1.4 by SEM; these factors are consistent with theoretical predictions [19].

**Data fidelity terms (comparisons by column):** Each estimator type (method obtained by varying the denoiser for one choice of data fidelity term) is improved by regularization for both the HIM and SEM samples. In comparing the methods using TR data and including regularization, it is almost uniformly true that the TRML data fidelity term gives the best performance, LQM second, and QM the worst; the only exception is unexpectedly poor performance of the TRML data fidelity term combined with the DnCNN denoiser for the SEM sample. For the HIM sample, the performance of the Gaussian data fidelity term is similar to QM and worse than TRML by a factor of 2.1 in MSE. For the SEM sample, the performance of the Gaussian data fidelity term is competitive with the best TR method – a bit worse than TRML for the TV and BM3D denoisers, and a bit better than LQM for the DnCNN denoiser. These empirical results with regularization are consistent with theoretical results from [18], [19] that show increasing utility of TR measurements over conventional measurements as $\eta$ increases.

For the SEM sample, the high MSE of QM is due to the large bias of this estimator when $\eta < 2$ [19]. Interestingly, the SSIM metric when using the QM data fidelity term is not bad, which is consistent with invariance properties of SSIM [46]. Biases are detailed below in the discussion of Figure 3.

**Denoisers (comparisons by row):** Within each estimator type, for the HIM sample, the DnCNN denoiser is the best, followed by BM3D and TV. Not only does the DnCNN denoiser reduce the MSE approximately by a factor of 1.3 compared with TV denoising under the TRML data fidelity, it also recovers finer details.

Denoiser comparisons for the SEM sample have different trends. TV denoising achieves the smallest MSE and highest SSIM for each of the data fidelity terms. DnCNN performance is better than BM3D except for the case of TRML data fidelity, where it is worse by a very small margin. Though the best overall quantitative performance is achieved by combining TV denoising with the TRML data fidelity term, we must acknowledge that for this sample, naïve application of TV denoising, BM3D, and DnCNN are similar in MSE. Visually, we observe that the naïve estimators oversmooth their results and result in worse SSIM. Arguably, TV regularization also results in oversmoothing and lack of contrast. This may justify employing a deep network, which can represent more complex structural properties of the images.

**Accuracy summary:** Table I summarizes the quantitative results from Figures 1 and 2. The PnP ADMM columns present the best metrics over the choices of denoisers and applicable data fidelity terms. The summary highlights the effectiveness of the PnP ADMM methods introduced in this paper for both the HIM and SEM samples and both conventional and time-resolved data. MSE reductions are by factors from 5.4 to 11.3.

**Computation times:** Table II shows the average running time of Algorithm 1 for different data fidelity terms and denoisers for the sample in Figure 1. For the same denoiser, the Gaussian data fidelity term gives the lowest running time due to the closed-form solution in (27). The running times for both the HIM and SEM samples. In comparing the methods using TR data and including regularization, it is almost universally true that the TRML data fidelity term gives the best performance, LQM second, and QM the worst; the only exception is unexpectedly poor performance of the TRML data fidelity term combined with the DnCNN denoiser for the SEM sample. For the HIM sample, the performance of the Gaussian data fidelity term is similar to QM and worse than TRML by a factor of 2.1 in MSE. For the SEM sample, the performance of the Gaussian data fidelity term is competitive with the best TR method – a bit worse than TRML for the TV and BM3D denoisers, and a bit better than LQM for the DnCNN denoiser. These empirical results with regularization are consistent with theoretical results from [18], [19] that show increasing utility of TR measurements over conventional measurements as $\eta$ increases.

For the SEM sample, the high MSE of QM is due to the large bias of this estimator when $\eta < 2$ [19]. Interestingly, the SSIM metric when using the QM data fidelity term is not bad, which is consistent with invariance properties of SSIM [46]. Biases are detailed below in the discussion of Figure 3.

**Denoisers (comparisons by row):** Within each estimator type, for the HIM sample, the DnCNN denoiser is the best, followed by BM3D and TV. Not only does the DnCNN denoiser reduce the MSE approximately by a factor of 1.3 compared with TV denoising under the TRML data fidelity, it also recovers finer details.

Denoiser comparisons for the SEM sample have different trends. TV denoising achieves the smallest MSE and highest SSIM for each of the data fidelity terms. DnCNN performance is better than BM3D except for the case of TRML data fidelity, where it is worse by a very small margin. Though the best overall quantitative performance is achieved by combining TV denoising with the TRML data fidelity term, we must acknowledge that for this sample, naïve application of TV denoising, BM3D, and DnCNN are similar in MSE. Visually, we observe that the naïve estimators oversmooth their results and result in worse SSIM. Arguably, TV regularization also results in oversmoothing and lack of contrast. This may justify employing a deep network, which can represent more complex structural properties of the images.

**Accuracy summary:** Table I summarizes the quantitative results from Figures 1 and 2. The PnP ADMM columns present the best metrics over the choices of denoisers and applicable data fidelity terms. The summary highlights the effectiveness of the PnP ADMM methods introduced in this paper for both the HIM and SEM samples and both conventional and time-resolved data. MSE reductions are by factors from 5.4 to 11.3.

**Computation times:** Table II shows the average running time of Algorithm 1 for different data fidelity terms and denoisers for the sample in Figure 1. For the same denoiser, the Gaussian data fidelity term gives the lowest running time due to the closed-form solution in (27). The running times...
of QM, LQM, and oracle data fidelity terms are comparable, whereas the TRML method is the slowest because of the complicated form of the gradient (42). Comparing the running times by choice of denoiser, DnCNN-related methods are more efficient thanks to the advantages of GPU acceleration. TV-related methods are the second fastest followed by BM3D-related methods, which are time-consuming due to grouping similar image patches.

Bias–variance decomposition: Figure 3 provides an additional comparison of the results obtained with the five data fidelity terms combined with TV regularization. Monte Carlo simulations were computed on three different HIM examples with $\eta \in [2, 8]$ using total dose $\lambda = 20$ split over $n = 200$ sub-acquisitions. By separating the pixels by their ground truth $\eta$ values into 20 bins of equal width, we can estimate the bias, variance, and MSE as functions of $\eta$. Using the Gaussian data fidelity term (which implies conventional rather than TR data) results in large bias for high $\eta$ and the largest variance of all

| Data fidelity | Gaussian | Oracle | QM | LQM | TRML |
|---------------|----------|--------|----|-----|------|
| **Denoiser**  |          |        |    |     |      |
| TV            | 1.7      | 21.0   | 20.9 | 22.2 | 264.6 |
| BM3D          | 77.2     | 133.8  | 137.2 | 147.6 | 931.1 |
| DnCNN         | 2.0      | 9.4    | 9.4  | 10.9 | 137.3 |

Fig. 1: HIM sample with ground truth $\eta \in [2, 8]$, total dose $\lambda = 20$, and $n = 200$. All images are shown on the same scale as in the ground truth image. Absolute error images for the sub-region in each green square are included for reconstructed micrographs.
the methods, causing high MSE. The QM method has a large bias for small \(\eta\), which is corrected by the LQM method. The QM and LQM methods almost overlap for moderate and higher \(\eta\). The performance of the TRML method is close to that of the oracle, which is unimplementable in practice.

VII. CONCLUSION

In this paper, we develop plug-and-play ADMM method for estimation of the mean SE yield \(\eta\) in particle beam microscopy. Because of the Neyman Type A likelihood for PBM data, using the negative log-likelihood as the data fidelity term would be computationally intractable. We introduce data fidelity terms with different computational complexity and accuracies applicable to conventional or time-resolved measurements, and we compare their efficacies when combined with three different denoisers. In synthetic experiments emulating helium ion microscopy and scanning electron microscopy, we demonstrate that our approaches outperform pixelwise (non-regularized) methods substantially in MSE, SSIM, and qualitative appearance; MSE reduction is by a factor of 5.4 to 11.3. In the case of using a DnCNN denoiser, our use of denoiser scaling allows us to use a single pre-trained network to handle a variety of noise levels.

VIII. ACKNOWLEDGMENT

The authors thank Akshay Agarwal for insightful discussions. The authors also acknowledge the Boston University
Fig. 3: Comparisons of different data fidelities with TV regularization as a function of \( \eta \). Conventional, QM, LQM, TRML and oracle estimators are simulated on multiple HIM examples with \( \eta \in [2, 8] \) using total dose \( \lambda = 20 \) split over \( n = 200 \) sub-acquisitions. (a) Bias. (b) Variance. (c) MSE.

Research Computing Services group for providing computational resource support which has contributed to the results reported in this paper.

REFERENCES

[1] D. McMullan, “Scanning electron microscopy 1928–1965,” *Scanning*, vol. 17, no. 3, pp. 175–185, 1995.

[2] B. W. Ward, J. A. Notte, and N. P. Economou, “Helium ion microscope: A new tool for nanoscale microscopy and metrology,” *J. Vacuum Sci. Technol. B*, vol. 24, no. 6, pp. 2871–2874, Nov. 2006.

[3] J. Notte and J. Huang, *The Helium Ion Microscope*. Cham: Springer International Publishing, 2016, pp. 3–30. [Online]. Available: https://doi.org/10.1007/978-3-319-41990-9_1

[4] J. Morgan, J. Notte, R. Hill, and B. Ward, “An introduction to the helium ion microscope,” *Microscopy Today*, vol. 14, no. 4, pp. 24–31, 2006.

[5] D. Emmrich, A. Beyer, A. Nadzeyka, S. Bauerdiick, J. C. Meyer, J. Kottakoski, and A. Gölzhäuser, “Nanopore fabrication and characterization by helium ion microscopy,” *Appl. Phys. Lett.*, vol. 108, no. 16, p. 163103, 2016.

[6] S. Bazargan, N. F. Heinig, J. F. Rios, and K. T. Leung, “Electronic resolution for focused ion beams,” *J. Vac. Sci. & Technol. B*, vol. 12, pp. 523–543, 2019.

[7] S. Boyd, N. Parikh, and E. Chu, *Distributed Optimization and Statistical Learning* (Foundations and Trends, 2011). Now Publishers Inc.

[8] J. Orloff, L. W. Swanson, and M. Utlaut, “Fundamental limits to imaging resolution for focused ion beams,” *J. Vac. Sci. Technol. B*, vol. 14, no. 6, pp. 3759–3763, 1996.

[9] J. A. Notte and J. Huang, “The Helium Ion Microscope,” *Ann. Rev. Analytical Chem.*, vol. 4, pp. 127–239, 2011.

[10] L. Watkins, S. W. Seidel, M. Peng, A. Agarwal, C. C. Yu, and V. K. Goyal, “Prevention beats removal: Avoiding stripe artifacts from current variation in particle beam microscopy through time-resolved sensing,” *Microsc. Microanal.*, vol. 27, no. 51, pp. 422–425, Aug. 2021.

[11] J. Orloff, L. W. Swanson, and M. Utlaut, “Fundamental limits to imaging resolution for focused ion beams,” *J. Vac. Sci. Technol. B*, vol. 14, no. 6, pp. 3759–3763, 1996.

[12] L. W. Swanson, M. Utlaut, and J. Orloff, “A new tool for nanoscale microscopy and metrology,” *J. Vacuum Sci. Technol. B*, vol. 17, no. 3, pp. 175–185, 1995.

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

[14] K. Zhang, W. Zuo, S. Gu, and L. Zhang, “Learning deep CNN denoiser prior for image restoration,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2017, pp. 3929–3938.

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

[16] S. Boyd, N. Parikh, and E. Chu, “Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. Now Foundations and Trends, 2011.
[31] 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, Sep. 2010.

[32] 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, no. 1, pp. 293–318, 1992.

[33] A. Matakos, S. Ramani, and J. A. Fessler, “Accelerated edge-preserving image restoration without boundary artifacts,” *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 2019–2029, May 2013.

[34] M. S. C. Almeida and M. Figueiredo, “Deconvolving images with unknown boundaries using the alternating direction method of multipliers,” *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3074–3086, Aug. 2013.

[35] 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.

[36] 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, Jul. 2017.

[37] F. F. Rahman, J. A. Notte, R. H. Livengood, and S. Tan, “Observation of synchronized atomic motions in the field ion microscope,” *Ultramicroscopy*, vol. 126, pp. 10–18, 2013.

[38] J. A. Notte, “Imaging with helium ions – A new detector regime with new challenges and new opportunities,” American Vacuum Society 60th Int. Symp. Exhib., Oct. 31 (presented, no proceedings), 2013.

[39] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, “On the Lambert W function,” *Adv. Comput. Math.*, vol. 5, no. 1, pp. 329–359, Dec. 1996.

[40] D. C. Martin and S. K. Katti, “Approximations to the Neyman type A distribution for practical problems,” *Biometrika*, vol. 18, no. 3, pp. 354–364, Sep. 1992.

[41] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2016, pp. 770–778.

[42] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[43] X. Xu, J. Liu, Y. Sun, B. Wohlberg, and U. S. Kamilov, “Boosting the performance of plug-and-play priors via denoiser scaling,” arXiv:2002.11546, Feb. 2020.

[44] J. Notte, B. Ward, N. Economou, R. Hill, R. Percival, L. Farkas, and S. McVey, “An introduction to the helium ion microscope,” in *AIP Conference Proceedings*, vol. 931. AIP, 2007, pp. 489–496.

[45] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, “Contour detection and hierarchical image segmentation,” *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.

[46] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.