IEEE Open Journal of Signal Processing

Received 23 January 2024; revised 14 April 2024; accepted 14 April 2024. Date of publication 26 April 2024; date of current version 17 May 2024. The review of this article was arranged by Associate Editor Shuai Wan.

Digital Object Identifier 10.1109/OJSP.2024.3394369

Short Paper

Denoiser-Based Projections for 2D Super-Resolution MRA

JONATHAN SHANI[1], TOM TIRER [2], RAJA GIRYES [1] (Senior Member, IEEE), AND TAMIR BENDORY [1] (Senior Member, IEEE)

[1]Tel Aviv University, Tel Aviv 69978, Israel
[2]Bar Ilan University, Ramat Gan 5290002, Israel

CORRESPONDING AUTHOR: TAMIR BENDORY (email: bendory@tauex.tau.ac.il).

The work of Tom Tirer was supported by ISF under Grant 1940/23. The work of Raja Giryes was supported in part by the ERC-StG under Grant 757497 and in part by KLA grants. The work of Tamir Bendory was supported in part by BSF under Grant 2020159, in part by NSF-BSF under Grant 2019752, and in part by ISF under Grant 1924/21.

ABSTRACT We study the 2D super-resolution multi-reference alignment (SR-MRA) problem: estimating an image from its down-sampled, circularly translated, and noisy copies. The SR-MRA problem serves as a mathematical abstraction of the structure determination problem for biological molecules. Since the SR-MRA problem is ill-posed without prior knowledge, accurate image estimation relies on designing priors that describe the statistics of the images of interest. In this work, we build on recent advances in image processing and harness the power of denoisers as priors for images. To estimate an image, we propose utilizing denoisers as projections and using them within two computational frameworks that we propose: projected expectation-maximization and projected method of moments. We provide an efficient GPU implementation and demonstrate the effectiveness of these algorithms through extensive numerical experiments on a wide range of parameters and images.

INDEX TERMS Method of momented, projected gradient descent, expectation minimization, MRA.

I. INTRODUCTION

2D super-resolution multi-reference alignment (SR-MRA) entails estimating an image $x \in \mathbb{R}^{L_{\text{high}} \times L_{\text{high}}}$ from its N circularly-translated, down-sampled, noisy copies:

$$y_i = PR_{s_i}x + \varepsilon_i, \quad i = 1, \dots, N, \tag{1}$$

where R_s denotes a 2D circular translation, P denotes a down-sampling operator that collects $L_{\text{low}} \times L_{\text{low}}$ equally-spaced samples of $R_s x$, and $\varepsilon_i \in \mathbb{R}^{L_{\text{low}} \times L_{\text{low}}}$ is a noise matrix whose entries are drawn i.i.d. from $\mathcal{N}(0, \sigma^2)$. The 2D translation s is composed of a horizontal translation s^1 and a vertical translation s^2, which are drawn i.i.d. from unknown distributions, ρ_1 and ρ_2, respectively. Explicitly, each observation $y_i \in \mathbb{R}^{L_{\text{low}} \times L_{\text{low}}}$ takes the form $y_i[n_1, n_2] = x[n_1 K - s_i^1, n_2 K - s_i^2] + \varepsilon_i[n_1, n_2]$, where $n_1, n_2 = 0, \dots, L_{\text{low}} - 1$, the shifts should be considered modulo L_{high}, and $K := \frac{L_{\text{high}}}{L_{\text{low}}}$ is assumed to be an integer. Our goal is to estimate x (the high resolution image) from N low-resolution observations $y_1, \dots, y_N$, when the translations $s_1, \dots, s_N$ are unknown.

The SR-MRA model, first studied for 1-D signals [1], is a special case of the multi-reference alignment (MRA) model: The problem of estimating a signal from its noisy copies, each acted upon by a random element of some group; see for example [2], [3], [4], [5], [6], [7], [8]. The MRA model is mainly motivated by the single-particle cryo-electron microscopy (cryo-EM) technology: an increasingly popular technique to construct 3-D molecular structures [9]. In Section VI, we introduce the mathematical model of cryo-EM in detail and discuss how the techniques proposed in this paper have the potential to make a significant impact on the molecular reconstruction problem of cryo-EM. The MRA model is also motivated by a variety of applications in biology [10], robotics [11], radar [12], and image processing [13].

The two leading computational techniques for solving MRA problems are the method of moments (MoM) and expectation-maximization (EM) [4], [5]. MoM is a classical parameter estimation technique, aiming to recover the parameters of interest from the observed moments. MoM requires a single pass over the observations, making it an efficient technique for large data sets (large N) but is not statistically efficient. The EM algorithm aims to maximize the likelihood function (or the posterior distribution in a Bayesian framework) [14]. As Theorem 2.1 shows, it is impossible to uniquely identify the image x only from the SR-MRA observations (1) as many different images result in the same likelihood function. Thus, to estimate the image accurately, we need to incorporate a prior. In this work, we harness a recent line of works that use denoisers as

priors and show how to incorporate them into the MoM and EM in the context of SR-MRA.

Utilizing existing state-of-the-art denoisers was proposed as part of the Plug-and-Play technique [15] and was proven highly effective for many imaging inverse problems [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33]. Essentially, the underlying idea is exploiting the impressive capabilities of existing denoising algorithms to replace explicit, traditional priors. Indeed, these methods are especially useful for natural images, whose statistics are too complicated to be described explicitly, but for which excellent denoisers have been devised. A natural question is why not just use the standard plug-and-play framework for SR-MRA instead of developing a dedicated strategy for EM and MoM as we do in this work. The simple answer is that, despite intensive efforts, we could not combine the plug-and-play approach with MoM and EM in a stable way. Therefore, we found that it is necessary to develop a dedicated strategy as we do in this work.

Our proposed computational framework, presented in Section III, uses denoisers as projection operators. Specifically, every few iterations of either MoM or EM, we apply a denoiser to the current estimate. This stage can be interpreted as projecting the image estimate onto the implicit space spanned by the denoiser. Section IV provides convergence analysis, and Section V demonstrates the effectiveness of our method by extensive numerical experiments on images.

The contribution of this paper may be summarized as follows: (i) We propose a novel framework for the SR-MRA problem that relies on using denoisers as priors throughout the optimization; (ii) we show that the method is generic by applying it to two popular methods for SR-MRA, namely MoM and EM; (iii) we provide an initial analysis for the convergence of our proposed algorithms; (iv) we open-source the code of our approach that efficiently implements our strategy on a GPU for getting efficient parallel processing.

II. BACKGROUND

Before introducing our framework, we elaborate on four pillars of this work: the non-uniqueness of the SR-MRA model (1), MoM, EM, and denoiser-based priors. Hereafter, we let $\rho := [\rho_1, \rho_2]^T$, and define the set of sought parameters as $\theta := (x, \rho)$. With a slight abuse of notation, we treat the image x and a measurement y_i as vectors in $\mathbb{R}^{L_{\text{high}}^2}$ and $\mathbb{R}^{L_{\text{low}}^2}$, respectively. We also define $\Theta = \mathbb{R}^{L_{\text{high}}^2} \times \Delta_{L_{\text{high}}} \times \Delta_{L_{\text{high}}}$, where $\Delta_{L_{\text{high}}}$ is the L_{high}-dimensional simplex, so that $\theta \in \Theta$.

SR-MRA observations do not identify the image x uniquely: The following theorem states that the likelihood function of (1) does not determine the image x uniquely. This, in turn, implies that a prior is necessary for accurate estimation of the sought image.

Theorem 2.1: The likelihood function $p(y_1, \ldots, y_N | x, \rho)$ does not determine uniquely the sought parameters x and ρ.

Proof: We follow the proof of [1, Theorem 3.1]. Recall that $y = PR_s x + \varepsilon$, and that $K := \frac{L_{\text{high}}}{L_{\text{low}}}$ is an integer. Let us define a set of K^2 sub-images, indexed by $n_1, n_2 = 0, \ldots, K-1$:

$$x_{n_1,n_2}[\ell_1, \ell_2] := x[n_1 + \ell_1 K, n_2 + \ell_2 K],$$

for $\ell_1, \ell_2 = 0, \ldots, L_{\text{low}} - 1$. Then, the SR-MRA model reads

$$y = R_t x_{n_1,n_2} + \varepsilon, \tag{2}$$

where R_t is a translation over the low-resolution grid of size $L_{\text{low}} \times L_{\text{low}}$. We denote the distribution of choosing the sub-image x_{n_1,n_2} and translating it by t as $\rho[n_1, n_2, t]$. Thus, the likelihood function of (1), for a single observation y, can be written, up to a constant, as

$$p(y; x, \rho) = \sum_{n_1,n_2=0}^{K-1} \sum_{t} \rho[n_1, n_2, t] e^{-\frac{1}{2\sigma^2}\|y - R_t x_{n_1,n_2}\|_2^2}.$$

Note that the likelihood function is invariant under any permutation of the sub-images (overall $K^2!$ permutations), and under any translation of each of them (overall $(L_{\text{low}}^2)^{K^2}$ possible translations). This implies that there are $K^2!(L_{\text{low}}^2)^{K^2}$ different images with the same likelihood function. □

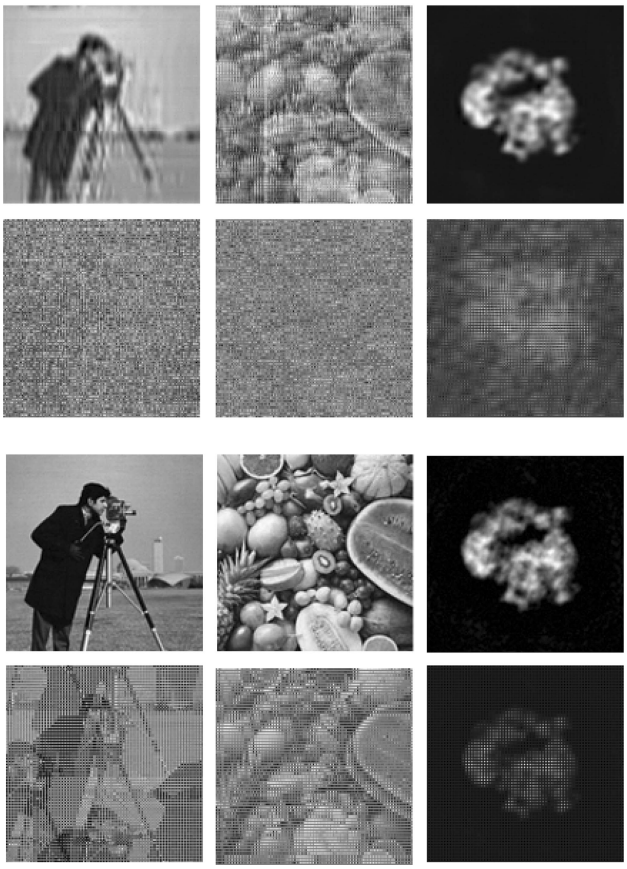

FIGURE 1. The first and third rows present recoveries using projected MoM and projected EM. Second and fourth rows present recoveries using MoM and EM, without projection. As Theorem 2.1 indicates, the projection is vital for accurate recovery. Images are of size $L_{\text{high}} = 128$ with a down-sampling factor of $K = 2$. MoM used the exact moments, and EM was applied with noise level $\sigma = 1/8$ and number of observations $N = 10^4$. In the projected versions, we set F = 5 (see Section III).

Fig. 1 compares image recovery using projected MoM and projected EM, which are our proposed algorithms (see Section III). Specifically, the first and third rows show recoveries using Algorithm 1 and Algorithm 2, respectively, while the second and fourth rows show the outputs of the standard MoM and EM, with no prior. Evidently, in light of Theorem 2.1, the latter results are of low quality.

The method of moments (MoM) aims to estimate the parameters of interest from the observable moments. Recall that the d-th observable moment is given by

$$\widehat{M}_d := \frac{1}{N} \sum_{i=1}^{N} y_i^{\otimes d}, \tag{3}$$

where $y^{\otimes d}$ is a tensor with $(L_{\text{low}}^2)^d$ entries, and the entry indexed by $\ell = (\ell_1, \ldots, \ell_d)$ is given by $\prod_{j=1}^{d} y[\ell_j]$. Computing the observable moments requires a single pass over the observations and results in a concise summary of the data. By the law of large numbers, for sufficiently large N, we have

$$\widehat{M_d} \approx \mathbb{E} y^{\otimes d} := M_d(\theta), \tag{4}$$

where the expectation is taken against the distribution of the translations and the noise. Finding the optimal parameters θ that fit the observable moments is usually performed by minimizing a least squares objective,

$$\min_{\theta \in \Theta} \sum_{d=1}^{D} \lambda_d \|\hat{M}_d - M_d(\theta)\|^2, \tag{5}$$

where $\lambda_1, \ldots, \lambda_D$ are some predefined weights.

Model (1) was studied in [5] when P is the identity operator (no down-sampling).[1] In particular, it was shown that if the discrete Fourier transform of the signal is non-vanishing, and ρ is almost any non-uniform distribution, then the signal and distribution are determined uniquely, up to an unavoidable translation symmetry, from the second moment of the observations. While this is not necessarily true when P is a down-sampling operator, we keep using the first two moments in this work (namely, $D = 2$).

Specifically, the first two moments of the SR-MRA model (1) are given by [34]:

$$M_1 = PC_x\rho, \tag{6}$$

$$M_2 = PC_xD_\rho C_x^T P^T + \sigma^2 PP^T, \tag{7}$$

where $C_x \in \mathbb{R}^{L_{\text{high}}^2 \times L_{\text{high}}^2}$ is a block circulant with circulant blocks (BCCB) matrix of the form:

$$C_x = \begin{pmatrix} c_0 & c_{L_{\text{high}}-1} & \ldots & c_2 & c_1 \\ c_1 & c_0 & c_{L_{\text{high}}-1} & \ldots & c_2 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ c_{L_{\text{high}}-1} & c_{L_{\text{high}}-2} & \ldots & c_1 & c_0 \end{pmatrix}, \tag{8}$$

where the block $c_i \in \mathbb{R}^{L_{\text{high}} \times L_{\text{high}}}$ represents a circulant matrix of the i-th column of the image x. Each column of C_x contains a copy of the image after 2D translation and vectorization. The matrix $D_\rho \in \mathbb{R}^{L_{\text{high}}^2 \times L_{\text{high}}^2}$ is a diagonal matrix, whose diagonal is a vectorization of the matrix $\rho_1 \rho_2^T$. Therefore, the least squares objective (5) can be explicitly written as

$$\min_{\theta \in \Theta} ||PC_xD_\rho C_x^T P^T + \sigma^2 PP^T - \hat{M}_2||_{\text{F}}^2 + \lambda ||PC_\rho x - \hat{M}_1||_2^2. \tag{9}$$

In this paper, we set $\lambda = \frac{1}{L_{\text{high}}^2(1+\sigma^2)}$, as suggested in [5].

Expectation-maximization (EM): The log-likelihood of (1) is given by

$$\log p(y_1, \ldots, y_n; \theta) = \sum_{i=1}^{N} \log \sum_{s^1, s^2 = 1}^{L_{\text{high}}} \rho[s] e^{-\frac{1}{2\sigma^2}\|y_i - PR_s x\|_2^2}. \tag{10}$$

A popular way to maximize the likelihood function is using the EM algorithm [14]. EM can also be used to maximize the posterior distribution when an analytical prior is available.

[1] The results of [5] concern 1-D, but they can be readily extended to 2-D.

EM is an iterative algorithm, where each iteration consists of two steps. The first, called the E-step, computes the expected value of the log-likelihood function with respect to the translations, given the current estimate of $\theta := (x, \rho)$. In the $(t+1)$-th iteration, it reads

$$Q(x, \rho | x_t, \rho_t) = \sum_{i=1}^{N} \sum_{s^1, s^2 = 1}^{L_{\text{high}}} w_t^{i,s} \left(\log \rho[s] - \frac{1}{2\sigma^2} ||y_i - PR_s x||_{\text{F}}^2 \right), \tag{11}$$

where

$$w_t^{i,s} = C_t^i \rho_t[s] e^{-\frac{1}{2\sigma^2}||y_i - PR_s x||_{\text{F}}^2}, \tag{12}$$

and C_t^i is a normalization factor so that $\sum_s w_t^{i,s} = 1$. The next step, called the M-step, maximizes (11) with respect to x and ρ:

$$(x_{t+1}, \rho_{t+1}) = \operatorname*{argmax}_{x, \rho} Q(x, \rho | x_t, \rho_t). \tag{13}$$

In our case, the maximum is attained by solving the linear system of equations:

$$Ax_{t+1} = b, \tag{14}$$

where

$$A = \sum_{i=1}^{N} \sum_{s} w_t^{i,s} R_s^{-1} P^{-1} PR_s,$$

$$b = \sum_{i=1}^{N} \sum_{s} w_t^{i,s} R_s^{-1} P^{-1} y_i,$$

and

$$\rho_{t+1}[s] = \frac{\sum_{i=1}^{N} w_k^{i,s}}{\sum_{i=1}^{N} \sum_{s'} w_k^{i,s'}}. \tag{15}$$

The EM algorithm iterates between the E and the M steps. While each EM iteration does not decrease the likelihood, for the SR-MRA model, the likelihood function is not convex, and thus the EM algorithm is not guaranteed to achieve its maximum. A common practice to improve estimation accuracy is either to initialize the EM iterations in the vicinity of the solution (if such an initial estimate is available) or to run it multiple times from different initializations, and choose the estimate corresponding to the maximal likelihood value. As we show in Section V, our projected EM algorithm (described in the next section) provides consistent results even with a single initialization.

Denoiser-based priors: Using denoisers as priors was proposed in the Plug-and-Play approach [15]. It aims to maximize a posterior distribution $p(\theta | y_1, \ldots, y_n)$, which is equivalent to the optimization problem:

$$\operatorname*{argmin}_{\theta \in \Theta} l(\theta) + \mathfrak{p}(\theta), \tag{16}$$

where $l(\theta) := -\log p(y_1, \ldots, y_n | \theta)$, $\mathfrak{p}(\theta) := -\log p(\theta)$, $p(\theta)$ is a prior on θ, and recall that $\theta = (x, \rho)$. This work only considers a prior on the image x, but a prior on the distribution might be learned as well [35].

Plug-and-Play proceeds by variable splitting,

$$\operatorname*{argmin}_{\theta, v \in \Theta} l(\theta) + \beta \mathfrak{p}(v) \quad \text{subject to} \quad \theta = v, \tag{17}$$

where β is a parameter that is manually tuned. The problem in (17) can be solved using ADMM by iteratively applying the following

three steps:

$$\theta^{k+1} = \underset{\theta\in\Theta}{\operatorname{argmin}}\, l(\theta) + \frac{\lambda}{2}||\theta - (v^k - u^k)||_2^2,$$
$$v^{k+1} = \underset{v\in\Theta}{\operatorname{argmin}}\, \frac{\lambda}{2\beta}||\theta^{k+1} + u^k - v||_2^2 + \mathfrak{p}(v),$$
$$u^{k+1} = u^k + (\theta^{k+1} - v^{k+1}), \tag{18}$$

where λ is the ADMM penalty parameter. The main insight here is that the second step aims to solve a Gaussian denoising problem, with a prior $\mathfrak{p}(v)$ and standard deviation of $\sqrt{\frac{\beta}{\lambda}}$. Instead of solving this problem for an analytical explicit prior, one can apply off-the-shelf denoisers, such as BM3D or neural nets.

While Plug-and-Play was used in many tasks, it requires tuning the parameters λ, β and the number of iterations for the first step; these parameters significantly impact its performance. In our setup, we did not find a set of parameters that yields accurate estimates for the SR-MRA problem in a wide range of scenarios. As an alternative, we propose a projection-based approach for EM and MoM.

Algorithm 1: Projected MoM.

Input: Measurements $y_1, \ldots, y_N$, denoiser $\mathcal{D}$, and the parameter F
Output: An estimate of the target image and distribution
1) Compute the empirical moments $\hat{M}_1, \hat{M}_2$ according to (3)
2) Until a stopping criterion is met:
a) Run F BFGS (or alternative gradient-based algorithm) steps to minimize (9)
b) Apply the denoiser $x \leftarrow \mathcal{D}(x, \sigma^{\mathrm{BM3D}})$
c) Update σ^{BM3D} according to (19)

Algorithm 2: Projected EM.

Input: Measurements $y_1, \ldots, y_N$, denoiser $\mathcal{D}$, and the parameter F
Output: An estimate of the target image and distribution
Until a stopping criterion is met:
1) Run F EM steps according to (12), (14), and (15).
2) Apply the denoiser $x \leftarrow \mathcal{D}(x, \sigma^{\mathrm{BM3D}})$
3) Update σ^{BM3D} according to (19)

III. DENOISER-BASED PROJECTED ALGORITHMS

This section introduces the main contribution of this work: incorporating a projection-based denoiser into the MoM and EM, and their implementation for SR-MRA. Importantly, there is no agreed way of including a prior into MoM, nor a standard way of incorporating natural images prior to EM. Projected MoM algorithm is presented in Algorithm 1, and projected EM in Algorithm 2.

The main idea is rather simple: every few iterations of either MoM or EM, we apply a predefined denoiser that acts on the current estimate of the image. Specifically, for MoM, we apply the denoiser every few steps of BFGS (with line-search) that minimizes the least squares objective (9); BFGS can be replaced by alternative gradient-based methods. For EM, we apply the denoiser every few EM steps.

As a denoiser, we chose to work with the celebrated BM3D denoiser [36] because of its simplicity and effectiveness for denoising natural images. However, replacing BM3D with alternative denoisers is straightforward. This might be especially relevant when dealing with specific applications, where a data-driven deep neural net may take the role of the denoiser. In Section VI, we discuss such potential applications for cryo-EM data processing.

Our algorithm requires only two parameters to be adjusted. The first parameter is the noise level of the denoiser, denoted hereafter by σ^{BM3D}. A low noise level increases the weight of the data (either likelihood or observed moments), while increasing the noise level strengthens the effect of the denoiser. In particular, we found that starting from a high noise level and gradually decreasing it with iterations leads to consistent and accurate estimations. Specifically, in the experiments of Section V, we set the noise level to

$$\sigma_t^{\mathrm{BM3D}} = 2^{-1/10}\sigma_{t-1}^{\mathrm{BM3D}}, \quad \sigma_1^{\mathrm{BM3D}} = 1, \tag{19}$$

where σ_t^{BM3D} is the noise level of the t-th application of BM3D.

The second parameter, which we denote by F, determines the number of EM iterations or BFGS iterations (or other gradient-based algorithms) for MoM between consecutive applications of the denoiser. Small or large F may put too much or too little weight on the prior (rather than on the data). Empirically, projected MoM is quite sensitive to this parameter and requires carefully tuning F, as presented in Section V, whereas EM is less sensitive. In contrast to Plug-and-Play, which was very sensitive to parameter tuning in our tests and did not converge well, we found it to be quite easy to tune the parameters F and σ and get consistent results for a wide range of images and parameters. The simplicity of parameter tuning, perhaps, stems from the fact that each parameter affects only one term (F the fidelity, and σ^{BM3D} the prior), whereas tuning the λ parameter in the Plug-and-Play approach (18) affects both terms.

IV. CONVERGENCE ANALYSIS

We turn to analyze the convergence of the proposed algorithms. We start with a couple of definitions and assumptions that will be used throughout the analysis. Then, we provide convergence analysis for projected MoM in a somewhat modified scheme, where the (nonconvex) MoM objective is minimized by a gradient descent step rather than BFGS iterations (which are essentially preconditioned gradient steps). Lastly, to tackle frameworks that are not based on any gradient method, we present a more general operator-based analysis. While some assumptions in this analysis are not guaranteed for EM, it provides mathematical reasoning for the convergence of frameworks that are composed of two "black-box" operators that generate convergent sequences, each on its own, which is relevant to EM.

Definition 4.1: We say that the map $\mathcal{M} : \mathcal{X} \to \mathcal{X}$ is C-Lipschitz in $\mathcal{X}$ if for all $x, z \in \mathcal{X}$ we have

$$\|\mathcal{M}(x) - \mathcal{M}(z)\|_2 \le C\|x - z\|_2.$$

In particular, if $C = 1$ then $\mathcal{M}$ is nonexpansive in $\mathcal{X}$ and if $C < 1$ then $\mathcal{M}$ is a contraction in $\mathcal{X}$.

Definition 4.2: The proximal operator associated with a function $\mathfrak{p}(\cdot)$ is defined by

$$\mathrm{Prox}_{\mathfrak{p}}(x) := \underset{z}{\operatorname{argmin}}\, \frac{1}{2}\|x - z\|_2^2 + \mathfrak{p}(z).$$

Observe the tight connection between the proximal operator and Gaussian denoising. Specifically, Gaussian denoiser $\mathcal{D}(\cdot;\sigma)$ that is associated (perhaps implicitly) with a prior function $\mathfrak{p}(\cdot)$ is typically obtained by an optimization problem similar to $\mathrm{Prox}_{\sigma^2\mathfrak{p}}(x)$.

The proximal operator has been originally explored under the assumption that $\mathfrak{p}$ is a lower semi-continuous (l.s.c.) convex function [37], for which it has been shown [38, Prop. 1] that: 1) it

is a subdifferential of a convex function and 2) it is 1-Lipschitz, i.e., nonexpansive. These properties are important for guaranteeing the convergence of algorithms that use it. A recent work [38] has extended the analysis to nonconvex $\mathfrak{p}$ (which may be more similar to practical denoisers), but these results do not imply nonexpansiveness without additional assumptions on the proximal operator.

Throughout this section, we make the following assumptions.

Assumption A1: ρ (the distribution parameters) is fixed.

Assumption A2: $\mathcal{D}$ is a proximal operator associated with some proper l.s.c. convex function.

Note that the data-terms that we consider (namely, the log-likelihood and the MoM objectives) are challenging nonconvex functions. They remain nonconvex even if ρ, which parameterizes the signal's distribution, is known or fixed. Similarly, the optimization problems remain nonconvex even when making the above simplifying assumption on the denoiser. Nevertheless, note that our approach can be readily applied with denoisers that are based on TV regularization or on ℓ_1-norm prior, which satisfy A2.

The two iterative estimation schemes in this paper can be expressed as $x_{k+1} = \mathcal{D}(\mathcal{T}(x_k))$, where the denoiser $\mathcal{D}$ follows an operation $\mathcal{T}$ that reduces some nonconvex data-fidelity term: the first scheme reduces the MoM objective with a gradient-based method and the second scheme reduces the negative log-likelihood via the EM method. We now turn to analyze two such schemes.

A. CONVERGENCE OF PROJECTED MOM

Under A1 and A2, let us provide convergence results for the iterative scheme

$$x_{k+1} = \mathcal{D}(x_k - \mu \nabla f(x_k)), \tag{20}$$

where f denotes the MoM objective and μ is the step-size. Notice that f is smooth but nonconvex. Compared to Algorithm 1, the main difference is merely replacing the BFGS steps with gradient steps.

To establish a convergence statement for (20), we will exploit convergence results of the proximal gradient method [39], [40]. To invoke them, we show that the MoM has a Lipschitz gradient. The following theorem shows that this is the case when the optimization is restricted to a closed domain $\mathcal{X}$. This assumption is reasonable when x is an image, as typically its pixels are restricted to the range $[0, 255]$.

Theorem 4.3: Assume that $\mathcal{X}$ is a closed convex set. Let ρ satisfy A1 and $\mathcal{D}$ satisfy A2 with respect to a function $\mu\mathfrak{p}(\cdot)$, with $\mu \in \mathbb{R}$ and domain $\mathcal{X}$. Consider the nonconvex MoM objective

$$f(x) = \|PC_x D_\rho C_x^T P^T + \sigma^2 PP^T - \hat{M}_2\|_{\mathrm{F}}^2 + \lambda \|PC_\rho x - \hat{M}_1\|_2^2. \tag{21}$$

Then, there exists $\mu > 0$ such that the sequence $\{x_k\}$ in (20) obeys: (1) the sequence $\|x_k - \mathcal{D}(x_k - \mu\nabla f(x_k))\|_2$ converges to zero, and (2) any limit of $\{x_k\}$ is a stationary point of $f + \mathfrak{p}$.

Proof: We need to show that, restricted to $\mathcal{X}$, ∇f is C-Lipschitz. Then by setting $\mu \in (0, \frac{2}{C})$ statements (1) and (2) in the theorem follow from [40, Theorem 10.15].

Note that $f(x)$ is infinitely differentiable. The Lagrange remainder (a consequence of the mean value theorem) of first-order Taylor's series of ∇f implies that for any $x, z \in \mathcal{X}$ there exists ξ on the line that connects x and z for which $\nabla f(x) = \nabla f(z) + \nabla^2 f(\xi)(x - z)$. Therefore,

$$\|\nabla f(x) - \nabla f(z)\|_2 \le \max_{\xi \in \mathcal{X}} \|\nabla^2 f(\xi)\|_{op} \|x - z\|_2,$$

where $\|\cdot\|_{op}$ is the operator norm. As f is a polynomial of order 4, the entries of $\nabla^2 f$ are polynomials of order 2 and are bounded on the closed set $\mathcal{X}$. Thus, the operator norm (magnitude of the largest eigenvalue of $\nabla^2 f$) is bounded on $\mathcal{X}$ due to the Gershgorin circle theorem. Denoting this finite value by C, we get that $\nabla f(x)$ is C-Lipschitz on $\mathcal{X}$. □

Note that our statement ensures that the limit is only a stationary point of the objective $f + \mathfrak{p}$, and not to a global minimum. This is inevitable, as f is a nonconvex function. Extensions to backtracking line search, instead of fixed step size, can be adopted from [39], [40].

B. COMPOSITION OF CONVERGENCE INDUCING OPERATORS

Unlike projected MoM, in projected EM the data-fidelity step is not a gradient step. It is well known that, on its own, the EM algorithm $x_{k+1} = \mathcal{T}(x_k)$ converges to a local minimum of the negative log-likelihood function. However, to the best of our knowledge, EM convergence behavior is not well understood, besides a few simple models. This motivates us to consider a framework of a composition of two "black-box" operators that generate convergent sequences when being applied alone. Indeed, it is also known that the sequence $x_{k+1} = \mathcal{D}(x_k)$, under A2, converges to a minimizer of its associated function $\mathfrak{p}$ (as a special case of the proximal gradient method on $0 + \mathfrak{p}$ [39], [40]). Furthermore, a gradient-step operator with a sufficiently small step size on the MoM objective also yields a convergent sequence. Therefore, the analysis we provide for EM may also apply to the MoM.

Denote by $\mathrm{Fix}(\mathcal{M})$ the set of fixed points of an operator $\mathcal{M}$, i.e., $\mathrm{Fix}(\mathcal{M}) = \{x : \mathcal{M}(x) = x\}$. As discussed above, both the data-fidelity operator $\mathcal{T}$ and the denoising operator $\mathcal{D}$ have fixed points associated with stationary points of a data-fidelity term f and an implicit prior term $\mathfrak{p}$, respectively.

Let us now define a class of operators that strictly attract points towards their set of fixed points, as done in [41].

Definition 4.4: We say that the map $\mathcal{M} : \mathcal{X} \to \mathcal{X}$ is γ-strongly quasi-nonexpansive in $\mathcal{X}$ with $\gamma > 0$ if for all $x \in \mathcal{X}$ and $z \in \mathrm{Fix}(\mathcal{M})$ we have

$$\|\mathcal{M}(x) - z\|_2^2 \le \|x - z\|_2^2 - \gamma \|\mathcal{M}(x) - x\|_2^2.$$

Note that this class includes the set of proximal operators (see [41, Fact 1c]). Thus, it is a weaker assumption. Even though it is not guaranteed that the data-fidelity operator $\mathcal{T}$ that is used in EM falls into this class for the entire domain, it might be reasonable to restrict the domain, where it holds to be a small set $\mathcal{X}$ around a fixed point. The following theorem shows how this property implies the convergence of the sequence $x_{k+1} = \mathcal{D}(\mathcal{T}(x_k))$ to a fixed point.

Theorem 4.5: Let $\mathcal{D}$ satisfy A2 with domain $\mathcal{X}$ and let $\mathcal{T}$ be γ-strongly quasi-nonexpansive in $\mathcal{X}$ such that $\mathrm{Fix}(\mathcal{D}) \cap \mathrm{Fix}(\mathcal{T}) \neq \emptyset$. Then, the sequence $x_{k+1} = \mathcal{D}(\mathcal{T}(x_k))$ converges to a fixed point $x_* = \mathcal{D}(\mathcal{T}(x_*))$.

Proof: As a proximal operator associated with a convex function, $\mathcal{D}$ is 1/2-average of a nonexpansive operator [40, Theorem 6.42], [42, Lemma 2.1]. Thus, from [41, Fact 1c], it is strongly quasi-nonexpansive. From [41, Prop. 1d] the composition of strongly quasi-nonexpansive operators $\mathcal{M} = \mathcal{D} \circ \mathcal{T}$ with $\mathrm{Fix}(\mathcal{D}) \cap \mathrm{Fix}(\mathcal{T})$ is strongly quasi-nonexpansive in $\mathcal{X}$ with some $\tilde{\gamma} > 0$ and $\mathrm{Fix}(\mathcal{M}) = \mathrm{Fix}(\mathcal{D}) \cap \mathrm{Fix}(\mathcal{T})$. Let $\tilde{x}_* \in \mathrm{Fix}(\mathcal{M})$, then

$$\|x_{k+1} - \tilde{x}_*\|_2^2 \le \|x_k - \tilde{x}_*\|_2^2 - \gamma \|x_{k+1} - x_k\|_2^2$$
$$\Rightarrow \|x_{k+1} - x_k\|_2^2 \le \tfrac{1}{\gamma}(\|x_k - \tilde{x}_*\|_2^2 - \|x_{k+1} - \tilde{x}_*\|_2^2).$$

Summing over k from 0 to K we have

$$\sum_{k=0}^{K} \|x_{k+1} - x_k\|_2^2 \leq \frac{1}{\gamma}(\|x_0 - \tilde{x}_*\|_2^2 - \|x_{K+1} - \tilde{x}_*\|_2^2)$$

$$\leq \frac{1}{\gamma}\|x_0 - \tilde{x}_*\|_2^2.$$

By taking K to infinity, we see that $\{x_k\}$ is a Cauchy sequence, and thus it converges to some limit point x_*, i.e., $x_k \to x_*$. Consequently, this limit point is a fixed-point: $x_* = \mathcal{D}(\mathcal{T}(x_*))$. □

The above result, despite including assumptions that are not guaranteed for EM on the entire domain, provides mathematical reasoning for the convergence of composition of a denoiser and a data-fitting operator, when each obeys convergence properties alone and both can agree on a fixed point (potentially, out of many fixed points that fit the measurements but can badly estimate the true signal). Note that the theorem assumption $\text{Fix}(\mathcal{D}) \cap \text{Fix}(\mathcal{T}) \neq \emptyset$ is reasonable, as we empirically show that projected EM converges to a noiseless point that agrees with the measurements.

V. NUMERICAL EXPERIMENTS

We applied the proposed algorithms to the 68 "natural" images of the CBSD-68 dataset and a cryo-EM image of the E. coli 70S ribosome, available at the Electron Microscopy Data Bank.[2] We set $L_{\text{high}} = 128$, and the images were down-sampled to $L_{\text{low}} = 64, 32, 16, 8$ (namely, down-sampling factors of $K = 2, 4, 6, 8$, respectively).

For all experiments, we used the BM3D denoiser [36], where the noise level decays as in (19). Unless specified otherwise, we set $F = 5$ for both algorithms. For the MoM, we minimized the least squares objective (5) using the BFGS algorithm with line-search [43], [44]. The distributions ρ_1, ρ_2 and their initial guesses were drawn from a uniform distribution on [0, 1], and normalized so that $\rho_1, \rho_2 \in \triangle^{L_{\text{high}}}$. The pixel values of all images are in [0, 1]. Each pixel in the initialization of the image estimate, for both algorithms, was drawn i.i.d. from a uniform distribution on [0, 1], which was then normalized to include the entire [0, 1] range. The algorithms were initialized with a single initial guess; considering more initializations did not make a significant effect on the results.

Due to the shift-invariance of (1), we measure the error by:

$$\text{error}(\hat{x}) = \min_s \frac{||R_s\hat{x} - x||_{\text{F}}}{||x||_{\text{F}}}, \tag{22}$$

where $\hat{x}$ is the estimated image, and R_s denotes a 2D translation. We define SNR as

$$\text{SNR} = \frac{||x||_{\text{F}}^2}{L_{\text{high}}^2\sigma^2}. \tag{23}$$

Our algorithms were implemented in Torch. The recovery of a high-dimensional signal requires an efficient implementation of the algorithms over GPUs. Actually, without this implementation, we were unable to recover images of size $L_{\text{high}} \geq 32$ within a reasonable running time, where reducing the resolution beyond that adversely affects denoising performance. The code to reproduce all experiments is available at https://github.com/JonathanShani/Denoiser_projection.

Visual examples: Fig. 2 presents a collection of visual recoveries from the CBSD-68 dataset using both algorithms, with a down-sampling factor of $K = 2$. For projected MoM, we assumed to have access to the population (exact) moments (which is the case when $N \gg \sigma^4$), and projected EM used $N = 10^4$ observations with a noise level of $\sigma = 1/8$ (corresponding to SNR≈15). Evidently, projected EM provides more accurate recoveries, although MoM uses exact moments. However, as we show next, the computational load of projected EM is much heavier. Fig. 3 presents recoveries of additional images using projected EM with sampling factors of $K = 2, 4, 8$, $N = 10^4$ observations and noise level of $\sigma = 1/8$. Remarkably, projected EM provides accurate estimates with $K = 8$, i.e., when the number of pixels is reduced from 128^2 to 16^2.

Quantitative comparisons: We compared projected EM and projected MoM in terms of estimation error and run-time. While the reported running times depend on implementation and hardware, the goal is to show the general trend.

Table 1 presents the error and running time, averaged over 20 trials, for each combination of the parameters L_{high}, L_{low}, N and σ. We used the Lena image of size $L_{\text{high}} = 128$ with down-sampling

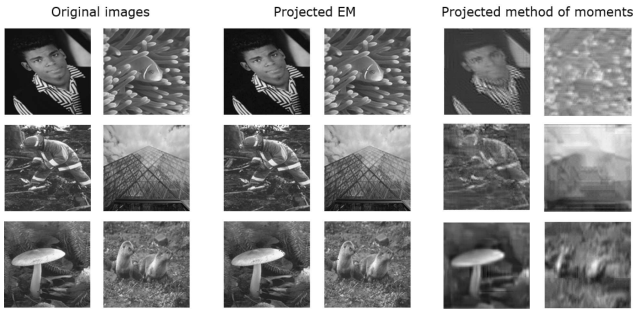

FIGURE 2. A gallery of 6 examples from the CBSD-68 dataset, and the corresponding recoveries using projected EM in a high-SNR regime ($\sigma = 1/8$) and projected MoM with accurate moments ($N \gg \sigma^4$); the images were down-sampled by a factor of $K = 2$. It can be seen that projected EM outperforms projected MoM. However, as we show in Table 1, the computational load of projected EM is much heavier.

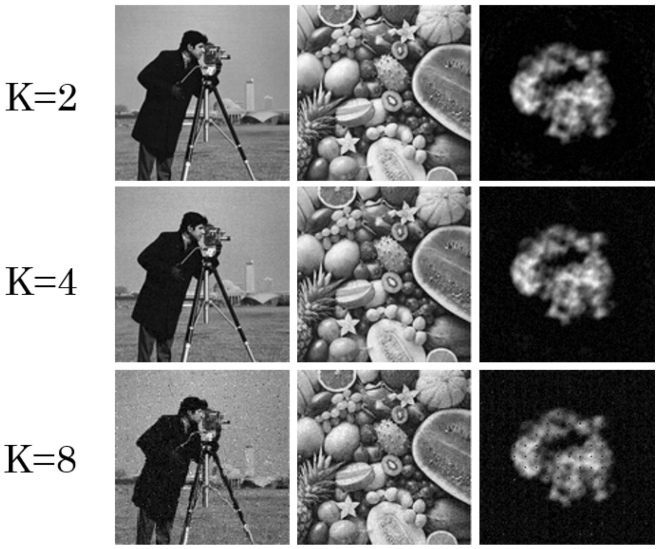

FIGURE 3. Recoveries using the projected EM algorithm with noise level of $\sigma = 1/8$ and $N = 10^4$ observations. The down-sampling factors are $K = 2$ (top row), $K = 4$ (middle row), and $K = 8$ (bottom row).

[2][Online]. Available: https://www.ebi.ac.uk/emdb/

TABLE 1. Comparing Projected MoM and EM for Various Target Resolutions L_{high} of the Lena Image, Observed Resolutions L_{low}, Observations Number N and σ

Model parameters				Projected MoM		Projected EM	
L_{high}	L_{low}	N	σ	Error	Time	Error	Time
128	32	100	0.125	0.332	**266.5421**	**0.0991**	871.6759
128	32	100	0.25	0.3459	**207.2537**	**0.196**	861.4668
128	32	100	0.5	**0.356**	**120.2959**	0.3836	837.9343
128	32	10000	0.125	0.339	**261.8394**	**0.0311**	1593.9534
128	32	10000	0.25	0.3302	**280.6154**	**0.0623**	1668.756
128	32	10000	0.5	0.3301	**277.3122**	**0.1387**	1660.7012
128	16	100	0.125	0.3627	**85.716**	**0.2475**	919.5462
128	16	100	0.25	**0.411**	**84.4553**	0.4154	984.6871
128	16	100	0.5	**0.3714**	**77.8746**	0.7497	893.2808
128	16	10000	0.125	0.3874	**122.5611**	**0.0685**	3769.9855
128	16	10000	0.25	0.381	**113.7814**	**0.141**	3179.0298
128	16	10000	0.5	0.3652	**94.8278**	**0.3001**	3148.3869
64	32	100	0.125	0.3319	103.8799	**0.0494**	**100.9239**
64	32	100	0.25	0.3427	**47.959**	**0.0988**	99.215
64	32	100	0.5	0.3439	**44.0597**	**0.2111**	87.1056
64	32	10000	0.125	0.2562	**101.4311**	**0.0158**	325.7816
64	32	10000	0.25	0.3382	**100.1693**	**0.0315**	264.8022
64	32	10000	0.5	0.3349	**98.7658**	**0.0647**	276.1571

factors of $K = 4, 8$, and of size $L_{\text{high}} = 64$ with $K = 2$; the number of observations was set to $N = 10^2, 10^4$, and the noise level to $\sigma = 0.125, 0.25, 0.5$ (corresponding to SNR$\approx$17, 4, 1, respectively). Both algorithms were limited to 100 iterations per trial. For each trial, we drew a fresh set of observations based on a new distribution and initial guesses. Projected EM shows a clear advantage for almost all values of N and σ in terms of estimation error (besides some cases when σ is large and N is small). However, the computational burden of EM is much heavier; there are cases (e.g., $L_{\text{high}} = 128$, $N = 10^4$, $\sigma = 0.25$) where the running time of projected EM algorithm is 30 times longer than the running time of projected MoM. The reason is that EM iterates over all observations, whereas MoM requires only a single pass over the observations [45], while the dimension of the variables in the least squares objective (9) is proportional to the dimension of the image.

Error as a function of the iterations: We next examined the behavior of the algorithms as a function of their iterations, averaged over all images in the CBSD-68 dataset of size $L_{\text{high}} = 128$. Projected MoM used exact moments, and projected EM used $N = 10^4$ observations and a noise level of $\sigma = 1/8$. Fig. 4 presents results for a variety of sampling factors. As can be seen, the algorithms with projections significantly outperform the unprojected versions for all down-sampling factors. Notably, the projected algorithms provide non-trivial estimates even for $K = 16$ (when each measurement is down-sampled from 128×128 to 8×8). We note that the denoiser leads to a locally non-monotonic error.

Projected MoM performance as a function of SNR: Fig. 5 (left panel) shows the average recovery error of projected MoM as a function of the SNR. These experiments require averaging over multiple trials, which makes it virtually impossible to produce results for EM in a reasonable running time. All experiments were conducted with the Lena image of size $L_{\text{high}} = 32$, with a sampling factor of $K = 2$, and $N = 10^5$ observations. For each SNR value, we

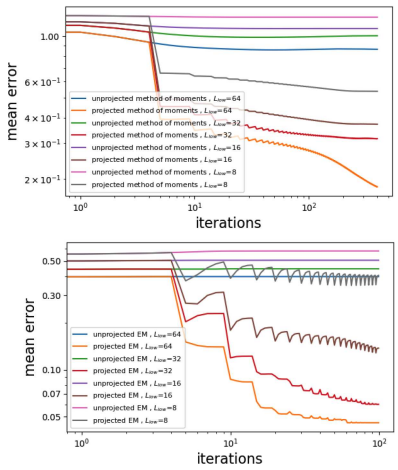

FIGURE 4. Error as a function of the iteration (averaged over all images in the CBSD-68 dataset, each of size $L_{\text{high}} = 128$), for different sampling factors. Top: projected MoM and the unprojected MoM, based on the perfect moments (corresponding to $N \to \infty$). Bottom: projected and unprojected EM, with $N = 10^4$ observations and noise level of $\sigma = 1/8$.

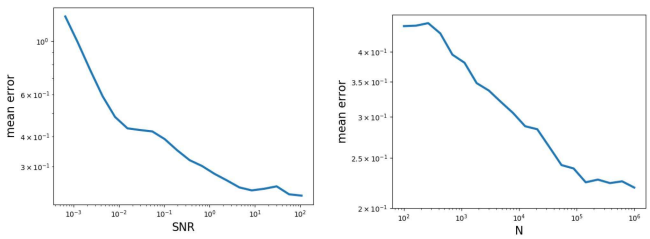

FIGURE 5. Mean error of projected MoM as a function of SNR with $N = 10^5$ observations (left) and number of observations with $\sigma = 1/8$ (right). The image size is $L_{\text{high}} = 32$ and the down-sampling factor is $K = 2$.

conducted 100 trials, each with a fresh distribution of translations. As can be seen, the error decreases monotonically with the SNR. Remarkably, the slope of the curve increases as the SNR decreases, hinting that the sample complexity of the problem (the number of observations required to achieve a desired accuracy) increases in the low SNR regime; this is a known phenomenon in the MRA literature for models with no projection [4], [5], [6].

Projected MoM performance as a function of the number of measurements: By the law of large numbers and according to (4), for any fixed SNR, the empirical moments almost surely converge to the analytic moments. To assess the impact of the number of observations on the estimation error, we used the image of Lena of size $L_{\text{high}} = 32$, down-sampling factor of $K = 2$, and noise level of $1/8$, corresponding to SNR $\approx$ 17.

Fig. 5 (right panel) presents the average error, over 100 trials, as a function of N. For each trial, we drew a fresh distribution. The error indeed decreases with N, as expected, but the slope is smaller than $-1/2$ (in logarithmic scale), as the law of large numbers predicts. This is due to the effect of the prior (manifested as a denoiser in our case), which does not depend on the observations (the data).

How frequently should we apply the denoiser? An important parameter of the proposed algorithms, denoted by F, determines after how many gradient steps (or EM iterations for projected EM) we apply the denoiser. For example, $F = 1$ means that we apply the

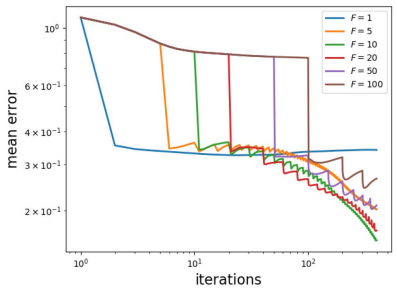

FIGURE 6. Mean error of projected MoM as a function of the iteration, for different values of F (number of gradient steps between consecutive denoising operations). The experiments were conducted on an image of size $L_{\text{high}} = 128$, assuming to have access to the population moments, and with a down-sampling factor of $K = 2$. The optimal value seems to be around $F = 10$; larger or smaller values of F lead to sub-optimal results.

denoiser after each gradient step, whereas $F = 100$ means that we run 100 gradient steps before applying the denoiser. This is especially important since the computational complexity of a single gradient step is significantly heavier than a BM3D application. To study the effect of this parameter for projected MoM, we used the image of Lena of size $L_{\text{high}} = 128$, a down-sampling factor of $K = 2$, and assumed to have access to the population (perfect) moments.

Fig. 6 shows the average error (over 10 trials) of MoM for different values of F; each trial was conducted with a fresh distribution. Plainly, the impact of F on the performance is significant. The optimal value seems to be around $F = 10$ in this case; larger and smaller values of F lead to sub-optimal results.

VI. CONCLUSION

Single-particle cryo-EM is an increasingly popular technology for high-resolution structure determination of biological molecules [9], [46]. The cryo-EM reconstruction problem is to estimate a 3-D structure (the electrostatic potential map of the molecule of interest) from multiple observations of the form $y = TR_\omega X + \varepsilon$, where X is the 3-D structure to be recovered, R_ω represents an unknown 3-D rotation by some angle $\omega \in SO(3)$, and T is a fixed, linear tomographic projection. The noise level is typically very high. The SR-MRA problem (1) can be interpreted as a toy model of the cryo-EM problem, where the image plays the role of the 3-D structure, and the 2-D translations and the sampling operator P replace, respectively, the unknown 3-D rotations and the tomographic projection.

The standard reconstruction algorithms in cryo-EM are based on EM [47], [48], aiming to maximize the posterior distribution based on analytical, explicit priors. Recently, techniques based on MoM were also designed for quick ab initio modeling [49], [50]. The techniques proposed in this paper can be readily integrated into these schemes, which may lead to improved accuracy and acceleration. A similar idea was recently suggested by [51], who used the regularization by denoising technique [16], and showed recoveries of moderate resolution with simulated data.

Current cryo-EM technology cannot be used to reconstruct small molecular structures (below ~ 40 kDa): this is one of its main drawbacks [52]. Recent papers suggested overcoming this barrier using variations of the MoM and EM [53], [54], [55]. However, as the problem is ill-conditioned, the resolution of the recovered structures is not high enough. We hope that modifications of the algorithms suggested in this paper, where the denoiser is replaced with a data-driven projection operator can be used to increase the resolution of the reconstructions and thus will allow elucidating multiple new structures of small biological molecules.

In our proposed scheme, the data fidelity (likelihood) term is computationally expensive due to the large number of measurements ($N > 100$) involved. A possible approach to effectively reduce the runtime is using online strategies, such as the one used in the online Plug-and-Play [56]. Note also that in this work we use BM3D as prior in our proposed framework. Yet, one may replace BM3D with a learned denoiser to further boost the performance, e.g., using a deep neural network. We defer this extension to future work.

REFERENCES

[1] T. Bendory, A. Jaffe, W. Leeb, N. Sharon, and A. Singer, "Super-resolution multi-reference alignment," *Inf. Inference: A. J. IMA*, vol. 11, no. 2, pp. 533–555, 2022.
[2] A. S. Bandeira, Y. Chen, R. R. Lederman, and A. Singer, "Non-unique games over compact groups and orientation estimation in cryo-EM," *Inverse Problems*, vol. 36, no. 6, 2020, Art. no. 064002.
[3] A. S. Bandeira, B. Blum-Smith, J. Kileel, J. Niles-Weed, A. Perry, and A. S. Wein, "Estimation under group actions: Recovering orbits from invariants," *Appl. Comput. Harmon. Anal.*, vol. 66, pp. 236–319, 2023.
[4] T. Bendory, N. Boumal, C. Ma, Z. Zhao, and A. Singer, "Bispectrum inversion with application to multireference alignment," *IEEE Trans. Signal Process.*, vol. 66, no. 4, pp. 1037–1050, Feb. 2018.
[5] E. Abbe, T. Bendory, W. Leeb, J. M. Pereira, N. Sharon, and A. Singer, "Multireference alignment is easier with an aperiodic translation distribution," *IEEE Trans. Inf. Theory*, vol. 65, no. 6, pp. 3565–3584, Jun. 2019.
[6] A. Perry, J. Weed, A. S. Bandeira, P. Rigollet, and A. Singer, "The sample complexity of multireference alignment," *SIAM J. Math. Data Sci.*, vol. 1, no. 3, pp. 497–517, 2019.
[7] T. Bendory, D. Edidin, W. Leeb, and N. Sharon, "Dihedral multi-reference alignment," *IEEE Trans. Inf. Theory*, vol. 68, no. 5, pp. 3489–3499, May 2022.
[8] C. Zhao, J. Liu, and X. Gong, "An adaptive variational model for multireference alignment with mixed noise," in *Proc. IEEE Int. Conf. Bioinf. Biomed.*, 2022, pp. 692–699.
[9] T. Bendory, A. Bartesaghi, and A. Singer, "Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities," *IEEE Signal Process. Mag.*, vol. 37, no. 2, pp. 58–76, Mar. 2020.
[10] W. Park, C. R. Midgett, D. R. Madden, and G. S. Chirikjian, "A stochastic kinematic model of class averaging in single-particle electron microscopy," *Int. J. Robot. Res.*, vol. 30, no. 6, pp. 730–754, 2011.
[11] D. M. Rosen, L. Carlone, A. S. Bandeira, and J. J. Leonard, "A certifiably correct algorithm for synchronization over the special euclidean group," in *Algorithmic Foundations of Robotics XII*. Berlin, Germany: Springer, 2020, pp. 64–79.
[12] J. P. Zwart, R. van der Heiden, S. Gelsema, and F. Groen, "Fast translation invariant classification of HRR range profiles in a zero phase representation," *IEE Proc.-Radar, Sonar Navigation*, vol. 150, no. 6, pp. 411–418, 2003.
[13] T. Bendory, I. Hadi, and N. Sharon, "Compactification of the rigid motions group in image processing," *SIAM J. Imag. Sci.*, vol. 15, no. 3, pp. 1041–1078, 2022.
[14] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc.: Ser. B. (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.
[15] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, 2013, pp. 945–948.
[16] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (RED)," *SIAM J. Imag. Sci.*, vol. 10, no. 4, pp. 1804–1844, 2017.
[17] S. H. Chan, X. Wang, and O. A. Elgendy, "Plug-and-play ADMM for image restoration: Fixed-point convergence and applications," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 84–98, Mar. 2017.
[18] F. Heide et al., "Flexisp: A flexible camera image processing framework," *ACM Trans. Graph.*, vol. 33, no. 6, pp. 1–13, Nov. 2014.

[19] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3929–3938.

[20] T. Tirer and R. Giryes, "Image restoration by iterative denoising and backward projections," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1220–1234, Mar. 2019.

[21] T. Tirer and R. Giryes, "Super-resolution via image-adapted denoising CNNs: Incorporating external and internal learning," *IEEE Signal Process. Lett.*, vol. 26, no. 7, pp. 1080–1084, Jul. 2019.

[22] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6360–6376, Oct. 2022.

[23] E. Ryu, J. Liu, S. Wang, X. Chen, Z. Wang, and W. Yin, "Plug-and-play methods provably converge with properly trained denoisers," in *Proc. Int. Conf. Mach. Learn.*, 2019, vol. 97, pp. 5546–5557.

[24] A. Rond, R. Giryes, and M. Elad, "Poisson inverse problems by the plug-and-play scheme," *J. Vis. Commun. Image Representation*, vol. 41, pp. 96–108, 2016.

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

[26] K. Wei, A. Aviles-Rivero, J. Liang, Y. Fu, C.-B. Schönlieb, and H. Huang, "Tuning-free plug-and-play proximal algorithm for inverse imaging problems," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 10158–10169.

[27] T. Meinhardt, M. Moller, C. Hazirbas, and D. Cremers, "Learning proximal operators: Using denoising networks for regularizing inverse imaging problems," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 1781–1790.

[28] G. T. Buzzard, S. H. Chan, S. Sreehari, and C. A. Bouman, "Plug-and-play unplugged: Optimization-free reconstruction using consensus equilibrium," *SIAM J. Imag. Sci.*, vol. 11, no. 3, pp. 2001–2020, 2018.

[29] Y. Sun, B. Wohlberg, and U. S. Kamilov, "An online plug-and-play algorithm for regularized image reconstruction," *IEEE Trans. Comput. Imag.*, vol. 5, no. 3, pp. 395–408, Sep. 2019.

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

[31] E. K. Ryu, J. Liu, S. Wang, X. Chen, Z. Wang, and W. Yin, "Plug-and-play methods provably converge with properly trained denoisers," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 5546–5557.

[32] Y. Sun, J. Liu, and U. S. Kamilov, "Block coordinate regularization by denoising," in *Proc. Ann. Conf. Neural Inf. Process. Syst.*, 2019.

[33] Z. Lai, K. Wei, Y. Fu, P. Härtel, and F. Heide, "∇-prox: Differentiable proximal algorithm modeling for large-scale optimization," *ACM Trans. Graph.*, vol. 42, no. 4, pp. 1–19, Jul. 2023.

[34] A. Abas, T. Bendory, and N. Sharon, "The generalized method of moments for multi-reference alignment," *IEEE Trans. Signal Process.*, vol. 70, pp. 1377–1388, 2022.

[35] N. Janco and T. Bendory, "An accelerated expectation-maximization algorithm for multi-reference alignment," *IEEE Trans. Signal Process.*, vol. 70, pp. 3237–3248, 2022.

[36] 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.

[37] J.-J. Moreau, "Proximité et dualité dans un espace hilbertien," *Bull. Soc. Math. France*, vol. 93, no. 2, pp. 273–299, 1965.

[38] R. Gribonval and M. Nikolova, "A characterization of proximity operators," *J. Math. Imag. Vis.*, vol. 62, no. 6, pp. 773–789, 2020.

[39] A. Beck and M. Teboulle, "Gradient-based algorithms with applications to signal recovery," in *Convex Optimization in Signal Processing and Communications*. Cambridge, U.K.: Cambridge University Press, pp. 42–88, 2009.

[40] A. Beck, *First-Order Methods in Optimization*. Philadelphia, PA, USA: SIAM, 2017.

[41] I. Yamada and N. Ogura, "Hybrid steepest descent method for variational inequality problem over the fixed point set of certain quasi-nonexpansive mappings," 2005.

[42] P. L. Combettes*, "Solving monotone inclusions via compositions of nonexpansive averaged operators," *Optimization*, vol. 53, no. 5/6, pp. 475–504, 2004.

[43] T. Bendory, T.-Y. Lan, N. F. Marshall, I. Rukshin, and A. Singer, "Multi-target detection with rotations," *Inverse Problems Imag.*, 2022, *arXiv:2101.07709*.

[44] S. Kreymer and T. Bendory, "Two-dimensional multi-target detection: An autocorrelation analysis approach," *IEEE Trans. Signal Process.*, vol. 70, pp. 835–849, 2022.

[45] N. Boumal, T. Bendory, R. R. Lederman, and A. Singer, "Heterogeneous multireference alignment: A single pass approach," in *Proc. 52nd Annu. Conf. Inf. Sci. Syst.*, 2018, pp. 1–6.

[46] C. Zhao et al., "Computational methods for in situ structural studies with cryogenic electron tomography," *Front. Cellular Infection Microbiol.*, vol. 13, 2023, Art. no. 1135013.

[47] S. H. Scheres, "RELION: Implementation of a Bayesian approach to cryo-EM structure determination," *J. Struct. Biol.*, vol. 180, no. 3, pp. 519–530, 2012.

[48] A. Punjani, J. L. Rubinstein, D. J. Fleet, and M. A. Brubaker, "cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination," *Nature Methods*, vol. 14, no. 3, pp. 290–296, 2017.

[49] E. Levin, T. Bendory, N. Boumal, J. Kileel, and A. Singer, "3D ab initio modeling in cryo-EM by autocorrelation analysis," in *Proc. IEEE 16th Int. Symp. Biomed. Imag.*, 2018, pp. 1569–1573.

[50] N. Sharon, J. Kileel, Y. Khoo, B. Landa, and A. Singer, "Method of moments for 3D single particle ab initio modeling with non-uniform distribution of viewing angles," *Inverse Problems*, vol. 36, no. 4, 2020, Art. no. 044003.

[51] D. Kimanius et al., "Exploiting prior knowledge about biological macromolecules in cryo-EM structure determination," *IUCrJ*, vol. 8, no. 1, pp. 60–75, 2021.

[52] R. Henderson, "The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules," *Quart. Rev. Biophys.*, vol. 28, no. 2, pp. 171–193, 1995.

[53] T. Bendory, N. Boumal, W. Leeb, E. Levin, and A. Singer, "Toward single particle reconstruction without particle picking: Breaking the detection limit," *SIAM J. Imag. Sci.*, vol. 16, no. 2, pp. 886–910, 2023.

[54] S. Kreymer, A. Singer, and T. Bendory, "A stochastic approximate expectation-maximization for structure determination directly from cryo-EM micrographs," 2023, *arXiv:2303.02157*.

[55] A. Zabatani, S. Kreymer, and T. Bendory, "Score-based diffusion priors for multi-target detection," in *Proc. IEEE 58th Annu. Conf. Inf. Sci. Syst.*, 2024, pp. 1–6.

[56] Y. Sun, B. Wohlberg, and U. S. Kamilov, "An online plug-and-play algorithm for regularized image reconstruction," *IEEE Trans. Comput. Imag.*, vol. 5, no. 3, pp. 395–408, Sep. 2019.