

# AN APPROXIMATE EXPECTATION-MAXIMIZATION FOR TWO-DIMENSIONAL MULTI-TARGET DETECTION

Shay Kreymer<sup>\*</sup>, Amit Singer<sup>†</sup>, and Tamir Bendory<sup>\*</sup>

<sup>\*</sup>School of Electrical Engineering, Tel Aviv University, Tel Aviv, Israel

<sup>†</sup>Department of Mathematics and PACM, Princeton University, Princeton, NJ, USA

## ABSTRACT

We consider the two-dimensional multi-target detection (MTD) problem of **recovering** a target image from a noisy measurement that contains multiple copies of the image, each randomly rotated and translated. The MTD model serves as a mathematical abstraction of the structure reconstruction problem in single-particle cryo-electron microscopy, the chief motivation of this study. We focus on high noise regimes, where accurate detection of image occurrences within a measurement is impossible. To estimate the image, we develop an approximate expectation-maximization framework that aims to maximize an approximation of the maximum likelihood estimator. We demonstrate image recovery in highly noisy environments, and show that our framework outperforms the previously studied autocorrelation analysis in a wide range of **noise levels**. The code to reproduce all numerical experiments is publicly available at <https://github.com/krshay/MTD-2D-EM>.

**Index Terms**— Expectation-maximization, multi-target detection, cryo-electron microscopy.

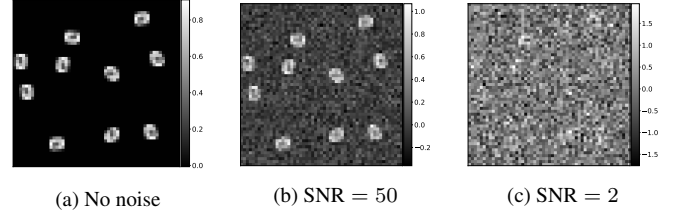
## 1. INTRODUCTION

We study the multi-target detection (MTD) problem of estimating a target image  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  from a noisy measurement that contains multiple copies of the image, each randomly rotated and translated [1, 2, 3, 4, 5, 6, 7]. We consider a measurement  $M \in \mathbb{R}^{N \times N}$  of the form

$$M[\vec{\ell}] = \sum_{i=1}^p F_{\phi_i}[\vec{\ell} - \vec{\ell}_i] + \varepsilon[\vec{\ell}], \quad (1)$$

where  $F_{\phi_i}[\vec{\ell}] := f_{\phi_i}(\vec{\ell}/n)$  is a discrete copy of  $f$ , rotated by angle  $\phi_i$  about the origin;  $n$  is the radius of the image in pixels;  $\{\phi_i\}_{i=1}^p \sim \text{Unif}[0, 2\pi)$  are uniformly distributed rotations;  $\{\vec{\ell}_i\}_{i=1}^p \in \{n+1, \dots, N-n\}^2$  are arbitrary translations; and  $\varepsilon[\vec{\ell}]$  is **i.i.d.** Gaussian noise with zero mean and variance  $\sigma^2$ . The rotations, translations, and the number of occurrences of  $f$  in  $M$ , denoted by  $p$ , are unknown. Importantly, since the rotations are unknown, it is possible to reconstruct the target image only up to a rotation.

**We assume the** the image  $f$  is supported on the unit disk  $D = \{\vec{x} \in \mathbb{R}^2 : |\vec{x}| \leq 1\}$  and has a finite expansion in the basis of Dirichlet Laplacian eigenfunctions (**see [3, 4, 5] for details**). In



**Fig. 1:** Three measurements at different SNRs: (a) no noise; (b) SNR = 50; (c) SNR = 2. Each measurement contains multiple rotated versions of the target image. We focus on the low SNR regime (e.g., panel (c)) in which the locations and rotations of image occurrences cannot be detected reliably, **and aim to estimate the target image directly from the measurement.**

particular, **it implies that**

$$f(r, \theta) = \sum_{(\nu, q): \lambda_{\nu, q} \leq \lambda} \alpha_{\nu, q} \psi_{\nu, q}(r, \theta), \quad (2)$$

where  $\psi_{\nu, q}(r, \theta) = J_{\nu}(\lambda_{\nu, q} r) e^{i\nu\theta}$ , for  $r \leq 1$ ,  $\nu \in \mathbb{Z}_{\geq 0}$ ,  $J_{\nu}$  is the  $\nu$ -th order Bessel function of the first kind,  $\lambda_{\nu, q} > 0$  is the  $q$ -th positive root of  $J_{\nu}$ , and  **$\alpha$  is the vector of expansion coefficients**. Hereafter, by estimating the image we mean estimating the vector of coefficients  $\alpha$ . **The basis of Dirichlet Laplacian eigenfunctions is steerable:** rotating  $f$  is equivalent to modulating the expansion coefficients  $\alpha_{\nu, q}$ . **Specifically,**  $f_{\phi}(r, \theta) := f(r, \theta + \phi)$  is given by

$$f_{\phi}(r, \theta) = \sum_{(\nu, q): \lambda_{\nu, q} \leq \lambda} \alpha_{\nu, q} \psi_{\nu, q}(r, \theta) e^{i\nu\phi}. \quad (3)$$

**In this work, we will focus on the analysis of a discrete version of  $f$ ,  $F_{\phi} : \mathbb{Z}^2 \rightarrow \mathbb{R}$ , which is defined by  $F_{\phi}[\vec{\ell}] = f_{\phi}(\vec{\ell}/n)$ , for  $\vec{\ell} \in \mathbb{Z}^2$ .**

We focus on the well-separated case of the 2-D MTD problem, which was introduced in [3, 4]. In this case, each translation in the measurement  $M$  is separated by at least a full image diameter from its neighbors. Specifically, we assume **that**

$$|\vec{\ell}_{i_1} - \vec{\ell}_{i_2}| > 4n, \quad \text{for all } i_1 \neq i_2. \quad (4)$$

Figure 1 presents an example of a measurement  $M$  at different signal-to-noise ratios (SNRs). We define  $\text{SNR} := \frac{\|F\|_F^2}{A\sigma^2}$ , where  $A$  is the area in pixels of  $F$ .

The MTD model serves as a mathematical abstraction of the cryo-electron microscopy (cryo-EM) technology for macromolecular structure determination [8, 9, 10]. In a cryo-EM experiment [11], individual copies of the target biomolecule are dispersed at unknown 2-D locations and 3-D orientations in a thin layer of vitreous ice, from which 2-D tomographic projection images are produced by an electron microscope. It is necessary to keep the electron dose low in order to minimize the irreversible structural damage. Consequently,

S.K. is supported by the Yitzhak and Chaya Weinstein Research Institute for Signal Processing. A.S. is partly supported by AFOSR Award FA9550-20-1-0266, the Simons Foundation Math+X Investigator Award, NSF BIG-DATA Award IIS1837992, NSF Award DMS-2009753, and NIH/NIGMS Award R01GM136780-01. T.B. is supported in part by NSF-BSF grant no. 2019752.

the projection images are considerably noisy. In the current data processing pipeline for cryo-EM [12, 13, 14], the 2-D projections are first detected and extracted from the micrograph, and later rotationally and translationally aligned to reconstruct the 3-D molecular structure. This approach fails for small molecules, which are difficult to detect and align [7, 8, 12, 15].

The MTD model was devised in [7] in order to study the recovery of small molecules directly from the micrograph, below the current detection limit of cryo-EM [8, 16]. In [3, 4, 5], an autocorrelation analysis technique was implemented for the two-dimensional MTD problem. Autocorrelation analysis is a special case of the method of moments, and it consists of finding an image that best matches the empirical autocorrelations of the measurement. We use the first three autocorrelations since the third-order autocorrelation is the lowest-order autocorrelation that determines a generic signal uniquely [1]. In this work, we propose to replace the method by using expectation-maximization (EM) [17]. Similarly to autocorrelation analysis, the EM algorithm estimates the target image  $F$  directly, circumventing estimating the image locations in the measurement. EM iteratively maximizes the likelihood function of the image (the translations and rotations are marginalized). Importantly, EM is guaranteed to increase the likelihood at each iteration [17].

However, for the MTD model, the number of admissible configurations for the locations grows exponentially with the measurement size. Therefore, the direct application of the EM algorithm to the MTD problem is computationally intractable, even for very small measurements. Thus, following [2], we suggest mitigating the computational burden by approximating the likelihood.

Previous works suggest that EM outperforms autocorrelation analysis in terms of estimation accuracy, for the one-dimensional MTD problem [2] and for the closely related multireference alignment problem [18]. In addition, EM is the most popular computational framework for cryo-EM [13, 14]. Moreover, a recent paper [19] shows that likelihood optimization in the low SNR regime reduces to a sequence of least squares optimization problems that match the moments of the estimate to the ground truth moments one by one, and by that suggests that EM has the potential to surpass the estimation accuracy achieved by autocorrelation analysis.

The main contribution of this paper is in developing an approximate EM framework for the two-dimensional MTD problem. We demonstrate a successful reconstruction in noisy regimes (see Section 3). We conduct extensive comparison experiments between the approximate EM framework and the autocorrelation analysis scheme, and suggest significant improvement in estimation accuracy. It is thus a significant step towards efficiently estimating a molecular structure directly from noisy cryo-EM experimental datasets [7].

## 2. APPROXIMATE EXPECTATION-MAXIMIZATION

Given a measurement  $M$  that follows the MTD model (1), the maximum marginal likelihood estimator (MMLE) for the vector of coefficients  $\alpha$ , that represents the target image  $F$ , is the maximizer of the likelihood function  $p(M|\alpha)$ . The locations and rotations of the target images within the measurement are treated as nuisance variables. The EM algorithm estimates the MMLE by iteratively applying the expectation (E) and maximization (M) steps [17]. Specifically, given the current estimate  $\alpha_k$ , the E-step computes the expected log-likelihood function, where the expectation is taken over all admissible configurations of locations and rotations. The estimate is then updated in the M-step by maximizing the function with respect to  $\alpha$ . Unfortunately, for the MTD model, the number of

admissible configurations for the locations grows exponentially with the problem size, rendering the direct application of EM computationally intractable. Hence, based on [2], we suggest to apply EM to an approximation of the likelihood function.

The approximate EM begins by partitioning the measurement  $M$  into  $N_d = (N/L)^2$  non-overlapping patches; each patch is of size  $L \times L$ , where  $L = 2n + 1$  is the diameter of the target image  $F$ . Then, we aim to estimate the image by the maximizer of the approximate likelihood function

$$p(M_0, M_1, \dots, M_{N_d-1}|\alpha) \approx \prod_{m=0}^{N_d-1} p(M_m|\alpha), \quad (5)$$

where  $M_m$  is the  $m^{\text{th}}$  patch, and we neglect the dependencies between patches. Our approximate EM algorithm works by applying the EM algorithm to estimate the approximate MMLE of (5).

The separation condition (4) implies that each patch contains parts of at most one rotated image. Specifically, to account for all  $4L^2 - 1$  different ways that a rotated image can be presented in a patch (including the possibility that there is no image in the patch), we model the  $m^{\text{th}}$  patch by

$$M_m = CT_{\vec{l}_m} ZF_{\phi_m} + \varepsilon_m, \quad \varepsilon_m \sim \mathcal{N}(0, \sigma^2 I_{L \times L}). \quad (6)$$

The operator  $Z$  zero-pads  $L$  entries to the right and to the bottom of the rotated copy of  $F$ , and  $T_{\vec{l}_m}$  circularly shifts the zero-padded image by  $\vec{l}_m = (\ell_{mx}, \ell_{my})$  positions, that is

$$T_{\vec{l}_m} ZF_{\phi_m} [i, j] = ZF_{\phi_m} [(i + \ell_{mx}) \bmod 2L, (j + \ell_{my}) \bmod 2L], \quad (7)$$

where  $\vec{l}_m \in \mathbb{L} := \{0, 1, \dots, 2L - 1\}^2$ , and is treated as a random variable. The operator  $C$  then crops the first  $L$  entries in the vertical and horizontal axes, and the result is further corrupted by additive white Gaussian noise. In addition, since the EM algorithm assigns probabilities to different rotations (in the expectation step), we need to discretize the search space in the angular direction:

$$\phi_m \in \Phi := \left\{ k \frac{2\pi}{K} \right\}_{k=0}^{K-1}, \quad (8)$$

where  $K$  is a parameter. Higher  $K$  means higher accuracy at the cost of running time (see Figure 4).

In the E-step, our algorithm calculates the expected log-likelihood function

$$Q(\alpha|\alpha_k) = \sum_{m=0}^{N_d-1} \sum_{\vec{l} \in \mathbb{L}} \sum_{\phi \in \Phi} p(\vec{l}, \phi | M_m, \alpha_k) \log p(M_m, \vec{l}, \phi | \alpha), \quad (9)$$

given the current coefficients estimate  $\alpha_k$ , where

$$p(M_m | \vec{l}, \phi, \alpha) \propto \exp \left( - \frac{\|M_m - CT_{\vec{l}} ZF_{\phi}\|_F^2}{2\sigma^2} \right), \quad (10)$$

with the normalization  $\sum_{\vec{l} \in \mathbb{L}} \sum_{\phi \in \Phi} p(M_m | \vec{l}, \phi, \alpha) = 1$ . From Bayes' rule, we have

$$p(\vec{l}, \phi | M_m, \alpha_k) = \frac{p(M_m | \vec{l}, \phi, \alpha_k) p(\vec{l}, \phi | \alpha_k)}{\sum_{\vec{l}' \in \mathbb{L}} \sum_{\phi' \in \Phi} p(M_m | \vec{l}', \phi', \alpha_k) p(\vec{l}', \phi' | \alpha_k)}, \quad (11)$$

which is the normalized likelihood function  $p(M_m | \vec{l}, \phi, \alpha_k)$ , weighted by the prior distribution  $p(\vec{l}, \phi | \alpha_k)$ . We assume that  $p(\vec{l}, \phi | \alpha_k) = p(\vec{l})p(\phi)$ , namely,  $\vec{l}$  and  $\phi$  are independent in  $\alpha_k$  and each other. We also assume that the rotations are drawn from a uniform distribution, and therefore

$$p(\vec{l}, \phi | \alpha_k) = p(\vec{l}, \phi) = p(\vec{l})p(\phi) := \rho[\vec{l}] \frac{1}{K}, \quad (12)$$

where we assume the rotations are uniformly distributed in the set  $\Phi$  from (8), and  $\rho[\vec{\ell}]$  is the distribution of image locations within the patch. We can rewrite (9) as (up to a constant)

$$Q(\alpha, \rho | \alpha_k, \rho_k) = \sum_{m=0}^{N_d-1} \sum_{\vec{\ell} \in \mathbb{L}} \sum_{\phi \in \Phi} p(M_m | \vec{\ell}, \phi, \alpha_k) \rho_k[\vec{\ell}] \times \left( \log p(M_m | \vec{\ell}, \phi, \alpha) + \log \rho[\vec{\ell}] \right), \quad (13)$$

where  $p(M_m | \vec{\ell}, \phi, \alpha)$  is given by (10).

The M-step updates the image estimate and  $\rho$  by maximizing  $Q(\alpha, \rho | \alpha_k, \rho_k)$  under the constraint that  $\rho$  lies on the simplex  $\Delta_{2L \times 2L}$  for  $\rho$ :

$$\alpha_{k+1}, \rho_{k+1} = \arg \max_{\alpha, \rho} Q(\alpha, \rho | \alpha_k, \rho_k) \text{ s.t. } \rho \in \Delta_{2L \times 2L}. \quad (14)$$

The constrained maximization of (14) can be achieved with the unconstrained maximization of the Lagrangian

$$\mathcal{L}(\alpha, \rho, \eta) = Q(\alpha, \rho | \alpha_k, \rho_k) + \eta \left( 1 - \sum_{\vec{\ell} \in \mathbb{L}} \rho[\vec{\ell}] \right), \quad (15)$$

where  $\eta$  is the Lagrange multiplier. As we will see later, the constraints are automatically satisfied at the computed local maximum of the Lagrangian. Since  $Q(\alpha, \rho | \alpha_k, \rho_k)$  is additively separable for  $\alpha$  and  $\rho$ , we maximize  $\mathcal{L}(\alpha, \rho, \eta)$  with respect to  $\alpha$  and  $\rho$  separately. At the maximum of  $\mathcal{L}(\alpha, \rho, \eta)$ , we have

$$0 = \frac{\partial \mathcal{L}}{\partial (\alpha)_{\nu, q}} = \sum_{m=0}^{N_d-1} \sum_{\vec{\ell} \in \mathbb{L}} \sum_{\phi \in \Phi} p(M_m | \vec{\ell}, \phi, \alpha_k) \rho_k[\vec{\ell}] \times \frac{\partial \log p(M_m | \vec{\ell}, \phi, \alpha)}{\partial (\alpha)_{\nu, q}}, \quad (16)$$

resulting in a set of linear equations which is solved to achieve the updated  $\alpha$ . In order to update  $\rho$ , we maximize  $\mathcal{L}(\alpha, \rho, \eta)$  with respect to  $\rho$ :

$$0 = \frac{\partial \mathcal{L}}{\partial \rho[\vec{\ell}]} = \sum_{m=0}^{N_d-1} \sum_{\phi \in \Phi} p(M_m | \vec{\ell}, \phi, \alpha_k) \rho_k[\vec{\ell}] \frac{1}{\rho[\vec{\ell}]} - \eta, \quad (17)$$

for  $\vec{\ell} \in \mathbb{L}$ . We thus obtain the update rule for  $\rho$  as

$$\rho[\vec{\ell}] = \frac{1}{\eta} \sum_{m=0}^{N_d-1} \sum_{\phi \in \Phi} p(M_m | \vec{\ell}, \phi, \alpha_k) \rho_k[\vec{\ell}], \quad (18)$$

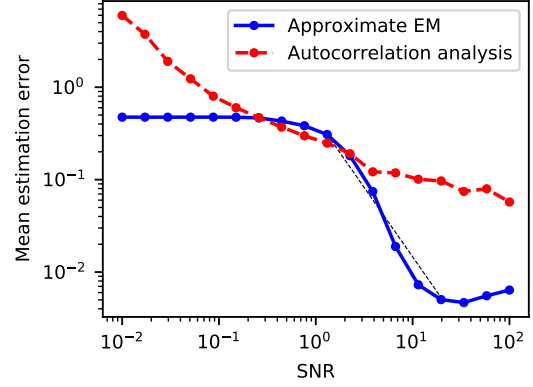
and we can immediately solve  $\eta = N_d$  from the normalization  $\sum_{\vec{\ell} \in \mathbb{L}} \rho[\vec{\ell}] = 1$ .

### 3. NUMERICAL EXPERIMENTS

In this section, we present numerical results for the approximate EM described in Section 2. As a baseline, we compare the results to the recovery achieved using autocorrelation analysis, based on the framework (and code) of [4, 5]. To take the in-plane rotation symmetry into account, we measure the estimation error by

$$\text{relative error}_\alpha := \min_{\phi \in [0, 2\pi)} \frac{\|\alpha^* - \alpha_\phi\|_2}{\|\alpha^*\|_2}, \quad (19)$$

where  $\alpha^*$  is the true vector of expansion coefficients, and  $\alpha_\phi$  is the vector of coefficients of the estimated image, rotated by angle  $\phi$ . In all experiments, the well-separated measurements were generated according to (1) with density  $\gamma = 0.04$ , where  $p = \gamma \frac{N^2}{L^2}$ . The target images are of diameter  $L = 5$  pixels. Each entry of the target images was drawn i.i.d. from a uniform distribution on  $[0, 1]$ .



**Fig. 2:** The mean estimation error of recovering the target image  $F$ , as a function of the SNR, by: (a) approximate EM; (b) autocorrelation analysis.

Then, each image was normalized such that  $\|F\|_F = 10$ , and expanded using its first 10 coefficients. We try to estimate the image from initial guesses that were drawn from the same distribution as the ground truth image, and  $\gamma_{\text{init}} = 0.03$ . We then calculate the estimation error of the image estimate whose final objective function is minimal. Figures 2, 3 and 4 present the mean error over 40 trials. The code to reproduce all experiments is publicly available at <https://github.com/krshay/MTD-2D-EM>.

#### 3.1. Recovery error as a function of the SNR

Figure 2 presents recovery error as a function of the SNR. The measurements are of size  $N = 2000$  pixels, and we use  $K = 8$  samples at the angular direction, and one random initial guesses for  $\alpha$ . To save computation time (see Section 4), we initialize the approximate EM algorithm using the estimate achieved by autocorrelation analysis. We achieve a significant improvement in recovery accuracy using approximate EM, even though the search space in the angular direction is coarsely sampled, for all SNR regimes.

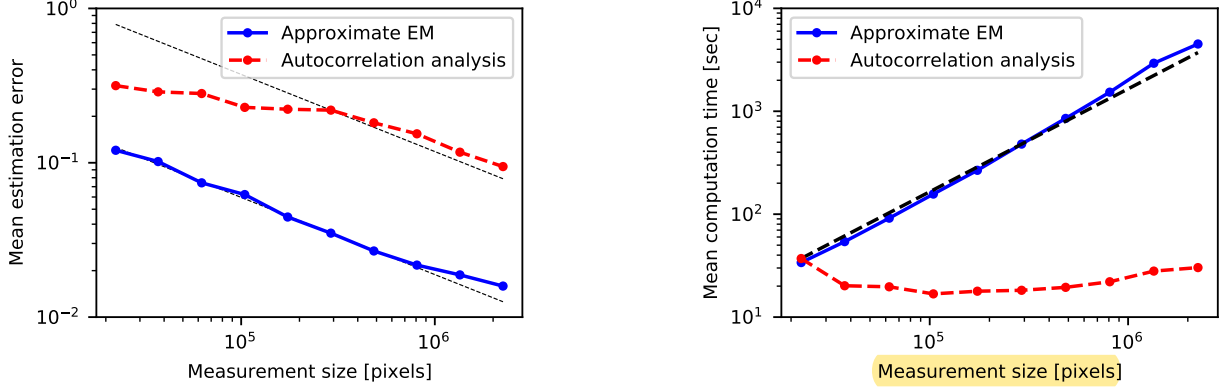
We further report results for the recovery errors in estimating a target image from a measurement with SNR = 2 (a low SNR environment), using approximate EM and autocorrelation analysis. We consider a measurement of size  $N = 10000$  pixels, and use  $K = 16$  samples at the angular direction. The relative error is 0.017 using approximate EM, and 0.073 using autocorrelation analysis.

#### 3.2. Recovery error as a function of the measurement size

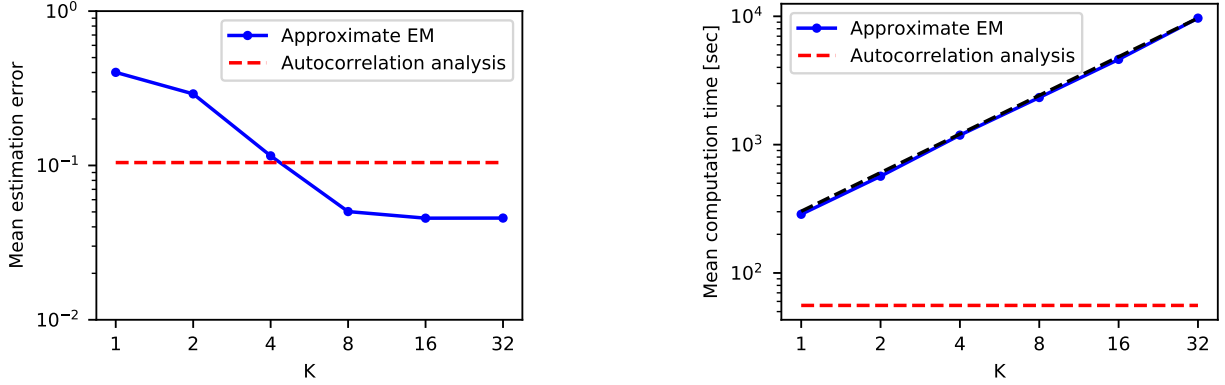
Figure 3 presents recovery error and running time as a function of the measurement size, with SNR = 5,  $K = 16$  samples at the angular direction, and 5 random initial guesses for  $\alpha$ . Using approximate EM, the error decays as  $1/\sqrt{N^2}$ . The same trend is visible also for autocorrelation analysis for sufficiently large measurements. We achieve a significant improvement in recovery accuracy using approximate EM, although the search space in the angular direction is coarsely sampled. However, the computation time using our method is significantly larger, and grows linearly with the measurement size.

#### 3.3. Recovery error as a function of discretization of rotations

Figure 4 presents recovery error and running time as a function of  $K$ , the size of the search space in the angular direction, for



**Fig. 3:** The mean estimation error of recovering the target image  $F$  (left) and running time (right), as a function of measurement size, by: (a) approximate EM; (b) autocorrelation analysis. For the estimation error, the black dashed lines illustrate a slope of  $-1/2$ , as predicted by the law of large numbers. For the computation time, the black dashed line illustrates a slope of 1, which implies a linear increase in computation time, as the number of patches  $N_d = N^2/L^2$  grows linearly in  $N^2$  ( $N^2$  is the total number of pixels in a measurement).



**Fig. 4:** The mean estimation error of recovering the target image  $F$  (left) and running time (right), as a function of  $K$ , the size of the search space in the angular direction, by approximate EM. The recovery error and running time using autocorrelation analysis are marked by the dashed horizontal red lines. For the running time, the black dashed line illustrates a slope of 1, which implies a linear increase in computation time, as the number of computations per patch depends linearly on  $K$ .

measurements with  $\text{SNR} = 5$ ,  $N = 1500$  pixels, and one random initial guess for  $\alpha$ . The results are compared against autocorrelation analysis. Remarkably, even when the EM searches over only 4 rotations, it achieves estimation errors similar to the estimation errors of autocorrelation analysis, which takes all possible rotations into account. As expected, the computation time grows linearly with the parameter  $K$ .

#### 4. CONCLUSION

This paper is motivated by the effort of reconstructing small 3-D molecular structures using cryo-EM, below the current detection limit. The main contribution of this paper is in introducing an approximate EM scheme for the two-dimensional MTD problem, and comparing it numerically to autocorrelation analysis. The numerical experiments show an improvement in estimation accuracy, at the cost of an increase in required computational time. This is since computing the autocorrelations is straightforward and requires only one pass over the data, while the EM algorithm scans through

the whole dataset in each iteration. As Figure 4 shows, the parameter  $K$  provides an accuracy-running time trade-off. A possible improvement is increasing the resolution of the search space as the iterations progress [2]; this is a standard procedure in current cryo-EM software packages [13, 14].

Our ultimate goal is developing an approximate EM scheme for recovering small molecular structures using cryo-EM [7]. Although it significantly outperforms autocorrelation analysis in terms of accuracy, a major drawback of the approximate EM method is its computation time. Therefore, in order to achieve a computationally efficient algorithm for the 3-D case of the MTD problem, parallel processing and randomized algorithms, such as stochastic or online EM [20, 21, 22, 23, 24], must be utilized. Another acceleration can be achieved by applying the EM on a lower dimensional representation [25]. Furthermore, in the future, EM may be replaced by more intricate techniques, such as variational inference [26] or variational auto-encoders [27]. Finally, adding a prior on the target image should improve the robustness and accuracy of the scheme, at the cost of a possible model bias.

## 5. REFERENCES

- [1] Tamir Bendory, Nicolas Boumal, William Leeb, Eitan Levin, and Amit Singer, “Multi-target detection with application to cryo-electron microscopy,” *Inverse Problems*, vol. 35, no. 10, pp. 104003, 2019.
- [2] Ti-Yen Lan, Tamir Bendory, Nicolas Boumal, and Amit Singer, “Multi-target detection with an arbitrary spacing distribution,” *IEEE Transactions on Signal Processing*, vol. 68, pp. 1589–1601, 2020.
- [3] Nicholas F Marshall, Ti-Yen Lan, Tamir Bendory, and Amit Singer, “Image recovery from rotational and translational invariants,” in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 5780–5784.
- [4] Tamir Bendory, Ti-Yen Lan, Nicholas F Marshall, Iris Rukshin, and Amit Singer, “Multi-target detection with rotations,” *arXiv preprint arXiv:2101.07709*, 2021.
- [5] Shay Kreymer and Tamir Bendory, “Two-dimensional multi-target detection: an autocorrelation analysis approach,” *arXiv preprint arXiv:2105.06765*, 2021.
- [6] Ye’Ela Shalit, Ran Weber, Asaf Abas, Shay Kreymer, and Tamir Bendory, “Generalized autocorrelation analysis for multi-target detection,” *arXiv preprint arXiv:2109.11813*, 2021.
- [7] Tamir Bendory, Nicolas Boumal, William Leeb, Eitan Levin, and Amit Singer, “Toward single particle reconstruction without particle picking: breaking the detection limit,” *arXiv preprint arXiv:1810.00226*, 2018.
- [8] Richard Henderson, “The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules,” *Quarterly Reviews of Biophysics*, vol. 28, no. 2, pp. 171–193, 1995.
- [9] Eva Nogales, “The development of cryo-EM into a mainstream structural biology technique,” *Nature methods*, vol. 13, no. 1, pp. 24–27, 2016.
- [10] Xiao-Chen Bai, Greg McMullan, and Sjors HW Scheres, “How cryo-EM is revolutionizing structural biology,” *Trends in Biochemical Sciences*, vol. 40, no. 1, pp. 49–57, 2015.
- [11] Joachim Frank, *Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state*, Oxford University Press, 2006.
- [12] Tamir Bendory, Alberto Bartesaghi, and Amit Singer, “Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities,” *IEEE Signal Processing Magazine*, vol. 37, no. 2, pp. 58–76, 2020.
- [13] Sjors HW Scheres, “RELION: implementation of a Bayesian approach to cryo-EM structure determination,” *Journal of Structural Biology*, vol. 180, no. 3, pp. 519–530, 2012.
- [14] Ali Punjani, John L Rubinstein, David J Fleet, and Marcus A Brubaker, “cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination,” *Nature methods*, vol. 14, no. 3, pp. 290–296, 2017.
- [15] Cecilia Aguerreberere, Mauricio Delbracio, Alberto Bartesaghi, and Guillermo Sapiro, “Fundamental limits in multi-image alignment,” *IEEE Transactions on Signal Processing*, vol. 64, no. 21, pp. 5707–5722, 2016.
- [16] Edoardo D’Imprima and Werner Kühlbrandt, “Current limitations to high-resolution structure determination by single-particle cryoEM,” *Quarterly Reviews of Biophysics*, vol. 54, 2021.
- [17] Arthur P Dempster, Nan M Laird, and Donald B Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.
- [18] Tamir Bendory, Nicolas Boumal, Chao Ma, Zhizhen Zhao, and Amit Singer, “Bispectrum inversion with application to multi-reference alignment,” *IEEE Transactions on signal processing*, vol. 66, no. 4, pp. 1037–1050, 2017.
- [19] Anya Katsevich and Afonso Bandeira, “Likelihood maximization and moment matching in low SNR gaussian mixture models,” *arXiv preprint arXiv:2006.15202*, 2020.
- [20] Søren Feodor Nielsen, “The stochastic EM algorithm: estimation and asymptotic results,” *Bernoulli*, pp. 457–489, 2000.
- [21] Jianfei Chen, Jun Zhu, Yee Whye Teh, and Tong Zhang, “Stochastic expectation maximization with variance reduction,” in *NeurIPS*, 2018, pp. 7978–7988.
- [22] Percy Liang and Dan Klein, “Online EM for unsupervised models,” in *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 2009, pp. 611–619.
- [23] Olivier Cappé and Eric Moulines, “On-line expectation-maximization algorithm for latent data models,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 71, no. 3, pp. 593–613, 2009.
- [24] Olivier Cappé, “Online EM algorithm for hidden Markov models,” *Journal of Computational and Graphical Statistics*, vol. 20, no. 3, pp. 728–749, 2011.
- [25] Nícha C Dvornek, Fred J Sigworth, and Hemant D Tagare, “SubspaceEM: A fast maximum-a-posteriori algorithm for cryo-EM single particle reconstruction,” *Journal of Structural Biology*, vol. 190, no. 2, pp. 200–214, 2015.
- [26] David M Blei, Alp Kucukelbir, and Jon D McAuliffe, “Variational inference: A review for statisticians,” *Journal of the American Statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.
- [27] Dan Rosenbaum, Marta Garnelo, Michal Zielinski, Charlie Beattie, Ellen Clancy, Andrea Huber, Pushmeet Kohli, Andrew W Senior, John Jumper, Carl Doersch, et al., “Inferring a continuous distribution of atom coordinates from cryo-EM images using VAEs,” *arXiv preprint arXiv:2106.14108*, 2021.