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

Shay Kreymer*, Amit Singer†, and Tamir Bendory*

*School of Electrical Engineering, Tel Aviv University, Tel Aviv, Israel
†Department of Mathematics and PACM, Princeton University, Princeton, NJ, USA

ABSTRACT

We consider the two-dimensional multi-target detection (MTD) problem of estimating 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 expectation-maximization framework that aims to maximize an approximation of the likelihood function. We demonstrate image recovery in highly noisy environments, and show that our framework outperforms the previously studied autocorrelation analysis in a wide range of parameters. 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 $f : \mathbb{R}^2 \to \mathbb{R}$ 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

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

where $F_{\phi_i}[\vec{\ell}] := f_{\phi_i}(\vec{\ell}/n)$ is a discrete copy of f, rotated by angle ϕ_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 σ^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.

Following [3, 4, 5], we assume that the image f is supported on the unit disk $\{\vec{x} \in \mathbb{R}^2 : |\vec{x}| \le 1\}$ and has a finite expansion in the basis of Dirichlet Laplacian eigenfunctions. In particular, the

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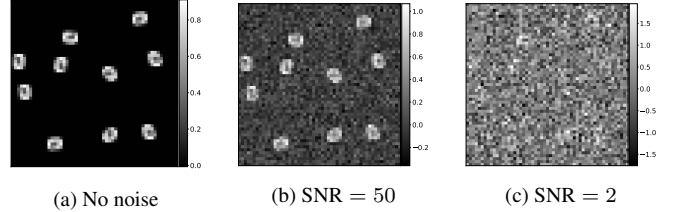
Fig. 1: Three measurements at different SNRs: (a) no noise; (b) SNR = 50; (c) SNR = 2. Each measurement contains multiple rotated versions of the target image. We focus on the low SNR regime (e.g., panel (c)) in which the locations and rotations of the image occurrences cannot be detected reliably.

image f can be expanded as

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

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

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

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

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

Figure 1 presents an example of a measurement M at different signal-to-noise ratios (SNRs). We define 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 irreversible structural damage. Consequently, the projection images are considerably noisy. In the current data pro-

cessing pipeline of cryo-EM [12, 13, 14, 15], 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, 16].

The MTD model was devised in [7] in order to study the recovery of small molecular structures using cryo-EM, below the current detection limit [17]. In [3, 4, 5], an autocorrelation analysis technique was devised for the 2-D MTD problem (1). 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, thus bypassing detecting the locations and rotations of individual image occurrences. In this work, we propose to replace autocorrelation analysis by expectation-maximization (EM): a classical algorithm to compute the maximum likelihood estimator [18]. Similarly to autocorrelation analysis, the EM algorithm estimates the target image F directly by marginalizing over the translations and rotations.

Previous works demonstrated that EM outperforms autocorrelation analysis in terms of estimation accuracy for the 1-D MTD problem [2] as well as for the closely related multireference alignment model [19, 20]. EM is also the most popular computational framework for reconstructing molecular structures using cryo-EM, e.g., [14, 15]. Moreover, a recent paper [21] 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.

At each iteration, EM assigns probabilities to all possible rotations and translations (see Section 2). Unfortunately, for the MTD model, the number of possible translations grows quickly with the measurement size, and therefore 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 developing an EM algorithm that maximizes an approximation of the likelihood function. In the approximate EM scheme, the number of possible translations is linear in N^2, making the algorithm tractable.

The main contribution of this paper is in developing an approximate EM framework for the two-dimensional MTD problem; see Section 2. In Section 3, we demonstrate a successful reconstruction in noisy regimes. We also conduct extensive numerical experiments that demonstrate significant improvement in estimation accuracy compared to autocorrelation analysis. Section 4 concludes the paper and introduces future work. In particular, we discuss potential implications for the ongoing effort to estimate small molecular structures using cryo-EM [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 α, that represents the target image F (2), is the maximizer of the likelihood function $p(M|\alpha)$. The translations 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 [18]. Specifically, given the current estimate α_k, the E-step computes the expected log-likelihood function, where the expectation is taken over over all admissible configurations of translations and rotations. The estimate is then updated in the M-step by maximizing the function

with respect to α. Unfortunately, for the MTD model, the number of possible translations grows quickly with the problem size, rendering direct application of EM computationally intractable. Hence, based on [2], we suggest to apply EM to an approximation of the likelihood function, in which the number of possible translations is linear in N^2.

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 maximizing the approximate likelihood function

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

where M_m is the mth patch, and we neglect statistical 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 M_m can contain either no target image, a full rotated target image, or part of a rotated image; overall there are $(2L-1)^2$ possible configurations. In particular, each patch can be modeled by

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

where the operator Z zero-pads L entries to the right and to the bottom of the rotated copy of F, and $T_{\vec{\ell}_m}$ circularly shifts the zero-padded image by $\vec{\ell}_m = (\ell_{mx}, \ell_{my}) \in \mathbb{L} := \{0, 1, \ldots, 2L-1\}^2$ positions, that is,

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

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 of rotations:

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

where K is a parameter chosen by the user. 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 of the model (6)

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

given the current estimate α_k, where

$$p(M_m|\vec{\ell}, \phi, \alpha) \propto \exp\left(-\frac{\|M_m - CT_{\vec{\ell}}ZF_\phi\|_{\text{F}}^2}{2\sigma^2}\right), \qquad (10)$$

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

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

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

bution in the set Φ from (8), and therefore

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

where $\rho[\vec{\ell}]$ is the distribution of 2-D translations in (6) (which should be estimated simultaneously with α). 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 ρ by maximizing $Q(\alpha, \rho | \alpha_k, \rho_k)$ under the constraint that ρ lies on the simplex $\Delta_{2L \times 2L}$:

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

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

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

where η is the Lagrange multiplier. As we will see next, the constraint is automatically satisfied at the maximum of the Lagrangian.

Since $Q(\alpha, \rho | \alpha_k, \rho_k)$ is additively separable for α and ρ, we maximize $\mathcal{L}(\alpha, \rho, \eta)$ with respect to α and ρ 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 update α. In order to update ρ, we maximize $\mathcal{L}(\alpha, \rho, \eta)$ with respect to ρ:

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

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

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

and $\eta = N_d$ from the normalization $\sum_{\vec{\ell} \in \mathbb{L}} \rho[\vec{\ell}] = 1$. The approximate EM algorithm is summarized in Algorithm 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 against autocorrelation analysis with the first three autocorrelations based on the framework (and code) of [3, 4, 5]. To take the in-plane rotation symmetry into account, we measure the estimation error by

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

where α^* is the true vector of expansion coefficients, and α_ϕ is the vector of coefficients of the estimated image, rotated by angle ϕ. In all experiments, the measurements were generated according to (1)

Algorithm 1: Approximate EM for 2-D MTD

Input: measurement M; noise variance σ^2; initial guesses α_0 and ρ_0; parameter K (8); stopping parameter ϵ
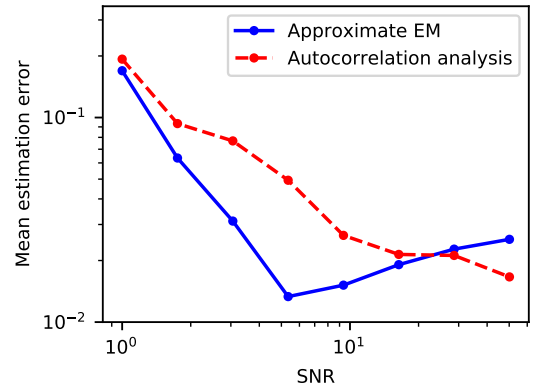
Output: an estimate for α and ρ

1 set $k \to 0$;
2 calculate Q_0 according to (9) and set $Q_{-1} \to -\infty$;
3 **while** $Q_k - Q_{k-1} > \epsilon$ **do**
4   calculate $p(M_m | \vec{\ell}, \phi, \alpha_k)$ according to (11) and ;
5   update α_{k+1} by solving (16);
6   update ρ_{k+1} according to (18);
7   calculate Q_{k+1} according to (9);
8   set $k \to k + 1$;

Fig. 2: The mean estimation error of recovering the target image F, as a function of the SNR, by approximate EM (Algorithm 1) and autocorrelation analysis.

with density $\gamma = 0.04$, where $p = \gamma \frac{N^2}{L^2}$. The rotations were drawn from a uniform distribution on $[0, 2\pi)$, while the search space of the EM was discretized with a parameter K according to (8). 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 $[0, 1]$. Then, each image was normalized such that $\|F\|_F = 10$, and expanded using its first 10 coefficients. The initializations of the EM and autocorrelation analysis iterations were drawn from the same distribution as the ground truth images, and $\gamma_{\text{init}} = 0.03$. If the algorithms were initialized from several random points, we calculated the estimation error of the image estimate whose final likelihood function is maximal (for approximate EM), or final objective function is minimal (for autocorrelation analysis). 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 $N = 3000$ pixels, and we use $K = 8$ possible rotations (8), and 5 random initial guesses for α. 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 of rotations is coarsely
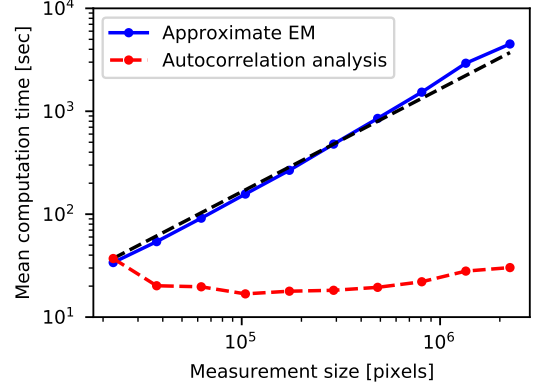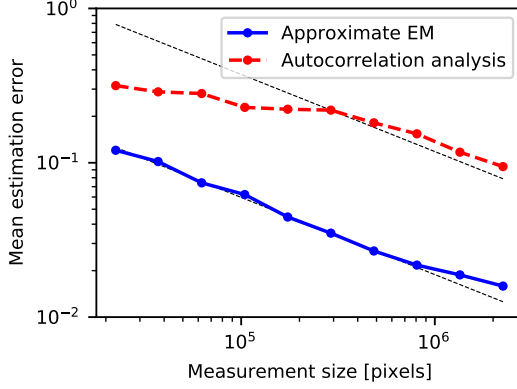
Fig. 3: The mean estimation error of recovering the target image F (left) and running time (right), as a function of measurement size N^2 by approximate EM and autocorrelation analysis. For the estimation error, the black dashed lines illustrate a slope of $-1/2$, as predicted by the law of large numbers. For the computation time, the black dashed line illustrates a slope of 1, implying a linear increase in computation time, as the number of patches $N_d = N^2/L^2$ grows linearly in N^2.
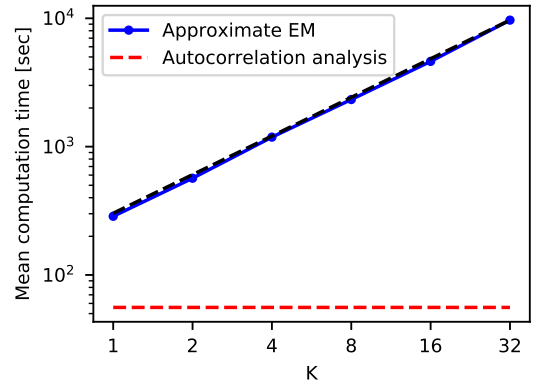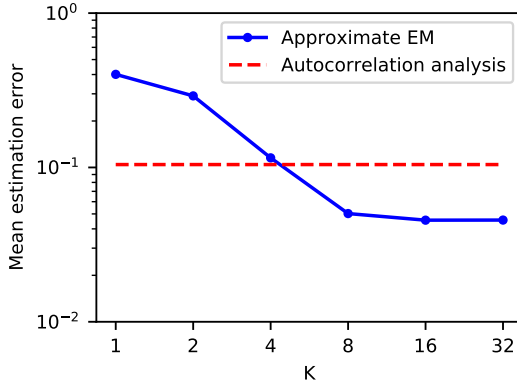
Fig. 4: The mean estimation error of recovering the target image F (left) and running time (right), as a function of K—the size of the search space of rotations (8)—by approximate EM. The recovery error and running time using autocorrelation analysis are marked by the dashed horizontal red lines. For 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. The results suggest that one can save running time by coarsely sampling the rotations without severely degrading the estimation quality.

sampled, for all SNR regimes.

In an additional numerical experiment with SNR $= 2$ (low SNR, see Figure 1), $N = 10000$ pixels, and $K = 16$, the relative error of approximate EM was 0.017, whereas the error of autocorrelation analysis was 0.073, more than four times larger.

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 N^2, with SNR $= 5$, $K = 16$ possible rotations, and 5 random initial guesses for α. 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 of rotations is coarsely sampled. However, the computation time using our method is greater, and grows linearly with the measurement size. In particular, autocorrelation analysis is faster since it requires a single pass over the measurement, while the computational complexity of matching an image to the observed autocorrelations scale with L^2 (which is much smaller than N^2).

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 of rotations, for measurements with SNR $= 5$, $N = 1500$, and one random initial guess for α. Remarkably, even when the EM searches over only 4 rotations (recall that the rotations are drawn from a continuous distribution), the obtained estimation errors are 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. This implies that one can save running time by coarsely sampling the rotations without severely degrading the estimation quality.

4. CONCLUSION

This paper is motivated by the effort of reconstructing small 3-D molecular structures using cryo-EM, below the current detection limit [7]. The main contribution of this paper is in introducing an approximate EM scheme for the 2-D MTD problem, and comparing it numerically to autocorrelation analysis. The numerical experiments show an improvement in estimation accuracy, but at the cost of computational time. 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 [14, 15].

Our ultimate goal is developing an approximate EM scheme for recovering small molecular structures using cryo-EM [7]. In order to achieve a computationally efficient algorithm for the 3-D case of cryo-EM, parallel processing and randomized algorithms, such as stochastic or online EM [22, 23, 24, 25, 26], must be utilized. A further acceleration can be achieved by applying the EM algorithm on a lower dimensional representation of the data [27]. Moreover, adding a prior on the target image is expected to improve the robustness and accuracy of the scheme, at the cost of a possible model bias. Another research direction is replacing EM by more intricate techniques that aim to approximate the posterior distribution, such as variational inference [28] or variational auto-encoders [29].

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] Amit Singer and Fred J Sigworth, "Computational methods for single-particle electron cryomicroscopy," *Annual Review of Biomedical Data Science*, vol. 3, pp. 163–190, 2020.

[14] 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.

[15] 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.

[16] Cecilia Aguerrebere, 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.

[17] Edoardo D'Imprima and Werner Kühlbrandt, "Current limitations to high-resolution structure determination by single-particle cryoEM," *Quarterly Reviews of Biophysics*, vol. 54, 2021.

[18] 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.

[19] Tamir Bendory, Nicolas Boumal, Chao Ma, Zhizhen Zhao, and Amit Singer, "Bispectrum inversion with application to multireference alignment," *IEEE Transactions on Signal Processing*, vol. 66, no. 4, pp. 1037–1050, 2017.

[20] Emmanuel Abbe, Tamir Bendory, William Leeb, João M Pereira, Nir Sharon, and Amit Singer, "Multireference alignment is easier with an aperiodic translation distribution," *IEEE Transactions on Information Theory*, vol. 65, no. 6, pp. 3565–3584, 2018.

[21] Anya Katsevich and Afonso Bandeira, "Likelihood maximization and moment matching in low SNR gaussian mixture models," *arXiv preprint arXiv:2006.15202*, 2020.

[22] Søren Feodor Nielsen, "The stochastic EM algorithm: estimation and asymptotic results," *Bernoulli*, pp. 457–489, 2000.

[23] Jianfei Chen, Jun Zhu, Yee Whye Teh, and Tong Zhang, "Stochastic expectation maximization with variance reduction," in *NeurIPS*, 2018, pp. 7978–7988.

[24] 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.

[25] 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.

[26] Olivier Cappé, "Online EM algorithm for hidden Markov models," *Journal of Computational and Graphical Statistics*, vol. 20, no. 3, pp. 728–749, 2011.

[27] Nicha 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.

[28] 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.

[29] 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.