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

Shay Kreymer and Tamir Bendory

School of Electrical Engineering, Tel Aviv University, Tel Aviv, Israel

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. In high noise regimes, accurate detection of the image occurrences within a measurement is impossible, rendering current algorithmic pipelines ineffective. We develop an approximate expectation-maximization framework to estimate the image directly from a measurement, and demonstrate image recovery in highly noisy environments. We conduct extensive numerical experiments and suggest improvement in estimation accuracy in comparison to previously studied autocorrelation analysis.

***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]. We consider a measurement $M \in \mathbb{R}^{N\times N}$ of the form

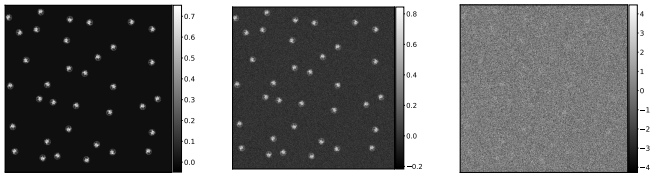
$$M[\vec{\ell}] = \sum_{i=1}^{p} F_{\phi_i}[\vec{\ell} - \vec{\ell}_i] + \varepsilon[\vec{\ell}], \tag{1}$$

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; $\{\phi_i\}_{i=1}^p \sim \mathrm{Unif}[0, 2\pi)$ 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. We discuss the well-separated case of the 2-D MTD problem, which was studied in [3] and [4]. In this case, we assume that each image F in the measurement M is separated by at least a full image diameter from its neighbors. Specifically, we assume that

$$|\vec{\ell}_{i_1} - \vec{\ell}_{i_2}| > 4n, \quad \text{for all } i_1 \neq i_2. \tag{2}$$

Section 2.1 introduces the image model of f in detail. Figure 1 presents an example of a measurement M at different signal-to-noise ratios (SNRs). We define SNR $:= \frac{\|F_0\|_{\mathrm{F}}^2}{A\sigma^2}$, where A is the area in pixels of F_0 (the unrotated image).

S.K. is supported by the Yitzhak and Chaya Weinstein Research Institute for Signal Processing. T.B. is supported in part by NSF-BSF grant no. 2019752.



(a) No noise. (b) SNR $= 50$. (c) SNR $= 0.1$.

Fig. 1: Three measurements at different noise levels: (a) no noise; (b) SNR $= 50$; (c) SNR $= 0.1$. Each measurement contains multiple rotated versions of the target image. In this work, our goal is to estimate the target image directly from the measurement. 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.

The MTD model serves as a mathematical abstraction of the cryo-electron microscopy (cryo-EM) technology for macromolecular structure determination [7, 8, 9]. In a cryo-EM experiment [10], 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, the projection images are considerably noisy.

In the current analysis workflow of cryo-EM data [11, 12, 13], 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 induce low contrast, and thus low SNR. This makes them difficult to detect and align [6, 7, 11, 14], rendering current cryo-EM algorithmic pipeline ineffective.

The MTD model was devised in [6] in order to study the recovery of small molecules directly from the micrograph, below the current detection limit of cryo-EM [7, 15]. In [4, 5], an autocorrelation analysis technique was implemented for the two-dimensional MTD problem. Autocorrelation analysis consists of finding an image that best explains the empirical autocorrelations of the measurement. Computing the autocorrelations is straightforward and requires only one pass over the data, which is advantageous for massively large datasets, such as cryo-EM datasets [11].

In this work, we propose an approximate version of the expectation maximization (EM) algorithm [16] to circumvent the estimation of the image locations in the measurement by estimating the target image F directly. In the approximate EM algorithm, the image is reconstructed by iteratively maximizing the data likelihood, which marginalizes over the rotations and translations; importantly, it does not estimate them explicitly. The approximate EM algorithm for the one-dimensional MTD problem was developed in [2]. In contrast to autocorrelation analysis, the approximate EM algorithm

scans through the whole dataset in each iteration, and hence requires much longer computational time. However, previous works [2, 17] suggest 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 well-separated MTD problem. We demonstrate a successful reconstruction in noisy regimes (see Section 3). Moreover, we conduct extensive comparison experiments between the approximate EM framework and the previously developed autocorrelation analysis scheme, and suggest significant improvement in estimation accuracy. It is thus a significant step towards efficiently estimating a molecular structure directly from a noisy cryo-EM datasets [6].

2. MATHEMATICAL FRAMEWORK

2.1. Image model

We consider an image $f : \mathbb{R}^2 \to \mathbb{R}$, which is supported on the unit disk $D = \{\vec{x} \in \mathbb{R}^2 : |\vec{x}| \leq 1\}$. We assume that f has a finite expansion in the basis of Dirichlet Laplacian eigenfunctions; this is a standard assumption in the literature [4, 18, 19], which is akin to assuming that the image is bandlimited. This implies that we can expand the image (see [3]) as

$$f(r,\theta) = \sum_{(\nu,q):\lambda_{\nu,q}\leq\lambda} \alpha_{\nu,q}\psi_{\nu,q}(r,\theta), \quad \text{for } r \leq 1, \tag{3}$$

in polar coordinates (r, θ), where λ is called the bandlimit frequency. The expansion coefficients are denoted by $\alpha_{\nu,q}$, and

$$\psi_{\nu,q}(r,\theta) = J_\nu\left(\lambda_{\nu,q} r\right) e^{i\nu\theta}, \quad \text{for } r \leq 1, \tag{4}$$

where $\nu \in \mathbb{Z}_{\geq 0}$, J_ν is the ν-th order Bessel function of the first kind, and $\lambda_{\nu,q} > 0$ is the q-th positive root of J_ν, where $\lambda_{\nu,q} = \lambda_{-\nu,q}$. This expansion is known as the Fourier-Bessel expansion. The number of required coefficients is given by the sampling criterion provided in [19, 20], which is the analog of the classical Nyquist sampling rate.

The basis of Dirichlet Laplacian eigenfunctions is steerable: rotating f is equivalent to modulating the expansion coefficients $\alpha_{\nu,q}$. Specifically, $f_\phi(r,\theta) := f(r,\theta+\phi)$ can be computed by multiplying each term in (3) by $e^{i\nu\phi}$:

$$f_\phi(r,\theta) = \sum_{(\nu,q):\lambda_{\nu,q}\leq\lambda} \alpha_{\nu,q}\psi_{\nu,q}(r,\theta)e^{i\nu\phi}. \tag{5}$$

As actual cryo-EM measurements are discretized on a Cartesian grid, we will focus on the analysis of a discrete version of f, under the assumption of point-wise sampling. The discrete image $F_\phi : \mathbb{Z}^2 \to \mathbb{R}$ is thus defined by

$$F_\phi[\vec{\ell}] = f_\phi(\vec{\ell}/n), \quad \text{for } \vec{\ell} \in \mathbb{Z}^2, \tag{6}$$

where n is a fixed integer that determines the sampling resolution. In our case, n is the radius of f, in pixels. Since f_ϕ is supported on the unit disk D, it follows that F_ϕ is supported on $\{\vec{\ell} \in \mathbb{Z}^2 : |\vec{\ell}| \leq n\}$.

Let $\Psi_{\nu,q} : \mathbb{Z}^2 \to \mathbb{C}$ be the discretization of the Dirichlet Laplacian eigenfunctions (4), $\Psi_{\nu,q}[\vec{\ell}] = \psi_{\nu,q}(\vec{\ell}/n)$. With this notation,

$$F_\phi[\vec{\ell}] = \sum_{(\nu,q):\lambda_{\nu,q}\leq\lambda} \alpha_{\nu,q}\Psi_{\nu,q}[\vec{\ell}]e^{i\nu\phi}. \tag{7}$$

2.2. Approximate expectation-maximization

Given the measurement M that follows the MTD model (1) in the well-separated case, the maximum marginal likelihood estimator (MMLE) for the vector of coefficients α, that represents the target image F, is the maximizer of the likelihood function $p(M|\alpha)$. Within the EM framework [16], the locations of the target images within the measurement are treated as random variables drawn from some distribution under the separation condition (2). The EM algorithm estimates the MMLE by iteratively applying the expectation (E) and maximization (M) steps. Specifically, given the current estimate α_k, the E-step constructs the expected log-likelihood function $Q(\alpha|\alpha_k) = \sum_{\vec{\ell}} p(\vec{\ell}|M,\alpha_k) \sum_\phi p(\phi|M,\alpha_k,\vec{\ell}) \log p(M,\vec{\ell},\phi|\alpha)$, where the summation is over all admissible configurations of locations and rotations. The estimate is then updated in the M-step by maximizing $Q(\alpha|\alpha_k)$ with respect to α. The major drawback of this approach is that the number of admissible configurations for the locations and rotations grows exponentially with the problem size. Therefore, the direct application of the EM algorithm is computationally intractable, even for very small measurements.

In our framework of the approximate EM algorithm, we first partition the measurement M into $N_d = (N/L)^2$ non-overlapping patches, each of size $L \times L$, where $L = 2n + 1$ is the diameter of the target image F. We denote the m^{th} patch by M_m. Overall, a rotated image can occur in $4L^2 - 1$ different ways when it is present in a patch. The separation condition (2) implies that each patch contains parts of at most one rotated image. The vector of coefficients is estimated by the maximizer of the approximate likelihood function

$$p(M_0, M_1, \ldots, M_{N_d-1}|\alpha) \approx \prod_{m=0}^{N_d-1} p(M_m|\alpha), \tag{8}$$

where we ignore the dependencies between patchs. Our approximate EM algorithm works by applying the EM algorithm to estimate the MMLE of (8), without any prior on the image.

Depending on the position of image occurrences, the patch M_m can be modeled by

$$M_m = CT_{\vec{l}_m} ZF_{\phi_m} + \varepsilon_m, \quad \varepsilon_m \sim \mathcal{N}(0, \sigma^2 I_{L\times L}). \tag{9}$$

Notably, even though the images' rotations in the measurement are continuously distributed, for the approximate EM scheme we must assume a discretization of the rotations, represented by the parameter K:

$$\phi_m \in \Phi := \left\{ k\frac{2\pi}{K} \right\}_{k=0}^{K-1}. \tag{10}$$

This is a drawback of the method relative to autocorrelation analysis; see Section 3.1. 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 $\vec{l}_m$ positions, that is

$$\begin{aligned} &T_{\vec{l}_m} ZF_{\phi_m}\left[\vec{\ell} := (i,j)\right] \\ &\quad = ZF_{\phi_m}\left[(i+\ell_{m_x}) \bmod 2L, (j+\ell_{m_y}) \bmod 2L\right], \end{aligned} \tag{11}$$

where $\vec{\ell}_m \in \mathbb{L} := \{0, 1, \ldots, 2L-1\}^2$, 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 the E-step, our algorithm constructs the expected log-likelihood function

$$Q(\alpha|\alpha_k) = \sum_{m=0}^{N_d-1} \sum_{\vec{\ell}\in\mathbb{L}} \sum_{\phi\in\Phi} p(\vec{\ell},\phi|M_m,\alpha_k) \log p(M_m,\vec{\ell},\phi|\alpha) \tag{12}$$

given the current coefficients estimate α_k, where

$$p(M_m|\vec{\ell},\phi,\alpha) \propto \exp\left(-\frac{\|M_m - CT_{\vec{l}}ZF_\phi\|_{\text{F}}^2}{2\sigma^2}\right), \tag{13}$$

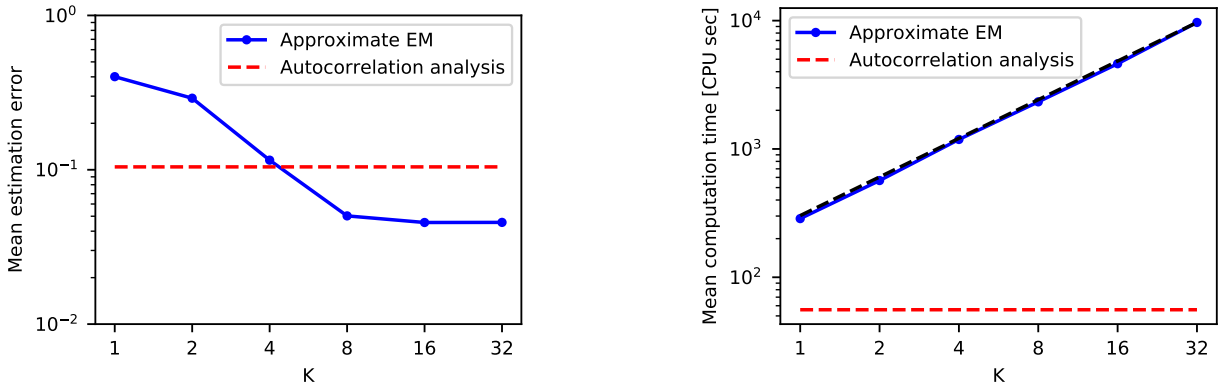


Fig. 2: The mean estimation error (left) and time (right) of recovering the vector of coefficients α, as a function of K, the number of assumed target images' possible rotations, by approximate EM. The recovery error and computation time using autocorrelation analysis is marked by the dashed horizontal red line. 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 computations per patch depends linearly on K.

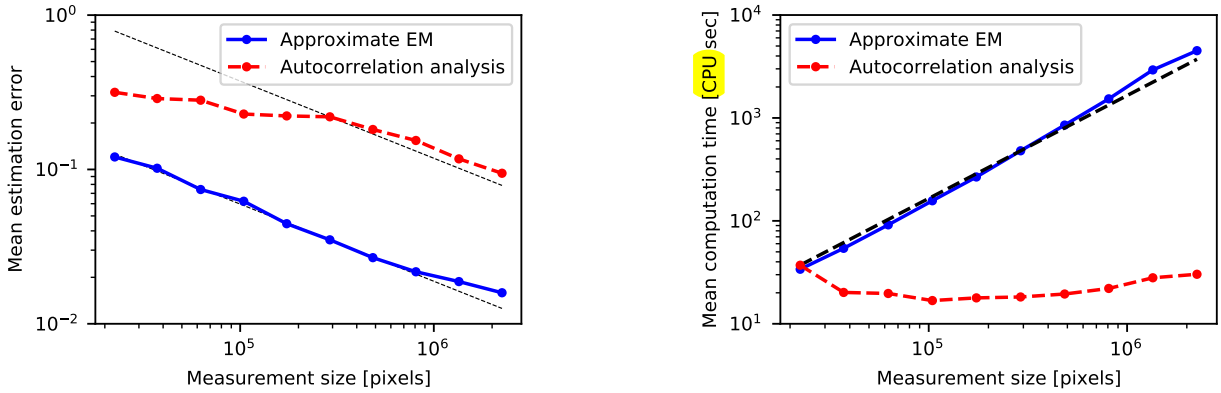


Fig. 3: The mean estimation error (left) and conputation time (right) of recovering the vector of coefficients α, as a function of measurement's size, by: (a) approximate EM; (b) autocorrelation analysis. For the estimation error, the black dashed lines illustrate a slope of $-1/2$, as predicted by the 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.

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

$$p(\vec{\ell},\phi|M_m,\alpha_k) = \frac{p(M_m|\vec{\ell},\phi,\alpha_k)p(\vec{\ell},\phi|\alpha_k)}{\sum_{\vec{\ell}'\in\mathbb{L}}\sum_{\phi'\in\Phi} p(M_m|\vec{\ell}',\phi',\alpha_k)p(\vec{\ell}',\phi'|\alpha_k)}, \tag{14}$$

which is the normalized likelihood function $p(M_m|\vec{\ell},\phi,\alpha_k)$, weighted by the prior distribution $p(\vec{\ell},\phi)$. In general, the prior distribution $p(\vec{\ell},\phi)$ is independent of the model α_k and can be estimated simultaneously with the image. Since the rotations and translations are independent, we have that

$$p(\vec{\ell},\phi|\alpha_k) = p(\vec{\ell},\phi) = p(\vec{\ell})p(\phi) := \rho[\vec{\ell}]\frac{1}{K}, \tag{15}$$

where we assume the rotations are uniformly distributed in the set Φ from (10), and $\rho[\vec{\ell}]$ is the distribution of image locations within the patch. This distribution is dependent on the density of the target images within the measurment.

We can rewrite (12) as (up to an irrelevant constant)

$$Q(\alpha,\rho|\alpha_k,\rho_k) = \sum_{m=0}^{N_d-1}\sum_{\vec{\ell}\in\mathbb{L}}\sum_{\phi\in\Phi} p(M_m|\vec{\ell},\phi,\alpha_k)\rho_k[\vec{\ell}] \times \left(\log p(M_m|\vec{\ell},\phi,\alpha) + \log\rho[\vec{\ell}]\right), \tag{16}$$

where $p(M_m|\vec{\ell},\phi,\alpha_k)$ and $p(M_m|\vec{\ell},\phi,\alpha)$ are given by (13).

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

$$\alpha_{k+1},\rho_{k+1} = \arg\max_{\alpha,\rho} Q(\alpha,\rho|\alpha_k,\rho_k) \text{ s.t. } \rho\in\Delta_{2L\times 2L}. \tag{17}$$

The constrained maximization of (17) can be achieved with the unconstrained maximization of the Lagrangian

$$\mathcal{L}(\alpha,\rho,\lambda) = Q(\alpha,\rho|\alpha_k,\rho_k) + \lambda\left(1-\sum_{\vec{\ell}\in\mathbb{L}}\rho[\vec{\ell}]\right), \tag{18}$$

where λ denotes the Lagrange multiplier. We note that the con-

straints in (17) involve the inequalities that the priors are non-negative. Such constrained maximization in general cannot be achieved by maximizing the Lagrangian, for the inequalities might be violated. As we will see later, however, these inequalities are automatically satisfied at the computed maximum (or local maximum) of the Lagrangian, which justifies this approach.

Since $Q(\alpha,\rho|\alpha_k,\rho_k)$ is additively separable for α and ρ, we maximize $\mathcal{L}(\alpha,\rho,\lambda)$ with respect to α and ρ separately. At the maximum of $\mathcal{L}(\alpha,\rho,\lambda)$, we have

$$0=\frac{\partial\mathcal{L}}{\partial(\alpha)_{\nu,q}}=\sum_{m=0}^{N_d-1}\sum_{\vec{\ell}\in\mathbb{L}}\sum_{\phi\in\Phi}p(M_m|\vec{\ell},\phi,\alpha_k)\rho_k[\vec{\ell}] \times\frac{\partial\log p(M_m|\vec{\ell},\phi,\alpha)}{\partial(\alpha)_{\nu,q}}, \quad (19)$$

resulting in a set of equations which is solved to achieve the updated α.

In order to update the priors, we maximize $\mathcal{L}(\alpha,\rho,\lambda)$ with respect to ρ:

$$0=\frac{\partial\mathcal{L}}{\rho[\vec{\ell}]}=\sum_{m=0}^{N_d-1}\sum_{\phi\in\Phi}p(M_m|\vec{\ell},\phi,\alpha_k)\rho_k[\vec{\ell}]\frac{1}{\rho[\vec{\ell}]}-\lambda, \quad (20)$$

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

$$\rho[\vec{\ell}]=\frac{1}{\lambda}\sum_{m=0}^{N_d-1}\sum_{\phi\in\Phi}p(M_m|\vec{\ell},\phi,\alpha_k)\rho_k[\vec{\ell}], \quad (21)$$

and we can immediately solve $\lambda = N_d$ 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 recovery procedure described in Section 2.2. As a baseline, we compare the results to the recovery achieved using autocorrelation analysis, based on the framework (and code) of [4] and [5]. To take the in-plane rotation symmetry into account, we measure the estimation error by

$$\text{relative error}_\alpha := \min_{\phi\in[0,2\pi)}\frac{\|\alpha^*-\alpha_\phi\|_2}{\|\alpha^*\|_2}, \quad (22)$$

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 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 images was drawn i.i.d. from a uniform distribution on $[0,1]$, and the target images were normalized such that $\|x\|_2=10$. Then, each image was normalized such that $\|F\|_{\text{F}}=10$, and expanded using it first 10 Fourier-Bessel coefficients. We try to estimate the vector of coefficients α from a varying number of initial guesses (that were drawn from the same distribution as the ground truth image) and $\gamma_{\text{init}}=0.03$. 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 discretization of rotations

Figure 2 presents recovery error and time as a function of as a function of K, the number of assumed target images' possible rotations, for measurements with SNR $=5$ and $N=1500$ pixels, and one random initial guess for α. The results are compared to the results using autocorrelation analysis. Remarkably, even for only $K=4$ assumed rotations in the EM framework we achieve estimation errors similar to the autocorrelation analysis scheme. As expected, the computation time grows linearly with the parameter K.

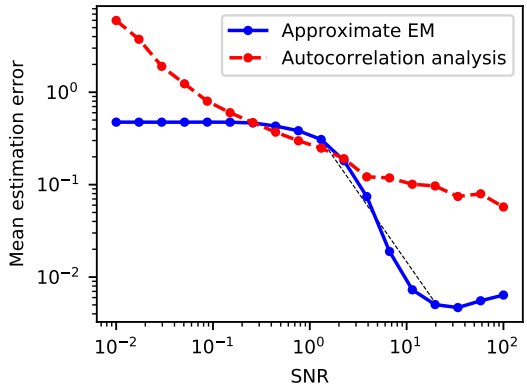


Fig. 4: The mean estimation error of recovering the vector of coefficients α, as a function of SNR, by: (a) approximate EM; (b) autocorrelation analysis.

3.2. Recovery error as a function of the measurement size

Figure 3 presents recovery error and time as a function of the measurement size, with SNR $=5$ and $K=16$, 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. This is the same estimation rate as if the translations and rotations were known (that is, the estimation rate of averaging over i.i.d. Gaussian variables). We achieve a significant improvement in recovery accuracy using approximate EM, even with a discretization of rotations. However, the computation time using the method is significantly larger, and grows linearly with the measurement's size.

3.3. Recovery error as a function of SNR

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 well-separated MTD problem, and comparing it numerically to the previously studied autocorrelation analysis. The numerical experiments suggest an improvement in estimation accuracy, together with an increase in required computational time.

In the future, we hope to extend the framework of this paper to deal with the more general case of cryo-EM—where the images are different tomographic projections of a 3-D molecular structure taken from unknown viewing directions [6]. A major drawback of the approximate EM method is the computation time, which is far greater than the computation time using autocorrelation analysis. Therefore, in order to achieve a computationally tractable algorithm for the 3-D case of the MTD problem, parallel processing and randomized algorithms, such as stochastic or online EM [21, 22, 23, 24, 25, 26], must be utilized.

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] 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.

[7] 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.

[8] Eva Nogales, "The development of cryo-EM into a mainstream structural biology technique," *Nature methods*, vol. 13, no. 1, pp. 24–27, 2016.

[9] 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.

[10] Joachim Frank, *Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state*, Oxford University Press, 2006.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

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

[16] 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.

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

[18] Zhizhen Zhao and Amit Singer, "Fourier–Bessel rotational invariant eigenimages," *JOSA A*, vol. 30, no. 5, pp. 871–877, 2013.

[19] Zhizhen Zhao, Yoel Shkolnisky, and Amit Singer, "Fast steerable principal component analysis," *IEEE Transactions on Computational Imaging*, vol. 2, no. 1, pp. 1–12, 2016.

[20] A Klug and RA Crowther, "Three-dimensional image reconstruction from the viewpoint of information theory," *Nature*, vol. 238, no. 5365, pp. 435–440, 1972.

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

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

[23] 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.

[24] 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.

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

[26] Daniel Quang and Xiaohui Xie, "EXTREME: an online EM algorithm for motif discovery," *Bioinformatics*, vol. 30, no. 12, pp. 1667–1673, 2014.