Multi-target Detection with an Arbitrary Spacing Distribution

Ti-Yen Lan, Tamir Bendory, Nicolas Boumal and Amit Singer

Abstract—**Motivated by the structure reconstruction problem in single-particle cryo-electron microscopy, we consider the multi-target detection model, where multiple copies of a target signal occur at unknown locations in a long measurement, further corrupted by additive Gaussian noise. At low noise levels, one can easily detect the signal occurrences and estimate the signal by averaging. However, in the presence of high noise, which is the focus of this paper, detection is impossible. Here, we propose two approaches—autocorrelation analysis and an approximate expectation maximization algorithm—to reconstruct the signal without the need to detect signal occurrences in the measurement. In particular, our methods apply to an arbitrary spacing distribution of signal occurrences. We demonstrate reconstructions with synthetic data and empirically show that the sample complexity of both methods scales as SNR^{-3} in the low SNR regime.**

Index Terms—**autocorrelation analysis, expectation maximization, frequency marching, cryo-EM, blind deconvolution.**

I. INTRODUCTION

We consider the multi-target detection (MTD) problem [10] to estimate a signal $x \in \mathbb{R}^L$ from a long, noisy measurement

$$y = s * x + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I_N), \quad (1)$$

where $y \in \mathbb{R}^N$ is the linear convolution of an unknown binary sequence $s \in \{0, 1\}^{N-L+1}$ with the signal, further corrupted by additive white Gaussian noise with zero mean and variance σ^2, and we assume $L \ll N$. Both x and s are treated as deterministic variables. The signal length L and the noise variance σ^2 are assumed to be known. The non-zero entries of s indicate the starting positions of the signal occurrences in y. We require the signal occurrences not to overlap, so consecutive non-zero entries of s are separated by at least L positions. Figure 1 gives an example of the measurement y that contains three signal occurrences at different noise levels.

MTD belongs to the wider class of blind deconvolution problems [25], which have applications in astronomy [24], [33], microscopy [28], [32], system identification [1], and motion deblurring [17], [26], among others. The main difference is that we are only interested in estimating the signal x: we treat s as a *nuisance variable*, while most of
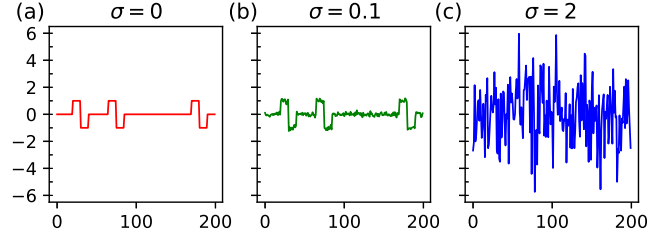
Fig. 1. An example of a measurement in the MTD model corrupted by additive Gaussian noise with (a) $\sigma = 0$, (b) $\sigma = 0.1$ (high SNR regime) and (c) $\sigma = 2$ (low SNR regime).

the literature on blind deconvolution aims to estimate both x and s. This distinction allows us to estimate the signal at higher noise levels, which are usually not addressed in the literature on blind deconvolution, specifically because s *cannot* be accurately estimated in such regimes. We give a theoretical argument for this claim below, and corroborate it with numerical experiments in Section IV.

This high-noise MTD model has been studied in [9], [10] under the assumption that the signal occurrences either are well separated or follow a Poisson distribution. In applications, however, the signal occurrences in the measurement might be arbitrarily close to each other. In this study, we extend the framework of [9], [10] to allow an arbitrary spacing distribution of signal occurrences by simultaneously estimating the signal and the *distance distribution* between consecutive signal occurrences.

The solution of the MTD problem is straightforward at high signal-to-noise ratio (SNR), such as the case shown in Figure 1(b). The signal can be estimated by first detecting the signal occurrences and then averaging. In the low SNR regime however, this simple method becomes problematic due to the difficulty of detection, as illustrated by Figure 1(c). More than difficulty, reliable detection becomes impossible beyond a critical noise level. This can be understood from [9, Proposition 3.1], which we reproduce here (the proof is based on the Neyman–Pearson lemma).

Proposition 1: Consider two known vectors $\theta_0 = x$ and $\theta_1 = 0$ in $\mathbb{R}^L$ and a random variable η taking value 0 or 1 with equal probability. We observe the random vector $X \in \mathbb{R}^L$ with the following distribution:

$$X \sim \begin{cases} \mathcal{N}(\theta_0, \sigma^2 I_L) & \text{if } \eta = 0 \\ \mathcal{N}(\theta_1, \sigma^2 I_L) & \text{if } \eta = 1. \end{cases}$$

For any deterministic estimator $\hat{\eta}$ of η,

$$\lim_{\sigma \to \infty} p(\hat{\eta} = \eta) \leq \frac{1}{2},$$

that is, the probability of success for the best deterministic estimator of η is no better than a random guess in the limit of $\sigma \to \infty$.

In our context, this simple fact can be interpreted as follows: even if we know the true signal x and an oracle provides windows of length L in the observation y which, with equal probability, contain either the noisy signal or pure noise, deciding whether a given window contains one or the other cannot be done significantly better than a random chance decision at high noise level. Results similar in spirit (but distinct in mathematical tools) are derived in [3].

In this work, we suggest to circumvent the estimation of s by estimating the signal x directly. Specifically, we propose two different approaches—autocorrelation analysis and an approximate version of the expectation maximization (EM) algorithm [18]. Since these methods do not require estimating s, Proposition 1 does not limit their performance in the low SNR regime.

Autocorrelation analysis relates the autocorrelations calculated from the noisy measurement to the signal. The signal is then estimated by fitting the autocorrelations through least squares. This approach is efficient as it requires only one pass over the data to calculate the autocorrelations; this is of particular importance as the data size grows.

In the second approach, the approximate EM algorithm, the signal is reconstructed by iteratively maximizing the data likelihood, which marginalizes over s; importantly, it does not estimate it explicitly. In contrast to autocorrelation analysis, the approximate EM algorithm scans through the whole dataset in each iteration, and hence requires much longer computational time.

In this study, we demonstrate the reconstruction of the underlying signal from the noisy measurement using the two proposed approaches. Our numerical experiments show that the approximate EM algorithm provides slightly more accurate estimates of the signal in the low SNR regime, whereas autocorrelation analysis is considerably faster, especially at low SNR. It is empirically shown that the sample complexity of both approaches scales as SNR^{-3} at low SNR, with details discussed later in the text. In the high SNR regime, the sample complexity of both methods scales as SNR^{-1}, the same as the sample complexity of the simple method that estimates the signal by first detecting the signal occurrences and then averaging.

The main contributions of this work are as follows.

1) We formalize MTD for an arbitrary signal spacing distribution, making it a more realistic model for applications.
2) We propose two algorithms to solve the MTD problem. In particular, our algorithm based on autocorrelations illustrates why, to recover x, we need not estimate all of s, but rather only a concise summary of it; and why it is possible, in principle at least, to solve this problem for arbitrary noise levels (given a sufficiently long observation). For our second algorithm, we note that the popular EM method is intractable, but we show how to implement an approximation of it, which performs well in practice.
3) For both algorithms, we design a coarse-to-fine multiresolution scheme to alleviate issues pertaining to nonconvexity. This is related to the ideas of frequency marching which are often used in cryo-electron microscopy (cryo-EM) [6], [34].

II. AUTOCORRELATION ANALYSIS

In what follows, we discuss autocorrelations of both the signal x (of length L) and the measurement y (of length N). To keep notation general, we here consider a sequence z of length m, and define its autocorrelations of order $q = 1, 2, \ldots$ for any integer shifts $l_1, l_2, \ldots, l_{q-1}$ as

$$a_z^q[l_1, l_2, \ldots, l_{q-1}] = \frac{1}{m} \sum_{i=0}^{m-1} z[i]z[i+l_1] \cdots z[i+l_{q-1}],$$

where z is zero-padded for indices out of the range $[0, m-1]$. We have $m = L$ when z represents the signal x and $m = N$ when z represents the measurement y. Since the autocorrelations only depend on the differences of the integer shifts and are invariant under any permutation of the shifts, for the second- and third-order autocorrelations we have the symmetries

$$a_z^2[l] = a_z^2[-l], \quad a_z^3[l_1, l_2] = a_z^3[l_2, l_1] = a_z^3[-l_1, l_2 - l_1]. \quad (2)$$

For applications of higher-order autocorrelations and their symmetries, see for example [4], [22], [31], [35].

In this section, we describe how the autocorrelations of the noisy measurement y are related to the underlying signal x. These relations are later used to estimate x without the need to identify the locations of signal occurrences, which are nuisance variables and difficult, if not impossible, to determine reliably at high noise levels. For completeness, we include a brief discussion of the special case where the signals are well separated and refer the reader to [9], [10] for details. The generalization to an arbitrary signal spacing distribution follows.

A. Well-separated signals

The signals are said to be well-separated when the consecutive non-zero entries of s are separated by at least $2L - 1$ positions. Under this condition, the autocorrelations of y with integer shifts $l_1, l_2, \ldots, l_{q-1}$ within the range $[-(L-1), L-1]$ are unaffected by the relative positions of signal occurrences. As a result, these autocorrelations of y provide fairly direct information about those of x, and therefore about x itself.

Due to the presence of Gaussian noise, the entries of y are stochastic. Taking the expectations of the first three autocorrelations of y with respect to the distribution of Gaussian noise, we obtain these relations (see Appendix A):

$$\mathbb{E}_\varepsilon\{a_y^1\} = \rho_0 a_x^1 \qquad (3)$$

$$\mathbb{E}_\varepsilon\{a_y^2[l]\} = \rho_0 a_x^2[l] + \sigma^2 \delta[l] \qquad (4)$$

$$\mathbb{E}_\varepsilon\{a_y^3[l_1, l_2]\} = \rho_0 a_x^3[l_1, l_2]$$
$$+ \rho_0 a_x^1 \sigma^2 (\delta[l_1] + \delta[l_2] + \delta[l_1 - l_2]), \quad (5)$$

where $0 \leq l < L$ and $0 \leq l_1 \leq l_2 < L$. Here, $\rho_0 = ML/N$ denotes the signal density, where M is the number of signal occurrences in y. We assume M grows with N at the constant rate ρ_0/L. The delta functions, defined by $\delta[0] = 1$ and $\delta[l \neq 0] = 0$, are due to the autocorrelations of the white Gaussian noise. As indicated by the studies in phase retrieval [7], [12], [15], in general a 1D signal x cannot be uniquely determined by its first two autocorrelations. It is thus necessary (and generically sufficient [10]) to include the third-order autocorrelations to uniquely determine x.

The expectations of the autocorrelations of y are estimated by averaging over the given noisy measurement. This data reduction requires only one pass over the data, which is a great computational advantage as the data size grows. As shown in Appendix A, for a given signal x, the average over the noisy entries of y gives the relations:

$$a_y^1 = \mathbb{E}_\varepsilon\{a_y^1\} + \eta^1$$
$$a_y^2[l] = \mathbb{E}_\varepsilon\{a_y^2[l]\} + \eta^2[l]$$
$$a_y^3[l_1, l_2] = \mathbb{E}_\varepsilon\{a_y^3[l_1, l_2]\} + \eta^3[l_1, l_2],$$

where $\eta^1, \eta^2[l]$ and $\eta^3[l_1, l_2]$ are random variables with zero mean and variances $\mathcal{O}(\sigma^2/N)$, $\mathcal{O}((\sigma^2+\sigma^4)/N)$ and $\mathcal{O}((\sigma^2+\sigma^6)/N)$ respectively. For autocorrelations of order q, the standard deviations scale as $\sigma^q/\sqrt{N}$ at high noise levels. Therefore, we need $\sqrt{N}/\sigma^q \gg 1$ in order for the a_y^q calculated from the noisy measurement to be a good estimator for $\mathbb{E}_\varepsilon\{a_y^q\}$, and thus to establish a reliable relation with the signal x such as (3)-(5). Since the SNR is proportional to σ^{-2}, the sample complexity therefore scales as SNR^{-q}. We also expect the error of the reconstructed signal to depend on the errors of the highest-order autocorrelations used in the analysis at high noise levels.

We estimate the signal density ρ_0 and signal x by fitting the first three autocorrelations of y via non-linear least squares:

$$f(\hat{\rho}_0, \hat{x}) = (a_y^1 - \hat{\rho}_0 a_{\hat{x}}^1)^2 + \frac{1}{L}\sum_{l=0}^{L-1}(a_y^2[l] - \hat{\rho}_0 a_{\hat{x}}^2[l] - \sigma^2\delta[l])^2$$

$$+ \frac{2}{L(L+1)}\sum_{l_2=0}^{L-1}\sum_{l_1=0}^{l_2}\left(a_y^3[l_1, l_2] - \hat{\rho}_0 a_{\hat{x}}^3[l_1, l_2]\right.$$

$$\left. - \hat{\rho}_0 a_{\hat{x}}^1 \sigma^2(\delta[l_1] + \delta[l_2] + \delta[l_1 - l_2])\right)^2. \quad (6)$$

The weights are chosen as the inverse of the number of terms in the respective sums. Since the autocorrelations have symmetries, as indicated in (2), the summations above are restricted to the non-redundant shifts. Due to the errors in estimating $\mathbb{E}_\varepsilon\{a_y^3\}$ with a_y^3, we expect the root-mean-square error of the reconstructed signal $\hat{x}$,

$$\text{RMSE}(\hat{x}) = \frac{||\hat{x} - x||_2}{||x||_2},$$

to scale as σ^3 in the low SNR regime.

We mention that there exist alternative methods to fit the observed autocorrelations. One possibility is to reformulate the problem as a tensor sensing problem: the autocorrelations are then linear functions of an $(L+1) \times (L+1) \times (L+1)$

tensor. Such an approach has been proven to be effective for the related problems of multi-reference alignment [5], [29] and phase retrieval [16]. Similar formulations are also found in tensor optimization [2], [30]. Nevertheless, this method requires lifting the original problem to a higher-dimensional space, which increases the computational burden and scales poorly with the dimension of the problem. In contrast, our least-squares formulation operates in the ambient dimension of the problem and thus might be applicable to high-dimensional setups.

B. Arbitrary spacing distribution

The condition of well-separated signals can be further relaxed to allow arbitrary spacing distribution by assuming that the signal occurrences in any subset of y follow the same spatial distribution. To this end, we define the pair separation function as follows.

Definition 1: For a given binary sequence s identifying M starting positions of signal occurrences (that is, $\sum_i s[i] = M$), the *pair separation function* $\xi[l]$ is defined as

$$\xi[l] = \frac{1}{M-1}\sum_{k=1}^{M-1}\delta[s_{k-1} + l - s_k], \quad (7)$$

where s_k indicates the index of the kth non-zero entry of the sequence s. It is not hard to see that $\sum_l \xi[l] = 1$.

In particular, we force $\xi[0] = \cdots = \xi[L-1] = 0$ since we exclude overlapping occurrences; then $\xi[L]$ is the fraction of pairs of consecutive signal occurrences that occur right next to each other (no spacing at all), $\xi[L+1]$ is the fraction of pairs of consecutive signal occurrences that occur with one signal-free entry in between them, etc.

In contrast to the well-separated model, autocorrelations of y may now involve correlating distinct occurrences of x, which may be in various relative positions. The crucial observation is that these autocorrelations depend only indirectly on s, namely, through $L-1$ entries of the unknown ξ, which have much smaller dimensions than s. As shown in Appendix A, for a given signal x, the autocorrelations of the given measurement y gives the relations:

$$a_y^1 = \mathbb{E}_\varepsilon\{a_y^1\} + \eta_1' = \rho_0 a_x^1 + \eta_1' \quad (8)$$

$$a_y^2[l] = \mathbb{E}_\varepsilon\{a_y^2[l]\} + \eta_2'[l]$$
$$= \rho_0 a_x^2[l] + \rho_0\sum_{j=L}^{L+l-1}\xi[j]a_x^2[j-l] + \sigma^2\delta[l] + \eta_2'[l] \quad (9)$$

$$a_y^3[l_1, l_2] = \mathbb{E}_\varepsilon\{a_y^3[l_1, l_2]\} + \eta_3'[l_1, l_2]$$
$$= \rho_0 a_x^3[l_1, l_2] + \rho_0\sum_{j=L}^{L+l_2-l_1-1}\xi[j]a_x^3[j-l_2, j+l_1-l_2]$$
$$+ \rho_0\sum_{j=L}^{L+l_1-1}\xi[j]a_x^3[l_2-l_1, j-l_1] + \rho_0 a_x^1 \sigma^2(\delta[l_1]$$
$$+ \delta[l_2] + \delta[l_1 - l_2]) + \eta_3'[l_1, l_2], \quad (10)$$

where $0 \leq l < L$ and $0 \leq l_1 \leq l_2 < L$. The random variables η_1', $\eta_2'[l]$ and $\eta_3'[l_1, l_2]$ have zero mean and variances $\mathcal{O}(\sigma^2/N)$, $\mathcal{O}((\sigma^2 + \sigma^4)/N)$ and $\mathcal{O}((\sigma^2 + \sigma^6)/N)$

respectively. Note that (8)–(10) reduce to (3)–(5) when $\xi[L] = \xi[L+1] = \cdots = \xi[2L-2] = 0$, as required by the condition of well-separated signals. Expressions (8)–(10) further simplify upon defining

$$\rho_1[i] = \rho_0 \xi[i+L], \quad i = 0, 1, \ldots, L-2. \qquad (11)$$

After calculating the first three autocorrelations of y from the noisy measurement, we estimate the signal x and the parameters ρ_0 and ρ_1 by fitting the autocorrelations of y through the non-linear least squares

$$f(\hat{x}, \hat{\rho}_0, \hat{\rho}_1) = (a_y^1 - \hat{\rho}_0 a_{\hat{x}}^1)^2$$
$$+ \frac{1}{L} \sum_{l=0}^{L-1} \left(a_y^2[l] - \hat{\rho}_0 a_{\hat{x}}^2[l] - \sigma^2 \delta[l] - \sum_{j=L}^{L+l-1} \hat{\rho}_1[j-L] a_{\hat{x}}^2[j-l] \right)^2$$
$$+ \frac{2}{L(L+1)} \sum_{l_2=0}^{L-1} \sum_{l_1=0}^{l_2} \left(a_y^3[l_1, l_2] - \hat{\rho}_0 a_{\hat{x}}^3[l_1, l_2] - \hat{\rho}_0 a_{\hat{x}}^1 \sigma^2 \big(\delta[l_1] \right.$$
$$+ \delta[l_2] + \delta[l_1 - l_2] \big) - \sum_{j=L}^{L+l_2-l_1-1} \hat{\rho}_1[j-L] a_{\hat{x}}^3[j-l_2, j+l_1-l_2]$$
$$- \left. \sum_{j=L}^{L+l_1-1} \hat{\rho}_1[j-L] a_{\hat{x}}^3[l_2-l_1, j-l_1] \right)^2. \qquad (12)$$

Numerically, we find that this problem can often be solved, meaning that, even though s cannot be estimated, we can still estimate x and the summarizing statistics ρ_0 and ρ_1. As discussed in Section II-A, the RMSE of the reconstructed signal is expected to scale as σ^3 in the low SNR regime owing to the errors in estimating $\mathbb{E}_\varepsilon\{a_y^3\}$ with a_y^3.

C. Frequency marching

To minimize the least squares in (6) and (12), we use the trust-regions method in Manopt [13] over the product of the Euclidean manifold with the constraints that $\hat{\rho}_0$ and $\hat{\rho}_1$ are positive. However, as the least squares problems are inherently non-convex, we observe that the iterates of the trust-regions method used for minimization are liable to stagnate in local minima. To alleviate this issue, we adopt the frequency marching scheme [6] in our optimization. The idea behind frequency marching is based on the following heuristics:

1) The coarse-grained (low-resolution) version of the original problem has a smoother optimization landscape so that it is empirically more likely for the iterates of the optimization algorithm to reach the global optimum of the coarse-grained problem.
2) Intuitively, the global optimum of the original problem can be reached more easily by following the path along the global optima of a series of coarse-grained problems with incremental resolution.

Our goal is to guide the iterates of the optimization algorithm to reach the global optimum of the original problem by successively solving the coarse-grained problems, which are warm-started with the solution from the previous stage.

The coarse-grained problems are characterized by the order of the Fourier series, $n_{\max} = 1, 2, \ldots, \lfloor L/2 \rfloor$, used to express the low-resolution approximate $x^{(n_{\max})}$ by

$$x^{(n_{\max})}[l] = c_0 + \sum_{n=1}^{n_{\max}} c_n \cos\left(\frac{2\pi nl}{L}\right) + d_n \sin\left(\frac{2\pi nl}{L}\right), \quad (13)$$

where $l = 0, 1, \ldots, L-1$. Instead of the entries of x, the least squares are minimized with respect to the Fourier coefficients in our frequency marching scheme. The order $n_{\max}$ is related to the spatial resolution by Nyquist rate:

$$\Delta x = \begin{cases} 1 & \text{if } n_{\max} = (L-1)/2, \\ L/2n_{\max} & \text{otherwise.} \end{cases}$$

The spatial resolution subdivides the signal x into $L' = L/\Delta x$ units and represents the "step size" of the shifts for the coarse-grained autocorrelations.

We define the coarse-grained autocorrelations of $x^{(n_{\max})}$ of order q for integer shifts $l_1, l_2, \ldots, l_{q-1}$ as

$$b_{x^{(n_{\max})}}^q[l_1, l_2, \ldots, l_{q-1}]$$
$$= a_{x^{(n_{\max})}}^q[\lfloor l_1 \Delta x \rceil, \lfloor l_2 \Delta x \rceil, \ldots, \lfloor l_{q-1} \Delta x \rceil],$$

where $\lfloor \cdot \rceil$ rounds the argument to the nearest integer. The coarse-grained autocorrelations of y are given by sub-sampling the original autocorrelations calculated from the full measurement. With $b[l]$ denoting the bin centered at $l\Delta x$, where $l = 0, 1, \ldots, L' - 1$, we estimate the coarse-grained autocorrelations of y by

$$b_y^1 = a_y^1$$
$$b_y^2[l] = B_2^{-1} \sum_{i \in b[l]} \left(a_y^2[i] - \sigma^2 \delta[i] \right)$$
$$b_y^3[l_1, l_2] = B_3^{-1} \sum_{\substack{i_1 \in b[l_1] \\ i_2 \in b[l_2] \\ i_1 \leq i_2}} \left(a_y^3[i_1, i_2] \right.$$
$$\left. - a_y^1 \sigma^2 \big(\delta[i_1] + \delta[i_2] + \delta[i_1 - i_2] \big) \right),$$

where $0 \leq l < L'$, $0 \leq l_1 \leq l_2 < L'$, and B_2 and B_3 represent the number of terms in the respective sums.

Following the discussion in Section II-B, we relate the first three autocorrelations of y to those of $x^{(n_{\max})}$, as defined in (13), by

$$b_y^1 \approx \rho_0 b_{x^{(n_{\max})}}^1$$
$$b_y^2[l] \approx \rho_0 b_{x^{(n_{\max})}}^2[l] + \sum_{j=L'}^{L'+l-1} \rho_1^{(n_{\max})}[j-L'] b_{x^{(n_{\max})}}^2[j-l]$$
$$b_y^3[l_1, l_2]\} \approx \rho_0 b_{x^{(n_{\max})}}^3[l_1, l_2]$$
$$+ \sum_{j=L'}^{L'+l_2-l_1-1} \rho_1^{(n_{\max})}[j-L'] b_{x^{(n_{\max})}}^3[j-l_2, j+l_1-l_2]$$
$$+ \sum_{j=L'}^{L'+l_1-1} \rho_1^{(n_{\max})}[j-L'] b_{x^{(n_{\max})}}^3[l_2-l_1, j-l_1].$$

The autocorrelations are related by approximation instead of equality to reflect the errors due to the low-resolution

approximation. Above, we define $\rho_1^{(n_{\max})}$ as the product of the signal density ρ_0 and the coarse-grained pair separation function:

$$\rho_1^{(n_{\max})}[i] = \rho_0 \xi^{(n_{\max})}[i + L'], \quad i = 0, 1, \ldots, L' - 2, \quad (14)$$

where $\xi^{(n_{\max})}$ is defined as

$$\xi^{(n_{\max})}[i] = \sum_{l=\max\{0, \lfloor (i-1/2)\Delta x \rfloor\}}^{\lfloor (i+1/2)\Delta x \rfloor - 1} \xi[l]. \quad (15)$$

In each stage of frequency marching, we estimate the Fourier coefficients, which are related to $x^{(n_{\max})}$ through (13), and the parameters $\rho_0, \rho_1^{(n_{\max})}$ by fitting the coarse-grained autocorrelations of y through the non-linear least squares

$$f(\{\hat{c}_0, \ldots, \hat{c}_{n_{\max}}\}, \{\hat{d}_1, \ldots, \hat{d}_{n_{\max}}\}, \hat{\rho}_0, \hat{\rho}_1^{(n_{\max})})$$
$$= (b_y^1 - \hat{\rho}_0 b_{\hat{x}^{(n_{\max})}}^1)^2 + \frac{1}{L'} \sum_{l=0}^{L'-1} \left(b_y^2[l] - \hat{\rho}_0 b_{\hat{x}^{(n_{\max})}}^2[l] \right.$$
$$\left. - \sum_{j=L'}^{L'+l-1} \hat{\rho}_1^{(n_{\max})}[j - L'] b_{\hat{x}^{(n_{\max})}}^2[j - l] \right)^2$$
$$+ \frac{2}{L'(L'+1)} \sum_{l_2=0}^{L'-1} \sum_{l_1=0}^{l_2} \left(b_y^3[l_1, l_2] - \hat{\rho}_0 b_{\hat{x}^{(n_{\max})}}^3[l_1, l_2] \right.$$
$$- \sum_{j=L'}^{L'+l_2-l_1-1} \hat{\rho}_1^{(n_{\max})}[j - L'] b_{\hat{x}^{(n_{\max})}}^3[j - l_2, j + l_1 - l_2]$$
$$\left. - \sum_{j=L'}^{L'+l_1-1} \hat{\rho}_1^{(n_{\max})}[j - L'] b_{\hat{x}^{(n_{\max})}}^3[l_2 - l_1, j - l_1] \right)^2.$$

Our frequency marching scheme increments the order $n_{\max}$ from 1 to $\lfloor L/2 \rfloor$, and the computed solution of each stage is used to initialize optimization in the next stage.

III. EXPECTATION MAXIMIZATION

In this section, as an alternative to the autocorrelations approach, we describe an approximate EM algorithm to address both the cases of well-separated signals and arbitrary spacing distribution. A frequency marching scheme is also designed to help the iterates of the EM algorithm converge to the global maximum of the data likelihood.

A. Well-separated signals

Given the measurement y that follows the MTD model (1), the maximum marginal likelihood estimator (MMLE) for the signal x is the maximizer of the likelihood function $p(y|x)$. Within the EM framework [18], the nuisance variable s is treated as a random variable drawn from some distribution under the condition of non-overlapping signal occurrences. The EM algorithm estimates the MMLE by iteratively applying the expectation (E) and maximization (M) steps. Specifically, given the current signal estimate x_k, the E-step constructs the expected log-likelihood function

$$Q(x|x_k) = \sum_s p(s|y, x_k) \log p(y, s|x),$$

where the summation runs over all admissible configurations of the binary sequence s. The signal estimate is then updated in the M-step by maximizing $Q(x|x_k)$ with respect to x. The major drawback of this approach is that the number of admissible configurations for s grows exponentially with the problem size. Therefore, the direct application of the EM algorithm is computationally intractable, even for very short measurements.

In our framework of the approximate EM algorithm, we first partition the measurement y into $N_d = N/L$ non-overlapping segments, each of length L, and denote the mth segment by y_m. Overall, the signal can occur in $2L - 1$ different ways when it is present in a segment. The signal is estimated by the maximizer of the approximate likelihood function

$$p(y_0, y_1, \ldots, y_{N_d-1}|x) \approx \prod_{m=0}^{N_d-1} p(y_m|x), \quad (16)$$

where we ignore the dependencies between segments. Our approximate EM algorithm works by applying the EM algorithm to estimate the MMLE of (16), without any prior on the signal. As we will see in Section IV, the validity of the approximation is corroborated by the results of our numerical experiments.

Depending on the position of signal occurrences, the segment y_m can be modeled by

$$y_m = CR_{l_m} Zx + \varepsilon_m, \quad \varepsilon_m \sim \mathcal{N}(0, \sigma^2 I_L).$$

Here, Z first zero-pads L entries to the left of x, and R_{l_m} circularly shifts the zero-padded sequence by l_m positions, that is,

$$(R_{l_m} Zx)[l] = (Zx)[(l + l_m) \bmod 2L],$$

where $l_m = 0, 1, \ldots, 2L - 1$ and is treated as a random variable. The operator C then crops the first L entries of the circularly shifted sequence, which are further corrupted by additive white Gaussian noise. In this generative model, $l_m = 0$ represents no signal occurrence in y_m, and $l_m = 1, \ldots, 2L-1$ enumerate the $2L - 1$ different ways a signal can appear in a segment.

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

$$Q(x|x_k) = \sum_{m=0}^{N_d-1} \sum_{l=0}^{2L-1} p(l|y_m, x_k) \log p(y_m, l|x) \quad (17)$$

given the current signal estimate x_k, where

$$p(y_m|l, x) \propto \prod_{i=0}^{L-1} \exp\left(-\frac{(y_m[i] - (CR_l Zx)[i])^2}{2\sigma^2} \right), \quad (18)$$

with the normalization $\sum_{l=0}^{2L-1} p(y_m|l, x) = 1$. From Bayes' rule, we have

$$p(l|y_m, x_k) = \frac{p(y_m|l, x_k)p(l|x_k)}{\sum_{l=0}^{2L-1} p(y_m|l, x_k)p(l|x_k)}, \quad (19)$$

which is the normalized likelihood function $p(y_m|l, x_k)$, weighted by the prior distribution $p(l|x_k)$. In general, the prior distribution $p(l|x_k)$ is independent of the model x_k and can be estimated simultaneously with the signal.

Denoting the prior distribution $p(l)$ by $\alpha[l]$, we rewrite (17) as (up to an irrelevant constant)

$$Q(x,\alpha|x_k,\alpha_k) = \sum_{m=0}^{N_d-1} \sum_{l=0}^{2L-1} p(l|y_m,x_k)$$
$$\times \left(\log p(y_m|l,x) + \log \alpha[l] \right).$$

We note that the dependence of $Q(x,\alpha|x_k,\alpha_k)$ on the current prior estimates α_k lies in $p(l|y_m,x_k)$ through (19). The M-step updates the signal estimate and the priors by maximizing $Q(x,\alpha|x_k,\alpha_k)$ under the constraint that the priors lie on the simplex Δ_{2L}:

$$x_{k+1},\alpha_{k+1} = \arg\max_{x,\alpha} Q(x,\alpha|x_k,\alpha_k) \text{ s.t. } \alpha \in \Delta_{2L}. \quad (20)$$

As shown in Appendix B, we obtain the update rules

$$x_{k+1}[j] = \frac{\sum_{m=0}^{N_d-1} \sum_{l=j+1}^{j+L} p(l|y_m,x_k)y_m[j+L-l]}{\sum_{m=0}^{N_d-1} \sum_{l=j+1}^{j+L} p(l|y_m,x_k)}, \quad (21)$$

where $0 \le j < L$, and

$$\alpha_{k+1}[l] = \frac{1}{N_d} \sum_{m=0}^{N_d-1} p(l|y_m,x_k), \quad (22)$$

where $0 \le l < 2L$. We repeat the iterations of the EM algorithm until the estimates stop improving, as judged within some tolerance.

B. Arbitrary spacing distribution

We extend the approximate EM approach to address arbitrary spacing distribution of signal occurrences by reformulating the probability model: each segment y_m can now contain up to two signal occurrences. In this case, the two signals can appear in $L(L-1)/2$ different combinations in a segment, which is explicitly modeled by

$$y_m = CR_{l_1^m}Zx + CR_{l_2^m}Zx + \varepsilon_m, \ \ \varepsilon_m \sim \mathcal{N}(0,\sigma^2 I_L),$$

where $L < l_1^m < 2L$ and $0 < l_2^m \le l_1^m - L$. Given the signal estimate x_k and the shifts $l_1^m = l_1, l_2^m = l_2$, the likelihood function $p(y_m|l_1,l_2,x_k)$ can be written as

$$p(y_m|l_1,l_2,x_k) \propto \prod_{i=0}^{L-1} \exp\left(-\frac{1}{2\sigma^2}\right.$$
$$\left. \times \left[y_m[i] - (CR_{l_1}Zx_k)[i] - (CR_{l_2}Zx_k)[i]\right]^2\right), \quad (23)$$

and we have the normalization condition

$$\sum_{l=0}^{2L-1} p(y_m|l,x_k) + \sum_{l_1=L+1}^{2L-1} \sum_{l_2=1}^{l_1-L} p(y_m|l_1,l_2,x_k) = 1.$$

Incorporating the terms with two signal occurrences, the E-step constructs the expected log-likelihood function as

$$Q(x,\alpha|x_k,\alpha_k)$$
$$= \sum_{m=0}^{N_d-1} \left[\sum_{l=0}^{2L-1} p(l|y_m,x_k)\left(\log p(y_m|l,x) + \log\alpha[l] \right)\right.$$
$$+ \sum_{l_1=L+1}^{2L-1} \sum_{l_2=1}^{l_1-L} p(l_1,l_2|y_m,x_k)$$
$$\left. \times \left(\log p(y_m|l_1,l_2,x) + \log\alpha[l_1,l_2] \right)\right], \quad (24)$$

where the prior $p(l_1,l_2)$ is denoted by $\alpha[l_1,l_2]$.

Under the assumption that the signal occurrences in any subset of y follow the same spatial distribution, we can parametrize the priors with the pair separation function ξ (see Definition 1). Recall that $(M-1)\xi[l]$ is the number of pairs of consecutive signal occurrences whose starting positions are separated by exactly l positions. The priors $\alpha[l_1,l_2]$ can be related to the probability that two signal occurrences appear in the combination specified by (l_1,l_2) in a segment of length L selected from the measurement y, which is estimated by

$$\alpha[l_1,l_2] = \frac{(M-1)\xi[l_1-l_2]}{N-L+1} \approx \frac{M}{N}\xi[l_1-l_2], \quad (25)$$

where $L < l_1 < 2L$ and $0 < l_2 \le l_1 - L$. Here, $(M-1)\xi[l_1 - l_2]$ is the number of segments that realize the configuration of signal occurrences specified by (l_1,l_2), and $N-L+1$ indicates the total number of segment choices. The priors $\alpha[l]$ can similarly be related to the probability that a signal occurs in the way specified by l in a segment of length L:

$$\alpha[l] = \alpha[2L-l]$$
$$= \frac{(M-1)\xi[2L-l] + (M-1)\xi[2L-l+1] + \cdots}{N-L+1}$$
$$\approx \frac{M}{N} \sum_{j=2L-l}^{\infty} \xi[j], \quad (26)$$

where $0 < l \le L$. An interesting observation is that the number of signal occurrences M can be estimated by

$$\alpha[L] \approx \frac{M}{N} \sum_{j=L}^{\infty} \xi[j] = \frac{M}{N},$$

since the signal occurrences are required not to overlap. The value of the prior $\alpha[0]$ is determined by the normalization

$$\sum_{l=0}^{2L-1} \alpha[l] + \sum_{l_1=L+1}^{2L-1} \sum_{l_2=1}^{l_1-L} \alpha[l_1,l_2] = 1. \quad (27)$$

From (25), (26) and (27), we see that the priors are uniquely specified by the positive parameters $\alpha[0]$, $\alpha[1]$, $\rho_1[0]$, $\rho_1[1]$, $\ldots, \rho_1[L-2]$, with ρ_1 defined in (11). Therefore, the normalization (27) can be rewritten as

$$\alpha[0] + (2L-1)\alpha[1] + \sum_{i=0}^{L-2} \frac{i+L}{L}\rho_1[i] = 1. \quad (28)$$

In the special case of well-separated signals, where $\xi[i] = 0$ for $L \leq i \leq 2L - 2$, we have $\alpha[1] = \alpha[2] = \cdots = \alpha[2L-1]$ and the normalization $\alpha[0] + (2L-1)\alpha[1] = 1$.

In the M-step, we update the signal estimate and the parameters $\alpha[0], \alpha[1], \rho_1$ by maximizing $Q(x, \alpha|x_k, \alpha_k)$ under the constraint that (28) is satisfied:

$$
\begin{aligned}
&x_{k+1}, \alpha_{k+1}[0], \alpha_{k+1}[1], \{\rho_1\}_{k+1} \\
&= \arg \max_{x, \alpha[0], \alpha[1], \rho_1} Q(x, \alpha|x_k, \alpha_k) \text{ s.t. (28) is satisfied.} \quad (29)
\end{aligned}
$$

As shown in (24), $Q(x, \alpha|x_k, \alpha_k)$ is additively separable for x and α, so the constrained maximization (29) can be reached by maximizing $Q(x, \alpha|x_k, \alpha_k)$ with respect to x and the parameters $\alpha[0], \alpha[1], \rho_1$ separately. Maximizing $Q(x, \alpha|x_k, \alpha_k)$ with respect to x yields the update rule (see Appendix C for derivation)

$$
\begin{aligned}
&x_{k+1}[j] \\
&= \Bigg[\sum_{m=0}^{N_d-1} \Bigg(\sum_{l=j+1}^{j+L} p(l|y_m, x_k) y_m[j + L - l] \\
&\quad + \sum_{l_1=L+1}^{j+L} \sum_{l_2=1}^{l_1-L} p(l_1, l_2|y_m, x_k) y_m[j + L - l_1] \\
&\quad + \sum_{l_1=L+j+1}^{2L-1} \sum_{l_2=j+1}^{l_1-L} p(l_1, l_2|y_m, x_k) y_m[j + L - l_2] \Bigg) \Bigg] \\
&\quad \times \Bigg[\sum_{m=0}^{N_d-1} \Bigg(\sum_{l=j+1}^{j+L} p(l|y_m, x_k) + \sum_{l_1=L+1}^{j+L} \sum_{l_2=1}^{l_1-L} p(l_1, l_2|y_m, x_k) \\
&\quad + \sum_{l_1=L+j+1}^{2L-1} \sum_{l_2=j+1}^{l_1-L} p(l_1, l_2|y_m, x_k) \Bigg) \Bigg]^{-1}, \quad (30)
\end{aligned}
$$

where $0 \leq j < L$. To update the parameters $\alpha[0], \alpha[1], \rho_1$, we note that the function $Q(x, \alpha|x_k, \alpha_k)$ is concave with respect to these parameters and the constraint (28) forms a compact convex set. Therefore, the constrained maximization is achieved using the Frank-Wolfe algorithm [20].

C. Frequency marching

Because the iterates of the EM algorithm are not guaranteed to converge to the maximum likelihood, we develop a frequency marching scheme to help the iterates converge to the global maximum. Recall that frequency marching converts the original optimization problem into a series of coarse-grained problems with gradually increasing resolution. The coarse-grained version of the EM algorithm is characterized by the spatial resolution

$$
\Delta x = \max(1, L/2n_{\max}),
$$

where $n_{\max} = 1, 2, \ldots, \lfloor (L+1)/2 \rfloor$. This spatial resolution models the signal shifts in the coarse-grained problem, with

the corresponding expected likelihood function given by

$$
\begin{aligned}
&Q^{(n_{\max})}(x, \alpha^{(n_{\max})}|x_k, \alpha_k^{(n_{\max})}) \\
&= \sum_{m=0}^{N_d-1} \Bigg[\sum_{l=0}^{2L'-1} p(\lfloor l\Delta_x \rceil|y_m, x_k) \Bigg(\log p(y_m|\lfloor l\Delta_x \rceil, x) \\
&\quad + \log \alpha^{(n_{\max})}[l] \Bigg) + \sum_{l_1=L'+1}^{2L'-1} \sum_{l_2=1}^{l_1-L'} p(\lfloor l_1\Delta_x \rceil, \lfloor l_2\Delta_x \rceil|y_m, x_k) \\
&\quad \times \Bigg(\log p(y_m|\lfloor l_1\Delta_x \rceil, \lfloor l_2\Delta_x \rceil, x) + \log \alpha^{(n_{\max})}[l_1, l_2] \Bigg) \Bigg],
\end{aligned}
$$

where $L' = L/\Delta_x$. This likelihood function is equal to the original one shown in (24) restricted to the rounded shifts, so the signal can be updated with (30) by ignoring the terms with irrelevant shifts.

Mimicking (25) and (26), we construct the expressions for the priors in the coarse-grained problem as

$$
\alpha^{(n_{\max})}[l_1, l_2] \approx \frac{M}{N} \xi^{(n_{\max})}[l_1 - l_2]
$$

$$
\alpha^{(n_{\max})}[l] = \alpha^{(n_{\max})}[2L' - l] \approx \frac{M}{N} \sum_{j=2L'-l}^{\infty} \xi^{(n_{\max})}[j],
$$

where $L' < l_1 < 2L'$, $0 < l_2 \leq l_1 - L'$, $0 < l \leq L'$, and $\xi^{(n_{\max})}$ is defined in (15). Since the priors are uniquely specified by the positive parameters $\alpha^{(n_{\max})}[0]$, $\alpha^{(n_{\max})}[1]$, $\rho_1^{(n_{\max})}[0]$, $\rho_1^{(n_{\max})}[1]$, $\ldots, \rho_1^{(n_{\max})}[L' - 2]$, with $\rho_1^{(n_{\max})}$ defined in (14), we can update the priors by maximizing $Q^{(n_{\max})}(x, \alpha^{(n_{\max})}|x_k, \alpha_k^{(n_{\max})})$ with respect to these parameters under the constraint that the parameters lie on the simplex defined by

$$
\alpha^{(n_{\max})}[0] + (2L' - 1)\alpha^{(n_{\max})}[1] + \sum_{i=0}^{L'-2} \frac{i + L'}{L'} \rho_1^{(n_{\max})}[i] = 1.
$$

Incrementing from $n_{\max} = 1$ to $n_{\max} = \lfloor (L+1)/2 \rfloor$, the estimated signal and priors in each stage of frequency marching are used to initialize the optimization problem in the next stage.

IV. NUMERICAL EXPERIMENTS

This section describes the construction of our synthetic data and the performance of the two proposed methods.[1] A measurement is generated by first sampling M integers from $[0, N - L]$, with the constraint that any two samples differ by at least $L + W$ entries, for some integer W. We use $W = L - 1$ for the case of well-separated signals, while $W = 0$ is chosen to test our methods for arbitrary spacing distribution of signal occurrences. The sampling is done by generating the random integers uniformly one by one, rejecting any integer that would violate the constraint with respect to previously accepted samples. The M integers indicate the starting positions of the signal occurrences, which are recorded as the non-zero entries of the binary sequence s. The noisy measurement y is given by the linear convolution of s with the signal x, with

---

[1] The code for all experiments is publicly available at https://github.com/tl578/multi-target-detection.
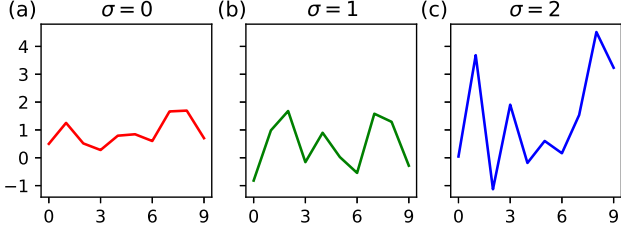
Fig. 2. (a) The signal used to generate our synthetic data, with the ℓ_2 norm scaled to $\sqrt{L}$. (b), (c) Signals corrupted by additive white Gaussian noise with zero mean and standard deviations $\sigma = 1$ and $\sigma = 2$.

each entry of y further corrupted by additive white Gaussian noise of zero mean and variance σ^2. Our synthetic data all have length $N = 10^6$, and are constructed with the signal shown in Figure 2(a), which has length $L = 10$. Also shown in Figures 2(b) and 2(c) are examples of the corrupted signals by different levels of noise. The number of signal occurrences M is adjusted so that the signal density $\rho_0 = ML/N$ is equal to 0.3 and 0.5 for the cases of well-separated signals and arbitrary spacing distribution respectively. Figure 3 shows the resulting pair separation function for our test case of arbitrary spacing distribution.
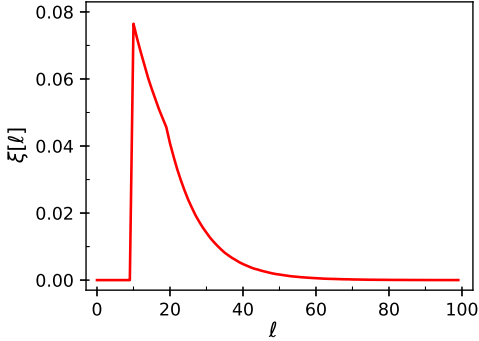

Fig. 3. Pair separation function to test arbitrary signal spacing distributions.

For both cases of signal spacing distribution, we reconstruct the signal from synthetic data using the two proposed methods with the frequency marching scheme over a wide range of σ. When the frequency marching scheme is not implemented, we observe that the algorithms usually suffer from large errors due to the non-convexity of the problems. A total of 20 instances of synthetic data are generated for each value of σ, and each instance is solved by the two methods using the frequency marching schemes. For our autocorrelation analysis methods, we run the optimization with 10 different random initializations and retain the one that minimizes the cost function, as defined in (6) or (12). For the EM approach, we also use 10 random initializations and retain the solution with the maximum data likelihood. Given the signal estimate x_k, we approximate the data likelihood for the cases of well-separated signals and arbitrary spacing distribution by

$$p(y|x_k) \approx \prod_{m=0}^{N_d-1} \sum_{l=0}^{2L-1} p(y_m|l, x_k)\alpha[l]$$

and

$$p(y|x_k) \approx \prod_{m=0}^{N_d-1} \left(\sum_{l=0}^{2L-1} p(y_m|l, x_k)\alpha[l] + \sum_{l_1=L+1}^{2L-1} \sum_{l_2=1}^{l_1-L} p(y_m|l_1, l_2, x_k)\alpha[l_1, l_2] \right)$$

respectively. This is the same approximation we use to formulate the EM algorithm.
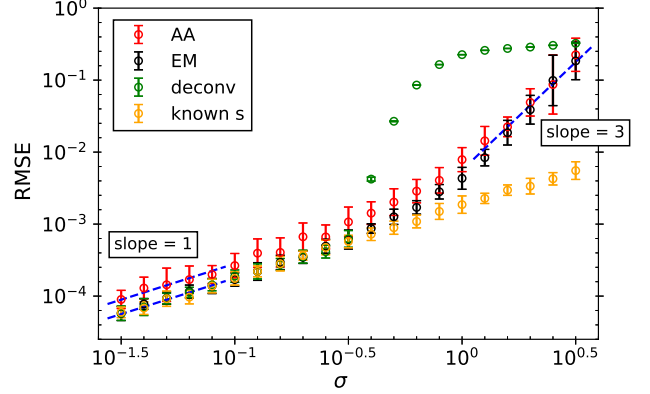

Fig. 4. The RMSEs of the reconstructed signal from data generated with well-separated occurrences. We show performance of our autocorrelation analysis (AA) and of expectation-maximization (EM). Also shown are the RMSEs of the signal estimated by the oracle-based deconvolution algorithm 'deconv' described in Section IV and the RMSEs of the estimated signal when the binary sequence s is known.

The RMSEs of the reconstructions for the case of well-separated signals are shown in Figure 4. As predicted in Section II-A, the RMSE for autocorrelation analysis scales in proportion to σ at high SNR, and to σ^3 at low SNR. Interestingly, the EM algorithm exhibits the same behavior, which empirically shows that the two approaches share the same scaling of sample complexity in the two extremes of noise levels. Figure 5(a) and 5(b) show an instance of the signals reconstructed from data generated with $\sigma = 10^{0.3}$ and $\sigma = 10^{0.5}$ respectively. The same scaling is not observed for the reconstructed signal density ρ_0 (not shown), although the relative errors are generally well below a few percents even in the noisiest cases.
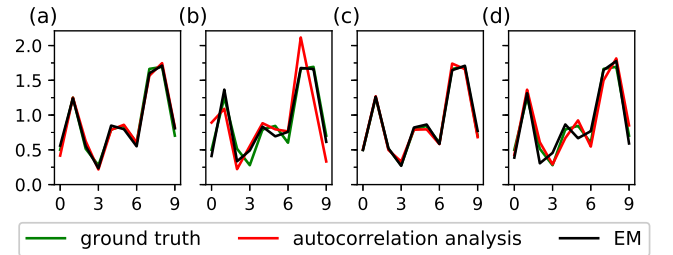

Fig. 5. (a), (b) An instance of the reconstructed signals from data generated with well-separated occurrences for $\sigma = 10^{0.3}$ and $\sigma = 10^{0.5}$ respectively. (c), (d) An instance of the reconstructed signals from data generated with an arbitrary spacing distribution for $\sigma = 10^{0.3}$ and $\sigma = 10^{0.5}$ respectively.

Figure 6 shows the RMSEs of the reconstructed signal for an arbitrary spacing distribution displayed in Figure 3.
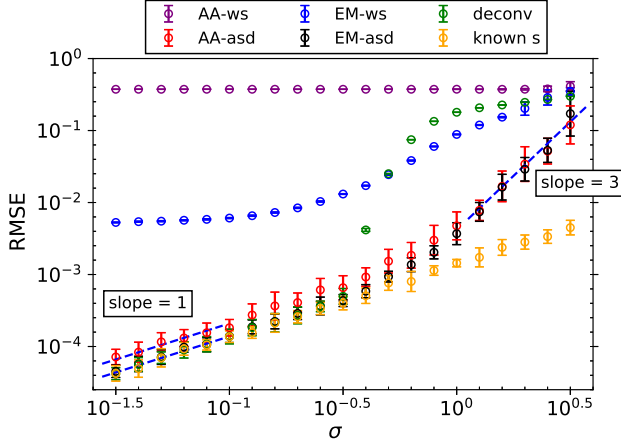
Fig. 6. The RMSEs of the reconstructed signal from data generated with an arbitrary spacing distribution. We show performance of our autocorrelation analysis (AA) and of expectation-maximization (EM), both under the (incorrect) assumptions of well-separated signals (ws) and the (correct) arbitrary spacing distribution model (asd). Also shown are the RMSEs of the signal estimated by the oracle-based deconvolution algorithm 'deconv' described in Section IV and the RMSEs of the estimated signal when the binary sequence s is known.

As a comparison, we also run the estimation algorithms under the (incorrect) assumption of well-separated signals on the same datasets. We first see that autocorrelation analysis assuming well-separated signals obtains poor reconstruction at all noise levels. Although the EM approach that assumes well-separated signals produces more accurate estimates, the nearly constant RMSEs at high SNR indicate the systematic error due to model misspecification. By contrast, the algorithms that assume an arbitrary spacing distribution achieve much better reconstruction, and the resulting RMSEs have the same scaling behaviors at the two extremes of SNR as shown in Figure 4. An instance of the signals reconstructed from data generated with $\sigma = 10^{0.3}$ and $\sigma = 10^{0.5}$ are shown in Figure 5(c) and 5(d) respectively.

One of the premises of this paper is that, at high noise levels, detection-based methods are destined to fail because detection cannot be done reliably. To support this theoretical argument further, in Figures 4 and 6 we display the performance of the following oracle, named 'deconv'. This method is given a strictly simpler task to solve: it must estimate x given the observation y, the noise level σ and the number of signal occurrences M, as well as the ℓ_2-distance between the ground truth x and every single window of length L in the observation y. Precisely, the oracle has access to the vector $z \in \mathbb{R}^{N-L+1}$ defined for $i = 0, \ldots, N-L$ by

$$z[i] = \sum_{l=0}^{L-1} (x[l] - y[i+l])^2.$$

The vector z provides this estimator the advantage of finding the likely locations of signal occurrences. If the norm of x is known, this oracle advantage is equivalent to knowledge of the cross-correlation between x and y. Then, the oracle-based estimator 'deconv' proceeds as follows: it selects the index i corresponding to the lowest value in z, excludes other indices too close to i as prescribed by the signal separation model

under consideration, and repeats this procedure M times to produce an estimator of s: the M starting positions of signal occurrences in y. The signal x is then estimated by averaging the M selected segments of y, effectively deconvolving y by the estimator of s. To our point: despite the unfair oracle advantage, this estimator is unable to estimate x to a competitive accuracy when the noise level is large.

Also plotted in Figures 4 and 6 are the RMSEs of the estimated signal when the binary sequence s is known. In this case, the RMSEs scale as σ and serve as the lower bounds of any algorithm. We can see that, at the low noise levels, our approximate EM algorithm performs nearly as well as the case when s is known.
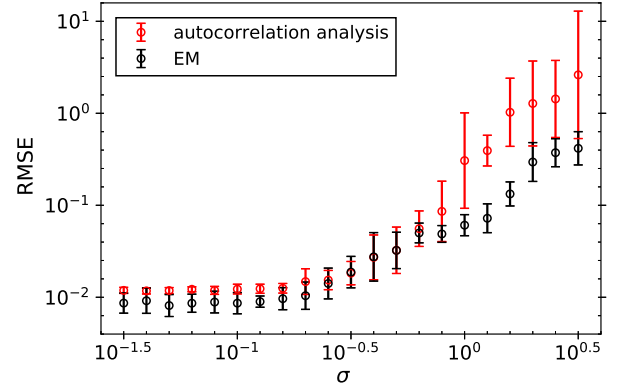

Fig. 7. The (relative) RMSEs of the reconstructed values of ρ_1 as defined by (11), for the case of arbitrary spacing distribution.

Figure 7 shows the RMSEs of the reconstructed values of ρ_1 for the experiments "AA-asd" and "EM-asd" shown in Figure 6. Although our methods are not able to reconstruct ρ_1 to high precision, the results shown in Figure 6 indicate the necessity to include the pair separation function in the model to achieve good signal reconstruction.
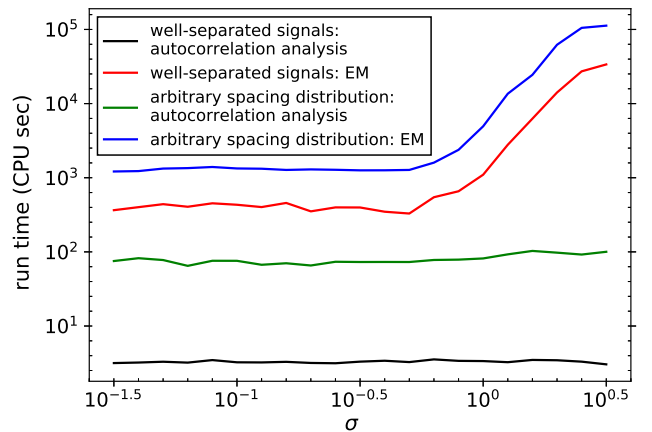

Fig. 8. The average run time to solve an instance of synthetic data for the two methods in both cases of signal spacing distribution.

Figure 8 shows the average computation time for the main methods in Figures 4 and 6. EM is slower because it cross-correlates all the observed segments with the signal estimate in each iteration. This issue becomes especially prominent in the case of arbitrary spacing distribution. On the other hand,

autocorrelation analysis requires shorter computation time by summarizing the data as autocorrelation statistics with one pass over the data. We note that the distinct computational speeds for autocorrelation analysis and the EM algorithm is also observed in the related problem of multi-reference alignment [8], [14]. The difference in run time and the similar reconstruction quality of the two methods at low SNR make autocorrelation analysis the preferred approach for large datasets.

V. DISCUSSION

In this study, we have presented two approaches—autocorrelation analysis and an approximate EM algorithm—to tackle the MTD problem without the need to detect the positions of the underlying signal occurrences. By incorporating the notion of pair separation function, we generalize the solution of the MTD problem from the special case of well-separated signals to allow an arbitrary spacing distribution of signal occurrence. It is empirically shown that the two methods have the same scaling of sample complexity in the two extremes of noise levels, in particular, SNR^{-3} at high noise level. Since the optimizations in both methods are non-convex, computing schemes based on frequency marching are designed to help the iterates converge to the global optimum.

Our study of the MTD model is primarily motivated by the goal to reconstruct the structures of small biomolecules using cryo-EM [11], [21]. In a cryo-EM experiment, individual copies of the target biomolecule are dispersed at unknown 2D locations and 3D orientations in a thin layer of vitreous ice, from which 2D tomographic projection images are produced by an electron microscope. To minimize the irreversible structural damage, it is necessary to keep the electron dose low. As a result, the projection images are considerably noisy, and high-resolution structure estimation requires averaging over a large number of noisy projections. In particular, as the size of the molecule gets smaller, the SNR of the data drops correspondingly.

Currently, the analysis workflow of cryo-EM data is roughly divided into two steps: The first step, known as particle picking, detects the locations of the biomolecules in the noisy projection images. The 3D structure of the biomolecule is reconstructed in the second step from the unoriented particle projections. When the sizes of biomolecules get smaller, however, reliable detection of their positions becomes challenging, which in turn hampers successful particle picking [3], [9]. It is estimated from first principles that, in order to obtain a 3 Å resolution reconstruction, the particle size should be at least 45 Å so that the particle occurrences can be accurately detected [23]. The results of [9], [10] and this work suggest that it is possible to bypass the need to detect particle positions but still reconstruct the structure at high resolution.

A key feature of cryo-EM that the MTD model fails to capture is that the "signals" are actually 2D projections of the underlying biomolecules at random 3D orientations, and the orientation distribution is usually non-uniform and unknown. This distinction makes the direct application of our approximate EM algorithm prohibitive, because each observed window now needs to cross-correlate with model projections with all possible 2D in-plane translations and 3D orientations, let alone the case where multiple copies of biomolecules may appear in a window. Further approximations such as representing the data and model projections in low-dimensional subspaces [19], [27] seem necessary to push this approach forward. We note that the use of frequency marching in the MTD problem is also related to this idea: the signal is mapped to low-dimensional subspaces with growing dimensions specified by $n_{\max}$.

On the other hand, the autocorrelations calculated from the noisy measurements in cryo-EM involve averages over both the 3D orientations and 2D in-plane translations. It was pointed out that the 3D structure reconstruction problem by fitting the first three autocorrelations of a biomolecule might be ill-conditioned [9], so information from higher-order autocorrelations seems required. We also expect the cross-terms between 2D projections of different copies to further complicate the reconstruction problem. The resolution of these technical challenges may help extend the use of cryo-EM to smaller biomolecules that are currently believed insurmountable, and is the subject of our ongoing studies.

APPENDIX A
DERIVATIONS OF RELATIONS (3)–(5) AND (8)–(10)

With the noisy measurement y defined in (1), we express its first order autocorrelation as

$$a_y^1 = \frac{1}{N} \sum_{i=0}^{N-1} (s * x)[i] + \frac{1}{N} \sum_{i=0}^{N-1} \varepsilon[i]$$
$$= \frac{M}{N} \sum_{i=0}^{L-1} x[i] + \frac{1}{N} \sum_{i=0}^{N-1} \varepsilon[i] = \rho_0 a_x^1 + \frac{1}{N} \sum_{i=0}^{N-1} \varepsilon[i].$$

Taking the expectation of a_y^1 with respect to the distribution of Gaussian noise, we obtain

$$\mathbb{E}_\varepsilon \{a_y^1\} = \rho_0 a_x^1,$$

with variance $\mathcal{O}(\sigma^2/N)$. We note that this expression applies to both the cases of well-separated signals and arbitrary spacing distribution.

The second order autocorrelation of y with shifts $l = 0, 1, \ldots, L-1$ can be written as

$$a_y^2[l] = \frac{1}{N} \sum_{i=0}^{N-1} y[i]y[i+l]$$
$$= \frac{1}{N} \sum_{i=0}^{N-1} \Big((s*x)[i](s*x)[i+l] + (s*x)[i]\varepsilon[i+l]$$
$$+ \varepsilon[i](s*x)[i+l] + \varepsilon[i]\varepsilon[i+l] \Big)$$
$$= a_{s*x}^2[l] + \frac{1}{N} \sum_{i=0}^{N-1} \varepsilon[i]\varepsilon[i+l]$$

$$+ \frac{1}{N} \sum_{k=0}^{M-1} \sum_{i=0}^{L-1} x[i] (\varepsilon[s_k - l + i] + \varepsilon[s_k + l + i]), \quad (31)$$

where $a^2_{s*x}[l]$ denotes the second order autocorrelation of the linear convolution $s*x$ with shifts l, and s_k indicates the index of the kth non-zero entry of the sequence s. Taking the expectation of $a^2_y[l]$ with respect to the distribution of Gaussian noise, we have

$$\mathbb{E}_\varepsilon\{a^2_y[l]\} = a^2_{s*x}[l] + \sigma^2 \delta[l].$$

The term that scales as $\mathcal{O}(\varepsilon^2)$ in (31) contributes to the variance by $\mathcal{O}(\sigma^4/N)$, while the terms that scale as $\mathcal{O}(\varepsilon)$ contribute to the variance by $\mathcal{O}(M\sigma^2/N^2)$. Since we assume that M grows with N at a constant rate, the resulting variance scales as $\mathcal{O}((\sigma^2 + \sigma^4)/N)$.

For well-separated signals,

$$a^2_{s*x}[l] = \frac{M}{N} \sum_{i=0}^{L-1} x[i]x[i+l] = \rho_0 a^2_x[l],$$

and we obtain the relation in (4). For arbitrary spacing distribution, we need to include the correlation terms between consecutive signal occurrences:

$$a^2_{s*x}[l] = \rho_0 a^2_x[l] + \frac{M}{N} \sum_{j=L}^{L+l-1} \xi[j] \sum_{i=0}^{L-1} x[i]x[i+j-l]$$

$$= \rho_0 a^2_x[l] + \rho_0 \sum_{j=L}^{L+l-1} \xi[j] a^2_x[j-l],$$

which gives the relation in (9). The quantity $\xi[j]$ denotes the number of consecutive 1's that are separated by exactly j entries in the sequence s, divided by $M-1$ so that $\sum_j \xi[j] = 1$.

As for the third order autocorrelation of y with shifts $0 \le l_1 \le l_2 < L$, we have

$$a^3_y[l_1, l_2] = \frac{1}{N} \sum_{i=0}^{N-1} y[i]y[i+l_1]y[i+l_2]$$

$$= \frac{1}{N} \sum_{i=0}^{N-1} \Big((s*x)[i](s*x)[i+l_1](s*x)[i+l_2]$$

$$+ (s*x)[i](s*x)[i+l_1]\varepsilon[i+l_2] + (s*x)[i]\varepsilon[i+l_1]\varepsilon[i+l_2]$$

$$+ \varepsilon[i](s*x)[i+l_1](s*x)[i+l_2] + \varepsilon[i](s*x)[i+l_1]\varepsilon[i+l_2]$$

$$+ (s*x)[i]\varepsilon[i+l_1](s*x)[i+l_2] + \varepsilon[i]\varepsilon[i+l_1](s*x)[i+l_2]$$

$$+ \varepsilon[i]\varepsilon[i+l_1]\varepsilon[i+l_2] \Big)$$

$$= a^3_{s*x}[l_1, l_2] + \frac{1}{N} \sum_{i=0}^{N-1} \varepsilon[i]\varepsilon[i+l_1]\varepsilon[i+l_2]$$

$$+ \frac{1}{N} \sum_{k=0}^{M-1} \sum_{i=0}^{L-1} \Big(x[i]\big(x[i+l_1] + \varepsilon[s_k + l_1 + i]\big)\varepsilon[s_k + l_2 + i]$$

$$+ \varepsilon[s_k - l_1 + i]x[i]\big(x[i+l_2 - l_1] + \varepsilon[s_k + l_2 - l_1 + i]\big)$$

$$+ \big(x[i-l_2] + \varepsilon[s_k - l_2 + i]\big)\varepsilon[s_k - l_2 + l_1 + i]x[i] \Big), \quad (32)$$

where $a^3_{s*x}[l_1, l_2]$ denotes the third order autocorrelation of the linear convolution $s*x$ with shifts (l_1, l_2), and x is zero-padded for indices out of the range $[0, L-1]$.

Taking the expectation of $a^3_y[l_1, l_2]$ with respect to the distribution of Gaussian noise, we obtain

$$\mathbb{E}_\varepsilon\{a^3_y[l_1, l_2]\}$$

$$= a^3_{s*x}[l_1, l_2] + \frac{M}{N} \sigma^2 \big(\delta[l_1] + \delta[l_2] + \delta[l_1 - l_2]\big) \sum_{i=0}^{L-1} x[i]$$

$$= a^3_{s*x}[l_1, l_2] + \rho_0 a^1_x \sigma^2 \big(\delta[l_1] + \delta[l_2] + \delta[l_1 - l_2]\big).$$

As for the variance, the term that scales as $\mathcal{O}(\varepsilon^3)$ in (32) contributes to the variance by $\mathcal{O}(\sigma^6/N)$, while the terms that scale as $\mathcal{O}(\varepsilon)$ contribute to the variance by $\mathcal{O}(M\sigma^2/N^2)$. Since we assume that M grows with N at a constant rate, the resulting variance scales as $\mathcal{O}((\sigma^2 + \sigma^6)/N)$.

For well-separated signals, we have

$$a^3_{s*x}[l_1, l_2] = \rho_0 a^3_x[l_1, l_2],$$

and we obtain the relation in (5). By including the correlations between consecutive signal occurrences, we expand $a^3_{s*x}[l_1, l_2]$ for the case of arbitrary spacing distribution as

$$a^3_{s*x}[l_1, l_2] = \rho_0 a^3_x[l_1, l_2]$$

$$+ \frac{M}{N} \sum_{j=L}^{L+l_2-l_1-1} \xi[j] \sum_{i=0}^{L-1} x[i+j-l_2]x[i+j+l_1-l_2]x[i]$$

$$+ \frac{M}{N} \sum_{j=L}^{L+l_1-1} \xi[j] \sum_{i=0}^{L-1} x[i+j-l_1]x[i]x[i+l_2-l_1]$$

$$= \rho_0 a^3_x[l_1, l_2] + \rho_0 \sum_{j=L}^{L+l_2-l_1-1} \xi[j] a^3_x[j-l_2, j+l_1-l_2]$$

$$+ \rho_0 \sum_{j=L}^{L+l_1-1} \xi[j] a^3_x[l_2 - l_1, j - l_1],$$

which gives the relation in (10).

APPENDIX B
DERIVATION OF (21) AND (22)

The constrained maximization in (20) can be achieved with the unconstrained maximization of the Lagrangian

$$\mathcal{L}(x, \alpha, \lambda) = Q(x, \alpha | x_k, \alpha_k) + \lambda \Big(1 - \sum_{l=0}^{2L-1} \alpha[l]\Big),$$

where λ denotes the Lagrange multiplier. We note that the constraints in (20) 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 $Q(x,\alpha|x_k,\alpha_k)$ is additively separable for x and α, we maximize $\mathcal{L}(x,\alpha,\lambda)$ with respect to x and α separately. At the maximum of $\mathcal{L}(x,\alpha,\lambda)$, we have

$$0 = \frac{\partial \mathcal{L}}{\partial x[j]} = \sum_{m=0}^{N_d-1} \sum_{l=0}^{2L-1} p(l|y_m,x_k)\frac{\partial \log p(y_m|l,x)}{\partial x[j]}, \quad (33)$$

where $0 \leq j < L$. With $p(y_m|l,x)$ give by (18), we can write

$$\frac{\partial \log p(y_m|l,x)}{\partial x[j]}$$
$$= -\frac{\partial}{\partial x[j]} \sum_{i=0}^{L-1} \frac{(y_m[i] - (CR_l Zx)[i])^2}{2\sigma^2}$$
$$= -\frac{\partial}{\partial x[j]} \sum_{i=0}^{L-1} \frac{(y_m[i] - (Zx)[(i+l) \bmod 2L])^2}{2\sigma^2}$$
$$= -\frac{1}{\sigma^2} \sum_{i=0}^{L-1} \Big(y_m[i] - (Zx)[(i+l) \bmod 2L] \Big)$$
$$\times \delta[(i+l) - (j+L)]. \quad (34)$$

Substituting this expression into (33), we obtain

$$\sum_{m=0}^{N_d-1} \sum_{l=j+1}^{j+L} p(l|y_m,x_k)(y_m[j+L-l] - x[j]) = 0.$$

Rearranging the terms gives the update rule shown in (21).

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

$$0 = \frac{\partial \mathcal{L}}{\partial \alpha[l]} = \frac{1}{\alpha[l]} \sum_{m=0}^{N_d-1} p(l|y_m,x_k) - \lambda,$$

where $0 \leq l < 2L$. We thus obtain the update rule for α as

$$\alpha[l] = \frac{1}{\lambda} \sum_{m=0}^{N_d-1} p(l|y_m,x_k),$$

and we can immediately solve $\lambda = N_d$ from the normalization $\sum_{l=0}^{2L-1} \alpha[l] = 1$.

APPENDIX C
DERIVATION OF (30)

As shown in (24), $Q(x,\alpha|x_k,\alpha_k)$ is additively separable for x and α. Moreover, the constraint (28) only involves the values of the parameters $\alpha[0], \alpha[1], \rho_1$, so the maximization of $Q(x,\alpha|x_k,\alpha_k)$ is unconstrained on x. At the maximum of $Q(x,\alpha|x_k,\alpha_k)$, we have

$$\frac{\partial Q(x,\alpha|x_k,\alpha_k)}{\partial x[j]} = 0$$
$$= \sum_{m=0}^{N_d-1} \Bigg[\sum_{l=0}^{2L-1} p(l|y_m,x_k)\frac{\partial \log p(y_m|l,x)}{\partial x[j]}$$
$$+ \sum_{l_1=L+1}^{2L-1} \sum_{l_2=1}^{l_1-L} p(l_1,l_2|y_m,x_k)\frac{\partial \log p(y_m|l_1,l_2,x)}{\partial x[j]} \Bigg], \quad (35)$$

where $0 \leq j < L$. With $p(y_m|l_1,l_2,x_k)$ given by (23), we obtain

$$\frac{\partial \log p(y_m|l_1,l_2,x)}{\partial x[j]}$$
$$= -\frac{\partial}{\partial x[j]} \sum_{i=0}^{L-1} \frac{(y_m[i] - (CR_{l_1} Zx)[i] - (CR_{l_2} Zx)[i])^2}{2\sigma^2}$$
$$= -\frac{1}{2\sigma^2}\frac{\partial}{\partial x[j]} \sum_{i=0}^{L-1} \Big(y_m[i] - (Zx)[(i+l_1) \bmod 2L]$$
$$- (Zx)[(i+l_2) \bmod 2L] \Big)^2$$
$$= -\frac{1}{\sigma^2} \sum_{i=0}^{L-1} \Big(y_m[i] - (Zx)[(i+l_1) \bmod 2L]$$
$$- (Zx)[(i+l_2) \bmod 2L] \Big)$$
$$\times \Big(\delta[(i+l_1) - (j+L)] + \delta[(i+l_2) - (j+L)] \Big). \quad (36)$$

Substituting (34) and (36) into (35), we obtain

$$0 = \sum_{m=0}^{N_d-1} \Bigg[\sum_{l=j+1}^{j+L} p(l|y_m,x_k)(y_m[j+L-l] - x[j])$$
$$+ \sum_{l_1=L+1}^{j+L} \sum_{l_2=1}^{l_1-L} p(l_1,l_2|y_m,x_k)(y_m[j+L-l_1] - x[j])$$
$$+ \sum_{l_1=L+j+1}^{2L-1} \sum_{l_2=j+1}^{l_1-L} p(l_1,l_2|y_m,x_k)(y_m[j+L-l_2] - x[j]) \Bigg].$$

Rearranging the terms, we obtain the update rule in (30).

REFERENCES

[1] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," Proceedings of the IEEE, vol. 85, no. 8, pp. 1310–1322, 1997.
[2] A. Aura, R. Carrillo, J.-P. Thiran, and Y. Wiaux, "Tensor optimization for optical-interferometric imaging," Monthly Notices of the Royal Astronomical Society, vol. 437, no. 3, pp. 2083–2091, 2013.
[3] C. Aguerrebere, M. Delbracio, A. Bartesaghi, and G. Sapiro, "Fundamental limits in multi-image alignment," IEEE Transactions on Signal Processing, vol. 64, no. 21, pp. 5707–5722, 2016.
[4] Y. Aizenbud, B. Landa, and Y. Shkolnisky, "Rank-one multi-reference factor analysis," arXiv:1905.12442, 2019.
[5] A. S. Bandeira, B. Blum-Smith, J. Kileel, A. Perry, J. Weed, and A. S. Wein, "Estimation under group actions: recovering orbits from invariants," arXiv:1712.10163, 2017.
[6] A. Barnett, L. Greengard, A. Pataki, and M. Spivak, "Rapid solution of the cryo-EM reconstruction problem by frequency marching," SIAM Journal on Imaging Sciences, vol. 10, no. 3, pp. 1170–1195, 2017.
[7] R. H. T. Bates, "Fourier phase problems are uniquely solvable in more than one dimension. I: underlying theory," Optik vol. 61, pp. 247–262, 1982.
[8] T. Bendory, N. Boumal, C. Ma, Z. Zhao and A. Singer, "Bispectrum inversion with application to multireference alignment," IEEE Transactions on Signal Processing, vol. 66, no. 4, pp. 1037–1050, 2018.
[9] T. Bendory, N. Boumal, W. Leeb, E. Levin, and A. Singer, "Toward single particle reconstruction without particle picking: Breaking the detection limit," arXiv:1810.00226, 2018.
[10] T. Bendory, N. Boumal, W. Leeb, E. Levin, and A. Singer, "Multitarget detection with application to cryo-electron microscopy," Inverse Problems, 2019.
[11] T. Bendory, A. Bartesaghi, and A. Singer, "Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities," arXiv:1908.00574, 2019.

[12] R. Beinert and G. Plonka. "Ambiguities in one-dimensional discrete phase retrieval from Fourier magnitudes," Journal of Fourier Analysis and Applications, vol. 21, no. 6, pp. 1169–1198, 2015.

[13] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, "Manopt, a Matlab toolbox for optimization on manifolds." Journal of Machine Learning Research, vol. 15, pp. 1455–1459, 2014.

[14] N. Boumal, T. Bendory, R. R. Lederman, and A. Singer, "Heterogeneous multireference alignment: A single pass approach," 2018 52nd Annual Conference on Information Sciences and Systems (CISS), pp. 1–6, 2018.

[15] Y. M. Bruck and L. G. Sodin, "On the ambiguity of the image reconstruction problem," Optics Communications, vol. 30, no. 3, pp. 304–308, 1979.

[16] E. J. Candes, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming." Communications on Pure and Applied Mathematics, vol. 66, no. 8, pp. 1241–1274, 2013.

[17] S. Cho and S. Lee, "Fast motion deblurring," ACM Transactions on Graphics, vol. 28, no. 5, 2009.

[18] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," Journal of the Royal Statistical Society, Series B, vol. 39, no. 1, pp. 1–38, 1977.

[19] N. C. Dvornek, F. J. Sigworth, and H. 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.

[20] M. Frank and P. Wolfe, "An algorithm for quadratic programming," Naval Research Logistics Quarterly, vol. 3, pp. 95–110, 1956.

[21] J. Frank, "Three-dimensional electron microscopy of macromolecular assemblies: Visualization of biological molecules in their native state," Oxford University Press, 2006.

[22] G. B. Giannakis and J. M. Mendel, "Identification of nonminimum phase systems using higher order statistics," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 37, no. 3, pp. 360–377, 1989.

[23] R. 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.

[24] S. M. Jefferies and J. C. Christou, "Restoration of astronomical images by iterative blind deconvolution," Astrophysical Journal vol. 415, pp. 862–874, 1993.

[25] D. Kondur and D. Hatzinakos, "Blind image deconvolution," IEEE Signal Processing Magazine, vol. 13, no. 3, pp. 43–64, 1996.

[26] A. Levin, "Blind motion deblurring using image statistics," Advances in Neural Information Processing Systems, 2006.

[27] B. Landa and Y. Shkolnisky, "The steerable graph Laplacian and its application to filtering image datasets," SIAM Journal on Imaging Sciences, vol. 11, no. 4, pp. 2254–2304, 2018.

[28] J. G. McNally, T. Karpova, J. Cooper, and J. A. Conchello, "Three-dimensional imaging by deconvolution microscopy," Methods, vol. 19, no. 3, pp. 373–385, 1999.

[29] A. Perry, J. Weed, A. Bandeira, P. Rigollet, and A. Singer, "The sample complexity of multi-reference alignment." SIAM Journal on Mathematics of Data Science, vol. 1, no. 3, pp.497–517, 2019.

[30] R. Holger, and . Stojanac, "Tensor theta norms and low rank recovery." arXiv:1505.05175, 2015.

[31] B. M. Sadler and G. B. Giannakis, "Shift- and rotation-invariant object reconstruction using the bispectrum," Journal of the Optical Society of America A, vol. 9, no. 1, pp. 57–69, 1992.

[32] P. Sarder and A. Nehorai, "Deconvolution methods for 3-D fluorescence microscopy images," IEEE Signal Processing Magazine, vol. 23, no. 3, pp. 32–45, 2006.

[33] T. J. Schulz, "Multi-frame blind deconvolution of astronomical images," Journal of the Optical Society of America A, vol. 10, no. 5, pp. 1064–1073, 1993.

[34] S. H. W. Scheres, "RELION: Implementation of a Bayesian approach to cryo-EM structure determination," Journal of Structural Biology, vol. 180, no. 3, pp. 519–530, 2012.

[35] A. Swami, G. B. Giannakis, and J. Mendel, "Linear modeling of multi-dimensional non-gaussian processes using cumulants," Multidimensional Systems and Signal Processing, vol. 1, no. 1, pp. 11–37, 1990.