Multi-target detection with application to cryo-electron microscopy

Tamir Bendory[a], Nicolas Boumal[b], William Leeb[c], Eitan Levin[a,b], and Amit Singer[a,b]

[a]The Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ, USA
[b]Department of Mathematics, Princeton University, Princeton, NJ, USA
[c]School of Mathematics, University of Minnesota, Minneapolis, MN, USA

February 26, 2019

Abstract

abstract

1 Introduction

We consider the *multi-target detection* problem of recovering a set of K signals that appear multiple times at unknown locations in a noisy measurement. Let $x_1, \dots, x_K \in \mathbb{R}^L$ be the sought signals and let $y \in \mathbb{R}^N$ be the observed data, where we assume N is far larger than L. Let $s[i]$ count the number of signals whose first element is positioned in $y[i]$. Each of those $s[i]$ is chosen according to some (possibly unknown) distribution over $\{1, \dots, K\}$. If signal occurrences overlap, they interfere additively. With additive white Gaussian noise, the measurement model can be written as

$$y = \sum_{k=1}^{K} s_k * x_k + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2 I_N), \tag{1.1}$$

where $*$ denotes linear convolution, and $s_k[i]$ indicates the number of starting positions of x_k in $y[i]$ so that $s = s_1 + \cdots + s_K$. Explicitly, with zero-based indexing,

$$y[i] = \sum_{k=1}^{K} \sum_{j=0}^{L-1} s_k[i-j] x_k[j] + \varepsilon[i].$$

The goal is to estimate $x_1, \ldots, x_K$ from y. In parts of the paper, we focus on the case $K = 1$, called the *homogeneous* case; the case $K > 2$ is called *heterogeneous*. This idealized setup appears in several scientific applications, including structural biology [13] (as we detail below), spike sorting [31], passive radar [20], and system identification [36].

In the low noise regime, a valid strategy is to first detect the signal occurrences in y (that is, estimate s), cluster them (that is, separate s into $s_1, \ldots, s_K$), and solve a standard deconvolution problem. Crucially, we focus on the high noise regime, where *reliable detection of signal occurrences is impossible* [13, 3]. This limitation does not, however, preclude estimation of the signals $x_1, \ldots, x_K$, as we show in this paper. In this setting, we consider $s_1, \ldots, s_K$ as *nuisance variables*.

In order to recover the signals in the high noise regime, we use autocorrelation analysis. For any noise level, the autocorrelations of the observation can be estimated to any desired accuracy for large enough N. This computation is straightforward and requires only one pass over the data. The underlying principle is to relate the autocorrelations of the observation y to the autocorrelations of $x_1, \ldots, x_K$. Below we describe two generative models for s. In these models, the relationship between the autocorrelations of y and those of $x_1, \ldots, x_K$ depend on $s_1, \ldots, s_K$ only through their sums, that is, the total number of occurrences of each signal. To estimate the signals and occurrence counts from the computed autocorrelations, we solve a nonlinear least-squares (LS) problem as explained in Section 5.

The multi-target detection problem is an instance of *blind deconvolution*—a long-standing problem arising in a variety of engineering and scientific applications, such as astronomy, communication, image deblurring, system identification and optics; see [24, 39, 5, 2], just to name a few. Different variants of the blind deconvolution problem have been subject recently to a thorough analysis [4, 33, 32, 29, 34, 27]. In clear contrast to multi-target detection, these works focus on the low noise regime and aim at estimating both unknown signals.

Models for target distribution

We consider two models for the distribution of signal occurrences in the observation, that is, for $s_1, \ldots, s_K$.

The well-separated model. As a first setup, we allow any generative model for s which meets the following separation requirement: s is binary, and

$$\text{If } s[i] = 1 \text{ and } s[j] = 1 \text{ for } i \neq j, \text{ then } |i - j| \geq 2L - 1. \tag{1.2}$$

In words: the starting positions of any two occurrences must be separated by at least $2L - 1$ positions, so that their end points are necessarily separated by at least $L - 1$ signal-free entries in the data. Furthermore, we require that the last signal occurrence in y is also followed by at least $L - 1$ signal-free (but still noisy) entries. This property ensures that correlating y with versions of itself shifted by at most $L-1$ entries does not

involve correlating distinct signal occurrences. Once s is determined, for each position i such that $s[i] = 1$, one of the signals x_k is selected independently at random, and accordingly we set $s_k[i] = 1$. As a result, the only properties of $s_1, \ldots, s_K$ that affect the autocorrelations of y are the total number of occurrences of the distinct signals: their individual and relative locations do not intervene.

The Poisson model. If the separation condition is violated, more knowledge about the location distribution is necessary to disentangle the autocorrelations of y. To that effect, we consider a Poisson model.

[Careful that γ changed by a factor of L: need to modify here too.] Specifically, for some parameter $\gamma > 0$, the total number of signal occurrences is drawn from a Poisson distribution with expectation $(N - L + 1)\gamma$. For each occurrence, a starting location i is drawn uniformly and independently at random from $\{0, \ldots, N - L\}$. Then, one of the signals $x_1, \ldots, x_K$ is selected independently at random from a fixed distribution and added to the observation y with first entry positioned at i. By definition of the Poisson process, the entries of s are independent and follow a Poisson distribution with expectation γ. In particular, this model allows for (additive) signal overlap. In Section **??**, we show that the autocorrelations of y are nonetheless equivalent to those arising under our well-separated model above.

Extensions

Extending the problem setup and autocorrelation analysis to signals in more than one dimension is straightforward: see discussion in Section 4.3 and numerical experiments in Section 5.

Likewise, it is easy to extend the model to situations where the signal occurrences are sampled from a general distribution rather than from a finite set of choices $x_1, \ldots, x_K$. In this setup, the goal is to estimate the distribution (possibly defined by a finite set of parameters). In particular, this allows for continuous distributions of targets. We adopt this perspective when deriving the autocorrelations in Section 3.

In the next section, we show how this flexibility allows us to model an important imaging problem in structural biology.

2 Motivation: single-particle reconstruction using cryo-electron microscopy

Cryo-electron microscopy (cryo-EM) has recently joined X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy as a high-resolution structural method for biological macromolecules; see for instance [18, 26, 10]. In contrast X-ray and NMR which aggregate information from an ensembles of particles, single particle cryo-EM produces images of individual particles and thus can, in principle, elucidate multiple

structures. In addition, it does not require the formation of crystalline arrays of macromolecules.

In a cryo–EM experiment, biological samples are rapidly frozen in a thin layer of vitreous ice. The microscope produces a 2-D tomographic image of the samples embedded in the ice, called a *micrograph*. Each micrograph contains tomographic projections of the samples at unknown locations and under unknown viewing directions. The goal is to construct 3-D models of the molecular structure from the micrographs. Importantly, to keep radiation damage within acceptable bounds, the dose must be kept low, leading to high noise levels.

All contemporary methods in the field split the reconstruction procedure into two main stages. The first stage consists in extracting the particle projections from the micrographs. This stage is called *particle picking*. The second stage aims to construct a 3-D model of the molecular structure from these projections. The quality of the reconstruction eventually hinges on the quality of the particle picking stage. Crucially, it can be shown that reliable detection of individual particles is impossible below a certain critical SNR. This fact has been recognized early on by the cryo-EM community. Particularly, in [22, 19], it was established that particle picking is impossible for molecules below a certain weight (below $\sim$50 kDa).

Another potential pitfall of particle picking pertains to *model bias*, whose importance in cryo-EM was stressed by a number of authors [40, 43, 23, 42]. In the classical "Einstein from noise" experiment, multiple realizations of pure noise are aligned to a picture of Einstein using template matching and then averaged. In [40], it was shown that the averaged noise rapidly becomes remarkably similar to the Einstein template. In the context of cryo–EM, this experiment exemplifies how prior assumptions about the particles may influence the reconstructed structure. This model bias is common to all particle picking methods based on template matching.

A recent work of the authors suggests to bypass the particle picking stage and reconstruct the 3-D structure directly from the micrograph [13]. In that paper, it was shown that—at least in principle—the limits particle picking imposes on molecule size do not necessarily translate into limits on particle reconstruction. The principle mathematical tool is *autocorrelation analysis*, described in detail in Section 3. This goal of the current paper is to provide a theoretical justification and numerical support for the method proposed in [13]. In this context, the models described above serve as an abstraction of the cryo-EM problem: the random signal X discussed in Section 1 can be thought of as 2-D random tomographic projections of the 3-D structure taken according to some unknown distribution of the particles within the ice.

We mention that [13] was not the first paper to employ autocorrelation analysis to cryo-EM. Zvi Kam [25] first proposed autocorrelation analysis for 3-D reconstruction, under the assumption of perfect particle picking: his method used autocorrelations of the picked, perfectly centered, particles. His method has been extended and employed by in X-ray free electron lasers and cryo-EM; see for instance [35, 28, 30, 44]. In order to investigate the computational and statistical properties of Kam's method, a series of papers have studied a simplified model, called *multi-reference alignment* [8, 14, 6, 38, 7,

1]. We follows the same line of research by considering the multi-target detection and clustering as an abstraction to the application of reconstructing 3-D structures directly from the micrograph as proposed in [13].

3 Autocorrelation analysis

In what follows, we consider autocorrelations of both the observation y and of the signal occurrences in y. As per our discussion of extensions, the signal occurrences may be sampled from a discrete set $\{x_1, \ldots, x_K\}$, or from a more general distribution. Accordingly, we define autocorrelations for a random signal z of length M . For our purposes, this will be applied both to signal occurrences (of length L) and to y (of length N).

For a random signal $z \in \mathbb{R}^M$, the autocorrelation of order $q = 1, 2, \ldots$ is given for any integer shifts $\ell_1, \ldots, \ell_{q-1}$ by

$$a_z^q[\ell_1, \ldots, \ell_{q-1}] = \mathbb{E}_z \left\{ \frac{1}{M} \sum_{i=-\infty}^{\infty} z[i] z[i+\ell_1] \cdots z[i+\ell_{q-1}] \right\}, \tag{3.1}$$

where the expectation is taken with respect to the distribution of z. Indexing of z out of the range $0, \ldots, M-1$ is zero-padded. Explicitly, the first-, second- and third-order autocorrelations are given by:

$$\begin{aligned} a_z^1 &= \mathbb{E}_z \left\{ \frac{1}{M} \sum_{i=0}^{M-1} z[i] \right\}, \\ a_z^2[\ell] &= \mathbb{E}_z \left\{ \frac{1}{M} \sum_{i=\max\{0,-\ell\}}^{M-1+\min\{0,-\ell\}} z[i] z[i+\ell] \right\}, \\ a_z^3[\ell_1, \ell_2] &= \mathbb{E}_z \left\{ \frac{1}{M} \sum_{i=\max\{0,-\ell_1,-\ell_2\}}^{M-1+\min\{0,-\ell_1,-\ell_2\}} z[i] z[i+\ell_1] z[i+\ell_2] \right\}. \end{aligned} \tag{3.2}$$

Since autocorrelations depend only on the differences between indices, they obey the following symmetries:

$$a_z^2[\ell] = a_z^2[-\ell],$$

and

$$a_z^3[\ell_1, \ell_2] = a_z^3[\ell_2, \ell_1] = a_z^3[-\ell_1, \ell_2 - \ell_1].$$

In what follows, x is the random variable corresponding to signal occurrences in y. In particular, for the model (1.1), x is sampled from $\{x_1, \ldots, x_K\}$ with probabilities $(\pi_1, \ldots, \pi_K)$, so that its autocorrelations are given in explicit form as:

$$a_x^q = \sum_{k=1}^{K} \pi_k a_{x_k}^q. \tag{3.3}$$

We are given one observation (one realization) of y. Thus, we cannot compute the autocorrelations of y exactly as they involve taking an expectation against the distribution of y. However, by the law of large numbers, as N grows to infinity with γ remaining constant, the empirical autocorrelations of y converge to the actual autocorrelations of y, that is,

$$\lim_{N\to\infty,\gamma \text{ constant}} \frac{1}{N} \sum_{i=-\infty}^{\infty} y[i]y[i+\ell_1]\cdots y[i+\ell_{q-1}] = a_y^q[\ell_1,\ldots,\ell_{q-1}]. \tag{3.4}$$

This provides a concrete means of estimating the quantities a_y^q. In the remainder of this section, we relate the observables a_y^q to the unknowns a_x^q, first under the well-separated model, then under the Poisson model.

3.1 Autocorrelations under the well-separated model

Under the separation condition (1.2), the relation between autocorrelations of the observation y and those of x is particularly simple, as we now show. It is useful to introduce some notation: let $|s| = \sum_i s[i]$ denote the number of signal occurrences in y, and let

$$\gamma = \frac{|s|L}{N}. \tag{3.5}$$

This γ is the fraction of entries of y occupied by signal occurrences. Henceforth, we call it the signal density. The separation condition imposes $\gamma \le \frac{L}{2L-1} \approx 1/2$.

Owing to the separation condition, when correlating y with shifted versions of itself for shifts in $0,\ldots,L-1$, any given occurrence of x in y is only ever correlated with itself, and never with another occurrence. As a result, the autocorrelations of y depend on the corresponding autocorrelations of x, the noise level σ and the density γ (which is a weak dependence on the support signal s). Specifically, we show the following identities in Section 7.1:

$$a_y^1 = \gamma a_x^1, \tag{3.6}$$

$$a_y^2[\ell] = \gamma a_x^2[\ell] + \sigma^2\delta[\ell], \tag{3.7}$$

$$a_y^3[\ell_1,\ell_2] = \gamma a_x^3[\ell_1,\ell_2] + \sigma^2\gamma a_x^1\big(\delta[\ell_1]+\delta[\ell_2]+\delta[\ell_1-\ell_2]\big), \tag{3.8}$$

where $\delta[0]=1$ and $\delta[\ell\neq 0]=0$, and indices ℓ,ℓ_1,ℓ_2 are in the range $0\le\ell\le L-1$. Terms proportional to σ^2 are due to the noise. If σ is known, they can be handled easily. If σ is unknown, one can either estimate it form the data, or one can ignore the few entries of the autocorrelations that are affected by σ—one in a_y^2 and $3L-2$ in a_y^3, a relatively small number in both cases.

[We show in Section 4.1 how to estimate γ from observations for the particular case where $K=1$ (all signal occurrences are the same).]

3.2 Autocorrelations under the Poisson model

[What we want to have in this section]

1. Recall the Poisson model (it's already described at the end of intro, so keep it short).

2. Define a proper notion of γ (careful with scaling L).

3. State formulas for a_y^q with $q = 1, 2, 3$ (ideally, including noise terms) to parallel what we did in the well-separated section. It would be better to do everything with autocorrelations rather than moments to keep the story uniform.

4. (optional) Claim: under some assumptions (knowing γ? Others?), the autocorrelations of y of order $q = 1, 2, 3$ under the Poisson model provide the same information about x as the corresponding autocorrelations under the well-separated model. Could also be in Section 4.1.

4 Theory

[Explain difference between homogeneous and heterogeneous]

4.1 Guarantees for the homogeneous case

In this section, we derive some theoretical results for the homogeneous case. [Note: I am using terms interchangeably]

A signal is determined uniquely by its second- and third-order autocorrelations. Indeed, assuming $z[0]$ and $z[L-1]$ are nonzero (otherwise, redefine the length of the signal), we can recover z explicitly using this identity for $k = 0, \ldots, L-1$:

$$z[k] = \frac{z[0]z[k]z[L-1]}{z[0]z[L-1]} = \frac{a_z^3[k, L-1]}{a_z^2[L-1]}. \tag{4.1}$$

This proves the following useful fact:

Proposition 4.1. *A signal $z \in \mathbb{R}^L$ is determined uniquely from a_z^2 and a_z^3. [Definitely only clear for the homogeneous case]*

A couple of remarks are in order. First, (4.1) is not numerically stable: if $z[0]$ or $z[L-1]$ are close to 0, recovery of z is sensitive to errors in the autocorrelations. In practice, we recover z by fitting it to its autocorrelations using a nonconvex least-squares (LS) procedure, which is empirically more robust to additive noise; we have observed similar phenomena for related problems [14, 16, 1].

The observed moments a_y^1, a_y^2 and a_y^3 of y do not immediately yield the moments of the signal x, as seen by (3.6), (3.7) and (3.8); rather, the two are related by the

noise level σ and the ratio γ. We will show, however, that x is still determined by the observed moments of y.

First, we observe that if the noise level σ is known, generally, one can estimate γ from the first two moments of the micrograph. The proof is provided in Appendix 7.2.

Proposition 4.2. *[Can we do this simultaneously for Poisson? or do we only need to do it for well-separated and invoke the equivalence claim? Not if that claim requires knowledge of γ, σ.] Let $\sigma > 0$ be fixed and assume that the separation condition* (1.2) *holds. If the mean of x is nonzero, then*

$$\gamma \overset{a.s.}{=} \lim_{N\to\infty} \frac{L(a_y^1)^2}{a_y^2[0] + 2\sum_{\ell=1}^{L-1} a_y^2[\ell] - \sigma^2}. \tag{4.2}$$

Using third-order autocorrelation information of y, both the ratio γ and the noise σ are determined. For the following results, when we say that a result holds for a "generic" signal x, we mean that the set of signals which cannot be determined by these measurements has Lebesgue measure zero. In particular, this means that we can recover almost all signals with the given measurements. The proof is provided in Appendix 7.3.

Proposition 4.3. *Assume $L \geq 3$ and assume that the separation condition* (1.2) *holds. In the limit of $N \to \infty$, the observed autocorrelations a_y^1, a_y^2 and a_y^3 determine the ratio γ and noise level σ uniquely for a generic signal x. If $\gamma > \frac{1}{4}$, then this holds for any signal x with nonzero mean.*

From Propositions 4.1 and 4.3 we deduce the following:

Corollary 4.4. *In the limit of $N \to \infty$ and under the separation condition* (1.2)*, the signal x, the ratio γ, and the noise level σ are determined from the first three autocorrelation functions of y if either the signal x is generic or x has nonzero mean and $\gamma > \frac{1}{4}$.*

As a side note, under the separation condition, the length L of the signal can also be determined from the autocorrelations in the asymptotic regime, by inspection of the support of a_y^2.

4.2 Elemetary limitations of the heterogeneous case

[Check and specify if this holds for both models] [Did we define densities γ_k?]

In the heterogeneous model (1.1), the unknowns are K signals of length L, together with their densities $\gamma_1, \ldots, \gamma_K$ (equivalently: the distribution π and overall density γ) and possibly the noise level σ. To estimate these parameters, we must collect at least as many independent equations. Within our framework, polynomial equations are provided by the observable autocorrelations, which correspond to mixed autocorrelations of the unknowns as per (3.3). In this section, following [16], we count how many equations the first three autocorrelations may provide in the best case (discounting symmetries).

This leads to a straightforward information-theoretic upper bound on the number K of signals which can be estimated, as a function of L. This is only an upper bound, though a bound of the same type was shown to be tight in a similar setting [7].

The first-order autocorrelation a_y^1 (3.6) provides one equation. For second-order autocorrelations $a_y^2[\ell]$ (3.7), if σ is known we obtain L equations with ℓ ranging from 0 to $L-1$. If σ is unknown, we may disregard $a_y^2[0]$ (then only entry affected by σ) and still collect $L-1$ equations. Similarly, for third-order autocorrelations, $a_y^3[\ell_1, \ell_2]$ (3.8) with $0 \leq \ell_1, \ell_2 \leq L-1$ such that $\ell_2 \leq \ell_1$ includes all relevant entries for our purpose (this accounts for symmetries), providing $\frac{(L+1)(L+2)}{2} - 2$ equations in total. If we further exclude any entries such that ℓ_1, ℓ_2 or $\ell_1 - \ell_2$ are zero to avoid the need to estimate σ, there are $\frac{(L-1)(L-2)}{2}$ remaining entries.

Hence, if σ is known we collect

$$1 + L + \frac{(L+1)(L+2)}{2} - 2 = \frac{1}{2}L(L+5)$$

equations, while if it unknown and we choose not to estimate it, then we collect

$$1 + (L-1) + \frac{(L-1)(L-2)}{2} = \frac{1}{2}L(L-1) + 1$$

equations in total. Of course, there may be redundancy in these equations: we aim only to provide a bound, under the assumption that these equations are independent.

Since we aim to estimate KL parameters for the K signals of length L plus K parameters for the densities γ_k (or the distribution π and overall density γ), there are $K(L+1)$ unknowns. As a result, an absolute upper bound on K such that the estimation problem may be solvable is

$$K \leq \frac{L(L+5)}{2(L+1)}, \tag{4.3}$$

for the case of σ known, and

$$K \leq \frac{L(L-1)+1}{2(L+1)}$$

for the case of σ unknown and not estimated. Overall, this indicates that, at best, approximately $L/2$ signals and their densities can be recovered from the first three mixed autocorrelations. Based on related results in [7], we expect that as many as $L/2$ signals can indeed be estimated, though possibly not with computationally tractable estimators.

4.3 Autocorrelations in higher dimensions

[Debate this later]

Autocorrelations in d dimensions is defined for $\ell_1, \ldots, \ell_{q-1} \in \mathbb{Z}^d$ as

$$a_z^q[\ell_1, \ldots, \ell_{q-1}] = \mathbb{E}\left\{ \frac{1}{m^d} \sum_{i \in \mathbb{Z}^d} z[i] z[i+\ell_1] \cdots z[i+\ell_{q-1}] \right\}. \tag{4.4}$$

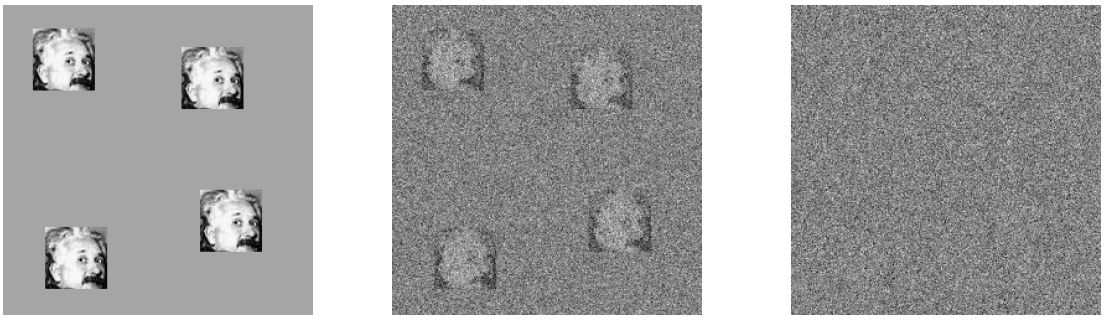
Interestingly, for dimensions greater than one almost all deterministic (that is, the distribution of z is a Delta function) signals are determined uniquely from their second-order autocorrelation, up to two symmetries: sign (or phase for complex signals) and reflection through the origin (with conjugation in the complex case) [21]. If the mean of signal is available and non-zero, the sign symmetry can be resolved. However, determining the reflection symmetry still requires additional information, beyond the second-order autocorrelation. The case of 1-D signals is essentially different: generally there are 2^{L-2} signals with the same second-order autocorrelation (after eliminating symmetries) [11, 12].

This uniqueness result for multi-dimensional signals (especially for 2-D images) is the basis of a popular imaging technique called coherent diffraction imaging (CDI). In CDI, an object is illuminated with a coherent wave, and the far field diffraction intensity pattern is measured, corresponding object's Fourier magnitude [37, 41]. If the support of the object is known and less than half of the support of the measured signal, then the data is equivalent to the second-order autocorrelation. The computational problem of recovering the signal is usually referred to as *phase retrieval* or *phase problem*. Despite the uniqueness result, recently it has been shown that, at least for 2-D images, the problem is ill-conditioned [9]. That is, there exist different images whose second-order autocorrelation agree up to machine precision.

5 Algorithms and numerical experiments

The technique we advocate allows recovery of a signal hidden in noisy micrographs without detecting the location of the signals embedded in these micrographs. To illustrate the underlying principles of the method, we present several numerical examples. The code to generate all figures is publicly available in `https://github.com/PrincetonUniversity/BreakingDetectionLimit`.

In the first experiment, we estimated an 50-by-50 pixel image of Einstein with mean zero from a growing number of micrographs, each of size 4096×4096 pixels. Each micrograph contains, on average, 700 occurrences of the target image at random locations. Thus, about 10% of each micrograph contains signal. The micrographs are contaminated with additive white Gaussian noise with standard deviation $\sigma = 3$, corresponding to SNR $= \frac{M\|x\|_F^2}{\sigma^2 N} \approx 1/370$. This high noise level is illustrated in Figure 1. To simplify the experiment, we assume the number of signal occurrences and the noise standard deviation are known. Micrographs are generated such that any two occurrences are always separated by at least 49 pixels in each direction in accordance with the separation condition (1.2).



(a) $\sigma = 0$ (b) $\sigma = 0.5$ (c) $\sigma = 3$

Figure 1: Example of micrographs of size 250×250 with additive white Gaussian noise of variance σ^2 for increasing values of σ. Each micrograph contains the same four occurrences of a 50×50 image of Einstein. In panel (c), the noise level is such that it is very challenging to locate the occurrences of the planted image. In fact, it can be shown that at low SNR, reliable detection of individual image occurrences is impossible, even if the true image is known. By analogy to cryo-EM, this depicts a scenario where particle picking cannot be done.

We compute the average second-order autocorrelation of the micrographs. This is a particularly simple computation which can be efficiently executed with a fast Fourier transform (FFT) in parallel. Given the noise level and number of image repetitions, the second-order autocorrelation of the image can be easily deduced from (3.7). Then, to estimate the target image, we resort to a standard phase retrieval algorithm called relaxed-reflect-reflect (RRR) [17], initialized randomly. Relative error is measured as the ratio of the root mean square error to the norm of the ground truth (square root of the sum of squared pixel intensities).

Figure 2 shows several estimated images for a growing number of micrographs. Figure 3 presents the normalized recovery error as a function of the amount of data available. This is computed after fixing the reflection symmetries (see Section 3). As evidenced by these figures, the ground truth image can be estimated increasingly well from increasingly many micrographs, without particle picking.

Appendix **??** provides additional details on the experiments.

In practice, we do not expect to know γ and maybe not even σ

Numerical experiment with three 1-D signals. For the 1-D experiment depicted in Figure **??**, we fix $K = 3$ signals of length $L = 21$. Following the forward model described at the beginning of this section, we generate an observation y of length $12.3 \cdot 10^9$. Each of the three signals appears, respectively (and approximately), $30.0 \cdot 10^6$, $20.0 \cdot 10^6$ and $10.0 \cdot 10^6$ times in y for a total of exactly $60 \cdot 10^6$ occurrences, such that at least $L-1$ zeros separate any two occurrences of any signals. This is done by randomly selecting $60 \cdot 10^6$ placements in y, one at a time with an accept/reject rule based on the separation constraint and locations picked so far. For each placement, one of the

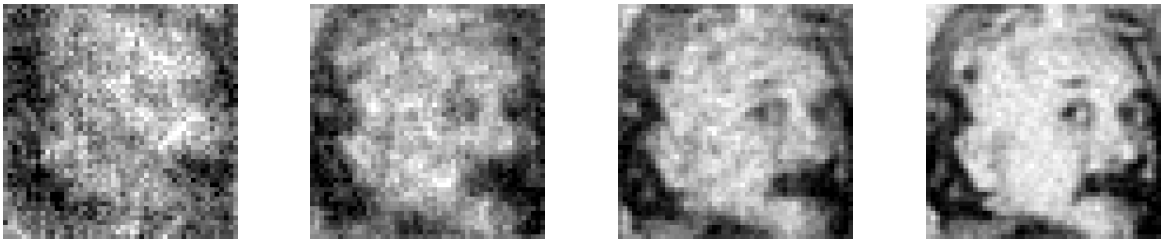


Figure 2: Recovery of Einstein from micrographs at noise level $\sigma = 3$ (see Figure 1(c)). Averaged autocorrelations of the micrographs allow to estimate the power spectrum of the target image. This does not require particle picking. A phase retrieval algorithm (RRR) produces the estimates here shown, initialized randomly. Estimates are obtained from $2 \times 10^2, 2 \times 10^3, 2 \times 10^4, 2 \times 10^5$ micrographs (growing across panels from left to right) of size 4096×4096, each containing 700 image occurrences on average.

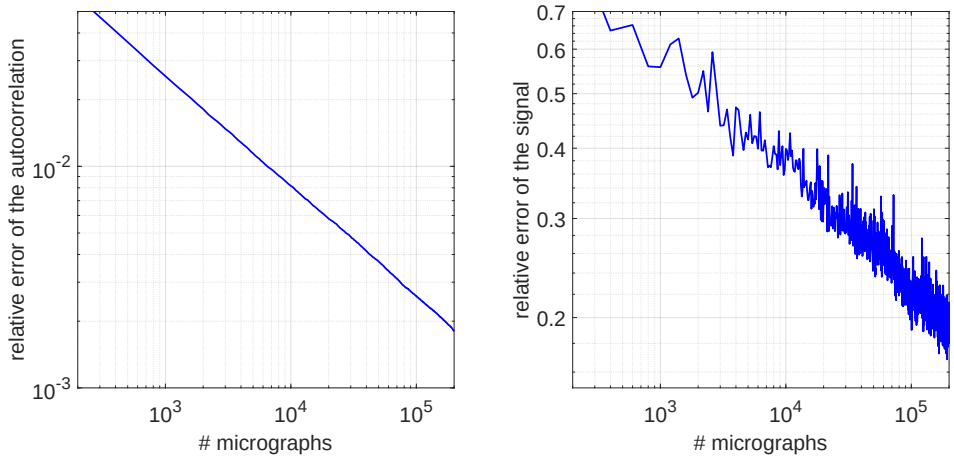


Figure 3: Relative error curves for the experiment of Figure 2.

three signals is picked at random according to the proportions $1/2, 1/3, 1/6$. Then, i.i.d. Gaussian noise with mean zero and standard deviation $\sigma = 3$ is added, to form the observed y. The resulting SNR of y is about $1/9$.

This is enough noise to make cross-correlations of y even with the true signals display peaks at essentially random locations, uninformative of the actual locations of the signal occurrences. Thus, we contend that it would be difficult for any algorithm to locate the signal occurrences, let alone to classify them according to which signal appears where.

Given the observation y, we proceed to compute the autocorrelations. The first-order autocorrelation is straightforward. For second-order autocorrelations, notice from equation (**??**) that $a_y^2[\ell]$ suffers no noise-induced bias for ℓ in 1 to $L-1$. Thus, we omit $\ell = 0$, which has the practical effect that we need not know σ to make sense of the computed quantities. Likewise, for third-order autocorrelations, $a_y^3[\ell_1, \ell_2]$ for $0 \leq \ell_1, \ell_2 \leq L-1$ such that $\ell_2 \leq \ell_1$ includes all relevant entries for our purpose (this accounts for symmetries), and we further exclude any such that ℓ_1, ℓ_2 or $\ell_1 - \ell_2$ are zero

to avoid the need to estimate σ—there are $\frac{(L-1)(L-2)}{2}$ remaining entries. We have

$$1 + (L-1) + \frac{(L-1)(L-2)}{2} = \frac{1}{2}L(L-1) + 1$$

coefficients in total. Since we aim to estimate KL parameters (for the K signals of length L) plus K parameters (for the densities γ_k), an absolute upper bound on K (simply to ensure we have at least as many equations as we have unknowns) is

$$K(L+1) \leq \frac{1}{2}L(L-1) + 1.$$

Thus, $(L-1)/2$ (up to a small approximation) is an absolute upper limit on K (compare with [16, 7]). [The last paragraph can be removed] In practice, the autocorrelations are computed on disjoint segments of y of length $100 \cdot 10^6$ and added up, without correction for the junction points. Segments are handled sequentially on a GPU, as GPUs are particularly well suited to execute simple instructions across large vectors of data. If multiple GPUs are available, segments can of course be handled in parallel.

Having computed the moments of interest, we now estimate signals $x_1, \ldots, x_K$ and coefficients $\gamma_1, \ldots, \gamma_K$ which agree with the data. We choose to do so by running an optimization algorithm on the following nonlinear least-squares problem:

$$\min_{\substack{\hat{x}_1,\ldots,\hat{x}_K \in \mathbb{R}^W \\ \hat{\gamma}_1,\ldots,\hat{\gamma}_K > 0}} w_1 \left(a_y^1 - \sum_{k=1}^{K} \hat{\gamma}_k a_{\hat{x}_k}^1 \right)^2 + w_2 \sum_{\ell=1}^{L-1} \left(a_y^2[\ell] - \sum_{k=1}^{K} \hat{\gamma}_k a_{\hat{x}_k}^2[\ell] \right)^2 + w_3 \sum_{\substack{2 \leq \ell_1 \leq L-1 \\ 1 \leq \ell_2 \leq \ell_1 - 1}} \left(a_y^3[\ell_1, \ell_2] - \sum_{k=1}^{K} \hat{\gamma}_k a_{\hat{x}_k}^3[\ell_1, \ell_2] \right)^2, \quad (5.1)$$

where $W \geq L$ is the length of the sought signals and the weights are set to $w_1 = 1/2, w_2 = 1/2n_2, w_3 = 1/2n_3$, where n_2, n_3 are the number of moments used: $n_2 = L-1$, $n_3 = \frac{(L-1)(L-2)}{2}$ (weights could also be set in accordance with variance estimates as in [16]).

Setting $W = L$ (as is a priori desired) is problematic because the above optimization problems appears to have numerous poor local optimizers. Thus, we first run the optimization with $W = 2L - 1$. This problem appears to have few poor local optima, perhaps because the additional degrees of freedom allow for more escape directions. Since we hope the signals estimated this way correspond to the true signals zero-padded to length W, we extract from each one a subsignal of length L that has largest ℓ_2-norm. This estimator is then used as initial iterate for (5.1), this time with $W = L$. We find that this procedure is reliable for a wide range of experimental parameters. To solve (5.1), we run the trust-region method implemented in Manopt [15], which allows to treat the positivity constraints on coefficients $\hat{\gamma}_k$. Notice that the cost function is a polynomial in the variables, so that it is straightforward to compute it and its derivatives.

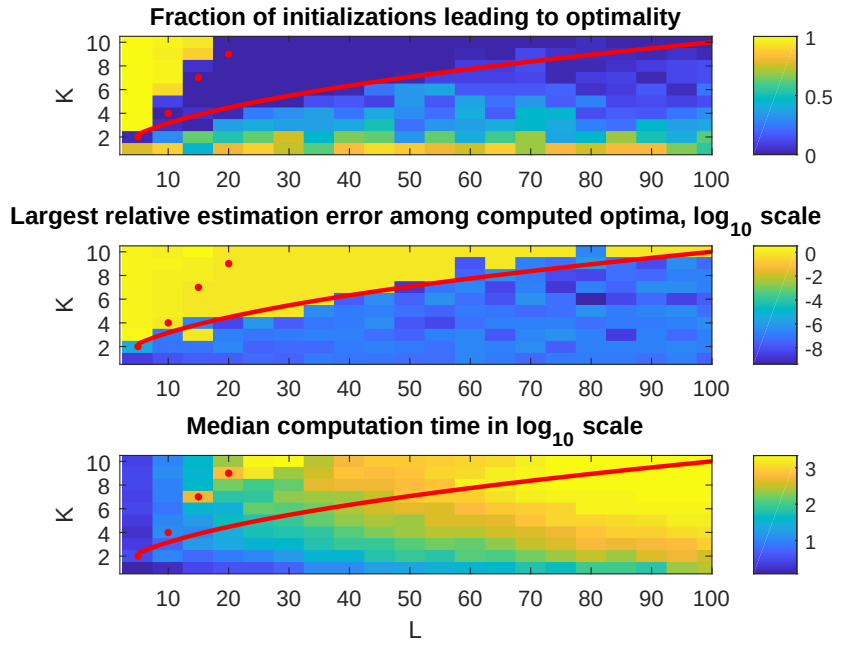


Figure 4: In the $N \to \infty$ regime (access to exact moments, excluding biased entries) and with known uniform densities, it seems K up to $\sqrt{L}$ (red curve) i.i.d. Gaussian signals of length L can be recovered from the known moments. CPU time in seconds. Strictly above red dots, recovery is impossible because the number of unknowns exceeds the number of computed moments. Similar to [16, Fig. 4.1], this experiment suggests a possible statistical-computational gap.

K-L figure

6 Relation with cryo-EM + experiments?

7 Proof

7.1 Derivation of the identities in Section 3

Let $x_{(1)}, \ldots, x_{(|s|)}$ denote the (independent) realizations of the random signal x in the observation y, starting at (deterministic) positions $s_{(1)}, \ldots, s_{(|s|)}$. Let I_{ij} be the indicator variable for whether position i is in the support of occurrence j, that is, it is one if i is in $\{s_{(j)}, \ldots, s_{(j)} + L - 1\}$, and zero otherwise. Then,

$$y[i] = \sum_{j=1}^{|s|} I_{ij} x_{(j)}[i - s_{(j)}] + \varepsilon[i]. \tag{7.1}$$

This gives a simple expression for the first autocorrelation of y. Indeed,

$$a_y^1 = \mathbb{E}_y \left\{ \frac{1}{N} \sum_{i=0}^{N-1} y[i] \right\} \tag{7.2}$$

$$= \frac{1}{N} \mathbb{E}_{x_{(1)},\ldots,x_{(|s|)},\varepsilon} \left\{ \sum_{i=0}^{N-1} \sum_{j=1}^{|s|} I_{ij} x_{(j)}[i - s_{(j)}] + \varepsilon[i] \right\}. \tag{7.3}$$

Now switch the sums over i and j, and observe that I_{ij} is zero unless $i = s_{(j)} + t$ for t in the range $0, \ldots, L-1$. Hence,

$$a_y^1 = \frac{1}{N} \sum_{j=1}^{|s|} \mathbb{E}_{x_{(j)}} \left\{ \sum_{t=0}^{L-1} x_{(j)}[t] \right\} + \frac{1}{N} \mathbb{E}_{\varepsilon} \left\{ \sum_{i=0}^{N-1} \varepsilon[i] \right\}. \tag{7.4}$$

Since the noise has zero mean and $x_{(1)}, \ldots, x_{(|s|)}$ are independent and all distributed as x, we further find:

$$a_y^1 = \frac{|s|L}{N} a_x^1 = \gamma a_x^1. \tag{7.5}$$

To address the second-order moments, we resort to the separation conditions. First, consider this expression:

$$\begin{aligned}
N \cdot a_y^2[\ell] &= \mathbb{E}_y \left\{ \sum_{i=0}^{N-\ell-1} y[i] y[i+\ell] \right\} \\
&= \sum_{i=0}^{N-\ell-1} \mathbb{E}_{x_{(1)},\ldots,x_{(|s|)},\varepsilon} \left\{ \left(\sum_{j=1}^{|s|} I_{ij} x_{(j)}[i - s_{(j)}] + \varepsilon[i] \right) \cdot \right. \\
&\qquad \left. \left(\sum_{j'=1}^{|s|} I_{i+\ell,j'} x_{(j')}[i + \ell - s_{(j')}] + \varepsilon[i+\ell] \right) \right\} \\
&= \sum_{i=0}^{N-\ell-1} \mathbb{E}_{x_{(1)},\ldots,x_{(|s|)},\varepsilon} \left\{ \sum_{j=1}^{|s|} \sum_{j'=1}^{|s|} I_{ij} I_{i+\ell,j'} x_{(j)}[i - s_{(j)}] x_{(j')}[i + \ell - s_{(j')}] \right. \\
&\qquad + \sum_{j=1}^{|s|} I_{ij} x_{(j)}[i - s_{(j)}] \varepsilon[i+\ell] \\
&\qquad + \sum_{j'=1}^{|s|} I_{i+\ell,j'} x_{(j')}[i + \ell - s_{(j')}] \varepsilon[i] \\
&\qquad \left. + \varepsilon[i] \varepsilon[i+\ell] \right\}.
\end{aligned}$$

The cross-terms vanish in expectation since ε is zero mean and independent from the signal occurrences. The last term vanishes in expectation unless $\ell = 0$ since distinct entries of ε are independent. For $\ell = 0$, $\mathbb{E}\{\varepsilon[i]^2\} = \sigma^2$. Finally, using the separation property, observe that if $I_{ij}I_{i+\ell,j'}$ is nonzero, then it is equal to one, $j = j'$ and $i = s_{(j)}+t$ for some t in $0, \ldots, L-\ell-1$. Then, switch the order of summations to get

$$N \cdot a_y^2[\ell] = \sum_{j=1}^{|s|} \mathbb{E}_{x_{(j)}} \left\{ \sum_{t=0}^{L-\ell-1} x_{(j)}[t]x_{(j)}[t+\ell] \right\} + (N-\ell)\sigma^2\delta[\ell], \tag{7.6}$$

where $\delta[0] = 1$ and $\delta[\ell \neq 0] = 0$. Since each $x_{(j)}$ is distributed as x, they all have the same autocorrelations as x and we finally get

$$a_y^2[\ell] = \gamma a_x^2[\ell] + \frac{N-\ell}{N}\sigma^2\delta[\ell] = \gamma a_x^2[\ell] + \sigma^2\delta[\ell]. \tag{7.7}$$

We now turn to the third-order autocorrelations. These involve the sum

$$\sum_{i=0}^{N-\max(\ell_1,\ell_2)-1} y[i]y[i+\ell_1]y[i+\ell_2]. \tag{7.8}$$

Using (7.1), we find that this quantity can be expressed as a sum of eight terms:

1. $\sum_i \sum_{j,j',j''=1}^{|s|} I_{ij}I_{i+\ell_1,j'}I_{i+\ell_2,j''}x_{(j)}[i-s_{(j)}]x_{(j')}[i+\ell_1-s_{(j')}]x_{(j'')}[i+\ell_2-s_{(j'')}]$
2. $\sum_i \sum_{j,j'=1}^{|s|} I_{ij}I_{i+\ell_1,j'}x_{(j)}[i-s_{(j)}]x_{(j')}[i+\ell_1-s_{(j')}]\varepsilon[i+\ell_2]$
3. $\sum_i \sum_{j,j''=1}^{|s|} I_{ij}I_{i+\ell_2,j''}x_{(j)}[i-s_{(j)}]\varepsilon[i+\ell_1]x_{(j'')}[i+\ell_2-s_{(j'')}]$
4. $\sum_i \sum_{j',j''=1}^{|s|} I_{i+\ell_1,j'}I_{i+\ell_2,j''}\varepsilon[i]x_{(j')}[i+\ell_1-s_{(j')}]x_{(j'')}[i+\ell_2-s_{(j'')}]$
5. $\sum_i \sum_{j=1}^{|s|} I_{ij}x_{(j)}[i-s_{(j)}]\varepsilon[i+\ell_1]\varepsilon[i+\ell_2]$
6. $\sum_i \sum_{j'=1}^{|s|} I_{i+\ell_1,j'}\varepsilon[i]x_{(j')}[i+\ell_1-s_{(j')}]\varepsilon[i+\ell_2]$
7. $\sum_i \sum_{j''=1}^{|s|} I_{i+\ell_2,j''}\varepsilon[i]\varepsilon[i+\ell_1]x_{(j'')}[i+\ell_2-s_{(j'')}]$
8. $\sum_i \varepsilon[i]\varepsilon[i+\ell_1]\varepsilon[i+\ell_2]$

Terms 2–4 and 8 vanish in expectation since odd moments of centered Gaussian variables are zero. For the first term, we use the fact that the separation condition implies

$$\begin{aligned} I_{ij}I_{i+\ell_1,j'}I_{i+\ell_2,j''} = 1 \iff \\ j = j' = j'' \text{ and } i = s_{(j)} + t \text{ with } t \in \{0, \ldots L - \max(\ell_1,\ell_2) - 1\}. \end{aligned} \tag{7.9}$$

(Otherwise, the product of indicators is zero.) This allows to reduce the summations over j, j', j'' to a single sum over j. Then, witching the order of summation with i, we get that the first term is equal to

$$\sum_{j=1}^{|s|} \sum_{t=0}^{L-\max(\ell_1,\ell_2)-1} x_{(j)}[t]x_{(j)}[t+\ell_1]x_{(j)}[t+\ell_2]. \tag{7.10}$$

In expectation over the realizations $x_{(j)}$, using again that they are i.i.d. with the same distribution as x, this first term yields $|s|La_x^3[\ell_1, \ell_2]$. Now consider the fifth term. Taking expectation against ε yields

$$\sum_{i=0}^{N-\max(\ell_1,\ell_2)-1} \sum_{j=1}^{|s|} I_{ij}x_{(j)}[i - s_{(j)}]\sigma^2\delta[\ell_1 - \ell_2]. \tag{7.11}$$

Switch the order of summation over i and j again to get

$$\sigma^2\delta[\ell_1 - \ell_2] \sum_{j=1}^{|s|} \sum_{t=0}^{L-1} x_{(j)}[t]. \tag{7.12}$$

Now taking expectation against the signal occurrences yields $|s|L\sigma^2 a_x^1\delta[\ell_1 - \ell_2]$. A similar reasoning for terms 6 and 7 yields this final formula for the third-order autocorrelations of y:

$$a_y^3[\ell_1, \ell_2] = \gamma a_x^3[\ell_1, \ell_2] + \gamma\sigma^2 a_x^1 \left(\delta[\ell_1] + \delta[\ell_2] + \delta[\ell_1 - \ell_2]\right). \tag{7.13}$$

7.2 Proof of Proposition 4.2

In the limit,

$$(a_y^1)^2 = \frac{\gamma^2}{L^2} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} x[i]x[j].$$

Similarly,

$$\sum_{\ell=1}^{L-1} a_y^2[\ell] = \frac{\gamma}{L} \sum_{\ell=1}^{L-1} \sum_{i=0}^{L-1-\ell} x[i]x[i+\ell],$$

and $a_y^2[0] = \frac{\gamma}{L} \sum_{i=0}^{L-1} x^2[i] + \sigma^2$. The proof is concluded by noting that $a_x^2[-\ell] = a_x^2[\ell]$.

7.3 Proof of Proposition 4.3

We prove that both σ and γ are identifiable from the observed first three moments of y. For convenience, we work with $\beta = \gamma/L$ rather than γ itself. To this end, we construct two quadratic equations satisfied by β and whose coefficients can be computed from observable quantities (in the limit). Then, we show that these equations are

independent, and hence that β is uniquely defined. Given β, we can estimate σ using Proposition 4.2.

Throughout the proof, it is important to distinguish between observed and unobserved values. We denote the observed values by E_i or a_y^1, a_y^2, a_y^3. We use F_i to denote functions of the signal's autocorrelations (which are not directly observable).

In the limit $N \to \infty$, almost surely, $a_y^1 = \beta(\mathbf{1}^T x)$ and $a_y^2[0] = \beta\|x\|^2 + \sigma^2$, where $\mathbf{1} \in \mathbb{R}^L$ is the vector of all-ones. (In this whole section, for clarity, we now omit to specify that identities hold almost surely in the limit.) Consider the product:

$$E_1 := a_y^1 a_y^2[0] = (\beta(\mathbf{1}^T x))(\beta\|x\|^2 + \sigma^2) = \sigma^2 a_y^1 + L\beta^2 F_1, \tag{7.14}$$

where $F_1 := a_x^3[0,0] + \sum_{j=1}^{L-1}(a_x^3[j,j] + a_x^3[0,j])$. The terms of F_1 can also be estimated from a_y^3, while taking the scaling and bias terms into account. This yields another observable:

$$\begin{aligned} E_2 &:= a_y^3[0,0] + \sum_{j=1}^{L-1}(a_y^3[j,j] + a_y^3[0,j]) \\ &= L\beta F_1 + (2L+1)\sigma^2 a_y^1. \end{aligned} \tag{7.15}$$

Therefore, from (7.14) and (7.15) we get:

$$E_2\beta - (2L+1)\sigma^2\beta a_y^1 = E_1 - \sigma^2 a_y^1. \tag{7.16}$$

Let $E_3 := a_y^2[0] + 2\sum_{j=1}^{L-1} a_y^2[j]$; recall from Proposition 4.2:

$$\sigma^2 = E_3 - (a_y^1)^2/\beta. \tag{7.17}$$

Plugging into (7.16) and rearranging, we get a first quadratic equation in β,

$$\mathcal{A}\beta^2 + \mathcal{B}\beta + \mathcal{C} = 0, \tag{7.18}$$

where

$$\begin{aligned} \mathcal{A} &= E_2 - (2L+1)a_y^1 E_3, \\ \mathcal{B} &= -E_1 + (2L+1)(a_y^1)^3 + a_y^1 E_3, \\ \mathcal{C} &= -(a_y^1)^3. \end{aligned}$$

Importantly, these coefficients are observable quantities. As we assume throughout this proof that x has nonzero mean, $a_y^1 \neq 0$ and we conclude that this equation is non-trivial.

Next, we derive the second quadratic equation for β. We notice that

$$E_4 := \frac{1}{L}(a_y^1)^3 = \frac{1}{L}\beta^3(\mathbf{1}^T x)^3 = \beta^3 F_2, \tag{7.19}$$

where $F_2 = \frac{1}{L}(\mathbf{1}^T x)^3$, and we can work out that:

$$F_2 = a_x^3[0,0] + 3\sum_{j=1}^{L-1}\left(a_x^3[j,j] + a_x^3[0,j]\right) + 6\sum_{1\le i<j\le L-1} a_x^3[i,j].$$

Once again, F_2 can be estimated from a_y^3, taking bias and scaling into account:

$$E_5 := a_y^3[0,0] + 3\sum_{j=1}^{L-1}\left(a_y^3[j,j] + a_y^3[0,j]\right) + 6\sum_{1\le i<j\le L-1} a_y^3[i,j] = L\beta F_2 + (6L-3)\sigma^2 a_y^1. \tag{7.20}$$

Consider the following ratio:

$$\frac{E_5}{E_4} = \frac{L}{\beta^2} + \frac{(6L-3)\sigma^2 a_y^1}{E_4}.$$

From the latter, we deduce:

$$\sigma^2 = \frac{E_5}{a_y^1(6L-3)} - \frac{LE_4}{\beta^2 a_y^1(6L-3)}.$$

Using (7.17) and rearranging, we get the second quadratic:

$$\mathcal{D}\beta^2 + \mathcal{E}\beta + \mathcal{F} = 0, \tag{7.21}$$

where

$$\begin{aligned}\mathcal{D} &= E_3 - \frac{E_5}{a_y^1(6L-3)},\\ \mathcal{E} &= -(a_y^1)^2,\\ \mathcal{F} &= \frac{LE_4}{a_y^1(6L-3)}.\end{aligned}$$

It is also non-trivial since $E_4 \neq 0$.

To complete the proof, we need to show that the two quadratic equations (7.18) and (7.21) are independent. To this end, it is enough to show that the ratios between coefficients differ. From (7.18) and (7.14), we have:

$$\frac{\mathcal{B}}{\mathcal{C}} = \frac{E_1 - (2L+1)(a_y^1)^3 - a_y^1 E_3}{(a_y^1)^3} = \frac{a_y^2[0] - (2L+1)(a_y^1)^2 - E_3}{(a_y^1)^2}.$$

In addition, using (7.19),

$$\frac{\mathcal{E}}{\mathcal{F}} = \frac{(3-6L)(a_y^1)^3}{LE_4} = 3 - 6L.$$

For contradiction, suppose that the quadratics are dependent. Then, $\frac{\mathcal{B}}{\mathcal{C}} = \frac{\mathcal{E}}{\mathcal{F}}$, that is,

$$a_y^2[0] - (2L+1)(a_y^1)^2 - E_3 = (a_y^1)^2(3-6L).$$

Rewriting the identity in terms of x and dividing by β we get:

$$4(L-1)\beta(\mathbf{1}^\top x)^2 - (\mathbf{1}^\top x)^2 + \|x\|^2 = 0. \tag{7.22}$$

For generic x, this polynomial equation is not satisfied so that the quadratic equations are independent. Furthermore, from the inequality $L\|x\|^2 \geq (\mathbf{1}^\top x)^2$ it follows immediately that the equations must be independent so long as

$$\beta > \frac{1}{4L}.$$

7.4 Proof of Proposition ??

7.4.1 First moment

To compute the first moment of y, we will first condition on $M = (M_1, \ldots, M_{N-L+1})$, and then average over M. We have:

$$\mathbb{E}[y[i]|M] = \sum_{j=0}^{L-1} \sum_{k=1}^{M_{i-j}} \mathbb{E}X_k^{i-j}[j] = \sum_{j=0}^{L-1} \sum_{k=1}^{M_{i-j}} m_y^1[j] = \sum_{j=0}^{L-1} M_{i-j} m_y^1[j]. \tag{7.23}$$

Now taking expectations over M we see:

$$\mathbb{E}Y[i] = \gamma \sum_{j=0}^{L-1} m_y^1[j] = \gamma L a_x^1. \tag{7.24}$$

7.4.2 Second moment

Again, we will condition on M first, and then take the expectation over M. Fix $i_1 \neq i_2$, and let $\Delta = i_2 - i_1$. Then:

$$Y_{i_1} Y_{i_2} = \sum_{j_1=0}^{L-1} \sum_{j_2=0}^{L-1} \sum_{k_1=1}^{M_{i_1-j_1}} \sum_{k_2=1}^{M_{i_2-j_2}} X_{k_1}^{i_1-j_1}[j_1] X_{k_2}^{i_2-j_2}[j_2]. \tag{7.25}$$

We break up the double sum over j_1 and j_2 into two terms: one where $j_2 \neq j_1 + \Delta$, and one where $j_2 = j_1 + \Delta$ or equivalently $i_1 - j_1 = i_2 - j_2$. In the first case, all the terms are independent, and so the expectation factors. In the second case, when $k_1 \neq k_2$ we have independence, but otherwise not. This gives (all expectations are conditional on

M):

$$\begin{aligned}
\mathbb{E}Y_{i_1}Y_{i_2} &= \sum_{j_1=0}^{L-1}\sum_{j_2=0}^{L-1}\sum_{k_1=1}^{M_{i_1-j_1}}\sum_{k_2=1}^{M_{i_2-j_2}} \mathbb{E}X_{k_1}^{i_1-j_1}[j_1]X_{k_2}^{i_2-j_2}[j_2] \\
&= \sum_{j_1-j_2\neq\Delta}\sum_{k_1}\sum_{k_2}\mathbb{E}X_{k_1}^{i_1-j_1}[j_1]X_{k_2}^{i_2-j_2}[j_2] \\
&\quad + \sum_{j_1=0}^{L-1}\sum_{k_1\neq k_2}\mathbb{E}X_{k_1}^{i_1-j_1}[j_1]X_{k_2}^{i_1-j_1}[j_1+\Delta] \\
&\quad + \sum_{j_1=0}^{L-1}\sum_{k_1=1}^{M_{i_1-j_1}}\mathbb{E}X_{k_1}^{i_1-j_1}[j_1]X_{k_1}^{i_1-j_1}[j_1+\Delta] \\
&= \sum_{j_1-j_2\neq\Delta} M_{i_1-j_1}M_{i_2-j_2}\mathcal{M}_1[j_1]\mathcal{M}_1[j_2] \\
&\quad + \sum_{j_1=0}^{L-1} M_{i_1-j_1}(M_{i_1-j_1}-1)\mathcal{M}_1[j_1]\mathcal{M}_1[j_1+\Delta] \\
&\quad + \sum_{j_1=0}^{L-1} M_{i_1-j_1}\mathcal{M}_2[j_1,j_1+\Delta]. \qquad (7.26)
\end{aligned}$$

Now take expectations over the Poisson random variables, using this fact:

Lemma 7.1. *If $M \sim Poisson(\gamma)$, then*

$$\mathbb{E}\binom{M}{k} = \frac{\gamma^k}{k!}. \qquad (7.27)$$

We get (now the expectation is over M and X):

$$\begin{aligned}
\mathbb{E}Y_{i_1}Y_{i_2} =& \sum_{j_1-j_2\neq\Delta} \mathbb{E}M_{i_1-j_1}M_{i_2-j_2}\mathcal{M}_1[j_1]\mathcal{M}_1[j_2] \\
&+ \sum_{j_1=0}^{L-1} \mathbb{E}M_{i_1-j_1}(M_{i_1-j_1}-1)\mathcal{M}_1[j_1]\mathcal{M}_1[j_1+\Delta] \\
&+ \sum_{j_1=0}^{L-1} \mathbb{E}M_{i_1-j_1}\mathcal{M}_2[j_1, j_1+\Delta] \\
=& \sum_{j_1-j_2\neq\Delta} \gamma^2\mathcal{M}_1[j_1]\mathcal{M}_1[j_2] + \sum_{j_1=0}^{L-1} \gamma^2\mathcal{M}_1[j_1]\mathcal{M}_1[j_1+\Delta] \\
&+ \sum_{j_1=0}^{L-1} \gamma\mathcal{M}_2[j_1, j_1+\Delta] \\
=& \left(\gamma\sum_{j=0}^{L-1}\mathcal{M}_1[j]\right)^2 + \gamma\sum_{j=0}^{L-1}\mathcal{M}_2[j, j+\Delta] \\
=& (\gamma\mathcal{L}_1)^2 + \gamma\mathcal{L}_2(\Delta). \qquad (7.28)
\end{aligned}$$

But the first term in the sum is just the square of the first moment of Y; so from the first two moments we can recover $\gamma\mathcal{L}_2(\Delta)$, which is just the expected power spectrum of the random vector X, i.e. the usual second moment we have been working with.

7.4.3 Third moment

For three distinct i_1, i_2 and i_3, we let $\Delta_1 = i_2 - i_1$ and $\Delta_2 = i_3 - i_1$. We have:

$$\begin{aligned}
&Y_{i_1}Y_{i_2}Y_{i_3} \\
&= \sum_{j_1=0}^{L-1}\sum_{j_2=0}^{L-1}\sum_{j_3=0}^{L-1}\sum_{k_1=1}^{M_{i_1-j_1}}\sum_{k_2=1}^{M_{i_2-j_2}}\sum_{k_3=1}^{M_{i_3-j_3}} X_{k_1}^{i_1-j_1}[j_1]X_{k_2}^{i_2-j_2}[j_2]X_{k_3}^{i_3-j_3}[j_3]. \qquad (7.29)
\end{aligned}$$

We will break up the outer three sums into disjoint sums with the following ranges of indices:

1. $j_2 = j_1 + \Delta_1$ and $j_3 = j_2 + \Delta_2 - \Delta_1$.
2. $j_2 = j_1 + \Delta_1$ and $j_3 \neq j_2 + \Delta_2 - \Delta_1$.
3. $j_2 \neq j_1 + \Delta_1$ and $j_3 = j_1 + \Delta_2$.
4. $j_2 \neq j_1 + \Delta_1$ and $j_3 \neq j_1 + \Delta_2$ and $j_3 = j_2 + \Delta_2 - \Delta_1$.
5. $j_2 \neq j_1 + \Delta_1$ and $j_3 \neq j_1 + \Delta_2$ and $j_3 \neq j_2 + \Delta_2 - \Delta_1$.

For Case 1, we have $\ell \equiv i_1 - j_1 = i_2 - j_2 = i_3 - j_3$. We further break up the sum:

$$\begin{aligned}
&\sum_{j=0}^{L-1} \sum_{k_1=1}^{M_\ell} \sum_{k_2=1}^{M_\ell} \sum_{k_3=1}^{M_\ell} X_{k_1}^\ell[j] X_{k_2}^\ell[j+\Delta_1] X_{k_3}^\ell[j+\Delta_2] \\
&= \underbrace{\sum_{j=0}^{L-1} \sum_{k_i \text{distinct}} X_{k_1}^\ell[j] X_{k_2}^\ell[j+\Delta_1] X_{k_3}^\ell[j+\Delta_2]}_{\text{(a)}} \\
&\quad + \underbrace{\sum_{j=0}^{L-1} \sum_{k_1=k_2\neq k_3} X_{k_1}^\ell[j] X_{k_2}^\ell[j+\Delta_1] X_{k_3}^\ell[j+\Delta_2]}_{\text{(b)}} \\
&\quad + \underbrace{\sum_{j=0}^{L-1} \sum_{k_1=k_3\neq k_2} X_{k_1}^\ell[j] X_{k_2}^\ell[j+\Delta_1] X_{k_3}^\ell[j+\Delta_2]}_{\text{(c)}} \\
&\quad + \underbrace{\sum_{j=0}^{L-1} \sum_{k_2=k_3\neq k_1} X_{k_1}^\ell[j] X_{k_2}^\ell[j+\Delta_1] X_{k_3}^\ell[j+\Delta_2]}_{\text{(d)}} \\
&\quad + \underbrace{\sum_{j=0}^{L-1} \sum_{k_1=k_2=k_3} X_{k_1}^\ell[j] X_{k_2}^\ell[j+\Delta_1] X_{k_3}^\ell[j+\Delta_2]}_{\text{(e)}} . \qquad (7.30)
\end{aligned}$$

For term (a), the expectation conditional on M is:

$$\sum_{j=0}^{L-1} M_\ell(M_\ell - 1)(M_\ell - 2)\mathcal{M}[j]\mathcal{M}[j+\Delta_1]\mathcal{M}[j+\Delta_2]. \qquad (7.31)$$

Using Lemma 7.1, the unconditional expectation of (a) is then:

$$\gamma^3 \sum_{j=0}^{L-1} \mathcal{M}_1[j]\mathcal{M}_1[j+\Delta_1]\mathcal{M}_1[j+\Delta_2]. \qquad (7.32)$$

For term (b), the expectation conditional on M is:

$$\sum_{j=0}^{L-1} M_\ell(M_\ell - 1)\mathcal{M}_2[j, j+\Delta_1]\mathcal{M}_1[j+\Delta_2] \qquad (7.33)$$

and then again using Lemma 7.1 we get the expected value:

$$\gamma^2 \sum_{j=0}^{L-1} \mathcal{M}_2[j, j+\Delta_1]\mathcal{M}_1[j+\Delta_2]. \qquad (7.34)$$

Similarly, the expected values of terms (c) and (d) are:

$$\gamma^2 \sum_{j=0}^{L-1} \mathcal{M}_2[j, j+\Delta_2]\mathcal{M}_1[j+\Delta_1]. \tag{7.35}$$

and

$$\gamma^2 \sum_{j=0}^{L-1} \mathcal{M}_2[j+\Delta_1, j+\Delta_2]\mathcal{M}_1[j]. \tag{7.36}$$

Finally, the expected value of term (e) is easily shown to be:

$$\gamma \sum_{j=0}^{L-1} \mathcal{M}_3[j, j+\Delta_1, j+\Delta_2]. \tag{7.37}$$

This concludes the computation for Case 1.

Moving onto Case 2, we have $\ell_1 \equiv i_1 - j_1 = i_2 - j_2$, and also define $\ell_2 \equiv i_3 - j_3$. By definition, $\ell_1 \neq \ell_2$. The sum is:

$$\begin{aligned}
&\sum_{j_1=0}^{L-1} \sum_{j_3 \neq j_1+\Delta_2} \sum_{1 \leq k_1, k_2 \leq M_{\ell_1}} \sum_{k_3=1}^{M_{\ell_2}} X_{k_1}^{\ell_1}[j_1] X_{k_2}^{\ell_1}[j_1+\Delta_1] X_{k_3}^{\ell_2}[j_3] \\
=&\sum_{j_1=0}^{L-1} \sum_{j_3 \neq j_1+\Delta_2} \sum_{k_3=1}^{M_{\ell_2}} \Bigg\{ \sum_{1 \leq k_1 \neq k_2 \leq M_{\ell_1}} X_{k_1}^{\ell_1}[j_1] X_{k_2}^{\ell_1}[j_1+\Delta_1] X_{k_3}^{\ell_2}[j_3] \\
&+ \sum_{k_1=1}^{M_{\ell_1}} X_{k_1}^{\ell_1}[j_1] X_{k_1}^{\ell_1}[j_1+\Delta_1] X_{k_3}^{\ell_2}[j_3] \Bigg\}.
\end{aligned} \tag{7.38}$$

Taking expectations conditional on M, we then get:

$$\begin{aligned}
&\sum_{j_1=0}^{L-1} \sum_{j_3 \neq j_1+\Delta_2} \Bigg(M_{\ell_1}(M_{\ell_1}-1) M_{\ell_2} \mathcal{M}_1[j_1]\mathcal{M}_1[j_1+\Delta_1]\mathcal{M}_1[j_3] \\
&+ M_{\ell_1} M_{\ell_2} \mathcal{M}_2[j_1, j_1+\Delta_1]\mathcal{M}_1[j_3] \Bigg).
\end{aligned} \tag{7.39}$$

Taking expectations over M and using Lemma 7.1 then gives:

$$\gamma^3 \sum_{j_1=0}^{L-1} \sum_{j_3 \neq j_1+\Delta_2} \mathcal{M}_1[j_1]\mathcal{M}_1[j_1+\Delta_1]\mathcal{M}_1[j_3] \tag{7.40}$$

$$+ \gamma^2 \sum_{j_1=0}^{L-1} \sum_{j_3 \neq j_1+\Delta_2} \mathcal{M}_2[j_1, j_1+\Delta_1]\mathcal{M}_1[j_3]. \tag{7.41}$$

Similarly, Cases 3 and 4 give the expressions:

$$\gamma^3 \sum_{j_1=0}^{L-1} \sum_{j_2 \neq j_1+\Delta_1} \mathcal{M}_1[j_1]\mathcal{M}_1[j_1+\Delta_2]\mathcal{M}_1[j_2] \tag{7.42}$$

$$+ \gamma^2 \sum_{j_1=0}^{L-1} \sum_{j_2 \neq j_1+\Delta_1} \mathcal{M}_2[j_1, j_1+\Delta_2]\mathcal{M}_1[j_2] \tag{7.43}$$

and

$$\gamma^3 \sum_{j_2=0}^{L-1} \sum_{j_1 \neq j_2} \mathcal{M}_1[j_1]\mathcal{M}_1[j_2+\Delta_1]\mathcal{M}_1[j_2+\Delta_2] \tag{7.44}$$

$$+ \gamma^2 \sum_{j_2=0}^{L-1} \sum_{j_1 \neq j_2} \mathcal{M}_2[j_2+\Delta_1, j_2+\Delta_2]\mathcal{M}_1[j_1]. \tag{7.45}$$

Finally, in Case 5 we have $i_1 - j_1$, $i_2 - j_2$, and $i_3 - j_3$ are all pairwise distinct. Consequently, the X variables are always independent, and the expectation conditional on M (letting $\ell_q = i_q - j_q$, $q = 1, 2, 3$),

$$\sum_{j_1, j_2, j_3} M_{\ell_1} M_{\ell_2} M_{\ell_3} \mathcal{M}_1[j_1]\mathcal{M}_1[j_2]\mathcal{M}_1[j_3]; \tag{7.46}$$

since the M_{ℓ_q}'s are pairwise independent, $q = 1, 2, 3$, the expectation over M then yields:

$$\gamma^3 \sum_{j_1, j_2, j_3} \mathcal{M}_1[j_1]\mathcal{M}_1[j_2]\mathcal{M}_1[j_3]. \tag{7.47}$$

Now we add all the terms from Cases 1 to 5. Expressions (7.32), (7.40), (7.42), (7.44), and (7.47) sum to the expression:

$$(\gamma \mathcal{L}_1)^3. \tag{7.48}$$

Note that this is obtained directly from the first moment. Expressions (7.34), (7.35), (7.36), (7.41),(7.43), and (7.45) sum to the expression:

$$\gamma \mathcal{L}_1 \cdot (\gamma \mathcal{L}_2(\Delta_1) + \gamma \mathcal{L}_2(\Delta_2) + \gamma \mathcal{L}_2(\Delta_2 - \Delta_1)). \tag{7.49}$$

Again, note that this is obtained directly from the first two moments. Finally, expression (7.37) is simply:

$$\gamma \mathcal{L}_3(\Delta_1, \Delta_2) \tag{7.50}$$

which is the usual third-order auto-correlation.

References

[1] Emmanuel Abbe, Tamir Bendory, William Leeb, João M Pereira, Nir Sharon, and Amit Singer. Multireference alignment is easier with an aperiodic translation distribution. *IEEE Transactions on Information Theory*, 2018.

[2] Karim Abed-Meraim, Wanzhi Qiu, and Yingbo Hua. Blind system identification. *Proceedings of the IEEE*, 85(8):1310–1322, 1997.

[3] Cecilia Aguerrebere, Mauricio Delbracio, Alberto Bartesaghi, and Guillermo Sapiro. Fundamental limits in multi-image alignment. *IEEE Transactions on Signal Processing*, 64(21):5707–5722, 2016.

[4] Ali Ahmed, Benjamin Recht, and Justin Romberg. Blind deconvolution using convex programming. *IEEE Transactions on Information Theory*, 60(3):1711–1732, 2014.

[5] GR Ayers and J Christopher Dainty. Iterative blind deconvolution method and its applications. *Optics letters*, 13(7):547–549, 1988.

[6] Afonso Bandeira, Philippe Rigollet, and Jonathan Weed. Optimal rates of estimation for multi-reference alignment. *arXiv preprint arXiv:1702.08546*, 2017.

[7] Afonso S Bandeira, Ben Blum-Smith, Joe Kileel, Amelia Perry, Jonathan Weed, and Alexander S Wein. Estimation under group actions: recovering orbits from invariants. *arXiv preprint arXiv:1712.10163*, 2017.

[8] Afonso S Bandeira, Moses Charikar, Amit Singer, and Andy Zhu. Multireference alignment using semidefinite programming. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 459–470. ACM, 2014.

[9] Alexander Barnett, Charles L Epstein, Leslie Greengard, and Jeremy Magland. Geometry of the phase retrieval problem. *arXiv preprint arXiv:1808.10747*, 2018.

[10] Alberto Bartesaghi, Alan Merk, Soojay Banerjee, Doreen Matthies, Xiongwu Wu, Jacqueline LS Milne, and Sriram Subramaniam. 2.2 å resolution cryo-em structure of β-galactosidase in complex with a cell-permeant inhibitor. *Science*, 348(6239):1147–1151, 2015.

[11] Robert Beinert and Gerlind Plonka. Ambiguities in one-dimensional discrete phase retrieval from Fourier magnitudes. *Journal of Fourier Analysis and Applications*, 21(6):1169–1198, 2015.

[12] Tamir Bendory, Robert Beinert, and Yonina C Eldar. Fourier phase retrieval: Uniqueness and algorithms. In *Compressed Sensing and its Applications*, pages 55–91. Springer, 2017.

[13] 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.

[14] Tamir Bendory, Nicolas Boumal, Chao Ma, Zhizhen Zhao, and Amit Singer. Bispectrum inversion with application to multireference alignment. *IEEE Transactions on Signal Processing*, 66(4):1037–1050, 2017.

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

[16] Nicolas Boumal, Tamir Bendory, Roy R Lederman, and Amit Singer. Heterogeneous multireference alignment: A single pass approach. In *Information Sciences and Systems (CISS), 2018 52nd Annual Conference on*, pages 1–6. IEEE, 2018.

[17] Veit Elser. Matrix product constraints by projection methods. *Journal of Global Optimization*, 68(2):329–355, 2017.

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

[19] Robert M Glaeser. Electron crystallography: present excitement, a nod to the past, anticipating the future. *Journal of structural biology*, 128(1):3–14, 1999.

[20] Sandeep Gogineni, Pawan Setlur, Muralidhar Rangaswamy, and Raj Rao Nadakuditi. Passive radar detection with noisy reference channel using principal subspace similarity. *IEEE Transactions on Aerospace and Electronic Systems*, 54(1):18–36, 2018.

[21] MHMH Hayes. The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 30(2):140–154, 1982.

[22] Richard Henderson. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Quarterly reviews of biophysics*, 28(2):171–193, 1995.

[23] Richard Henderson. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proceedings of the National Academy of Sciences*, 110(45):18037–18041, 2013.

[24] Stuart M Jefferies and Julian C Christou. Restoration of astronomical images by iterative blind deconvolution. *The Astrophysical Journal*, 415:862, 1993.

[25] Zvi Kam. The reconstruction of structure from electron micrographs of randomly oriented particles. *Journal of Theoretical Biology*, 82(1):15–39, 1980.

[26] Werner Kühlbrandt. The resolution revolution. *Science*, 343(6178):1443–1444, 2014.

[27] Han-Wen Kuo, Yenson Lau, Yuqian Zhang, and John Wright. Geometry and symmetry in short-and-sparse deconvolution. *arXiv preprint arXiv:1901.00256*, 2019.

[28] Ruslan P Kurta, Jeffrey J Donatelli, Chun Hong Yoon, Peter Berntsen, Johan Bielecki, Benedikt J Daurer, Hasan DeMirci, Petra Fromme, Max Felix Hantke, Filipe RNC Maia, et al. Correlations in scattered X-ray laser pulses reveal nanoscale structural features of viruses. *Physical review letters*, 119(15):158102, 2017.

[29] Kiryung Lee, Yanjun Li, Marius Junge, and Yoram Bresler. Blind recovery of sparse signals from subsampled convolution. *IEEE Transactions on Information Theory*, 63(2):802–821, 2017.

[30] Eitan Levin, Tamir Bendory, Nicolas Boumal, Joe Kileel, and Amit Singer. 3D ab initio modeling in cryo-EM by autocorrelation analysis. In *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*, pages 1569–1573. IEEE, 2018.

[31] Michael S Lewicki. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network: Computation in Neural Systems*, 9(4):R53–R78, 1998.

[32] Xiaodong Li, Shuyang Ling, Thomas Strohmer, and Ke Wei. Rapid, robust, and reliable blind deconvolution via nonconvex optimization. *Applied and Computational Harmonic Analysis*, 2018.

[33] Yanjun Li, Kiryung Lee, and Yoram Bresler. Identifiability in blind deconvolution with subspace or sparsity constraints. *IEEE Transactions on Information Theory*, 62(7):4266–4275, 2016.

[34] Shuyang Ling and Thomas Strohmer. Blind deconvolution meets blind demixing: Algorithms and performance bounds. *IEEE Transactions on Information Theory*, 63(7):4497–4520, 2017.

[35] Haiguang Liu, Billy K Poon, Dilano K Saldin, John CH Spence, and Peter H Zwart. Three-dimensional single-particle imaging using angular correlations from X-ray laser data. *Acta Crystallographica Section A: Foundations of Crystallography*, 69(4):365–373, 2013.

[36] Lennart Ljung. System identification. In *Signal analysis and prediction*, pages 163–173. Springer, 1998.

[37] Jianwei Miao, Pambos Charalambous, Janos Kirz, and David Sayre. Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. *Nature*, 400(6742):342, 1999.

[38] Amelia Perry, Jonathan Weed, Afonso Bandeira, Philippe Rigollet, and Amit Singer. The sample complexity of multi-reference alignment. *arXiv preprint arXiv:1707.00943*, 2017.

[39] Ofir Shalvi and Ehud Weinstein. New criteria for blind deconvolution of non-minimum phase systems (channels). *IEEE Transactions on information theory*, 36(2):312–321, 1990.

[40] Maxim Shatsky, Richard J Hall, Steven E Brenner, and Robert M Glaeser. A method for the alignment of heterogeneous macromolecules from electron microscopy. *Journal of structural biology*, 166(1):67–78, 2009.

[41] Yoav Shechtman, Yonina C Eldar, Oren Cohen, Henry Nicholas Chapman, Jianwei Miao, and Mordechai Segev. Phase retrieval with application to optical imaging: a contemporary overview. *IEEE signal processing magazine*, 32(3):87–109, 2015.

[42] Marin van Heel. Finding trimeric HIV-1 envelope glycoproteins in random noise. *Proceedings of the National Academy of Sciences*, 110(45):E4175–E4177, 2013.

[43] Marin van Heel, Michael Schatz, and Elena Orlova. Correlation functions revisited. *Ultramicroscopy*, 46(1–4):307–316, 1992.

[44] Benjamin von Ardenne, Martin Mechelke, and Helmut Grubmüller. Structure determination from single molecule X-ray scattering with three photons per image. *Nature communications*, 9(1):2375, 2018.