Toward single particle reconstruction without particle picking

Tamir Bendory, Nicolas Boumal, William Leeb and Amit Singer

May 17, 2018

Abstract

Here comes the abstract

1 Introduction

Single particle reconstruction (SPR) using cryo–electron microscopy (cryo–EM) is an innovative technology for reconstructing the 3-D structure of macromolecules. In recent years, structures of many molecules, previously regarded as insurmountable, are now being reconstructed to near-atomic resolution; see for instance [32, 8, 45]. This technological advancement was recognized by the 2017 Nobel Prize in Chemistry [1].

In a cryo–EM experiment, multiple biological samples of the (ideally) same molecule are rapidly frozen in a thin layer of vitreous ice. Within the ice, the molecules are randomly oriented and positioned. Then, the microscope produces a 2D tomographic projection image, called a *micrograph*, of the multiple samples embedded in the ice. The micrograph is dominated by noise due to the small electron doses that can be applied to the specimen without causing radiation damage. The cryo–EM problem is to estimate the structure of the molecule from the micrograph (or, typically, several micrographs).

Particle picking. All contemporary methods in the field split the reconstruction procedure into several subroutines. The first step in the pipeline is known as *particle picking*, in which one identifies the locations of the 2-D tomographic projections of the samples embedded within the noisy micrograph. The output of an ideal particle picker is a series of 2-D images, each containing one centered tomographic projection associated with an unknown viewing direction and an in-plane rotation. This series of images is then used to estimate the structure.

Many automatic and semi-automatic methods for particle picking have been proposed, based on edge detection, template matching and deep learning; see for instance [24, 38, 49, 21, 41, 27]. However, most of these procedures are prone to *model bias*. For instance, in the popular framework of RELION [41], the user manually marks hundreds of spots on the micrograph, believed to contain projections. Therefore, the algorithm's performance depends on the prior assumptions of the users about the particle's structure; the same holds true for deep learning based approaches which require constructing labeled sets of data. Other methods use disks or differences of Gaussians as templates [33, 47]. Nowadays, it is also still popular to pick particles manually. This method, while it exploits the researcher's experience, is both tedious and subject to model bias.

The significance of model bias in cryo–EM was stressed in [43, 28] and demonstrated by the "Einstein from noise" example. In this experiment, the image of Einstein is correlated with multiple images of pure white noise. The noisy images are aligned to Einstein's image using cross-correlation and then averaged. In [43], it was shown that with merely 1000 such noisy images, the averaged image is remarkably similar to the template image of Einstein rather than to pure noise. In the context of cryo–EM, this example demonstrates how our prior assumptions about the particle may influence the reconstructed structure. This is true in general for any method which is based on template matching. [TB: Do we have a nice experiment to exemplify the phenomenon?]

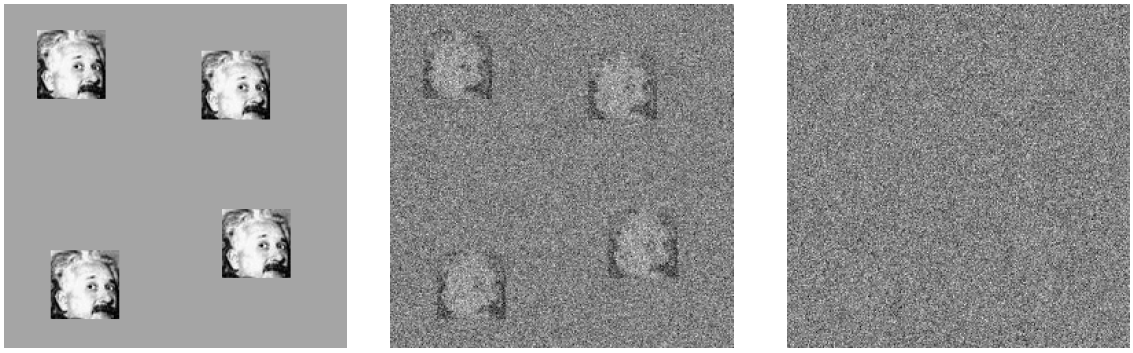
Besides the model bias, particle pickers have several limitations due to the high noise level, in particular for small particles and in low contrast. As a result, the projections in the 2D images are typically not centered; including the correct centers as parameters dramatically increases the complexity of reconstruction algorithms.

Simplified model for SPR. In this paper, we present a preliminary study that aims to show that estimating a particle directly from the micrograph, without identifying the location of individual particle images in the micrograph, might be possible.

In particular, we consider the toy problem of estimating a set of signals $x_1, \ldots, x_K$ from their multiple occurrences in unknown locations in a noisy data sequence y. Here, y plays the role of the micrograph and the K signals can be thought of as K different viewing directions of the particle, each of which appears multiple times in the micrograph. A precise mathematical formulation of the model and the estimation problem is provided in Section 4. To be clear, we do not consider here many prominent features of real SPR experiments and do not aim to reconstruct any 3-D structure; instead, we solve a simpler problem that we believe captures key elements of the true SPR problem. We also mention in passing that similar models emerge in different scientific fields, such as spike sorting [35], passive radar [23] and system identification [37].

Figure 1 shows a simple example of our problem with one signal ($K = 1$). Each panel demonstrates an image of size 250×250, containing 4 occurrences of an Einstein image of size 50×50. The left panel shows the data without noise. In the middle panel, i.i.d. Gaussian noise with standard deviation $\sigma = 0.5$ was added. At this noise level, Einstein's locations are still easily identified. In this scenario it is likely that standard detection algorithms can locate the copies of the signal within the micrograph. In the right panel, we show the same data swamped in i.i.d. Gaussian noise with standard deviation $\sigma = 3$. Clearly, identifying the locations of Einstein in this micrograph is much more challenging. In fact, it can be shown that in the low signal–to–noise (SNR) regime, detection of individual image occurrences is impossible—even if the true image is known—and therefore particle picking is impossible. This phenomenon is explained in detail in Appendix C. [WL: I think this example should go in the previous section on particle picking. The main issue we're trying to overcome is when detection is impossible; the fact that we don't use a template seems to be an additional advantage of our method, but not the main selling point. So I think the impossibility of detection in sufficiently low SNR must be emphasized earlier.]

Autocorrelation analysis. We now return to the general problem of estimating $x_1, \ldots, x_K$ from the noisy observation y. In order to recover the signals $x_1, \ldots, x_K$ from the data, we use



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

RR

Figure 1: Example of data sequences (micrographs) of size 250×250 with additive i.i.d. Gaussian noise with variance σ^2. Each micrograph contains four occurrences of a 50×50 image of Einstein.

autocorrelation analysis. The method consists of two stages. First, we estimate a mixture (i.e., linear combination) of the low–order autocorrelation functions of the signals from the data. These quantities can be estimated, to any desired accuracy, if individual occurrences are separated by the maximum signal size and each signal appears sufficiently many times in the data. There is no need to detect individual occurrences. In the second stage, the signals are estimated from the mixed autocorrelations using a nonconvex least-squares (LS) algorithm or a phase retrieval algorithm; see Section 4 for details. This method requires only one pass over the data and can recover the signals in any SNR, if the signals appear enough times in the data. As a side note, we mention that expectation-maximization—a popular framework in SPR—is intractable for this problem (see Appendices B and C for more details).

2 Results

In this section we describe some numerical experiments and show results. The experimental details and the algorithms are discussed at length in Section 4.

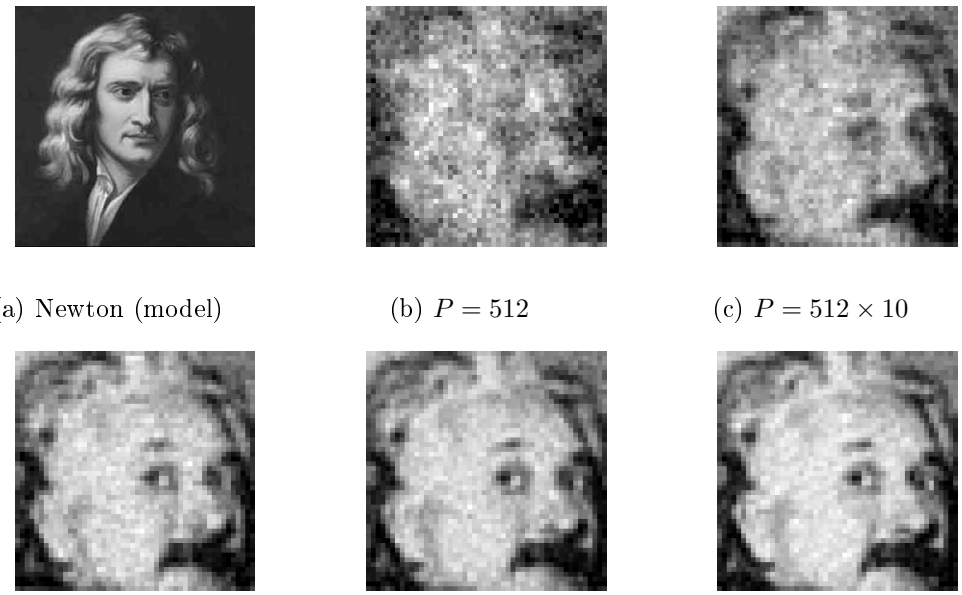
In the first experiment, we estimated Einstein's image of size 50×50 from multiple micrographs of size 4096×4096 pixels, contaminated with additive i.i.d. Gaussian noise with standard deviation $\sigma = 3$ ($\mathsf{SNR} = 1/20$). An illustrative example of a micrograph appears in the right panel of Figure 1. Each micrograph contains multiple occurrences of Einstein's image in random locations, while keeping a separation of at least 49 pixels between the images in each direction. From the micrographs, we estimated the signal's power spectrum, which is equivalent to the second-order autocorrelation. Then, in order to estimate the signal itself, we applied a standard phase retrieval algorithm called relaxed-reflect-reflect (RRR) [20].

As an initial guess, we picked an image of the physicist Issac Newton; see Figure 2a. If the algorithm was prone to model bias, we would expect to get as an output an image that resembles Newton, similar to the "Einstein from noise" effect. However, the experiment exhibits the desired phenomenon: the more data we collect, the better the reconstruction quality.

Figure 2 demonstrates several recovery results for different number of recorded micrographs. Figure 3 presents the normalized recovery error and the power spectrum estimation error as a function of the number of micrographs. To measure the error, we use the root mean square error (RMSE) defined as

$$\text{RMSE} := \frac{\|x - \hat{x}\|_{\text{F}}}{\|x\|_{\text{F}}}, \tag{2.1}$$

where x and $\hat{x}$ are, respectively, the underlying and estimated image, and $\|\cdot\|_{\text{F}}$ stands for the Frobenius norm. As expected, the RMSE of estimating the power spectrum decreases linearly with slope $-1/2$ in logarithmic scale. [TB: Few comments: 1. We have a movie in the supplementary material. 2. Does the error also have the right slope? We'll know once we'll gather enough micrographs 3. Need to improve the micrograph's generation code]



(a) Newton (model) (b) $P = 512$ (c) $P = 512 \times 10$

(d) $P = 512 \times 100$ (e) $P = 512 \times 200$ (f) $P = 512 \times 840$

Figure 2: Recovery of Einstein's image. The RRR algorithm is initialized with the image of the physicist Issac Newton in the upper-left panel. We show estimations of Einstein using P number of micrographs, each containing 700 image occurrences on average with an SNR of 1/20. The signal's estimation RMSE is $0.5874, 0.3731, 0.2517, 0.2137, 0.1534$. The power spectrum estimation RMSE is $0.0381, 0.012, 0.0038, 0.0027, 0.0013$.

Results from a similar experiment with three ($K = 3$) one-dimensional signals are presented in Figure 4. We define the ratio of the space occupied by the ith signal as

$$\gamma_i = \frac{M_i L}{N}, \tag{2.2}$$

where M_i is the number of times signal x_i occurs, L is the length of each signal and N is the micrograph length. These ratios γ_i represent the density of the the signals in the data. In the 1-D experiment, we do not assume knowledge of either the ratios γ_i or the noise level σ.

In order to estimate the signals, we compute the first three autocorrelation functions of the data and then estimate the signals and their corresponding γ_i using a nonconvex least-squares. As can be seen, given enough signal occurrences, we can estimate the signals accurately. The estimation quality of the triangle signal is poorer than the other two signals, a phenomenon that is explained using Proposition C.1. [TB: 1. Figure 4 will be replaced with a "progress figure" (like we did for Einstien) for each signals and a plot of the recovery error for all three signals. 2. We may want to replace the triangle signal]

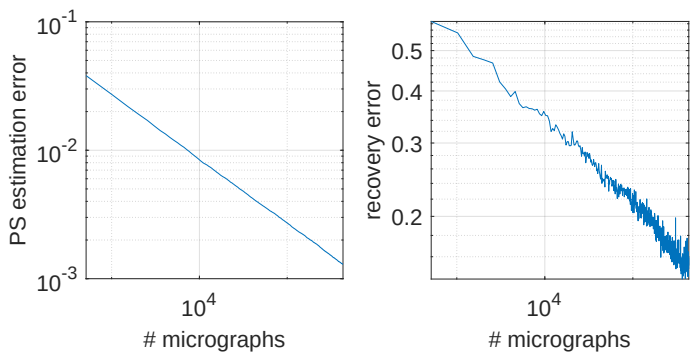


Figure 3: The left panel shows the RMSE of estimating the power spectrum of the signal from the data as a function of the number of collected micrographs. As expected, the curve is linear with slope $-1/2$ in logarithmic scale. The right panel shows the RMSE (in logarithmic scale) of recovering Einstein's image. [TB: The program continues running so these figures will be extended to the right (more micrographs).]

3 Discussion

All current algorithmic pipelines for SPR using cryo–EM start with a particle picking stage to detect the 2-D tomographic projections of the samples in the micrograph. The particle picking algorithms tend to be prone to model bias and cannot work reliably on small particles in low SNR environment. Bypassing the particle picking stage and constructing a 3-D model directly from the data—without assuming prior knowledge about the particle to be estimated—can be used to reconstruct ab initio models to initialize a refinement algorithm. Alternatively, it can be applied to generate templates for a particle picker which does not suffer from model bias.

In this paper, we examined a simplified model that, we believe, captures important features of the SPR problem. While the field is currently dominated by Bayesian methods such as EM, they are intractable for such problems. As an alternative, we propose to use an autocorrelation analysis technique. Methods based on autocorrelation analysis require only one pass over the

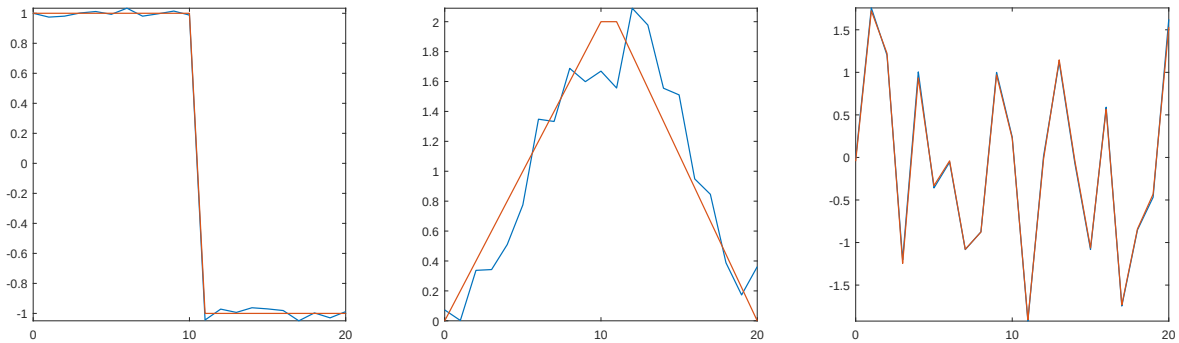


Figure 4: An experiment for estimating three one-dimensional signals simultaneously with noise level $\sigma = 3$. Red signals are the ground truth (targets) and the blue signals are our estimations. Individual RMSE of the estimates: 0.0239393/0.208925/0.0335956. The estimated γ's are 0.02574/0.01693/0.00818 and the true ones 0.02561/0.01707/0.00853. The SNR is 1/12. [TB: 1. We will replace the figure with "progress" plots 2. Need to replace the triangle signal 3. Need to add error curves]

data and therefore are usually computationally efficient. [TB: Do we want to mention Kam's method?] That being said, the SPR model is far more complicated than the model presented here. In future research, we hope to bridge this gap.

Our results rely on two core assumptions that are not necessarily met by an SPR experiment. First, we modeled the background information as i.i.d. additive noise. In practice, the background information may be structured or signal dependent. Second, we assumed that the signal occurrences are all separated by the length of the signal; see (4.2). This separation can be induced by careful experimental design [TB: ???]. If the signals are not separated, one can introduce additional parameters that model the distribution of the spacing between signal occurrences and then compute explicitly the relation between the autocorrelation functions of the data and the signals. It is yet to be studied under what conditions on this distribution one can estimate the signals.

4 Methods

Mathematical model. Let $x_1, \ldots, x_K \in \mathbb{R}^L$ be the target signals and let $y \in \mathbb{R}^N$ be the observed data. The forward model can be posed as a mixture of *blind deconvolution* problems between binary signals and the target signals x_i:

$$y = \sum_{i=1}^{K} x_i * s_i + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I). \tag{4.1}$$

The nonzero values of each $s_i \in \{0,1\}^N$ determine the position of x_i's occurrences. We denote the set of these nonzero values by $\mathcal{S}_i$ and its cardinality by $|\mathcal{S}_i| = M_i$. By assuming that all $\mathcal{S}_i$'s are disjoint, we let $s = \sum_{i=1}^K s_i$, $\mathcal{S} = \bigcup_{i=1}^K \mathcal{S}_i$ and $|\mathcal{S}| := M = \sum_{i=1}^K M_i$. Literature survey on blind deconvolution and related problems is given in Appendix B.

In order to estimate the mixture of autocorrelations, we assume that the support of s is

not clustered. In particular, we assume that

$$|i-j| \geq L-1, \quad \text{for all } i,j \in \mathcal{S} \text{ such that } i \neq j. \tag{4.2}$$

The goal of the problem is to estimate $x_1, \ldots, x_K$ from y.

Aperiodic autocorrelation functions. For $z \in \mathbb{R}^L$ and $k \geq 2$, the autocorrelation of order k is defined for any integer shifts $\ell_1, \ldots, \ell_{k-1}$ by

$$a_z^k[\ell_1, \ldots, \ell_{k-1}] = \sum_{i=-\infty}^{+\infty} z[i]z[i+\ell_1]\ldots z[i+\ell_{k-1}], \tag{4.3}$$

where the values of z outside the indices $0, \ldots, L-1$ are taken to be zero. Explicitly, the mean (which can be thought of as a first-order autocorrelation) and the second and third autocorrelations are given by

$$\begin{aligned} a_z^1 &= \sum_{i=0}^{L-1} z[i], \\ a_z^2[\ell] &= \sum_{i=\max\{0,-\ell\}}^{L-1+\min\{0,-\ell\}} z[i]z[i+\ell], \\ a_z^3[\ell_1,\ell_2] &= \sum_{i=\max\{0,-\ell_1,-\ell_2\}}^{L-1+\min\{0,-\ell_1,-\ell_2\}} z[i]z[i+\ell_1]z[i+\ell_2]. \end{aligned} \tag{4.4}$$

Note that the autocorrelation functions are symmetric so that $a_z^2[\ell] = a_z^2[-\ell]$ and also

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

Estimating the autocorrelation function of a single signal. We first consider the problem of estimating the autocorrelations of a single signal from the data. The main principles carry through for $K > 1$, as will be shown next.

For the purpose of the analysis, we consider the asymptotic regime where M grows proportionally to N. Specifically, we define the ratio of the measurement occupied by the signal as

$$\gamma = \frac{ML}{N}. \tag{4.5}$$

[WL: γ should really be the limit of this ratio as N grows to infinity; for any finite value of N this ratio may be different. TB: The accurate definition is given in (A.1). We may want to leave it as is here.] Under the spacing constraint (4.2), we have $\gamma \leq \frac{L}{2L-1} \approx 1/2$.

The main pillar of this work is the following simple observation. If the support signal s satisfies the spacing constraint (4.2), then the first L entries of the data autocorrelations converge to a scaled, biased, version of the signal's autocorrelation:

$$\begin{aligned} \lim_{N\to\infty} a_y^1 &= \gamma a_x^1, \\ \lim_{N\to\infty} a_y^2[\ell] &= \gamma a_x^2[\ell] + \sigma^2\delta[\ell], \\ \lim_{N\to\infty} a_y^3[\ell_1,\ell_2] &= \gamma a_x^3[\ell_1,\ell_2] + \sigma^2\gamma a_x^1\big(\delta[\ell_1,0] + \delta[0,\ell_2] + \delta[\ell_1,\ell_2]\big), \end{aligned} \tag{4.6}$$

for $\ell, \ell_1, \ell_2 = 0, \ldots L-1$, and where δ denotes the Kronecker delta function.

Estimating the autocorrelation function of a multiple signals. As before, we consider the asymptomatic regime where $M_1, \ldots, M_K$ grow proportional to N and define

$$\gamma_k = \frac{M_k L}{N}, \quad \gamma = \sum_{k=1}^{K} \gamma_k. \tag{4.7}$$

If the support s satisfies the spacing constraint (4.2), then similarly to (4.6) one can estimate the mixture of the K signals' autocorrelations:

$$\begin{aligned}
\lim_{N\to\infty} a_y^1 &= \sum_{k=1}^{K} \gamma_k a_{x_k}^1, \\
\lim_{N\to\infty} a_y^2[\ell] &= \sum_{k=1}^{K} \gamma_k a_{x_k}^2[\ell] + \sigma^2 \delta[\ell], \\
\lim_{N\to\infty} a_y^3[\ell_1, \ell_2] &= \sum_{k=1}^{K} \gamma_k a_{x_k}^3[\ell_1, \ell_2] + \sigma^2 \left(\sum_{k=1}^{K} \gamma_k a_{x_k}^1 \right) (\delta[\ell_1, 0] + \delta[0, \ell_2] + \delta[\ell_1, \ell_2]),
\end{aligned} \tag{4.8}$$

This relation is proven in Appendix A.

Numerical experiments details. For the 2-D experiment shown in Figures 2 and 3, we generated P micrographs of size 4096×4096 pixels. In each micrograph, we placed Einstein's image of size 50×50 in random locations, while keeping the separation condition (4.2). This is done by randomly selecting 4000 placements in the micrograph, one at a time with an accept/reject rule based on the separation constraint and locations picked so far. On average, 700 images were placed in each micrograph. Then, an i.i.d. Gaussian noise with standard deviation $\sigma = 3$ was added, inducing an SNR of approximately 1/20. A piece of one such micrograph is presented in the right panel of Figure 1.

In this experiment, we assume the knowledge of the number of Einstein's occurrences and the noise level. In this situation, the second-order autocorrelation (or, equivalently, the Fourier magnitudes of the signal) determines the image almost surely up to reflection through the origin [25, 11]. Therefore, we estimate the signal's Fourier magnitude from the Fourier magnitude of the micrographs, at the cost of one FFT per micrograph. To estimate the signal, we use a phase retrieval algorithm called relaxed-reflect-reflect (RRR) that iterates by [20]

$$z \to z + P_2(2P_1(z) - z) - P_1(z). \tag{4.9}$$

The RRR algorithm estimates a $2L \times 2L$ image and the exact solution contains Einstein's image in the upper-left corner and zero elsewhere. The operator $P_2(z)$ combines the Fourier phases of the current estimation z with the Fourier magnitudes of the signal (estimated from the data). The operator $P_1(z)$ zeros out all entries of z outside the $L \times L$ upper-left corner. In order to compare the performance across multiple cases and different noise levels, the algorithm stopped after a fixed number of 1000 iterations and the iteration with the smallest error compared to the ground truth (also taking into account the reflection ambiguity) was chosen as the solution. While this cannot be done in practice since we do not have access to the ground truth, this procedure enables us to compare a large number of instances in different noisy environments [TB: Note the last two sentences!].

For the 1-D experiment, we worked with three signals of length $L = 21$ and generated the data in the same way as in the 2-D example. The only difference is that here, for each placement, one of the three signals is picked at random proportionally to the desired number of occurrences of each. In this experiment, we computed the first three autocorrelation functions. We do not assume to know the number of occurrences of each signal γ_i ahead and we removed the terms depending on σ (see (4.6)) so we do not need to know σ.

To estimate the signal, we employ an optimization algorithm on the following nonlinear least-squares problem that estimates the signals and their number of occurrences simultaneously:

$$\min_{\substack{\hat{x}_1,\ldots,\hat{x}_K \in \mathbb{R}^W \\ \hat{\gamma}_1,\ldots,\hat{\gamma}_K > 0}} w_1 \left(a_y^1 - \sum_{k=1}^{K} \hat{\gamma}_k a_{\hat{x}_k}^1 \right)^2 + w_2 \sum_{\ell=1}^{L-1} \left(a_y^2[\ell] - \sum_{k=1}^{K} \hat{\gamma}_k a_{\hat{x}_k}^2[\ell] \right)^2 + w_3 \sum_{\substack{2 \le \ell_1 \le L-1 \\ 1 \le \ell_2 \le \ell_1 - 1}} \left(a_y^3[\ell_1, \ell_2] - \sum_{k=1}^{K} \hat{\gamma}_k a_{\hat{x}_k}^3[\ell_1, \ell_2] \right)^2, \quad (4.10)$$

where $W \geq L$ is the length of the sought signals and $w_1 = \frac{1}{2}, w_2 = \frac{1}{2}(L-1), w_3 = \frac{1}{(L-1)(L-2)}$; see [17] see for discussion on how to choose the weights properly.

Setting $W = L$ 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 fewer 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 whose autocorrelation functions are the closest to the measured ones in the sense of (4.10). This estimator is then used as initial iterate for (4.10), this time with $W = L$. We find that this procedure is reliable for a wide range of experimental parameters. To solve (4.10), we run the trust-region method implemented in Manopt [16], which allows enforcement of the positivity constraints [reference] on the coefficients $\hat{\gamma}_k$. Notice that the cost function is a polynomial in the variables, so that it is straightforward to evaluate it and its derivatives.

References

[1] `https://www.nobelprize.org/nobel_prizes/chemistry/laureates/2017/`.

[2] Emmanuel Abbe, Tamir Bendory, William Leeb, João Pereira, Nir Sharon, and Amit Singer. Multireference alignment is easier with an aperiodic translation distribution. *arXiv preprint arXiv:1710.02793*, 2017.

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

[4] 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.

[5] Christophe Andrieu, Éric Barat, and Arnaud Doucet. Bayesian deconvolution of noisy filtered point processes. *IEEE Transactions on Signal Processing*, 49(1):134–146, 2001.

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

[7] Jean-Marc Azais, Yohann De Castro, and Fabrice Gamboa. Spike detection from inaccurate samplings. *Applied and Computational Harmonic Analysis*, 38(2):177–195, 2015.

[8] 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.

[9] 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.

[10] Tamir Bendory. Robust recovery of positive stream of pulses. *IEEE Transactions on Signal Processing*, 65(8):2114–2122, 2017.

[11] 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.

[12] Tamir Bendory, Nicolas Boumal, Chao Ma, Zhizhen Zhao, and Amit Singer. Bispectrum inversion with application to multireference alignment. *arXiv preprint arXiv:1705.00641*, 2017.

[13] Tamir Bendory, Shai Dekel, and Arie Feuer. Robust recovery of stream of pulses using convex optimization. *Journal of Mathematical Analysis and Applications*, 442(2):511–536, 2016.

[14] Albert Benveniste, Maurice Goursat, and Gabriel Ruget. Robust identification of a nonminimum phase system: Blind adjustment of a linear equalizer in data communications. *IEEE Transactions on Automatic Control*, 25(3):385–399, 1980.

[15] Brett Bernstein and Carlos Fernandez-Granda. Deconvolution of point sources: A sampling theorem and robustness guarantees. *arXiv preprint arXiv:1707.00808*, 2017.

[16] 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.

[17] Nicolas Boumal, Tamir Bendory, Roy R Lederman, and Amit Singer. Heterogeneous multireference alignment: a single pass approach. *arXiv preprint arXiv:1710.02590*, 2017.

[18] Olivier Cappé, Arnaud Doucet, Marc Lavielle, and Eric Moulines. Simulation-based methods for blind maximum-likelihood filter identification. *Signal processing*, 73(1-2):3–25, 1999.

[19] Quentin Denoyelle, Vincent Duval, and Gabriel Peyré. Support recovery for sparse super-resolution of positive measures. *Journal of Fourier Analysis and Applications*, 23(5):1153–1194, 2017.

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

[21] Joachim Frank and Terence Wagenknecht. Automatic selection of molecular images from electron micrographs. *Ultramicroscopy*, 12(3):169–175, 1983.

[22] Georgios B Giannakis and Jerry M Mendel. Identification of nonminimum phase systems using higher order statistics. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(3):360–377, 1989.

[23] 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*, 2017.

[24] George Harauz and Amelia Fong-Lochovsky. Automatic selection of macromolecules from electron micrographs by component labelling and symbolic processing. *Ultramicroscopy*, 31(4):333–344, 1989.

[25] 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.

[26] Monson H Hayes and James H McClellan. Reducible polynomials in more than one variable. *Proceedings of the IEEE*, 70(2):197–198, 1982.

[27] Ayelet Heimowitz, Amit Singer, et al. Apple picker: Automatic particle picking, a low-effort cryo-em framework. *arXiv preprint arXiv:1802.00469*, 2018.

[28] 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.

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

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

[31] John Kormylo and J Mendel. Identifiability of nonminimum phase linear stochastic systems. *IEEE transactions on automatic control*, 28(12):1081–1090, 1983.

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

[33] Robert Langlois, Jesper Pallesen, Jordan T Ash, Danny Nam Ho, John L Rubinstein, and Joachim Frank. Automated particle picking for low-contrast macromolecules in cryo-electron microscopy. *Journal of structural biology*, 186(1):1–7, 2014.

[34] Eitan Levin, Tamir Bendory, Nicolas Boumal, Joe Kileel, and Amit Singer. 3D ab initio modeling in cryo-em by autocorrelation analysis. *arXiv preprint arXiv:1710.08076*, 2017.

[35] 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.

[36] KS Lii, M Rosenblatt, et al. Deconvolution and estimation of transfer function phase and coefficients for nongaussian linear processes. *The annals of statistics*, 10(4):1195–1208, 1982.

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

[38] Toshihiko Ogura and Chikara Sato. Automatic particle pickup method using a neural network has high accuracy by applying an initial weight derived from eigenimages: a new reference free method for single-particle analysis. *Journal of structural biology*, 145(1-2):63–75, 2004.

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

[40] Lawrence R Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[41] Sjors HW Scheres. Semi-automated selection of cryo-em particles in relion-1.3. *Journal of structural biology*, 189(2):114–122, 2015.

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

[43] 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.

[44] Amit Singer. Mathematics for cryo-electron microscopy. *arXiv preprint arXiv:1803.06714*, 2018.

[45] Martin TJ Smith and John L Rubinstein. Beyond blob-ology. *Science*, 345(6197):617–619, 2014.

[46] Jitendra Tugnait. Identification of nonminimum phase linear stochastic systems. In *The 23rd IEEE Conference on Decision and Control*, number 23, pages 342–347, 1984.

[47] NR Voss, CK Yoshioka, M Radermacher, CS Potter, and B Carragher. Dog picker and tiltpicker: software tools to facilitate particle selection in single particle electron microscopy. *Journal of structural biology*, 166(2):205–213, 2009.

[48] Alex Wein. *Statistical Estimation in the Presence of Group Actions*. PhD thesis, 2018.

[49] Yanan Zhu, Qi Ouyang, and Youdong Mao. A deep learning approach to single-particle recognition in cryo-electron microscopy. *arXiv preprint arXiv:1605.05543*, 2016.

A Proof of (4.8)

Throughout the proof, we consider the case of one signal $K = 1$. The extension to $K > 1$ is straightforward by averaging the contributions of all signal with appropriate weights; see [17].

We will let the number of instances of the signal M grow with N, and write $M = M_N$ to emphasize this. We assume M_N grows proportionally with N, and define:

$$\gamma = \lim_{N\to\infty} \frac{M_N L}{N} < 1. \tag{A.1}$$

We will assume that $M_N = \Omega(N)$, so that $\gamma > 0$. In the sequel, we will suppress the explicit dependence of M on N for notational convenience.

We start by considering the mean of the data:

$$a_y^1 = \frac{1}{N}\sum_{i=0}^{N-1} y[i] = \frac{1}{N/L}\sum_{j=0}^{M-1}\frac{1}{L}\sum_{i=0}^{L-1} x[i] + \underbrace{\frac{1}{N}\sum_{i=0}^{N-1}\varepsilon[i]}_{\text{noise term}} \xrightarrow{a.s.} \gamma a_x^1, \tag{A.2}$$

where the noise term converges to zero almost surely (a.s.) by the strong law of large numbers.

We proceed with the (second-order) autocorrelation for fixed $\ell \in [0, \ldots, L-1]$. We can compute:

$$\begin{aligned} a_y^2[\ell] &= \frac{1}{N}\sum_{i=0}^{N-1-\ell} y[i]y[i+\ell] \\ &= \underbrace{\frac{1}{N}\sum_{j=1}^{M}\sum_{i=0}^{L-\ell-1} x[i]x[i+\ell]}_{\text{signal term}} + \underbrace{\frac{1}{N}\sum_{i=0}^{N-1-\ell}\varepsilon[i]\varepsilon[i+\ell]}_{\text{noise term}} + \underbrace{\frac{1}{N}\sum_{j=1}^{M}\sum_{i=0}^{L-1} x[i]\varepsilon[s_j+i+\ell]}_{\text{cross-term}}. \end{aligned} \tag{A.3}$$

The cross-terms are linear in the noise, and are easily shown to vanish almost surely in the limit $N \to \infty$, by the strong law of large numbers. As for the signal term, we break it into M different sums, each containing one copy of the signal. This gives:

$$\frac{1}{N}\sum_{j=1}^{M}\sum_{i=0}^{L-\ell-1} x[i]x[i+\ell] = \frac{ML}{N}\frac{1}{L}\sum_{i=0}^{L-\ell-1} x[i]x[i+\ell] \xrightarrow{N\to\infty} \gamma a_x^2[\ell]. \tag{A.4}$$

We next analyze the pure noise term. When $\ell \neq 0$, we can break the noise term into a sum of independent terms:

$$\frac{1}{N}\sum_{i=0}^{N-1-\ell}\varepsilon[i]\varepsilon[i+\ell] = \frac{1}{\ell}\sum_{i=0}^{\ell-1}\frac{1}{N/\ell}\sum_{j=0}^{N/\ell-1}\varepsilon[j\ell+i]\varepsilon[(j+1)\ell+i]. \tag{A.5}$$

Each sum $\frac{1}{N/\ell}\sum_{j=0}^{N/\ell-1}\varepsilon[j\ell+i]\varepsilon[(j+1)\ell+i]$ is an average of N/ℓ independent terms with expectation zero, and thus converges to zero almost surely as $N\to\infty$. If $\ell=0$, then we have:

$$\frac{1}{N}\sum_{i=0}^{N-1}\varepsilon^2[i]\xrightarrow{a.s.}\sigma^2. \tag{A.6}$$

We now analyze the third-order autocorrelation. Let us fix $\ell_1\geq\ell_2\geq 0$. We have:

$$\begin{aligned}
a_y^3[\ell_1,\ell_2] &= \frac{1}{N}\sum_{i=0}^{N-1-\ell_1} y[i]y[i+\ell_1]y[i+\ell_2] \\
&= \underbrace{\frac{ML}{N}\frac{1}{M}\sum_{j=1}^{M}\frac{1}{L}\sum_{i=0}^{L-1-\ell_1} x[i]x[i+\ell_1]x[i+\ell_2]}_{(1)} + \underbrace{\frac{1}{N}\sum_{i=0}^{N-1-\ell_1}\varepsilon[i]\varepsilon[i+\ell_1]\varepsilon[i+\ell_2]}_{(2)} \\
&\quad + \underbrace{\frac{1}{N}\sum_{j=1}^{M}\sum_{i=0}^{L-1} x[i]\varepsilon[s_j+i+\ell_1]\varepsilon[s_j+i+\ell_2]}_{(3)} + \underbrace{\frac{1}{N}\sum_{j=1}^{M}\sum_{i=0}^{L-1}\varepsilon[s_j+i-\ell_1]x[i]\varepsilon[s_j+i+\ell_2-\ell_1]}_{(4)} \\
&\quad + \underbrace{\frac{1}{N}\sum_{j=1}^{M}\sum_{i=0}^{L-1}\varepsilon[s_j+i-\ell_2]\varepsilon[s_j+i+\ell_1-\ell_2]x[i]}_{(5)} + \underbrace{\frac{1}{N}\sum_{j=1}^{M}\sum_{i=0}^{L-\ell_1+\ell_2-1}\varepsilon[s_j+i]x[i+\ell_1-\ell_2]x[i]}_{(6)} \\
&\quad + \underbrace{\frac{1}{N}\sum_{j=1}^{M}\sum_{i=0}^{L-\ell_2-1} x[i]\varepsilon[s_j+i+\ell_1]x[s_j+i+\ell_2]}_{(7)} + \underbrace{\frac{1}{N}\sum_{j=1}^{M}\sum_{i=0}^{L-\ell_1-1} x[i]x[i+\ell_1]\varepsilon[s_j+i+\ell_2]}_{(8)}.
\end{aligned} \tag{A.7}$$

Terms (6), (7) and (8) are linear in ε, and can easily be shown to converge to 0 almost surely by the law of large numbers, by similar arguments as used previously. Term (1) converges to $\gamma a_x^3[\ell_1,\ell_2]$ almost surely, for the same reasons as (A.4). To deal with terms (2)–(5), we must distinguish between different values of ℓ_1 and ℓ_1.

Case 1: $0<\ell_2<\ell_1$. Here, all summands with elements of ε involve products of distinct entries, which have expected value 0. Consequently, the usual argument shows that terms (2)–(5) all converge to 0 almost surely as $N\to\infty$.

Case 2: $0=\ell_2<\ell_1$. Term (2) is an average of products of the form $\varepsilon[i]^2\varepsilon[i+\ell_1]$, which have mean zero; consequently, term (2) converges to 0 almost surely. The same argument as for Case 1 shows that (3) and (5) also converge to 0. For term (4), we write:

$$\begin{aligned}
&\frac{1}{N}\sum_{j=1}^{M}\sum_{i=0}^{L-1}\varepsilon[s_j+i-\ell_1]x[i]\varepsilon[s_j+i+\ell_2-\ell_1] \\
&= \frac{ML}{N}\frac{1}{L}\sum_{i=0}^{L-1}x[i]\frac{1}{M}\sum_{j=1}^{M}\varepsilon[s_j+i-\ell_1]^2 \xrightarrow{N\to\infty} \gamma\frac{1}{L}\sum_{i=0}^{L-1}x[i]\sigma^2 = \gamma a_x^1\sigma^2.
\end{aligned} \tag{A.8}$$

Case 3: $0 < \ell_2 = \ell_1$. An argument nearly identical to that for Case 2 shows that terms (2), (4) and (5) converge to 0, while term (3) converges to $\gamma a_x^1 \sigma^2$.

Case 4: $0 = \ell_2 = \ell_1$. The same argument as for term (4) in Case 2 shows that terms (3), (4) and (5) all converge to $\gamma a_x^1 \sigma^2$. Term (2) is an average of $\varepsilon[i]^3$, which is mean zero; consequently, it converges to 0.

This completes the proof of (4.8).

B Related work

System identification. For $K = 1$, our problem can be interpreted as a special case of the system identification problem. Similarly to (4.1), the system identification forward model takes the form $y = x * w + \varepsilon$, where x is the unknown signal (the "system"), w is an unknown, random, input sequence, and ε is an additive noise. The goal of this problem is to estimate x, usually referred to as "identifying the system." The question of identifiability of x under this observation model is addressed for certain Gaussian and non-Gaussian w in [14, 31]. In the special case where $w \in \{0,1\}^N$, satisfying the spacing requirement (4.2), we obtain our model in the case of a single signal ($K = 1$). The same observation model is used for blind deconvolution, a longstanding problem arising in a variety of engineering and scientific applications such as astronomy, communication, image deblurring, system identification and optics; see [29, 42, 6, 3], just to name a few.

Likelihood-based methods. Likelihood-based methods estimate x as the maximizer of some function $f(x|y)$, where f is derived from the likelihood function of x given the observed signal y. For example, f may be the likelihood itself, or a related function with a similar form (leading to the class of "quasi-likelihood" methods). If some prior is assumed on x, then $f(x|y)$ can be taken to be the posterior distribution of x given the data; this is the simplest form of Bayesian inference.

Optimizing the function $f(x|y)$ exactly is often intractable, and thus heuristic methods are used instead. One proposed technique is to use Markov Chain Monte Carlo (MCMC) [18]. Another paper considers parameterized models for multiple distinct signals, as in our framework ($K > 1$) [5]. Their proposed solution is an MCMC algorithm tailored for their specific parametrized problem.

In special cases, including the case where w is binary, expectation maximization (EM) has been used [18]. The EM method for discrete w is based upon a certain "forward-backward" procedure used in hidden Markov models [40]. However, the complexity of this procedure is nonlinear in N, and therefore its usage is limited for big data sets. Indeed, on each iteration of EM, a probability must be assigned to any feasible combination of positions for the current signal estimate in M locations on the grid $\{1, \ldots, N\}$. In total, even when excluding forbidden combinations due to the spacing constraint, there are $O(N^M)$ such combinations, and the problem becomes computationally intractable when M grows with N and N is large.

Because likelihood methods are computationally expensive, methods based on recovery from moments, which are akin to our method, have also been previously used for system identification. Methods based on the third- and fourth-order moments are described and analyzed in [36, 22, 46].

C Theory

The impossibility of detection in low SNR. If x is known and $K = 1$, then the locations s_i can be estimated via linear programming in the high SNR regime [7, 19, 13, 10, 15]. However, in the low SNR regime, estimating the binary sparse signal s is impossible. To see this, suppose that an oracle provides us M windows of length $W > L$, each containing exactly one copy of x. Suppose too that the oracle provides us with x itself. That is to say, we get a series of windows of length W, each one containing a signal x at an unknown location; and our only task is to estimate the locations. This is an easier problem than detecting the support of s. Nevertheless, even this simpler problem is impossible in the low SNR regime [4]. Consequently, detecting the nonzero values of s is impossible in low SNR.

Estimating a signal from its third-order autocorrelation. A one-dimensional signal is determined uniquely by its second- and third-order autocorrelations. Indeed, since $z[0]$ and $z[L-1]$ are non-zero by definition, we have the formula:

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

In particular, we have proven the following proposition:

Proposition C.1. *Let $z \in \mathbb{R}^L$ and suppose that $z[0]$ and $z[L-1]$ are nonzero. Then z is determined uniquely from a_z^2 and a_z^3.*

Some remarks are in order. First, formula (C.1) is not numerically stable if $z[0]$ and/or $z[L-1]$ are close to 0. In practice, we recover z by fitting it to its autocorrelations using a nonconvex least-squares procedures, which is empirically more robust to additive noise; we have seen similar phenomena for related problems [12, 17].

Second, if the spacing condition (4.2) holds, then the length of the signal can be determined from the autocorrelations. In particular, if (4.2) holds for some spacing $W \geq L$, then $a_z^2[i] = 0$ for all $i > L-1$.

Note too that the second-order autocorrelation is not by itself sufficient to determine the signal uniquely [9, 11]. However, for dimensions greater than one, almost all signals are determined uniquely up to sign (phase for the complex signals) and reflection through the origin (with conjugation in the complex case) [25, 26]. The sign ambiguity can be resolved by the mean of the signal if it is not zero. However, determining the reflection symmetry still requires additional information, beyond the second-order autocorrelation.

Identifiability of parameters from the moments of y. The observed moments a_y^1, a_y^2 and a_y^3 of y do not immediately give the moments of the signal x, as seen by formula (4.6); rather, the two are related by the noise level σ and the ratio $\gamma = \lim_{N\to\infty} ML/N$, where $M = M_N$ grows with N. We will show, however, that x is still identifiable from the observed moments of y. In general, we say a parameter is "identifiable" if its value is uniquely determined in the limit $N \to \infty$.

First, we observe that if the noise level σ is known, one can estimate γ from the first two moments of the observed vector y.

Proposition C.2. *Let $K = 1$ and $\sigma > 0$ be fixed. If the mean of x is nonzero, then*

$$\gamma = \lim_{N\to\infty} \frac{(a_y^1)^2}{\sum_{j=0}^{L-1} a_y^2[j] - \sigma^2} \quad a.s.$$

Proof. The proof follows from plugging the explicit expressions of (4.6) into the right hand side of the equality. □

Using third-order autocorrelation information of y, both the ratio γ and the noise σ are identifiable. For the following results, when we say that a result holds for a "generic" signal x, we mean that it holds for all x inside a set $\Omega \subset \mathbb{R}^L$, whose complement $\mathbb{R}^L \setminus \Omega$ has Lebesgue measure zero.

Proposition C.3. *Let $K = 1$, and $\sigma > 0$ be fixed. Then, 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 $\gamma \geq \frac{1}{4L(L-1)}$, then this holds for any signal x with nonzero mean.*

Proof. See Appendix D. □

From Propositions C.1 and C.3 we can directly deduce the following:

Corollary C.4. *Let $K = 1$ and $\sigma > 0$ be fixed. Then the signal x, the ratio γ, and the noise level σ are identifiable from the first three autocorrelation functions of y if:*

- *Either the signal x is generic; or*
- *Both $x[0]$ and $x[L-1]$ are nonzero, x has nonzero mean, and $\gamma \geq \frac{1}{4L(L-1)}$.*

Open theoretical questions. Our method of estimating x uses the third-order moments of the observations. These empirical moments are used to obtain consistent estimators of population parameters related to the the mean and second- and third-order autocorrelations of x, to which we fit the signal x. Consequently, the number of signal occurrences M should grow at least as fast as $1/\mathsf{SNR}^3$ to achieve a constant estimation error. In the related problem of multireference alignment [39, 2], this is optimal in the low SNR regime; we conjecture that the same is true for our problem.

Another interesting question is how many signals $x_1, \ldots, x_K$ can be demixed from their mixed autocorrelation functions. In [17], we empirically observed that $K \sim \sqrt{L}$ signals can be estimated simultaneously from their mixed second- and third-order autocorrelations, using the least-squares procedure. In [48], this result is shown theoretically for a different, and much less efficient, algorithm. In our current setting, the additional parameters γ and σ make the problem more challenging; however, we conjecture that the number of estimable signals still grows like $\sqrt{L}$.

D Proof of Proposition C.3

We will prove that both σ and γ are identifiable from the observed first three moments of y. For convenience, we will work with $\beta = \gamma/L$ rather than γ itself. We will construct two quadratic equations satisfied by β from observed quantities, independent of σ. Then, we will

show that these equations are independent, and hence that β is uniquely defined. Given β, we can estimate σ using Proposition C.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, while using F_i for functions of the signal's autocorrelations.

Recall that $a_y^1 = \beta(\mathbf{1}^T x)$ and and $a_y^2[0] = \beta\|x\|^2 + \sigma^2$, where $\mathbf{1} \in \mathbb{R}^L$ stands for vector of ones. Taking the product:

$$\begin{aligned} E_1 &:= a_y^1 a_y^2[0] = (\beta(\mathbf{1}^T x))(\beta\|x\|^2 + \sigma^2) \\ &= \sigma^2 a_y^1 + \beta^2 F_1, \end{aligned} \tag{D.1}$$

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 be also estimated from a_y^3, while taking the scaling and bias terms into account:

$$E_2 := \beta F_1 + (2L+1)\sigma^2 a_y^1. \tag{D.2}$$

Therefore, from (D.1) and (D.2) we get

$$E_2\beta - (2L+1)\sigma^2\beta a_y^1 = E_1 - \sigma^2 a_y^1. \tag{D.3}$$

Let $a_y^2 := \sum_{j=0}^{L-1} a_y^2[j]$ and recall from Proposition C.2:

$$\sigma^2 = a_y^2 - (a_y^1)^2/(\beta L). \tag{D.4}$$

Plugging into (D.3) and rearranging we get

$$\mathcal{A}\beta^2 + \mathcal{B}\beta + \mathcal{C} = 0, \tag{D.5}$$

where

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

Importantly, these coefficients are observable quantities.

We are now proceeding to derive the second quadratic equation. We notice that

$$E_3 = \frac{1}{L}(a_y^1)^3 = \frac{1}{L}\beta^3(\mathbf{1}^T x)^3 = \frac{1}{L}\beta^3 F_2, \tag{D.6}$$

where

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

On the other hand, from a_y^3 we can directly estimate F_2 up to scale and bias

$$E_4 = \beta F_2 + (6L-3)\sigma^2 a_y^1. \tag{D.7}$$

Taking the ratio:

$$\frac{E_4}{E_3} = \frac{L}{\beta^2} + \frac{(6L-3)L\sigma^2 a_y^1}{E_3},$$

we conclude:

$$\sigma^2 = \frac{E_4}{a_y^1 L(6L-3)} - \frac{E_3}{\beta^2 a_y^1(6L-3)}.$$

Using (D.4) and rearranging we get the second quadratic:

$$\mathcal{D}\beta^2 + \mathcal{E}\beta + \mathcal{F} = 0, \tag{D.8}$$

where

$$\begin{aligned}
\mathcal{D} &= a_y^2 - \frac{E_4}{a_y^1 L(6L-3)}, \\
\mathcal{E} &= -(a_y^1)^2/L, \\
\mathcal{F} &= \frac{E_3}{a_y^1(6L-3)}.
\end{aligned}$$

To complete the proof, we need to show that the two quadratic equations (D.5) and (D.8) are independent. To this end, it is enough to show that the ratio between the coefficients is not the same. From (D.5) and (D.1), we have

$$\begin{aligned}
\frac{\mathcal{B}}{\mathcal{C}} &= \frac{LE_1 - (2L+1)(a_y^1)^3 - La_y^1 a_y^2}{(a_y^1)^3} \\
&= \frac{La_y^2[0] - (2L+1)(a_y^1)^2 - La_y^2}{(a_y^1)^2}.
\end{aligned}$$

In addition, using (D.6)

$$\frac{\mathcal{E}}{\mathcal{F}} = \frac{(3-6L)(a_y^1)^3}{LE_3} = 3 - 6L.$$

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

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

Rearranging the equation and writing in terms of x we get

$$4(L-1)\beta(a_x^1)^2 - \sum_{i=1}^{L-1} a_x^2[i] = 0. \tag{D.9}$$

For generic x, this polynomial equation is not satisfied. Therefore, the equations are independent. More than that, for any nonzero x, $(a_x^1)^2 > \sum_{i=1}^{L-1} a_x^2[i]$. Therefore, if $4(L-1)\beta \geq 1$, or,

$$\beta \geq \frac{1}{4(L-1)},$$

the condition (D.9) cannot be satisfied for any signal.