Toward single particle reconstruction without particle picking

Tamir Bendory, Nicolas Boumal, William Leeb and Amit Singer

May 14, 2018

Abstract

Here comes the abstract

1 Introduction

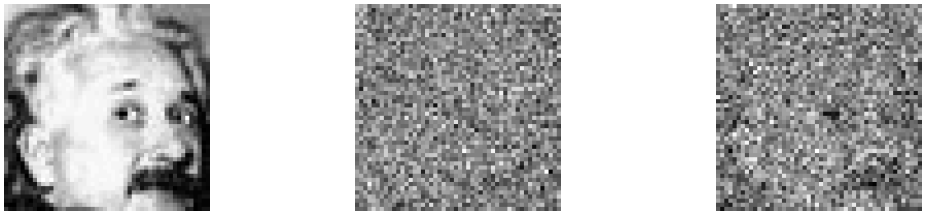
Single particle reconstruction (SPR) using cryo–electron microscopy (cryo–EM) is an innovative technology for reconstructing the 3-D structure of macromolecules. In recent years, structures of many molecules, previously regarded as insurmountable, are now being reconstructed to near-atomic resolution; see for instance [34, 10, 51]. 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 to several subroutines. The first step in the pipeline is the so-called *particle picking*, in which one aims to detect the 2-D tomographic projections of the samples from the noisy micrograph. The output of 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 [26, 44, 55, 23, 47, 29]. However, most of these procedures are prone to *model bias*. For instance, in the popular framework of RELION [47], 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 [35, 53]. Nowadays, it 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 the model bias in cryo–EM was stressed in [49, 30] and demonstrated by the "Einstein from noise" example. In this experiment, the image of Einstein is correlated with multiple images of pure noise. The noisy images are aligned to Einstein's image using cross-correlation and then averaged. In Figure 1 we conducted this experiment with different number of noise images. With merely 1000 noise images, Einstein's face is clearly observed in the averaged image. With $10,000$ image, the output is remarkably similar to the template image of Einstein, rather than to pure noise. Crucially, the more noisy images we use, the more severe the model bias towards the template. In the context of cryo–EM, this example demonstrates how our prior assumptions about the particle may influence the reconstructed structure.



(a) Einstein image (template) (b) $n = 10$ (c) $n = 10^2$

(d) $n = 10^3$ (e) $n = 10^4$ (f) $n = 10^5$

Figure 1: Einstein from noise experiment. As a template, we use Einstein's image of size 50×50 pixels. We drawn n images of the same size whose entries are composed of i.i.d. Gaussian entries with the same mean and variance as Einstein's image. Each noise image was aligned with respect to the template by the maximal value of their cross correlation. After the alignment, all the images were averaged. As can be seen, the more pure noise images we use, the closer the averaged image to the template.

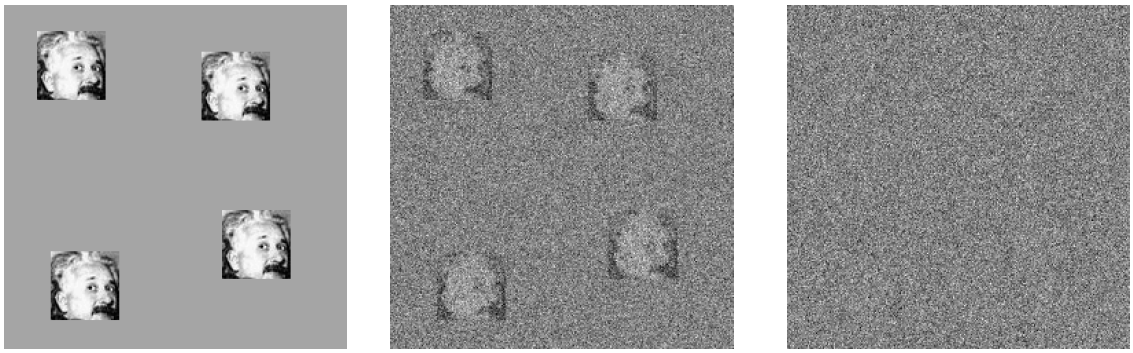
Besides the model bias, particle pickers have several limitations due to the high noise level, in particular for small particles and in low contrast. As a result, the projections in the 2D images are typically not centered; including the correct centers as parameters dramatically increases the complexity 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 [37], passive radar [25] and system identification [43].

Figure 2 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, appearing in Figure 1a. 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, however, Einstein is 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.



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

Figure 2: 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 that appears in Figure 1a.

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 autocorrelation analysis. In a nutshell, 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 Appendix C for more details).

2 Results

In this section we describe some numerical experiments and show results. The experimental details and the algorithms used 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 $\sigma = 3$. An illustrative example of a micrograph appears in the right panel of Figure 2. In each micrograph, multiple occurrences of Einstein's image were placed, while keeping a separation of 50 pixels between the images in each direction. From the micrographs, we estimated the 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).

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

Figure 3 demonstrates several recovery results for different number of micrographs collected. Figure 4 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

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

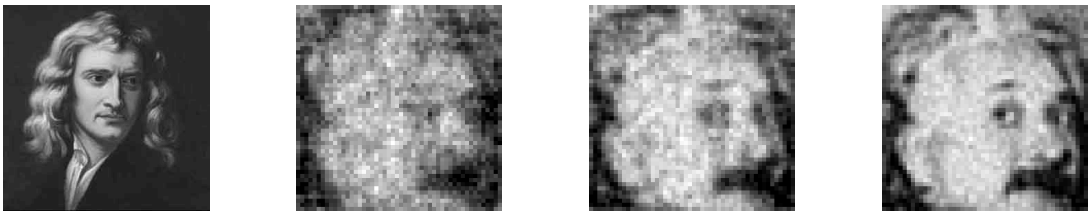
where $\hat{x}$ is the estimated image and $\|\cdot\|_{\text{F}}$ stands for the Frobenius norm. As expected, the RMSE of estimating the power spectrum decreases linearly with slope $-1/2$ in logarithmic scale. [Few comments: 1. We have a movie in the supplementary material. 2. Does the error also have the right slope? 3. We can put more images to exemplify the progress 4. Need to improve the micrograph's generation code]

Results from a similar experiment with $\sigma = 3$ for $K = 3$ 1-D signals is presented in Figure 5. 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 signal's occurrences, L is the length of the signal and N is the micrograph length. These ratios present the *density* of the ith signal in the data. In the experiment, we do not assume to know these ratios, neither the noise level σ.

In order to estimate the signals, we computed the first three autocorrelation functions of the data and then estimated the signals and their corresponding γ_i using a nonconvex least-squares. As can be seen, given enough signal occurrences, we can estimate accurately the signals. The estimation quality of the triangle signal is poorer than the other two signals, a phenomenon that is explained using Proposition C.1. [1. We will replace the figure with a



(a) Newton (model) (b) $P = 1024$ (c) $P = 1024 \times 10$ (d) $P = 1024 \times 100$

Figure 3: Recovery of Einstein's image. The RRR algorithm is initialized with the image of the physicist Issac Newton. We show estimations of Einstein using $P = 1024, 1024 \times 10$ and 1024×100 number of micrographs, each contains 800 image occurrences on average (to verify). The RMSE is 0.5514, 0.3494 and 0.2112, respectively. The power spectrum estimation RMSE is, respectively, 0.0269, 0.0083 and 0.0027. [Can add more images to exemplify the progress]

"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]

3 Discussion

All current algorithmic pipeline for SPR using cryo–EM start with a particle picking technique which is prone to model bias. Bypassing the particle picking stage and constructing a 3-D model directly from the data—without assuming prior knowledge on the particle to be estimated—can be used to the construction of ab initio models or to generate multiple denoised projections that can be used for 2-D classification. It might be even possible to use such templates for the particle picking stage, now with templates that were extracted from the data itself.

In this paper, we examined a simplified model that, we believe, captures important features of the SPR problem. While Bayesian algorithms are currently dominate the field, they are intractable for our problem. As alternative, we propose to use autocorrelation analysis technique that shares some common lines with Kam's method for ab inito modeling [32, 36, 50]. That being said, the SPR model is far more complicated than the model presented here. In a future research, we hope to bridge this gap.

Our results rely on two core assumptions that are not necessarily met by in an SPR experiment. First, we modeled the background information as i.i.d. additive noise. In practice, the background information may be structured or depend on the signal. Yet, autocorrelation analysis is easily adapted to different noise statistics. [Here, an example with colored noise would be of great help.]

In addition, 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 [???]. If the signals are not separated, one can introduce a new variable that represents the distribution of the spacing between signal occurrences and then compute explicitly the relation between the autocorrelation of the data and the signals. It is yet to be studied under what conditions on this distribution one can estimate the signals.

[need a lot of signal occurrences]

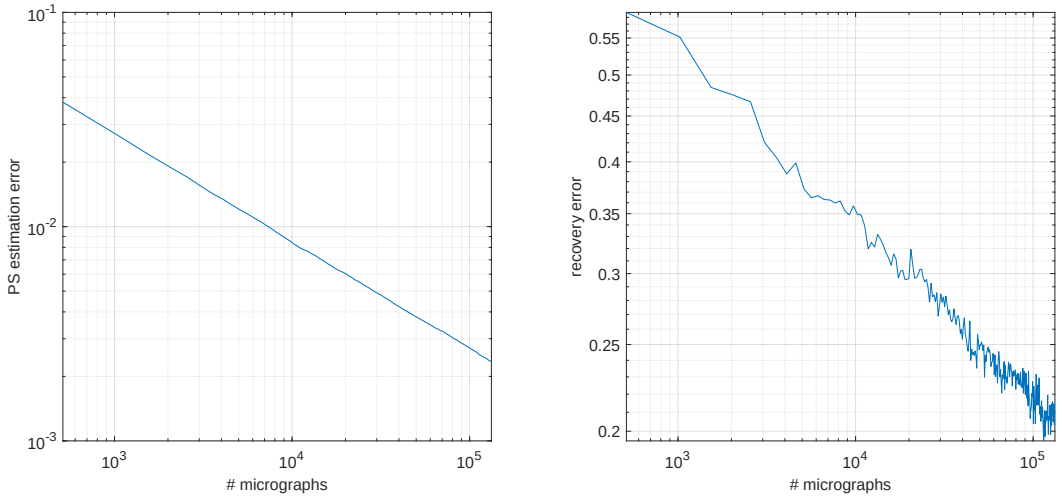


Figure 4: The left panel shows the recovery error (in log-log scale) of Einstein's image defined in (**??**) as a function of the number of the collected micrographs. The right panel presents the error of estimating the power spectrum of the signal from the data. As expected, in log-log scale, the curve has a $-1/2$ slope.

4 Methods

Mathematical model Let $x_1, \dots, x_K \in \mathbb{R}^L$ be the sought signals and let $y \in \mathbb{R}^N$ be the 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 the occurrences of the corresponding x_i. 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 $|\mathcal{S}| := M = \sum_{i=1}^K M_i$. Neither the M_i's nor M are assumed to be known. 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,

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

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

Aperiodic autocorrelation functions For the purpose of this paper, we need the first three (aperiodic) autocorrelation functions. The first-order autocorelation is the mean of the

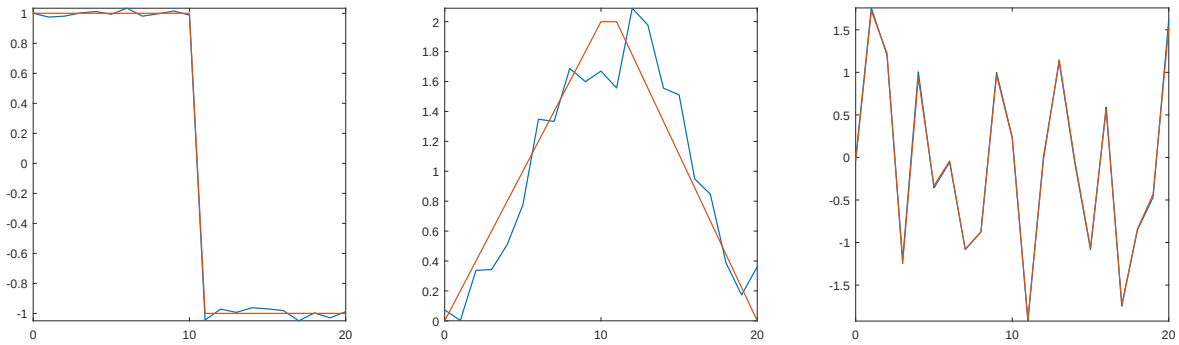


Figure 5: Red signals are the ground truth (targets) and the blue signals are our estimations. Individual RMSE of the estimates: 0.0239393/0.208925/0.0335956. The estimated γ's are 0.02574/0.01693/0.00818 and the true ones 0.02561/0.01707/0.00853. The SNR is 1/12. [to replace with a progress plot']

signals. 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 indexing of z out of the bounds $0, \ldots, L-1$ is zero-padded, as usual. Explicitly, the first three autocorrelations are

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

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

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

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 $M, N \to \infty$, while preserving fixed ratio. Specifically, we define the ratio of the measurement occupied by the signal as

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

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

Estimating the autocorrelation function of a multiple signals As before, we consider the asymptomatic regime where $M_1, \dots, M_K, N \to \infty$, while preserving fixed ratios

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

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

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

This relation is proven in Appendix A.

Numerical experiments details For the 2-D experiment, we generated P micrographs of size 4000×4000 with Einstein's image of size 50×50. In each micrograph, we placed the image in random locations while keeping the separation condition (4.2). In particular, to picked a random location on the micrograph and used an accept/reject rule based on the separation constraint and locations picked so far. We stopped the process when we reached the desired number of occurrences, or the micrograph was dense enough so that we failed to find an open space (randomly) repeatedly. Then, an i.i.d. Gaussian noise was added. On average, each micrograph contains 860 images and the signal–to–noise ratio (SNR) is approximately 1/18. Different micrographs 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.

Given the micrographs y, we proceed to compute the moments. In Appendix C we provide an argument on the number of equations we get from the autocorrelation function. To ease the computation, we assume to know the number of Einstein's occurrences and the noise

level. This allow us to compute only the first and second moment. In dimensional greater than one, almost all images are uniquely determined from their second-order aperiodic up to reflection through the origin [27, 13]. To estimate the signal, we use a popular algorithm called relaxed-reflect-reflect (RRR) iterates as

$$x_{\ell+1} = x_\ell + \beta(P_2(2P_1(x_\ell) - x_\ell) - P_1(x_\ell)). \tag{4.9}$$

The RRR estimates a $2L \times 2L$ image. Exact solution contains Einstein's image in the upper-left corner and zero else where. The operation $P_1(z)$ combines the Fourier phases of z with the Fourier transform of the second-order autocorrelation, which in the absence of noise, is exactly the Fourier magnitude of the solution. The operator $P_1(z)$ zeros out all entries of z outside the upper-left corner.

For the 1D experiment, we worked with signals of length $L = 21$ and generated the data in the same way as in the 2D example. However, here we computed the first three autocorrelation functions. We removed the biased terms (see (4.6)) so we do not need to know σ ahead. To estimate the signal, we employ an optimization algorithm on the following nonlinear least-squares problem that estimate the signals and their number of occurrences simultaneously:

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

where $W \ge L$ is the length of the sought signals and $w_1 = 1/2, w_2 = 1/2(L-1), w_3 = \frac{1}{(L-1)(L-2)}$; see [19] see for discussion in choosing the weights properly.

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 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 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 [18], which allows to treat the positivity constraints [I might need a reference for this] 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.

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

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

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

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

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

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

[13] 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.

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

[15] 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.

[16] 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.

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

[18] 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.

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

[20] 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.

[21] Yuejie Chi. Guaranteed blind sparse spikes deconvolution via lifting and convex optimization. *IEEE Journal of Selected Topics in Signal Processing*, 10(4):782–794, 2016.

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

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

[24] 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.

[25] 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.

[26] 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.

[27] 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.

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

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

[30] 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.

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

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

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

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

[35] 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.

[36] 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.

[37] 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.

[38] Xiaodong Li, Shuyang Ling, Thomas Strohmer, and Ke Wei. Rapid, robust, and reliable blind deconvolution via nonconvex optimization. *arXiv preprint arXiv:1606.04933*, 2016.

[39] 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.

[40] 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.

[41] Shuyang Ling and Thomas Strohmer. Self-calibration and biconvex compressive sensing. *Inverse Problems*, 31(11):115002, 2015.

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

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

[44] 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.

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

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

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

[48] 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.

[49] 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.

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

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

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

[53] 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.

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

[55] 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 Autocorrelation estimations

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 appropriates weights, see [19].

Let us define

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

With a bit abuse of notation, M_N stresses that M is a function of N. Indeed, by assuming $M_N = \Omega(N)$, we deduce $\gamma > 0$. We start by considering the first autocorrelation of the data

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

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

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

$$a_y^2[\ell] = \frac{1}{N}\sum_{i=0}^{N-1-\ell} y[i]y[i+\ell] \\ \underbrace{\frac{1}{N}\sum_{j=1}^{M_N}\sum_{i=0}^{L-\ell-1} x[i]x[i+\ell]}_{\text{signal term}} + \underbrace{\frac{1}{N}\sum_{i=0}^{N-1-\ell}\varepsilon[i]\varepsilon[i+\ell]}_{\text{noise term}}, \tag{A.3}$$

where the cross terms between the signal and the noise vanish almost surely in the limit $N \to \infty$.

We treat the signal and noise terms separately. We first break the signal term into M_N different sums, each contains one copy of the signal, and get

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

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

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

We are now moving to analyze the third-order autocorrelation. Let us fix $\ell_1 \geq \ell_2$ and recall that

$$a_y^3[\ell_1, \ell_2] = \sum_{i=0}^{N-1-\ell_1} y[i]y[i+\ell_1]y[i+\ell_2].$$

Writing explicitly in terms of signal and noise, the sum can be broken into eight terms. The first contains only signal terms (does not see noise) and converges to γa_x^3 from the same reasons as (A.4). Three other terms contain the product of two signal entries and one noise term. Since the noise is independent of the signal and has zero mean, these terms go to zero almost surely.

We next analyze the contribution of the term composed of triple products of noise terms. For $\ell_1 \neq 0$, this sum can be formulate as follows:

$$\sum_{i=0}^{N-1-\ell_1}\varepsilon[i]\varepsilon[i+\ell_1]\varepsilon[i+\ell_2] = \frac{1}{\ell_1}\sum_{i=0}^{\ell_1-1}\frac{1}{N/\ell_1}\sum_{j=0}^{N/\ell_1-1}\varepsilon[j\ell_1+i]\varepsilon[(j+1)\ell_1+i]\varepsilon[j\ell_1+i+\ell_2].$$

For each fixed i, we sum of over N/ℓ_1 independent variables that goes to zero almost surely. For $\ell_1 = \ell_2 = 0$, we get a some of N independent variables, each one is a triple product of Gaussian variables with zero mean and therefore has zero expectation.

To complete the analysis, we consider the three terms composed of the product of two noise terms and one signal entry. Most of these terms converge to zero almost surely because of independency between the noise entries. For $\ell_1 = 0, \ell_2 = 0$ and $\ell_1 = \ell_2$, a simple computation shows that the sum converges to $\gamma\sigma^2 a_x^1$; c.f. [19].

B Related literature

Blind deconvolution. Blind deconvolution is a longstanding problem, arising in a variety of engineering and scientific applications, such as astronomy, communication, image deblurring, system identification and optics; see [31, 48, 7, 3], just to name a few. To make the problem well-posed, we must assume some prior knowledge or structure. In our case, the prior information is that s is a binary signal that satisfies the separation constraint (4.2). Other settings of blind deconvolution problems have been analyzed recently, see for instance [5, 39, 38, 41, 42, 21] where the focus is on high SNR regimes.

An important feature of the problem under consideration is that while both x_i's and s_i's are unknown, the goal is merely to estimate the x_i's. The s_i's are referred to as *nuisance variables*. Indeed, in many blind deconvolution applications the sole purpose is to recover one of the unknown signals. For instance, in image deblurring, both the blurring kernel and the high-resolution image are unknown, but the primary goal is only to sharpen the image.

The impossibility of detection in low SNR. If x is known and $K = 1$, then a sparse signal can be estimated by linear programming in the high SNR regime, e.g., [8, 22, 15, 12, 17]. However, in the low SNR regime, estimating the binary sparse signal s is impossible. To see that, suppose that an oracle provides us M windows of length $W > L$, each contains one copy of x. That is to say, we get a series of windows of length W, each one contains a signal at an unknown location. Estimating the position of the known signal within each window is an easier problem than detecting the support of s. Nevertheless, even this problem is impossible in the low SNR regime [4]. Therefore, we conclude that detecting the nonzero values of s is impossible in low SNR. As aforementioned, this work focuses on this regime and examines under what conditions we can estimate the signals, despite the impossibility of detecting their individual occurrences.

System identification. For $K = 1$, our problem can be also interpreted as a special case of the system identification problem. Similarly to (4.1), the system identification forward model takes the form

$$y = x * w + \varepsilon, \tag{B.1}$$

where x is the unknown signal ("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 [16, 33]. 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 ($K = 1$).

Likelihood-based methods seek to maximize the likelihood function for x, given the observed signal y. Solving this optimization exactly is typically intractable, and thus heuristic methods are used instead. One proposed technique is to use Markov Chain Monte Carlo (MCMC); in special cases, including the case where w is binary, EM has been used [20]. The EM method is based upon a certain "forward-backward" procedure used in hidden Markov models [46]. However, the complexity of this procedure is still nonlinear in N, and therefore its usage is limited for big data sets. Another paper considers parameterized models for multiple distinct signals, as in our heterogeneity framework ($K > 1$) [6]. Their proposed solution is an MCMC algorithm tailored for their specific parametrized problem.

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 [40, 24, 52].

C Theory

A one-dimensional signal is determined uniquely and stably by its third-order autocorrelation as proven in the following simple proposition.

Proposition C.1. *Let $z \in \mathbb{R}^L$ and suppose that $z[0]$ and $z[L-1]$ are nonzero. Then:*

- ***Uniqueness:*** *z is determined uniquely from a_z^2 and a_z^3.*
- ***Finite sensitivity:*** *Suppose we can only measure $\tilde{a_z^3}[k, L-1] = a_z^3[k, L-1] + \upsilon$ and that $|z[0]z[L-1]| \geq \delta > 0$. Then, $\hat{z}[k] = \frac{\tilde{a_z^3}[k,L-1]}{a_z^2[L-1]}$ satisfies $|\hat{z}[k] - z[k]| \leq \frac{|\upsilon|}{\delta}$.*

Proof. By assumption $a_z^2[L-1] = z[0]z[L-1] \neq 0$. Then, the uniqueness results, for all $k = 0, \ldots L-1$, follows from:

$$a_z^3[k, L-1] = z[0]z[k]z[L-1].$$

In addition,

$$\hat{z}[k] = \frac{\tilde{a_z^3}[k, L-1]}{a_z^2[L-1]} = z[k] + \frac{\upsilon}{a_z^2[L-1]} \quad \Rightarrow \quad |\hat{z}[k] - z[k]| \leq \frac{|\upsilon|}{\delta}.$$

□

A few remarks are in order. First, the second result of Proposition C.1 shows that there exists a very simple estimator that has finite sensitivity. In the next numerical experiments we use an estimator based on nonconvex LS that shows empirical robustness to additive noise, in accordance with related problems [14, 19]. Second, these results carry through to signals of any dimension. Third, if the spacing condition (4.2) holds, then the length of the signal can be determined from the autocorrelations and therefore the assumption that the first and last entries are nonzero is met. In particular, if (4.2) holds for some spacing $W \geq L$, then $a_z^2[i] = 0$ for all $i > L-1$. Finally, computing the dth autocorrelation amplifies the variance

of the noise by a factor d in the low SNR regime. Therefore, if we can estimate a_z^3 up to small perturbation, it implies that we can estimate a_z^2 accurately as the proposition assumes.

Considering the third-order autocorrelation is also a necessary condition to determine a signal from its autocorrelations. Indeed, the second-order autocorrelation of a one-dimensional signal does not determine a signals uniquely [11, 13]. On the other hand, 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) [27, 28]. The sign ambiguity can be resolved by the mean of the signal if it is not zero. However, in order to determine the reflection symmetry, one needs to use additional information.

If the noise level σ^2 is known, one can estimate the ratio M/N from the first two moments.

Proposition C.2. *Let $K = 1$, $N \to \infty$ and σ fixed. If the mean of x is nonzero, then*

$$\frac{M}{N} = \frac{1}{L} \frac{(a_y^1)^2}{\sum_{j=0}^{L-1} a_y^2[j] - \sigma^2}.$$

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

If we use third-order autocorrelation information, then it is possible to estimate both the ratio M/N and σ simultaneously.

Proposition C.3. *Let $K = 1$, $N \to \infty$ and σ fixed. Then, a_y^1, a_y^2 and a_y^3 determine the ratio M/N and σ uniquely for a generic signal x. If $\frac{M}{N} \geq \frac{1}{4(L-1)}$, then it holds for any signal with nonzero mean.*

Proof. See Appendix D. □

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

Corollary C.4. *Let $K = 1$, $N \to \infty$ and σ is fixed. Then, the signal, the ratio M/N and σ can be recovered from the first three autocorrelation functions if:*

- *x is generic;*
- *$x[0], x[L-1] \neq 0$, x has nonzero mean and $\frac{M}{N} \geq \frac{1}{4(L-1)}$.*

Expectation-maximization. Interestingly, expectation-maximization (EM)—a popular algorithm for similar estimation problems, such as Gaussian mixture models and multireference alignment—is intractable for this problem. This is true even if $K = 1$ and the number of signal occurrences M is known. In particular, in each iteration, EM assigns a probability to any feasible combination of positioning 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.

Open theoretical questions. The minimal order of data statistics used to get an accurate estimation of a signal is important to understand, in the asymptotic SNR regime, the sample complexity of the problem. In methods which are based on detection and averaging, the number of signals occurrences must scale like $1/\mathsf{SNR}$. On the contrary, the required number signal occurrences using the autocorrelation analysis should scale like $1/\mathsf{SNR}^d$ to retain a constant estimation error. Accordingly, in our method, M must scale like $1/\mathsf{SNR}^3$. We believe that similarly to the closely-related problem of multirefrence alignment [45, 2], this estimation rate is optimal in the low SNR regime.

Another interesting question is how many signals can be demixed from the autocorrelation functions. For second-order moments, notice from equation (4.8) that $a_y^2[\ell]$ suffers no 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 estimate the moments. Likewise, for third-order moments, $a_y^3[\ell_1, \ell_2]$ for $0 \leq \ell_1, \ell_2 \leq L-1$ such that $\ell_2 \leq \ell_1$ includes all relevant moments for our purpose, and we further exclude any such that ℓ_1, ℓ_2 or $\ell_1 - \ell_2$ are zero to avoid biased elements—there are $\frac{(L-1)(L-2)}{2}$ remaining moments. As a result, it is unnecessary to estimate σ. We have

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

moments in total.

In the related problem of demixing signals from their periodic autocorrelation functions, there are strong evidences that the number of signals that, theoretically, can be demixed is equal to the number of different equations [9]. Based on these evidence, we conjecture the same here, namely, $L/2$ signals can be estimated simultaneously. That being said, we believe the number of classed that can actually be demixed by efficient algorithm is significantly smaller and scale like $\sqrt{L}$; see [19, 54].

D Proof of Proposition C.3

We aim to prove that one can estimate both σ and $\beta = M/N$ from the observed first three moments. To this end, we construct two quadratic equations of β from the observed (measured) quantities, independent of σ. Then, we show that these equations are independent and therefore 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. to this end, 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.