Estimation below the detection limit

Tamir Bendory

March 23, 2018

Abstract

Here comes the abstract

1 Introduction

In this paper, we consider the problem of estimating a set of signals $x_1, \ldots, x_K$ from their multiple occurrences in unknown, random, locations in a data sequence y. The data may also contain background information – independent of the signals – which is modeled as noise. For one-dimensional signals, the data can be thought of as a long time series and the signals as repetitive short events. For two-dimensional signals, y presents a big image, containing many smaller images. The problem is then to estimate the signals $x_1, \ldots x_K$ from y. This model appears, in different noise levels, in many applications, including spike sorting [22], passive radar [17] and system identification [27]. In Section 2 we provide a precise mathematical formulation of the model and the estimation problem.

If the noise level is negligible, estimating the signals is easy. In this scenario, standard detection and clustering algorithms can produce multiple copies of each signal that can be then averaged. Even in higher noise level regimes, clever methods based on template matching, such as those used in structural biology [20] and radar [17], may work. [Do we want to add another 1D example to demonstare the problem, similarly to Figure 1.1 in the bispectrum paper?] However, in the low signal–to–noise (SNR) regime, detection of individual signal occurrences is impossible as explained in Section 2. Figure 1 illustrates the problem in different noise levels and one underlying signal ($K = 1$). Figure 1b shows a 21×21 image, which is a downsampled version of a projection of TKTK, taken from ASPIRE package [1] [The image is taken from the example folder]. Figure 1a shows a excerpt of a big image (the data image) that contains many repetitions of the projection. In Figures 1c and 1e, the same excerpt is shown with the addition of i.i.d. Gaussian noise with standard deviations of $\sigma = 0.2$ and $\sigma = 1$, respectively. Figures 1d and 1f show estimates of the projection from noisy data with $\sigma = 1$. These examples demonstrate that our method can work even if the data may seem as a pure noise. In Section 4 we provide the details of this experiment and show more corroborating experiments.

In this work we focus on the low SNR regime. In order to estimate the signal, we use autocorrelation analysis of the data. In a nutshell, the method consists of two stages. First, we estimate a mix of the low–order autocorrelation functions of the signals from the data. These quantities can be estimated, to any desired accuracy, if the signals appear enough times in the measurement, without the need to detect individual occurrences. Then, the signals are estimated from the mixed autocorrelations using a non-convex least-squares (LS). In Section 3, we elaborate on the technique and prove some of its properties. Interestingly, expectation-maximization (EM) – a popular framework for similar estimation problems, such as Gaussian mixture models and multireference alignment – is intractable for this problem. Even if the number of signal occurrences M is known and $K = 1$, at each iteration EM needs to assign probabilities to each possible combination of M copies of the current signal estimation in the measurement, which is of order of $O(N^M)$.

This work is primely motivated by cryo-electron microscopy (cryo–EM), which is an innovative single particle reconstruction technology. The acquired data in a cryo-EM experiment is contaminated with high noise levels. Therefore, any molecule reconstruction algorithm must take the challenging SNR level into account. In the last part of this manuscript, we draw connections with the estimation problem under consideration and the cryo–EM problem.

2 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. For each x_i, we associate a binary signal $s_i \in \{0,1\}^N$, referred to as the *support signal* and let $s = \sum_{i=1}^K s_i$. The nonzero values of s_i indicates the locations of x_i in y. If $s_i[n] = 1$, then $y[n+j] = x_i[n+j] + \varepsilon[n+j]$ for $j = 0, \dots, L-1$. We denote the cardinality of s_i by M_i and $M = \sum_{i=1}^K M_i$. Neither the M_i's nor M is assumed to be known.

The support signals are generated by the following generative model. The signal s is initialized with zeros. First, an index i_1 is drawn uniformly from $\{1, \dots, N\}$ and we set $s[i_1] = 1$. A second index i_2 is then drawn uniformly from $\{1, \dots, N\}$. If $|i_2 - i_1| \geq L$, then we set $s[i_2] = 1$, otherwise we keep $s[i_2] = 0$ and draw a new index from uniform distribution. We then proceed adding nonzero entries to s, while keeping L entries separation until some halting criterion is obtained. We note that if the support is sparse enough, this generative model can be approximated by a simple Bernoulli process which takes the values of one and zero with probabilities M/N and $1 - M/N$, respectively. Once s is determined, each one of its nonzero entries is associated with one of the s_i's. In particular, for each $s[k] \neq 0$ we set $s_i[k] = 1$ and $s_j[k] = 0$ for $i \neq j$, where i is drawn form an unknown probability over $\{1, \dots, K\}$.

The simplest way to present the forward model is a mix of blind deconvolution problems between the support signals and the target signals

$$y = \sum_{i=1}^{K} x_i * s_i + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I). \tag{2.1}$$

We model the background information as i.i.d. Gaussian noise with zero mean and σ^2 variance. The goal is to estimate $x_1, \dots, x_K$ from y.

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 [21, 28, 5, 2], just to name a few. Clearly, the problem is ill-posed without additional information. In our case, the prior information is that s is a binary, sparse, signal. Other settings of blind deconvolution problems have been analyzed recently under different settings, see for instance [4, 24, 23, 25, 26, 14] where the focus is on high SNR regimes.

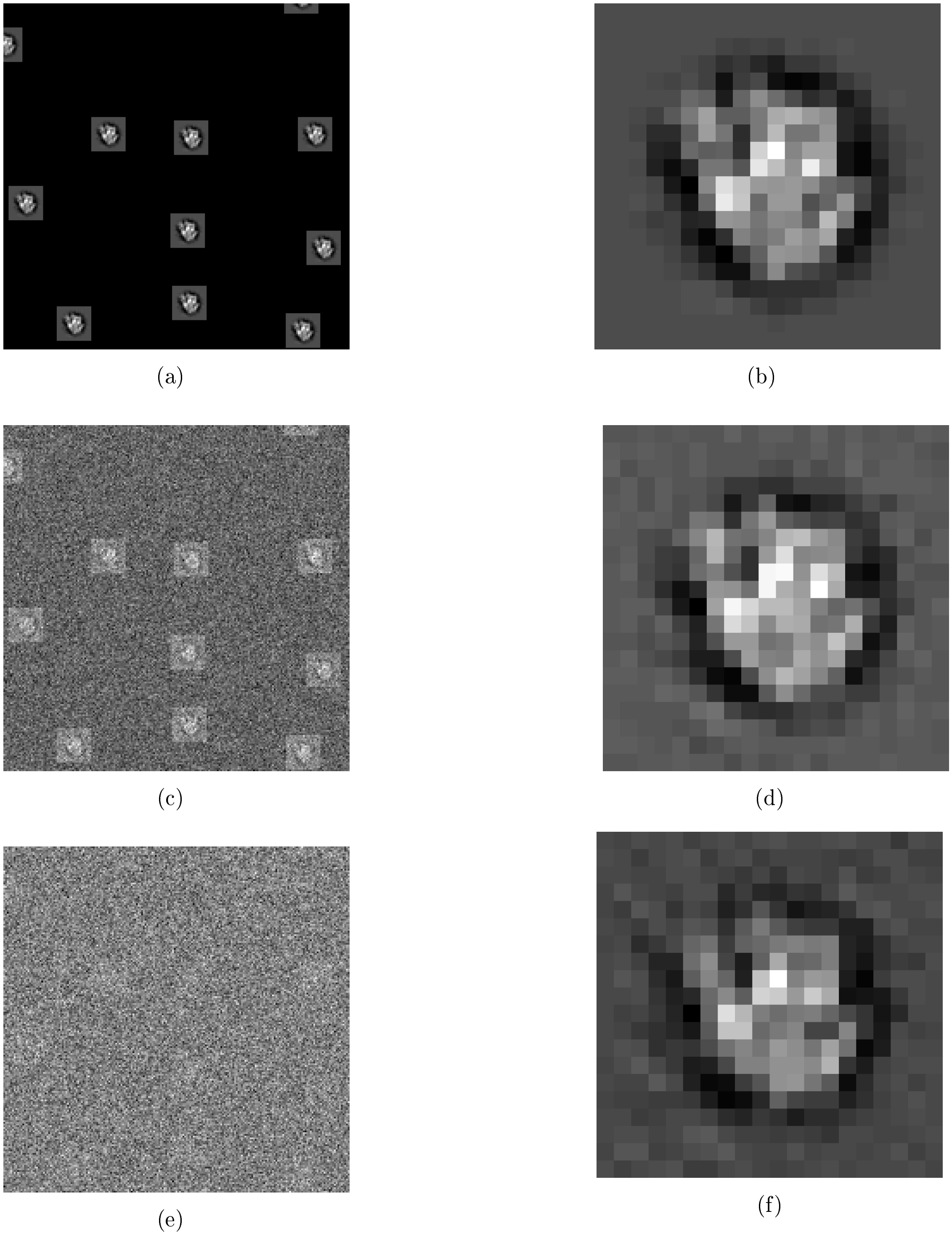

Figure 1: Figure (a) shows an excerpt of 2D data with multiple signal occurrences and no noise. Each small image is the 21×21 image shown in Figure (b). Figures (c) and (e) show the same data, now contaminated with i.i.d. Gaussian noise with $\sigma = 0.2$ and $\sigma = 1$, respectively. Figures (d) and (f) show estimates of the image (b) from data with noise level $\sigma = 1$ (as in Figure (e)). The signal appeared $M = 561 \times 10^6$ and $M = 110 \times 10^6$ times and the normalized recovery error is 0.078 and 0.135, respectively.

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 goal is merely to estimate one of the unknown signals. For instance, in image deblurring, both the blurring kernel and the high-resolution image are unknown, but the prime goal is only to sharpen the image. If x is known and $K = 1$, then s can be estimated by linear programming in the high SNR regime [15, 16, 11, 8, 12]. However, in the low SNR regime, estimating s is impossible. To see that, suppose that an oracle provides us M windows of length $W > L$, each contains one copy of x. Put it differently, we get a series of windows, each one contains a signal at unknown location. Estimating the first entry of the signal within each window is an easier problem than detecting the support of s. However, even this problem – called alignment or synchronization – is impossible in the low SNR regime. For instance, the variance of any estimator is, at best, proportional to σ^2 and independent of the number of windows, even if x is known [3]. Therefore, we conclude that that detecting the nonzero values of s is impossible in low SNR. In the next section we show that if M is large enough, then estimating the signals is possible, in any SNR level and to any accuracy, although we cannot detect their occurrences in y.

3 Autocorrelation analysis

Our method for estimating the signals is a two-stages procedure. First, we use the autocorrelation functions of the data to estimate a mix (i.e., linear combination) of the K signal's autocorrelations. The mixed autocorrelation can be estimated to any accuracy, in any SNR level, if M is large enough and the separation condition on the support is met. Then, we use a nonconvex LS to estimate the signals from their mixed autocorrelation. The precise recovery procedure we are using, based on nonconvex optimization, will be elaborated in the next section.

3.1 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 signals. For $z \in \mathbb{R}^L$ and $k \geq 2$, the higher order autocorrelations are defined by

$$a_z^k[\ell_1, \ldots, \ell_{k-1}] = \sum_{i=0}^{L-1-\max\{\ell_1,\ldots,\ell_{k-1}\}} z[i]z[i+\ell_1]\ldots z[i+\ell_{k-1}].$$

Explicitly, the first three autocorrelations are

$$\begin{aligned} a_z^1 &= \sum_{i=0}^{L-1} z[i], \\ a_z^2[\ell] &= \sum_{i=0}^{L-1-\ell} z[i]z[i+\ell], \\ a_z^3[\ell_1, \ell_2] &= \sum_{i=0}^{L-1-\max\{\ell_1,\ell_2\}} z[i]z[i+\ell_1]z[i+\ell_2]. \end{aligned} \tag{3.1}$$

Note that the autocorrelation functions are symmetric so that $a_z^2[\ell] = a_z^2[-\ell]$ and $a_z^3[\ell_1, \ell_2] = a_z^3[-\ell_1, -\ell_2]$. Additionally, if the moments of the signal depend only on the difference between the indices (Toeplitz structure), the they are equivalent to the autocorrelation functions.

A one-dimensional signal is determined uniquely by its third-order auto-correlation:

Proposition 3.1. *Let $z \in \mathbb{R}^L$ and suppose that $z[0]$ and $z[L-1]$ are nonzeros. Then, z is determined uniquely from a_z^2 and a_z^3.*

Proof. From the second-order autocorrelation, we ge

$$a_z^2[L-1] = z[0]z[L-1] \neq 0.$$

Then, for all $k = 0, \dots L-1$, we can compute $z[k]$ from the third-order autocorrelation

$$a_z^3[k, L-1] = z[0]z[k]z[L-1]. \tag{3.2}$$

□

This result can be extended to any dimension. We note that the length of the signal can be easily derived from the length of the second-order autocorrelation of the signal. Therefore, the assumption that $z[0]$ and $z[L-1]$ are nonvanishing is met in practice. We also note that for one-dimensional signals, the second-order autocorrelation does not determined the signals uniquely [7, 9]. For dimensions greater than one, almost all signals are determined uniquely from their aperiodic autocorrelations, up to sign (phase in the complex case) and reflection through the origin (with conjugation in the complex case) [18, 19]. 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, such as higher-order statistics.

High-order autocorrelations have been used for a variety of tasks in signals processing, bla bla. In particular, [Here we should refer to Gianakis's paper]

The uniqueness for $K > 1$ was explored for the related case of periodic autocorrelation functions. Recall that the kth periodic autocorrelation of $z \in \mathbb{R}^L$ is given by

$$b_z^k[\ell_1, \dots, \ell_{k-1}] = \sum_{i=0}^{W-1} z[i]z[(i+\ell_1) \mod L] \dots z[(i+\ell_{k-1}) \mod L].$$

If $z[n] = 0$ for $n = L/2, \dots L-1$, then $b_z^k[\ell_1, \dots, \ell_{k-1}] = a_z^k[\ell_1, \dots, \ell_{k-1}]$. In [6], it was shown that a mix of K third-order periodic autocorrelations determine a finite list of K generic signals if $L/6 \gtrsim K$. Empirical evidences hints that this finite list includes only group symmetries [13].

3.2 Estimating autocorrelations from the data

In order to estimate the mixed autocorrelation of the K signals, we first compute the first L entries of the data's autocorrelations. For the purpose of the analysis, we consider the asymptomatic regime where $M_1, \dots, M_K, N \to \infty$. We define the ratio of the measurement occupied by each one of the signal by

$$\gamma_i = \frac{M_i L}{N}, \tag{3.3}$$

and $\gamma = \sum_{i=1}^{K} \gamma_i$. Under the spacing constraint, 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 s satisfies the spacing constraint of L entries, then the first L entries of the data autocorrelations converge to a mix of scaled versions of the signals autocorrelations:

$$\begin{aligned}
\lim_{N\to\infty} a_y^1 &= \sum_{i=1}^{K} \gamma_i a_{x_i}^1, \\
\lim_{N\to\infty} a_y^2[\ell] &= \sum_{i=1}^{K} \gamma_i a_{x_i}^2[\ell] + \sigma^2 \delta[\ell], \\
\lim_{N\to\infty} a_y^3[\ell_1,\ell_2] &= \sum_{i=1}^{K} \gamma_i a_{x_i}^3[\ell_1,\ell_2] + \sigma^2 \left(\sum_{i=1}^{K} \gamma_i a_{x_i}^1 \right) (\delta[\ell_1,0] + \delta[0,\ell_2] + \delta[\ell_1,\ell_2]),
\end{aligned} \tag{3.4}$$

for $\ell, \ell_1, \ell_2 = 0, \ldots L-1$, and where δ denotes the Kronecker delta function. These relations are proven in Appendix A. The analysis is similar to [10, 13], yet a particular caution should be taken with the statistical dependencies of the noise entries. The relations (3.4) imply that for $K = 1$ and given M and σ, one can estimate the signal for any noise level if M is large enough. Next, we show that for $K = 1$, M and σ are also uniquely determined from the autocorrelations.

An interesting consequence of our analysis is that, at least for $K = 1$, the third autocorrelation of the data suffices to determine the signal. The minimal order of data statistics that we use 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 scales like σ^2. Taking the kth order autocorrelation function amplifies the variance of the noise by a factor of k. Therefore, the required number signal occurrences should scale like σ^{2k} to retain some constant estimation error. Accordingly, in our method, M scales like σ^6. In Section 4, we show empirically the third-order autocorrelation suffices also for a mixture of $K > 1$ signals without prior knowledge of the M_i's and σ.

If the noise level σ^2 is known and $K = 1$, one can estimate M/N from only the first two moments.

Proposition 3.2. *Let $K = 1$, $N \to \infty$ and σ fixed. 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 plugging the explicit expressions of (3.4) into the right hand side of the equality. □

If we use third-order autocorrelation information, then it possible to estimate both M/N and σ simultaneously. In contrast to Proposition 3.2, this result does not come with explicit formula.

Proposition 3.3. *Let $K = 1$, $N \to \infty$ and σ fixed. Then, a_y^1, a_y^2 and a_y^3 determine M/N and σ uniquely for generic signal. If $\frac{M}{N} \geq \frac{1}{4(L-1)}$, then it holds for any nonzero x.*

Proof. See Appendix B. □

From Propositions 3.1 and 3.3 we can directly deduces the following.

Corollary 3.4. *For $K = 1$, $N \to \infty$ and fixed σ, a generic signal obeying $x[0], x[L-1] \neq 0$, the ratio M/N and σ can be recovered from the first three autocorrelation functions. If $\frac{M}{N} \geq \frac{1}{4(L-1)}$, then it holds for any nonzero x.*

4 Numerical experiments

4.1 One-dimensional experiments

Here will come the 1D experiment with heterogeneity

4.2 Two-dimensional experiments

In this section, we provide the details of the 2D experiment shown in Figure 1. In the experiment, we generated a sequence of 767×10^3 2D images ("micrograph") of size 2000×2000. Each micrograph had around ~ 900 signal occurrences (depends on the stopping criterion that we need to explain) and in total $M = 561 \times 10^6$. The noise level was $\sigma = 1$.

To accelerate the experiment, we estimated only the first two autocorrelations of the signal. As mentioned in Section 3.2, this data determines a 2D signal uniquely, up to possible reflection symmetry. In this experiment, we chose the correct reflection manually. To estimate the signal we used the Relax-Reflect-Reflect (RRR) algorithm, which is a known phase retrieval algorithm. Recall that the second-order autocorrelation is equivalent to the power spectrum of the signal. Starting form random initialization of size $(2L-1) \times (2L-1)$, its kth iteration take the form of

$$x_k = x_{k-1} + \beta(P_2(2P_1(x_{k-1}) - x_{k-1}) - P_1(x_{k-1}),$$

where P_1 is a projection that keeps the values of the $L \times L$ entries in the upper-left corner and zeros out all other entries, P_2 is a projection that keeps the Fourier phase of x_{k-1} and impose the correct Fourier magnitudes, and we set $\beta = 0.5$. The solution of the RRR scheme was used to initialized a LS estimation on the first two moments with signals of length $(2L-1) \times (2L-1)$ for the 200 iterations. Finally, we chose the $L \times L$ upper left corner and refined it using LS with 500 iterations.

5 Conclusion

Here we conclude the paper. Points to mention;

- cryo – EM
- structured background
- without separation
- sample complexity

References

[1] `http://spr.math.princeton.edu`.

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

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

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

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

[6] Afonso 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.

[7] 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.

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

[9] 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.

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

[11] 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.

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

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

[14] 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.

[15] Yohann De Castro and Fabrice Gamboa. Exact reconstruction using beurling minimal extrapolation. *Journal of Mathematical Analysis and applications*, 395(1):336–354, 2012.

[16] Vincent Duval and Gabriel Peyré. Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, 15(5):1315–1355, 2015.

[17] 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.

[18] 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.

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

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

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

[22] 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.

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

[24] 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.

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

[26] Shuyang Ling and Thomas Strohmer. Blind deconvolution meets blind demixing: Algorithms and performance bounds. *IEEE Transactions on Information Theory*, 2017.

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

[28] 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.

A Autocorrelation estimations

To analyze the asymptotic behavior of the data autocorrelation functions, we consider the case of one signal $K = 1$. The extension to $K > 1$ is straightforward by averaging the contributions of all signal with the appropriates weights, see [13].

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 means 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:

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

where the cross terms between the signal and the noise almost surely vanish in the limit.

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 [to recheck indicies]

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

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 $\ell = 0$,

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

We are now moving to 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 partial quantities. 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 partial sums 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 interdependency 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. [13].

B Proof of Proposition 3.3

We aim to prove that one can estimate both σ and $\beta = M/N$ from the observed first three moments. Using observed (measured) quantities, we construct two quadratic equations of β, independent of σ. Then, we show that these equations are independent and therefore M is uniquely defined. Given β, we can estimate σ using Proposition 3.2. 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 and 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:

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

where $F_1 := L\left(a_x^3[0,0] + 2\sum_{j=1}^{L-1}(a_x^3[j,j]\right)$. The terms of F_1 can be also estimated from a_y^3 with bias:

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

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

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

Let $a_y^2 := \sum_{j=0}^{L-1} a_y^2[j]$ and recall from Proposition 3.2

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

Plugging into (B.3) and rearranging we get

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

where

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

Importantly, these coefficients are are observable quantities.

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

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

where [to verify]

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

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

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

Taking the ratio:

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

we conclude:

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

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

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

where

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

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

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

In addition, using (B.6)

$$\frac{\mathcal{E}}{\mathcal{F}} = -\frac{(6L-3)(a_y^1)^3}{E3} = 3 - 6L.$$

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

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

Rearranging the equation and writing in terms of x we get

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

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 (B.9) cannot be satisfied for any signal.