Estimation below the detection limit

Tamir Bendory, Nicolas Boumal, William Leeb and Amit Singer

May 2, 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 locations in a data sequence y[This is the most important sentence of the paper, we need to polish it up]. The data may also contain background information—independent of the signals—which we model as noise. A precise mathematical formulation of the model and the estimation problem is provided in Section 2. For one-dimensional signals, the data sequence can be thought of as a long time series and the K signals as repeated short events. This model appears in many applications, including spike sorting [31], passive radar [24] and system identification [37]. In the last part of this paper, we also propose to interpreted this estimation problem as a toy model for the task of single particle reconstruction using cryo-electron microscopy (cryo–EM).

Figure 1 shows a simple example of data sequences with one signal ($K = 1$). The upper row presents the data without noise, in the left panel, and its cross-correlation with the signal itself in the right panel. The cross-correlation exhibits two peaks, corresponding to the two signal's locations in the data. In the middle row, the same data is shown, now contaminated with i.i.d. Gaussian noise with zero mean and standard deviation of $\sigma = 0.5$. While it is harder to identify the signal occurrences in the left panel, the cross-correlation still produces fairly adequate estimates. In practice one usually does not possess a precise template of the signal to be recovered, however, clever methods based on template matching, such as those used in structural biology [27] and radar [24], may work. The bottom panels show the same data swamped in i.i.d. Gaussian noise with standard deviation $\sigma = 3$. In this low signal–to–noise (SNR) regime, detection of individual signal occurrences is impossible, even if the true signal is known. This phenomenon—explained in more detail in Section 2–raises the following fundamental question:

Can we estimate the signals accurately without detecting their individual occurrences?

In this work, we provide a positive answer to this question. For a single signal ($K = 1$), we prove that in the asymptotic regime in which the signal appears infinitely many times, and under a spacing condition, one can estimate the signal *to any desired accuracy in any SNR level*. We show empirically that the same holds true for multiple signals if K is not too large.

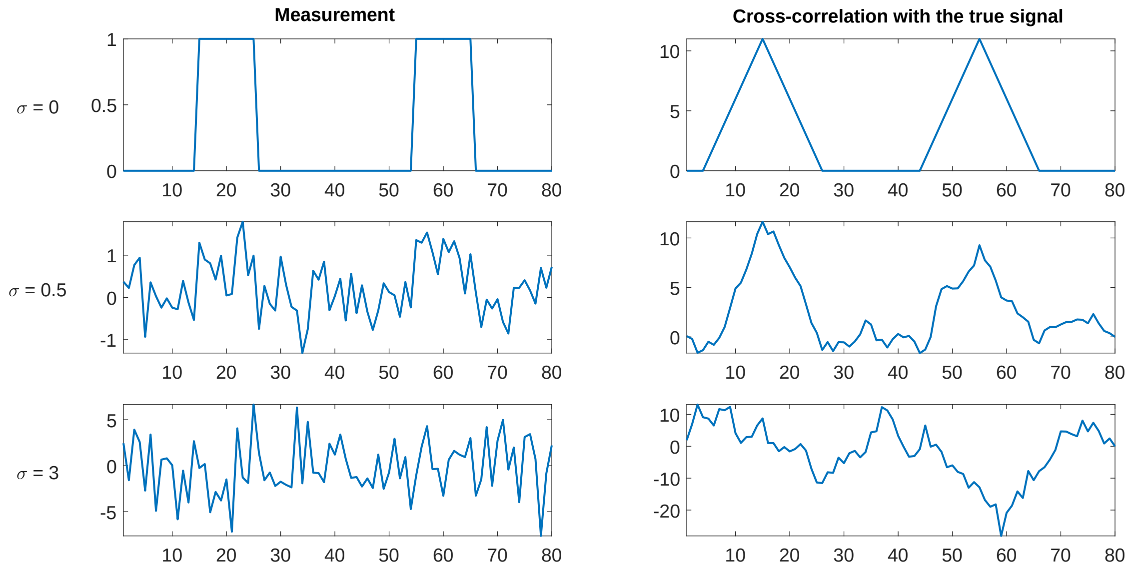

Figure 1: The upper left panel shows a data sequence (measurement) of length $N = 80$ with one rectangular signal of length $L = 11$ that appears twice. The upper right panel shows the cross-correlation of the data with the true rectangular signal. The second and third rows present the data sequence with additive Gaussian noise with mean zero and standard deviation $\sigma = 0.5$ and $\sigma = 3$, respectively, and its cross-correlation with the signal. Note to the different scales in different panels. Importantly, in the low SNR regime the cross-correlation does not provide information on the locations of the signal in the data.

[Do we want a sentence like: Based on recent results on the multireference alignment problem, we conjecture that it remains true as long as $K \lesssim L/6$. ? I am not sure it is necessary.]

Our framework is based on 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 at least $L - 1$ entries and each signal appears sufficiently many times in the data. There is no need to detect individual occurrences. Furthermore, we do not assume the knowledge of the number of signal occurrences M or the noise level. In the second stage, the signals are estimated from the mixed autocorrelations using a nonconvex least-squares (LS). Section 3 elaborates on the technique and proves some of its properties, while Section 4 shows numerical demonstrations and discusses computational aspects.

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 $\{1, \dots, N\}$. In total, even when excluding forbidden combinations due to the spacing constraint, there are $O(N^M)$ such combinations. [cumbersome]

2 Model and related literature

Let $x_1, \ldots, x_K \in \mathbb{R}^L$ 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:

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

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.

In order to estimate the mixture of autocorrelations, we assume that the support of s is not clustered. In particular,

$$\text{For all } i, j \in \mathcal{S},\ i \neq j, \quad |i - j| \geq L - 1. \tag{2.2}$$

The goal of the problem is to estimate $x_1, \ldots, x_K$ 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 [28, 39, 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 (2.2). Other settings of blind deconvolution problems have been analyzed recently, see for instance [5, 33, 32, 35, 36, 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.

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.

For $K = 1$, our problem can be also interpreted as a special case of the system identification problem. Similarly to (2.1), the system identification forward model takes the form

$$y = x * w + \varepsilon, \tag{2.3}$$

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, 29]. In the special case where $w \in \{0,1\}^N$, satisfying the spacing requirement (2.2), 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 [38]. 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 [34, 23, 40].

3 Autocorrelation analysis

Our method for estimating the signals is composed of two stages. First, we use the autocorrelation functions of the data to estimate a mixture (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 spacing condition (2.2) is met. Then, we use a nonconvex LS to estimate the signals from their mixed autocorrelations. In this section, we elaborate on the autocorrelation functions and their estimations, while the precise recovery procedure, based on nonconvex optimization, will be discussed in detail 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 autocorrelation of order k is defined for any integer shifts $\ell_1, \ldots, \ell_{k-1}$ by

$$a_z^k[\ell_1, \ldots, \ell_{k-1}] = \sum_{i=-\infty}^{+\infty} z[i]z[i+\ell_1]\ldots z[i+\ell_{k-1}], \tag{3.1}$$

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

$$\begin{aligned}
a_z^1 &= \sum_{i=0}^{L-1} z[i], \\
a_z^2[\ell] &= \sum_{i=\max\{0,-\ell\}}^{L-1+\min\{0,-\ell\}} z[i]z[i+\ell], \\
a_z^3[\ell_1,\ell_2] &= \sum_{i=\max\{0,-\ell_1,-\ell_2\}}^{L-1+\min\{0,-\ell_1,-\ell_2\}} z[i]z[i+\ell_1]z[i+\ell_2].
\end{aligned} \tag{3.2}$$

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

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

Additionally, if the moments of the signal depend only on the difference between the indices (Toeplitz structure), then they are equivalent to the autocorrelation functions.

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

Proposition 3.1. *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 3.1 shows that there exists a very simple estimator that has finite sensitivity. In the next section, we propose 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 (2.2) 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 (2.2) 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) [25, 26]. 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.

[NB: I would remove most if this paragraph; we can have a similar discussion but for our case (after having discussing which moments we keep); make the statement, and finish with a one-sentence reference to the MRA paper. TB: I didn't touch it yet. We may want to put the degrees-of-freedom calculations from Section 4 here] The invertibility of the autocorrelations for $K > 1$ was explored for the related case of periodic autocorrelation functions. Recall that the periodic autocorrelations are defined similarly to (3.1), with two differences: the sum goes to $L-1$ and all indices should be taken modulo L. If $z[n] = 0$ for $n = L/2, \dots L-1$ ("zero-padded"), then the periodic and aperiodic autocorrelation coincide. In [9], 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, at least as long as $K \leq \sqrt{L}$ [19].

3.2 Estimating autocorrelations from the data with a single signal $K = 1$

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 in the next section.

In order to estimate the autocorrelations of the signal, we first compute the first L entries of the data's autocorrelations. 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

$$\gamma = \frac{ML}{N}. \tag{3.3}$$

Under the spacing constraint (2.2), 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 (2.2), then the first L entries of the data autocorrelations converge to a scaled, biased, version of the signal's autocorrelation:

$$\begin{aligned} \lim_{N\to\infty} a_y^1 &= \gamma a_x^1, \\ \lim_{N\to\infty} a_y^2[\ell] &= \gamma a_x^2[\ell] + \sigma^2\delta[\ell], \\ \lim_{N\to\infty} a_y^3[\ell_1,\ell_2] &= \gamma a_x^3[\ell_1,\ell_2] + \sigma^2\gamma a_x^1\big(\delta[\ell_1,0] + \delta[0,\ell_2] + \delta[\ell_1,\ell_2]\big), \end{aligned} \tag{3.4}$$

for $\ell, \ell_1, \ell_2 = 0, \dots L-1$, and where δ denotes the Kronecker delta function. These relations are proven in Appendix A. The analysis is similar to [14, 19], yet a particular caution should be taken with the statistical dependencies of the noise entries. The relations (3.4), together with Proposition 3.1, imply that given M and σ, one can estimate the signal for any noise level if M is large enough. Next, we show that M and σ are also uniquely determined from the autocorrelations.

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

Proposition 3.2. *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 (3.4) 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 3.3. *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 B. □

From Propositions 3.1 and 3.3 we can directly deduce the following:

Corollary 3.4. *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 $\frac{M}{N} \geq \frac{1}{4(L-1)}$.*

3.3 Estimating autocorrelations from the data with multiple signals $K > 1$

As before, we consider the asymptomatic regime where $M_1, \ldots, M_K, N \to \infty$, while preserving fixed ratios

$$\gamma_k = \frac{M_k L}{N}, \quad \gamma = \sum_{k=1}^{K} \gamma_k. \tag{3.5}$$

If the support s satisfies the spacing constraint (2.2), then one can estimate the mixture of the K signals' autocorrelations, similarly to (3.4):

$$\begin{aligned}
\lim_{N\to\infty} a_y^1 &= \sum_{k=1}^{K} \gamma_k a_{x_k}^1, \\
\lim_{N\to\infty} a_y^2[\ell] &= \sum_{k=1}^{K} \gamma_k a_{x_k}^2[\ell] + \sigma^2 \delta[\ell], \\
\lim_{N\to\infty} a_y^3[\ell_1, \ell_2] &= \sum_{k=1}^{K} \gamma_k a_{x_k}^3[\ell_1, \ell_2] + \sigma^2 \left(\sum_{k=1}^{K} \gamma_k a_{x_k}^1 \right) (\delta[\ell_1, 0] + \delta[0, \ell_2] + \delta[\ell_1, \ell_2]),
\end{aligned} \tag{3.6}$$

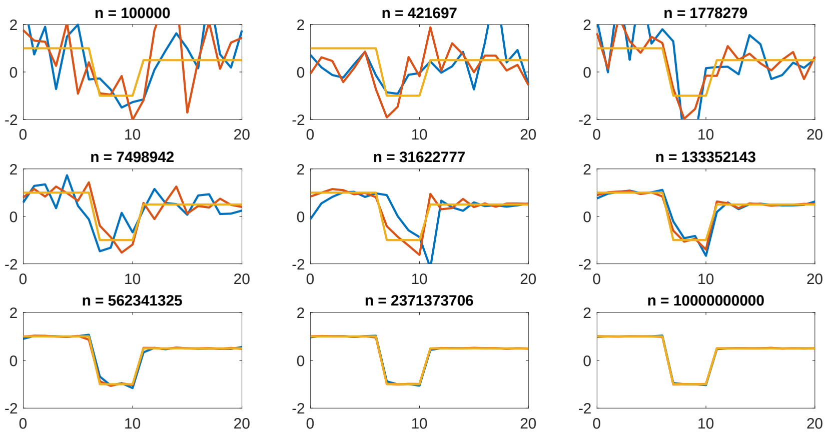

Figure 2: [...] [This figure is with new ROI method based on loss functions]

This relation is proven in Appendix A.

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 σ^2. Taking the dth order autocorrelation function amplifies the variance of the noise by a factor of d. Therefore, the required number signal occurrences should scale like σ^{2d} to retain some constant estimation error. Accordingly, in our method, M must scale 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. [question mark on the entire paragraph] [NB: We could move here the paragraphs from 1D XP about which moments we compute, to argue that we can expect to recover at most up to $K = L/2$ signals.] [TB: I didn't touch for now.]

4 Numerical experiments

[Revise to explain $K = 1$ experiment first, then explain $K = 3$.]

For the 1D experiment, we fix $K = 3$ signals of length $L = 21$, as depicted in Figure [ref]. Following the data model described in Section [ref], we generate an observation y of length $24.6 \cdot 10^9$. Each of the three signals appears, respectively (and approximately) $300 \cdot 10^6$, $200 \cdot 10^6$ and $100 \cdot 10^6$ times in y, such that at least $L - 1$ zeros separate two occurrences of any signals. This is done by randomly selecting $600 \cdot 10^6$ placements in y, one at a time with an accept/reject rule based on the separation constraint and locations picked so far. For each placement, one of the three signals is picked at random proportionally to the desired number of occurrences of each. Then, i.i.d. Gaussian noise with mean zero and standard deviation $\sigma = 3$ is added, to form the observed y. The SNR of y is about $1/12$. This is enough noise to make cross-correlations of y even with the true signals display peaks at random locations, uninformative of the actual locations of the signal occurrences. Thus, we contend that it would be difficult for any algorithm to locate the signal occurrences, let alone to classify them according to which signal appears where.

Given the observation y, we proceed to compute the moments. The first-order moment

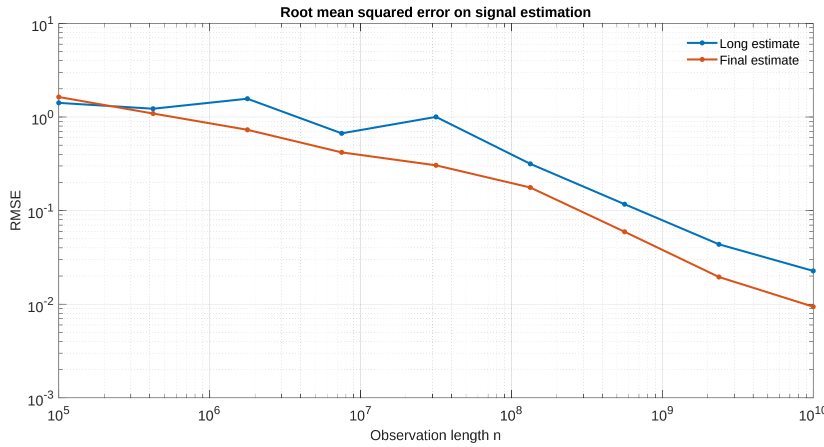

Figure 3: [...] [This figure is with new ROI method based on loss functions]

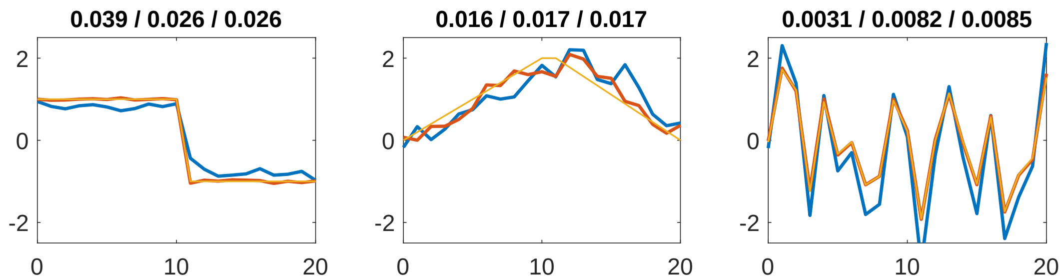

Figure 4: [Thin yellow is ground truth; blue is ROI of first optimization; red is final estimation. Relative error for blue is .37, and .13 for red. Individual relative errors of the red estimates: 0.0239393 / 0.208925 / 0.0335956]

is straightforward. For second-order moments, notice from equation (3.6) 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 $0 \le \ell_1, \ell_2 \le L-1$ such that $\ell_2 \le \ell_1$ includes all relevant moments for our purpose [Do we want to include a figure to explain why that is?], 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 [TB: We may want to put this calculation is the last paragraph of Section 3.1]

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

moments in total. In practice, these are computed on disjoint segments of y of length $100 \cdot 10^6$ and added up, without correction for the junction points. Segments are handled sequentially on a GPU, as GPUs are particularly well suited to execute simple instructions across large vectors of data. If multiple GPUs are available, segments can of course be handled in parallel.

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

$$\min_{\substack{\hat{x}_1,\ldots,\hat{x}_K \in \mathbb{R}^W \\ \hat{\gamma}_1,\ldots,\hat{\gamma}_K > 0}} w_1 \left(a_y^1 - \sum_{k=1}^{K} \hat{\gamma}_k a_{\hat{x}_k}^1 \right)^2 + w_2 \sum_{\ell=1}^{L-1} \left(a_y^2[\ell] - \sum_{k=1}^{K} \hat{\gamma}_k a_{\hat{x}_k}^2[\ell] \right)^2 + \\ w_3 \sum_{\substack{2 \le \ell_1 \le L-1 \\ 1 \le \ell_2 \le \ell_1 - 1}} \left(a_y^3[\ell_1, \ell_2] - \sum_{k=1}^{K} \hat{\gamma}_k a_{\hat{x}_k}^3[\ell_1, \ell_2] \right)^2. \tag{4.1}$$

where $W \geq L$ is the length of the sought signals and [explain w_i's: currently they are $w_1 = 1/2, w_2 = 1/2n_2, w_3 = 1/2n_3$, where n_2, n_3 are the number of moments used: $n_2 = L-1$, $n_3 = \frac{(L-1)(L-2)}{2}$. Issue is: this is not very smart..]. 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 (with cyclic indexing [we should understand / explain this]) that has largest ℓ_2-norm. This estimator is then used as initial iterate for (4.1), this time with $W = L$. We find that this procedure is reliable for a wide range of experimental parameters. To solve (4.1), 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. [Should we do variable projection for the gammas, that is, exploit the fact the problem is a regular least squares in the gammas (up to the positivity constraints) to substitute the explicit optimum for them? Not sure it's worth the effort. – Ok, it's probably not a good idea, because even with fixed gammas to the correct value, optimization takes a while.] [Do we still need to stress at this point that the optimization part has complexity independent of length of observation? Should be pretty clear at this point already.]

[TB: 2D experiment?]

5 Open questions

In this paper, we considered the problem of estimating a set of signals from their multiple occurrences in unknown locations in the data. We focused on low SNR environments in which detection of individual occurrences is impossible. Our technique is based on computing the third-order autocorrelation function of the data with asymptotic estimation rate proportional to σ^6/M [TB: it's not clear to me yet whether we discuss the estimation rate in the paper]. Based on related results in multireference alignment [2], we believe that it is the optimal estimation rate, however, this is yet to be proven.

Our results rely on two core assumptions that are not necessarily met by applications. 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. It raises the question under what conditions on the background structure and statistics one can still estimate the signals in the low SNR regime.

In addition, we assumed that the signal occurrences are all separated by $L-1$ entries, see (2.2). If the signals are not separated, one can introduce a new variable p that represents the distribution of the spacing between signal occurrences. The first entry $p[1]$ will represent the probability that two consecutive signals are separated by only one entry, $p[2]$ the probability for spacing of two entries and so on. Using this auxiliary variable p, one can write explicitly the relation between the autocorrelation functions of the data and those of the signal in a similar way to (3.6). An interesting question is under what conditions on p and the signals, one can estimate the signals from the data.

6 Connection with the cryo–EM problem

Cryo–EM is an innovative technology for reconstructing the 3D 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 [30, 10]. 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. Importantly, 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). 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 2D tomographic projections of the samples from the noisy micrograph. The output of ideal particle picker is a series of 2D images, each contains one centered tomographic projection associated with an unknown 3D orientation. This series of images is then used to estimate the structure. However, due to the high noise level, the performance of particle pickers is not ideal. For instance, the projections in the 2D images are typically not centered, increasing dramatically the number of parameters involved in the estimation problem. In addition, the information from particles that are too close to each other is usually neglected. Hence, valuable information that can be harnessed is omitted.

The model considered in this paper can be interpreted as a preliminary step towards investigation of the possibility to *estimate the molecule's structure directly from the micrograph*, that is, by skipping the particle picking step. In our model, the noisy data y is the analog of the micrograph. The K sought signals correspond to K different molecule's projections, each taken from a different, unknown, viewing direction. The position of individual projections are nuisance variables of the cryo–EM problem, in analogy to individual signal's locations in our setup. That being said, the cryo–EM data is far more complicated that the simplified model of this paper. In a future research, we hope to bridge this gap.

References

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

[2] Emmanuel Abbe, João M Pereira, and Amit Singer. Estimation in the group action channel. *arXiv preprint arXiv:1801.04366*, 2018.

[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 non-minimum 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 super-resolution of positive measures. *Journal of Fourier Analysis and Applications*, 23(5):1153–1194, 2017.

[23] 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.

[24] 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.

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

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

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

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

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

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

[34] 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.

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

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

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

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

[39] 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.

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

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:

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

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 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. 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 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, 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:

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

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:

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

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:

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

Plugging into (B.3) and rearranging we get

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

where

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

Importantly, these coefficients are observable quantities.

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

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

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

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

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

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

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 (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

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

In addition, using (B.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

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

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.