Estimation below the detection limit

Tamir Bendory

March 13, 2018

Abstract

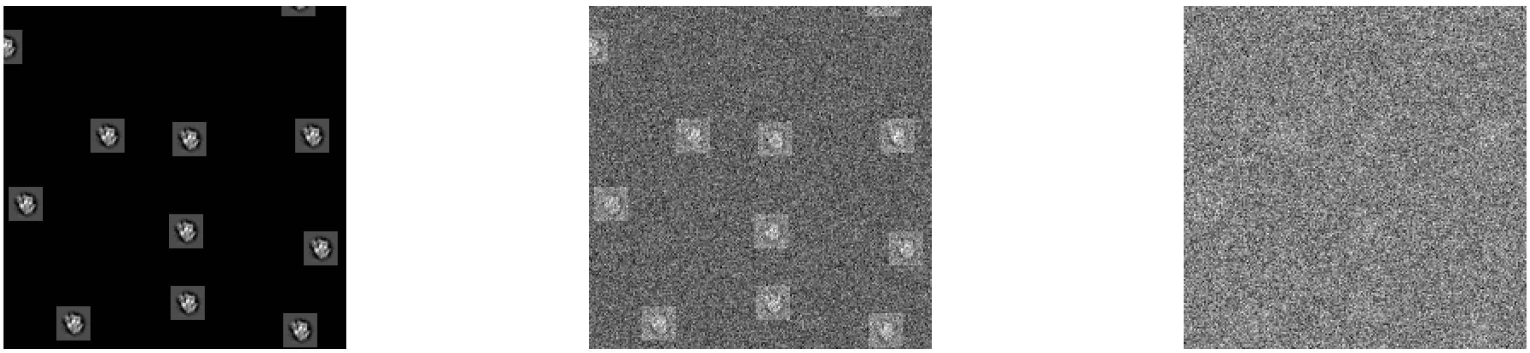
Here comes the abstract

1 Introduction

In this paper, we consider the problem of estimating a set of signals $x_1, \ldots, x_K$ which appear multiple times 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, the data is a big image that contains 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 [21], passive radar [16] and system identification [26]. In Section 2 we provide a precise mathematical formulation of the problem.

If the noise level is negligible, estimating the signals is easy. One can detect the signal occurrences in the data sequence and then estimate the signals using clustering and averaging. Even in higher noise level regimes, clever methods based on template matching, such as those used in structural biology [19] and radar [16], may work. However, in the low signal–to–noise (SNR) regime, detection of individual signal occurrences is impossible as clarified in Section 2. In Figure 1 we demonstrate an instance of the problem in different noise levels. Figure 1a shows a excerpt of a big image that contains many repetitions of 21×21 small image. The small image is a downsampled version of TKTK, which is shown in Figure 1d. In Figures 1b and 1c the same excerpt is shown with the addition of i.i.d. Gaussian noise with standard deviations of 0.2 and 1, respectively.

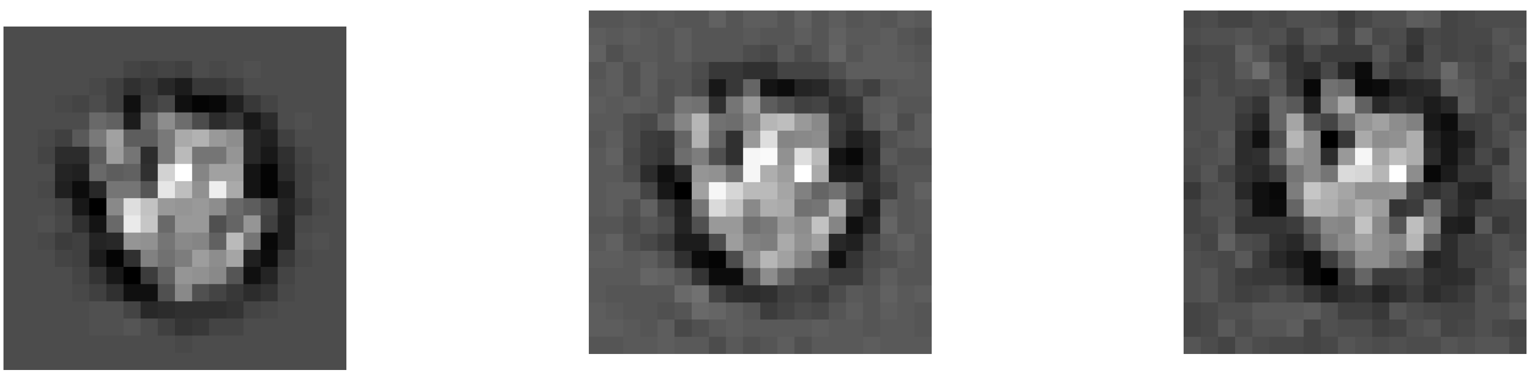
In this work we focus on the low SNR regime. To estimate the signal we use autocorrelation analysis of the data. In a nutshell, the method consists of two stages. First, we estimate a mix of the low order autocorrelations of the signals from the data. These quantities can be estimated to any desired accuracy if the signals appear enough times in the measurements, without the need to detect individual occurrences. Then, the signals are estimated from the mixed autocorrelations. In Section 3, we elaborate on our method and show that if K not too large, it is possible to estimate the signals, in any SNR level and to any desired accuracy. This holds if each signals appears enough times in the data and the signals are sufficiently spaced. Interestingly, alternative methods which are often employed in similar estimation problems, such as expectation-maximization (EM) or Monte-Carlo Markov Chain (MCMC), seems to be intractable for this problem. In Figure 1e we show an example for estimation of the signals where the data is contaminated with noise with standard deviation of $\sigma = 1$ as presented in Figure 1c. The normalized error between the signal and its estimation is TKTK. In Section 4 we provide the details of this experiment and show additional experiments with $K > 1$.



(a) Excerpt of clean data

(b) Excerpt of noisy data with $\sigma = 0.2$

(c) Excerpt of noisy data with $\sigma = 1$



(d) Clean signal

(e) 710K micrographs, $M = 561Mega$, error = 0.05062

(f) Recovery with 140K micrographs and . $110Mega$, error = ?

Figure 1: Example

This work is primely motivated by innovative technologies for single particle reconstruction, such as cryo-electron microscopy (cryo–EM) and X-ray Free–Electron Laser (XFEL). The acquired data in these modalities 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 will revisit the cryo–EM problem and draw connections with the estimation problem under consideration.

2 Model

Let $x_1, \ldots, x_K \in \mathbb{R}^L$ be the sought signals and let $y \in \mathbb{R}^L$ 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 non-zero 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 $j=0,\ldots,L-1$. We denote the cardinality of s_i by M_i and $M=\sum_{i=1}^{K} M_i$. We do not assume the knowledge of M_i's, neither of M.

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 $\{1,\ldots,N\}$ and we set $s[i_1]=1$. A second index i_2 is then drawn uniformly from $\{1,\ldots,N\}$. 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 non-zero entries to s, while keeping the separation condition until some halting criterion is obtained. Ultimately, each non-zero value of s is associated with one of the s_i's uniformly over $\{1,\ldots,K\}$. 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.

The simplest way to present the forward model is a mix of blind deconvolution problems between the support signals and the unknown 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 $x_1,\ldots,x_K$ 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 [20, 27, 4, 1], just to name a few. Clearly, without additional information on the signal, blind deconvolution is ill-posed. 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 [3, 23, 22, 24, 25, 13] where the focus is on high SNR regimes.

In our setup, the support signals $s_1,\ldots,s_K$ are the *latent* variables of the problem. 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 [14, 15, 10, 7, 11]. 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 contain one occurrence 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 signals within each window is an essentially easier problem the detecting the support of s since the windows are given to us by the oracle. Apparently, 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 [2]. Therefore, we conclude that that detecting the non-zero 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 relies on two pillars. First, we use the autocorrelation functions of the data to estimate a mix (i.e., linear combination) of the K signals 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 held. Then, we use least-squares (LS) estimator to estimate the signals from their mixed autocorrelation. From pedagogical reasons, we start with the second stage of recovering signals from their autocorrelations. In this section we mainly discuss the uniqueness of the mapping from the autocorrelation to the signals. Concrete algorithms will be discussed in the next section.

For the purpose of this paper, we need the first three (aperiodic) autocorrelation functions of a signal $z \in \mathbb{R}^W$, defined as

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

Note that the autocorrelation functions are symmetric so that $a_z^2[\ell] = a_z^2[-\ell]$ and $a_z^3[\ell_1,\ell_2] = a_z^3[-\ell_1,-\ell_2]$.

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

Proposition 3.1. *Suppose that $z[0]$ and $z[L-1]$ are non-zeros. Then, a signal $z \in \mathbb{R}^W$ 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, from the third-order autocorrelation we can compute $z[k]$ for all $k = 0, \dots L-1$ by

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

$\square$

We note that the length of the signal can be easily derived from the second-order autocorrelation of the signal. Therefore, the assumption that $z[0]$ and $z[L-1]$ is met in practice. We also note that for one-dimensional signals, the second-order autocorrelation does not determined the signals uniquely [6, 8]. This is not the case for dimensions greater than one, in which 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) [17, 18, 8]. 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 (Z_2 symmetry), one needs to use the third moment.

The heterogeneous case $K > 1$ was explored for periodic autocorrelation functions. The aperiodic autocorrelations are the periodic autocorrelations of signals padded by zeros. Therefore, the following results hold for our setup as well [we should be prudent here to make sure that this set does not "fall" in the zero measure]. In [5], it was shown that a mix of K third-order autocorrelations determine a finite list of K generic signals when

- $K = 2$ when $L > 1$,
- $K = 3$ when $L > 12$,
- $K = 4$ when $L > 18$,
- $5 \leq K \leq 15$ when $L \geq 6K - 5$.

The authors conjecture that for $K > 15$, $L \geq 6K - 5$ is enough. Recently, Joe told use that he has proven that there is unique mapping in the same regime. This also holds for arbitrary weights γ_i which are rational fractions.

[Where we should refer to Gianakis's paper?]

We are moving forward to discuss the second question of how to estimate the autocorrelation of the signals from the data. The analysis is conducted in the asymptomatic regime where $M_1, \ldots, M_K, N \to \infty$. We define the ratios

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

and $\gamma = \sum_{i=1}^{K} \gamma_i$. Under the separation condition, we have $\gamma \leq \frac{L}{2L-1} \approx 1/2$. The main insight is that if s satisfies the separation condition, then the first L entries of the data autocorrelations converge to a γ scaled version of a mix of the signals autocorrelations:

$$\lim_{N\to\infty} a_y^1 = \sum_{i=1}^{K} \gamma_i a_{x_i}^1, \tag{3.4}$$

$$\lim_{N\to\infty} a_y^2[\ell] = \sum_{i=1}^{K} \gamma_i a_{x_i}^2[\ell] + \sigma^2\delta[\ell], \tag{3.5}$$

$$\lim_{N\to\infty} a_y^3[\ell_1, \ell_2] = \sum_{i=1}^{K} \gamma_i a_{x_i}^3[\ell_1, \ell_2] + \sigma^2 \left(\sum_{i=1}^{K} \gamma_i a_{x_i}^1\right) (\delta[\ell_1, 0] + \delta[0, \ell_2] + \delta[\ell_1, \ell_2]), \tag{3.6}$$

for $\ell, \ell_1, \ell_2 = 0, \ldots L-1$. These relations are proved in Appendix A. The analysis is similar to [9, 12], yet a special caution should be taken with the noise dependencies. This means that given $M_1, \ldots, M_K$ and σ^2 and K does not exceed the limit for uniqueness, the one can estimate the signals from the third-order autocorrelation of the data. In other words, the signals can be estimated in asymptotic estimation rate of σ^6/N.

If the noise level σ^2 is known, then for $K = 1$, one can estimate M from only the first two moments, namely, with estimation rate of σ^4/N.

Proposition 3.2. *Let $K = 1$. Then,*

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

Proof. The relation is proved by plugging the definitions of the signal autocorrelations into the right-hand side of the equation (3.1). □

If we assume that s is sparse, then by ignoring the separation condition, one can describe the generative process of s as a Bernoulli process with parameter M/N. Therefore, in low SNRone can estimate the parameter that control this statistical process, although cannot estimate individual entries.

Estimating σ and M?

4 Numerical experiments

5 Conclusion

Here we conclude the paper: cryo – EM, structured background, without separation, heterogeneity In particular, in cryo–EM, multiple projections of a molecule (signal), taken from unknown viewing directions, are recorded on two-dimensional detector array. Current algorithms try to detect these projections in a low SNR regime and then use them for the reconstruction process. Since the Then, the reconstruction problem is to detect these projection, and then use them to estimat The data acquired in both technologies is composed of multiple projections of a molecule (the signal), taken from unknown viewing direction, and drawn in high noise level. The reconstitution problem is then to estimate the molecule from this data.

References

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

[2] 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.

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

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

[5] 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.

[6] 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.

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

[8] 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.

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

[10] 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.

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

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

[13] 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.

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

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

[16] 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.

[17] 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.

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

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

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

[21] 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.

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

[23] 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.

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

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

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

[27] 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 one signal $K = 1$. The extension to $K > 1$ is straightforward by averaging the contributions of all signal with the appropriates weights, see [12].

Let us define

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

By assuming $M_N = \Omega(N)$, we also have $\gamma > 0$. 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 moment for fixed $\ell \in [0, \ldots, L-1]$. Then, 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

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

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 therefore converge 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, this sum can be broken into eight partial sums. The first contains only signal terms 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, then these terms go to zero almost surely.

We next analyze the contribution of triple product 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 is a triple product of Gaussian variables with zero mean. Therefore, it is also converges to zero.

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 converges 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 $\gamma\sigma^2 a_x^1$.