Multi-target detection and clustering with application to cryo-electron microscopy

Tamir Bendory[a], Nicolas Boumal[b], William Leeb[c], Eitan Levin[a,b], and Amit Singer[a,b]

[a]The Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ, USA
[b]Department of Mathematics, Princeton University, Princeton, NJ, USA
[c]School of Mathematics, University of Minnesota, Minneapolis, MN, USA

February 14, 2019

Abstract

abstract

1 Introduction

1.1 Problem formulation

In this paper, we consider the problem of recovering a signal that appear multiple times at unknown locations in a noisy measurement. Let $x \in \mathbb{R}^L$ be the sought signal and let $y \in \mathbb{R}^N$ be the observed data, where we assume N is far larger than L. Let $s \in \{0,1\}^{N-L+1}$ be a binary signal indicating (with ones) the starting positions of all occurrences of x in y, so that, with additive white Gaussian noise the measurement model reads:

$$y = x * s + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I_N), \tag{1.1}$$

where $*$ denotes linear convolution. While both x and s are unknown, the goal is only to estimate x from y. We name this problem *multi-target detection*.

The multi-target detection problem is an instance of *blind deconvolution*—a long-standing problem arising in a variety of engineering and scientific applications, such as astronomy, communication, image deblurring, system identification and optics; see [24, 39, 5, 2], just to name a few. Different variants of the blind deconvolution problem have been subject recently to a thorough analysis, focusing on low noise regimes [4, 33, 32, 29, 34, 27]. A key difference is that in multi-target detection the parameters of the signal s (the locations of its nonzero values) are *nuisance variables*—that is, we aim only at estimating only x. Multi-target detection appears in several scientific applications, including spike sorting [31], passive radar [20] and system identification [36].

In the low noise regime, one can estimate s and then estimate x by standard deconvolution techniques. However, in the high noise regime, estimating s is impossible—that is, detection is impossible [13, 3]. To guarantee recovery in any noise regime, we require a separation between adjacent signal's occurrences, that is,

$$\text{If } s[i] = 1 \text{ and } s[j] = 1 \text{ for } i \neq j, \text{ then } |i - j| \geq 2L - 1. \tag{1.2}$$

In words: the starting positions of any two occurrences must be separated by at least $2L - 1$ positions, so that their end points are necessarily separated by at least $L - 1$ signal-free entries in the data. In then next section we present a model that alleviates this condition.

1.2 Extensions

The basic model (1.1) can be extended in various ways, some discussed below.

Multi-target detection and clustering with separation. The multi-target detection model can be naturally generalized to the problem of estimating a set of K signals $x_1, \ldots, x_K \in \mathbb{R}^L$. Let s be a binary signal satisfying the separation condition (1.2). For each $s[i] = 1$, a random variable v_i is drawn from some (possibly unknown) distribution over $\{1, \ldots, K\}$. Then, x_{v_i} is placed in y with element 0 at location i. We refer to this model as *multi-target detection and clustering*.

This model can be written as a mix of blind deconvolution problems

$$y = \sum_{i=1}^{K} x_i * s_i + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I_N), \tag{1.3}$$

where again s_i is a binary signal, which its non-zero values are the subset of the non-zero values of s associated with x_i. Generally, we refer to this problem as *multi-target detection and clustering with separation.*

The well-separated model with a random signal The homogeneous and heterogeneous models can be generalized as follows. We let the signal $X \in \mathbb{R}^L$ to be a random vector drawn from some fixed distribution. For each non-zero element of s, a random vector X from the distribution is then placed in y, with element 0 at location i. If the distribution is a Delta function that this model coincides with the multi-target detection model; if it is a sum of Diracs then it reduces to multi-target detection and clustering.

Poisson model for the support signal. The models we described so far assume a well-separated support. The separation condition can be alleviated by assuming a Poisson process. We consider the following observation model. Let $X \in \mathbb{R}^L$ be a random vector drawn from some fixed distribution. Points are chosen in $\{1, \dots, N-L+1\}$ according to a Poisson process with parameter $\gamma(N-L)$. For each point i that is chosen from 1 to $N-L+1$, a random vector X from the distribution is then placed in y, with element 0 at location i, with overlapping vectors being added together.

If M_i denotes the number of hits at location i, then by definition of the Poisson process M_i's are i.i.d. and $M_i \sim \text{Poisson}(\gamma)$. Conditional on the value of $M = (M_1, \dots, M_{N-L+1})$, if we let $X_1^i, \dots, X_{M_i}^i$ denote the random vectors with position 0 located at i, then $X_{k_1}^i$ and $X_{k_2}^i$ are independent for $k_1 \neq k_2$. With this notation, we can write each entry as:

$$y[i] = \sum_{j=0}^{L-1} \sum_{k=1}^{M_{i-j}} X_k^{i-j}[j]. \tag{1.4}$$

1.3 Motivation: single-particle reconstruction using cryo-electron microscopy

Cryo-electron microscopy (cryo-EM) has recently joined X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy as a high-resolution structural method for biological macromolecules; see for instance [18, 26, 10]. In contrast X-ray and NMR which aggregate information from an ensembles of particles, single particle cryo-EM produces images of individual particles and thus can, in principle, elucidate multiple structures. In addition, it does not require the formation of crystalline arrays of macromolecules.

In a cryo–EM experiment, biological samples are rapidly frozen in a thin layer of vitreous ice. The microscope produces a 2-D tomographic image of the samples embedded in the ice, called a *micrograph*. Each micrograph contains tomographic projections of the samples at unknown locations and under unknown viewing directions. The goal is to construct 3-D models of the molecular structure from the micrographs. Importantly, to keep radiation damage within acceptable bounds, the dose must be kept low, leading to high noise levels.

All contemporary methods in the field split the reconstruction procedure into two main stages. The first stage consists in extracting the particle projections from the micrographs. This stage is called *particle picking*. The second stage aims to construct a 3-D model of the molecular structure from these projections. The quality of the reconstruction eventually hinges on the quality of the particle picking stage. Crucially, it can be shown that reliable detection of individual particles is impossible below a certain critical SNR. This fact has been recognized early on by the cryo-EM community. Particularly, in [22, 19], it was established that particle picking is impossible for molecules below a certain weight (below ∼50 kDa).

Another potential pitfall of particle picking pertains to *model bias*, whose importance in cryo-EM was stressed by a number of authors [40, 43, 23, 42]. In the classical "Einstein from noise" experiment, multiple realizations of pure noise are aligned to a picture of Einstein using template matching and then averaged. In [40], it was shown that the averaged noise rapidly becomes remarkably similar to the Einstein template. In the context of cryo–EM, this experiment exemplifies how prior assumptions about the particles may influence the reconstructed structure. This model bias is common to all particle picking methods based on template matching.

A recent work of the authors suggests to bypass the particle picking stage and reconstruct the 3-D structure directly from the micrograph [13]. In that paper, it was shown that—at least in principle—the limits particle picking imposes on molecule size do not necessarily translate into limits on particle reconstruction. The principle mathematical tool is *autocorrelation analysis*, described in detail in Section 2. This goal of the current paper is to provide a theoretical justification and numerical support for the method proposed in [13]. In this context, the models described above serve as an abstraction of the cryo-EM problem: the random signal X discussed in Section 1.2 can be thought of as 2-D random tomographic projections of the 3-D structure taken according to some unknown distribution of the particles within the ice.

We mention that [13] was not the first paper to employ autocorrelation analysis to cryo-EM. Zvi Kam [25] first proposed autocorrelation analysis for 3-D reconstruction, under the assumption of perfect particle picking: his method used autocorrelations of the picked, perfectly centered, particles. His method has been extended and employed by in X-ray free electron lasers and cryo-EM; see for instance [35, 28, 30, 44]. In order to investigate the computational and statistical properties of Kam's method, a series of papers have studied a simplified model, called *multi-reference alignment* [8, 14, 6, 38, 7, 1]. We follows the same line of research by considering the multi-target detection and clustering as an abstraction to the application of reconstructing 3-D structures directly from the micrograph as proposed in [13].

2 Autocorrelation analysis

In order to recover the signal, we employ autocorrelation analysis. The underlying principle is to relate the autocorrelations of the observation to the autocorrelations of the signal without trying to detect individual signal occurrences (i.e., the signal s). In the models described above, for any noise level, these autocorrelations can be estimated to any desired accuracy in the limit of $N \to \infty$. The autocorrelations of the observation are straightforward to compute and require only one pass over the data. After estimation of the density of particles in the micrographs, these directly yield estimates for the mixed autocorrelations of the signals. To estimate the signals them from their estimated autocorrelations, we solve a nonlinear inverse problem as explained in Section 4.

Let $Z \in \mathbb{R}^m$ be a random signal drawn from some fixed distribution. The autocorrelation of order $q = 1, 2, \ldots$ is given for any integer shifts $\ell_1, \ldots, \ell_{q-1}$ by

$$a_Z^q[\ell_1, \ldots, \ell_{q-1}] = \mathbb{E}\left\{ \frac{1}{m} \sum_{i=-\infty}^{\infty} Z[i]Z[i+\ell_1] \cdots Z[i+\ell_{q-1}] \right\}, \tag{2.1}$$

where the expectation is taken with respect to the distribution of Z. The indexing of Z out of the range $0, \ldots, m-1$ is zero-padded. Explicitly, the first-, second- and third-order autocorrelations are given by:

$$\begin{aligned} a_Z^1 &= \frac{1}{m} \sum_{i=0}^{m-1} \mathbb{E}\left\{ Z[i] \right\}, \\ a_Z^2[\ell] &= \frac{1}{m} \sum_{i=\max\{0,-\ell\}}^{m-1+\min\{0,-\ell\}} \mathbb{E}\left\{ Z[i]Z[i+\ell] \right\}, \\ a_Z^3[\ell_1, \ell_2] &= \frac{1}{m} \sum_{i=\max\{0,-\ell_1,-\ell_2\}}^{m-1+\min\{0,-\ell_1,-\ell_2\}} \mathbb{E}\left\{ Z[i]Z[i+\ell_1]Z[i+\ell_2] \right\}. \end{aligned} \tag{2.2}$$

The autocorrelation functions have symmetries. Specifically,

$$\mathbb{E}\left\{ a_Z^2[\ell] \right\} = \mathbb{E}\left\{ a_Z^2[-\ell] \right\},$$

and [to verify]

$$\mathbb{E}\left\{ a_Z^3[\ell_1, \ell_2] \right\} = \mathbb{E}\left\{ a_Z^3[\ell_2, \ell_1] \right\} = \mathbb{E}\left\{ a_Z^3[-\ell_1, \ell_2 - \ell_1] \right\}.$$

For our purposes, this will be applied both to x (of length L) and to y (of length N).

Under the separation condition, the relation between autocorrelations of the micrograph and those of X is particularly simple, as we now show. It is useful to introduce some notation: let M denote the number of occurrences of X in y (that is, the number of 1's in s), and let

$$\gamma = \frac{ML}{N} \tag{2.3}$$

denote the density of X in y (that is, the fraction of entries of y occupied by occurrences of X.) In this section, we assume the separation condition (1.2) that imposes $\gamma \leq \frac{L}{2L-1} \approx 1/2$. In Section 3.2 we show that the autocorrelations in the Poisson model are equivalent to those presented in this section.

For shifts in $0, \ldots, L-1$, the autocorrelation functions of y depend on the corresponding autocorrelations of x, the noise level σ and the support signal s. Importantly, under the separation condition (1.2), the dependency on s is only through the density γ. We consider the asymptotic regime where γ remains constant; that is, as N goes to infinity, M also goes to infinity at the same rate (in other words, as we see an increasingly large observation, it contains increasingly many signal occurrences, with constant signal density). In that regime, the law of large numbers can be used to show the following statement:

$$\lim_{N\to\infty} a_y^1 \stackrel{a.s.}{=} \gamma a_X^1, \tag{2.4}$$

[should is be $\mathbb{E}\{a_X^1\}$?] where equality holds almost surely (a.s.), meaning it holds with probability one. The randomness is over the Gaussian noise ε; s may be deterministic. Thus, given enough data, if γ is known, we can estimate a_X^1 from y. We show in Section 3.1 how to estimate γ as well in the homogeneous model.

We have a similar observation for the second-order autocorrelation: $a_y^2[\ell]$ computes the correlation between y and a copy of y shifted by ℓ entries. Considering ℓ only in the range $0, \ldots, L-1$, one can see that any given occurrence of x in y is only ever correlated with itself, and never with another occurrence. As a result,

$$\lim_{N\to\infty} a_y^2[\ell] \stackrel{a.s.}{=} \gamma a_X^2[\ell] + \sigma^2\delta[\ell], \tag{2.5}$$

for $\ell = 0, \ldots, L-1$, where $\delta[\ell]$ equals one for $\ell = 0$ and zero otherwise. The last part captures the autocorrelation of the noise. Notice that, even if σ is unknown, entries $\ell = 1, \ldots, L-1$ still provide useful information about $a_X^2[\ell]$.

Along the same lines, one can establish a relation for third-order autocorrelations:

$$\lim_{N\to\infty} a_y^3[\ell_1, \ell_2] \stackrel{a.s.}{=} \gamma a_x^3[\ell_1, \ell_2] + \sigma^2\gamma a_X^1\big(\delta[\ell_1, 0] + \delta[0, \ell_2] + \delta[\ell_1, \ell_2]\big), \tag{2.6}$$

for $\ell_1, \ell_2 = 0, \ldots, L-1$, where $\delta[\ell_1, \ell_2] := \delta[\ell_1 - \ell_2]$. Here too, few entries are affected by σ in the limit.

3 Theory

3.1 Theory for the homogeneous case

In this section, we derive some theoretical results for the homogeneous case. [Note: I am using terms interchangeably]

A signal is determined uniquely by its second- and third-order autocorrelations. Indeed, assuming $z[0]$ and $z[L-1]$ are nonzero (otherwise, redefine the length of the signal), we can recover z explicitly using this identity for $k = 0, \ldots, L-1$:

$$z[k] = \frac{z[0]z[k]z[L-1]}{z[0]z[L-1]} = \frac{a_z^3[k, L-1]}{a_z^2[L-1]}. \tag{3.1}$$

This proves the following useful fact:

Proposition 3.1. *A signal $z \in \mathbb{R}^L$ is determined uniquely from a_z^2 and a_z^3.*

A couple of remarks are in order. First, (3.1) is not numerically stable: if $z[0]$ or $z[L-1]$ are close to 0, recovery of z is sensitive to errors in the autocorrelations.

In practice, we recover z by fitting it to its autocorrelations using a nonconvex least-squares (LS) procedure, which is empirically more robust to additive noise; we have observed similar phenomena for related problems [14, 16, 1].

The observed moments a_y^1, a_y^2 and a_y^3 of y do not immediately yield the moments of the signal x, as seen by (2.4), (2.5) and (2.6); rather, the two are related by the noise level σ and the ratio γ. We will show, however, that x is still determined by the observed moments of y.

First, we observe that if the noise level σ is known, generally, one can estimate γ from the first two moments of the micrograph. The proof is provided in Appendix B.

Proposition 3.2. *Let $\sigma > 0$ be fixed and assume that the separation condition* (1.2) *holds. If the mean of x is nonzero, then*

$$\gamma \overset{a.s.}{=} \lim_{N\to\infty} \frac{L(a_y^1)^2}{a_y^2[0] + 2\sum_{\ell=1}^{L-1} a_y^2[\ell] - \sigma^2}. \tag{3.2}$$

Using third-order autocorrelation information of y, both the ratio γ and the noise σ are determined. For the following results, when we say that a result holds for a "generic" signal x, we mean that the set of signals which cannot be determined by these measurements has Lebesgue measure zero. In particular, this means that we can recover almost all signals with the given measurements. The proof is provided in Appendix C.

Proposition 3.3. *Assume $L \geq 3$ and assume that the separation condition* (1.2) *holds. In the limit of $N \to \infty$, the observed autocorrelations a_y^1, a_y^2 and a_y^3 determine the ratio γ and noise level σ uniquely for a generic signal x. If $\gamma > \frac{1}{4}$, then this holds for any signal x with nonzero mean.*

From Propositions 3.1 and 3.3 we deduce the following:

Corollary 3.4. *In the limit of $N \to \infty$ and under the separation condition* (1.2)*, the signal x, the ratio γ, and the noise level σ are determined from the first three autocorrelation functions of y if either the signal x is generic or x has nonzero mean and $\gamma > \frac{1}{4}$.*

As a side note, under the separation condition, the length L of the signal can also be determined from the autocorrelations in the asymptotic regime, by inspection of the support of a_y^2.

3.2 Equivalence between the autocorrelations of under the separation and the Poisson process [?]

In this section we omit the affect of the noise as it remains the same for the well-separated model. We aim to show that the moments of

Let us denote by m_Y^q the moments of Y:

$$\begin{aligned} m_X^1[i] &= \mathbb{E}X[i], \quad 0 \le i \le L-1, \\ m_X^2[i,j] &= \mathbb{E}X[i]X[j], \quad 0 \le i,j \le L-1, \\ m_X^3[i,j,k] &= \mathbb{E}X[i]X[j]X[k], \quad 0 \le i,j,k \le L-1. \end{aligned} \tag{3.3}$$

Note that

$$\begin{aligned} a_X^1 &= \frac{1}{L}\sum_{i=0}^{L-1} m_X^1[i], \\ a_X^2[\ell] &= \frac{1}{L}\sum_{i=0}^{L-1} m_X^2[i,i+\ell], \\ a_X^3[\ell_1,\ell_2] &= \frac{1}{L}\sum_{i=0}^{L-1} m_X^3[i,i+\ell_1,i+\ell_2]. \end{aligned} \tag{3.4}$$

In the following, we show that the moments of the observed data under the Poisson process can be written in terms of the autocorrelations of the well-separated model.

Proposition 3.5. *For any i, the moments under the Poisson process are equal to*

$$\begin{aligned} m_y^1[i] &= \gamma L a_X^1 \\ m_y^2[i,i+\ell] &= (\gamma a_X^1)^2 + \gamma a_X^2[\ell], \quad \ell = 0,\dots,L-1, \\ m_y^3[i,i+\ell_1,i+\ell_2] &= (\gamma a_X^1)^3 + \gamma a_X^1 \cdot (\gamma a_X^2[\ell_1] + \gamma a_X^2[\ell_2] + \gamma a_X^2[\ell_2-\ell_1]) + \gamma a_X^3[\ell_1,\ell_2]. \end{aligned} \tag{3.5}$$

Corollary 3.6. *[re-write] From the first three moments of the Poisson process model, one can recover the first three moments from the strongly-separated model.*

3.3 Autocorrelations in higher dimensions

Autocorrelations in d dimensions is defined for $\ell_1,\dots,\ell_{q-1} \in \mathbb{Z}^d$ as

$$a_Z^q[\ell_1,\dots,\ell_{q-1}] = \mathbb{E}\left\{\frac{1}{m^d}\sum_{i\in\mathbb{Z}^d} z[i]z[i+\ell_1]\cdots z[i+\ell_{q-1}]\right\}. \tag{3.6}$$

Interestingly, for dimensions greater than one almost all deterministic (that is, the distribution of Z is a Delta function) signals are determined uniquely from their second-order autocorrelation, up to two symmetries: sign (or phase for complex signals) and reflection through the origin (with conjugation in the complex case) [21]. If the mean of signal is available and non-zero, the sign symmetry can be resolved. However, determining the reflection symmetry still requires additional information, beyond the second-order autocorrelation. The case of 1-D signals is essentially different: generally there are 2^{L-2} signals with the same second-order autocorrelation (after eliminating symmetries) [11, 12].

This uniqueness result for multi-dimensional signals (especially for 2-D images) is the basis of a popular imaging technique called coherent diffraction imaging (CDI). In CDI, an object is illuminated with a coherent wave, and the far field diffraction intensity pattern is measured, corresponding object's Fourier magnitude [37, 41]. If the support of the object is known and less than half of the support of the measured signal, then the data is equivalent to the second-order autocorrelation. The computational problem of recovering the signal is usually referred to as *phase retrieval* or *phase problem*. Albeit the uniqueness result, recently it has been shown that, at least for 2-D images, the problem is ill-conditioned [9]. That is, there exist different images whose second-order autocorrelation agree up to machine precision.

3.4 How many signals can be recovered in the heterogeneous model?

In the heterogeneous models, the autocorrelation of K signals are mixed together. To this end, we count how many equations the first three autocorrelations provide, after omitting symmetries. While we do not provide a rigorous proof, it was shown in similar problems that number of equations captures exactly how many signals can be decomposed for a mix of their autocorrelations [7].

[Probably should be re-written] Clearly, the first-order autocorrelation provides us one equation. For second-order autocorrelations, notice that $a_y^2[\ell]$ suffers no noise-induced bias for ℓ in 1 to $L-1$. Thus if σ is known we obtain L equations, while if not we omit $\ell = 0$. This has the practical effect that we need not know σ to make sense of the computed quantities, while loosing one one equation. Likewise, for third-order autocorrelations, $a_y^3[\ell_1, \ell_2]$ for $0 \leq \ell_1, \ell_2 \leq L-1$ such that $\ell_2 \leq \ell_1$ includes all relevant entries for our purpose (this accounts for symmetries): $\frac{(L+1)(L+2)}{2} - 2$ in total. If we further exclude any such that ℓ_1, ℓ_2 or $\ell_1 - \ell_2$ are zero to avoid the need to estimate σ—there are $\frac{(L-1)(L-2)}{2}$ remaining entries. Hence, if σ is known we have

$$1 + L + \frac{(L+1)(L+2)}{2} - 2 = \frac{1}{2}L(L+5)$$

and if it unknown

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

coefficients in total. Since we aim to estimate KL parameters (for the K signals of length L) plus K parameters (for the densities γ_k), an absolute upper bound on K is

$$K \leq \frac{L(L+5)}{2(L+1)}, \tag{3.7}$$

$$K \leq \frac{L(L-1)+1}{2(L+1)}.$$

Thus, we conjecture that approximately $L/2$ signals can be decomposed from the first three mixed autocorrelations.

4 Algorithms and numerical experiments

The technique we advocate allows recovery of a signal hidden in noisy micrographs without detecting the location of the signals embedded in these micrographs. To illustrate the underlying principles of the method, we present several numerical examples. The code to generate all figures is publicly available in `https://github.com/PrincetonUniversity/BreakingDetectionLimit`.

In the first experiment, we estimated an 50-by-50 pixel image of Einstein with mean zero from a growing number of micrographs, each of size 4096×4096 pixels. Each micrograph contains, on average, 700 occurrences of the target image at random locations. Thus, about 10% of each micrograph contains signal. The micrographs are contaminated with additive white Gaussian noise with standard deviation $\sigma = 3$, corresponding to SNR $= \frac{M\|x\|_F^2}{\sigma^2 N} \approx 1/370$. This high noise level is illustrated in Figure 1. To simplify the experiment, we assume the number of signal occurrences and the noise standard deviation are known. Micrographs are generated such that any two occurrences are always separated by at least 49 pixels in each direction in accordance with the separation condition (1.2).

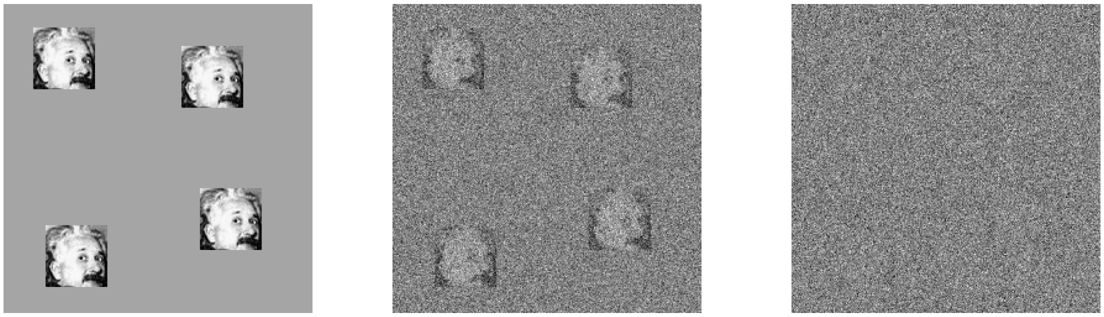
We compute the average second-order autocorrelation of the micrographs. This is a particularly simple computation which can be efficiently executed with a fast Fourier transform (FFT) in parallel. Given the noise level and number of image repetitions, the second-order autocorrelation of the image can be easily deduced from (2.5). Then, to estimate the target image, we resort to a standard phase retrieval algorithm called relaxed-reflect-reflect (RRR) [17], initialized randomly. Relative error is measured as the ratio of the root mean square error to the norm of the ground truth (square root of the sum of squared pixel intensities).

Figure 2 shows several estimated images for a growing number of micrographs. Figure 3 presents the normalized recovery error as a function of the amount of data available. This is computed after fixing the reflection symmetries (see Section 2). As evidenced by these figures, the ground truth image can be estimated increasingly well from increasingly many micrographs, without particle picking.

Appendix **??** provides additional details on the experiments.

In practice, we do not expect to know γ and maybe not even σ

Numerical experiment with three 1-D signals. For the 1-D experiment depicted in Figure **??**, we fix $K = 3$ signals of length $L = 21$. Following the forward model described at the beginning of this section, we generate an observation y of length $12.3 \cdot 10^9$. Each of the three signals appears, respectively (and approximately), $30.0 \cdot 10^6$,



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

Figure 1: Example of micrographs of size 250×250 with additive white Gaussian noise of variance σ^2 for increasing values of σ. Each micrograph contains the same four occurrences of a 50×50 image of Einstein. In panel (c), the noise level is such that it is very challenging to locate the occurrences of the planted image. In fact, it can be shown that at low SNR, reliable detection of individual image occurrences is impossible, even if the true image is known. By analogy to cryo-EM, this depicts a scenario where particle picking cannot be done.

$20.0 \cdot 10^6$ and $10.0 \cdot 10^6$ times in y for a total of exactly $60 \cdot 10^6$ occurrences, such that at least $L-1$ zeros separate any two occurrences of any signals. This is done by randomly selecting $60 \cdot 10^6$ 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 according to the proportions $1/2, 1/3, 1/6$. Then, i.i.d. Gaussian noise with mean zero and standard deviation $\sigma = 3$ is added, to form the observed y. The resulting SNR of y is about $1/9$.

This is enough noise to make cross-correlations of y even with the true signals display peaks at essentially 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 autocorrelations. The first-order autocorrelation is straightforward. For second-order autocorrelations, notice from equation (**??**) that $a_y^2[\ell]$ suffers no noise-induced 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 make sense of the computed quantities. Likewise, for third-order autocorrelations, $a_y^3[\ell_1, \ell_2]$ for $0 \leq \ell_1, \ell_2 \leq L-1$ such that $\ell_2 \leq \ell_1$ includes all relevant entries for our purpose (this accounts for symmetries), and we further exclude any such that ℓ_1, ℓ_2 or $\ell_1 - \ell_2$ are zero to avoid the need to estimate σ—there are $\frac{(L-1)(L-2)}{2}$ remaining entries. We have

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

coefficients in total. Since we aim to estimate KL parameters (for the K signals of length L) plus K parameters (for the densities γ_k), an absolute upper bound on K

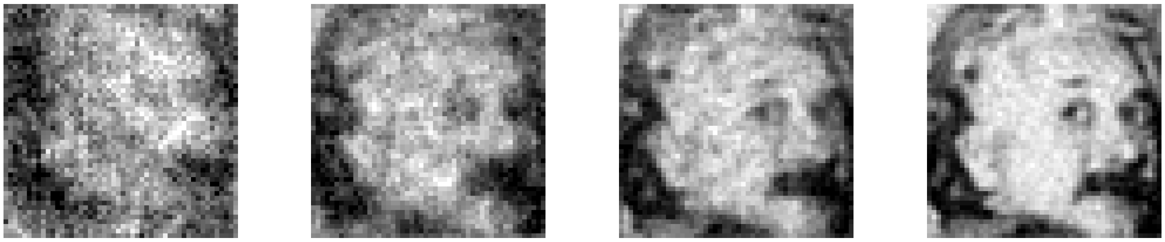


Figure 2: Recovery of Einstein from micrographs at noise level $\sigma = 3$ (see Figure 1(c)). Averaged autocorrelations of the micrographs allow to estimate the power spectrum of the target image. This does not require particle picking. A phase retrieval algorithm (RRR) produces the estimates here shown, initialized randomly. Estimates are obtained from $2 \times 10^2, 2 \times 10^3, 2 \times 10^4, 2 \times 10^5$ micrographs (growing across panels from left to right) of size 4096×4096, each containing 700 image occurrences on average.

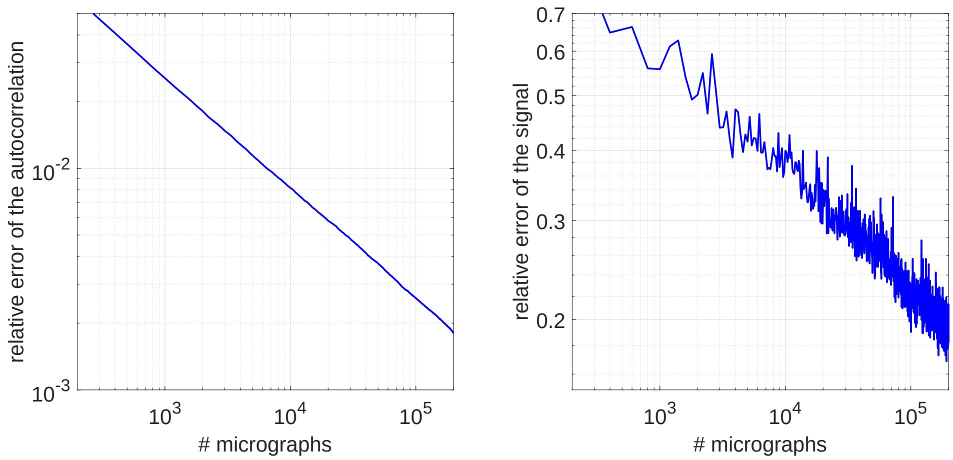


Figure 3: Relative error curves for the experiment of Figure 2.

(simply to ensure we have at least as many equations as we have unknowns) is

$$K(L+1) \leq \frac{1}{2}L(L-1) + 1.$$

Thus, $(L-1)/2$ (up to a small approximation) is an absolute upper limit on K (compare with [16, 7]). [The last paragraph can be removed] In practice, the autocorrelations 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 $x_1, \ldots, x_K$ and coefficients $\gamma_1, \ldots, \gamma_K$ which agree with the data. We choose to do so by running an optimization algorithm on the following nonlinear least-squares problem:

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

where $W\ge L$ is the length of the sought signals and the weights are set to $w_1=1/2, w_2=1/2n_2, w_3=1/2n_3$, where n_2,n_3 are the number of moments used: $n_2=L-1$, $n_3=\frac{(L-1)(L-2)}{2}$ (weights could also be set in accordance with variance estimates as in [16]).

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 few 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 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 [15], which allows to treat the positivity constraints 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.

K-L figure

5 Relation with cryo-EM + experiments?

References

[1] Emmanuel Abbe, Tamir Bendory, William Leeb, João M Pereira, Nir Sharon, and Amit Singer. Multireference alignment is easier with an aperiodic translation distribution. *IEEE Transactions on Information Theory*, 2018.

[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.

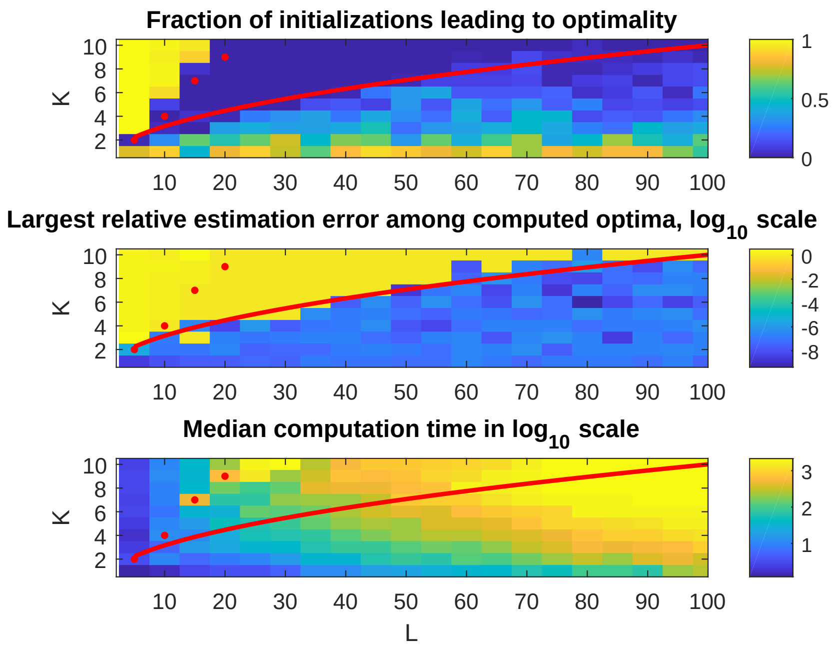


Figure 4: In the $N \to \infty$ regime (access to exact moments, excluding biased entries) and with known uniform densities, it seems K up to $\sqrt{L}$ (red curve) i.i.d. Gaussian signals of length L can be recovered from the known moments. CPU time in seconds. Strictly above red dots, recovery is impossible because the number of unknowns exceeds the number of computed moments. Similar to [16, Fig. 4.1], this experiment suggests a possible statistical-computational gap.

[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 Bandeira, Philippe Rigollet, and Jonathan Weed. Optimal rates of estimation for multi-reference alignment. *arXiv preprint arXiv:1702.08546*, 2017.

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

[8] Afonso S Bandeira, Moses Charikar, Amit Singer, and Andy Zhu. Multireference alignment using semidefinite programming. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 459–470. ACM, 2014.

[9] Alexander Barnett, Charles L Epstein, Leslie Greengard, and Jeremy Magland. Geometry of the phase retrieval problem. *arXiv preprint arXiv:1808.10747*, 2018.

[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, Robert Beinert, and Yonina C Eldar. Fourier phase retrieval: Uniqueness and algorithms. In *Compressed Sensing and its Applications*, pages 55–91. Springer, 2017.

[13] Tamir Bendory, Nicolas Boumal, William Leeb, Eitan Levin, and Amit Singer. Toward single particle reconstruction without particle picking: Breaking the detection limit. *arXiv preprint arXiv:1810.00226*, 2018.

[14] Tamir Bendory, Nicolas Boumal, Chao Ma, Zhizhen Zhao, and Amit Singer. Bispectrum inversion with application to multireference alignment. *IEEE Transactions on Signal Processing*, 66(4):1037–1050, 2017.

[15] 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.

[16] Nicolas Boumal, Tamir Bendory, Roy R Lederman, and Amit Singer. Heterogeneous multireference alignment: A single pass approach. In *Information Sciences and Systems (CISS), 2018 52nd Annual Conference on*, pages 1–6. IEEE, 2018.

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

[18] Joachim Frank. *Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state*. Oxford University Press, 2006.

[19] Robert M Glaeser. Electron crystallography: present excitement, a nod to the past, anticipating the future. *Journal of structural biology*, 128(1):3–14, 1999.

[20] 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*, 54(1):18–36, 2018.

[21] 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.

[22] Richard Henderson. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Quarterly reviews of biophysics*, 28(2):171–193, 1995.

[23] Richard Henderson. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proceedings of the National Academy of Sciences*, 110(45):18037–18041, 2013.

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

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

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

[27] Han-Wen Kuo, Yenson Lau, Yuqian Zhang, and John Wright. Geometry and symmetry in short-and-sparse deconvolution. *arXiv preprint arXiv:1901.00256*, 2019.

[28] Ruslan P Kurta, Jeffrey J Donatelli, Chun Hong Yoon, Peter Berntsen, Johan Bielecki, Benedikt J Daurer, Hasan DeMirci, Petra Fromme, Max Felix Hantke, Filipe RNC Maia, et al. Correlations in scattered X-ray laser pulses reveal nanoscale structural features of viruses. *Physical review letters*, 119(15):158102, 2017.

[29] Kiryung Lee, Yanjun Li, Marius Junge, and Yoram Bresler. Blind recovery of sparse signals from subsampled convolution. *IEEE Transactions on Information Theory*, 63(2):802–821, 2017.

[30] Eitan Levin, Tamir Bendory, Nicolas Boumal, Joe Kileel, and Amit Singer. 3D ab initio modeling in cryo-EM by autocorrelation analysis. In *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*, pages 1569–1573. IEEE, 2018.

[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. *Applied and Computational Harmonic Analysis*, 2018.

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

[35] Haiguang Liu, Billy K Poon, Dilano K Saldin, John CH Spence, and Peter H Zwart. Three-dimensional single-particle imaging using angular correlations from X-ray laser data. *Acta Crystallographica Section A: Foundations of Crystallography*, 69(4):365–373, 2013.

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

[37] Jianwei Miao, Pambos Charalambous, Janos Kirz, and David Sayre. Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. *Nature*, 400(6742):342, 1999.

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

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

[40] Maxim Shatsky, Richard J Hall, Steven E Brenner, and Robert M Glaeser. A method for the alignment of heterogeneous macromolecules from electron microscopy. *Journal of structural biology*, 166(1):67–78, 2009.

[41] Yoav Shechtman, Yonina C Eldar, Oren Cohen, Henry Nicholas Chapman, Jianwei Miao, and Mordechai Segev. Phase retrieval with application to optical imaging: a contemporary overview. *IEEE signal processing magazine*, 32(3):87–109, 2015.

[42] Marin van Heel. Finding trimeric HIV-1 envelope glycoproteins in random noise. *Proceedings of the National Academy of Sciences*, 110(45):E4175–E4177, 2013.

[43] Marin van Heel, Michael Schatz, and Elena Orlova. Correlation functions revisited. *Ultramicroscopy*, 46(1–4):307–316, 1992.

[44] Benjamin von Ardenne, Martin Mechelke, and Helmut Grubmüller. Structure determination from single molecule X-ray scattering with three photons per image. *Nature communications*, 9(1):2375, 2018.

A Derivation of the identities in Section 2

We consider the asymptotic regime $N, M \to \infty$ and assume that $M = \Omega(N)$, so that

$$\gamma := \frac{ML}{N} > 0.$$

Any vectors with indices out of range are given value 0. We focus here on the homogeneous model (1.1). The extension to random signal X is straight-forward by taking the expectation with respect to the distribution of X.

We start by considering the mean of the data:

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

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

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

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

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

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

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

$$\frac{1}{N}\sum_{i=0}^{N-1-\ell}\varepsilon[i]\varepsilon[i+\ell] = \frac{1}{\ell}\sum_{i=0}^{\ell-1}\frac{1}{N/\ell}\sum_{j=0}^{N/\ell-2}\varepsilon[j\ell+i]\varepsilon[(j+1)\ell+i].$$

Each sum $\frac{1}{N/\ell}\sum_{j=0}^{N/\ell-2}\varepsilon[j\ell+i]\varepsilon[(j+1)\ell+i]$ is an average of N/ℓ independent terms with expectation zero, hence converges to zero almost surely as $N \to \infty$. If $\ell = 0$, then we have:

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

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

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

Terms (6), (7) and (8) are linear in ε, and can easily be shown to converge to 0 almost surely by the law of large numbers, by similar arguments as used previously. Term (1) converges to $\gamma a_x^3[\ell_1, \ell_2]$ almost surely, for the same reasons as (A.1). To deal with terms (2)–(5), we must distinguish between different values of ℓ_1 and ℓ_2.

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

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

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

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

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

B Proof of Proposition 3.2

In the limit,

$$(a_y^1)^2 = \frac{\gamma^2}{L^2} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} x[i]x[j].$$

Similarly,

$$\sum_{\ell=1}^{L-1} a_y^2[\ell] = \frac{\gamma}{L} \sum_{\ell=1}^{L-1} \sum_{i=0}^{L-1-\ell} x[i]x[i+\ell],$$

and $a_y^2[0] = \frac{\gamma}{L} \sum_{i=0}^{L-1} x^2[i] + \sigma^2$. The proof is concluded by noting that $a_x^2[-\ell] = a_x^2[\ell]$.

C Proof of Proposition 3.3

We prove that both σ and γ are identifiable from the observed first three moments of y. For convenience, we work with $\beta = \gamma/L$ rather than γ itself. To this end, we construct two quadratic equations satisfied by β and whose coefficients can be computed from observable quantities (in the limit). Then, we show that these equations are independent, and hence that 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. We denote the observed values by E_i or a_y^1, a_y^2, a_y^3. We use F_i to denote functions of the signal's autocorrelations (which are not directly observable).

In the limit $N \to \infty$, almost surely, $a_y^1 = \beta(\mathbf{1}^T x)$ and $a_y^2[0] = \beta\|x\|^2 + \sigma^2$, where $\mathbf{1} \in \mathbb{R}^L$ is the vector of all-ones. (In this whole section, for clarity, we now omit to specify that identities hold almost surely in the limit.) Consider the product:

$$E_1 := a_y^1 a_y^2[0] = (\beta(\mathbf{1}^T x))(\beta\|x\|^2 + \sigma^2) = \sigma^2 a_y^1 + L\beta^2 F_1, \tag{C.1}$$

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 also be estimated from a_y^3, while taking the scaling and bias terms into account. This yields another observable:

$$\begin{aligned} E_2 &:= a_y^3[0,0] + \sum_{j=1}^{L-1}(a_y^3[j,j] + a_y^3[0,j]) \\ &= L\beta F_1 + (2L+1)\sigma^2 a_y^1. \end{aligned} \tag{C.2}$$

Therefore, from (C.1) and (C.2) we get:

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

Let $E_3 := a_y^2[0] + 2\sum_{j=1}^{L-1} a_y^2[j]$; recall from Proposition 3.2:

$$\sigma^2 = E_3 - (a_y^1)^2/\beta. \tag{C.4}$$

Plugging into (C.3) and rearranging, we get a first quadratic equation in β,

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

where

$$\begin{aligned}
\mathcal{A} &= E_2 - (2L+1)a_y^1 E_3,\\
\mathcal{B} &= -E_1 + (2L+1)(a_y^1)^3 + a_y^1 E_3,\\
\mathcal{C} &= -(a_y^1)^3.
\end{aligned}$$

Importantly, these coefficients are observable quantities. As we assume throughout this proof that x has nonzero mean, $a_y^1 \neq 0$ and we conclude that this equation is non-trivial.

Next, we derive the second quadratic equation for β. We notice that

$$E_4 := \frac{1}{L}(a_y^1)^3 = \frac{1}{L}\beta^3(\mathbf{1}^T x)^3 = \beta^3 F_2, \tag{C.6}$$

where $F_2 = \frac{1}{L}(\mathbf{1}^T x)^3$, and we can work out that:

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

Once again, F_2 can be estimated from a_y^3, taking bias and scaling into account:

$$E_5 := a_y^3[0,0] + 3\sum_{j=1}^{L-1}\left(a_y^3[j,j] + a_y^3[0,j]\right) + 6\sum_{1\le i<j\le L-1} a_y^3[i,j] = L\beta F_2 + (6L-3)\sigma^2 a_y^1. \tag{C.7}$$

Consider the following ratio:

$$\frac{E_5}{E_4} = \frac{L}{\beta^2} + \frac{(6L-3)\sigma^2 a_y^1}{E_4}.$$

From the latter, we deduce:

$$\sigma^2 = \frac{E_5}{a_y^1(6L-3)} - \frac{LE_4}{\beta^2 a_y^1(6L-3)}.$$

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

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

where

$$\begin{aligned}
\mathcal{D} &= E_3 - \frac{E_5}{a_y^1(6L-3)}, \\
\mathcal{E} &= -(a_y^1)^2, \\
\mathcal{F} &= \frac{LE_4}{a_y^1(6L-3)}.
\end{aligned}$$

It is also non-trivial since $E_4 \neq 0$.

To complete the proof, we need to show that the two quadratic equations (C.5) and (C.8) are independent. To this end, it is enough to show that the ratios between coefficients differ. From (C.5) and (C.1), we have:

$$\frac{\mathcal{B}}{\mathcal{C}} = \frac{E_1 - (2L+1)(a_y^1)^3 - a_y^1 E_3}{(a_y^1)^3} = \frac{a_y^2[0] - (2L+1)(a_y^1)^2 - E_3}{(a_y^1)^2}.$$

In addition, using (C.6),

$$\frac{\mathcal{E}}{\mathcal{F}} = \frac{(3-6L)(a_y^1)^3}{LE_4} = 3 - 6L.$$

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

$$a_y^2[0] - (2L+1)(a_y^1)^2 - E_3 = (a_y^1)^2(3-6L).$$

Rewriting the identity in terms of x and dividing by β we get:

$$4(L-1)\beta(\mathbf{1}^\top x)^2 - (\mathbf{1}^\top x)^2 + \|x\|^2 = 0. \tag{C.9}$$

For generic x, this polynomial equation is not satisfied so that the quadratic equations are independent. Furthermore, from the inequality $L\|x\|^2 \geq (\mathbf{1}^\top x)^2$ it follows immediately that the equations must be independent so long as

$$\beta > \frac{1}{4L}.$$

D Proof of Proposition 3.5

D.1 First moment

To compute the first moment of y, we will first condition on $M = (M_1, \ldots, M_{N-L+1})$, and then average over M. We have:

$$\mathbb{E}[y[i]|M] = \sum_{j=0}^{L-1} \sum_{k=1}^{M_{i-j}} \mathbb{E}X_k^{i-j}[j] = \sum_{j=0}^{L-1} \sum_{k=1}^{M_{i-j}} m_y^1[j] = \sum_{j=0}^{L-1} M_{i-j} m_y^1[j]. \tag{D.1}$$

Now taking expectations over M we see:

$$\mathbb{E}Y[i] = \gamma \sum_{j=0}^{L-1} m_y^1[j] = \gamma L a_x^1. \tag{D.2}$$

D.2 Second moment

Again, we will condition on M first, and then take the expectation over M. Fix $i_1 \neq i_2$, and let $\Delta = i_2 - i_1$. Then:

$$Y_{i_1} Y_{i_2} = \sum_{j_1=0}^{L-1} \sum_{j_2=0}^{L-1} \sum_{k_1=1}^{M_{i_1-j_1}} \sum_{k_2=1}^{M_{i_2-j_2}} X_{k_1}^{i_1-j_1}[j_1] X_{k_2}^{i_2-j_2}[j_2]. \tag{D.3}$$

We break up the double sum over j_1 and j_2 into two terms: one where $j_2 \neq j_1 + \Delta$, and one where $j_2 = j_1 + \Delta$ or equivalently $i_1 - j_1 = i_2 - j_2$. In the first case, all the terms are independent, and so the expectation factors. In the second case, when $k_1 \neq k_2$ we have independence, but otherwise not. This gives (all expectations are conditional on M):

$$\begin{aligned}
\mathbb{E}Y_{i_1} Y_{i_2} &= \sum_{j_1=0}^{L-1} \sum_{j_2=0}^{L-1} \sum_{k_1=1}^{M_{i_1-j_1}} \sum_{k_2=1}^{M_{i_2-j_2}} \mathbb{E}X_{k_1}^{i_1-j_1}[j_1] X_{k_2}^{i_2-j_2}[j_2] \\
&= \sum_{j_1-j_2\neq\Delta} \sum_{k_1} \sum_{k_2} \mathbb{E}X_{k_1}^{i_1-j_1}[j_1] X_{k_2}^{i_2-j_2}[j_2] \\
&\quad + \sum_{j_1=0}^{L-1} \sum_{k_1\neq k_2} \mathbb{E}X_{k_1}^{i_1-j_1}[j_1] X_{k_2}^{i_1-j_1}[j_1+\Delta] \\
&\quad + \sum_{j_1=0}^{L-1} \sum_{k_1=1}^{M_{i_1-j_1}} \mathbb{E}X_{k_1}^{i_1-j_1}[j_1] X_{k_1}^{i_1-j_1}[j_1+\Delta] \\
&= \sum_{j_1-j_2\neq\Delta} M_{i_1-j_1} M_{i_2-j_2} \mathcal{M}_1[j_1] \mathcal{M}_1[j_2] \\
&\quad + \sum_{j_1=0}^{L-1} M_{i_1-j_1}(M_{i_1-j_1}-1) \mathcal{M}_1[j_1] \mathcal{M}_1[j_1+\Delta] \\
&\quad + \sum_{j_1=0}^{L-1} M_{i_1-j_1} \mathcal{M}_2[j_1, j_1+\Delta].
\end{aligned} \tag{D.4}$$

Now take expectations over the Poisson random variables, using this fact:

Lemma D.1. *If $M \sim Poisson(\gamma)$, then*

$$\mathbb{E}\binom{M}{k} = \frac{\gamma^k}{k!}. \tag{D.5}$$

We get (now the expectation is over M and X):

$$
\begin{aligned}
\mathbb{E}Y_{i_1}Y_{i_2} = &\sum_{j_1-j_2\neq\Delta} \mathbb{E}M_{i_1-j_1}M_{i_2-j_2}\mathcal{M}_1[j_1]\mathcal{M}_1[j_2] \\
&+\sum_{j_1=0}^{L-1}\mathbb{E}M_{i_1-j_1}(M_{i_1-j_1}-1)\mathcal{M}_1[j_1]\mathcal{M}_1[j_1+\Delta] \\
&+\sum_{j_1=0}^{L-1}\mathbb{E}M_{i_1-j_1}\mathcal{M}_2[j_1,j_1+\Delta] \\
=&\sum_{j_1-j_2\neq\Delta}\gamma^2\mathcal{M}_1[j_1]\mathcal{M}_1[j_2]+\sum_{j_1=0}^{L-1}\gamma^2\mathcal{M}_1[j_1]\mathcal{M}_1[j_1+\Delta] \\
&+\sum_{j_1=0}^{L-1}\gamma\mathcal{M}_2[j_1,j_1+\Delta] \\
=&\left(\gamma\sum_{j=0}^{L-1}\mathcal{M}_1[j]\right)^2+\gamma\sum_{j=0}^{L-1}\mathcal{M}_2[j,j+\Delta] \\
=&(\gamma\mathcal{L}_1)^2+\gamma\mathcal{L}_2(\Delta). \qquad \text{(D.6)}
\end{aligned}
$$

But the first term in the sum is just the square of the first moment of Y; so from the first two moments we can recover $\gamma\mathcal{L}_2(\Delta)$, which is just the expected power spectrum of the random vector X, i.e. the usual second moment we have been working with.

D.3 Third moment

For three distinct i_1, i_2 and i_3, we let $\Delta_1 = i_2 - i_1$ and $\Delta_2 = i_3 - i_1$. We have:

$$
\begin{aligned}
&Y_{i_1}Y_{i_2}Y_{i_3} \\
&=\sum_{j_1=0}^{L-1}\sum_{j_2=0}^{L-1}\sum_{j_3=0}^{L-1}\sum_{k_1=1}^{M_{i_1-j_1}}\sum_{k_2=1}^{M_{i_2-j_2}}\sum_{k_3=1}^{M_{i_3-j_3}} X_{k_1}^{i_1-j_1}[j_1]X_{k_2}^{i_2-j_2}[j_2]X_{k_3}^{i_3-j_3}[j_3]. \qquad \text{(D.7)}
\end{aligned}
$$

We will break up the outer three sums into disjoint sums with the following ranges of indices:

1. $j_2 = j_1 + \Delta_1$ and $j_3 = j_2 + \Delta_2 - \Delta_1$.
2. $j_2 = j_1 + \Delta_1$ and $j_3 \neq j_2 + \Delta_2 - \Delta_1$.
3. $j_2 \neq j_1 + \Delta_1$ and $j_3 = j_1 + \Delta_2$.
4. $j_2 \neq j_1 + \Delta_1$ and $j_3 \neq j_1 + \Delta_2$ and $j_3 = j_2 + \Delta_2 - \Delta_1$.
5. $j_2 \neq j_1 + \Delta_1$ and $j_3 \neq j_1 + \Delta_2$ and $j_3 \neq j_2 + \Delta_2 - \Delta_1$.

For Case 1, we have $\ell \equiv i_1 - j_1 = i_2 - j_2 = i_3 - j_3$. We further break up the sum:

$$
\begin{aligned}
&\sum_{j=0}^{L-1} \sum_{k_1=1}^{M_\ell} \sum_{k_2=1}^{M_\ell} \sum_{k_3=1}^{M_\ell} X_{k_1}^{\ell}[j] X_{k_2}^{\ell}[j+\Delta_1] X_{k_3}^{\ell}[j+\Delta_2] \\
&= \underbrace{\sum_{j=0}^{L-1} \sum_{k_i \text{distinct}} X_{k_1}^{\ell}[j] X_{k_2}^{\ell}[j+\Delta_1] X_{k_3}^{\ell}[j+\Delta_2]}_{\text{(a)}} \\
&\quad + \underbrace{\sum_{j=0}^{L-1} \sum_{k_1=k_2\neq k_3} X_{k_1}^{\ell}[j] X_{k_2}^{\ell}[j+\Delta_1] X_{k_3}^{\ell}[j+\Delta_2]}_{\text{(b)}} \\
&\quad + \underbrace{\sum_{j=0}^{L-1} \sum_{k_1=k_3\neq k_2} X_{k_1}^{\ell}[j] X_{k_2}^{\ell}[j+\Delta_1] X_{k_3}^{\ell}[j+\Delta_2]}_{\text{(c)}} \\
&\quad + \underbrace{\sum_{j=0}^{L-1} \sum_{k_2=k_3\neq k_1} X_{k_1}^{\ell}[j] X_{k_2}^{\ell}[j+\Delta_1] X_{k_3}^{\ell}[j+\Delta_2]}_{\text{(d)}} \\
&\quad + \underbrace{\sum_{j=0}^{L-1} \sum_{k_1=k_2=k_3} X_{k_1}^{\ell}[j] X_{k_2}^{\ell}[j+\Delta_1] X_{k_3}^{\ell}[j+\Delta_2]}_{\text{(e)}} .
\end{aligned}
\tag{D.8}
$$

For term (a), the expectation conditional on M is:

$$\sum_{j=0}^{L-1} M_\ell (M_\ell - 1)(M_\ell - 2) \mathcal{M}[j] \mathcal{M}[j+\Delta_1] \mathcal{M}[j+\Delta_2]. \tag{D.9}$$

Using Lemma D.1, the unconditional expectation of (a) is then:

$$\gamma^3 \sum_{j=0}^{L-1} \mathcal{M}_1[j] \mathcal{M}_1[j+\Delta_1] \mathcal{M}_1[j+\Delta_2]. \tag{D.10}$$

For term (b), the expectation conditional on M is:

$$\sum_{j=0}^{L-1} M_\ell (M_\ell - 1) \mathcal{M}_2[j, j+\Delta_1] \mathcal{M}_1[j+\Delta_2] \tag{D.11}$$

and then again using Lemma D.1 we get the expected value:

$$\gamma^2 \sum_{j=0}^{L-1} \mathcal{M}_2[j, j+\Delta_1] \mathcal{M}_1[j+\Delta_2]. \tag{D.12}$$

Similarly, the expected values of terms (c) and (d) are:

$$\gamma^2 \sum_{j=0}^{L-1} \mathcal{M}_2[j, j+\Delta_2]\mathcal{M}_1[j+\Delta_1]. \tag{D.13}$$

and

$$\gamma^2 \sum_{j=0}^{L-1} \mathcal{M}_2[j+\Delta_1, j+\Delta_2]\mathcal{M}_1[j]. \tag{D.14}$$

Finally, the expected value of term (e) is easily shown to be:

$$\gamma \sum_{j=0}^{L-1} \mathcal{M}_3[j, j+\Delta_1, j+\Delta_2]. \tag{D.15}$$

This concludes the computation for Case 1.

Moving onto Case 2, we have $\ell_1 \equiv i_1 - j_1 = i_2 - j_2$, and also define $\ell_2 \equiv i_3 - j_3$. By definition, $\ell_1 \neq \ell_2$. The sum is:

$$\begin{aligned}
&\sum_{j_1=0}^{L-1} \sum_{j_3 \neq j_1+\Delta_2} \sum_{1 \leq k_1, k_2 \leq M_{\ell_1}} \sum_{k_3=1}^{M_{\ell_2}} X_{k_1}^{\ell_1}[j_1] X_{k_2}^{\ell_1}[j_1+\Delta_1] X_{k_3}^{\ell_2}[j_3] \\
=&\sum_{j_1=0}^{L-1} \sum_{j_3 \neq j_1+\Delta_2} \sum_{k_3=1}^{M_{\ell_2}} \Bigg\{ \sum_{1 \leq k_1 \neq k_2 \leq M_{\ell_1}} X_{k_1}^{\ell_1}[j_1] X_{k_2}^{\ell_1}[j_1+\Delta_1] X_{k_3}^{\ell_2}[j_3] \\
&+ \sum_{k_1=1}^{M_{\ell_1}} X_{k_1}^{\ell_1}[j_1] X_{k_1}^{\ell_1}[j_1+\Delta_1] X_{k_3}^{\ell_2}[j_3] \Bigg\}.
\end{aligned} \tag{D.16}$$

Taking expectations conditional on M, we then get:

$$\begin{aligned}
&\sum_{j_1=0}^{L-1} \sum_{j_3 \neq j_1+\Delta_2} \Bigg(M_{\ell_1}(M_{\ell_1}-1) M_{\ell_2} \mathcal{M}_1[j_1]\mathcal{M}_1[j_1+\Delta_1]\mathcal{M}_1[j_3] \\
&+ M_{\ell_1} M_{\ell_2} \mathcal{M}_2[j_1, j_1+\Delta_1]\mathcal{M}_1[j_3] \Bigg).
\end{aligned} \tag{D.17}$$

Taking expectations over M and using Lemma D.1 then gives:

$$\gamma^3 \sum_{j_1=0}^{L-1} \sum_{j_3 \neq j_1+\Delta_2} \mathcal{M}_1[j_1]\mathcal{M}_1[j_1+\Delta_1]\mathcal{M}_1[j_3] \tag{D.18}$$

$$+ \gamma^2 \sum_{j_1=0}^{L-1} \sum_{j_3 \neq j_1+\Delta_2} \mathcal{M}_2[j_1, j_1+\Delta_1]\mathcal{M}_1[j_3]. \tag{D.19}$$

Similarly, Cases 3 and 4 give the expressions:

$$\gamma^3 \sum_{j_1=0}^{L-1} \sum_{j_2 \neq j_1+\Delta_1} \mathcal{M}_1[j_1]\mathcal{M}_1[j_1+\Delta_2]\mathcal{M}_1[j_2] \tag{D.20}$$

$$+ \gamma^2 \sum_{j_1=0}^{L-1} \sum_{j_2 \neq j_1+\Delta_1} \mathcal{M}_2[j_1, j_1+\Delta_2]\mathcal{M}_1[j_2] \tag{D.21}$$

and

$$\gamma^3 \sum_{j_2=0}^{L-1} \sum_{j_1 \neq j_2} \mathcal{M}_1[j_1]\mathcal{M}_1[j_2+\Delta_1]\mathcal{M}_1[j_2+\Delta_2] \tag{D.22}$$

$$+ \gamma^2 \sum_{j_2=0}^{L-1} \sum_{j_1 \neq j_2} \mathcal{M}_2[j_2+\Delta_1, j_2+\Delta_2]\mathcal{M}_1[j_1]. \tag{D.23}$$

Finally, in Case 5 we have $i_1 - j_1$, $i_2 - j_2$, and $i_3 - j_3$ are all pairwise distinct. Consequently, the X variables are always independent, and the expectation conditional on M (letting $\ell_q = i_q - j_q$, $q = 1, 2, 3$),

$$\sum_{j_1, j_2, j_3} M_{\ell_1} M_{\ell_2} M_{\ell_3} \mathcal{M}_1[j_1]\mathcal{M}_1[j_2]\mathcal{M}_1[j_3]; \tag{D.24}$$

since the M_{ℓ_q}'s are pairwise independent, $q = 1, 2, 3$, the expectation over M then yields:

$$\gamma^3 \sum_{j_1, j_2, j_3} \mathcal{M}_1[j_1]\mathcal{M}_1[j_2]\mathcal{M}_1[j_3]. \tag{D.25}$$

Now we add all the terms from Cases 1 to 5. Expressions (D.10), (D.18), (D.20), (D.22), and (D.25) sum to the expression:

$$(\gamma \mathcal{L}_1)^3. \tag{D.26}$$

Note that this is obtained directly from the first moment. Expressions (D.12), (D.13), (D.14), (D.19),(D.21), and (D.23) sum to the expression:

$$\gamma \mathcal{L}_1 \cdot (\gamma \mathcal{L}_2(\Delta_1) + \gamma \mathcal{L}_2(\Delta_2) + \gamma \mathcal{L}_2(\Delta_2 - \Delta_1)). \tag{D.27}$$

Again, note that this is obtained directly from the first two moments. Finally, expression (D.15) is simply:

$$\gamma \mathcal{L}_3(\Delta_1, \Delta_2) \tag{D.28}$$

which is the usual third-order auto-correlation.