Multireference alignment meets blind deconvolution

Tamir Bendory

***Abstract*—Here comes the abstract**

***Index Terms*—Blind deconvolution, multireference alignment, bispectrum, cryo-EM**

I. Introduction

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 [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], just to name a few. In blind deconvolution, we aim at estimating a signal from its convolution with an unknown kernel. Clearly, without additional information, the problem is ill-posed.

In this paper, we consider a particular case, in which one signal is of finite support, while the other is a binary sparse signal. Let us define the (zero-padded) signal

$$x_{\rm zp} = [x, \underbrace{0, 0, \ldots, 0}_{N-L \text{ zeros}}] \in \mathbb{R}^N, \quad N \gg L,$$

and a binary signal $s \in \{0,1\}^N$. Then, our goal is to estimate x from

$$y = x_{\rm zp} * s + \varepsilon, \tag{I.1}$$

where ε is an i.i.d. normal noise ε with mean zero and variance σ^2. As can be seen, the measurement $y \in \mathbb{R}^N$ is composed of repetitions of an underlying signal $x \in \mathbb{R}^L$, located at different positions. So, we can also write the model as

$$y[n] = \sum_{i=1}^{K} x_{\rm zp}[n - n_i] + \varepsilon[n], \tag{I.2}$$

where the n_i's are the non-zero values of s.

While most works on blind deconvolution tries to estimate both unknown signals (see Section II-A), in most cases, the goal is merely to estimate one of the signals. For instance, in image deblurring, both the blurring kernel and the high-resolution image are unknown, but the main goal is only to sharpen the image. In (I.1), the signal s is referred to as the latent or hidden variable of the problem. Our goal is to estimate x and we focus on the low signal–to–noise ratio (SNR) regime. As will be demonstrated, in this case, estimating s is impossible, even if $x_{\rm zp}$ is known. However, we show, maybe surprisingly, that one can estimate x accurately in some cases. We support this statement by theoretical analysis and numerical experiments.

Figure I.1 shows examples of the problem in different noise levels. In the example, the signal x appears twice in y so that $\|s\|^2 = 2$. The first row demonstrates a high SNR regime with $\sigma = 0.1$. As can be seen, in this regime the problem is rather easy. In this case, one can easily the support of s (namely, the repetition of x in y) by simple detection algorithms. This is demonstrated by the correlation function between $x_{\rm zp}$ and y that exhibits two unmistakable peaks in the support of x. Once several copies of the signal were detected and cropped from y, then one can improve the SNR by averaging. We are interest in the more challenging regime of high noise level, as exemplified in the bottom row ($\sigma = 3$) in which the signal is completely swamped in the noise, and one cannot detect the signal's appearances. In this case, one need to use the prior information that the signal appears many times across the measurement.

Our approach for this blind deconvolution problem is based on tools that were developed for the multireference alignment (MRA) problem. MRA is the problem of estimating a signal from its circularly-shifted noisy measurements [14], [15] In Section II-B, we survey recent works in this field. We use a recent suggestion to estimate the signal from feature that are invariant under cyclic-translation [15]. In essence, our method is based on *approximate reduction* of the blind deconvolution problem into the MRA model. This is done be splitting y into shorter windows and treat them as the MRA measurements. If each window contains exactly one signal, then the problem reduces exactly the MRA problem. However, this is not the case. Some windows contain only noise, while others contain both noise and parts of a signal. We analyze the estimation rate of this method and its deviation from the MRA model.

A motivation for this work arises from the cryo–electron microscopy (EM) problem. Cryo–EM received a great deal of attention in structural biology in recent years as it allows visualizing molecules that were not stained in any way, showing them in their native environment. In Section II-C we draw the connections between blind deconvolution and the cryo–EM problem.

The outline of this paper is as follows. In Section II we introduce background on blind deconvolution (Section II-A) and MRA (Section II-B). In Section II-C we discuss the connection with the cryo–EM problem. In Section III we introduce the algorithm and analyze it in Section IV. Section V is devoted for numerical experiments. Section VI concludes the manuscript.

II. background

In this section, we introduce the two main ingredients of this work, blind deconvolution and MRA. Then, we draw connections with the cryo–EM problem.

A. Blind deconvolution

Blind deconvolution is the problem of recovering a signal from its convolution with an unknown kernel h. That is, we aim to recover x from

$$y = x * h, \tag{II.1}$$

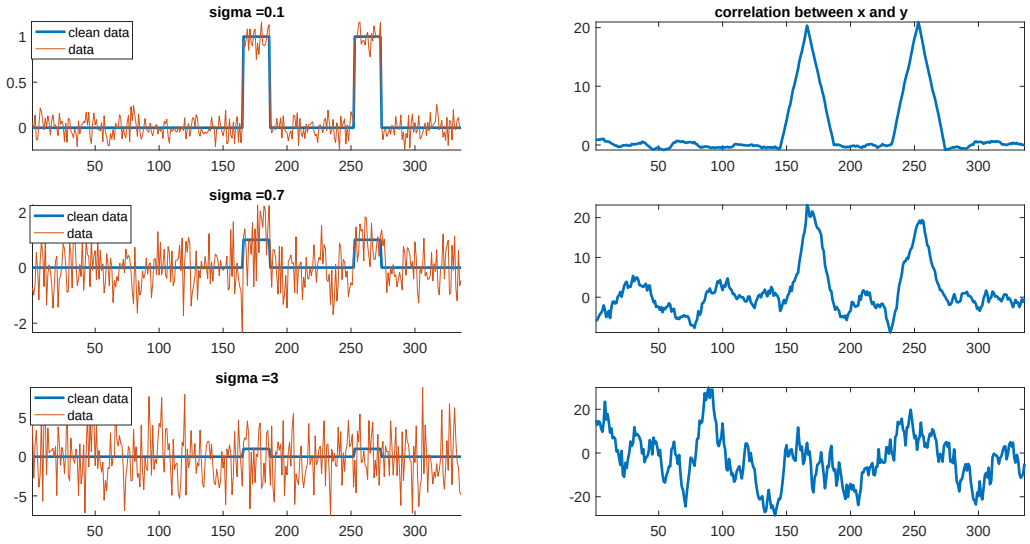

Fig. I.1. The left column presents the clean data and the measurement in different noise levels. In this example, a rectangular signal of length $L = 21$ appears twice in a measurement of length $N = 336$ with different noise levels. Note to the different scale of the y-axis. The right column presents the correlation between the rectangular signal and the measurement for the associated problems.

where both x and h are unknown. In the Fourier domain, it takes the form of

$$Fy = Fx \odot Fh, \tag{II.2}$$

where Fz stands for the Fourier transform of z and $\odot$ for entry-wise product. Clearly, without prior information of the signal, it is impossible to estimate the signal. For instance, even if we assume that the DFT of both signals are non-vanishing, then Fy remains unchanged if we replace $(Fx)[k], (Fh)[k]$ by $a(Fx)[k], \frac{1}{a}(Fh)[k]$ for any $a \in \mathbb{C}^N \backslash \{0\}$ and any k.

The blind deconvolution has been analyzed recently with a various convex and non-convex algorithms. These works assume that the signal lies in a known subspace spanned by vectors with random entries [16], [17], [18]. A somewhat more related line of works assume sparsity in a random linear subspaces [19], [20], [21]. The papers [22], [23], [24], [25] deals with the fundamental question of identifiability of the problems, but none of these paper consider our setup, namely, when the sparse signal is in the canonical basis.

Our problem is very different from all previous works in at least two ways. First, while we do assume that one of the signals is sparse, our goal is to estimate the other signal. We do not claim that we can estimate the sparse signal. Second, we are mainly care of accurate estimation of x in the low SNR regime. In this regime, it seems that estimating the sparse signal s is impossible. This is contrast to previous works that focused on the high SNR regime.

We briefly mention that the deconvolution problem, namely, when the kernel is assumed to be known, has been analyzed throughly in recent years under sparsity constraints [26], [27], [28], [29], [30], [31], [32], [33], [34]. Nevertheless, the problem under consideration is essentially different from two reasons. First, of course, we consider the blind setting when the kernel is unknown. Second, these works considered the recovery of sparse signal, with Gaussian-like kernel. In our problem, the convolution itself kernel is sparse, while we aim to recover a general signal. .

B. Multireference alignment

Our approach for blind deconvolution is based on reducing the problem, approximately, to an MRA problem. In MRA, we aim at estimating a signal $x \in \mathbb{R}^L$ from its circularly shifted noisy measurements

$$y_j = R_{r_j} x + \varepsilon_j, \quad j = 1, \dots, N, \tag{II.3}$$

where R_r translates a signal by r locations, namely, $(R_r x)[i] = x[(i-r) \mod L]$, and ε is an i.i.d. normal noise with mean zero and variance σ^2. This problem finds applications in radar, image processing and structural biology [35], [36], [37]. The algorithmic and statistical characterizations of this problem have been analyzed thoroughly in the last couple of years, see [15], [14], [38], [39], [40], [41], [42].

In this work, we exploit the algorithmic framework proposed in [15]. In order to estimate the signal, this paper suggests to use features of the signal that are invariant under cyclic translation. Apparently, three moments are required. The first is mean of the signal, denoted by μ_x, or, equivalently, it DC component $(Fx)[0]$. The second feature is the power spectrum of the signal $P_x[k] = |(Fx)[k]|^2$ for $k = 0, 1, \dots L-1$. Since in general one cannot estimate a signal from its power spectrum [43], one needs a third order feature. This third order feature, called bispectrum, is defined as [44]

$$B_x[k_1, k_2] = (Fx)[k_1]\overline{(Fx)[k_2]}(Fx)[k_2 - k_1]. \tag{II.4}$$

These features can be estimated directly from the measured

data by

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\mu_{y_j} &\to \mu_x,\\ \frac{1}{N}\sum_{i=1}^{N}P_{y_j} &\to P_x + L\sigma^2,\\ \frac{1}{N}\sum_{i=1}^{N}B_{y_j} &\to B_x + \mu_x\sigma^2L^2A,\end{aligned} \tag{II.5}$$

where A is a known deterministic matrix. Therefore, by proper debiasing, one can get a reliable estimating of these features. For large enough σ, the variance of the estimators is dominated by the bispectrum and thus goes as σ^6/N. This estimation rate is optimal under the assumption that distribution of translations is uniform [38]. If the distribution is non-uniform, then one can do better using only the first two moments, see [41]. Given enough measurements, one can obtain reliable estimation of the features, and then to estimate the underlying signal, up to cyclic translation, using a variety of algorithms [15]. In this work, we use a simple non-convex least-squares estimator.... [Need to see what do we actually solve...]

C. Connections with the cryo–EM problem

This work is partially motivated by the imaging technique called single particle cryo–EM, enabling the visualization of molecules at near atomic resolution [45], [46]. Cryo–EM gained a lot of popularity in recent years in the structural biology community. As a result, Dubochet, Frank and Henderson were recently awarded the the Nobel prize for their contribution in the development of this technique [47].

In cryo–EM, many samples of a molecule are frozen in a thin sheet of ice. An electron beam then passes through the ice and the samples and recorded by an array of detectors. Cryo–EM contains several such images, which are called *micrographs*. Each micrograph contains many two-dimensional (2D) tomographic projections of the samples. The cryo–EM problem is then to estimate the molecules from the micrographs. This process is illustrated by in Figures II.1 and II.2.

Compared to more traditional tomographic methods, like computerized tomography (CT), the cryo–EM problem arises two main challenges. First, the 3D orientation of each sample is unknown since the samples are rapidly frozen in the ice sheet. The experimentalist does not have control over these orientations. Second, the SNR in each projection is very low since the electron dosage must be restricted to limit radiation damage to the molecules. In real cryo–EM data sets, there are some additional challenges, like the contrast transfer function of the microscope that corrupts the acquired image.

The first stage of a cryo–EM reconstruction algorithm is the so called *particle picking* stage. In this stage, the projections are detected in the micrographs, and then cropped to a series of 2D images. Many algorithms were proposed to this detection problem based on support vector machine classifiers, deep learning and template matching and more; see for instance [48], [49], [50], [51]. The most popular techniques are based

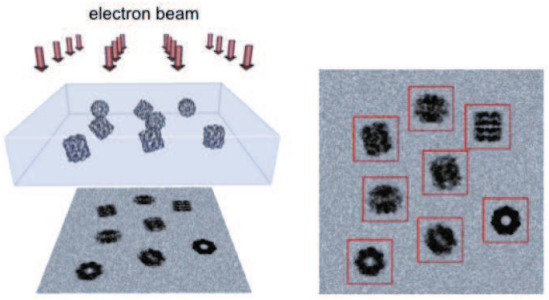

Fig. II.1. A schematic draw of the micrograph generation. Courtesy of ?

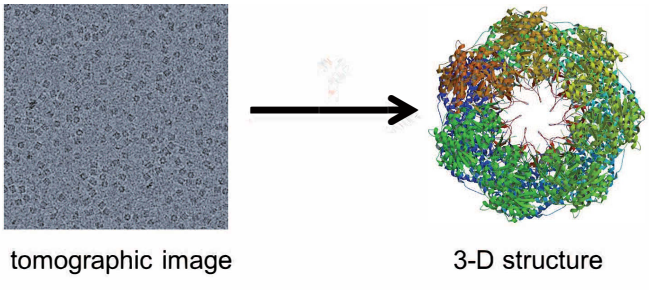

Fig. II.2. The cryo–EM problem: estimating a molecule from the micrograph. Courtesy of ?

on the latter, where the templates can manually or automatically selected. The underlying idea is that the cross-correlation should be high in the presence of a signal in the template. However, this is not true anymore in the presence of high noise level, even if the signal itself is known, as demonstrated in Figure I.1. An illuminating examples are given in [52], [53], where the famous picture of Albert Einstein is used to match pure noise and rotated images of the mathematician Hermann Weyl.

After the particle picking, ideally, one gets a set of measurements of the form

$$y_j = \mathcal{P}(g_j \circ x) + \varepsilon_j, \quad j = 1, \dots, N, \tag{II.6}$$

where x is the three-dimensional (3D) molecule to be estimated, $g_j \in SO(3)$ are unknown 3D rotations that act on the molecule and $\mathcal{P}$ is a linear tomographic projection. This model has been analyzed theoretically from different point–of-views [53], [54]. The MRA model (II.3) can be interpreted as a simplification of the model (II.6), where the unknown cyclic-shifts correspond to the unknown rotations (but there is no analog for the projection operation).

In practice, particle picking algorithms are far from being optimal, mainly because the high noise level prevents reliable detection. One result is that the cropped images are usually not centered, and therefore an unknown 2D translation should be added to the model (II.6). In addition, many samples in the micrograph are non detected or ignored. After the particle picking stage, the molecule is estimated from the 2D images, where the most popular algorithms are based on expectation–maximization (see for instance [55]). Other techniques, such

as the common-lines[56], [57] method or auto-correlation analysis [58], [59], can be used for ab initio modeling.

Our model can be seen as a highly simplified cryo–EM model, where y corresponds to the micrograph and the signal repeated x to a "one-dimensional molecule" (the tomographic projection $\mathcal{P}$ does not appear in the model). We are interest in the question whether we can estimate the signal x, even if the SNR is so low that we cannot detect the locations of x, in y. In the perspective of the cryo–EM problem. the success to estimate the signal in the very low SNR regime may indicate that the molecule can be estimated directly from the cryo–EM micrographs, even if the SNR is so low that particle picking algorithm work poorly.

III. Algorithm

The blind deconvolution algorithm is inspired by the recent MRA approach using translation invariant features proposed in [15]. The algorithm requires setting a parameter $W \in \mathbb{N}$ obeying $W > L$ which is the length of the *analysis window*. We assume henceforth, to ease notation, that W divides N. The algorithm begins with splitting the measurement $y \in \mathbb{R}^N$ into a series of analysis windows $y_i \in \mathbb{R}^W$, $i = 1, \ldots, N/W$, where

$$y_i[n] = y[(i-1)W + n], \quad n = 1, \ldots, W. \tag{III.1}$$

By assumption, there are exactly N/W windows. Of course, one can easily modify the algorithm using overlapping windows, however, our numerical experiments indicate it does not improve the numerical performance. For each window, we then compute its first three translation-invariant features, namely, mean, power spectrum and bispectrum, and then average them over all windows according to (II.5) Following (II.5), we estimate the features of the signal by average over the features of the analysis windows, namely,

$$\begin{aligned}
\hat{\mu}_x &= \frac{W}{N} \sum_{i=1}^{N/W} \mu_{y_i}, \qquad \text{(III.2)} \\
\hat{P}_x &= \frac{W}{N} \sum_{i=1}^{N/W} P_{y_i} - L\sigma^2, \\
\hat{B}_x &= \frac{W}{N} \sum_{i=1}^{N/W} B_{y_i - \hat{\mu}_{x_W}}.
\end{aligned}$$

As discussed in [15], we can replace the mean by more robust estimator, like the median. Finally, once we estimated these features, we look for a signal $\hat{x} \in \mathbb{R}^W$ that is consistent with the data by solving ??

Recall that the current estimate $\hat{x}$ is a signal of length W, whereas the underlying signal of length L. Ideally, it is cyclic-shifted version of x, padded with $W - L$ zeros. To crop the signal correctly, we simply search for the segment of length L with maximal energy. Formally, let $T_\ell : \mathbb{R}^W \to \mathbb{R}^L$ be the operator that takes L consecutive entries of the signal, starting at ℓ. The signal is treated as periodic. Then, by letting

$$\ell_{\max} = \arg\max_{\ell=0,\ldots,W-1} \|T_\ell \hat{x}\|_2, \tag{III.3}$$

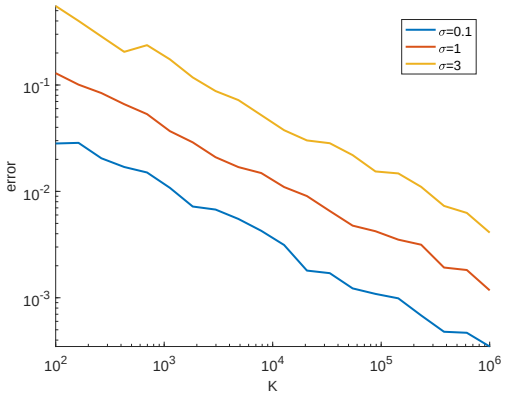

Fig. III.1. The normalized error for norm estimation $\frac{|\|\hat{x}\| - \|x\||}{\|x\|}$, where $\|\hat{x}\|$ is the norm estimation. The error was averaged over 200 experiments for each value of k.

we map

$$\hat{x} \leftarrow T_{\ell_{\max}} \hat{x}. \tag{III.4}$$

The scaling of $\hat{x}$ is incorrect, since the averaging in done over many windows containing pure noise with no information on the signal itself. To rescale the estimated signal, we estimate the norm of the signal separately. The norm estimation can be done as follows. For short, we denote $x_K[n] = \sum_{i=1}^K x_{\text{zp}}[n - n_i]$ and note that $\|x_K\|_2^2 = K\|x\|_2^2$ if a separation of $|n_i - n_j| > L$ is satisfied. Then,

$$\|y\|_2^2 = \|x_K\|_2^2 + \|\varepsilon\|_2^2 + 2\varepsilon^T x_K, \tag{III.5}$$

and therefore

$$\begin{aligned}
\mathbb{E}\left\{\|y\|_2^2\right\} &= \|x_K\|_2^2 + \mathbb{E}\left\{\|\varepsilon\|_2^2\right\} + 2\mathbb{E}\left\{\varepsilon^T x_K\right\} \\
&= K\|x\|_2^2 + N\sigma^2.
\end{aligned} \tag{III.6}$$

As $N \to \infty$, we then obtain

$$\|x\|_2^2 \approx \frac{\|y\|_2^2 - N\sigma^2}{K}. \tag{III.7}$$

The estimation error for different values of σ is depicted in Figure III.1.

The output of the algorithm is a signal of length W, that, ideally, contains x, shifted by some unknown shift, and padded by $W - L$ zeros. The algorithm is summarized in Algorithm 1.

Algorithm 1 Blind devolution by invariant features

Input: The measurement $y \in \mathbb{R}^N$
Output: A signal $\tilde{x} \in \mathbb{R}^W$

1) Split y into a set of N/W analysis windows $y_i \in \mathbb{R}^W$ according to (III.1)
2) Compute the average mean, power spectrum and bispectrum according to (II.5)
3) Estimate the signal $\hat{x} \in \mathbb{R}^W$ by ?
4) Map $\hat{x} \leftarrow T_{\ell_{\max}} \hat{x}$ according to (III.4)
5) Compute the estimated norm x_{norm} by (III.7)
6) Rescale the signal $\tilde{x} \leftarrow \frac{\tilde{x} \cdot x_{\text{norm}}}{\|\tilde{x}\|_2}$

IV. Analysis

In this section, we analyze the performance of Algoritmh 1. Our analysis is based on the asymptotic regime in which $N, \sigma, K \to \infty$ and L, W are fixed. Since we work in high noise regime, We focus on the converges rate of the noise terms. In addition, we analyze the error occurred due to the proposed technique independent of the noise. We assume that positions n_i are well-separated so that $|n_i - n_j| > 2W$. The factor 2 is merely to ease the analysis is not really required by the algorithm. We also assume that each n_i is uniformly distributed within the window.

Our framework is based on splitting the measurement into N/W windows $y_i \in \mathbb{R}^W$. Since in the high noise regime, we cannot detect the appearances of the signal in y, the devision is somewhat arbitrary. The windows y_i, $i = 1, \ldots, N/W$ can be classified into three classes. The first class contains a signal in an unknown location and noise. These are "good" windows since they are perfectly match the MRA model. The second class contains only noise – no signal. As will be demonstrated, these windows cause only a scaling problem which is easily fixed by estimation the signal's norm separately (III.7). The third class of signals contains only part of a signal and the noise. These are the "bad" windows, as they do not contain the entire information of the signal, but, also, do not obey the Gaussian statistics of the noise. These windows induce estimation error, independently of the noise.

The performance of the algorithm relies on the factors. The first is the ratio between the length of the analysis window and the signal, namely,

$$P := W/L > 1.$$

Clearly, the larger this ratio is, the less windows are with the class of "bad" windows; the windows with part of the signal. On the other hand, if we impose large P, we restrict ourselves to work with very sparse signals. The second factor is defined as

$$S = \frac{N}{WK},$$

where WK is a measure for the "active" windows. We refer to S as the "sparsity factor". Clearly, $1 < S < \infty$. The larger S is, the more sparse the signal is.

Let us define the zero padded signal

$$x_W = [x, \underbrace{0, 0, \ldots, 0}_{N-W \text{ zeros}}] \in \mathbb{R}^W.$$

We want to estimate the features of this signal from the features of the y_i's. In Algorithm 1 there is another step of trimming this padded signal to a shorter signal of length L that contains most of the information, but we currently ignore this stage.

We next analyze the estimation error of the invariant features, estimated by (III.2). In the ideal case, when one has n windows with signal and noise, them the asymptotic estimation rate of the mean, power spectrum and bispectrum is σ^2/n, σ^4/n, σ^6/n, respectively. We are interest in how this rate is changed in the model under consideration.

For each feature, the sum in (III.2) can be split into three classed. For instance, the estimated bispectrum can be written as

$$\hat{B}_{x_W} = B_{\text{signal}} + B_{\text{clutter}} + B_{\text{noise}}, \tag{IV.1}$$

where

- B_{noise} - sum of the bispectra of the segments that contain pure noise (no signal),
- B_{signal} - sum of the bispectra of the segments that contain a full signal (one appearance of x) and noise,
- B_{clutter} - sum of the bispectra of the segments that contain only part of x and noise.

The same of course holds for the mean and the power spectrum.

We analyze each class separately. First, we observe that for the noise class, all features converges to zero, namely, $\hat{\mu}_{\text{noise}} \to 0$, $\hat{P}_{\text{noise}} \to 0$ and $\hat{B}_{\text{noise}} \to 0$. Since we have, at least, $N/W - 2K = N/W(1 - 2/S)$ such windows, and the converges is controlled by the bispectrum, the estimation rate is $\mathcal{O}\left(\frac{W}{N}\left(1 - \frac{2}{S}\right)\sigma^6\right)$. This imply that for large S, the estimation rate of this factor reduces since we "inject" pure noise to the estimation, with no relevant information on the signal.

The class group contains windows with a full signal. Indeed, the signal may appear anywhere, but since we work with features that are invariant to cyclic-translation, that exact position is irrelevant as long as we have the full information of the signal. Since we assumed that the positions n_i are distributed uniformly in the window, there are (asymptotically) $K(1 - L/W) = K(1 - 1/P)$ windows in this class. Therefore, B_{signal} converges to a scaled version of B_x at rate $\mathcal{O}\left(\frac{\sigma^6}{K(1-1/P)}\right)$. The scaling will be corrected in the last stage of Algorithm 1. This implies that the larger P, the faster the estimation rate of this group. This makes a lot of sense since it means that will find more windows with full signal and less cropped signals.

Finally, we analyze the third class, the "clutter" — the cropped signals. Since each such signal appears in two windows, we have $2KL/W = 2K/P$ clutter segments. Of course, the noise in this segments goes to zero at rate $\mathcal{O}\left(\frac{P\sigma^6}{2K}\right)$. However, we have an additional error term arises from the cropped signals. This term does not reduce to zero, but can be bounded. To analyze the effect of this term, let us assume that N/σ^6 is great enough so that $B_{\text{noise}} \to 0$ and also $\|B_{y_i}\|_{\text{F}} \approxeq \|B_x\|_{\text{F}}$. Then,

$$\begin{aligned}\left\|\hat{B}_{x_W} - B_{\text{signal}}\right\|_{\text{F}} \approx \|B_{\text{clutter}}\|_{\text{F}} &= \left\|\frac{W}{N}\sum_{i\in\text{clutter}} B_{y_i}\right\|_{\text{F}} \\ &\leq \frac{W}{N}\frac{2KL}{W}\|B_x\|_{\text{F}} = \frac{2}{SP}\|B_x\|_{\text{F}}.\end{aligned} \tag{IV.2}$$

Therefore, for large sparsity term S or large P, we get an accurate estimation of a scaled version of the signal's bispectrum. Similar analysis holds for the mean and power spectrum.

Few insights from the analysis: (need to verified numerically)

- Larger S reduces the clutter error, however, increases the variance of the noise estimation.
- Larger P reduces the clutter error, however, increases the variance of the noise where the signal appears.
- We must keep P large enough to see enough full signals.

V. Numerical experiments

We use the definition of signal–to–noise ratio (SNR) as

$$\mathrm{SNR} = \frac{\|y_c\|_2^2}{\|y_c - y\|_2^2}, \tag{V.1}$$

where y_c is the measurement before adding the noise. The error is measured as

$$\text{error} = \frac{\|\hat{x} - x\|}{\|x\|}. \tag{V.2}$$

Two representative examples are for recovery in low SNR are given in Figure V.1. In these experiments, the signal appears $5 \cdot 10^6$ times in the noise measurement with noise level $\sigma = 2.5$ and $\mathrm{SNR} \approx 1/180$. The figure also presents a segment from the correlation between the underlying signal and the measurement. As can be seen, even if the signal was known, it is impossible to detect its appearances from the data. In other words, even in the end of the task, after we accurately estimated the signal, it is impossible to identify its locations.

An experiment for constant noise level $\sigma = 1$ and varying k appears in Figure V.2. The error stops decreasing at some level from obvious reasons.

VI. Conclusion

Here we conclude the paper

References

[1] S. M. Jefferies and J. C. Christou, "Restoration of astronomical images by iterative blind deconvolution," *The Astrophysical Journal*, vol. 415, p. 862, 1993.

[2] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Transactions on information Theory*, vol. 40, no. 2, pp. 340–349, 1994.

[3] T. F. Chan and C.-K. Wong, "Total variation blind deconvolution," *IEEE transactions on Image Processing*, vol. 7, no. 3, pp. 370–375, 1998.

[4] P. Campisi and K. Egiazarian, *Blind image deconvolution: theory and applications*. CRC press, 2016.

[5] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE signal processing magazine*, vol. 13, no. 3, pp. 43–64, 1996.

[6] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding blind deconvolution algorithms," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2354–2367, 2011.

[7] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Transactions on information theory*, vol. 36, no. 2, pp. 312–321, 1990.

[8] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 1964–1971.

[9] D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 233–240.

[10] G. Ayers and J. C. Dainty, "Iterative blind deconvolution method and its applications," *Optics letters*, vol. 13, no. 7, pp. 547–549, 1988.

[11] T. Michaeli and M. Irani, "Blind deblurring using internal patch recurrence," in *European Conference on Computer Vision*. Springer, 2014, pp. 783–798.

[12] Y. Lin and D. D. Lee, "Relevant deconvolution for acoustic source estimation," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, vol. 5. IEEE, 2005, pp. v–529.

[13] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proceedings of the IEEE*, vol. 85, no. 8, pp. 1310–1322, 1997.

[14] A. S. Bandeira, M. Charikar, A. Singer, and A. Zhu, "Multireference alignment using semidefinite programming," in *Proceedings of the 5th conference on Innovations in theoretical computer science*. ACM, 2014, pp. 459–470.

[15] T. Bendory, N. Boumal, C. Ma, Z. Zhao, and A. Singer, "Bispectrum inversion with application to multireference alignment," *arXiv preprint arXiv:1705.00641*, 2017.

[16] A. Ahmed, B. Recht, and J. Romberg, "Blind deconvolution using convex programming," *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1711–1732, 2014.

[17] X. Li, S. Ling, T. Strohmer, and K. Wei, "Rapid, robust, and reliable blind deconvolution via nonconvex optimization," *arXiv preprint arXiv:1606.04933*, 2016.

[18] S. Ling and T. Strohmer, "Blind deconvolution meets blind demixing: Algorithms and performance bounds," *IEEE Transactions on Information Theory*, 2017.

[19] K. Lee, Y. Li, M. Junge, and Y. Bresler, "Blind recovery of sparse signals from subsampled convolution," *IEEE Transactions on Information Theory*, vol. 63, no. 2, pp. 802–821, 2017.

[20] S. Ling and T. Strohmer, "Self-calibration and biconvex compressive sensing," *Inverse Problems*, vol. 31, no. 11, p. 115002, 2015.

[21] Y. Chi, "Guaranteed blind sparse spikes deconvolution via lifting and convex optimization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 782–794, 2016.

[22] S. Choudhary and U. Mitra, "Sparse blind deconvolution: What cannot be done," in *Information Theory (ISIT), 2014 IEEE International Symposium on*. IEEE, 2014, pp. 3002–3006.

[23] Y. Li, K. Lee, and Y. Bresler, "Identifiability in blind deconvolution with subspace or sparsity constraints," *IEEE Transactions on Information Theory*, vol. 62, no. 7, pp. 4266–4275, 2016.

[24] M. Kech and F. Krahmer, "Optimal injectivity conditions for bilinear inverse problems with applications to identifiability of deconvolution problems," *SIAM Journal on Applied Algebra and Geometry*, vol. 1, no. 1, pp. 20–37, 2017.

[25] Y. Li, K. Lee, and Y. Bresler, "A unified framework for identifiability analysis in bilinear inverse problems with applications to subspace and sparsity models," *arXiv preprint arXiv:1501.06120*, 2015.

[26] T. Bendory, S. Dekel, and A. Feuer, "Robust recovery of stream of pulses using convex optimization," *Journal of Mathematical Analysis and Applications*, vol. 442, no. 2, pp. 511–536, 2016.

[27] T. Bendory, "Robust recovery of positive stream of pulses," *IEEE Transactions on Signal Processing*, vol. 65, no. 8, pp. 2114–2122, 2017.

[28] T. Bendory, A. D. Bar-Zion, D. Adam, S. Dekel, and A. Feuer, "Stable support recovery of stream of pulses with application to ultrasound imaging." *IEEE Trans. Signal Processing*, vol. 64, no. 14, pp. 3750–3759, 2016.

[29] C. Boyer, Y. De Castro, and J. Salmon, "Adapting to unknown noise level in sparse deconvolution," *Information and Inference: A Journal of the IMA*, p. iaw024, 2017.

[30] B. Bernstein and C. Fernandez-Granda, "Deconvolution of point sources: A sampling theorem and robustness guarantees," *arXiv preprint arXiv:1707.00808*, 2017.

[31] Y. De Castro and F. Gamboa, "Exact reconstruction using beurling minimal extrapolation," *Journal of Mathematical Analysis and applications*, vol. 395, no. 1, pp. 336–354, 2012.

[32] J.-M. Azais, Y. De Castro, and F. Gamboa, "Spike detection from inaccurate samplings," *Applied and Computational Harmonic Analysis*, vol. 38, no. 2, pp. 177–195, 2015.

[33] V. Duval and G. Peyré, "Exact support recovery for sparse spikes deconvolution," *Foundations of Computational Mathematics*, vol. 15, no. 5, pp. 1315–1355, 2015.

[34] ——, "Sparse spikes deconvolution on thin grids," *arXiv preprint arXiv:1503.08577*, 2015.

[35] J. P. Zwart, R. van der Heiden, S. Gelsema, and F. Groen, "Fast translation invariant classification of hrr range profiles in a zero phase representation," *IEE Proceedings-Radar, Sonar and Navigation*, vol. 150, no. 6, pp. 411–418, 2003.

[36] H. Foroosh, J. B. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE transactions on image processing*, vol. 11, no. 3, pp. 188–200, 2002.

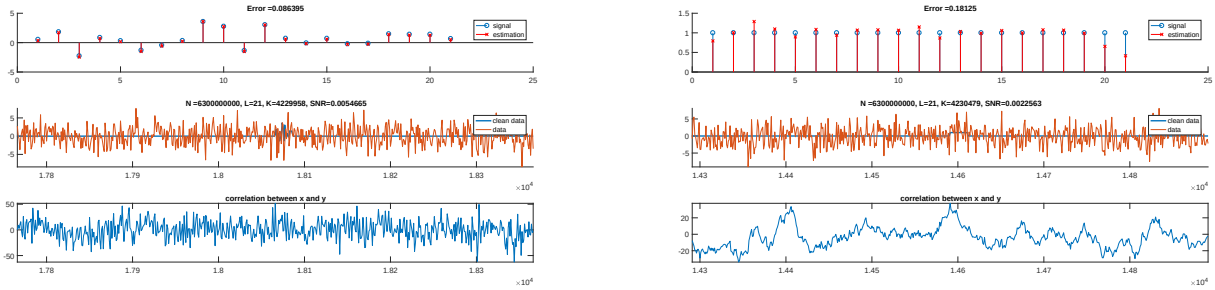

Fig. V.1. Two representative examples of accurate estimation of an i.i.d./ random signal (left) and a rectangular signal (right), both of length $L = 21$, from noisy measurements with $\sigma = 2.5$ and $\mathrm{SNR} \approx 1/180$. The signal appearances $K = 5 \cdot 10^6$ times in the measurements of length $N = 6WK$ for $W = 10L$. The upper figure show the estimated signal versus the original one. Below that, we plot a typical segment of the noisy data. The bottom figures present a segment of the correlation between the underlying signals and the data.

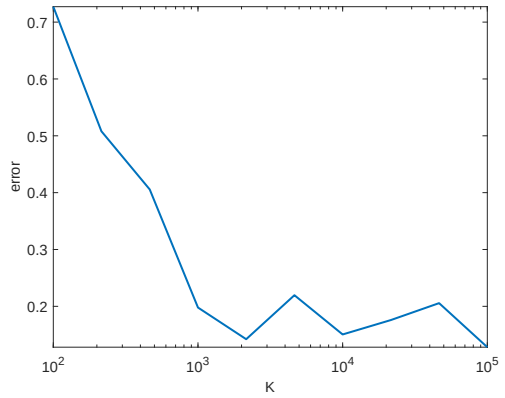

Fig. V.2. The recovery error as a function of k, the number of signal appearances (approximately!) for $\sigma = 1$. We worked with i.i.d. signal of length $L = 1$ (same signal for all experiments), the window size was chosen as $W = 10L$ and the number of measurements was chosen as $N = 6Wk$. In average, $1/\mathrm{SNR} \approx 30$.

[37] R. Diamond, "On the multiple simultaneous superposition of molecular structures by rigid body transformations," *Protein Science*, vol. 1, no. 10, pp. 1279–1287, 1992.

[38] A. Bandeira, P. Rigollet, and J. Weed, "Optimal rates of estimation for multi-reference alignment," *arXiv preprint arXiv:1702.08546*, 2017.

[39] A. Perry, J. Weed, A. Bandeira, P. Rigollet, and A. Singer, "The sample complexity of multi-reference alignment," *arXiv preprint arXiv:1707.00943*, 2017.

[40] N. Boumal, T. Bendory, R. R. Lederman, and A. Singer, "Heterogeneous multireference alignment: a single pass approach," *arXiv preprint arXiv:1710.02590*, 2017.

[41] E. Abbe, T. Bendory, W. Leeb, J. Pereira, N. Sharon, and A. Singer, "Multireference alignment is easier with an aperiodic translation distribution," *arXiv preprint arXiv:1710.02793*, 2017.

[42] E. Abbe, J. Pereira, and A. Singer, "Sample complexity of the boolean multireference alignment problem," *arXiv preprint arXiv:1701.07540*, 2017.

[43] T. Bendory, R. Beinert, and Y. C. Eldar, "Fourier phase retrieval: Uniqueness and algorithms," *arXiv preprint arXiv:1705.09590*, 2017.

[44] J. W. Tukey, "The spectral representation and transformation properties of the higher moments of stationary time series," in *The Collected Works of John W. Tukey*, D. R. Brillinger, Ed. Wadsworth,, 1984, vol. 1, ch. 4, pp. 165–184.

[45] A. Bartesaghi, A. Merk, S. Banerjee, D. Matthies, X. Wu, J. L. Milne, and S. Subramaniam, "2.2 å resolution cryo-em structure of β-galactosidase in complex with a cell-permeant inhibitor," *Science*, vol. 348, no. 6239, pp. 1147–1151, 2015.

[46] D. Sirohi, Z. Chen, L. Sun, T. Klose, T. C. Pierson, M. G. Rossmann, and R. J. Kuhn, "The 3.8 å resolution cryo-EM structure of Zika virus," *Science*, vol. 352, no. 6284, pp. 467–470, 2016.

[47] https://www.nobelprize.org/nobel_prizes/chemistry/laureates/2017/press.html.

[48] J. Frank and T. Wagenknecht, "Automatic selection of molecular images from electron micrographs," *Ultramicroscopy*, vol. 12, no. 3, pp. 169–175, 1983.

[49] S. H. Scheres, "Semi-automated selection of cryo-em particles in relion-1.3," *Journal of structural biology*, vol. 189, no. 2, pp. 114–122, 2015.

[50] P. Arbeláez, B.-G. Han, D. Typke, J. Lim, R. M. Glaeser, and J. Malik, "Experimental evaluation of support vector machine-based and correlation-based approaches to automatic particle selection," *Journal of Structural Biology*, vol. 175, no. 3, pp. 319–328, 2011.

[51] Y. Zhu, Q. Ouyang, and Y. Mao, "A deep learning approach to single-particle recognition in cryo-electron microscopy," *arXiv preprint arXiv:1605.05543*, 2016.

[52] M. Shatsky, R. J. Hall, S. E. Brenner, and R. M. Glaeser, "A method for the alignment of heterogeneous macromolecules from electron microscopy," *Journal of structural biology*, vol. 166, no. 1, pp. 67–78, 2009.

[53] A. S. Bandeira, Y. Chen, and A. Singer, "Non-unique games over compact groups and orientation estimation in cryo-em," *arXiv preprint arXiv:1505.03840*, 2015.

[54] R. Hadani and A. Singer, "Representation theoretic patterns in three dimensional cryo-electron microscopy i: The intrinsic reconstitution algorithm," *Annals of mathematics*, vol. 174, no. 2, p. 1219, 2011.

[55] S. H. Scheres, "RELION: implementation of a bayesian approach to cryo-em structure determination," *Journal of structural biology*, vol. 180, no. 3, pp. 519–530, 2012.

[56] M. Van Heel, "Angular reconstitution: a posteriori assignment of projection directions for 3d reconstruction," *Ultramicroscopy*, vol. 21, no. 2, pp. 111–123, 1987.

[57] A. Singer and Y. Shkolnisky, "Three-dimensional structure determination from common lines in cryo-em by eigenvectors and semidefinite programming," *SIAM journal on imaging sciences*, vol. 4, no. 2, pp. 543–572, 2011.

[58] Z. Kam, "The reconstruction of structure from electron micrographs of randomly oriented particles," *Journal of Theoretical Biology*, vol. 82, no. 1, pp. 15–39, 1980.

[59] E. Levin, T. Bendory, N. Boumal, J. Kileel, and A. Singer, "3d ab initio modeling in cryo-em by autocorrelation analysis," *arXiv preprint arXiv:1710.08076*, 2017.