A note on the blind deconvolution of multiple sparse signals from unknown subspaces

Augustin Cosse*[a]

[a]Courant Institute of Mathematical Sciences and Center for Data Science, NYU, NYC
Département de Mathématiques et Applications, Ecole Normale Supérieure, Paris.

ABSTRACT

This note studies the recovery of multiple sparse signals, $\boldsymbol{x}_n \in \mathbb{R}^L$, $n = 1, \ldots, N$, from the knowledge of their convolution with an unknown point spread function $\boldsymbol{h} \in \mathbb{R}^L$. When the point spread function is known to be nonzero, $|h[k]| > 0$, this blind deconvolution problem can be relaxed into a linear, ill-posed inverse problem in the vector concatenating the unknown inputs $\boldsymbol{x}_n$ together with the inverse of the filter, $\boldsymbol{d} \in \mathbb{R}^L$ where $d[k] := 1/h[k]$. When prior information is given on the input subspaces, the resulting overdetermined linear system can be solved efficiently via least squares (see Ling et al. 2016[1]). When no information is given on those subspaces, and the inputs are only known to be sparse, it still remains possible to recover these inputs along with the filter by considering an additional ℓ_1 penalty. This note certifies exact recovery of both the unknown PSF and unknown sparse inputs, from the knowledge of their convolutions, as soon as the number of inputs N and the dimension of each input, L , satisfy $L \gtrsim N$ and $N \gtrsim T_{\max}^2$, up to log factors. Here $T_{\max} = \max_n\{T_n\}$ and T_n, $n = 1, \ldots, N$ denote the supports of the inputs $\boldsymbol{x}_n$. Our proof system combines the recent results on blind deconvolution via least squares to certify invertibility of the linear map encoding the convolutions, with the construction of a dual certificate following the structure first suggested in Candès et al. 2007.[2] Unlike in these papers, however, it is not possible to rely on the norm $\|(A_T^* A_T)^{-1}\|$ to certify recovery. We instead use a combination of the Schur Complement and Neumann series to compute an expression for the inverse $(A_T^* A_T)^{-1}$. Given this expression, it is possible to show that the poorly scaled blocks in $(A_T^* A_T)^{-1}$ are multiplied by the better scaled ones or vanish in the construction of the certificate. Recovery is certified with high probablility on the choice of the supports and distribution of the signs of each input $\boldsymbol{x}_n$ on the support. The paper follows the line of previous work by Wang et al. 2016[3] where the authors guarantee recovery for subgaussian $\times$ Bernoulli inputs satisfying $\mathbb{E}x_n[k] \in [1/10, 1]$ as soon as $N \gtrsim L$. Examples of applications include seismic imaging with unknown source or marine seismic data deghosting, magnetic resonance autocalibration or multiple channel estimation in communication. Numerical experiments are provided along with a discussion on the sample complexity tightness.

Keywords: Blind deconvolution, ℓ_1-minimization, Compressed sensing, Convex programming, Schur complement.

1. INTRODUCTION

This paper is interested in the recovery of unknown sparse signals $\boldsymbol{x}_n \in \mathbb{R}^L$, $1 \leq n \leq N$ from the result of their convolution with an unknown filter $\boldsymbol{h} \in \mathbb{R}^L$. This particular question is an instance of the more general blind deconvolution problem (see for example Ahmed et al. 2014,[4] 2015,[5] Ling et al. 2015[6] or Li et al. 2016.[7]). When multiple inputs are considered, this general problem reads mathematically as

$$\begin{aligned} \text{find} \quad & \boldsymbol{x}_n, \boldsymbol{h} \in \mathbb{R}^L \\ \text{subject to} \quad & \boldsymbol{y}_n = \boldsymbol{h} * \boldsymbol{x}_n, \quad n = 1, \ldots, N \end{aligned} \tag{1}$$

Problem (1) is inherently hard and ill-posed as it consists in recovering $LN + L$ unknowns from the knowledge of only NL measurements. This problem is also known to suffer from non convexity in the general case. It can in

* acosse@cims.nyu.edu

fact be recast as an instance of a rank one matrix recovery problem which is known to be hard in the general case. In formulation (1), $\boldsymbol{h} \in \mathbb{R}^L$ is a totally unknown non zero point spread function, and $\boldsymbol{x}_n \in \mathbb{R}^L$, $n = 1, \ldots, N$ denote the unknown signals to be recovered. The general question of identifiability (a.k.a uniqueness of the solution) of bilinear inverse problems is discussed in Li et al. 2015.[8] The authors prove identifiability for almost all filters and almost all input matrices $\boldsymbol{X} = [x_1, \ldots, x_N]$ whose columns have joint sparsity s (i.e at most s non zero rows) as soon as $L > s$ and $N > s$. They further show that a necessary condition for identifiability is given by $LN > L + sN - 1$. Note that identifiability here is different from computationally tractable recovery as shown through problems (P3) and (P4) in this last paper which involve ℓ_0 minimization.

One way to make problem (1) better posed and computationally tractable consists in adding assumptions on either the filter, or the inputs and to linearize the problem, or, on a different note, write it as a semidefinite program. Examples of semidefinite programming relaxations can be found in Ahmed et al. 2014,[4] 2015,[5] and 2016[9] as well as Ling et al. 2015.[6] Other approaches include non convex optimization and gradient descent (see Li et al. 2016[7]) as well as power iterations (see Li et al. 2017[10]). When the filter $\boldsymbol{h}$ is known to be non zero, it is possible to write (1) as a linear program by multiplying both sides by the inverse of the filter. The idea of optimizing over the inverse of the filter first appeared in Gribonval et al. 2012[11] as well as Billen et al. 2014.[12] Those papers were then followed by theory in Ling et al. 2016[1] as well as Wang et al. 2016.[3] In the former, the authors suggest to solve problem (1) as a linear problem on the inputs as well as the inverse of the filter. They show that whenever sufficient information is given on the input subspaces and when the inputs are known to live in sufficiently low dimensional subspaces, the linearization of problem (1) reduces to an overdetermined system of equations that can be solved efficiently via least squares. The proof uses concentration to get an estimate on the condition number of the underlying matrix encoding the constraints following from the linearization of (1). Our paper should also be compared to the work of Wang et al. 2016[3] where the authors study the analysis version of formulation (4) below, and work with the inverse of the circulant matrix representing the convolution. In this last paper, recovery is certified with high probability for $N \gtrsim L$ on the genericity of the input matrix $\boldsymbol{X}$ which is assumed to follow a Bernoulli $\times$ subgaussian distribution with expected entries $X_{ij} \in [1/10, 1]$.

In this note, we remove the subspace assumptions and consider inputs that are only known to be sparse sequences of spikes. Let F denote the (normalized) discrete Fourier matrix associated with the DFT,

$$F(\omega, \ell) = \frac{1}{\sqrt{L}} e^{-j2\pi(\omega-1)(\ell-1)/L}, \qquad 1 \leq \omega, \ell \leq L \tag{2}$$

in particular, we thus have,

$$\hat{y}_n[\ell] = \sqrt{L}(\mathcal{F}h)[\ell](\mathcal{F}x_n)[\ell], \quad 1 \leq \ell \leq L,\ 1 \leq n \leq N \tag{3}$$

Where we let $\hat{y} = \mathcal{F}y$ and $\mathcal{F}$ denotes the DFT. We consider the linear relaxation first introduced in Ling et al. 2016[1] to which we add an additional ℓ_1 penalty accounting for the sparsity of the inputs. The convex relaxation for (1) thus read as

$$\begin{aligned} &\text{minimize} && \sum_{n=1}^{N} \|x_n\|_{\ell_1} \\ &\text{subject to} && \operatorname{diag}(\hat{d})\hat{y}_n = \sqrt{L}F x_n, \qquad n = 1, \ldots, N \\ & && \langle w_1, \hat{d}\rangle + \langle w_2, x\rangle = c \end{aligned} \tag{4}$$

Here we let $\hat{d}$ be defined as the inverse of the Fourier transform of the filter, $\hat{d}_i := 1/\hat{h}_i$, and use x to denote the concatenation of all the inputs, $x := (x_1^*, \ldots, x_n^*)^*$. The last constraint is needed to prevent recovery of the trivial solution. Finally c and $w = (w_1, w_2)$ with $w_1 \in \mathbb{C}^L$, $w_2 \in \mathbb{R}^{LN}$, can be taken arbitrary, thus leading to the recovery of the solution $x_0^w := \frac{c}{\langle w, x_0\rangle} x_0 = c_1 x_0$ in the optimal case.

Problem (4) has multiple applications in both engineering and science among which one finds microscopy (see Campisi et al. 2016[13]), seismic imaging (see Liu 2016[14] or Larue 2005[15]), medical imaging, astronomy (see Schultz 1993[16]), communications (see Romberg 2013[17]) and synthetic aperture radar. We briefly discuss two particular applications below.

1.1 Marine seismic data deghosting

In geophysics, the simplest model for a seismic record assumes a stationary source propagating along a given direction. Following this assumption, the recorded wavefield reads as the convolution of the source signature with the sequence of reflection coefficients corresponding to the direction of propagation,

$$y(t) = h(t) * w(t) \tag{5}$$

where $h(t)$ is a weighted sequence of spikes encoding the reflection coefficients and $w(t)$ denotes the source waveform. Gaussian or Ricker sources are common (see for example Bernstein et al. 2017[18]). Most of these models rely on perfect knowledge of the source which is rarely available. Imaging with faulty knowledge on the source is frequent. As a particular example of this, in marine seismic acquisition, imaging is performed at the surface of the water, or underwater. In this framework, the desired trace is corrupted by the sea surface spurious reflections on both the source and receiver sides. The resulting delayed reflections are known as "ghosts" (see for example Parkes et al. 2013[19]). In the presence of ghosts, a reasonable model is given by a serial convolution of the sequence of reflection coefficients with the source wavelet first, followed by the unknown ghost point spread function. Eq. (5) thus turns into

$$y(t) = \left(\int_{t=0}^{L} w(t-\tau)h(\tau)\, d\tau\right) * g(t) = h * w * g \tag{6}$$

In this equation, $h(\tau)$ denotes the unknown sequence of reflection coefficients (sequence of dirac measures), and the source signature $w(t-\tau)$ is corrupted by the unknown ghost reflections. When imaging is performed along multiple directions, it is possible to identify problem (6) to formulation (1) and therefore to use the ℓ_1 minimization program (4) in order to recover the sequences of reflection coefficients along with the corrupted source signature.

1.2 Magnetic Resonance Auto-Calibration

Measurements in Magnetic Resonance Imaging usually suffer from phase distortion due to, among other factors, phase effects from the radio frequency field B_1, flow, or other off-resonance effects. In this framework, the acquired k-space data reads as

$$\hat{y}_n(\boldsymbol{k}) = \int d\boldsymbol{x} |\rho(\boldsymbol{x})| c_n(\boldsymbol{x}) e^{i2\pi \boldsymbol{k}\cdot\boldsymbol{x}}, \qquad \boldsymbol{k} := (k_1, k_2) \in \Omega \tag{7}$$

In this equation $c_n(\boldsymbol{x})$ denotes the coil sensitivity map and $\rho(\boldsymbol{x})$ is the magnetization image. Some of the attempts at removing the phase distortion so far include phase constrained imaging, which only removes part of the phase ambiguity, SENSE and GRAPPA which mitigate the ambiguity by computing correlations in k-space, as well as the more recent ESPIRiT approach by Uecker et al. 2014[20–22] which uses correlation of patches across multiple slices, and rewrites the calibration problem as an eigenproblem in the phases. Due to data inconsistencies or correlation across slices, however, the eigenvector is not always uniquely defined, whence the more recent regularization attempts.[22] All these approaches have yielded very interesting experimental results, yet they lack a proper theoretical framework that would enable to certify when they are working and when they are not. Magnetic Resonance Images, on the other hand are known to be compressible in appropriate transform domains such as wavelets or DCT and $\rho(\boldsymbol{x})$ can thus be identified with the inputs of formulation (4).

1.3 Main Result

The main result of this note certifies recovery of the filter (or equivalently its underlying point spread function) as well as the inputs as soon as the number of inputs N and the dimension of each input L satisfy $\{L, N\} \gtrsim \beta\{N, T_{\max}^2\}$. Here $T_{\max}^2$ denotes the maximum support size among the inputs $\boldsymbol{x}_n$. Recovery holds with high probability on the choice of the supports and distribution of signs over these supports. This result is summarized through the following Theorem,

Theorem 1.1. *Consider the blind deconvolution problem (1). Assume that x_n has support T_n of size $|T_n|$ chosen uniformly at random among $\{1,\ldots,L\}$. Further assume that the signs of $(x_n)_{T_n}$ are distributed according to a Bernoulli +1/-1 random variable with equal (1/2) probability. Finally, assume that*

$$R := \frac{\sup_{n,k} |\hat{x}_n[k]|^2}{\inf_k \sum_{n=1}^N |\hat{x}_n[k]|^2} = \mathcal{O}(\frac{1}{N}), \quad \textit{as well as} \quad |\langle w, z_0\rangle|^2 > 0 \tag{8}$$

where $z_0 := (x_0, \hat{d}_0)$. *Then as soon as* $L \gtrsim N\beta$, *and* $N \gtrsim T_{\max}^2\beta$, *with probability at least* $1-\mathcal{O}(L^{-\beta})$, *problem* (1) *can be solved through the* ℓ_1 *formulation* (4).

Theorem 1.1. shows that despite the fact that the linearization (4) is different from traditional compressed sensing where the matrix A is usually close to an isometry, and does not recover the exact pair (x_0, h) but only a scaled version of this pair, it remains possible to use compressed sensing on this formulation. Randomness in the input signs as well as in the supports is needed to certify invertibility of the linear operator encoding the constraints of (4) and to bound the value of the certificate on T^c. Before proceeding with the proof of Theorem 1.1., we briefly compare our proof to the least squares approach of Ling et al. 2016[1] as well as the proof of Candès et al. 2007.[2]

1.4 Why it works

Proving that the optimality conditions are satisfied for the linearization (4) requires controlling the inverse $(A_T^*A_T)^{-1}$. One way to control this inverse would be to lower bound the minimal eigenvalue of $A_T^*A_T$ and then use the bound $\|(A_T^*A_T)^{-1}\| \leq 1/\lambda_{\min}$. Such an approach would give at least an $\mathcal{O}(1)$ bound on $\|(A_T^*A_T)^{-1}\|$. In the optimality conditions, however, the bound on $(A_T^*A_T)^{-1}$ gets multiplied by $\mathcal{O}(L)$ which cannot be compensated by the $\mathcal{O}(1)$ resulting from the eigenvalue bound. It is however possible to overcome this limitation through the use of the Schur complement by noting that one of the blocks making up the diagonal of $A_T^*A_T$ has eigenvalues $\mathcal{O}(1)$ whether the other has eigenvalues $\mathcal{O}(L)$. Fortunately, the poorly scaled $\mathcal{O}(1)$ block gets multiplied by zero in the certificate construction from Candès et al. 2006,[23] and the remaining $\mathcal{O}(L)$ block ((which becomes $\mathcal{O}(1/L)$ in the inverse) can thus be used to compensate the $\mathcal{O}(L)$ appearing in the restriction of the certificate to the complement T^c.

2. PROOF OF THEOREM 1.1.

A now popular approach at certifying recovery of a vector $c_1(x_0, d_0) = (x_0^w, d_0^w)$ through convex programs such as (4) relies on the construction of a dual vector $Y(t)$ satisfying the optimality conditions for problem (4). This idea is summarized in the proposition below. Proposition 1 is by now more than standard in the fields of convex optimization and compressed sensing. We direct readers that are interested in a proof to, for example Candès et al. 2006.[23] The proposition below is a slightly modified version of the conditions in this paper.

Proposition 1 (Dual certificate). *Consider problem* (4). *Let* Γ *denote the support of* d_0 *in the vector* $z_0 = (x_0, d_0)$ *and let* T *denote the union* $T = \cup_n T_n \cup \Gamma$ *where each* T_n *denote the support of the corresponding input* x_n, $x \in \mathbb{R}^{LN}$, $h \in \mathbb{R}^L$. *Further let* $w = (w_1, w_2, \ldots, w_N, w_{N+1})$ *where* $w_n \in \mathbb{R}^L$. *Problem* (4) *has a unique solution as soon as there exists trigonometric polynomials* q_n, *defined from appropriate coefficients* $\lambda_{k,n}$,

$$q_n[\ell] = \sum_{k=1}^L \frac{1}{\sqrt{L}} \lambda_{k,n} e^{-j2\pi(k-1)(\ell-1)/L} \qquad 1 \leq \ell \leq L,\ 1 \leq n \leq N \tag{9}$$

as well as a constant $\alpha \in \mathbb{C}$ *such that*

$$\begin{cases} q_n(\ell) + \alpha w_n[\ell] = \mathrm{sign}(c_1 x_0) & \ell \in T,\ n = 1, \ldots, N \\ |q_n(\ell) + \alpha w_n[\ell]| < 1 & \ell \in T^c,\ n = 1, \ldots, N \end{cases} \tag{10}$$

Together with $\sum_{n=1}^N \lambda_{k,n} y_n(k) + w_{N+1}[k]\alpha = 0$ *for all* $1 \leq k \leq L$.

We now show how to find the coefficients $\lambda_{n,k}$ together with the constant α from proposition 1. The constraints from the linear program (4) can be encoded through the linear map A defined as

$$A := \begin{bmatrix} \sqrt{L}F & \ldots & 0 & -\mathrm{diag}(y_1) \\ 0 & \ddots & \vdots & \vdots \\ 0 & \ldots & \sqrt{L}F & -\mathrm{diag}(y_N) \\ w_1 & \ldots & w_N & w_{N+1} \end{bmatrix} \tag{11}$$

Given A, it can be seen that the conditions of proposition 1 are equivalent to exhibiting a vector $Y = [\tilde{q}_1, \ldots, \tilde{q}_N, 0_L]$ in the range of A^* such that $Y(t) = \mathrm{sign}(c_1 x_0)$ on $T_1 \cup \ldots \cup T_N$, and $|Y(t)| < 1$ on T^c. Let A_T denote the linear map obtained by retaining from A the columns corresponding to $T = \cup_i T_i \cup \Gamma$. In other words, we have

$$A_T := \begin{bmatrix} \sqrt{L}F_1 & \ldots & 0 & -\mathrm{diag}(y_1) \\ 0 & \ddots & \vdots & \vdots \\ 0 & \ldots & \sqrt{L}F_N & -\mathrm{diag}(y_N) \\ (w_1)_{T_1} & \ldots & (w_N)_{T_N} & w_{N+1} \end{bmatrix} \tag{12}$$

Where we let $F_n := (F)_{T_n}$. The restriction of A on the complement, A_{T^c} can be defined in the exact same way. Following those definitions, we consider the candidate certificate,

$$Y = A^* A_T (A_T^* A_T)^{-1} \boldsymbol{z}_0 \tag{13}$$

where $\boldsymbol{z}_0 = (\boldsymbol{s}_0, \mathbf{0}_L)$ and $\boldsymbol{s}_0$ is defined as,

$$\boldsymbol{s}_0 = \begin{cases} \mathrm{sign}(c_1 \boldsymbol{x}_0) & \text{on } T_1 \cup \ldots \cup T_N \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

Note that as soon as the matrix $(A_T^* A_T)$ is invertible, we immediately have $Y(t) = \mathrm{sign}(c_1 x_0)$ for all $t \in T_1 \cup \ldots \cup T_N$ as well as $Y(t) = 0$ on Γ. We are thus left with showing that $|Y(t)| < 1$ on T^c.

We start by showing invertibility of $(A_T^* A_T)$. To do this, we follow the approach of Ling et al. 2016.[1] We write the matrix $A_T^* A_T$ as $A_T^* A_T = P^* Z^* Q^* Q Z P$ for appropriate matrices P, Q and Z. We first derive a bound on $\lambda_{\min}(\mathbb{E} Z^* Z)$. As soon as we know a bound on the minimal eigenvalue of $\mathbb{E} Z^* Z$, usual concentration bounds can be used to derive a corresponding bound on the minimal eigenvalue of $A_T^* A_T$ with high probability. This is the point of the following lemma,

Lemma 2.1 (Invertibility of $A_T^* A_T$). *Consider the linear map A_T defined in (12) in the framework of problem (4). Assume that the supports $T_1, \ldots T_N$ are chosen uniformly at random among $\{1, \ldots, L\}$, independently for each $1 \leq n \leq N$. Further assume that the Fourier matrices F_n in (11) are replaced by random modulated matrices $\tilde{F}_n = F_n D_n$ where $D_n = diag(\sigma_{k,n})$ and $\sigma_{k,n}$ are independent Bernoulli random variables taking the value $+1$ or -1 with equal probability. Then as soon as $L \gtrsim \beta N$ and $N \gtrsim \beta T_{\max}^2$, with probability at least $1 - (NT_{\max} + L)^{-\beta} - \mathcal{O}((LN)^{\beta})$,*

$$\lambda_{\min}(A_T^* A_T) > 0 \tag{15}$$

Given this first lemma, the bound on $|Y(t)|$, $t \in T^c$ can be derived by using the following proposition.

Proposition 2 (Complement). *Consider the linearization (4) in the framework of problem (1). Let A_{T^c} be defined just as A_T by retaining only the columns belonging to T^c. Assume that $A_T^* A_T$ is invertible. Assume that the supports of $T_1, \ldots T_N$ are chosen uniformly at random among $\{1, \ldots, L\}$, independently for each $1 \leq n \leq N$. Further assume that the Fourier matrices F_n in (11) are replaced by random modulated matrices $\tilde{F}_n = F_n D_n$ where $D_n = diag(\sigma_{k,n})$ and $\sigma_{k,n}$ are independent Bernoulli random variables taking the values $+1$ or -1 with equal probability. Then with probability at least $1 - \mathcal{O}(L^{-\beta})$, as soon as $L \gtrsim \beta N$, $N \gtrsim \beta T_{\max^2}$ we have*

$$|Y(t)| < 1 \qquad \forall t \in T^c \tag{16}$$

It is not possible to directly make use of the bound on $\lambda_{\min}(A_T^* A_T)$ to derive the norm of $(A_T^* A_T)^{-1}$ as is suggested in Candès et al. 2007[2] because the lower bound is too small in this case and would lead to a $\mathcal{O}(\sqrt{L/N})$ bound on $|Y(t)|$, $t \in T^c$. Instead of this, the proof of proposition 2 relies on the decomposition provided by the Schur complement, as well as on the Neumann series in order to bound the norm of the inverse whenever $\{L, N\} \gtrsim \beta T_{\max}^2$.

The proof of Theorem 1.1 follows as a consequence of proposition 2 as well as from the fact that as soon as $A_T^* A_T$ is invertible, the first and last conditions in proposition 1 are automatically satisfied. The remaining part of this section proceeds with the proofs of lemma 2.1 and proposition 2.

2.1 Proof of lemma 2.1

In this section, we show invertibility of $A_T^* A_T$ by deriving a lower bound on the eigenvalues of this matrix. Following Ling et al. 2016,[1] we consider the decomposition

$$A_T^* A_T = P^* Z^* Q^* Q Z P \tag{17}$$

with Q, Z and P defined as

$$Q = \begin{bmatrix} \|x_1\| \boldsymbol{I}\sqrt{L} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \|x_N\| \boldsymbol{I}\sqrt{L} \end{bmatrix}, \qquad Z = \begin{bmatrix} \frac{\sqrt{L}}{\sqrt{N}} \mathrm{diag}(F_1 v_1) & -F_1 & \cdots & \mathbf{0} \\ \frac{\sqrt{L}}{\sqrt{N}} \mathrm{diag}(F_2 v_2) & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sqrt{L}}{\sqrt{N}} \mathrm{diag}(F_N v_N) & \mathbf{0} & \cdots & -F_N \end{bmatrix} \tag{18}$$

$$P = \begin{bmatrix} \frac{\sqrt{N}}{\sqrt{L}} \mathrm{diag}(\boldsymbol{h}) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \frac{1}{\|x_1\|} \boldsymbol{I}_L & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \frac{1}{\|x_N\|} \boldsymbol{I}_L \end{bmatrix} \tag{19}$$

where we let $v_n = x_n / \|x_n\|$, $n = 1, \ldots, N$ and $v := (v_1, \ldots, v_N)$. Using this decomposition, we have the following relations for each of the blocks in $\mathbb{E} Z^* Z$. First, the k^{th} element on the diagonal of the $L/N \sum_{n=1}^N (\mathrm{diag}(F_{n,T} v_n))^2$ block reads as

$$\mathbb{E} \sum_{n=1}^N L/N \, (\mathrm{diag}(F_n v_n))^2 = \mathbb{E} \frac{L}{N} \sum_{n=1}^N \left| \sum_{\ell=1}^L \langle f_k^*[\ell], v_n[\ell] \rangle \right|^2 \tag{20}$$

$$= \mathbb{E} \frac{L}{N} \sum_{n=1}^N \sum_{k=1}^L (v_n^* f_k)(f_k^* v_n) = 1. \tag{21}$$

Using the random modulation of the columns. Accordingly, the k^{th} row in $\mathbb{E} \frac{\sqrt{L}}{\sqrt{N}} \mathrm{diag}(F_n v_n)^* F_{n,T}$ thus reads as

$$-\mathbb{E} \frac{\sqrt{L}}{\sqrt{N}} \mathrm{diag}(F_n v_n)^* F_{n,T} = -\mathbb{E} \frac{\sqrt{L}}{\sqrt{N}} (v_n^* \tilde{f}_k) \tilde{f}_k^* \tag{22}$$

$$= -\frac{1}{\sqrt{NL}} v_n \mathbf{1}_L^* \tag{23}$$

Where we use $\tilde{f}_k$ to denote the k^{th} modulated Fourier row and $f_k[\ell]f_k[\ell]^* = 1/L$. The expected matrix $\mathbb{E}Z^*Z$ reads as

$$\mathbb{E}Z^*Z = \begin{bmatrix} \boldsymbol{I}_L & -\frac{1}{\sqrt{NL}}1_{|T_1|}v_1^* & -\frac{1}{\sqrt{NL}}1_{|T_2|}v_2^* & \dots & -\frac{1}{\sqrt{NL}}\mathbf{1}_{|T_N|}v_N^* \\ -\frac{1}{\sqrt{NL}}v_1 1_{|T_1|}^* & \boldsymbol{I}_{|T_1|} & 0 & \dots & 0 \\ -\frac{1}{\sqrt{NL}}v_2 1_{|T_2|} & 0 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ -\frac{1}{\sqrt{NL}}v_N \mathbf{1}_{|T_N|}^* & 0 & \dots & 0 & \boldsymbol{I}_{|T_n|} \end{bmatrix} \tag{24}$$

The eigenvalues of (24) are given by 0, 1 and 2. 0 has multiplicity 1. The bound on $\lambda_{\min}$ can then be derived by using the following proposition which is a straightforward extension to proposition 5.4 in Ling et al. 2016.[1]

Proposition 3. *Consider the matrix Z defined in (18). Then as soon as $L \gtrsim \beta T_{\max}$, $N \gtrsim \beta$ with probability at least $1-(T_{\max}N + L)^{-\beta} - (LN)^{-\beta+1} = 1 - \mathcal{O}(L^{-\beta})$,*

$$\|Z^*Z - \mathbb{E}(Z^*Z)\| \leq \frac{1}{2} \tag{25}$$

Since the eigenvalues of $\mathbb{E}Z^*Z$ are 0, 1 and 2, only the smallest eigenvalue of Z^*Z can vanish. Fortunately this eigenvalue can be compensated by the weight vector w. Following Proposition 3, with probability at least $1-\mathcal{O}(L^{-\beta})$, we have

$$A_T^*A_T \succeq \lambda_{\min}(Q^*Q)P^*Z^*ZP + ww^* \succeq \min_n L\|x_n\|^2 P^*Z^*ZP + ww^* \tag{26}$$

define the vector $u_1 := \frac{1}{\sqrt{2}}\left[\frac{1}{\sqrt{L}}\mathbf{1}_L, \frac{1}{\sqrt{N}}\boldsymbol{v}\right]$, note that $\|u_1\| = 1$ then we have

$$|u_1^*(P^{-1})^*w)|^2 = c_1^2 \frac{|\langle w_2, d_0\rangle|^2 + |\langle w_1, x_0\rangle|^2}{N} \tag{27}$$

using (27) and following the approach in Ling et al. 2016,[1] we get the following bound on $\lambda_{\min}$,

$$\lambda_{\min}(x_{\min}^2 P^*Z^*ZP + ww^*) \geq \min_n \left\{ L\|x_n\|^2, \frac{|\langle w_2, d_0\rangle|^2 + |\langle w_1, x_0\rangle|^2}{N} \right\} \min\left\{ \frac{N}{L}h_{\min}^2, \frac{1}{\|x_n\|_\infty^2} \right\} \tag{28}$$

Which finally gives

$$\lambda_{\min}(x_{\min}^2 P^*Z^*ZP + ww^*) > 0 \tag{29}$$

as soon as $\min\left\{h_{\min}^2, |\langle w, z_0\rangle|^2, \|x_n\|^2\right\} > 0$. We can thus assume that $(A_T^*A_T)^{-1}$ is well defined. We now proceed with the proof of Proposition 2.

2.2 Proof of Proposition 2

Let $t_0 \in T^c$, recall that $Y = A^*A_T(A_T^*A_T)^{-1}s_0$. We want to show that $|Y(t)| < 1$ for all $t \in T^c$. Further let $v_0 = e_{t_0}^*(A^*A_T)$ for $t_0 \in T^c$. The vector e_{t_0} is the canonical basis vector with t_0^{th} entry set to 1.

$$Y(t_0) = \langle v_0, (A_T^*A_T)^{-1}s_0\rangle \tag{30}$$

We first prove the following lemma which bounds the operator norm of v_0.

LEMMA 2.2. *Consider the linear map A in the framework of problem (4). Let A_T be defined as in (12) and let v_0 be defined as above. We have*

$$\|v_0\| \lesssim \beta \sup_n T\|h\|_\infty^2\|x_n\|_\infty\sqrt{L} \tag{31}$$

with probability at least $1 - \mathcal{O}(L^{-\beta})$ up to log factors.

Proof. The product $(A^*A_T)_{T^c}$ can be expanded as

$$(A^*A_T) = L\left[\begin{array}{ccc|c} F_{T_1^c}^*F_1 & \dots & 0 & F_{T_1^c}^*\mathrm{diag}(y_1)\\ \vdots & \ddots & \vdots & \vdots\\ 0 & \dots & F_{T_N^c}^*F_N & F_{T_N^c}^*\mathrm{diag}(y_n)\\ \hline 0 & \dots & 0 & 0\end{array}\right] + \boldsymbol{w}\boldsymbol{w}_T^* \tag{32}$$

Noting that $\Gamma \in T$. Let f_k denote the k^{th} column of F^* and $\tilde{f}_k$ denote the corresponding signed vector $\tilde{f}_k = \sigma_{n,k}f_k$. For any $t_0 \in T^c$, we have

$$v_0 = \left(\sqrt{L}\sum_{k=1}^L \sigma_{t_0}f_k[t_0]y_n[k]e_k^* + L\sum_{k=1}^L f_k[t_0]\sum_{t\in T}\sigma_{t_0}\sigma_t f_k^*[t]e_t^*\right) + w[t_0]w_T^* = (v_0)_1 + w[t_0]w_T^*,\ \ t\in T_n,\quad t_0 \in T_n^c \tag{33}$$

The second term vanishes by orthogonality of the rows t, t' for $t \neq t'$ of the DFT matrix without the need to make use of randomness. For the first term, using the definition of $y_n[k]$, we have

$$\mathbb{E}(v_0)_1 = L\mathbb{E}\sum_{k=1}^L f_k[t_0]\sigma_{t_0}e_k^*h[k]\left(\sum_{\ell\in T}\sigma_\ell(f_k^*[\ell]x_n[\ell])\right) = 0 \tag{34}$$

Where we use $\mathbb{E}\sigma_\ell = 0$ (or independence between σ_{t_0} and σ_ℓ). To compute the variance, let Y_ℓ be defined as

$$Y_\ell := L\sigma_{t_0}f_k[t_0]h[k]\sigma_\ell(f_k^*[\ell]x_n[\ell]),\quad \ell \in T \tag{35}$$

And let $Z_0 = \sum_{\ell=1}^L Y_\ell$. From this, one can write

$$\sigma^2 := \sum_T \mathbb{E}Y_\ell^2 \leq TL^2\frac{1}{L}\|h\|_\infty^2\frac{1}{L}\|x_n\|_\infty^2 \leq \sup_n T\|h\|_\infty^2\|x_n\|_\infty^2 \tag{36}$$

Now we can apply the scalar Bernstein inequality (see for example section 2.2.2, lemma 2.2.9 in van der Vaart et al. 1996[24]). Let $M := \max |Y_\ell|$ We get,

$$|Z_0 - \mathbb{E}Z_0| \lesssim \max\left\{tM, \sigma^2 t\right\} \tag{37}$$

with probability at least $1 - e^{-t}$. Setting $t = \beta \log(L \vee N)$

$$\begin{aligned} |Z_0 - \mathbb{E}Z_0| &\lesssim \max\left\{\sqrt{T}\|x_n\|_\infty \|h\|_\infty \beta \log(L \vee N), \sqrt{T}\|x_n\|_\infty \|h\|_\infty \sqrt{\beta \log(L \vee N)}\right\} && (38)\\ &\lesssim \|h\|_\infty \sqrt{T}\|x_n\|_\infty \beta \log(L \vee N) && (39)\end{aligned}$$

with probability at least $1 - (LN)^{-\beta}$. Finally, considering the normalization $\|w\| = \mathcal{O}(1)$

$$\begin{aligned} \|v_0\| = \sup_{\|g\|\leq 1} \langle v_0, g\rangle &\leq \sqrt{L}\|g\|\|(v_0)_1\|_\infty + \mathcal{O}\left(\frac{\sqrt{NT}}{\sqrt{L+NT}}\right) && (40)\\ &\lesssim \sqrt{LT}\|h\|_\infty \|x_n\|_\infty + \mathcal{O}\left(\frac{\sqrt{NT}}{\sqrt{L+NT}}\right) && (41)\\ &\leq \sup_{n,\ell} L\sqrt{T}\|h\|_\infty |\hat{x}_n[\ell]| + \mathcal{O}\left(\frac{\sqrt{NT}}{\sqrt{L+NT}}\right) && (42)\end{aligned}$$

□

The second term in (40) is the contribution from the weight vector $w[t_0]w_T^*$ in (33). Finally in (42) we replace $\sqrt{T}\sup_n \|x_n\|$ with $\sqrt{T}\sup_n \sqrt{L}|\hat{x}_n[\ell]|$ using Parseval.

We now show how one can compensate for the L appearing in the leading term of (42) by means of a tighter bound on $(A_T^* A_T)^{-1}$ given by the Schur complement and the Neumann series.

Let us introduce the following decomposition for the matrix $A_T^* A_T$,

$$A_T^* A_T = \left[\begin{array}{ccc|c} LF_1^*F_1 & 0 & 0 & \sqrt{L}F_1^*\mathrm{diag}(y_1) \\ 0 & \ddots & \vdots & \vdots \\ 0 & \dots & LF_N^*F_N & \sqrt{L}F_N^*\mathrm{diag}(y_N) \\ \hline \sqrt{L}\mathrm{diag}(y_1)^*F_1 & \dots & \sqrt{L}\mathrm{diag}(y_N)^*F_N & \sum_n \mathrm{diag}(y_n)^2 \end{array}\right] + w_T w_T^* = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{43}$$

Assuming that A and D are invertible and using the Schur complement, we then have,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = M_1 M_2 M_3 \tag{44}$$

where

$$M_1 = \begin{pmatrix} I_{NT} & 0 \\ -D^{-1}C & I_L \end{pmatrix}, \quad M_2 = \begin{pmatrix} \left(A - BD^{-1}C\right)^{-1} & 0 \\ 0 & D^{-1} \end{pmatrix} \tag{45}$$

$$M_3 = \begin{pmatrix} I_{NT} & -BD^{-1} \\ 0 & I_L \end{pmatrix} \tag{46}$$

Where I_{NT} and I_L respectively denote the $NT \times NT$ and $L \times L$ identity matrices. Let $\mathrm{sign}((x_0^w)_T)$ denote the sign of $c_1 x_0$ on the support $T_1 \cup \ldots T_N$. Developping (45) and (46) gives,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} s_0 = \begin{bmatrix} \left(A - BD^{-1}C\right)^{-1} \mathrm{sign}((x_0^w)_T) \\ -D^{-1}C\left(A - BD^{-1}C\right)^{-1} \mathrm{sign}((x_0^w)_T) \end{bmatrix} \tag{47}$$

Using the results of (32) and (33), the certificate on T^c is correspondingly given by

$$Y = -(A_{T^c}^* A_T)D^{-1}C\left(A - BD^{-1}C\right)^{-1} s_0 + w_{T^c} w_T^* \begin{bmatrix} \left(A - BD^{-1}C\right) \mathrm{sign}((x_0^w)_T) \\ -D^{-1}C\left(A - BD^{-1}C\right) \mathrm{sign}((x_0^w)_T) \end{bmatrix} \tag{48}$$

We start by bounding the first term in (48). Let us switch to the decomposition $w = (w_A, w_B)$ of the weight vector w, with $w_A \in \mathbb{C}^{TN}$ and $w_D \in \mathbb{C}^L$, by analogy to the decomposition (43). Let $\tilde{w}_D := \left(\sum_{n=1}^N \mathrm{diag}(y_n)^2\right)^{-1} w_D$. From (43), we have

$$A = LI_{TN} + w_A w_A^*, \tag{49}$$

$$D = \sum_n \mathrm{diag}(|y_n|)^2 + w_D w_D^*, \quad \text{and } D^{-1} = \left(\sum_n \mathrm{diag}(y_n)\right)^{-2} - \frac{\tilde{w}_D \tilde{w}_D^*}{1 + \langle \tilde{w}_D, w_D \rangle} \tag{50}$$

Note that $\sum_n \mathrm{diag}(|y_n|)^2$ as well as $\left(\sum_n \mathrm{diag}(|y_n|)^2\right)^{-1}$ are both strictly postive definite and so $\langle \tilde{w}_D, w_D \rangle > 0$. The next lemma bound the norm of $D^{-1}C$ where C is defined from (43) as

$$C = [\mathrm{diag}(y_1)^* F_1, \ldots, \mathrm{diag}(y_N)^* F_N] \tag{51}$$

Lemma 2.3 (bound on $\|D^{-1}C\|$). *Consider the matrices B and D defined in (43) in the framework of problem (4). For these matrices, the product $D^{-1}C$ can be bounded as*

$$\|D^{-1}C\| \leq \frac{\|h\|_\infty \sup_{n,k} |\hat{x}_n[k]|}{h_{\min}^2 \inf_{k,n} \sum_{n=1}^N \hat{x}_n[k]^2}. \tag{52}$$

Before proving lemma 2.3, note that when multiplying the bound (52) with the one from (42), we get

$$\begin{align}
|Y(t)| &\leq \sup_{n,\ell} L\sqrt{T}\|h\|_\infty |\hat{x}_n[\ell]| \frac{\|h\|_\infty \sup_{k,n} |\hat{x}_n[k]|}{h_{\min}^2 \inf \sum_{n=1}^N \hat{x}_n[k]^2} \|(A - BD^{-1}C)^{-1}\| \|s_0\| \tag{53}\\
&\leq \sqrt{T}L \frac{\|h\|_\infty^2 \left(\sup_{\ell,n} |\hat{x}_n[\ell]|^2\right)}{h_{\min}^2 \inf \sum_{n=1}^N \hat{x}_n[k]^2} \|(A - BD^{-1}C)^{-1}\| \|s_0\| \tag{54}\\
&\leq \sqrt{T}L \frac{h_{\max}^2}{N h_{\min}^2} \|(A - BD^{-1}C)^{-1}\| \|s_0\| \tag{55}
\end{align}$$

where we let $h_{\min}^2 = \inf |\hat{h}[k]|^2$ and correspondingly $h_{\max}^2 = \sup |\hat{h}[k]|^2$, and use condition (8). We now prove lemma 2.3.

Proof. The inverse D^{-1} of D is defined as

$$D^{-1} = \left(\sum_{n=1}^{N} \mathrm{diag}(y_n)^2 + w_D w_D^*\right)^{-1} = \mathrm{diag}\left(\frac{1}{\sum_{n=1}^{N} y_n^2[k]}\right) - \frac{\tilde{w}_D \tilde{w}_D^*}{1 + \langle \tilde{w}_D, w_D \rangle} \tag{56}$$

The operator norm of this matrix can be bounded through Weyl's inequalities as

$$\lambda_{\max}(D^{-1}) \leq \sup_k \frac{1}{\sum_{n=1}^{N} y_n^2[k]} - \frac{1}{1 + \langle \tilde{w}_D, w_D \rangle} \|\tilde{w}_D\|^2 \tag{57}$$

$$\leq \sup_k \frac{1}{\sum_{n=1}^{N} y_n^2[k]} - \left(1 + \sup_k \frac{1}{\sum_n y_n[k]^2}\right)^{-1} \|\tilde{w}_D\|^2 \tag{58}$$

Noting that

$$\langle \tilde{w}_D, w_D \rangle = \sum_k (w_D)_k^2 d_k = \sum_k (w_D)_k^2 \frac{1}{\sum_n y_n[k]^2} \leq \|w_D\|^2 \sup \frac{1}{\sum_n y_n[k]^2}$$

as well as

$$\|\tilde{w}_D\|^2 \geq \inf_k \left(\frac{1}{\sum_n y_n[k]^2}\right)^2 \|w_D\|^2$$

So in particular, we have

$$\lambda_{\max}(D^{-1}) \leq \sup \frac{1}{\sum_{n=1}^{N} y_n^2[k]} - \alpha \inf \frac{1}{\sum_{n=1}^{N} y_n^2[k]} \tag{59}$$

where $\alpha = \inf \frac{1}{\sum_{n=1}^{N} y_n^2[k]} / \left(1 + \sup \frac{1}{\sum_{n=1}^{N} y_n^2[k]}\right)$ and $0 < \alpha < 1$. On the other hand, we have

$$|\hat{y}_n[k]|^2 = L|\hat{h}[k]|^2 \left(\sum_t f_k[t] x_n[t] \sigma_t\right)^2 \geq h_{\min}^2 L |\hat{x}[k]|^2. \tag{60}$$

Taking the sum yields

$$\sum_{n=1}^{N} |\hat{y}_n[k]|^2 \geq h_{\min}^2 L \sum_{n=1}^{N} |\hat{x}_n[k]|^2 \tag{61}$$

Combining (61) with (59) finally gives

$$\|D^{-1}\| \leq \frac{1}{h_{\min}^2 L \sum_{n=1}^{N} |\hat{x}_n[k]|^2} \tag{62}$$

The norm of $\mathrm{diag}(y_n)^* F_n$ can be bounded by noting that

$$\|\mathrm{diag}(\hat{y}_n)^*\sqrt{L}F_n\|^2 \leq L\|F_n\|^2 \sup_k |\hat{y}_n[k]|^2 \\ \leq L^2 \sup_k |\hat{h}[k]|^2|\hat{x}_n[k]|^2 \tag{63}$$

Combining the results from (59) and (63), we get

$$\|D^{-1}C\| \leq \frac{\|\hat{h}\|_\infty \sup_k |\hat{x}_n[k]|}{\hat{h}^2_{\min}\sum_{n=1}^N |\hat{x}_n[k]|^2} \tag{64}$$

which concludes the proof of lemma (2.3). □

To complete the bound on $|Y(t)|$, it remains to show that the norm of $\left(A - BD^{-1}C\right)^{-1}$ is less than $\mathcal{O}(1/L)$. This is the point of lemma (2.4) below. Note that as soon as $\|BD^{-1}C - w_A w_A^*\| < L$, the Neumann series gives

$$\left(A - BD^{-1}C\right)^{-1} = \frac{1}{L}I + \frac{1}{L}\sum_{k=1}^{\infty}\left(BD^{-1}C - w_A w_A^*\right)^k \tag{65}$$

Lemma 2.4 (Bound on $\|(A - BD^{-1}C)^{-1}\|$). *Consider the matrices A, B, C and D defined in (43) in the framework of the linearization (4). The norm of the inverse $(A - BD^{-1}C)^{-1}$ can be bounded as*

$$\|(A - BD^{-1}C)^{-1}\| \leq \frac{1}{L} + \mathcal{O}(\frac{\sqrt{T}}{L\sqrt{N}}) \tag{66}$$

Proof. As explained above, we only need to show that $\|BD^{-1}C - w_A w_A^*\| < L$. Since $\|w_A\|^2 = 1$, as soon as $L > 1$, we can just focus on the contribution from $BD^{-1}C$. For this last term, we have

$$BD^{-1}C = L\begin{bmatrix} F_1^*\mathrm{diag}(y_1) \\ \vdots \\ F_N^*\mathrm{diag}(y_N)\end{bmatrix} D^{-1}\left[\mathrm{diag}(y_1)^*F_1, \ldots, \mathrm{diag}(y_N)^*F_N\right] \tag{67}$$

$$+ w_A w_D^* D^{-1}\sqrt{L}\left[\mathrm{diag}(y_1)^*F_1, \ldots, \mathrm{diag}(y_N)^*F_N\right] + \sqrt{L}\begin{bmatrix} F_1^*\mathrm{diag}(y_1) \\ \vdots \\ F_N^*\mathrm{diag}(y_N)\end{bmatrix} D^{-1}w_D w_A^* + w_A w_D^* D^{-1} w_D w_A^* \tag{68}$$

Using the bound (62), the first term can be bounded as

$$L\begin{bmatrix} F_1^*\mathrm{diag}(y_1) \\ \vdots \\ F_N^*\mathrm{diag}(y_N)\end{bmatrix} D^{-1}\left[\mathrm{diag}(y_1)^*F_1, \ldots, \mathrm{diag}(y_N)^*F_N\right] \tag{69}$$

$$\preceq L\begin{bmatrix} F_1^*\mathrm{diag}(\frac{|y_1[k]|^2}{\underline{d}})F_1 & \ldots & F_1^*\mathrm{diag}(\frac{y_1[k]y_N[k]}{\underline{d}})F_n \\ \vdots & \ddots & \vdots \\ F_N^*\mathrm{diag}(\frac{y_N[k]y_1^*[k]}{\underline{d}})F_1 & \ldots & F_N^*\mathrm{diag}(\frac{|y_N[k]|^2}{\underline{d}})F_N \end{bmatrix} \tag{70}$$

$$\preceq \frac{L}{\underline{d}} \begin{pmatrix} F_1^* & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & F_N^* \end{pmatrix} \begin{bmatrix} \mathrm{diag}(|y_1[k]|^2) & \dots & \mathrm{diag}(y_1[k]y_N[k]) \\ \vdots & \ddots & \vdots \\ \mathrm{diag}(y_N[k]y_1^*[k]) & \dots & \mathrm{diag}(|y_N[k]|^2) \end{bmatrix} \begin{pmatrix} F_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & F_N \end{pmatrix} \tag{71}$$

Where we let $\underline{d}$ denote the infimum

$$\underline{d} := \inf_k L|h[k]|^2 \sum_{n=1}^{N} |\hat{x}_n[k]|^2 \geq c_2(L) \tag{72}$$

Clearly we have $\|F_n^*\| = \sqrt{T}$ and $\|F_n\| = 1$ and we can thus write,

$$\left\| L \begin{bmatrix} F_1^* \mathrm{diag}(y_1) \\ \vdots \\ F_N^* \mathrm{diag}(y_N) \end{bmatrix} D^{-1} \left[\mathrm{diag}(y_1)^* F_1, \dots, \mathrm{diag}(y_N)^* F_N \right] \right\| \tag{73}$$

$$\leq \frac{L\sqrt{T}}{\underline{d}} \left\| \begin{bmatrix} \mathrm{diag}(|y_1[k]|^2) & \dots & \mathrm{diag}(y_1[k]y_N[k]) \\ \vdots & \ddots & \vdots \\ \mathrm{diag}(y_N[k]y_1^*[k]) & \dots & \mathrm{diag}(|y_N[k]|^2) \end{bmatrix} \right\| \tag{74}$$

$$\leq \frac{L\sqrt{T}}{\underline{d}} \left\| \begin{bmatrix} \mathrm{diag}(y_1) \\ \vdots \\ \mathrm{diag}(y_N) \end{bmatrix} \left[\mathrm{diag}(y_1^*), \dots, \mathrm{diag}(y_N)^* \right] \right\| \tag{75}$$

The last line of (75) then gives

$$\left\| L \begin{bmatrix} F_1^* \mathrm{diag}(y_1) \\ \vdots \\ F_N^* \mathrm{diag}(y_N) \end{bmatrix} D^{-1} \left[\mathrm{diag}(y_1)^* F_1, \dots, \mathrm{diag}(y_N)^* F_N \right] \right\| \leq \frac{L\sqrt{T}}{\underline{d}} \sqrt{N} \sup_n \|\mathrm{diag}(y_n)\|^2 \tag{76}$$

Using the definition of $y_n[k]$,

$$|y_n[k]|^2 \leq L|h[k]|^2 |\langle f_k^*, x_n \rangle|^2 \leq L|h[k]|^2 \sup_{n,k} |\hat{x}_n[k]|^2 \tag{77}$$

In particular as soon as $N \geq T$, and the condition (8) is satisfied, we have

$$\left\| L \begin{bmatrix} F_1^* \mathrm{diag}(y_1) \\ \vdots \\ F_N^* \mathrm{diag}(y_N) \end{bmatrix} D^{-1} \left[\mathrm{diag}(y_1)^* F_1, \dots, \mathrm{diag}(y_N)^* F_N \right] \right\| \leq \mathcal{O}\left(\frac{L\sqrt{T}}{\sqrt{N}} \right) \tag{78}$$

For each of the remaining terms, we respectively have,

$$\sqrt{L}\|w_A w_D^* D^{-1}\left[\text{diag}(y_1)^* F_1, \ldots, \text{diag}(y_n)^* F_n\right]\| \leq \sqrt{L}\left\|D^{-1}\left[\text{diag}(y_1)^* F_1, \ldots, \text{diag}(y_n)^* F_n\right]\right\| \tag{79}$$

$$\leq \sqrt{L}\|D^{-1}\| \sup_n \|\text{diag}(y_n)^* F_n\| \tag{80}$$

$$\leq \sqrt{L}\|D^{-1}\| \sup_{n,k} |y_n[k]| \tag{81}$$

$$\leq \sqrt{L}\frac{\sup_{n,k}|y_n[k]|}{h_{\min}^2 \inf L\sum_{n=1}^N \hat{x}_n[k]^2} \tag{82}$$

Where we use $\|w_A\|$,$\|w_D\| = 1$. This last expression is at most on the order of L/N. if $h_{\min}^2 = \mathcal{O}(1/L)$ and $|\hat{x}_n[k]|^2$ is $\mathcal{O}(1/L)$ and $\mathcal{O}(\sqrt{L}/N)$ if $h_{\min}^2 \sim \mathcal{O}(1)$. In a similar way, we have

$$\left\|\begin{bmatrix} F_1^*\text{diag}(y_1) \\ \vdots \\ F_N^*\text{diag}(y_N)\end{bmatrix} D^{-1} w_D w_A^*\right\| \leq \sqrt{TNL}\frac{\sup_{n,k}|y_n[k]|}{h_{\min}^2 \inf L\sum_{n=1}^N \hat{x}_n[k]^2} \tag{83}$$

Which is at most on the order of $L\sqrt{T}/\sqrt{N}$. Finally, for the last term in (68), we have,

$$\|w_A w_D^* D^{-1} w_D w_A^*\| = \|w_A w_D^*\| \left\|\left(\text{diag}\left(\frac{1}{\sum_{n=1}^N y_n^2[i]}\right) - \frac{\tilde{w}_D\tilde{w}_D^*}{1+\langle\tilde{w}_D, w_D\rangle}\right)\right\| \|w_D w_A^*\| \tag{84}$$

$$\leq \left\|\left(\text{diag}\left(\frac{1}{\sum_{n=1}^N y_n^2[i]}\right) - \frac{\tilde{w}_D\tilde{w}_D^*}{1+\langle\tilde{w}_D, w_D\rangle}\right)\right\| \tag{85}$$

$$\leq \frac{1}{h_{\min}^2 L\sum_{n=1}^N \hat{x}_n[k]^2} \tag{86}$$

In the last line, we use (59). This term is at most an order of L/N when $|\hat{x}_n[k]|^2 \sim \mathcal{O}(1/L)$ and $h_{\min}^2 \sim \mathcal{O}(1/L)$ whenever condition (8) is satisfied. This concludes the proof of lemma 2.4

□

Given lemmas 2.3 and 2.4, we can now bound the two terms of (48). For the first term, using the Neumann series, plugging (66) into (48) and using $\|s_0\| = \sqrt{NT}$, we get

$$|Y(t)| \lesssim \frac{1}{L}\sqrt{T}L\frac{\|h\|_\infty^2\left(\sup_{\ell,n}|\hat{x}_n[\ell]|^2\right)}{h_{\min}^2 \inf_k \sum_{n=1}^N \hat{x}_n[k]^2}\|s_0\|, \quad \forall t \in T^c \tag{87}$$

$$\lesssim T\sqrt{N}\frac{\|h\|_\infty^2\left(\sup_{\ell,n}|\hat{x}_n[\ell]|^2\right)}{h_{\min}^2 \inf \sum_{n=1}^N \hat{x}_n[k]^2}, \quad \forall t \in T^c \tag{88}$$

For the second term in (48), using $\|w\| = 1$, this second term reads as

$$\frac{1}{\sqrt{LN+L}}\|w_T\| \max\left\{\left\|\left(A - BD^{-1}C\right)^{-1}\right\|\|s_0\|, \left\|D^{-1}C\left(A - BD^{-1}C\right)\right\|\|s_0\|\right\} \tag{89}$$

Using the bounds from (64) and (66), as well as the fact that $\|w\| = 1$, (89) reads as

$$\frac{\sqrt{NT}}{\sqrt{LN+L}} \max\left\{\frac{1}{L}, \frac{1}{L}\frac{\|\hat{h}\|_\infty \sup_k |\hat{x}_n[k]|}{\hat{h}_{\min}^2 \sum_{n=1}^N |\hat{x}_n[k]|^2}\right\} \tag{90}$$

The second term in the maximum is at most on the order of $\mathcal{O}(1)$ whenever both $|\hat{x}_n[k]| = \mathcal{O}(1/\sqrt{L})$ and $\|\hat{h}\|_\infty = \mathcal{O}(1/\sqrt{L})$. This last term can be made sufficiently small as soon as $L \gtrsim T$.

This concludes the proof of proposition 2 and thus Theorem 1.1.

3. NUMERICAL EXPERIMENTS

In this section we provide numerical experiments to validate the result of Theorem 1.1. We start by looking at the magnitude of the certificate on T and T^c for values of L, N and T leading to exact recovery. The result is shown in Fig. 1. The figure on the top shows the modulus on T^c. The figure at the bottom displays the modulus of the entries corresponding to the inverse of the filter ($|s_0|_\Gamma = 0$) as well as the support of the x_n ($|(s_0)|_{T_n} = 1$). From those two figures, one can see that the construction (13) can be used to guarantee the recovery whenever the ℓ_1-minimization program (4) succeeds.

We then study the evolution of the recovery for various values of the dimensions L, N and T. We consider that exact recovery happens as soon as

$$\frac{1}{2}\left(\frac{\|h-h_0\|}{\|h_0\|} + \frac{\|x-x_0\|}{\|x_0\|}\right) < .01 \tag{91}$$

Where h and x are obtained from the solutions to (4) after appropriate normalization (see section 1.). For each pair (L, T) and (N, T), we solve the ℓ_1 minimization problem (4) 10 times. Each outcome is labeled as a success (1) or failure (0) depending on the value of the criterion (91). The labels for each sequence of 10 experiments are then averaged and displayed in Fig 2. The L vs T figure seems to illustrate a linear dependence between L and T, whether the N vs T phase transition seems to highlight the fact that as soon as a few inputs are considered ($N \gtrsim 1$), solvability of the problem mostly depends on L and T. Steepness of the phase transition in (N, T) is due to the value of L used to conduct the experiments. Here $L = 80$ so that when $T \geq 60$, the inputs can simply not be considered as sparse anymore and recovery does not happen independently of the value of N. Together the two phase transition plots seem to suggest that it might be possible to get tighter $L \gtrsim T_{\max}$ and $N \gtrsim 1$ sample complexities.

Finally, we study recovery as a function of the magnitude of the entries of $\boldsymbol{h}$. We start by solving the ℓ_1 formulation (4) for various values of the range of $\boldsymbol{h}$. The result is shown in Fig (3) (Left). From this figure, it again seems that the recovery might be independent from the magnitude of the entries of h. Finally, we study recovery for various bounds on the minimal entry in h, keeping the range to a constant (2) value. For each of the lower bounds on the entries of h, evolution of the recovery is shown in Fig. 3 (Right). Despite a few numerical fluctuations, this figure again seems to suggests that recovery might be independent of the magnitude of the entries of h.

4. CONCLUSION

In this note we study the blind recovery of sequences of spikes, $\boldsymbol{x}_n \in \mathbb{R}^L$ that are measured through some unknown point spread function $\boldsymbol{h} \in \mathbb{R}^L$. We show that an ℓ_1-minimization approach coupled to an appropriate linearization of the problem following Ling et al. 2006[1] can be certified to recover the sparse sequences as well as the point spread function as soon as the number of input sequences N and dimension of each input sequence L obey $N \gtrsim T_{\max}^2$ and $L \gtrsim N$ where $T_{\max}$ denotes the maximum support over all sequences. This result holds with high probability on the choice of each support as well as on the signs of each measure on the support. Our proof system relies on the construction of a dual certificate sharing the structure first introduced in Candès et al.

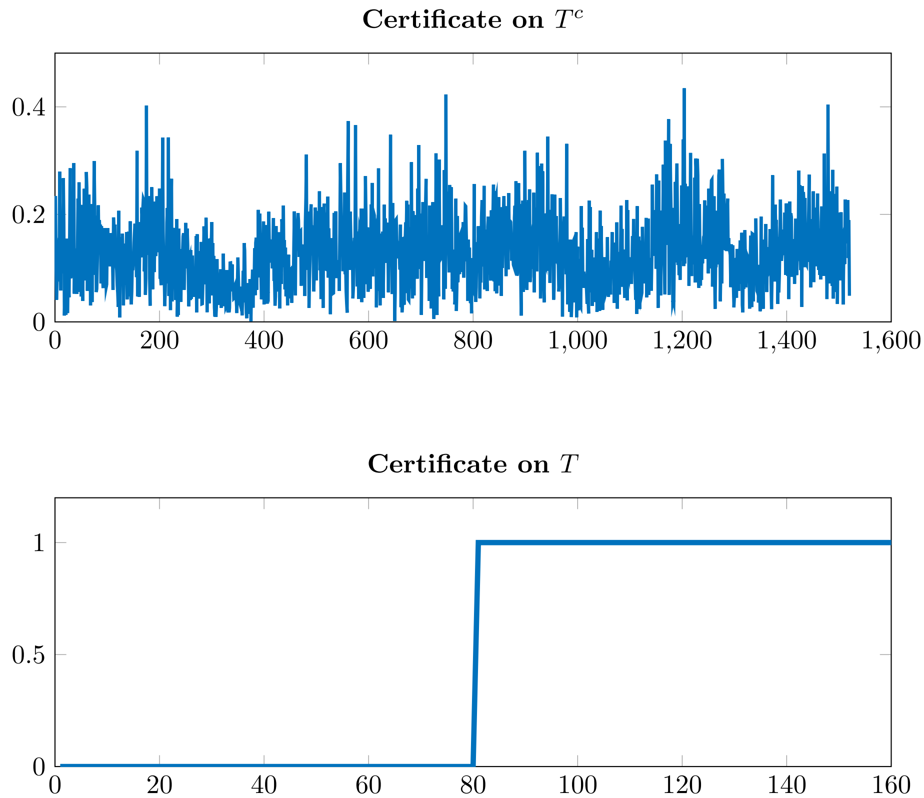

Figure 1. Simulations of the certificate on T (bottom) and T^c (top). Here $L = 80$, $N = 20$ and $S = 4$. The certificate can be seen to satisfy the optimality conditions of proposition 1.

2007.[2] This certificate is shown to satisfy the usual optimality conditions from duality theory. Randomness of the signs is needed to show concentration of the minimal eigenvalue of $A_T^* A_T$ and thus invertibility of this same matrix, as well as to bound the norm $\|(A^* A_T)_{T^c}\|$. Potential applications of our result include marine seismic data deghosting (see Parkes et al. 2013.[19]), Magnetic resonance autocalibration (see Uecker et al.[22]) or multiple channel estimation as in Romberg et al. 2013.[17] Numerical experiments seem to suggest that the dependence in the range of h might not be needed and that the scalings could be improved to $L \gtrsim T_{\max}$, $N \gtrsim 1$.

ACKNOWLEDGMENTS

AC was supported by the FNRS, FSMP, BAEF and Francqui Foundations. He is grateful to Ali Ahmed for suggesting the problem as well as for meaningful discussions on an early version of this paper. He is grateful to Laurent Demanet for suggesting the conference and to Carlos Fernandez-Granda for mentioning the work of Uecker et al. on MRI auto-calibration.

REFERENCES

[1] Ling, S. and Strohmer, T., "Self-calibration via linear least squares," *arXiv preprint arXiv:1611.04196* (2016).

[2] Candes, E. and Romberg, J., "Sparsity and incoherence in compressive sampling," *Inverse problems* **23**(3), 969 (2007).

[3] Wang, L. and Chi, Y., "Blind deconvolution from multiple sparse inputs," *IEEE Signal Processing Letters* **23**(10), 1384–1388 (2016).

[4] Ahmed, A., Recht, B., and Romberg, J., "Blind deconvolution using convex programming," *Information Theory, IEEE Transactions on* **60**(3), 1711–1732 (2014).

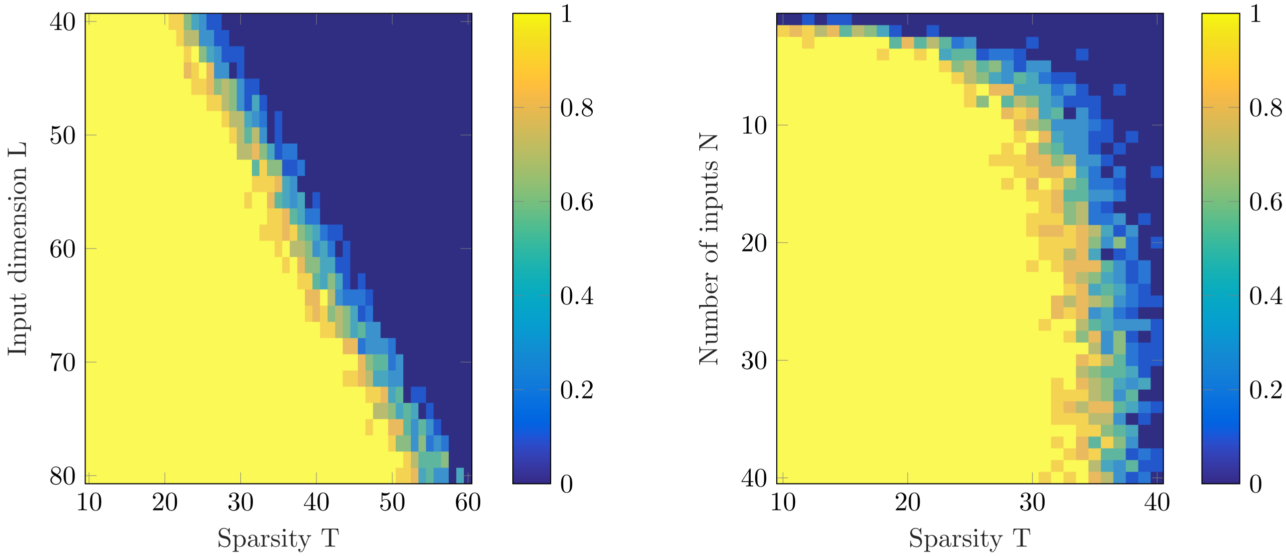

Figure 2. Phase transition for the recovery through the ℓ_1 minimization program (4). 1 denotes exact recovery whereas 0 is used to represent failure. The first phase transition (Left) illustrates the evolution of the recovery at $N = 20$, for $40 \leq L \leq 80$ and $10 \leq T \leq 60$. For each pair, (N, T), the experiments were reapeated 10 times each and the binary outcomes (success/failure) were averaged. The second figure (Right) illustrates the evolution of the recovery at $L = 60$ for $1 \leq N \leq 40$ and $10 \leq T \leq 40$. Just as for the first figure, experiments were repeated 10 times for each pair (N, T) and the binary outcomes (success/failure) were averaged. Those two Figues seem to illutrate the fact that scalings on the order of $L \gtrsim T_{\max}$ and $N \gtrsim 1$ might be obtained (see the discussion in Section 3.)

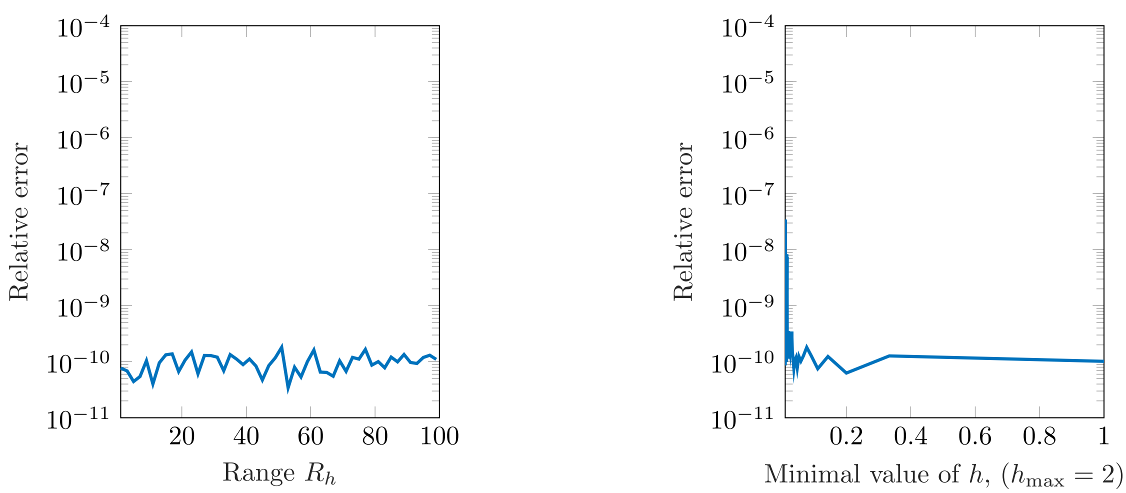

Figure 3. Evolution of the relative error (91) as a function of the range of h, for $1 \leq R_h \leq 100$ (Left) and the smallest entry in h (Right). Instabilities appear at small values of h due to the non-linearity of problem (3).

[5] Ahmed, A., Cosse, A., and Demanet, L., "A convex approach to blind deconvolution with diverse inputs," *Proceedings of the 6th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP'15)* (2015).

[6] Ling, S. and Strohmer, T., "Blind deconvolution meets blind demixing: Algorithms and performance bounds," *arXiv preprint arXiv:1512.07730* (2015).

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

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

[9] Ahmed, A. and Demanet, L., "Leveraging diversity and sparsity in blind deconvolution," *arXiv preprint arXiv:1610.06098* (2016).

[10] Li, Y., Lee, K., and Bresler, Y., "Blind gain and phase calibration for low-dimensional or sparse signal sensing via power iteration," *SampTA* (2017).

[11] Gribonval, R., Chardon, G., and Daudet, L., "Blind calibration for compressed sensing by convex optimization," in [*Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*], 2713–2716, IEEE (2012).

[12] Bilen, C., Puy, G., Gribonval, R., and Daudet, L., "Convex optimization approaches for blind sensor calibration using sparsity," *Signal Processing, IEEE Transactions on* **62**(18), 4847–4856 (2014).

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

[14] Liu, E., Iqbal, N., McClellan, J. H., and Al-Shuhail, A. A., "Sparse blind deconvolution of seismic data via spectral projected-gradient," *arXiv preprint arXiv:1611.03754* (2016).

[15] Larue, A., Van Der Baan, M., Mars, J. I., and Jutten, C., "Sparsity or whiteness: what criterion to use for blind deconvolution of seismic data?," in [*SEG Technical Program Expanded Abstracts 2005*], 1642–1645, Society of Exploration Geophysicists (2005).

[16] Schulz, T. J., "Multiframe blind deconvolution of astronomical images," *JOSA A* **10**(5), 1064–1073 (1993).

[17] Romberg, J., Tian, N., and Sabra, K., "Multichannel blind deconvolution using low rank recovery," in [*SPIE Defense, Security, and Sensing*], 87500E–87500E, International Society for Optics and Photonics (2013).

[18] Bernstein, B. and Fernandez-Granda, C., "Deconvolution of point sources: a sampling theorem and robustness guarantees," *to appear* (2017).

[19] Parkes, G. and Hatton, L., "The marine seismic source," **4** (2013).

[20] Uecker, M., Lai, P., Murphy, M. J., Virtue, P., Elad, M., Pauly, J. M., Vasanawala, S. S., and Lustig, M., "Espirit—an eigenvalue approach to autocalibrating parallel mri: where sense meets grappa," *Magnetic resonance in medicine* **71**(3), 990–1001 (2014).

[21] Uecker, M. and Lustig, M., "Estimating absolute-phase maps using espirit and virtual conjugate coils," *Magnetic resonance in medicine* **77**(3), 1201–1207 (2017).

[22] Uecker, M., "Autocalibrating and calibrationless parallel magnetic resonance imaging as a bilinear inverse problem," *arXiv preprint arXiv:1705.04081* (2017).

[23] Candès, E., Romberg, J., and Tao, T., "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory* **52**, 489–509 (February 2006).

[24] Van Der Vaart, A. W. and Wellner, J. A., "Weak convergence," in [*Weak Convergence and Empirical Processes*], 16–28, Springer (1996).