CHANNEL PROTECTION USING RANDOM MODULATION

Ali Ahmed, Humera Hameed

Department of Electrical Engineering,
Information Technology University,Lahore, Pakistan
{ali.ahmed,phdee17001}@itu.edu.pk

ABSTRACT

This paper shows that modulation protects a bandlimited signal against convolutive interference. A signal $s(t)$, bandlimited to BHz, is modulated (pointwise multiplied) with a known random sign sequence $r(t)$, alternating at a rate Q, and the resultant *spread spectrum* signal $s(t) \odot r(t)$ is convolved against an M-tap channel impulse response $h(t)$ to yield the observed signal $y(t) = (s(t) \odot r(t)) \circledast h(t)$, where $\odot$ and $\circledast$ denote pointwise multiplication, and circular convolution, respectively.

We show that both $s(t)$, and $h(t)$ can be provably recovered using a simple gradient descent scheme by alternating the binary waveform $r(t)$ at a rate $Q \gtrsim B + M$(to within log factors and a signal coherences) and sampling $y(t)$ at a rate Q. We also present a comprehensive set of phase transitions to depict the trade-off between Q, M, and B for successful recovery. Moreover, we show stable recovery results under noise.

Index Terms— Blind deconvolution, gradient descent, modulation, random signs, channel protection

1. INTRODUCTION

This paper shows that a simple modulation of a bandlimited signal protects it against unknown convolutive channel interference. We take a periodic signal $s(t)$ in $t \in [0, 1)$, bandlimited to B Hertz. We can express $s(t)$ in Fourier basis as

$$s(t) = \sum_{k=-B}^{B} x[k]e^{i2\pi kt}, \ t \in [0,1) \qquad (1)$$

Shannon-Nyquist sampling theorem tell us that the signal $s(t)$ can be captured by uniformly sampling at a rate $Q \geq K = 2B + 1$. We modulate $s(t)$ by pointwise multiplying it with a known random binary waveform $r(t)$, alternating at a rate Q, and use $r(t) \odot s(t)$ to denote the modulated signal, where $\odot$ represents the pointwise multiplication. The modulated signal is then filtered via an unknown LTI system characterized by an impulse response $h(t)$ resulting in the convolution $y(t) = (r(t) \odot s(t)) \circledast h(t)$, where $\circledast$ denotes the circular convolution. We observe $y(t)$, and aim to recover both $s(t)$, and $h(t)$. Our main result shows that $s(t)$, and $h(t)$ can be recovered using a simple gradient descent algorithm. Moreover, we are able to quantify exactly the alternating rate Q of the binary waveform,

and the sampling rate, which is also fixed at Q, of $y(t)$ sufficient to achieve the exact recovery of $s(t)$, and $h(t)$.

Modulation is prevalently used in signal processing and communications to, for example, effectively use available spectral bands by multiplexing signals in different frequency bands [1]. Perhaps more closely related to our framework is code division multiple access [2], where also a binary waveform is pointwise multiplied with data signal primarily to mitigate additive interferences from other user data signals. In this paper, however, we are claiming that binary modulation also protects a signal against more involved and unknown convolutive interference. In other words, a premodulation step enables us to deconvolve the signal, and impulse response without knowing either of them. In general, deconvolution of two unknown inputs is referred to as blind deconvolution (BD) and is severely ill-posed. BD is one of the fundamental problems in signal processing, communications, system theory, etc, [3, 4] and this paper shows that a simple modulation step, which already happens to be a part of the signal preprocessing assembly line prior to transmission in many wireless communication systems, enables us to undo the convolution with an unknown impulse response at the receiver.

1.1. Discrete Form

The bandlimited signal $s(t)$ in (1) can be captured taking $Q \geq K = 2B+1$ equally spaced samples at time instants $\mathcal{T}_Q := \{\frac{q-1}{Q}, q \in [Q]\}$, where, in general, $[Q] = \{1, 2, 3, \ldots, Q\}$. Let $\boldsymbol{F}$ be a $Q \times Q$ DFT matrix with entries $F[q,k] = \frac{1}{\sqrt{Q}}e^{-j2\pi kq/Q}$, $(q,k) \in [Q] \times [Q]$. In general, we will denote by $\boldsymbol{F}_J$, a submatrix formed by selecting the first J columns of $\boldsymbol{F}$. Similarly, the notation $\boldsymbol{F}_J^*$ denotes a submatrix consisting of the first J columns of $\boldsymbol{F}^*$ where $\boldsymbol{F}^*$ denotes hermitian of $\boldsymbol{F}$. Then the samples of $s(t)$ can be expressed as

$$\boldsymbol{s} = \boldsymbol{F}_K^* \boldsymbol{x},$$

where the entries of the K-vector $\boldsymbol{x}$ are the Fourier coefficients. Let $\boldsymbol{r}$ be a Q-vector of samples of binary waveform $r(t)$ in $\mathcal{T}_Q$. We model the impulse response $h(t)$ as an M-tap filter

$$h(t) = \sum_{m=1}^{M} h[m]\delta(t - t_m), \qquad (2)$$

where each $t_m \in \mathcal{T}_Q$ is unique and known. Assuming $h[m]$ to be the mth entry of an M-vector $\boldsymbol{h}$. The rate Q samples of $y(t)$ on the grid $t \in \mathcal{T}_Q$ are collected in Q-vector $\boldsymbol{y}$ below

$$\boldsymbol{y} = (\boldsymbol{r} \odot \boldsymbol{s}) \circledast \boldsymbol{h}, \qquad (3)$$

where $\circledast$ denotes the Q-point circular convolution[1], and $\boldsymbol{y} \in \mathbb{C}^Q$. The objective is to recover the unknown $\boldsymbol{x}$, and $\boldsymbol{h}$ from $\boldsymbol{y}$.

Formally, we take the measurements in the Fourier domain

$$\hat{\boldsymbol{y}} = \boldsymbol{F}\boldsymbol{y} = \sqrt{Q}(\boldsymbol{F}(\boldsymbol{r} \odot \boldsymbol{s}_0) \odot \boldsymbol{F}_M \boldsymbol{h}_0) + \boldsymbol{e}$$
$$= \sqrt{Q}(\boldsymbol{F}_Q \boldsymbol{R} \boldsymbol{F}_K^* \boldsymbol{x}_0 \odot \boldsymbol{F}_M \boldsymbol{h}_0) + \boldsymbol{e}, \qquad (4)$$

where $\boldsymbol{R} := \mathrm{diag}(\boldsymbol{r})$ is a $Q{\times}Q$ diagonal matrix, $\boldsymbol{h}_0, \boldsymbol{x}_0$ are the ground truths, and $\boldsymbol{e} \in \mathbb{C}^Q$ denote the additive noise in the Fourier domain. To deconvolve, we minimize the measurement loss by taking a gradient step in each of the unknowns $\boldsymbol{h}_0$, and $\boldsymbol{x}_0$ while keeping the other fixed. This paper details a particular set of conditions on the sample complexity, subspace dimensions, and the signals/filters under which this computationally feasible alternating gradient descent scheme provably succeeds.

Notice that the measurements in (4) are non-linear in the unknowns $(\boldsymbol{h}_0, \boldsymbol{x}_0)$, however, are linear in the rank-1 outer-product $\boldsymbol{h}_0 \bar{\boldsymbol{x}}_0^*$ where $\bar{\boldsymbol{x}}_0^*$ denotes conjugate hermitian of $\boldsymbol{x}_0$. To see this, let $\boldsymbol{f}_q^* \in \mathbb{C}^M$ be the qth row of $\boldsymbol{F}_M$ and $\hat{\boldsymbol{g}}_q \in \mathbb{C}^K$ be the qth row of the $Q \times K$ matrix $\sqrt{Q}(\boldsymbol{F}_Q \boldsymbol{R} \boldsymbol{F}_K^*)$. The qth entry $\hat{y}[q]$ of measurements $\hat{\boldsymbol{y}}$ in (4) is then simply

$$\hat{y}[q] = \boldsymbol{f}_q^* \boldsymbol{h}_0 \bar{\boldsymbol{x}}_0^* \hat{\boldsymbol{g}}_q + e[q] = \langle \boldsymbol{f}_q \hat{\boldsymbol{g}}_q^*, \boldsymbol{h}_0 \bar{\boldsymbol{x}}_0^* \rangle + e[q]; \qquad (5)$$

the linearity of the measurements in $\boldsymbol{h}_0 \bar{\boldsymbol{x}}_0^*$ is clear from the last equality above. We also define a linear map $\mathcal{A} : \mathbb{C}^{M \times K} \to \mathbb{C}^Q$ that maps $\boldsymbol{h}_0 \boldsymbol{x}_0^*$ to the vector $\hat{\boldsymbol{y}}$. The action of $\mathcal{A}$ on a rank-1 matrix $\boldsymbol{h}\boldsymbol{x}^*$ returns

$$\mathcal{A}(\boldsymbol{h}\boldsymbol{x}^*) := \{\boldsymbol{f}_q^* \boldsymbol{h} \bar{\boldsymbol{x}}^* \hat{\boldsymbol{g}}_q\}_q, \ q \in [Q],$$
$$\text{and therefore, } \hat{\boldsymbol{y}} = \mathcal{A}(\boldsymbol{h}_0 \boldsymbol{x}_0^*) + \boldsymbol{e}, \qquad (6)$$

where we used the definition of $\mathcal{A}$ to compactly express (5).
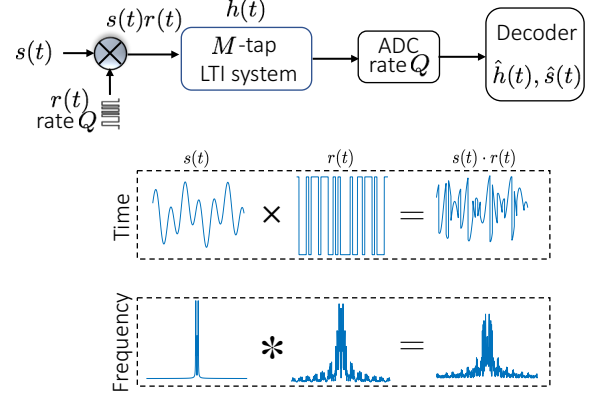
1.2. Implementation Potential

Binary modulation of an analog signal can be easily implemented using switches that flip the signs of the signal in real time; the setup is shown in Figure 1.

Fast rate binary switches can be easily implemented; see, for example, [5], and [6–8] for the use of binary switches in other applications in signal processing. The implementation potential combined with the ubiquity of blind deconvolution make this result interesting in system theory, applied communications, and signal processing, among other.

1.3. Recent Literature

It is important to point out critical differences in the structural assumptions compared to the contemporary recent literature [9], and [10–14] on blind deconvolution, where the convolved signals are assumed to live in known subspaces spanned by the columns of a random Gaussian matrices. Such random subspaces do not naturally arise in applications. We relinquish such restrictive Gaussian subspace assumption, and give a provable blind deconvolution result by only assuming random/generic sign (modulated) signal $\boldsymbol{s} = \boldsymbol{F}_K^* \boldsymbol{x}$ that resides in realistic subspace spanned by the columns of the DFT matrix.

---

[1] The Q-point circular convolution $\boldsymbol{x}_1 \circledast \boldsymbol{x}_2$ of $\boldsymbol{x}_1 \in \mathbb{C}^Q$, and $\boldsymbol{x}_2 \in \mathbb{C}^M$, where $M \leq Q$, is the circular convolution of Q-vector $\boldsymbol{x}_1$, and $\boldsymbol{x}_2$ zero padded to length Q. Mathematically, $\boldsymbol{x}_1 \circledast \boldsymbol{x}_2 = \boldsymbol{F}^* \mathrm{diag}(\hat{\boldsymbol{x}}_1) \boldsymbol{F}_M \boldsymbol{x}_2$, where $\hat{\boldsymbol{x}}_1 = \sqrt{Q} \boldsymbol{F} \boldsymbol{x}_1$.

Fig. 1. *Analog implementation for real time protection against channel intereference. A continuosus time signal $s(t)$, bandlimited to B Hz, is modulated with a random binary waveform $r(t)$ alternating at a rate Q. The modulated signal drives an unknown LTI system characterized by an M-tap impulse response $h(t)$. The resulting signal is sampled at a rate Q. Operate the modulator and ADC at a rate $Q \gtrsim \max(B, M)$ (to within a constant, log factors and coherences), and recover $s(t)$, and $h(t)$ using gradient descent. Underneath, the preprocessing is shown in time, and frequency domain. Modulation in time domain spreads the spectrum, and the resulting higher frequency signal remains protected against the distortions caused by an unknown LTI system.*

1.4. Coherence Parameters

Our main theoretical results depend on some signal dispersion measures that characterize how diffuse signals are in the Fourier domain. Intuitively, concentrated (not diffuse) signals in the Fourier domain in *annihilate* the measurements in (4) making it relatively difficult (more stringent sample complexity requirements) to recover such signals. We refer to the signal diffusion measures as coherence parameters, defined and discussed below.

For arbitrary vectors $\boldsymbol{h} \in \mathbb{C}^M$, $\boldsymbol{x} \in \mathbb{C}^K$, we define coherences

$$\mu_h^2 := Q \frac{\|\boldsymbol{F}_M \boldsymbol{h}\|_\infty^2}{\|\boldsymbol{h}\|_2^2}, \quad \nu_x^2 := Q \frac{\|\boldsymbol{F}_K^* \boldsymbol{x}\|_\infty^2}{\|\boldsymbol{x}\|_2^2}. \qquad (7)$$

Similar coherence parameters appear in the related recent literature on blind deconvolution [10, 13], and elsewhere in compressed sensing [15, 16], in general. Without loss of generality, we assume that $\|\boldsymbol{h}_0\|_2 = \sqrt{d_0}$, and $\|\boldsymbol{x}_0\|_2 = \sqrt{d_0}$. For brevity, we will denote the coherence parameters $\mu_{h_0}^2$, and $\nu_{x_0}^2$ of the fixed ground truth vectors $(\boldsymbol{h}_0, \boldsymbol{x}_0)$ by

$$\mu^2 := \mu_{h_0}^2, \text{ and } \nu^2 := \nu_{x_0}^2. \qquad (8)$$

In words, coherence parameter μ_h^2 is the peak value of the frequency spectrum of a fixed norm vector $\boldsymbol{h}$. A higher value roughly indicates a concentrated spectrum and vice versa. It is easy to check that $1 \leq \mu_h^2 \leq Q$.

On the other hand, ν_x^2 quantifies the dispersion (not in the Fourier domain) of the signal $\boldsymbol{s} = \boldsymbol{F}_K^* \boldsymbol{x}$. A signal concentrated in time (mostly zero) remains somewhat oblivious to the random sign flips $\boldsymbol{r} \odot \boldsymbol{s}$, and as a result is not as well-dispersed in the frequency domain. Let $\boldsymbol{g}_q^*$ be the rows of $\boldsymbol{F}_K^*$. By definition, $\nu_x^2 \|\boldsymbol{x}\|_2^2 \geq Q|\boldsymbol{g}_q^* \boldsymbol{x}|^2$ for any $q \in [Q]$. Summing over $q \in [Q]$ on both sides, and using

the isometry of F_K^* gives us the inequality $\nu_x^2 \geq 1$. The upper bound $\nu_x^2 \leq Q$ is easy to see using Cauchy Schwartz inequality, hence, $1 \leq \nu_x^2 \leq Q$.

2. RECOVERY VIA GRADIENT DESCENT

Given measurements $\hat{y}$ of the ground truth (h_0, x_0), we employ a regularized gradient descent algorithm that aims to minimize a loss function:

$$\tilde{F}(h, x) := F(h, x) + G(h, x). \tag{9}$$

w.r.t. h, and x, where the functions $F(h, x)$, and $G(h, x)$ account for the measurement loss, and regularization, respectively; and are defined below

$$F(h, x) := \|\mathcal{A}(hx^*) - \hat{y}\|_2^2 = \|\mathcal{A}(hx^* - h_0x_0^*) - e\|_2^2 =$$

$$\|\mathcal{A}(hx^* - h_0x_0^*)\|_2^2 + \|e\|_2^2 - 2\mathrm{Re}(\langle \mathcal{A}^*(e), hx^* - h_0x_0^* \rangle), \tag{10}$$

and

$$G(h, x) := \rho \Bigg[G_0 \left(\frac{\|h\|_2^2}{2d} \right) + G_0 \left(\frac{\|x\|_2^2}{2d} \right) + \sum_{q=1}^{Q} G_0 \left(\frac{Q|f_q^*h|^2}{8d\mu^2} \right)$$

$$+ \sum_{q=1}^{Q} G_0 \left(\frac{Q|g_q^*x|^2}{8d\nu^2} \right) \Bigg], \tag{11}$$

where $G_0(z) = \max\{z - 1, 0\}^2$. We set $\rho \geq d^2 + \|e\|_2^2$, and $0.9d_0 \leq d \leq 1.1d_0$. Evidently, the regularizer $G(h, x)$ restricts the coherences μ_h^2, ν_x^2; and norms of (h, x) under (to with in a constant) those of the ground truth (h_0, x_0).

The proposed alternating gradient descent algorithm takes alternate Wirtinger gradient (of the loss function $\tilde{F}(h, x)$) steps in each of the unknowns h, and x while fixing the other; see Algorithm 1 below for the pseudo code. The Wirtinger gradients are defined as[2]

$$\nabla \tilde{F}_h := \frac{\partial \tilde{F}}{\partial \bar{h}} = \overline{\frac{\partial \tilde{F}}{\partial h}}, \text{ and } \nabla \tilde{F}_x := \frac{\partial \tilde{F}}{\partial \bar{x}} = \overline{\frac{\partial \tilde{F}}{\partial x}}. \tag{12}$$

Similar algorithms with provable recovery results appeared earlier beginning with [17], and then [18] for phase retrieval, and in [11] for blind deconvolution, however, with observation model different from (3) considered here. For example, [11] assumes that s lives in the span of the columns of a tall Gaussian matrix. We on the other hand do not make such a restrictive structural assumption, which is unfortunately rarely met in practice, and only assume that s lives in a realistic subspace, spanned by the columns of the DFT matrix F_K^* in this case. In addition, we comparatively assume much limited and structured randomness in the form of random signs of one of the convolved inputs. This random sign assumption might either be naturally satisfied or can be enforced in real time through an easy to implement signal modulation. This easy to realize modulation plus the deterministic subspace model broadens the applicability of our result in actual physical applications; for more details, see [19]. Limited randomness, however, makes the analysis and proof of our theoretical results significantly more challenging than [11]. We have to resort to advanced techniques such as generic chaining [20] to control the some of the resultant random processes that arise in the proofs [19].

---

[2]For a complex function $f(z)$, where $z = u + \iota v \in \mathbb{C}^L$, and u, $v \in \mathbb{R}^L$, the Wirtinger gradient is defined as $\frac{\partial f}{\partial \bar{z}} = \frac{1}{2} \left(\frac{\partial f}{\partial u} + \iota \frac{\partial f}{\partial v} \right)$.

Algorithm 1 Wirtinger gradient descent with a step size η
___
Input: Obtain (u_0, v_0) via Algorithm 2 below.
for $t = 1, \ldots$ **do**
  $u_t \leftarrow u_{t-1} - \eta \nabla \tilde{F}_h(u_{t-1}, v_{t-1})$
  $v_t \leftarrow v_{t-1} - \eta \nabla \tilde{F}_m(u_{t-1}, v_{t-1})$
end for
___

Finally, a suitable initialization (u_0, v_0) for Algorithm 1 is computed using Algorithm 2 below. In short, the left and right singular vectors of $\mathcal{A}^*(\hat{y})$ when projected in the set of sufficiently incoherent (measured in terms of the coherence μ, ν of the original vectors h_0, and x_0) vectors supply us with the initializers (u_0, v_0).

Algorithm 2 Initialization
___
Input: Compute $\mathcal{A}^*(\hat{y})$, and find the leading singular value d, and the corresponding left and right singular vectors $\hat{h}_0$, and $\hat{x}_0$, respectively.
Solve the following optimization programs
$u_0 \leftarrow \underset{h}{\mathrm{argmin}} \|h - \sqrt{d}\hat{h}_0\|_2$, subject to $\sqrt{Q}\|F_M h\|_\infty \leq 2\sqrt{d}\mu$,
and
$v_0 \leftarrow \underset{x}{\mathrm{argmin}} \|x - \sqrt{d}\hat{x}_0\|_2$, subject to $\sqrt{Q}\|F_K^* x\|_\infty \leq 2\sqrt{d}\nu$.
Output: (u_0, v_0).
___

2.1. Main Results

Our main result shows that given the convolution measurements (4), a *suitably initialized* Wirtinger gradient-descent Algorithm 1 converges to the true solution, i.e., $(u_t, v_t) \approx (h_0, x_0)$ under an appropriate choice of Q. To state the main theorem, we need to introduce some neighborhood sets. For vectors $h \in \mathbb{C}^M$, and $x \in \mathbb{C}^K$, we define the following sets of neighboring points of (h, x) based on either, magnitude, coherence, or the distance from the ground truth.

$$\mathcal{N}_{d_0} := \{(h, x) | \|h\|_2 \leq 2\sqrt{d_0}, \|x\|_2 \leq 2\sqrt{d_0}\}, \tag{13}$$

$$\mathcal{N}_\mu := \{(h, x) | \sqrt{Q}\|F_M h\|_\infty \leq 4\mu\sqrt{d_0}\}, \tag{14}$$

$$\mathcal{N}_\nu := \{(h, x) | \sqrt{Q}\|F_K^* x\|_\infty \leq 4\nu\sqrt{d_0}\}, \tag{15}$$

$$\mathcal{N}_\varepsilon := \{(h, x) | \|hx^* - h_0x_0^*\|_F \leq \varepsilon d_0\}. \tag{16}$$

Our main result on blind deconvolution from modulated inputs (4) is stated below.

Theorem 1 (Theorem 1 in [19]). *Fix $0 < \varepsilon \leq 1/15$. Set $s_0 = F_K^* x_0 \in \mathbb{C}^Q$; and $x_0 \in \mathbb{C}^K$, and $h_0 \in \mathbb{C}^M$ be arbitrary vectors. Assume $Q \geq M$. Let the signal coherence parameters be as defined in (8). Let r be independently generated Q-vector with standard iid Rademacher entries. We observe the Q-point circular convolutions of the random sign vectors $r \odot s_0$ with h_0, leading to observations (4) contaminated with additive noise e. Assume that the initial guess (u_0, v_0) of (h_0, x_0) belongs to $\frac{1}{\sqrt{3}}\mathcal{N}_{d_0} \cap \frac{1}{\sqrt{3}}\mathcal{N}_\mu \cap \frac{1}{\sqrt{3}}\mathcal{N}_\nu \cap \mathcal{N}_{\frac{2}{5}\varepsilon}$, and that*

$$Q \geq c\frac{\nu^2}{\varepsilon^4}(\mu^2 K + \nu^2 M)\log^4(Q), \tag{17}$$

then Algorithm 1 will create a sequence $(u_t, v_t) \in \mathcal{N}_{d_0} \cap \mathcal{N}_\mu \cap \mathcal{N}_\nu \cap \mathcal{N}_\varepsilon$, which converges geometrically to (h_0, x_0) with probability at least

$$1 - 2\exp\left(-c\varepsilon^4 Q/\mu^2\nu^4\right), \tag{18}$$
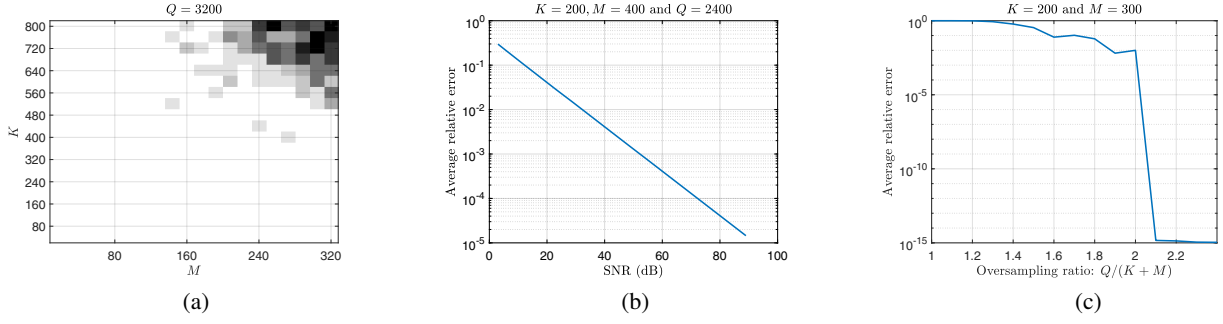
Fig. 2. *(a) Phase transition of K vs. M for fixed Q show that for modulated inputs allow recovery with in log factor sum of K, and M, (b) The performance in the presence of additive meaurement noise, and (c) The number of samples vs. the relative error.*

and there holds

$$\max\{\sin \angle(\boldsymbol{u}_t, \boldsymbol{h}_0), \sin \angle(\boldsymbol{v}_t, \boldsymbol{x}_0)\} \leq$$

$$\frac{1}{d_t}\left(\tfrac{2}{3}(1-\eta\omega)^{t/2}\varepsilon d_0 + 50\|\mathcal{A}^*(\boldsymbol{e})\|_{2\to 2}\right), \quad (19)$$

and $|d_t - d_0| \leq \frac{2}{3}(1-\eta\omega)^{t/2}\varepsilon d_0 + 50\|\mathcal{A}^*(\boldsymbol{e})\|_{2\to 2}$, where $d_t = \|\boldsymbol{u}_t\|_2\|\boldsymbol{v}_t\|_2$, $\omega > 0$, η is the fixed step size. Fix $\alpha \geq 1$. For noise $\boldsymbol{e} \sim Normal(\boldsymbol{0}, \frac{\sigma^2 d_0^2}{2Q}\boldsymbol{I}_Q) + \iota Normal(\boldsymbol{0}, \frac{\sigma^2 d_0^2}{2Q}\boldsymbol{I}_Q)$, $\|\mathcal{A}^*(\boldsymbol{e})\|_{2\to 2} \leq \frac{2\varepsilon}{50}d_0$ with probability at least $1 - O((Q)^{-\alpha})$ whenever

$$Q \geq c_\alpha \frac{\sigma^2}{\varepsilon^2}\max(M, K\log(Q))\log(Q). \quad (20)$$

Proof of the theorem above can be found in the preprint available on arxiv [19]. The above theorem claims that starting from a good enough initial guess the gradient descent algorithm converges super linearly to the ground truth. The theorem below guarantees that the required good enough initialization: $(\boldsymbol{u}_0, \boldsymbol{v}_0) \in \frac{1}{\sqrt{3}}\mathcal{N}_{d_0} \cap \frac{1}{\sqrt{3}}\mathcal{N}_\mu \cap \frac{1}{\sqrt{3}}\mathcal{N}_\nu \cap \mathcal{N}_{\frac{2}{5}\varepsilon}$ is supplied by Algorithm 2.

Theorem 2 (Theorem 2 in [19]). *The initialization obtained via Algorithm 2 satisfies $(\boldsymbol{u}_0, \boldsymbol{v}_0) \in \frac{1}{\sqrt{3}}\mathcal{N}_{d_0} \cap \frac{1}{\sqrt{3}}\mathcal{N}_\mu \cap \frac{1}{\sqrt{3}}\mathcal{N}_\nu \cap \mathcal{N}_{\frac{2}{5}\varepsilon}$, and $0.9d_0 \leq d \leq 1.1d_0$ holds with probability at least $1 - 2\exp\left(-c\varepsilon^2 Q/\mu^2\nu^4\right)$ whenever*

$$Q \geq c\frac{\nu^2}{\varepsilon^2}\left(\mu^2\nu_{\max}^2 K + \nu^2 M\right)\log^4 Q.$$

Proof of the theorem above can be found in the preprint available on arxiv [19].

2.2. Discussion

Theorem 1, and 2 together prove that randomly modulated unknown Q-vector $\boldsymbol{s}_0$, and unknown M-vector $\boldsymbol{h}_0$ can be recovered from their circular convolutions $\boldsymbol{h}_0 \circledast (\boldsymbol{r} \odot \boldsymbol{s}_0)$, under suitably large Q. We will refer to the bounds in (17), (20) as *sample complexity bounds*. Observe that the number of unknowns in the system of equations (4) is $K + M$. Combining (17), and (20), it becomes clear that number Q of measurements required for successful recovery scale with $K + M$ (within coherences, and log factors). The bound on Q above is information theoretically optimal (within log factors and coherence terms). This sample complexity result almost matches the results in [10, 11] except for an extra ν^2, and a log factor. However, the important difference is, as mentioned in the introduction, that unlike

[10–14], the inputs are not assumed to reside in Gaussian subspaces rather only have random signs.

A direct application of our main result shows that $\boldsymbol{s}$, and hence $s(t)$ can be recovered from the received signal $y(t) = (s(t) \odot r(t)) \circledast h(t)$ without knowing channel impulse response by operating the random binary waveform $r(t)$ at rate $Q \gtrsim \nu^2(\mu^2 K + \nu^2 M)\log^4 Q$, and sampling the received signal $y(t)$ at a rate Q. The coherences ν^2, and μ^2 are simply the peak values in time $\|\boldsymbol{s}\|_\infty^2$, and frequency domain $\|\boldsymbol{F}_M\boldsymbol{h}\|_\infty^2$, respectively.

3. NUMERICAL SIMULATIONS

In this section, we numerically investigate the sample complexity bounds using phase transitions. We also report stable recovery in the presence of additive measurement noise.

We present phase transitions to numerically investigate the constraints in (17), and (20) on the dimensions Q, M, and K for the gradient descent algorithm to succeed with high probability. The shade represents the probability of failure, which is computed over hundred independent experiments. For each experiment, we generate Gaussian random vectors $\boldsymbol{h}_0$, and $\boldsymbol{x}_0$, and generate $\boldsymbol{s}_0 = \boldsymbol{F}_K^*\boldsymbol{x}_0$. The synthetic measurements are then generated following the model (3). We run Algorithm 1 initialized via Algorithm 2, and classify the experiment as successful if the relative error

$$\text{Relative Error} := \frac{\|\hat{\boldsymbol{h}}\hat{\boldsymbol{x}}^* - \boldsymbol{h}_0\boldsymbol{x}_0^*\|_F}{\|\boldsymbol{h}_0\boldsymbol{x}_0^*\|_F} \quad (21)$$

is below 10^{-2}. The probability of success at each point is computed over hundred such independent experiments.

Phase diagram in Figure 2(a) investigate successful recovery for a fixed $Q = 3200$, and under varying K, and M. The shade in the phase transition represents probability of failure. For example, in the phase transition, successful recovery occurs almost always when the measurements are a factor of 2.8 above the number of unknowns, that is, $Q \geq 2.8(K + M)$.

Noise performance of the algorithm is depicted in Figure 2(b). Additive Gaussian noise $\boldsymbol{e}$ is added in the measurements as in (4). As before, we synthetically generate $\boldsymbol{h}_0$, and $\boldsymbol{x}_0$ as Gaussian vectors. We plot (second) relative error (log scale) in (21) of the recovered vectors $\hat{\boldsymbol{h}}$, and $\hat{\boldsymbol{x}}$ averaged over hundred independent experiments vs. SNR $:= 10\log_{10}\left(\|\boldsymbol{h}_0\boldsymbol{x}_0^*\|_F^2/\|\boldsymbol{e}\|_2^2\right)$, and Figure 2(c) shows average relative error (log scale) vs. oversampling ratio $:= Q/(K + M)$ under no noise. Oversampling ratio is a factor by which the number (Q measurements exceed the number $K + M$ of unknowns. The second plot shows that the relative error degrades *gracefully* by reducing SNR, and the third shows that relative error almost reduces to zero when the oversampling ratio exceeds 2.1.

4. REFERENCES

[1] Andrea Goldsmith, *Wireless communications*, Cambridge university press, 2005.

[2] R. Michael Buehrer, *Code Division Multiple Access (CDMA)*, Synthesis Lectures on Communications. Morgan & Claypool Publishers, 2006.

[3] Eric Moulines, Pierre Duhamel, J-F Cardoso, and Sylvie Mayrargue, "Subspace methods for the blind identification of multichannel fir filters," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 516–525, 1995.

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

[5] Jason N Laska, Sami Kirolos, Marco F Duarte, Tamer S Ragheb, Richard G Baraniuk, and Yehia Massoud, "Theory and implementation of an analog-to-information converter using random demodulation," in *IEEE Int'l Symposium Circuits Syst. ISCAS*. IEEE, 2007, pp. 1959–1962.

[6] J. Tropp and J. Laska and M. Duarte and J. Romberg, and R. Baraniuk, "Beyond nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Trans. Inform. Theory*, vol. 56, no. 1, pp. 520–544, 2010.

[7] Ali Ahmed and Justin Romberg, "Compressive multiplexing of correlated signals," *IEEE Trans. Inform. Th.*, vol. 1, pp. 479–498, 2015.

[8] Ali Ahmed and Justin Romberg, "Compressive sampling of ensembles of correlated signals," *arXiv preprint arXiv:1501.06654*, 2015.

[9] Ali Ahmed, Augustin Cosse, and Laurent Demanet, "A convex approach to blind deconvolution with diverse inputs," in *IEEE 6th Int'l Workshop Comput. Adv. Multi-Sensor Adaptive Process.(CAMSAP)*. IEEE, 2015, pp. 5–8.

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

[11] Xiaodong Li, Shuyang Ling, Thomas Strohmer, and Ke Wei, "Rapid, robust, and reliable blind deconvolution via nonconvex optimization," *Appl. Comput. Harmonic Anal.*, 2018.

[12] Kiryung Lee, Felix Krahmer, and Justin Romberg, "Spectral methods for passive imaging: Non-asymptotic performance and robustness," *arXiv preprint arXiv:1708.04343*, 2017.

[13] Ali Ahmed and Laurent Demanet, "Leveraging diversity and sparsity in blind deconvolution," *IEEE Trans. Inform. Theory*, vol. 64, no. 6, pp. 3975–4000, 2018.

[14] Ali Ahmed, "A convex approach to blind MIMO communications," *IEEE Wireless Commun. Lett.*, 2018.

[15] E. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.

[16] Emmanuel Candes and Justin Romberg, "Sparsity and incoherence in compressive sampling," *Inverse problems*, vol. 23, no. 3, pp. 969, 2007.

[17] Prateek Jain, Praneeth Netrapalli, and Sujay Sanghavi, "Low-rank matrix completion using alternating minimization," in *Proc. forty-fifth annual ACM symposium Theory comput.* ACM, 2013, pp. 665–674.

[18] Emmanuel J Candes, Xiaodong Li, and Mahdi Soltanolkotabi, "Phase retrieval via wirtinger flow: Theory and algorithms," *IEEE Trans. Inform. Theory*, vol. 61, no. 4, pp. 1985–2007, 2015.

[19] Ali Ahmed, "Blind deconvolution using modulated inputs," *arxiv preprint*, 2018.

[20] M Talagrand, "The generic chaining. springer monographs in mathematics," 2005.