Nonconvex Demixing from Bilinear Measurements

Jialin Dong and Yuanming Shi
School of Information Science and Technology
ShanghaiTech University, Shanghai 201210, China
Email: {dongjl, shiym}@shanghaitech.edu.cn

***Abstract*—We consider the problem of demixing a sequence of source signals from the sum of bilinear measurements. It is a generalized mathematical model of blind demixing with deconvolution, which has wide applications in communication, image processing and dictionary learning, etc. However, state-of-art algorithms for blind demixing either fail to scale to large problem sizes or require proper regularization with tedious algorithmic parameters for optimality guarantees. To address the limitations of exiting methods, we propose a provable nonconvex demixing procedure via Wirtinger flow, much like vanilla gradient descent, to harness the benefits of regularization free, fast convergence rate, and optimality guarantees. This is achieved by exploiting the benign geometry of blind demixing, thereby revealing that Wirtinger flow enforces the iterates in the region of strong convexity and qualified level of smoothness.**

I. Introduction

Demixing a sequence of source signals from the sum of bilinear measurements provides a generalized mathematical modeling framework for blind demixing with deconvolution [1], [2]. It arises in fields as diverse as communication, imaging, statistics, and machine learning, etc. In particular, similar to the idea in the work [3], [4] which support low-latency communication only based on connectivity information, the blind demixing approach turns out to be a promising approach that eludes the channel estimation information to reduce the channel signaling overhead, thereby supporting low-latency communication with unknown channels for Internet-of-things applications [2]. Although blind demixing can be regarded as a variant of blind deconvolution [5], which extends the problem of "single-source" to the "multi-source" setting, it is nontrivial to accomplish the extension. The main reason is that the "incoherence" between different sources brings unique challenges to develop effective algorithms for blind demixing with theoretical guarantees [1], [6]. Furthermore, the bilinear measurements in the blind demixing problem hamper the extension of the results developed in [7] with linear measurements for demixing.

A growing body of recent work has has been contributed to the blind demixing problem from both empirical and theoretical points of view. Specifically, by lifting the original bilinear model into the the linear model with multiple rank-one matrices, followed by convex relaxation, the semidefinite programming approach was developed in [1]. Despite attractive statistical guarantee, such convex relaxation method fails to scale to large problem sizes. To address the scaling issue, the Riemannian optimization algorithm was developed in [2] by exploiting the manifold geometry of fixed-rank matrices. However, with complicated iterative strategies in the Riemannian optimization algorithms, it is challenging to provide statistical analysis. Ling *et al.* in [6] thus proposed an efficient regularized gradient descent algorithm to optimize the nonconvex loss function with a regularization that account for incoherence. Although the regularization procedure in [6] provides appealing computational and optimality properties, it usually introduces tedious algorithmic parameters that require to be carefully tuned. To address the limitations of the state-of-art algorithms, efficient and provable methods are vital to solve blind demixing problem.

Wirtinger flow algorithm consisting of spectral initialization and vanilla gradient descent updates turns out to be a good candidate for the high-dimensional statistical estimation problem. In particular, the optimality guarantee for phase retrieval was provided in [8] via Wirtinger flow, i.e., vanilla gradient descent without regularization. However, the theory in [9] only ensures that iterates of such algorithm remain in ℓ_2 ball, in which the step size is thus chosen conservative based on the generic optimization theory. The statistical and computational efficiency was further improved in [9] via the truncated Wirtinger flow by carefully controlling search directions, much like regularized gradient descent. To harness all benefits of regularization free, fast convergence rate, and optimality guarantees, Ma *et al.* in [10] has recently uncovered that the Wirtinger flow algorithm implicitly enforces iterates within the intersection between ℓ_2 ball and the incoherence region, i.e., the region of incoherence and contraction, for the problems of phase retrieval, low-rank matrix completion, and blind deconvolution. Furthermore, by exploiting the local geometry in such region, i.e., strong convexity and qualified level of smoothness, the step size of the iterative algorithm can be chosen more aggressively, yielding fast convergence rates.

Inspired by the pioneering paper [10], to address the scaling issue of the convex method [1] and the tedious regularization issue of the nonconvex method [6], we propose to solve the blind demixing problem via Wirtinger flow with optimality guarantees, regularization free and fast iterative procedure. This is achieved by exploiting the benign geometry of the blind demixing problem. Specifically, we shall characterize the local geometry of blind demixing in the region of incoherence and contraction, i.e., strong convexity and qualified level of smoothness, followed by revealing that iterates of Wirtinger

This work was supported in part by the National Nature Science Foundation of China under Grant 61601290, and the Shanghai Sailing Program under Grant 16YF1407700.

978-1-5386-4780-6/18/$31.00©2018 IEEE

flow always pertain in such region. Note that it is nontrivial to finish the theoretical derivations since the "incoherence" between multiple sources leads to distortion to the statistical property in the "single-source" scenario of blind deconvolution. The blind demixing case is thus much more complicated and extra technical efforts are involved.

II. Problem Formulation

In this section, we present the blind demixing problem. As the problem is highly intractable without any further structural assumptions, the coupled signals is designed to belong to known subspaces [1], [6].

Let $\boldsymbol{A}^*$ denote the transpose of matrix $\boldsymbol{A}$. Suppose we have m bilinear measurements y_j's, which are represented in the frequency domain as

$$y_j = \sum_{i=1}^{s} \boldsymbol{b}_j^* \boldsymbol{h}_i^\natural \boldsymbol{x}_i^{\natural *} \boldsymbol{a}_{ij}, \ 1 \leq j \leq m, \tag{1}$$

where $\boldsymbol{a}_{ij} \in \mathbb{C}^K$ and $\boldsymbol{b}_j \in \mathbb{C}^K$ are known design vectors. Specifically, each $\boldsymbol{a}_{ij}$ follows an i.i.d. complex Gaussian distribution, i.e., $\boldsymbol{a}_{ij} \sim \mathcal{N}(0, \frac{1}{2}\boldsymbol{I}_K) + i\mathcal{N}(0, \frac{1}{2}\boldsymbol{I}_K)$ and the first K columns of the unitary discrete Fourier transform (DFT) matrix $\boldsymbol{F} \in \mathbb{C}^{m\times m}$ with $\boldsymbol{F}\boldsymbol{F}^* = \boldsymbol{I}_m$ form the matrix $\boldsymbol{B} := [\boldsymbol{b}_1, \cdots, \boldsymbol{b}_m]^* \in \mathbb{C}^{m\times K}$. Based on the above bilinear model, our goal is to simultaneously recover the underlying signals $\boldsymbol{h}_i \in \mathbb{C}^K$'s and $\boldsymbol{x}_i \in \mathbb{C}^K$'s by solving the following blind demixing problem [1], [2], [6]

$$\underset{\boldsymbol{h}_i, \boldsymbol{x}_i, i=1,\cdots,s}{\text{minimize}} \ f(\boldsymbol{h}, \boldsymbol{x}) := \sum_{j=1}^{m} \left| \sum_{i=1}^{s} \boldsymbol{b}_j^* \boldsymbol{h}_i \boldsymbol{x}_i^* \boldsymbol{a}_{ij} - \boldsymbol{y}_j \right|^2. \tag{2}$$

To simplify the presentation, we denote $f(\boldsymbol{z}) := f(\boldsymbol{h}, \boldsymbol{x})$ with

$$\boldsymbol{z} = [\boldsymbol{z}_1^* \cdots \boldsymbol{z}_s^*]^* \in \mathbb{C}^{2sK}, \text{ where } \boldsymbol{z}_i = [\boldsymbol{h}_i^* \ \boldsymbol{x}_i^*]^* \in \mathbb{C}^{2K}. \tag{3}$$

We further define the discrepancy between $\boldsymbol{z}_i$'s and the ground truth $\boldsymbol{z}_i^\natural$'s as the following distance function

$$\operatorname{dist}(\boldsymbol{z}, \boldsymbol{z}^\natural) := \min_{\alpha_i \in \mathbb{C}} \left(\sum_{i=1}^{s} \left\| \frac{1}{\overline{\alpha}_i} \boldsymbol{h}_i - \boldsymbol{h}_i^\natural \right\|_2^2 + \left\| \alpha_i \boldsymbol{x}_i - \boldsymbol{x}_i^\natural \right\|_2^2 \right)^{1/2} \tag{4}$$

where each α_i, $i = 1, \cdots, s$, is the alignment parameter.

III. Main Results

In this section, we shall present the Wirtinger flow algorithm with statistical analysis for the blind demixing problem.

A. Wirtinger Flow Algorithm

The Wirtinger flow algorithm [8] is a two-stage approach consisting of spectral initialization and vanilla gradient descent updates. Specifically, the gradient step in the second step of Wirtinger flow is characterized by the notion of Wirtinger derivatives [8], i.e., the derivatives of real valued functions over complex variables. For each $i = 1, \cdots, s$, $\nabla_{\boldsymbol{h}_i} f(\boldsymbol{h}, \boldsymbol{x})$ and $\nabla_{\boldsymbol{x}_i} f(\boldsymbol{h}, \boldsymbol{x})$ denote the Wirtinger gradient of $f(\boldsymbol{z})$ with respect to $\boldsymbol{h}_i$ and $\boldsymbol{x}_i$ as follows:

$$\nabla_{\boldsymbol{h}_i} f(\boldsymbol{z}) = \sum_{j=1}^{m} \left(\sum_{k=1}^{s} \boldsymbol{b}_j^* \boldsymbol{h}_k \boldsymbol{x}_k^* \boldsymbol{a}_{kj} - y_j \right) \boldsymbol{b}_j \boldsymbol{a}_{ij}^* \boldsymbol{x}_i, \tag{5a}$$

$$\nabla_{\boldsymbol{x}_i} f(\boldsymbol{z}) = \sum_{j=1}^{m} \overline{\left(\sum_{k=1}^{s} \boldsymbol{b}_j^* \boldsymbol{h}_k \boldsymbol{x}_k^* \boldsymbol{a}_{kj} - y_j \right)} \boldsymbol{a}_{ij} \boldsymbol{b}_j^* \boldsymbol{h}_i. \tag{5b}$$

The Wirtinger flow for blind demixing problem is presented in

Algorithm 1: Wirtinger flow for blind demixing

Given: $\{\boldsymbol{a}_{ij}\}_{1\leq i\leq s, 1\leq j\leq m}$, $\{\boldsymbol{b}_j\}_{1\leq j\leq m}$, and $\{y_j\}_{1\leq j\leq m}$.

Spectral Initialize:
for all $i = 1, \cdots, s$ **do in parallel**
  Let $\sigma_1(\boldsymbol{M}_i)$, $\check{\boldsymbol{h}}_i^0$ and $\tilde{\boldsymbol{x}}_i^0$ be the leading singular value, left singular vector and right singular vector of matrix $\boldsymbol{M}_i := \sum_{j=1}^{m} y_j \boldsymbol{b}_j \boldsymbol{a}_{ij}^*$, respectively.
  Set $\boldsymbol{h}_i^0 = \sqrt{\sigma_1(\boldsymbol{M}_i)} \check{\boldsymbol{h}}_i^0$ and $\boldsymbol{x}_i^0 = \sqrt{\sigma_1(\boldsymbol{M}_i)} \tilde{\boldsymbol{x}}_i^0$.
end for
for all $t = 1, \cdots, T$ **do**
  for all $i = 1, \cdots, s$ **do in parallel**
    $\begin{bmatrix} \boldsymbol{h}_i^{t+1} \\ \boldsymbol{x}_i^{t+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{h}_i^{t} \\ \boldsymbol{x}_i^{t} \end{bmatrix} - \eta \begin{bmatrix} \frac{1}{\|\boldsymbol{x}_i^t\|_2^2} \nabla_{\boldsymbol{h}} f(\boldsymbol{h}^t, \boldsymbol{x}^t) \\ \frac{1}{\|\boldsymbol{h}_i^t\|_2^2} \nabla_{\boldsymbol{x}} f(\boldsymbol{h}^t, \boldsymbol{x}^t) \end{bmatrix}$
  end for
end for

Algorithm 1, in which T is the maximum number of iterates and η is the step size.

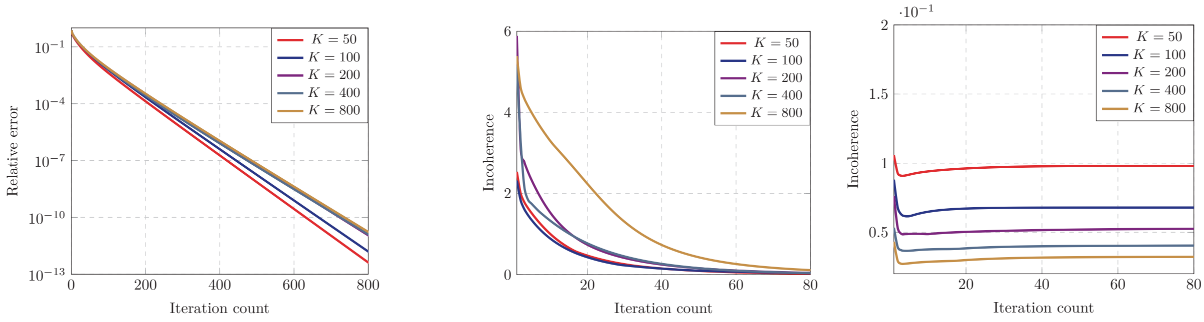
We now provide some numerical evidence by testing the performance of the Wirtinger flow algorithm for blind demixing (2). Specifically, for each $K \in \{50, 100, 200, 400, 800\}$, $s = 10$ and $m = 10K$, generate the design vectors $\boldsymbol{a}_{ij}$ and $\boldsymbol{b}_j$ for each $1 \leq i \leq s, 1 \leq j \leq m$, according to the descriptions in Section II. The underlying signals $\boldsymbol{h}_i^\natural, \boldsymbol{x}_i^\natural \in \mathbb{C}^K$, $1 \leq i \leq s$, are generated as random vectors with unit norm. With the chosen step size as $\eta = 0.1$ in all settings, Fig. 1(a) shows the relative error $\sum_{i=1}^{s} \|\boldsymbol{h}_i^t \boldsymbol{x}_i^{t*} - \boldsymbol{h}_i^\natural \boldsymbol{x}_i^{\natural *}\| / \sum_{i=1}^{s} \|\boldsymbol{h}_i^\natural \boldsymbol{x}_i^{\natural *}\|$ vs. the iteration count, where $\|\cdot\|_F$ denotes the Frobenius norm. Consequently, with constant step size, Wirtinger flow enjoys extraordinary linear convergence rate which rarely changes as the problem sizes vary. The main purpose of this paper is to theoretically analyze the above empirical observations.

B. Theoretical Results

Before stating main theorem, we need to introduce the incoherence parameter [6], which characterizes the incoherence between $\boldsymbol{b}_j$ and $\boldsymbol{h}_i$ for $1 \leq i \leq s, 1 \leq j \leq m$.

Definition 1 (Incoherence for blind deconvolution)**.** *Let the incoherence parameter μ be the smallest number such that* $\max_{1\leq i\leq s, 1\leq j\leq m} |\boldsymbol{b}_j \boldsymbol{h}_i^\natural| \leq \frac{\mu}{\sqrt{m}} \|\boldsymbol{h}_i^\natural\|_2$.

The incoherence between $\boldsymbol{a}_{ij}$ and $\boldsymbol{x}_i$ and the incoherence between $\boldsymbol{b}_j$ and $\boldsymbol{h}_i$ for $1 \leq i \leq s, 1 \leq j \leq m$ specify the smoothness of the loss function (2). Within the region of incoherence and contraction (in Section IV-A) that enjoys the



(a) (b) (c)

Fig. 1. Numerical experiments.

qualified level of smoothness, the step size for iterative refinement procedure can be chosen more aggressively according to generic optimization theory [10]. Based on the definition of coherence, our theory shall show that iterates of Algorithm 1 retains in the region of coherence and contraction endowed with strong convexity and the qualified level of smoothness, provided near-optimal sample complexity.

Using the similar arithmetic operators in [10], we present the main theorem. Without loss of generality, we assume $\|\boldsymbol{h}_i^\natural\|_2 = \|\boldsymbol{x}_i^\natural\|_2 = 1$ for $i = 1, \cdots, s$.

Theorem 1. *Suppose the step size $\eta > 0$ is taken to be some sufficiently small constant, i.e., $\eta \asymp s^{-1}$, then the iterates in Algorithm 1 satisfy*

$$\operatorname{dist}(\boldsymbol{z}^t, \boldsymbol{z}^\natural) \le C_1\left(1 - \frac{\eta}{16}\right)^t \frac{1}{\log^2 m}, \tag{6a}$$

$$\max_{1\le i\le s, 1\le j\le m} \left| \boldsymbol{a}_{ij}\left(\alpha_i^t \boldsymbol{x}_i^t - \boldsymbol{x}_i^\natural\right)\right| \le C_3 \frac{1}{\sqrt{s}\log^{3/2} m}, \tag{6b}$$

$$\max_{1\le i\le s, 1\le j\le m} \left| \boldsymbol{b}_j^* \frac{1}{\overline{\alpha_i^t}} \boldsymbol{h}_i^t \right| \le C_4 \frac{\mu}{\sqrt{m}} \log^2 m, \tag{6c}$$

for all $t \le 0$, with probability at least $1 - c_1 m^{-4} - c_1 m e^{-c_2 K}$ if the number of measurements $m \ge C\mu^2 s^2 K \log^8 m$, holds for some constants $c_1, c_2, C_1, C_3, C_4 > 0$ and sufficiently large constant $C > 0$. Here, we denote α_i^t for $i = 1, \cdots, s$ as the alignment parameter such that

$$\alpha_i^t := \arg\min_{\alpha_i \in \mathbb{C}} \left\| \frac{1}{\overline{\alpha_i}} \boldsymbol{h}_i^t - \boldsymbol{h}_i^\natural \right\|_2^2 + \left\| \alpha_i \boldsymbol{x}^t - \boldsymbol{x}_i^\natural \right\|_2^2.$$

Theorem 1 endorses the empirical results shown in Figure 1(a). For further illustration, we plot the incoherence measure $\max_{1\le i\le s, 1\le j\le m} |\boldsymbol{a}_{ij}^*(\tilde{\boldsymbol{x}}_i^t - \boldsymbol{x}_i^\natural)|$ (in Fig. 1(b)) and $\max_{1\le i\le s, 1\le j\le m} |\boldsymbol{b}_j^* \tilde{\boldsymbol{h}}_i^t|$ (in Fig. 1(c)) of the gradient iterates vs. iteration count, under the same settings as Fig. 1(a). We observe that both incoherence measures remain bounded by befitting values for all iterations.

IV. Trajectory Analysis for Blind Demixing

In this section, we prove the main theorem via trajectory analysis for blind demixing via Wirtinger flow. We shall reveal that iterates of Wirtinger flow, i.e., Algorithm 1, stay in the region of incoherence and contraction by exploiting the local geometry of blind demixing.

A. Characterizing Local Geometry in the Region of Incoherence and Contraction

We first introduce the notation of Wirtinger Hessian. Specifically, let $\overline{\boldsymbol{A}}$ denote the entrywise conjugate of matrix $\boldsymbol{A}$. The Wirtinger Hessian of $f(\boldsymbol{z})$ with respect to $\boldsymbol{z}_i$ can be written as

$$\nabla^2_{\boldsymbol{z}_i} f := \begin{bmatrix} \boldsymbol{C} & \boldsymbol{D} \\ \boldsymbol{D}^* & \overline{\boldsymbol{C}} \end{bmatrix}, \tag{7}$$

where $\boldsymbol{C} := \frac{\partial}{\partial \boldsymbol{z}_i}\left(\frac{\partial f}{\partial \boldsymbol{z}_i}\right)^*$, $\boldsymbol{D} := \frac{\partial}{\partial \overline{\boldsymbol{z}}_i}\left(\frac{\partial f}{\partial \boldsymbol{z}_i}\right)^*$. The Wirtinger Hessian of $f(\boldsymbol{z})$ with respect to $\boldsymbol{z}$ is thus represented as $\nabla^2 f(\boldsymbol{z}) := \operatorname{diag}(\{\nabla^2_{\boldsymbol{z}_i} f\}_{i=1}^s)$, where the operation $\operatorname{diag}(\{\boldsymbol{A}\}_{i=1}^s)$ generates a block diagonal matrix with the diagonal elements being matrices $\boldsymbol{A}_1, \cdots, \boldsymbol{A}_s$. In addition, we say $(\boldsymbol{h}_i, \boldsymbol{x}_i)$ is aligned with $(\boldsymbol{h}_i', \boldsymbol{x}_i')$, if the following condition is satisfied

$$\|\boldsymbol{h}_i - \boldsymbol{h}_i'\|_2^2 + \|\boldsymbol{x}_i - \boldsymbol{x}_i'\|_2^2 = \min_{\alpha_i \in \mathbb{C}} \left\{ \left\| \frac{1}{\overline{\alpha_i}} \boldsymbol{h}_i - \boldsymbol{h}_i' \right\|_2^2 + \|\alpha_i \boldsymbol{x}_i - \boldsymbol{x}_i'\|_2^2 \right\}. \tag{8}$$

The first step for trajectory analysis is to characterize the local geometry in the region of incoherence and contraction (RIC), where the objective function $f(\boldsymbol{z})$ enjoys restricted strong convexity and qualified level of smoothness. This is precisely presented in the following lemma.

Lemma 1. *(Restricted strong convexity and smoothness for blind demixing). Let $\delta > 0$ be a sufficiently small constant. If the number of measurements satisfies $m \gg \mu^2 s^2 K \log^5 m$, then with probability at least $1 - O(m^{-9})$, the Wirtinger Hessian $\nabla^2 f(\boldsymbol{z})$ obeys $\boldsymbol{u}^* \left[\boldsymbol{D}\nabla^2 f(\boldsymbol{z}) + \nabla^2 f(\boldsymbol{z})\boldsymbol{D}\right] \boldsymbol{u} \ge \|\boldsymbol{u}\|_2^2$ and $\left\|\nabla^2 f(\boldsymbol{z})\right\| \le 2 + s$ simultaneously for all*

$$\boldsymbol{u} = \begin{bmatrix} \boldsymbol{u}_1 \\ \vdots \\ \boldsymbol{u}_s \end{bmatrix} \quad \textit{with } \boldsymbol{u}_i = \begin{bmatrix} \boldsymbol{h}_i - \boldsymbol{h}_i' \\ \boldsymbol{x}_i - \boldsymbol{x}_i' \\ \overline{\boldsymbol{h}_i - \boldsymbol{h}_i'} \\ \overline{\boldsymbol{x}_i - \boldsymbol{x}_i'} \end{bmatrix},$$

and $\boldsymbol{D} = \mathrm{diag}\left(\{\boldsymbol{W}_i\}_{i=1}^s\right)$
with $\boldsymbol{W}_i = \mathrm{diag}\left(\left[\overline{\beta}_{i1}\boldsymbol{I}_K \ \overline{\beta}_{i2}\boldsymbol{I}_K \ \overline{\beta}_{i1}\boldsymbol{I}_K \ \overline{\beta}_{i2}\boldsymbol{I}_K\right]^*\right)$,

where $\boldsymbol{z}$ *satisfies*

$$\max_{1\leq i\leq s} \max\left\{\|\boldsymbol{h}_i - \boldsymbol{h}_i^\natural\|_2, \|\boldsymbol{x}_i - \boldsymbol{x}_i^\natural\|_2\right\} \leq \frac{\delta}{\sqrt{s}}, \tag{9a}$$

$$\max_{1\leq i\leq s, 1\leq j\leq m} |\boldsymbol{a}_{ij}^*\left(\boldsymbol{x}_i - \boldsymbol{x}_i^\natural\right)| \leq \frac{2C_3}{\sqrt{s}\log^{3/2} m}, \tag{9b}$$

$$\max_{1\leq i\leq s, 1\leq j\leq m} |\boldsymbol{b}_j^*\boldsymbol{h}_i| \leq \frac{2C_4\mu}{\sqrt{m}}\log^2 m; \tag{9c}$$

$(\boldsymbol{h}_i, \boldsymbol{x}_i)$ is aligned with $(\boldsymbol{h}_i', \boldsymbol{x}_i')$, and they hold that $\max\{\|\boldsymbol{h}_i - \boldsymbol{h}_i^\natural\|_2, \|\boldsymbol{h}_i' - \boldsymbol{h}_i^\natural\|_2, \|\boldsymbol{x}_i - \boldsymbol{x}_i^\natural\|_2, \|\boldsymbol{x}_i' - \boldsymbol{x}_i^\natural\|_2\} \leq \frac{\delta}{\sqrt{s}}$, *for* $i = 1, \cdots, s$ *and* $\boldsymbol{W}_i$*'s satisfy that for* $\beta_1, \beta_2 \in \mathbb{R}$,

$$\max_{1\leq i\leq s} \max\left\{|\beta_{i1} - 1|, |\beta_{i2} - 1|\right\} \leq \frac{\delta}{\sqrt{s}}. \tag{10}$$

Therein, $C_3, C_4 \geq 0$ *are numerical constants.*

Condition (9) identifies the local geometry of the blind demixing. Specifically, (9a) identifies a neighborhood close to the ground truth in ℓ_2-norm. In addition, (9b) and (9a) specify the incoherence region with respect to the vectors $\boldsymbol{a}_{ij}$ and $\boldsymbol{b}_j$ for $1 \leq i \leq s, 1 \leq j \leq m$, respectively. Based on the strong convexity and qualified level of smoothness in such region, more aggressive step size can be chosen.

Remark 1. *For the proof of Lemma 1, extension operation is required due to multiple sources in blind demixing. Furthermore, the sum of multiple "incoherence" signals in (5a) and (5b) calls for new statistical guarantee for the spectral norm of random matrices over an "incoherence" region, which is compared to Lemma 59 in [10] for blind deconvolution with a single source.*

Based on the local geometry in the region of incoherence and contraction, we further establish contraction of the error measured by the distance function (4).

Lemma 2. *Suppose the number of measurements satisfies* $m \gg \mu^2 s^2 K \log^5 m$ *and the step size* $\eta > 0$ *is some sufficiently small constant, e.g.,* $\eta \asymp s^{-1}$*. Then with probability at least* $1 - O(m^{-9})$, $\mathrm{dist}(\boldsymbol{z}^{t+1}, \boldsymbol{z}^\natural) \leq (1 - \eta/16)\mathrm{dist}(\boldsymbol{z}^t, \boldsymbol{z}^\natural)$, *provided that*

$$\mathrm{dist}(\boldsymbol{z}^t, \boldsymbol{z}^\natural) \leq \xi, \tag{11a}$$

$$\max_{1\leq i\leq s, 1\leq j\leq m} \left|\boldsymbol{a}_{ij}^*\left(\tilde{\boldsymbol{x}}_i^t - \boldsymbol{x}_i^\natural\right)\right| \leq \frac{2C_3}{\sqrt{s}\log^{3/2} m}, \tag{11b}$$

$$\max_{1\leq i\leq s, 1\leq j\leq m} \left|\boldsymbol{b}_j^*\tilde{\boldsymbol{h}}_i^t\right| \leq \frac{2C_4\mu}{\sqrt{m}}\log^2 m, \tag{11c}$$

for some constants $C_3, C_4 > 0$ *and the constant* $\xi > 0$ *being sufficiently small. Here* $\tilde{\boldsymbol{h}}_i^t$ *and* $\tilde{\boldsymbol{x}}_i^t$ *are defined as* $\tilde{\boldsymbol{h}}_i^t = \frac{1}{\overline{\alpha_i^t}}\boldsymbol{h}_i^t$ *and* $\tilde{\boldsymbol{x}}_i^t = \alpha_i^t \boldsymbol{x}_i^t$ *for* $i = 1, \cdots, s$.

As a result, if $\boldsymbol{z}^t$ satisfies the condition (11) for all $0 \leq t \leq T = O(m^5)$, then there is

$$\mathrm{dist}(\boldsymbol{z}^t, \boldsymbol{z}^\natural) \leq \rho^t \mathrm{dist}(\boldsymbol{z}^0, \boldsymbol{z}^\natural) \leq \rho^t c_1, \ 0 < t \leq T, \tag{12}$$

with probability at least $1 - O(m^{-5})$, where $\rho := 1 - \eta/16$. Similar to the case of blind deconvolution [10], Theorem 1 holds true as soon as the condition (11) holds for all iterates $0 \leq t \leq T$. In the following, we thus focus on establishing the condition (11) for all iterates $0 \leq t \leq T$.

B. Establishing Iterates in the Region of Incoherence and Contraction

Before moving to establishing iterates in the region of incoherence and contraction, we first introduce the leave-one-out sequences, denoted by $\{\boldsymbol{h}_i^{t,(l)}, \boldsymbol{x}_i^{t,(l)}\}_{t\geq 0}$ for each $1 \leq i \leq s$, $1 \leq l \leq m$. In particular, $\{\boldsymbol{h}_i^{t,(l)}, \boldsymbol{x}_i^{t,(l)}\}_{t\geq 0}$ is the gradient iterates operating on the loss function $f^{(l)}(\boldsymbol{h}, \boldsymbol{x}) := \sum_{j:j\neq l} |\sum_{i=1}^s \boldsymbol{b}_j^*\boldsymbol{h}_i\boldsymbol{x}_i^*\boldsymbol{a}_{ij} - \boldsymbol{y}_j|^2$. For simplicity, we denote $\boldsymbol{z}^{t,(l)} = [\boldsymbol{z}_1^{t,(l)*} \cdots \boldsymbol{z}_s^{t,(l)*}]^*$ where $\boldsymbol{z}_i^{t,(l)} = [\boldsymbol{h}_i^{t,(l)*} \ \boldsymbol{x}_i^{t,(l)*}]^*$ and $f\left(\boldsymbol{z}^{t,(l)}\right) := f^{(l)}\left(\boldsymbol{h}, \boldsymbol{x}\right)$. We further define the alignment parameters $\alpha_i^{t,(l)}$, signals $\tilde{\boldsymbol{h}}_i^{t,(l)}$ and $\tilde{\boldsymbol{x}}_i^{t,(l)}$ in the context of leave-one-out sequence, which are similar to the ones in [10].

We continue the proof by induction. For brief, with $\tilde{\boldsymbol{z}}^t = [\tilde{\boldsymbol{h}}_i^{t*} \ \tilde{\boldsymbol{x}}_i^{t*}]^*$, the set of induction hypotheses of local geometry is listed in the following,

$$\mathrm{dist}(\boldsymbol{z}^t, \boldsymbol{z}^\natural) \leq C_1\frac{1}{\log^2 m}, \tag{13a}$$

$$\mathrm{dist}(\boldsymbol{z}^{t,(l)}, \tilde{\boldsymbol{z}}^t) \leq C_2\frac{s\mu}{\sqrt{m}}\sqrt{\frac{\mu^2 K\log^9 m}{m}}, \tag{13b}$$

$$\max_{1\leq i\leq s, 1\leq j\leq m} \left|\boldsymbol{a}_{il}\left(\tilde{\boldsymbol{x}}_i^t - \boldsymbol{x}_i^\natural\right)\right| \leq C_3\frac{1}{\sqrt{s}\log^{3/2} m}, \tag{13c}$$

$$\max_{1\leq i\leq s, 1\leq j\leq m} \left|\boldsymbol{b}_l^*\tilde{\boldsymbol{h}}_i^t\right| \leq C_4\frac{\mu}{\sqrt{m}}\log^2 m, \tag{13d}$$

where C_1, C_3 are some sufficiently small constants, while C_2, C_4 are some sufficiently large constants. We aim to specify that the induction hypotheses (13) hold for $(t+1)$-th iteration with high probability, if these hypotheses hold up to the t-th iteration. Since (13a) has been identified in (11a), we begin with the hypothesis (13b) in the following lemma.

Lemma 3. *Suppose the number of measurements satisfies* $m \gg \mu^2 s^{3/2} K \log^{13/2} m$ *and the step size* $\eta > 0$ *is some sufficiently small constant, e.g.,* $\eta \asymp s^{-1}$*. Under the hypotheses (13) for the t-th iteration, one has*

$$\mathrm{dist}(\boldsymbol{z}^{t+1,(l)}, \tilde{\boldsymbol{z}}^{t+1}) \leq C_2\frac{s\mu}{\sqrt{m}}\sqrt{\frac{\mu^2 K\log^9 m}{m}}, \tag{14}$$

$$\max_{1\leq l\leq m} \left\|\tilde{\boldsymbol{z}}^{t+1,(l)}, \tilde{\boldsymbol{z}}^{t+1}\right\|_2 \lesssim C_2\frac{s\mu}{\sqrt{m}}\sqrt{\frac{\mu^2 K\log^9 m}{m}}, \tag{15}$$

with probability at least $1 - O(m^{-9})$.

Before proceeding to the hypothesis (13c), let us show the incoherence of the leave-one-out iterate $\boldsymbol{x}_i^{t+1,(l)}$ with respect to $\boldsymbol{a}_{il}$ for all $1 \leq i \leq s$. Based on the triangle inequality, Lemma 2 and Lemma 3, it follows that $\|\tilde{\boldsymbol{x}}_i^{t+1,(l)} - \boldsymbol{x}_i^\natural\|_2 \leq 2C_1/\sqrt{s}\log^2 m$, provided that the sample complexity $m \gg \mu^2\sqrt{sK}\log^{13/2} m$. Using the above inequality, the standard

Gaussian concentration inequality in [10] and the statistical independence, it follows that

$$\max_{1\le i\le s, 1\le l\le m} \left|\boldsymbol{a}_{il}\left(\tilde{\boldsymbol{x}}_i^{t+1,(l)}-\boldsymbol{x}_i^\natural\right)\right| \le C_3 \frac{1}{\log^{3/2} m} \tag{16}$$

with probability exceeding $1-O(m^{-9})$. Based on the inequality (16), the incoherence of the leave-one-out iterate $\boldsymbol{x}_i^{t+1,(l)}$ with respect to $\boldsymbol{a}_{il}$ further follows from Cauchy-Schwarz, Lemma 3 and some concentration results provided in [10]: $\max_{1\le i\le s, 1\le l\le m} \left|\boldsymbol{a}_{il}\left(\tilde{\boldsymbol{x}}_i^{t+1}-\boldsymbol{x}_i^\natural\right)\right| \le C_3 \frac{1}{\sqrt{s}\log^{3/2} m}$, as long as $m \gg \mu^2 sK \log^6 m$.

It remains to justify the incoherence of $\boldsymbol{h}_i^{t+1}$ with respect to $\boldsymbol{b}_l$ for all $1 \le i \le s$. The result is summarized as follows.

Lemma 4. *Suppose the induction hypotheses (13) hold true for t-th iteration and the number of measurements obeys $m \gg \mu^2 s^2 K \log^8 m$. Then with probability at least $1-O(m^{-9})$ $\max_{1\le i\le s, 1\le j\le m} \left|\boldsymbol{b}_l^*\tilde{\boldsymbol{h}}_i^{t+1}\right| \le C_4 \frac{\mu}{\sqrt{m}}\log^2 m$, provided that C_4 is sufficiently large and the step size $\eta > 0$ is taken to be some sufficiently small constant, e.g., $\eta \asymp s^{-1}$.*

Remark 2. *For the proofs of Lemma 3 and 4, since the term $\sum_{k=1}^s \boldsymbol{b}_j^*\boldsymbol{h}_k\boldsymbol{x}_k^*\boldsymbol{a}_{kj}$ in (5a) is incoherent to the term $\boldsymbol{b}_j\boldsymbol{a}_{ij}^*\boldsymbol{x}_i$ in (5a) as $k \ne i$ (similar to the terms in (5b)), different techniques are needed to derive the statistical guarantees, thereby identifying induction hypotheses (13).*

C. Establishing Initial Point in the Region of Incoherence and Contraction

In order to finish the induction step, we still need to show that the spectral initialization $\boldsymbol{z}_i^0$ and $\boldsymbol{z}_i^{0,(l)}$ for $1 \le i \le s$ hold for the induction hypotheses (13) of local geometry. The related lemmas are summarized as follows.

Lemma 5. *With probability at least $1-O(m^{-9})$, there exists some constant $C > 0$ such that*

$$\min_{\alpha_i\in\mathbb{C}|\alpha_i|=1}\left\{\left\|\alpha_i\boldsymbol{h}_i^0-\boldsymbol{h}_i^\natural\right\|+\left\|\alpha_i\boldsymbol{x}_i^0-\boldsymbol{x}_i^\natural\right\|\right\} \le \frac{\xi}{\sqrt{s}} \text{ and}$$
$$\min_{\alpha_i\in\mathbb{C}|\alpha_i|=1}\left\{\left\|\alpha_i\boldsymbol{h}_i^{0,(l)}-\boldsymbol{h}_i^\natural\right\|+\left\|\alpha_i\boldsymbol{x}_i^{0,(l)}-\boldsymbol{x}_i^\natural\right\|\right\} \le \frac{\xi}{\sqrt{s}}, \tag{17}$$

and $||\alpha_i^0|-1| < 1/4$, for each $1 \le i \le s, 1 \le l \le m$, provided that $m \ge C\mu^2 sK \log^2 m/\xi^2$.

From the definition of distance function (4), we immediately imply that $\mathrm{dist}(\boldsymbol{z}^0, \boldsymbol{z}^\natural) \le C_1 \log^2 m$, as long as $m \gg \mu^2 sK \log^6 m$. This establishes the inductive hypothesis (13a) for $t = 0$. We further show the identification of (13b) and (13d) for $t = 0$.

Lemma 6. *Suppose that $m \gg \mu^2 s^2 K \log^3 m$. Then with probability at least $1-O(m^{-9})$,*

$$\mathrm{dist}\left(\boldsymbol{z}^{0,(l)}, \tilde{\boldsymbol{z}}^0\right) \le C_2 \frac{s\mu}{\sqrt{m}}\sqrt{\frac{\mu^2 K \log^5 m}{m}} \text{ and} \tag{18}$$

$$\max_{1\le i\le m} |\boldsymbol{b}_l^*\tilde{\boldsymbol{h}}_i| \le C_4 \frac{\mu \log^2 m}{\sqrt{m}}. \tag{19}$$

Remark 3. *For the proofs of above two lemmas, you may recall the spectral initialization where $\boldsymbol{M}_i := \sum_{j=1}^m y_j \boldsymbol{b}_j \boldsymbol{a}_{ij}^*$. Considering the "incoherence" between multiple sources, we decompose $\boldsymbol{M}_i$ into the term of $j = i$ and the sum of terms of $j \ne i$. Further transformation is required to apply the variant of Wedin's sinΘ theorem [10].*

Finally, we specify (13c) regarding the incoherence of $\boldsymbol{x}_0$ with respect to the vector $\boldsymbol{a}_{ij}$ for each $1 \le i \le s, 1 \le j \le m$.

Lemma 7. *Suppose the sample complexity $m \gg \mu^2 s^2 K \log^6 m$. Then with probability at least $1-O(m^{-9})$, $\max_{1\le i\le s, 1\le j\le m} \left|\boldsymbol{a}_{ij}\left(\tilde{\boldsymbol{x}}_i^0-\boldsymbol{x}_i^\natural\right)\right| \le C_3 \frac{1}{\sqrt{s}\log^{3/2} m}$.*

V. Discussion

In this paper, we explored the local geometry of blind demixing with Wirtinger flow. We uncovered that, start with spectral initialization, the iterates of Wirtinger flow keep stay within the region of incoherence and contraction. The restricted strong convexity and qualified level of smoothness of such region leads to more aggressive step size for gradient descent, thereby accelerating convergence rates. The provable Wirtinger flow algorithm can thus solve the blind demixing problem with lower computational complexity, fewer algorithmic parameters than the state-of-art algorithms.

References

[1] S. Ling and T. Strohmer, "Blind deconvolution meets blind demixing: Algorithms and performance bounds," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4497–4520, Jul. 2017.

[2] J. Dong, K. Yang, and Y. Shi, "Blind demixing for low-latency communication," in *Proc. IEEE Wireless Commun. Networking Conf. (WCNC)*, Barcelona, Spain, Apr. 2018.

[3] Y. Shi, B. Mishra, and W. Chen, "Topological interference management with user admission control via riemannian optimization," *IEEE Trans. Wireless Commun.*, vol. 16, pp. 7362–7375, Nov. 2017.

[4] Y. Shi, J. Zhang, and K. B. Letaief, "Low-rank matrix completion for topological interference management by Riemannian pursuit," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4703–4717, Jul. 2016.

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

[6] S. Ling and T. Strohmer, "Regularized gradient descent: A nonconvex recipe for fast joint blind deconvolution and demixing," *Inform. Inference*, to appear, 2018.

[7] M. B. McCoy and J. A. Tropp, "Sharp recovery bounds for convex demixing, with applications," *Found. Comput. Math.*, vol. 14, pp. 503–567, Jun. 2014.

[8] E. J. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1985–2007, Apr. 2015.

[9] Y. Chen and E. Candes, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," in *Adv. Neural. Inf. Process. Syst. (NIPS)*, pp. 739–747, 2015.

[10] C. Ma, K. Wang, Y. Chi, and Y. Chen, "Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion and blind deconvolution," *arXiv preprint arXiv:1711.10467*, 2017.