BILINEAR INVERSE PROBLEMS WITH SPARSITY: OPTIMAL IDENTIFIABILITY CONDITIONS AND EFFICIENT RECOVERY

BY

YANJUN LI

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Doctoral Committee:

Professor Yoram Bresler, Chair
Professor Minh Do
Professor Olgica Milenkovic
Professor Pierre Moulin
Professor Justin Romberg, Georgia Institute of Technology
Professor Rayadurgam Srikant

ABSTRACT

Bilinear inverse problems (BIPs), the resolution of two vectors given their image under a bilinear mapping, arise in many applications. Without further constraints, BIPs are usually ill-posed. In practice, parsimonious structures of natural signals (e.g., subspace or sparsity) are exploited. However, there are few theoretical justifications for using such structures for BIPs. We consider two types of BIPs, blind deconvolution (BD) and blind gain and phase calibration (BGPC), with subspace or sparsity structures. Our contributions are twofold: we derive optimal identifiability conditions, and propose efficient algorithms that solve these problems.

In previous work, we provided the first algebraic sample complexities for BD that hold for Lebesgue almost all bases or frames. We showed that for BD of a pair of vectors in $\mathbb{C}^n$, with subspace constraints of dimensions m_1 and m_2, respectively, a sample complexity of $n \geq m_1 m_2$ is sufficient. This result is suboptimal, since the number of degrees of freedom is merely m_1+m_2-1. We provided analogous results, with similar suboptimality, for BD with sparsity or mixed subspace and sparsity constraints. In Chapter 2, taking advantage of the recent progress on the information-theoretic limits of unique low-rank matrix recovery, we finally bridge this gap, and derive an optimal sample complexity result for BD with generic bases or frames. We show that for BD of an arbitrary pair (resp. all pairs) of vectors in $\mathbb{C}^n$, with sparsity constraints of sparsity levels s_1 and s_2, a sample complexity of $n > s_1 + s_2$ (resp. $n > 2(s_1 + s_2)$) is sufficient. We also present analogous results for BD with subspace constraints or mixed constraints, with the subspace dimension replacing the sparsity level. Last but not least, in all the above scenarios, if the bases or frames follow a probabilistic distribution specified in Chapter 2, the recovery is not only unique, but also stable against small perturbations in the measurements, under the same sample complexities.

In previous work, we proposed studying the identifiability in bilinear in-

verse problems up to transformation groups. In particular, we studied several special cases of blind gain and phase calibration, including the cases of subspace and joint sparsity models on the signals, and gave sufficient and necessary conditions for identifiability up to certain transformation groups. However, there were gaps between the sample complexities in the sufficient conditions and the necessary conditions. In Chapter 3, under a mild assumption that the signals and models are generic, we bridge the gaps by deriving tight sufficient conditions with optimal or near optimal sample complexities.

Recently there has been renewed interest in solutions to BGPC with careful analysis of error bounds. In Chapter 4, we formulate BGPC as an eigenvalue/eigenvector problem, and propose to solve it via power iteration, or in the sparsity or joint sparsity case, via truncated power iteration (which we show is equivalent to a sparsity-projected gradient descent). Under certain assumptions, the unknown gains, phases, and the unknown signal can be recovered simultaneously. Numerical experiments show that power iteration algorithms work not only in the regime predicted by our main results, but also in regimes where theoretical analysis is limited. We also show that our power iteration algorithms for BGPC compare favorably with competing algorithms in adversarial conditions, e.g., with noisy measurement or with a bad initial estimate.

A problem related to BGPC is multichannel blind deconvolution (MBD) with a circular convolution model, i.e., the recovery of an unknown signal f and multiple unknown filters x_i from circular convolutions $y_i = x_i \circledast f$ $(i = 1, 2, \ldots, N)$. In Chapter 5, we consider the case where the x_i's are sparse, and convolution with f is invertible. Our nonconvex optimization formulation solves for a filter h on the unit sphere that produces sparse outputs $y_i \circledast h$. Under some technical assumptions, we show that all local minima of the objective function correspond to the inverse filter of f up to an inherent sign and shift ambiguity, and all saddle points have strictly negative curvatures. This geometric structure allows successful recovery of f and x_i using a simple manifold gradient descent algorithm with random initialization. Our theoretical findings are complemented by numerical experiments, which demonstrate superior performance of the proposed approach over the previous methods.

To my loved ones.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Prof. Yoram Bresler. He kindly opened the door to this great school for me and patiently guided me through my PhD grind. I am immensely grateful for his advice, support, trust, and encouragement. He has been an incredible teacher, and shares with me his knowledge, experience, insight and philosophy. He invested a tremendous amount of time and energy in helping me grow as a student, as a researcher, and as an individual. He has broadened my understanding of mathematical signal processing, and has taught me to think both rigorously as a mathematician and practically as an engineer. I owe what I have learned in graduate school mostly to him.

I would also like to thank my doctoral committee members for their invaluable mentorship. I look to them for inspiration and motivation. Prof. R. Srikant, Prof. Pierre Moulin, and Prof. Minh Do have taught me the essentials in signal processing, machine learning, statistics, and optimization. These are the knowledge and skills I relied on to finish my dissertation. Prof. Olgica Milenkovic is the only person who attended my qualifying exam, preliminary exam, and final exam. She has helped shape my thesis proposal from the very beginning. The work by Prof. Justin Romberg on blind deconvolution ignited my interest in bilinear inverse problems.

My thanks also go to my colleagues, who has made graduate school so much more fun. I have benefited tremendously from working closely with Kiryung Lee, who coauthored my first five journal publications. His knowledge, dedication, and optimism has helped me through the toughest days. Luke Pfister was my first office-mate. I have fond memories of the numerous conversations we had over the years, to which I owe my improved English. He has been a great influence on how to do research, such as how to typeset in LaTeX, how to code in Python, and how to run numerical experiments on a computer cluster. Bihan Wen has helped broaden my research, and I

have learned a lot about image processing from our joint work. He is a loyal friend, and goes above and beyond when I need his help.

My advisors during my internships at GE Global Research and NEC Labs, Jayakrishnan Unnikrishnan and Kai Zhang, offered me opportunities to expand my expertise in machine learning. Staff members at ECE and CSL has provided their help and support over the years.

My deepest gratitude goes to my parents. Their love and support have always been and will always be my source of positive energy. I would also like to thank my fiancée, Xingchen Kang, for making every day special.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION 1
1.1 Blind Deconvolution 1
1.2 Blind Gain and Phase Calibration 5
1.3 Multichannel Sparse Blind Deconvolution 7

CHAPTER 2 IDENTIFIABILITY AND STABILITY IN BLIND DECONVOLUTION 12
2.1 Problem Statement 12
2.2 Main Results 18
2.3 Identifiability in Low-Rank Matrix Recovery 23
2.4 Proof of the Main Results 31

CHAPTER 3 IDENTIFIABILITY IN BLIND GAIN AND PHASE CALIBRATION 38
3.1 Problem Statement 38
3.2 Main Results 42
3.3 Proof of the Main Results 46

CHAPTER 4 BLIND GAIN AND PHASE CALIBRATION VIA SPARSE SPECTRAL METHODS 55
4.1 Introduction 55
4.2 Power Iteration Algorithms for BGPC 60
4.3 Main Results 67
4.4 Fundamental Estimates 76
4.5 Proofs of the Main Results 82
4.6 Numerical Experiments 90

CHAPTER 5 MULTICHANNEL SPARSE BLIND DECONVOLUTION VIA MANIFOLD GRADIENT DESCENT 101
5.1 MSBD on the Sphere 101
5.2 Global Geometric View 103
5.3 Optimization Method 111
5.4 Extensions 114
5.5 Numerical Experiments 115

CHAPTER 6 CONCLUSION 125

APPENDIX A LEMMAS FOR CHAPTER 2 127
A.1 Concentration of Measure 127
A.2 Useful Lemmas about Minkowski Dimension 130

APPENDIX B PROOFS OF LEMMAS IN CHAPTER 4 133
B.1 Gap in Eigenvalues 133
B.2 Bounds of Perturbation Due to Randomness in A 135
B.3 Bounds of Perturbation Due to Noise 145
B.4 Scalar Concentration Bounds 149

APPENDIX C PROOFS FOR CHAPTER 5 151
C.1 Proofs for Section 5.2 151
C.2 Proofs for Section 5.3 163

REFERENCES 168

CHAPTER 1

INTRODUCTION

Bilinear inverse problems (BIPs), i.e., the simultaneous recovery of two variables x and y given bilinear measurement $z = \mathcal{A}(x, y)$, have attracted much attention recently [1]. However, theoretical understanding of the identifiability – or uniqueness of the solution to a BIP has been lacking until recently. Furthermore, there has been tremendous interest in efficient algorithms for BIPs with theoretical guarantees. In this dissertation, we present near optimal identifiability results for two BIPs, blind deconvolution (BD) [2] and blind gain and phase calibration (BGPC) [3]. We also study guaranteed efficient algorithms for BGPC [4], and for a related problem – multichannel blind deconvolution (MBD).

1.1 Blind Deconvolution

Blind deconvolution (BD) is the bilinear inverse problem of recovering the signal and the filter simultaneously given the their convolutioin or circular convolution. It arises in many applications, including blind image deblurring [5], blind channel equalization [6], speech dereverberation [7], and seismic data analysis [8]. Without further assumptions, BD is an ill-posed problem, and does not yield a unique solution. In Chapter 2, we focus on subspace or sparsity assumptions on the signal and the filter. These priors, which render BD better-posed by reducing the search space, were shown to be effective constraints or regularizers in various applications [9, 10, 11, 12, 13, 14]. However, despite the success in practice, the theoretical results on uniqueness in BD with a subspace or sparsity constraint are limited.

Recently, the "lifting" scheme – recasting bilinear or quadratic inverse problems, such as blind deconvolution and phase retrieval, as rank-1 matrix recovery from linear measurements – has attracted considerable attention

[14, 15]. Choudhary and Mitra [16] showed that identifiability in BD (or in any bilinear inverse problem) hinges on the set of rank-2 matrices in a certain nullspace. In particular, they showed a negative result that the solution to BD with a canonical sparsity prior, that is, sparsity over the natural basis, is *not* identifiable [17]. However, the authors did not analyze the identifiability of signals that are sparse over other dictionaries. Eldar et al. [18] derived tight sufficient conditions for low-rank matrix recovery. However, the authors did not exploit any sparsity priors, and the results do not apply to structured measurements that arise in BD.

Using the lifting framework, Ahmed et al. [14], Ling and Strohmer [19], and Lee et al. [20, 21] proposed algorithms to solve BD with with subspace constraints, mixed constraints, and sparsity constraints, respectively. Chi [22] solved BD with mixed constraints, where the sparse spikes do not necessarily lie on a grid.[1] They all showed successful recovery using convex programming or alternating minimization, which implies identifiability and stability. These results are constructive, being demonstrated by establishing performance guarantees of algorithms. However, the guarantees are only shown to hold with high probability. The probability of failure is *nonzero*, and decays in a power-law form as the size of the problem increases.

In previous work [23], we addressed the identifiability up to scaling in single channel blind deconvolution under subspace or sparsity constraints. We presented the first algebraic sample complexities for BD with fully deterministic signal models. In particular, we showed that for BD of a pair of vectors in $\mathbb{C}^n$, with generic subspace constraints of dimensions m_1 and m_2, the bilinear mapping is injective if $n \geq m_1 m_2$. This sufficient condition is suboptimal for two reasons. First, it has been shown that the information-theoretic limit (necessary condition) of such a problem is $n \geq m_1 + m_2 - 1$ [24, Theorem V.1]. Secondly, the number of degrees of freedom in the unknown pair of vectors is $m_1 + m_2 - 1$. Similarly, the sample complexities for BD with sparsity or with mixed constraints are $n \geq 2s_1 s_2$ and $n \geq 2s_1 m_2$, respectively, where s_1 and s_2 denote the sparsity levels of the signal and the filter. Here the cost for the unknown support is an extra factor of 2. These results suffer from the same suboptimality as the results for the subspace constraints, in comparison to

[1] The off-grid signal in [22] is not sparse over a fixed dictionary, and hence should not be confused with the setting in Chapter 2. The identifiability result corresponding to this scenario is an interesting open problem.

the number of degrees of freedom of the continuous-valued unknowns.

In Chapter 2, we finally bridge this gap. We show nearly optimal sufficient conditions for identifiability and stability in blind deconvolution that match the number of degrees of freedom in the unknowns. Results are given for the cases of subspace constraints, sparsity constraints, or mixed constraints, and for complex or real signal and filter. For example, a sample complexity of $n > s_1 + s_2$ is sufficient to recover a pair of signals, which are s_1 and s_2-sparse with respect to generic dictionaries, from their circular convolution. This sufficient condition almost matches the necessary condition in [24]. The results of Chapter 2 provide the first *tight* sample complexity bounds, without large constants or log factors, for unique and stable recovery in BD. Such tight bounds were not achieved (either for unique or for stable recovery) in any of the previous works [14, 19, 21, 22].

The tight sample complexities in the identifiability results apply to Lebesgue almost all bases or frames.[2] Given a sufficient number of measurements, the conditions for unique recovery are violated only on a set of Lebesgue measure zero. In this sense, these results are deterministic, requiring no probabilistic assumptions. As an immediate corollary though, if the bases or frames are drawn from any probability distribution that is absolutely continuous with respect to the Lebesgue measure (e.g., the entries are jointly Gaussian with a non-singular covariance, or i.i.d. following a uniform distribution, etc.), then the results in Chapter 2 hold: they imply that the signal and the filter are identifiable with probability 1, which is better than being identifiable with high probability as in previous works [14, 19, 21, 22].

The unique recovery results are complemented by matching stability results. If the bases or frames follow a distribution specified later in Chapter 2, then under the same sample complexities as in the identifiability results, the recovery is stable with high probability against small perturbations in the measurements. In Chapter 2, the probability of failure decays in an exponential form as the size of the problem increases, faster than the power-law decay in previous works [14, 19, 21, 22].

Although all the main results of Chapter 2 are stated and proved for 1D circular convolution, they translate to 2D or higher-dimensional circular con-

[2]Results of similar nature, in that they apply to "almost all" objects of interest, have been derived for FIR multichannel deconvolution [25] and for low-rank matrix recovery [18].

volutions, by replacing the 1D discrete Fourier transform (DFT) with 2D or higher-dimensional DFTs. These sample complexity bounds are theoretical confirmations that subspace and sparsity assumptions are effective regularizers for blind deconvolution problems, such as blind image deblurring and blind channel equalization. The solutions are indeed unique and stable as long as the number of measurements exceeds the number of unknowns. Although the emphasis of Chapter 2 is not on any practical method, it provides a guideline for solving BD with subspace or sparsity priors. Algorithms that succeed only in regimes with suboptimal dependence on subspace dimensions or sparsity levels (e.g., requiring a sample complexity of $n = \Omega(s_1 s_2)$ to recover a pair of signals of s_1 and s_2), are *not* due to a fundamental limitation, but due to the suboptimality of the method or its analysis. On the other hand, our results encourage the pursuit of algorithms that are guaranteed to succeed in the optimal regime [14, 26, 21].

One of the main technical tools for the derivation of our results are results on information-theoretic limits of low-rank matrix recovery. Inspired by the brilliant work of Riegler et al. [27] on such limits for real matrix recovery from noise-free observations, we extend the results to complex matrix recovery from noisy observations, and apply them to blind deconvolution. The contributions of our extension include: (i) we refine the covering number argument used in [27] to achieve stability under the same sample complexity; (ii) we provide a simpler proof that gets rid of some unnecessary technicalities; (iii) we derive a concentration of measure bound with better constants, and an analogous result in the complex case, which is a non-trivial extension. These results may be of independent interest.

After our paper [2] was submitted and posted on arXiv, Kech and Krahmer [28] proved slightly improved identifiability and stability results for blind deconvolution using techniques from algebraic geometry. Their sample complexities, proved to be both necessary and sufficient, differ from ours by an additive term of at most five samples. For example, we show that a sample complexity of $n > 2(s_1 + s_2)$ is sufficient for the uniform identifiability of every pair of signals of sparsity s_1 and s_2, respectively. In comparison, Kech and Krahmer gave an optimal bound $n \geq 2(s_1 + s_2) - 2$, which differs from our sample complexity by three samples. For BD with sparsity constraints, Kech and Krahmer only considered undercomplete or square dictionaries, in contrast to our analysis, which applies also to overcomplete dictionaries.

1.2 Blind Gain and Phase Calibration

Blind gain and phase calibration (BGPC) is a bilinear inverse problem that arises in many applications. It is the joint recovery of an unknown gain and phase vector $\lambda \in \mathbb{C}^n$ and signal vectors $\phi_1, \phi_2, \ldots, \phi_N \in \mathbb{C}^n$ given the entrywise product $Y = \text{diag}(\lambda)\Phi$, where $\Phi = [\phi_1, \phi_2, \ldots, \phi_N] \in \mathbb{C}^{n \times N}$. An example of BGPC is the joint estimation of albedo[3] and the lighting conditions in inverse rendering [29]. Another example is the joint recovery of source signals, and unknown gains and phases of sensors, in array processing [30], where the directions of arrival of source signals are properly discretized using a grid. Multichannel blind deconvolution (with the circular convolution model), i.e., the joint recovery of the signal and multiple channels, is also a BGPC problem. BGPC has been studied extensively, and numerous solutions have been proposed, in the context of direction of arrival estimation [31, 32, 33, 34, 35] or radar imaging [36, 37].

1.2.1 Identifiability

One of the fundamental questions is: When does BGPC admit a unique solution? Despite the massive research efforts in BGPC, there are few results on identifiability in terms of sample complexity. Several works provided partial answers to the uniqueness of BGPC in the context of certain applications. In each of these works, the problem formulation and treatment were tailored to the application. For example, Nguyen et al. [29] showed a sufficient condition for unique inverse rendering. Morrison et al. [38] proposed an algorithm for synthetic aperture radar (SAR) autofocus and showed a necessary condition for their algorithm. Both problems fall into the category of BGPC problems with subspace constraints.

In previous work [1, 39], by deriving general necessary and sufficient conditions for identifiability in a bilinear inverse problem up to a transformation group, we addressed the uniqueness in all BGPC problems in a common framework. Results were derived for several different scenarios, and were given in terms of sample complexities: the number of samples required for a unique solution. In particular, we considered the subspace constraint and

[3] Albedo, also known as reflection coefficient, is the ratio of reflected radiation from a surface to incident radiation upon it.

joint sparsity constraint scenarios for the signals, and derived sufficient conditions for the identifiability up to scaling (or other groups of equivalence transformations). We also gave necessary conditions in the form of tight lower bounds on sample complexities. The sufficient conditions and the necessary conditions coincide in some cases, but have gaps in other cases, which lead to some conjectures on how to bridge these gaps.

A limitation of the previous works [1, 39], is that the sample complexities in the sufficient conditions are suboptimal. For example, for BGPC with a subspace constraint of dimension m, the sample complexity in the sufficient condition is $N \geq m$. However, the necessary condition says that the sample complexity only needs to satisfy $N \geq \frac{n-1}{n-m}$. This less demanding sample complexity coincides with the bound obtained by counting the number of degrees of freedom and the number of measurements, and also agrees with the empirical phase transition [1]. The sufficient condition for identifiability in BGPC with a joint sparsity constraint at sparsity level s suffers from similar suboptimality: the sufficient condition is $N \geq s$, versus the necessary condition $N \geq \frac{n-1}{n-s}$.

In Chapter 3, we close the gaps between the sufficient and necessary conditions. In the subspace constraint scenario, the subspace model and the signals are assumed to be generic. Then we show that the sample complexity in the necessary condition is actually sufficient for almost all signals. Therefore, the sample complexity is optimal. This proves one of our conjectures in [1]. We also generalize this result to the joint-sparsity case, and derive a sample complexity that is near optimal. These results provide favorable uniqueness bounds for real-world applications. For example, in sensor array processing, if the number of sensors is four times the number of sources, then our results imply that two snapshots are sufficient to calibrate the gains and phases.

1.2.2 Efficient Solution

There exists a long line of research regarding the solutions for each application of BGPC. However, fundamental sample complexities for the uniqueness of solutions to BGPC [1, 3], and error bounds for efficient algorithms [40, 41] have been established only recently. A main drawback of the guaranteed

algorithms of [40, 41] is that the recovery error is sensitive to the choice of certain linear constraints. We refer readers to Section 4.1.4 for a detailed discussion of prior art.

In Chapter 4, we overcome the drawbacks of previous algorithms by reformulating the BGPC problem as an eigenvalue/eigenvector problem. In the subspace case, we use algorithms that find principal eigenvectors such as the power iteration algorithm (also known as the power method) [42, Section 8.2.1], to find the concatenation of the gain and phase vector and the vectorized signal matrix in the form of the principal component of a structured matrix. In the sparsity case, the problem resembles sparse principal component analysis (sparse PCA) [43]. We then propose to solve the sparse eigenvector problem using truncated power iteration [44].

The main contribution of Chapter 4 is the theoretical analysis of the error bounds of power iteration and truncated power iteration for BGPC in the subspace and joint sparsity cases, respectively. When the measurement matrix is random, and the signals and the noise are adversarial, our algorithms stably recover the unknown gains and phases, and the unknown signals with high probability under near optimal sample complexities. Since truncated power iteration relies on a good initial estimate, we also propose a simple initialization algorithm, and prove that the output is sufficiently good under certain technical conditions. The fundamental estimates derived in Chapter 4 can be applied to other algorithms for BGPC, and possibly to algorithms for similar problems.

We complement the theoretical results with numerical experiments, which show that the algorithms can indeed solve BGPC in the optimal regime. We also demonstrate that the algorithms are robust against noise and an inaccurate initial estimate. Experiments with different initialization schemes show that our initialization algorithm significantly outperforms the baseline. Then we apply the power iteration algorithm to inverse rendering, and showcase its effectiveness in real-world applications.

1.3 Multichannel Sparse Blind Deconvolution

Blind deconvolution, which aims to recover unknown vectors x and f from their convolution $y = x \circledast f$, has been extensively studied, especially in the

context of image deblurring [5, 45, 46, 47]. Recently, algorithms with theoretical guarantees have been proposed for single channel blind deconvolution [14, 19, 22, 26, 21, 48, 49]. In order for the problem to be well-posed, these previous methods assume that *both* x and f are constrained, to either reside in a known subspace or be sparse over a known dictionary. However, these methods cannot be applied if f (or x) is unconstrained, or does not have a subspace or sparsity structure.

In many applications in communications [50], imaging [51], and computer vision [52], convolutional measurements $y_i = x_i \circledast f$ are taken between a single signal (resp. filter) f and multiple filters (resp. signals) $\{x_i\}_{i=1}^N$. We call such problems multichannel blind deconvolution (MBD).[4] Importantly, in this multichannel setting, one can assume that only $\{x_i\}_{i=1}^N$ are structured, and f is unconstrained. While there has been abundant work on single channel blind deconvolution (with both f and x constrained), research in MBD (with f unconstrained) is relatively limited. Traditional MBD works assumed that the channels x_i's are FIR filters [53, 54, 55] or IIR filters [56], and proposed to solve MBD using subspace methods. Despite the fact that MBD with a linear (i.e., standard, non-circular) convolution model is known to have a unique solution under mild conditions [25], the problem is generally ill-conditioned [57]. Recent works improved the conditioning of such problems by introducing subspace or low-rank structures for the multiple channels [57, 58].

In Chapter 5, while retaining the unconstrained form of f, we consider a different structure of the multiple channels $\{x_i\}_{i=1}^N$: sparsity. The resulting problem is termed multichannel sparse blind deconvolution (MSBD). The sparsity structure arises in many real-world applications.

Opportunistic underwater acoustics: Underwater acoustic channels are sparse in nature [59]. Estimating such sparse channels with an array of receivers using opportunistic sources (e.g., shipping noise) involves a blind deconvolution problem with multiple unknown sparse channels [60, 61].

Reflection seismology: Thanks to the layered earth structure, reflectivity in seismic signals is sparse. It is of great interest to simultaneous recover the filter (also known as the wavelet), and seismic reflectivity along the multiple propagation paths between the source and the geophones [62].

[4]Since convolution is a commutative operation, we use "signal" and "filter" interchangeably.

Functional MRI: Neural activity signals are composed of brief spikes and are considered sparse. However, observations via functional magnetic resonance imaging (fMRI) are distorted by convolving with the hemodynamic response function. A blind deconvolution procedure can reveal the underlying neural activity [63].

Super-resolution fluorescence microscopy: In super-resolution fluorescence microscopic imaging, photoswitchable probes are activated stochastically to create multiple sparse images and allow microscopy of nanoscale cellular structures [64, 65]. One can further improve the resolution via a computational deconvolution approach, which mitigates the effect of the point spread function (PSF) of the microscope [66]. It is sometimes difficult to obtain the PSF (e.g., due to unknown aberrations), and one needs to jointly estimate the microscopic images and the PSF [67].

Previous approaches to MSBD have provided efficient iterative algorithms to compute maximum likelihood (ML) estimates of parametric models of the channels $\{x_i\}_{i=1}^N$ [61], or maximum a posteriori (MAP) estimates in various Bayesian frameworks [62, 52]. However, these algorithms usually do not have theoretical guarantees or sample complexity bounds.

Recently, guaranteed algorithms for MSBD have been developed. Wang and Chi [41] proposed a convex formulation of MSBD based on ℓ_1 minimization, and gave guarantees for successful recovery under the condition that f has one dominant entry that is significantly larger than other entries. In our previous work [4], we solved a nonconvex formulation using projected gradient descent (truncated power iteration), and proposed an initialization algorithm to compute a sufficiently good starting point. However, in that work, theoretical guarantees were derived only for channels that are sparse with respect to a Gaussian random dictionary, but not channels that are sparse with respect to the standard basis.

We would like to emphasize that, while earlier papers on MBD [53, 54, 55, 56] consider a linear convolution model, more recent guaranteed methods for MSBD [41, 4] consider a circular convolution model. By zero padding the signal and the filter, one can rewrite a linear convolution as a circular convolution. In practice, circular convolution is often used to approximate a linear convolution when the filter has a compact support or decays fast [68], and the signal has finite length or satisfies a circular boundary condition [45]. The accelerated computation of circular convolution via the fast Fourier

transform (FFT) is especially beneficial in 2D or 3D applications [45, 67].

Multichannel blind deconvolution with a circular convolution model is also related to blind gain and phase calibration [1, 69, 70, 40]. Suppose that a sensing system takes Fourier measurements of unknown signals and the sensors have unknown gains and phases, i.e., $\tilde{y}_i = \text{diag}(\tilde{f})\mathcal{F}x_i$, where x_i are the targeted unknown sparse signals, $\mathcal{F}$ is the discrete Fourier transform (DFT) matrix, and the entries of $\tilde{f}$ represent the unknown gains and phases of the sensors. The simultaneous recovery of $\tilde{f}$ and x_i's is equivalent to MSBD in the frequency domain.

In Chapter 5, we consider MSBD with circular convolution. In addition to the sparsity prior on the channels $\{x_i\}_{i=1}^N$, we impose, without loss of generality, the constraint that f has unit ℓ_2 norm, i.e., f is on the unit sphere. (This eliminates the scaling ambiguity inherent in the MBD problem.) We show that our sparsity promoting objective function has a nice geometric landscape on the the unit sphere: **(S1)** all local minima correspond to signed shifted versions of the desired solution, and **(S2)** the objective function is strongly convex in neighborhoods of the local minima, and has strictly negative curvature directions in neighborhoods of local maxima and saddle points. Similar geometric analysis has been conducted for dictionary learning [71], phase retrieval [72], and single channel sparse blind deconvolution [49]. Recently, Mei et al. [73] analyzed the geometric structure of the empirical risk of a class of machine learning problems (e.g., nonconvex binary classification, robust regression, and Gaussian mixture model). Chapter 5 is the first such analysis for MSBD.

Properties **(S1)** and **(S2)** allow simple manifold optimization algorithms to find the ground truth in the nonconvex formulation. Unlike the second-order methods in previous works [74, 72], we take advantage of recent advances in the understanding of first-order methods [75, 76], and prove that a simple manifold gradient descent algorithm, with random initialization and a fixed step size, can accurately recover a signed shifted version of the ground truth in polynomial time almost surely. This is the first guaranteed algorithm for MSBD that does *not* rely on restrictive assumptions on f (e.g., dominant entry [41], spectral flatness [4]), or on $\{x_i\}_{i=1}^N$ (e.g., jointly sparse, Gaussian random dictionary [4]).

Recently, many optimization methods have been shown to escape saddle points of objective functions with benign landscapes, e.g., gradient descent

[77, 78], stochastic gradient descent [79], perturbed gradient descent [80], Natasha [81, 82], and FastCubic [83]. Similarly, optimization methods over Riemannian manifolds that can escape saddle points include manifold gradient descent [76], the trust region method [74, 72], and the negative curvature method [84]. Our main result shows that these algorithms can be applied to MSBD thanks to the favorable geometric properties of our objective function.

CHAPTER 2

IDENTIFIABILITY AND STABILITY IN BLIND DECONVOLUTION

2.1 Problem Statement

2.1.1 Notations

We use lower-case letters x, y, z to denote vectors, and upper-case letters D and E to denote matrices. We use F to denote the normalized (unitary) discrete Fourier transform (DFT) matrix. Unless otherwise stated, all vectors are column vectors. The dimensions of all vectors and matrices are made clear in the context. We use superscript letters to denote subvectors or submatrices. For example, the scalar $x^{(j)}$ represents the jth entry of x. The vector $D^{(j,:)}$ represents the jth row of the matrix D. The colon notation is borrowed from MATLAB. The transpose and conjugate transpose to a matrix A are denoted by A^T and A^*, respectively. The inner product of two matrices A and M are denoted by $\langle A, M\rangle = \mathrm{trace}(A^*M)$. We use $\|\cdot\|_0$ to denote the ℓ_0 "norm", or number of nonzero entries. We use $\|\cdot\|_2$ to denote the ℓ_2 norm of a vector or the spectral norm of a matrix, and $\|\cdot\|_{\mathrm{F}}$ to denote the Frobenious norm of a matrix. We use $\odot$ to denote entrywise product. Circular convolution is denoted by $\circledast$.

We say a subset $\Omega_{\mathcal{M}}$ of a linear vector space is a cone, if for every $M \in \Omega_{\mathcal{M}}$ and every $\sigma > 0$, the scaled vector $\sigma M \in \Omega_{\mathcal{M}}$. The real and imaginary parts of a complex vector are denoted by $\mathrm{Re}(x)$ and $\mathrm{Im}(x)$, respectively. If $\Omega_{\mathcal{X}}$ is a subset of $\mathbb{C}^m$, then we use $\mathrm{Re}(\Omega_{\mathcal{X}}) = \{\mathrm{Re}(x) : x \in \Omega_{\mathcal{X}}\}$, and $\mathrm{Im}(\Omega_{\mathcal{X}}) = \{\mathrm{Im}(x) : x \in \Omega_{\mathcal{X}}\}$ to denote the real and imaginary parts of $\Omega_{\mathcal{X}}$. The unit ball in $\mathbb{R}^m$ (with respect to the ℓ_2 norm) centered at the origin is denoted by $\mathcal{B}_{\mathbb{R}^m}$. Then $x + R\mathcal{B}_{\mathbb{R}^m}$ denotes the ball in $\mathbb{R}^m$ of radius R centered at x. Similarly, the unit ball in $\mathbb{C}^{m_1\times m_2}$ (with respect to the Frobenius norm) centered at the origin is denoted by $\mathcal{B}_{\mathbb{C}^{m_1\times m_2}}$. Then $M + R\mathcal{B}_{\mathbb{C}^{m_1\times m_2}}$ denotes

the ball in $\mathbb{C}^{m_1 \times m_2}$ of radius R centered at M. We use $V_{\mathbb{C}^m}(R) = \int_{R\mathcal{B}_{\mathbb{C}^m}} \mathrm{d}x$ to denote the volume of a ball of radius R in $\mathbb{C}^m$. Here, the multiple integral of a real-valued function $f(x)$ over $\Omega_{\mathcal{X}} \subset \mathbb{C}^m$ is defined as the multiple integral of $f(y^{(1:m)} + \sqrt{-1}y^{(m+1:2m)})$ over $\{y \in \mathbb{R}^{2m} : y^{(1:m)} + \sqrt{-1}y^{(m+1:2m)} \in \Omega_{\mathcal{X}}\}$.

We say a property holds for (Lebesgue) almost all vectors/matrices, or generic vectors/matrices, if the property holds for all vectors/matrices except for a set of Lebesgue measure zero.

2.1.2 Blind Deconvolution

In this chapter, we study the blind deconvolution (BD) problem with the circular convolution model. It is the joint recovery of two vectors $u_0 \in \mathbb{C}^n$ and $v_0 \in \mathbb{C}^n$, namely the signal and the filter,[1] given their circular convolution $z = u_0 \circledast v_0$, subject to subspace or sparsity constraints. The constraint sets $\Omega_{\mathcal{U}}$ and $\Omega_{\mathcal{V}}$ are subsets of $\mathbb{C}^n$. With these definitions, the BD problem is written as follows:

$$\begin{aligned} \text{Find} \quad & (u, v), \\ \text{s.t.} \quad & u \circledast v = z, \\ & u \in \Omega_{\mathcal{U}}, \ v \in \Omega_{\mathcal{V}}. \end{aligned}$$

We further assume that the constraint sets, which add to BD the prior information of the signal and the filter, are subspaces or sets of sparse vectors over a dictionary. For example, in blind image deblurring, the image (signal) can be assumed to be sparse over a dictionary (e.g., wavelets). The point spread function (filter) either has a small support and hence belongs to a subspace, or follows a simple parametric model that can be linearized by manifold embedding [85]. Another example is blind echo cancellation, where one can model a multipath channel as a sparse vector. With channel coding, the transmitted signal resides in the column space of the coding matrix. For more examples of subspace or sparsity priors in BD, we refer the readers to [14, 19, 21, 26] and the references therein. Specifically, we consider the following scenarios for the constraints:

1. *(Subspace Constraints)* The signal u and the filter v reside in lower-

[1] Due to symmetry, the name "signal" and "filter" can be used interchangeably.

dimensional subspaces spanned by the columns of $D \in \mathbb{C}^{n\times m_1}$ and $E \in \mathbb{C}^{n\times m_2}$, respectively, with $m_1, m_2 < n$. The matrices D and E have full column ranks. The signal $u = Dx$ for some $x \in \mathbb{C}^{m_1}$. The filter $v = Ey$ for some $y \in \mathbb{C}^{m_2}$.

2. *(Sparsity Constraints)* The signal u and the filter v are sparse over given dictionaries formed by the columns of $D \in \mathbb{C}^{n\times m_1}$ and $E \in \mathbb{C}^{n\times m_2}$, with sparsity level s_1 and s_2, respectively. Here m_1 and m_2 do not have to be smaller than n. The matrices D and E are bases or frames that satisfy the spark condition [86]: the spark, namely the smallest number of columns that are linearly dependent, of D (resp. E) is greater than $2s_1$ (resp. $2s_2$). The signal $u = Dx$ for some $x \in \mathbb{C}^{m_1}$ with $\|x\|_0 \leq s_1$. The filter $v = Ey$ for some $y \in \mathbb{C}^{m_2}$ with $\|y\|_0 \leq s_2$.

3. *(Mixed Constraints)* The signal u is sparse over a given dictionary $D \in \mathbb{C}^{n\times m_1}$, and the filter v resides in a lower-dimensional subspace spanned by the columns of $E \in \mathbb{C}^{n\times m_2}$, with $m_2 < n$. The matrix D satisfies the spark condition, and E has full column rank. The signal $u = Dx$ for some $x \in \mathbb{C}^{m_1}$ with $\|x\|_0 \leq s_1$. The filter $v = Ey$ for some $y \in \mathbb{C}^{m_2}$.[2]

In all three scenarios, the vectors x, y, and z reside in Euclidean spaces $\mathbb{C}^{m_1}$, $\mathbb{C}^{m_2}$ and $\mathbb{C}^n$. Given the measurement $z = (Dx_0) \circledast (Ey_0)$, the blind deconvolution problem can be rewritten in the following form:

$$\begin{aligned} \text{(BD)} \qquad \text{Find} \quad & (x, y), \\ \text{s.t.} \quad & (Dx) \circledast (Ey) = z, \\ & x \in \Omega_{\mathcal{X}},\ y \in \Omega_{\mathcal{Y}}. \end{aligned}$$

If D and E satisfy the full column rank condition or the spark condition, then the uniqueness of (u, v) is equivalent to the uniqueness of (x, y). Indeed, the full rank or spark conditions are satisfied for Lebesgue almost all D and E. Therefore, the results about the recovery of (x, y) in BD with generic bases or frames imply the corresponding results for (u, v). For simplicity, we will discuss problem (BD) from now on. The constraint sets $\Omega_{\mathcal{X}}$ and $\Omega_{\mathcal{Y}}$ depend on the constraints on the signal and the filter. For subspace constraints, $\Omega_{\mathcal{X}}$

[2] We can also consider the scenario where u resides in a subspace spanned by the columns of D, and v is sparse over E. By symmetry, the analysis will be almost identical, and thus omitted.

and $\Omega_{\mathcal{Y}}$ are $\mathbb{C}^{m_1}$ and $\mathbb{C}^{m_2}$, respectively. For sparsity constraints, $\Omega_{\mathcal{X}}$ and $\Omega_{\mathcal{Y}}$ are $\{x \in \mathbb{C}^{m_1} : \|x\|_0 \leq s_1\}$ and $\{y \in \mathbb{C}^{m_2} : \|y\|_0 \leq s_2\}$, respectively.

2.1.3 Identifiability up to Scaling

An important question concerning the blind deconvolution problem is to determine when it admits a unique solution. The BD problem suffers from scaling ambiguity. For any nonzero scalar $\sigma \in \mathbb{C}$ such that $\sigma x_0 \in \Omega_{\mathcal{X}}$ and $\frac{1}{\sigma} y_0 \in \Omega_{\mathcal{Y}}$, $(D(\sigma x_0)) \circledast (E(\frac{1}{\sigma} y_0)) = (Dx_0) \circledast (Ey_0) = z$. Therefore, BD does not yield a unique solution if $\Omega_{\mathcal{X}}, \Omega_{\mathcal{Y}}$ contain such scaled versions of x_0, y_0 (which is the case for the subspace or sparsity constraint sets in the previous section). Any valid definition of unique recovery in BD must address this issue. Our approach is as follows. If every solution (x, y) is a scaled version of (x_0, y_0), then we say that (x_0, y_0) can be uniquely identified up to scaling.[3] We also consider the case when this property is satisfied by all pairs (x_0, y_0) of interest. Thus we define identifiability as follows.

Definition 2.1.1.

1. ***Weak identifiability:*** *We say that the pair $(x_0, y_0) \in \Omega_{\mathcal{X}} \times \Omega_{\mathcal{Y}}$, in which $x_0 \neq 0$ and $y_0 \neq 0$, is identifiable up to scaling, if every solution $(x, y) \in \Omega_{\mathcal{X}} \times \Omega_{\mathcal{Y}}$ satisfies $x = \sigma x_0$ and $y = \frac{1}{\sigma} y_0$ for some nonzero σ.*

2. ***Strong identifiability:*** *We say that the set $\Omega_{\mathcal{X}} \times \Omega_{\mathcal{Y}}$ is identifiable up to scaling, if every pair $(x_0, y_0) \in \Omega_{\mathcal{X}} \times \Omega_{\mathcal{Y}}$ that satisfies $x_0 \neq 0$ and $y_0 \neq 0$ is identifiable up to scaling.*

For blind deconvolution, there exists a linear operator $\mathcal{G}_{DE} : \mathbb{C}^{m_1 \times m_2} \to \mathbb{C}^n$ such that

$$\mathcal{G}_{DE}(xy^T) = (Dx) \circledast (Ey). \tag{2.1}$$

Given the measurement $z = \mathcal{G}_{DE}(x_0 y_0^T) = (Dx_0) \circledast (Ey_0)$, one can recast the BD problem as the recovery of the rank-1 matrix $M_0 = x_0 y_0^T \in \Omega_{\mathcal{M}} = \{xy^T :$

[3] Unconstrained BD also suffers from shift ambiguity. If the signal and the filter are circularly shifted by ℓ and $-\ell$, respectively, their circular convolution remains the same. However, the BD problem with generic basis or frames does not suffer from shift ambiguity. If the signal and the filter are shifted, then they no longer reside in the same generic subspaces, or are no longer sparse with respect to the same generic dictionaries, as before.

$x \in \Omega_{\mathcal{X}}, y \in \Omega_{\mathcal{Y}}\}$. Using this so-called "lifting" [14] procedure, the lifted BD problem has the following form:

$$\begin{aligned} \text{(Lifted BD)} \qquad \text{Find} \quad & M, \\ \text{s.t.} \quad & \mathcal{G}_{DE}(M) = z, \\ & M \in \Omega_{\mathcal{M}}. \end{aligned}$$

The uniqueness of M_0 is equivalent to the identifiability of (x_0, y_0) up to scaling. In (Lifted BD), *weak identifiability* means the recovery of M_0 is unique, or M_0 is the only point in $\Omega_{\mathcal{M}}$ that maps to $\mathcal{G}_{DE}(M_0)$. *Strong identifiability* means the recovery of all matrices in $\Omega_{\mathcal{M}}$ is unique, that is $\mathcal{G}_{DE}$ is injective on $\Omega_{\mathcal{M}}$, i.e., there exists $\mathcal{G}_{DE}^{-1} : \mathcal{G}_{DE}(\Omega_{\mathcal{M}}) \to \Omega_{\mathcal{M}}$.

Since $\Omega_{\mathcal{X}}$ and $\Omega_{\mathcal{Y}}$ are cones, the lifted constraint set $\Omega_{\mathcal{M}}$ is also a cone. As shown later, for the linear operator $\mathcal{G}_{DE}$ and the cone constraint set $\Omega_{\mathcal{M}}$, identifiability on $\Omega_{\mathcal{M}}$ is essentially the same as identifiability on the constraint set restricted to the unit ball $\Omega_{\mathcal{M}} \bigcap \mathcal{B}_{\mathbb{C}^{m_1 \times m_2}}$. From now on, we use the shorthand notation

$$\Omega_{\mathcal{B}} \coloneqq \Omega_{\mathcal{M}} \bigcap \mathcal{B}_{\mathbb{C}^{m_1 \times m_2}}. \tag{2.2}$$

Hence $\sigma\Omega_{\mathcal{B}} = \Omega_{\mathcal{M}} \bigcap \sigma\mathcal{B}_{\mathbb{C}^{m_1 \times m_2}}$.

2.1.4 Stable Recovery

Noise is ubiquitous in real-world applications. In a noisy setting, the measurement in matrix recovery is $z = \mathcal{G}_{DE}(M_0) + \xi$, where $M_0 = x_0 y_0^T$ denotes the true rank-1 matrix, and ξ denotes noise or other perturbation in the measurement. In order to estimate M_0 from the measurement z, we consider the following constrained least squares problem:

$$\begin{aligned} \text{(Noisy BD)} \qquad \min_{M} \quad & \left\| \mathcal{G}_{DE}(M) - z \right\|_2, \\ \text{s.t.} \quad & M \in \sigma\Omega_{\mathcal{B}}, \end{aligned}$$

where $\sigma\Omega_{\mathcal{B}} = \{xy^T : x \in \Omega_{\mathcal{X}}, y \in \Omega_{\mathcal{Y}}, \left\| xy^T \right\|_{\mathrm{F}} \leq \sigma\}$. For all practical purposes, the solution to a blind deconvolution problem is bounded. Therefore, we solve (Noisy BD) subject to the constraint set restricted to a ball, whose

radius σ is sufficiently large. For example, σ can be set based on conservative upper estimates of the energy of x_0 and y_0.

We introduce the following two notions of stability of recovery:

Definition 2.1.2.

1. ***Single point stability:*** *We say that the recovery of $M_0 \in \sigma\Omega_{\mathcal{B}}$, using measurement operator $\mathcal{G}_{DE}$ and constraint set $\sigma\Omega_{\mathcal{B}}$, is stable, if for all $M \in \sigma\Omega_{\mathcal{B}}$ such that $\|\mathcal{G}_{DE}(M) - \mathcal{G}_{DE}(M_0)\|_2 \leq \delta$, we have $\|M - M_0\|_2 \leq \varepsilon$.*

2. ***Uniform stability:*** *We say that the recovery on $\sigma\Omega_{\mathcal{B}}$ is uniformly stable if for all $M_1, M_2 \in \sigma\Omega_{\mathcal{B}}$ that satisfy $\|\mathcal{G}_{DE}(M_1) - \mathcal{G}_{DE}(M_2)\|_2 \leq \delta$, we have $\|M_1 - M_2\|_2 \leq \varepsilon$.*

In both definitions, $\varepsilon = \varepsilon(\delta)$ is a function of δ that vanishes as δ approaches 0.

It is easy to see that the stability as defined above, would guarantee the accuracy of the constrained least squares estimation. Let $M_1 = x_1 y_1^T$ denote the solution to (Noisy BD). Suppose the perturbation ξ is small, i.e., $\|\xi\|_2 \leq \frac{\delta}{2}$ for some small $\delta > 0$. Then the deviation of $\mathcal{G}_{DE}(M_1)$ from $\mathcal{G}_{DE}(M_0)$ is small, i.e.,

$$\begin{aligned}
&\|\mathcal{G}_{DE}(M_1) - \mathcal{G}_{DE}(M_0)\|_2 \\
&\leq \|\mathcal{G}_{DE}(M_1) - z\|_2 + \|z - \mathcal{G}_{DE}(M_0)\|_2 \\
&\leq 2\,\|\mathcal{G}_{DE}(M_0) - z\|_2 = 2\,\|\xi\|_2 \leq \delta.
\end{aligned}$$

By the definition of single point stability (or uniform stability), we have $\|M_1 - M_0\|_2 \leq \varepsilon(\delta)$, which is also a small quantity.

If the recovery of M_0 is stable, then for every $\varepsilon > 0$, there exists $\delta > 0$ such that for every $M \in \sigma\Omega_{\mathcal{B}}$ that satisfies $\|\mathcal{G}_{DE}(M) - \mathcal{G}_{DE}(M_0)\|_2 \leq \delta$, we have $\|M - M_0\|_2 \leq \varepsilon$. If the recovery is uniformly stable on $\sigma\Omega_{\mathcal{B}}$, then for every $\varepsilon > 0$, there exists $\delta > 0$ such that for all $M_1, M_2 \in \sigma\Omega_{\mathcal{B}}$ that satisfy $\|\mathcal{G}_{DE}(M_1) - \mathcal{G}_{DE}(M_2)\|_2 \leq \delta$, we have $\|M_1 - M_2\|_2 \leq \varepsilon$. If $\mathcal{G}_{DE}$ satisfies *strong identifiability*, i.e., $\mathcal{G}_{DE}$ is invertible when restricted to $\Omega_{\mathcal{M}}$, then *single point stability* at M_0 implies that $\mathcal{G}_{DE}^{-1}$ is continuous at $\mathcal{G}_{DE}(M_0)$. *Finally uniform stability* on $\sigma\Omega_{\mathcal{B}}$ implies that $\mathcal{G}_{DE}^{-1}$ is uniformly continuous on $\mathcal{G}_{DE}(\sigma\Omega_{\mathcal{B}})$.

Suppose $\Omega_{\mathcal{M}}$ is a cone, and we need to evaluate stability on $\sigma\Omega_{\mathcal{B}} = \Omega_{\mathcal{M}} \bigcap \sigma\mathcal{B}_{\mathbb{C}^{m_1\times m_2}}$. We can scale M_0 and the radius of the ball by $\frac{1}{\sigma}$ simultaneously. If for all $M \in \Omega_{\mathcal{B}}$ such that $\left\|\mathcal{G}_{DE}(M) - \mathcal{G}_{DE}(\frac{M_0}{\sigma})\right\|_2 \leq \delta$, we have $\left\|M - \frac{M_0}{\sigma}\right\|_2 \leq \varepsilon(\delta)$, then for all $M \in \sigma\Omega_{\mathcal{B}}$ such that $\|\mathcal{G}_{DE}(M) - \mathcal{G}_{DE}(M_0)\|_2 \leq \delta$, we have $\|M - M_0\|_2 \leq \sigma\varepsilon(\frac{\delta}{\sigma})$. Therefore, we only need to consider the stability of recovery on the constraint set restricted to the unit ball, $\Omega_{\mathcal{B}}$.

In the next section, we present the main results on the identifiability and stability in blind deconvolution, i.e., the optimal sample complexities that guarantee unique and stable recovery in (Lifted BD) and (Noisy BD), respectively.

2.2 Main Results

We present the weak and strong identifiability results for blind deconvolution in Section 2.2.1, and present single point and uniform stable recovery results in Section 2.2.2. These results are proved in Section 2.4, which depends heavily on the matrix recovery results in Section 2.3.

2.2.1 Identifiability Results

Subspace membership and sparsity have been used as priors in blind deconvolution for a long time. Previous works either use these priors without theoretical justification [9, 10, 11, 12, 13], or impose probabilistic models and show successful recovery with high probability [14, 19, 21, 22]. The sufficient conditions for the identifiability in BD in our prequel paper [23] are (except for a special class of so-called sub-band structured signals or filters) suboptimal. In this section, we present sufficient conditions for identifiability in BD, as defined in Section 2.1.1, with minimal assumptions. First, the weak identifiability results in the following theorem are sharp to within an additive term of two samples.

Theorem 2.2.1 (Weak Identifiability)**.** *If $n > d$, then for Lebesgue almost all $D \in \mathbb{C}^{n\times m_1}$ and $E \in \mathbb{C}^{n\times m_2}$, the pair $(x_0, y_0) \in \Omega_{\mathcal{X}} \times \Omega_{\mathcal{Y}}$ ($x_0 \neq 0$, $y_0 \neq 0$) is identifiable up to scaling. Here, d is the sample complexity bound, which is m_1+m_2, s_1+m_2, and s_1+s_2 in the subspace, mixed, and sparsity constraints scenarios, respectively.*

The above sufficient condition is appealing since it approaches the information-theoretic limit of blind deconvolution. For example, it has been shown that the necessary condition for blind deconvolution (or any bilinear inverse problem) with two unknown vectors of sparsity s_1 and s_2 is $n \geq s_1 + s_2 - 1$, which is a fundamental limit [24, Theorem V.1]. Therefore, to within two samples difference, our sufficient sample complexity presented above is optimal. Moreover, our sample complexity almost matches the number of degrees of freedom in the unknowns, which is $m_1 + m_2 - 1$, $s_1 + m_2 - 1$, and $s_1 + s_2 - 1$, for BD with subspace, mixed, and sparsity constraints, respectively.

This result is a sufficient condition for weak identifiability. Unlike our results on BD with generic bases or frames in [23], which guarantee the injectivity of the bilinear mapping of circular convolution, this result only guarantees the identifiability of one pair (x_0, y_0) in the constraint set. A sufficient condition for strong identifiability, which applies uniformly to all pairs (x_0, y_0) in the constraint set, is presented next. In comparison to the optimal result in Theorem 2.2.1, the cost for strong identifiability is a factor of 2 in the sample complexity.

Theorem 2.2.2 (Strong Identifiability)**.** *If $n > 2d$, then for Lebesgue almost all $D \in \mathbb{C}^{n\times m_1}$ and $E \in \mathbb{C}^{n\times m_2}$, all pairs $(x_0, y_0) \in \Omega_{\mathcal{X}} \times \Omega_{\mathcal{Y}}$ ($x_0 \neq 0$, $y_0 \neq 0$) are identifiable up to scaling. Here, d is the same as in Theorem 2.2.1.*

Interestingly, the sample complexity of Theorem 2.2.2 doubles that of Theorem 2.2.1. The extra samples are reasonable: (1) weak identifiability means that any **one** point other than (x_0, y_0) must map to a point different from $(Dx_0) \circledast (Ey_0)$; (2) strong identifiability means that any **two** distinct points in the set must map to different points in $\mathbb{C}^n$. A similar phenomenon in compressed sensing is well known: weak recovery of an s-sparse vector requires $s+1$ generic samples [87], but strong recovery (injectivity) requires $2s$ generic samples [86].

The above results hold true for Lebesgue almost all complex matrices D and E. However, in many real-world applications, both the signal and the filter are real vectors. Therefore, it is worthwhile to consider the special case where $D \in \mathbb{R}^{n\times m_1}$, $E \in \mathbb{R}^{n\times m_2}$, $x \in \mathbb{R}^{m_1}$, and $y \in \mathbb{R}^{m_2}$. We show that the same sample complexities still hold in this special case.

Theorem 2.2.3. *In the case where D, E, x, and y are real, the sample complexities in Theorems 2.2.1 and 2.2.2 hold for Lebesgue almost all $D \in \mathbb{R}^{n\times m_1}$ and $E \in \mathbb{R}^{n\times m_2}$.*

The proofs of Theorems 2.2.1 and 2.2.2 are presented in Section 2.4.1, and depend on Theorem 2.3.2, Corollaries 2.3.6 and 2.3.7, and Lemma 2.4.1. Theorem 2.2.3 is proved similarly in Section 2.4.2, with a variation of Theorem 2.3.2, i.e., Lemma 2.4.2.

All the results hold for Lebesgue almost all matrices D and E. When the sample complexity is met, the identifiability is violated only on a set of Lebesgue measure zero in the space of matrices D and E. Therefore, if D and E are drawn from a distribution that is absolutely continuous with respect to the Lebesgue measure (e.g., D and E are independent random matrices whose entries are i.i.d. following a Gaussian distribution), then the identifiability result holds almost surely.

2.2.2 Stability Results

The previous section gives the sample complexities that guarantee the identifiability in BD. Next, we show that the same sample complexity can guarantee stability. Recall that $\mathcal{G}_{DE}$ and $\Omega_{\mathcal{B}}$ are defined in (2.1) and (2.2), respectively. Here we only consider single point stability and uniform stability on $\Omega_{\mathcal{B}}$, which correspond to Definition 2.1.2 with $\sigma = 1$. As argued before, stability on $\Omega_{\mathcal{B}}$ implies stability on an arbitrary bounded set.

Theorem 2.2.4. *Assume that $D \in \mathbb{C}^{n\times m_1}$ and $E \in \mathbb{C}^{n\times m_2}$ are independent random matrices, such that the random vectors $\{(FD)^{(j,:)*}\}_{j=1}^{n}$ are i.i.d. following a uniform distribution on $R\mathcal{B}_{\mathbb{C}^{m_1}}$, and $\{(FE)^{(j,:)*}\}_{j=1}^{n}$ are i.i.d. following a uniform distribution on $R\mathcal{B}_{\mathbb{C}^{m_2}}$.*

1. *If $n > d$ and $\delta < \sqrt{n}R^2$, then with probability at least $1-C'(\frac{\delta^2}{R^4})^{n-d}(\frac{1}{\varepsilon^2})^n$, we have single point stability on $\Omega_{\mathcal{B}}$.*

2. *If $n > 2d$ and $\delta < \sqrt{n}R^2$, then with probability at least $1-C''(\frac{\delta^2}{R^4})^{n-2d}(\frac{1}{\varepsilon^2})^n$, we have uniform stability on $\Omega_{\mathcal{B}}$.*

Here, d is the same sample complexity bound as in Theorems 2.2.1 and 2.2.2. Except for a log factor, C' and C'' only depend on n, m_1, m_2, s_1, and s_2.

Table 2.1: A summary of the constants in Theorem 2.2.4.

	d	C'	C''
Subspace constraints	$m_1 + m_2$	$\frac{C^n}{n^{n-d}}$	$\frac{(4C)^n}{n^{n-2d}}$
Mixed constraints	$s_1 + m_2$	$\binom{m_1}{s_1}^2 \frac{C^n}{n^{n-d}}$	$\binom{m_1}{s_1}^4 \frac{(4C)^n}{n^{n-2d}}$
Sparsity constraints	$s_1 + s_2$	$\binom{m_1}{s_1}^2 \binom{m_2}{s_2}^2 \frac{C^n}{n^{n-d}}$	$\binom{m_1}{s_1}^4 \binom{m_2}{s_2}^4 \frac{(4C)^n}{n^{n-2d}}$

Define $C = 648\ m_1 m_2 \left(1 + 2\ln \frac{2\sqrt{n}R^2}{3\delta}\right)$. *The explicit expressions for* d, C', *and* C'' *in the scenarios of subspace, mixed, or sparsity constraints are summarized in Table 2.1.*

Theorem 2.2.4 is proved in Section 2.4.3. Its proof hinges on a key step (2.8) in the proof of Lemma 2.3.3, which is also crucial to the proofs of the identifiability results.

The stability results of Theorem 2.2.4 correspond to the identifiability results for the complex case, in Theorems 2.2.1 and 2.2.2. Similar stability results can be derived for the case where D, E, x, and y are real, which correspond to the identifiability results in Theorem 2.2.3. They are omitted here for brevity.

In the discussion below, we interpret the single point stability result in Theorem 2.2.4. The uniform stability result can be interpreted similarly. Here, to make sure that the probability of stable recovery $1 - C'(\frac{\delta^2}{R^4})^{n-d}(\frac{1}{\varepsilon^2})^n$ is non-trivial, let $\varepsilon = \varepsilon(\delta) > C'^{\frac{1}{2n}} \left(\frac{\delta}{R^2}\right)^{\alpha}$, where $\alpha = 1 - \frac{d}{n} \in (0, 1)$, and $\varepsilon(\delta)$ vanishes as δ approaches 0.

Reconstruction signal-to-noise ratio (RSNR) and measurement signal-to-noise ratio (MSNR) are defined respectively by:

$$\text{RSNR} = \frac{\|M_0\|_2^2}{\|M - M_0\|_2^2},$$
$$\text{MSNR} = \frac{\|\mathcal{G}_{DE}(M_0)\|_2^2}{\|\mathcal{G}_{DE}(M) - \mathcal{G}_{DE}(M_0)\|_2^2}.$$

Consider the case when the error bounds are tight: $\|M - M_0\|_2 = \varepsilon$, and $\|\mathcal{G}_{DE}(M) - \mathcal{G}_{DE}(M_0)\|_2 = \delta$. Since the matrix M_0 resides in the unit ball, RSNR is on the order of $\frac{1}{\varepsilon^2}$. Since $\{(FD)^{(j,:)*}\}_{j=1}^n$ and $\{(FE)^{(j,:)*}\}_{j=1}^n$ are

uniformly distributed on balls of radius R, the norm of the measurement $\mathcal{G}_{DE}(M_0)$ is on the order of R^2. Hence MSNR is on the order of $\frac{R^4}{\delta^2}$. Theorem 2.2.4 can then be interpreted as follows: the probability of failure (unstable reconstruction) is roughly $\mathrm{RSNR}^n \cdot \mathrm{MSNR}^{-(n-d)}$.

Let $\varepsilon(\delta) = C'^{\frac{1}{2n}} \left(\frac{\delta}{R^2}\right)^{\frac{\alpha}{2}}$, where $\alpha = 1 - \frac{d}{n}$, then the probability of single point stability in Theorem 2.2.4 reduces to $1 - (\frac{\delta}{R^2})^{n-d}$. If $n > d$, then as δ approaches 0, the recovery error $\varepsilon(\delta)$ vanishes, and the probability $1-(\frac{\delta}{R^2})^{n-d}$ converges to 1. This means that if D and E are random with the distributions specified in Theorem 2.2.4, then the recovery of M_0 is unique with probability 1, which is also a corollary of Theorem 2.2.1.

Next, we establish stability for the special case where the operator $\mathcal{G}_{DE}$ is an isometry in the mean. Given any matrix $M = xy^T$, we have

$$\mathcal{G}_{DE}^* \mathcal{G}_{DE}(M) = n \sum_{j=1}^{n} (FD)^{(j,:)*} (FD)^{(j,:)} M (FE)^{(j,:)T} \overline{(FE)^{(j,:)}},$$

the expectation of which is

$$\begin{aligned}
&\mathbb{E}\left[\mathcal{G}_{DE}^* \mathcal{G}_{DE}(M)\right] \\
&= \frac{n^2}{m_1 m_2} \mathbb{E}\left[\left\|(FD)^{(j,:)*}\right\|_2^2\right] \cdot \mathbb{E}\left[\left\|(FE)^{(j,:)*}\right\|_2^2\right] M \\
&= \frac{n^2}{m_1 m_2} \cdot \frac{m_1 R^2}{m_1 + 2} \cdot \frac{m_2 R^2}{m_2 + 2} M.
\end{aligned}$$

The first line follows from the fact that the distribution of $(FD)^{(j,:)*}$ and $(FE)^{(j,:)*}$ are independent and isotropic. The second line is due to the fact that $(FD)^{(j,:)*}$ and $(FE)^{(j,:)*}$ are uniformly distributed on $R\mathcal{B}_{\mathbb{C}^{m_1}}$ and $R\mathcal{B}_{\mathbb{C}^{m_2}}$, respectively. It follows that by setting $R = \left(\frac{(m_1+2)(m_2+2)}{n^2}\right)^{\frac{1}{4}}$, we have $\mathbb{E}\left[\mathcal{G}_{DE}^* \mathcal{G}_{DE}(M)\right] = M$.

Next, as an example, we analyze the uniform stability of the subspace constraints scenario, with this special choice of R. This will provide insight into how the constants vary with n, m_1, and m_2. Let $\varepsilon(\delta) = 2C''^{\frac{1}{2n}} \left(\frac{\delta}{R^2}\right)^{\beta}$, where $\beta = 1 - \frac{2(m_1+m_2)}{n}$. Substituting the expressions for R and C'', and ignoring the log factor, we have $\varepsilon(\delta) = O\left((m_1 m_2)^{\frac{1-\beta}{2}} n^{\frac{\beta}{2}} \delta^{\beta}\right)$. By Theorem 2.2.4, in the subspace constraints scenario, if $n > 2(m_1 + m_2)$, i.e., $\beta \in (0, 1)$,

then with probability at least $1 - 0.25^n$, we have

$$\|M_1 - M_2\|_2 \lesssim (m_1 m_2)^{\frac{1-\beta}{2}} n^{\frac{\beta}{2}} \|\mathcal{G}_{DE}(M_1) - \mathcal{G}_{DE}(M_2)\|_2^{\beta},$$

for all $M_1, M_2 \in \Omega_{\mathcal{B}}$. Hence, $\mathcal{G}_{DE}^{-1}$ is Hölder continuous of order β on $\mathcal{G}_{DE}(\Omega_{\mathcal{B}})$.

We conclude this section by emphasizing the differences between the identifiability results in Section 2.2.1 and the stability results in Section 2.2.2:

1. The identifiability results address the identifiability on cone constraint sets, whereas the stability results address the stability on the same constraint sets restricted to a ball of an arbitrary but finite radius. From a practical point of view, because the radius can be arbitrarily large, this restriction is of no significant consequence.

2. The identifiability results hold for generic (Lebesgue almost all) matrices D and E. The stability results hold with high probability when D and E follow some specific distributions.

2.3 Identifiability in Low-Rank Matrix Recovery

Using the lifted formulation, blind deconvolution with subspace or sparsity constraints has been reduced to the recovery, subject to constraints, of a rank-1 matrix from linear measurements that have a particular structure. The identifiability question in BD is thus reduced to identifiabilty in the latter recovery problem. In this section we address the more general question of identifiability in low-rank matrix recovery. Our results express the sample complexity for identifiability in terms of the Minkowski dimension of the set in which the matrix to be recovered lives. These results are applied to the BD problem in Section 2.4 to derive the main results of this chapter.

Recently, Riegler et al. [27] derived sample complexity results for low-rank matrix recovery, and for the recovery of matrices of low description complexity, that match the number of degrees of freedom. They considered the case where the matrices are real. Define the measurement operator $\mathcal{A}\colon \mathbb{R}^{m_1 \times m_2} \to \mathbb{R}^n$ as

$$z = \mathcal{A}(M_0) = \left[\langle A_1, M_0\rangle, \langle A_2, M_0\rangle, \cdots, \langle A_n, M_0\rangle\right]^T \in \mathbb{R}^n,$$

where $A_j \in \mathbb{R}^{m_1 \times m_2}$ $(j = 1, 2, \cdots, n)$ denote the measurement matrices. Denoting by $\Omega_{\mathcal{M}} \subset \mathbb{R}^{m_1 \times m_2}$ the constraint set (which is assumed to be nonempty and bounded) for the unknown matrix, the matrix recovery problem is

$$\begin{aligned} \text{(MR)} \qquad \text{Find} \quad & M, \\ \text{s.t.} \quad & \mathcal{A}(M) = z, \\ & M \in \Omega_{\mathcal{M}}. \end{aligned}$$

The conditions for unique solution to the matrix recovery problem (MR) are expressed in terms of the Minkowski dimension of the constraint set $\Omega_{\mathcal{M}}$, which is defined as follows.

Definition 2.3.1. *The lower and upper Minkowski dimensions of the nonempty bounded set $\Omega_{\mathcal{M}} \subset \mathbb{R}^{m_1 \times m_2}$ are*

$$\begin{aligned} \underline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}}) &:= \liminf_{\rho \to 0} \frac{\log N_{\Omega_{\mathcal{M}}}(\rho)}{\log \frac{1}{\rho}}, \\ \overline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}}) &:= \limsup_{\rho \to 0} \frac{\log N_{\Omega_{\mathcal{M}}}(\rho)}{\log \frac{1}{\rho}}, \end{aligned}$$

where $N_{\Omega_{\mathcal{M}}}(\rho)$ denotes the covering number of $\Omega_{\mathcal{M}}$ given by

$$N_{\Omega_{\mathcal{M}}}(\rho) = \min\left\{ k \in \mathbb{N} : \Omega_{\mathcal{M}} \subset \bigcup_{1 \le i \le k} (M_i + \rho \mathcal{B}_{\mathbb{R}^{m_1 \times m_2}}), \quad M_i \in \mathbb{R}^{m_1 \times m_2} \right\}.$$

If $\underline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}}) = \overline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}})$, then it is simply called the Minkowski dimension, denoted by $\dim_{\mathrm{B}}(\Omega_{\mathcal{M}})$.

The Minkowski dimension of the constraint set $\Omega_{\mathcal{M}}$ can be used to represent its description complexity. Riegler et al. showed that the solution to (MR) is unique if the sample complexity is greater than the description complexity. For almost all measurement matrices $A_1, A_2, \cdots, A_n \in \mathbb{R}^{m_1 \times m_2}$, the recovery of $M_0 \in \Omega_{\mathcal{M}}$ is unique if $n > \underline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}})$ (see [27, Theorem 1]). An even more amazing result is that the same sample complexity can be achieved by rank-1 measurement matrices. For almost all $a_j \in \mathbb{R}^{m_1}$ and $b_j \in \mathbb{R}^{m_2}$ $(j = 1, 2, \cdots, n)$, the recovery of $M_0 \in \Omega_{\mathcal{M}}$ from measurements $\langle a_j b_j^T, M_0 \rangle = a_j^T M_0 b_j$ $(j = 1, 2, \cdots, n)$ is unique if $n > \underline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}})$ (see [27, Theorem 2 and Lemma 3]).

In this section, we state and prove the extension of this result to the case where the matrices are complex. The Minkowski dimension of the constraint set of complex matrices $\Omega_{\mathcal{M}} \subset \mathbb{C}^{m_1 \times m_2}$ can be defined as in Definition 2.3.1, with the real number field $\mathbb{R}$ replaced by the complex number field $\mathbb{C}$. As will be shown in the next section, by simply changing the number field from real to complex, the Minkowski dimension of a set doubles. Meanwhile, by taking n complex-valued measurements, the number of real-valued measurements also doubles (from n to $2n$). Theorem 2.3.2 shows that, together with the fact that the Minkowski dimension doubles for the complex case, we need the same number of complex-valued measurements in complex matrix recovery as we need real-valued measurements in real matrix recovery.

Before the rigorous statement and proof, we provide an intuitive explanation for why the sample complexity matches the Minkowski dimension, which also serves as a road map to our proof. Weak or strong identifiability for almost all measurement operators means that the set of degenerate $\{a_j,\ b_j\}_{j=1}^n$ that map some nonzero matrix to zero, has Lebesgue measure zero. Alternatively, we can show that the set of "bad" $\{a_j,\ b_j\}_{j=1}^n$ that map some nonzero matrix to some point in a small ball of radius δ, has Lebesgue measure that vanishes as δ approaches zero. This Lebesgue measure turns out to be proportional to two quantities – the covering number of the constraint set $\Omega_{\mathcal{M}}$ with balls of radius δ, and the volume of a ball of δ in the ambient space $\mathbb{C}^n$ of measurements, i.e., the measure is roughly proportional to $(\frac{1}{\delta})^{\dim_{\mathrm{B}}(\Omega_{\mathcal{M}})} \cdot \delta^{2n}$. Therefore, it vanishes as δ approaches zero if $2n > \dim_{\mathrm{B}}(\Omega_{\mathcal{M}})$.

Theorem 2.3.2. *Suppose the set $\Omega_{\mathcal{M}} \subset \mathbb{C}^{m_1 \times m_2}$ is non-empty and bounded. For almost all sets of vectors $a_j \in \mathbb{C}^{m_1}$ and $b_j \in \mathbb{C}^{m_2}$ $(j = 1, 2, \cdots, n)$, there does not exist a matrix $M \in \Omega_{\mathcal{M}} \backslash \{0\}$ such that $\left\langle a_j b_j^T, M \right\rangle = a_j^* M \overline{b_j} = 0$ for $j = 1, 2, \cdots, n$, if $2n > \underline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}})$.*

Proof. We prove Theorem 2.3.2 using the following lemma.

Lemma 2.3.3. *Suppose the set $\Omega_{\mathcal{M}} \subset \mathbb{C}^{m_1 \times m_2}$ is non-empty and bounded. Let the vectors $\{a_j\}_{j=1}^n$ and $\{b_j\}_{j=1}^n$ be independent random vectors, where $\{a_j\}_{j=1}^n$ are i.i.d. following a uniform distribution on $R\mathcal{B}_{\mathbb{C}^{m_1}}$, and $\{b_j\}_{j=1}^n$ are i.i.d. following a uniform distribution on $R\mathcal{B}_{\mathbb{C}^{m_2}}$. If $2n > \underline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}})$, then*

$$P \coloneqq \mathbb{P}\left[\exists M \in \Omega_{\mathcal{M}} \backslash \{0\},\ s.t.\ a_j^* M \overline{b_j} = 0 \text{ for } j = 1, \cdots, n\right] = 0.$$

We use $\mathscr{N}(\Omega, \{a_j\}_{j=1}^n, \{b_j\}_{j=1}^n)$ to denote the event that there exists $M \in \Omega$ such that $a_j^* M \overline{b_j} = 0$ for $j = 1, 2, \cdots, n$. Here, we prove that such an event does not happen for almost all $\{a_j\}_{j=1}^n, \{b_j\}_{j=1}^n$ by proving it happens with probability zero for random $\{a_j\}_{j=1}^n, \{b_j\}_{j=1}^n$ following uniform distributions, thanks to the equivalence between the uniform measure and the Lebesgue measure. To be more specific, restricted to the same support $R\mathcal{B}_{\mathbb{C}^{m_1}} \times R\mathcal{B}_{\mathbb{C}^{m_2}}$, the Lebesgue measure is absolutely continuous with respect to the uniform distribution.[4] If the probability of the event $\mathscr{N}(\Omega_{\mathcal{M}}\backslash\{0\}, \{a_j\}_{j=1}^n, \{b_j\}_{j=1}^n)$ is zero, then the Lebesgue measure of the set of $\{a_j\}_{j=1}^n$ and $\{b_j\}_{j=1}^n$, over which the event happens, is zero too. It follows that, for almost all $a_j \in R\mathcal{B}_{\mathbb{C}^{m_1}}$ and $b_j \in R\mathcal{B}_{\mathbb{C}^{m_2}}$ $(j = 1, 2, \cdots, n)$, the event $\mathscr{N}(\Omega_{\mathcal{M}}\backslash\{0\}, \{a_j\}_{j=1}^n, \{b_j\}_{j=1}^n)$ does not happen. This argument is true for arbitrary radius R. Hence if $2n > \underline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}})$, then by Lemma 2.3.3 the event $\mathscr{N}(\Omega_{\mathcal{M}}\backslash\{0\}, \{a_j\}_{j=1}^n, \{b_j\}_{j=1}^n)$ does not happen, and therefore this event does not happen for almost all $a_j \in \mathbb{C}^{m_1}$ and $b_j \in \mathbb{C}^{m_2}$ $(j = 1, 2, \cdots, n)$, i.e., there does not exist a matrix $M \in \Omega_{\mathcal{M}}\backslash\{0\}$ such that $a_j^* M \overline{b_j} = 0$ for $j = 1, 2, \cdots, n$. Therefore, we only need to prove Lemma 2.3.3, thus completing the proof of Theorem 2.3.2. □

Proof of Lemma 2.3.3. The set $\Omega_{\mathcal{M}}\backslash\{0\}$ can be written as

$$\Omega_{\mathcal{M}}\backslash\{0\} = \bigcup_{L \in \mathbb{Z}^+} \Omega_{\mathcal{M},L}, \tag{2.3}$$

where $\Omega_{\mathcal{M},L} \coloneqq \{M \in \Omega_{\mathcal{M}} : \frac{1}{L} \leq \|M\|_2 \leq L\}$. By a union bound, we have

$$P \leq \sum_{L \in \mathbb{Z}^+} P_L,$$

where

$$P_L \coloneqq \mathbb{P}\left[\exists M \in \Omega_{\mathcal{M},L}, \text{ s.t. } a_j^* M \overline{b_j} = 0 \text{ for } j = 1, 2, \cdots, n\right].$$

In order to show that $P = 0$, it suffices to prove that $P_L = 0$ for all $L \in \mathbb{Z}^+$.

Let L be an arbitrary positive integer. We form a minimal cover of $\Omega_{\mathcal{M},L}$ with balls of radius ρ centered at the points $\{M_{\rho,L,i}\}_{i=1}^{N_{\Omega_{\mathcal{M},L}}(\rho)}$. These points may or may not be in $\Omega_{\mathcal{M},L}$. However, by the minimality of the cover, the

[4] Because the uniform measure is also absolutely continuous with respect to the Lebesgue measure, the two measures are equivalent.

intersection of $\Omega_{\mathcal{M},L}$ with each ball is nonempty, hence there exists another set of points $\{M'_{\rho,L,i}\}_{i=1}^{N_{\Omega_{\mathcal{M},L}}(\rho)}$ such that

$$M'_{\rho,L,i} \in \Omega_{\mathcal{M},L} \bigcap (M_{\rho,L,i} + \rho\mathcal{B}_{\mathbb{C}^{m_1\times m_2}}),$$

for $i = 1, 2, \cdots, N_{\Omega_{\mathcal{M},L}}(\rho)$. Now we cover $\Omega_{\mathcal{M},L}$ with balls of radius 2ρ centered at $\{M'_{\rho,L,i}\}_{i=1}^{N_{\Omega_{\mathcal{M},L}}(\rho)}$, which are points in $\Omega_{\mathcal{M},L}$ (a property that will be needed for inequality (2.8) below), because

$$(M_{\rho,L,i} + \rho\mathcal{B}_{\mathbb{C}^{m_1\times m_2}}) \subset (M'_{\rho,L,i} + 2\rho\mathcal{B}_{\mathbb{C}^{m_1\times m_2}}),$$

and

$$\Omega_{\mathcal{M},L} \subset \bigcup_{1\le i\le N_{\Omega_{\mathcal{M},L}}(\rho)} (M_{\rho,L,i} + \rho\mathcal{B}_{\mathbb{C}^{m_1\times m_2}}) \subset \bigcup_{1\le i\le N_{\Omega_{\mathcal{M},L}}(\rho)} (M'_{\rho,L,i} + 2\rho\mathcal{B}_{\mathbb{C}^{m_1\times m_2}}).$$

Defining $\delta = R^2\rho$, we have

$$\begin{aligned}
P_L \le & \sum_{i=1}^{N_{\Omega_{\mathcal{M},L}}(\rho)} \mathbb{P}\Big[\exists M \in (M'_{\rho,L,i} + 2\rho\mathcal{B}_{\mathbb{C}^{m_1\times m_2}}), \\
& \qquad \text{s.t. } a_j^* M \overline{b_j} = 0 \text{ for } j = 1, 2, \cdots, n\Big] && (2.4)\\
\le & \sum_{i=1}^{N_{\Omega_{\mathcal{M},L}}(\rho)} \mathbb{P}\Big[\exists M \in (M'_{\rho,L,i} + 2\rho\mathcal{B}_{\mathbb{C}^{m_1\times m_2}}), \\
& \qquad \text{s.t. } \left|a_j^* M \overline{b_j}\right| \le \delta \text{ for } j = 1, 2, \cdots, n\Big] && (2.5)\\
\le & \sum_{i=1}^{N_{\Omega_{\mathcal{M},L}}(\rho)} \mathbb{P}\Big[\left|a_j^* M'_{\rho,L,i} \overline{b_j}\right| \le 3\delta \text{ for } j = 1, 2, \cdots, n\Big] && (2.6)\\
= & \sum_{i=1}^{N_{\Omega_{\mathcal{M},L}}(\rho)} \prod_{j=1}^{n} \mathbb{P}\Big[\left|a_j^* M'_{\rho,L,i} \overline{b_j}\right| \le 3\delta\Big] && (2.7)\\
\le & N_{\Omega_{\mathcal{M}}}\left(\frac{\delta}{R^2}\right) (3\delta)^{2n} g(3\delta, \frac{1}{L}, L, R)^n. && (2.8)
\end{aligned}$$

Inequality (2.4) uses a union bound. The event in (2.4) implies the event in (2.5), which then implies the event in (2.6). Inequality (2.6) is due to the

following chain of inequalities, of which the last is implied by $|a_j^* M \overline{b_j}| \leq \delta$:

$$\begin{aligned}\left|a_j^* M'_{\rho,L,i} \overline{b_j}\right| \leq & \left|a_j^* (M'_{\rho,L,i} - M) \overline{b_j}\right| + \left|a_j^* M \overline{b_j}\right| \\ \leq & \|a_j\|_2 \left\|M'_{\rho,L,i} - M\right\|_2 \|b_j\|_2 + \left|a_j^* M \overline{b_j}\right| \\ \leq & \|a_j\|_2 \left\|M'_{\rho,L,i} - M\right\|_{\mathrm{F}} \|b_j\|_2 + \left|a_j^* M \overline{b_j}\right| \\ \leq & 2R^2\rho + \delta \;=\; 3\delta.\end{aligned}$$

Equation (2.7) is due to the independence between random vector pairs $\{a_j,\, b_j\}_{j=1}^n$. Inequality (2.8) uses the fact that $N_{\Omega_{\mathcal{M},L}}(\rho) \leq N_{\Omega_{\mathcal{M}}}(\rho) = N_{\Omega_{\mathcal{M}}}\left(\frac{\delta}{R^2}\right)$, and the concentration of measure inequality $\mathbb{P}\left[\left|a_j^* M'_{\rho,L,i} \overline{b_j}\right| \leq \delta\right] \leq \delta^2 g(\delta, \frac{1}{L}, L, R)$ in Lemma A.1.2 in Appendix A.1. (By construction, $M'_{\rho,L,i}$, as points in $\Omega_{\mathcal{M},L}$, satisfy the norm bounds $\frac{1}{L} \leq \left\|M'_{\rho,L,i}\right\|_2 \leq L$.) Here $g(\delta, \frac{1}{L}, L, R)$ is a function of δ defined in (A.6) in Appendix A.1 , which satisfies $\lim\limits_{\delta\to 0} \frac{\log g(\delta, \frac{1}{L}, L, R)}{\log \frac{1}{\delta}} = 0$.

Next, we show that (2.8) implies $P_L = 0$. Assume the contrary, i.e. $P_L > 0$. Since P_L does not depend on δ, we have $\liminf\limits_{\delta\to 0} \frac{\log P_L}{\log \frac{1}{\delta}} = 0$. By (2.8) and the assumed sample complexity $2n > \underline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}})$,

$$\begin{aligned}0 &= \liminf_{\delta\to 0} \frac{\log P_L}{\log \frac{1}{\delta}} \\ &\leq \liminf_{\delta\to 0} \frac{\log N_{\Omega_{\mathcal{M}}}\left(\frac{\delta}{R^2}\right) + 2n \log(3\delta) + n \log g(3\delta, \frac{1}{L}, L, R)}{\log \frac{1}{\delta}} \\ &= \underline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}}) - 2n < 0,\end{aligned}$$

which is a contradiction. Since L is arbitrary, we have $P_L = 0$ for all $L \in \mathbb{Z}^+$. This completes the proof of Lemma 2.3.3. □

Corollaries 2.3.4 and 2.3.5 are direct consequences of Theorem 2.3.2.

Corollary 2.3.4 (Weak Identifiability, Bounded)**.** *Suppose the constraint set $\Omega_{\mathcal{M}} \subset \mathbb{C}^{m_1 \times m_2}$ is nonempty and bounded. For almost all $a_j \in \mathbb{C}^{m_1}$ and $b_j \in \mathbb{C}^{m_2}$ $(j = 1, 2, \cdots, n)$, the recovery of M_0 from measurements $\left\langle a_j b_j^T, M_0 \right\rangle$ $(j = 1, 2, \cdots, n)$ is unique if $2n > \underline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}})$.*

Proof. Define the set $\Omega_{\mathcal{M}} - M_0 = \{M_1 - M_0 | M_1 \in \Omega_{\mathcal{M}}\}$. Saying that the recovery of M_0 from $a_j^* M_0 \overline{b_j}$ $(j = 1, 2, \cdots, n)$ is unique, is equivalent to saying that there does not exist a matrix M in $(\Omega_{\mathcal{M}} - M_0)\backslash\{0\}$ such that $\left\langle a_j b_j^T, M \right\rangle = 0$ $(j = 1, 2, \cdots, n)$.

Since the set $\Omega_{\mathcal{M}} - M_0$ is the shift of the set $\Omega_{\mathcal{M}}$ by M_0, we have that $\underline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}} - M_0) = \underline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}})$. Therefore, Corollary 2.3.4 follows from Theorem 2.3.2. □

Corollary 2.3.5 (Strong Identifiability, Bounded)**.** *Suppose the constraint set $\Omega_{\mathcal{M}} \subset \mathbb{C}^{m_1 \times m_2}$ is nonempty and bounded. For almost all $a_j \in \mathbb{C}^{m_1}$ and $b_j \in \mathbb{C}^{m_2}$ $(j = 1, 2, \cdots, n)$, the recovery of all matrices $M_0 \in \Omega_{\mathcal{M}}$ from measurements $\left\langle a_j b_j^T, M_0 \right\rangle$ $(j = 1, 2, \cdots, n)$ is unique if $n > \overline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}})$.*

Proof. Define the set $\Omega_{\mathcal{M}} - \Omega_{\mathcal{M}} = \{M_1 - M_2 | M_1, M_2 \in \Omega_{\mathcal{M}}\}$. Saying that the recovery of all matrices in $\Omega_{\mathcal{M}}$ is unique, is equivalent to saying that there does not exist a matrix M in $(\Omega_{\mathcal{M}} - \Omega_{\mathcal{M}}) \backslash \{0\}$ such that $\left\langle a_j b_j^T, M \right\rangle = 0$ $(j = 1, 2, \cdots, n)$.

By Lemma A.2.1 in Appendix A.2,

$$\underline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}} - \Omega_{\mathcal{M}}) \leq \overline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}} - \Omega_{\mathcal{M}}) \leq 2\overline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}}).$$

Therefore, Corollary 2.3.5 follows from Theorem 2.3.2. □

The proof of Theorem 2.3.2 is adapted from the proofs of [27, Theorem 2 and Lemma 3]. We make several refinements to this approach:

1. We simplify the expression of $\Omega_{\mathcal{M}} \backslash \{0\}$ as a union of subsets (see (2.3)). We define the subsets only by the spectral norm bounds, and remove technical discussions unrelated to our analysis of identifiability. This simplification also results in an easy proof of stability in Section 2.4.3.

2. We adjust the radius of balls in the covering number argument from δ to δ/R^2 (see (2.4) – (2.8)). This does not make any difference to the identifiability results, but has a big impact on the stability results. As will be shown by the proofs in Section 2.4.3, this change of radius results in tighter error bounds in Section 2.2.2, which can be interpreted in terms of signal-to-noise ratios.

3. We extend the analysis from the real case to the complex case, thus enabling its application to blind deconvolution. Despite the similarity in proofs, the extension is not a trivial application of the canonical isomorphism between $\mathbb{C}^n$ and $\mathbb{R}^{2n}$ (see Lemmas A.1.1 and A.1.2 in Appendix A.1).

The proofs in this chapter can serve as a simpler proof of the sample complexity for the real matrix recovery problem, which is $n > \underline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}})$, by making the following modifications:

1. Changing the number field from complex to real.

2. Using a different concentration of measure inequality in (2.8):

$$\mathbb{P}\left[\left|a_j^T M'_{\rho,L,i} b_j\right| \leq \delta\right] \leq \delta f(\delta, \frac{1}{L}, L, R),$$

which is formally stated and proved in Lemma A.1.1, where $f(\delta, \frac{1}{L}, L, R)$ is a function of δ that satisfies $\lim_{\delta \to 0} \frac{\log f(\delta, \frac{1}{L}, L, R)}{\log \frac{1}{\delta}} = 0$. Hence $P_L \leq N_{\Omega_{\mathcal{M}}}\left(\frac{\delta}{R^2}\right)\ (3\delta)^n f(3\delta, \frac{1}{L}, L, R)^n$. If $n > \underline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}})$, then $P_L = 0$ for all $L \in \mathbb{Z}^+$.

Owing to the linearity of the measurements in the matrix recovery problem, the above results can be easily extended to the case where the constraint set is a cone. To avoid verbosity, we only prove Corollary 2.3.6. Corollary 2.3.7 can be proved in a similar fashion.

Corollary 2.3.6 (Weak Identifiability, Unbounded)**.** *Suppose the constraint set $\Omega_{\mathcal{M}} \subset \mathbb{C}^{m_1 \times m_2}$ is a cone. For almost all $a_j \in \mathbb{C}^{m_1}$ and $b_j \in \mathbb{C}^{m_2}$ $(j = 1, 2, \cdots, n)$, the recovery of M_0 from measurements $\left\langle a_j b_j^T, M_0 \right\rangle = a_j^* M_0 \overline{b_j}$ $(j = 1, 2, \cdots, n)$ is unique if $2n > \underline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{B}})$, where $\Omega_{\mathcal{B}} = \Omega_{\mathcal{M}} \bigcap \mathcal{B}_{\mathbb{C}^{m_1 \times m_2}}$.*

Corollary 2.3.7 (Strong Identifiability, Unbounded)**.** *Suppose the constraint set $\Omega_{\mathcal{M}} \subset \mathbb{C}^{m_1 \times m_2}$ is a cone. For almost all $a_j \in \mathbb{C}^{m_1}$ and $b_j \in \mathbb{C}^{m_2}$ $(j = 1, 2, \cdots, n)$, the recovery of all matrices $M_0 \in \Omega_{\mathcal{M}}$ from measurements $\left\langle a_j b_j^T, M_0 \right\rangle = a_j^* M_0 \overline{b_j}$ $(j = 1, 2, \cdots, n)$ is unique if $n > \overline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{B}})$, where $\Omega_{\mathcal{B}} = \Omega_{\mathcal{M}} \bigcap \mathcal{B}_{\mathbb{C}^{m_1 \times m_2}}$.*

Proof of Corollary 2.3.6. We prove uniqueness by contradiction. Suppose that the recovery of M_0 is not unique, i.e., there exists $M_1 \in \Omega_{\mathcal{M}}$ such that $\left\langle a_j b_j^T, M_1 \right\rangle = \left\langle a_j b_j^T, M_0 \right\rangle$ $(j = 1, 2, \cdots, n)$. Let $\sigma \coloneqq 2 \max\{\|M_0\|_{\mathrm{F}}, \|M_1\|_{\mathrm{F}}\} > 0$. Since $\Omega_{\mathcal{M}}$ is a cone, we have

$$\frac{1}{\sigma} M_0,\ \frac{1}{\sigma} M_0 \in \Omega_{\mathcal{B}},$$

$$\left\langle a_j b_j^T, \frac{1}{\sigma} M_1 \right\rangle = \left\langle a_j b_j^T, \frac{1}{\sigma} M_0 \right\rangle, \quad j = 1, 2, \cdots, n.$$

Therefore, when the matrix recovery problem is restricted to a nonempty bounded constraint set $\Omega_{\mathcal{B}}$, the recovery of $\frac{1}{\sigma}M_0$ is not unique. This, however, contradicts the sample complexity $2n > \underline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{B}})$ and Corollary 2.3.4. $\square$

Corollaries 2.3.6 and 2.3.7 show that the solution to the matrix recovery problem with a cone constraint set is unique, if the solution to the corresponding problem restricted to the unit ball is unique.

2.4 Proof of the Main Results

2.4.1 Proof of Theorems 2.2.1 and 2.2.2

The identifiability of (x_0, y_0) up to scaling in (BD) is equivalent to the uniqueness of $M_0 = x_0 y_0^T$ in (Lifted BD). Note that

$$z = \mathcal{G}_{DE}(M_0) = (Dx_0) \circledast (Ey_0) = \sqrt{n} F^*[(FDx_0) \odot (FEy_0)],$$

$$\begin{aligned}\frac{1}{\sqrt{n}}(Fz)^{(j)} &= (FD)^{(j,:)} x_0 (FE)^{(j,:)} y_0 \\ &= (FD)^{(j,:)} x_0 y_0^T (FE)^{(j,:)T} = a_j^* M_0 \overline{b_j},\end{aligned}$$

where $a_j = (FD)^{(j,:)*}$ is the conjugate transpose of the jth row of FD, and $b_j = (FE)^{(j,:)*}$ is the conjugate transpose of the jth row of FE. Rewriting (Lifted BD) in the frequency domain:

$$\begin{aligned}(\text{Lifted BD})_f \quad &\text{Find} \quad M, \\ &\text{s.t.} \quad a_j^* M \overline{b_j} = \frac{1}{\sqrt{n}}(Fz)^{(j)}, \quad 1 \le j \le n, \\ &M \in \Omega_{\mathcal{M}} = \{xy^T : x \in \Omega_{\mathcal{X}}, y \in \Omega_{\mathcal{Y}}\}.\end{aligned}$$

Clearly, the constraint set $\Omega_{\mathcal{M}}$ is a cone. Since $a_j = (FD)^{(j,:)*}$ and $b_j = (FE)^{(j,:)*}$, there exists a bijection between the pair $(D, E) \in \mathbb{C}^{n\times m_1} \times \mathbb{C}^{n\times m_2}$ and the set of vector pairs $\{a_j \in \mathbb{C}^{m_1}, b_j \in \mathbb{C}^{m_2}\}_{j=1}^n$. By Corollary 2.3.6, the recovery of M_0 is unique for almost all $D \in \mathbb{C}^{n\times m_1}$ and $E \in \mathbb{C}^{n\times m_2}$ if $2n > \underline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{B}})$. By Corollary 2.3.7, the recovery of all matrices in $\Omega_{\mathcal{M}}$ is unique for almost all $D \in \mathbb{C}^{n\times m_1}$ and $E \in \mathbb{C}^{n\times m_2}$ if $n > \overline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{B}})$.

Hence, Theorems 2.2.1 and 2.2.2 follow from the upper bounds on Minkowski dimensions in Lemma 2.4.1.

Lemma 2.4.1. *The upper Minkowski dimensions of* $\Omega_{\mathcal{B}} = \Omega_{\mathcal{M}} \bigcap \mathcal{B}_{\mathbb{C}^{m_1 \times m_2}}$ *in (Lifted BD) with subspace, mixed, and sparsity constraints are bounded by* $2(m_1 + m_2)$*,* $2(s_1 + m_2)$*, and* $2(s_1 + s_2)$*, respectively.*

Proof of Lemma 2.4.1. For simplicity, we only prove the upper bound for the mixed constraint set. The bounds for the other two scenarios can be proved in a similar fashion. First of all,

$$\begin{aligned}\Omega_{\mathcal{B}} &= \{xy^T : x \in \Omega_{\mathcal{X}}, y \in \Omega_{\mathcal{Y}}, \left\|xy^T\right\|_{\mathrm{F}} \leq 1\} \\ &= \{xy^T : x \in \Omega_{\mathcal{X}}, y \in \Omega_{\mathcal{Y}}, \|x\|_2 \leq 1, \|y\|_2 \leq 1\} \\ &= \{xy^T : x \in \Omega_{\mathcal{X}} \bigcap \mathcal{B}_{\mathbb{C}^{m_1}}, y \in \Omega_{\mathcal{Y}} \bigcap \mathcal{B}_{\mathbb{C}^{m_2}}\}.\end{aligned}$$

By Lemmas A.2.2 and A.2.3, we have

$$\begin{aligned}&\overline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}} \bigcap \mathcal{B}_{\mathbb{C}^{m_1 \times m_2}}) \\ &\leq \overline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{X}} \bigcap \mathcal{B}_{\mathbb{C}^{m_1}}) + \overline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{Y}} \bigcap \mathcal{B}_{\mathbb{C}^{m_2}}) \\ &\leq \overline{\dim}_{\mathrm{B}}\left(\mathrm{Re}\left(\Omega_{\mathcal{X}} \bigcap \mathcal{B}_{\mathbb{C}^{m_1}}\right)\right) + \overline{\dim}_{\mathrm{B}}\left(\mathrm{Im}\left(\Omega_{\mathcal{X}} \bigcap \mathcal{B}_{\mathbb{C}^{m_1}}\right)\right) \\ &+ \overline{\dim}_{\mathrm{B}}\left(\mathrm{Re}\left(\Omega_{\mathcal{Y}} \bigcap \mathcal{B}_{\mathbb{C}^{m_2}}\right)\right) + \overline{\dim}_{\mathrm{B}}\left(\mathrm{Im}\left(\Omega_{\mathcal{Y}} \bigcap \mathcal{B}_{\mathbb{C}^{m_2}}\right)\right). \qquad (2.9)\end{aligned}$$

Recall that, in the mixed constraints scenario, the filter satisfies a subspace constraint, and $\Omega_{\mathcal{Y}} = \mathbb{C}^{m_2}$. The restriction to the unit ball is $\Omega_{\mathcal{Y}} \bigcap \mathcal{B}_{\mathbb{C}^{m_2}} = \mathcal{B}_{\mathbb{C}^{m_2}}$, whose real and imaginary parts are $\mathcal{B}_{\mathbb{R}^{m_2}}$. By a standard volume argument (see [88, Lemma 4.1]),

$$N_{\mathcal{B}_{\mathbb{R}^{m_2}}}(\rho) \leq \left(\frac{3}{\rho}\right)^{m_2}, \qquad \forall \rho \leq 1. \qquad (2.10)$$

Hence

$$\begin{aligned}&\overline{\dim}_{\mathrm{B}}\left(\mathrm{Re}\left(\Omega_{\mathcal{Y}} \bigcap \mathcal{B}_{\mathbb{C}^{m_2}}\right)\right) = \overline{\dim}_{\mathrm{B}}\left(\mathrm{Im}\left(\Omega_{\mathcal{Y}} \bigcap \mathcal{B}_{\mathbb{C}^{m_2}}\right)\right) \\ &= \overline{\dim}_{\mathrm{B}}(\mathcal{B}_{\mathbb{R}^{m_2}}) = \limsup_{\rho \to 0} \frac{\log N_{\mathcal{B}_{\mathbb{R}^{m_2}}}(\rho)}{\log \frac{1}{\rho}} \leq \limsup_{\rho \to 0}\, m_2 \frac{\log \frac{3}{\rho}}{\log \frac{1}{\rho}} = m_2. \qquad (2.11)\end{aligned}$$

Meanwhile, the signal satisfies a sparsity constraint, and $\Omega_{\mathcal{X}} = \{x \in \mathbb{C}^{m_1} : \|x\|_0 \leq s_1\}$. The restriction to the unit ball is $\Omega_{\mathcal{X}} \bigcap \mathcal{B}_{\mathbb{C}^{m_1}} = \{x \in \mathbb{C}^{m_1} :$

$\|x\|_0 \leq s_1, \|x\|_2 \leq 1\}$, whose real and imaginary parts are

$$\begin{aligned}\mathrm{Re}\left(\Omega_{\mathcal{X}} \bigcap \mathcal{B}_{\mathbb{C}^{m_1}}\right) &= \mathrm{Im}\left(\Omega_{\mathcal{X}} \bigcap \mathcal{B}_{\mathbb{C}^{m_1}}\right) \\ &= \{x \in \mathbb{R}^{m_1} : \|x\|_0 \leq s_1, \|x\|_2 \leq 1\},\end{aligned}$$

which is the union of unit balls in s_1-dimensional subspaces. Denote this set by $\Gamma_{s_1,1}^{m_1}$. By a standard volume argument,

$$N_{\Gamma_{s_1,1}^{m_1}}(\rho) \leq \binom{m_1}{s_1}\left(\frac{3}{\rho}\right)^{s_1} \leq \left(\frac{em_1}{s_1}\right)^{s_1}\left(\frac{3}{\rho}\right)^{s_1}, \quad \forall \rho \leq 1,$$

where the second inequality follows from Stirling's approximation. Hence

$$\begin{aligned}&\overline{\dim}_{\mathrm{B}}\left(\mathrm{Re}\left(\Omega_{\mathcal{X}} \bigcap \mathcal{B}_{\mathbb{C}^{m_1}}\right)\right) = \overline{\dim}_{\mathrm{B}}\left(\mathrm{Im}\left(\Omega_{\mathcal{X}} \bigcap \mathcal{B}_{\mathbb{C}^{m_1}}\right)\right) \\ &= \overline{\dim}_{\mathrm{B}}(\Gamma_{s_1,1}^{m_1}) = \limsup_{\rho \to 0} \frac{\log N_{\Gamma_{s_1,1}^{m_1}}(\rho)}{\log \frac{1}{\rho}} \leq \limsup_{\rho \to 0}\, s_1 \frac{\log \frac{1}{\rho} + \log \frac{3em_1}{s_1}}{\log \frac{1}{\rho}} = s_1.\end{aligned} \tag{2.12}$$

Combining (2.9), (2.11), and (2.12), we have that the upper Minkowski dimension of the mixed constraint set is bounded by $2(s_1 + m_2)$. □

2.4.2 Proof of Theorem 2.2.3

Next, we prove Theorem 2.2.3, which establishes results corresponding to those of Theorems 2.2.1 and 2.2.2 in the case where D, E, x, and y are real. When D are E are real matrices, $a_j = (FD)^{(j,:)*}$ and $b_j = (FE)^{(j,:)*}$ are complex vectors, but they are no longer generic. Therefore, Corollaries 2.3.6 and 2.3.7 cannot be applied directly to this case.

Proof of Theorem 2.2.3. By (2.9) in the proof of Theorem 2.4.1, when x, y, and $M = xy^T$ are real, the Minkowski dimensions of the restricted constraint sets are half those in Theorem 2.4.1. For subspace, mixed, and sparsity constraints, the upper Minkowski dimensions of the restricted constraint sets are bounded by $m_1 + m_2$, $s_1 + m_2$, and $s_1 + s_2$, respectively. To maintain the same sample complexities, we need to show a result analogous to Theorem 2.3.2, in which $a_j = (FD)^{(j,:)*}$ and $b_j = (FE)^{(j,:)*}$, D and E are real matrices, and $n > \underline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}})$ is sufficient.

Lemma 2.4.2. *Suppose $\Omega_{\mathcal{M}} \subset \mathbb{R}^{m_1 \times m_2}$ is a nonempty bounded set. Let $D \in \mathbb{R}^{n \times m_1}$ and $E \in \mathbb{R}^{n \times m_2}$, $a_j = (FD)^{(j,:)*}$ and $b_j = (FE)^{(j,:)*}$ $(j = 1, 2, \cdots, n)$. For almost all $D \in \mathbb{R}^{n \times m_1}$ and $E \in \mathbb{R}^{n \times m_2}$, there does not exist a matrix $M \in \Omega_{\mathcal{M}} \backslash \{0\}$ such that $\left\langle a_j b_j^T, M \right\rangle = a_j^* M \overline{b_j} = 0$ for $j = 1, 2, \cdots, n$, if $n > \underline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}})$.*

The proof of Lemma 2.4.2 is very similar to that of Theorem 2.3.2. In fact, the only difference is the following: the mapping between the real matrices D, E and the complex vectors $\{a_j\}_{j=1}^n$, $\{b_j\}_{j=1}^n$ is no longer a bijection. The vectors a_1 and b_1 are real vectors. Due to the conjugate symmetry of DFT, the vectors a_j and a_{n+2-j} is a conjugate pairs, i.e. $a_j = \overline{a_{n+2-j}}$. The same is true for b_j and b_{n+2-j}. Therefore, (roughly) the first half of the DFT measurements contain all the information of real-valued unknowns. There exists a bijection between D, E and the vectors $\{a_j\}_{j=1}^{\lceil (n+1)/2 \rceil}$, $\{b_j\}_{j=1}^{\lceil (n+1)/2 \rceil}$.

Due to this subtlety, in the probabilistic argument (analogous to Lemma 2.3.3) we assume $\{a_j\}_{j=1}^{\lceil (n+1)/2 \rceil}$, $\{b_j\}_{j=1}^{\lceil (n+1)/2 \rceil}$ are independent random vectors as follows:

- When n is even, $\{a_1,\ a_{\frac{n}{2}+1}\}$ and $\{b_1,\ b_{\frac{n}{2}+1}\}$ are real random vectors following uniform distributions on $R\mathcal{B}_{\mathbb{R}^{m_1}}$ and $R\mathcal{B}_{\mathbb{R}^{m_2}}$, respectively. The vectors $\{a_j\}_{j=2}^{\frac{n}{2}}$ and $\{b_j\}_{j=2}^{\frac{n}{2}}$ are complex random vectors following uniform distributions on $R\mathcal{B}_{\mathbb{C}^{m_1}}$ and $R\mathcal{B}_{\mathbb{C}^{m_2}}$, respectively.

- When n is odd, a_1 and b_1 are real random vectors following uniform distributions on $R\mathcal{B}_{\mathbb{R}^{m_1}}$ and $R\mathcal{B}_{\mathbb{R}^{m_2}}$, respectively. The vectors $\{a_j\}_{j=2}^{\frac{n+1}{2}}$ and $\{b_j\}_{j=2}^{\frac{n+1}{2}}$ are complex random vectors following uniform distributions on $R\mathcal{B}_{\mathbb{C}^{m_1}}$ and $R\mathcal{B}_{\mathbb{C}^{m_2}}$, respectively.

We apply corresponding changes to the proof of Lemma 2.3.3. (As before, we define $\delta = \rho R^2$.) When bounding the probability P_L, (2.7) and (2.8) now become:

- When n is even,

$$\begin{aligned} P_L &\le \sum_{i=1}^{N_{\Omega_{\mathcal{M},L}}(\rho)} \prod_{j=1}^{\frac{n}{2}+1} \mathbb{P}\left[\left|a_j^* M'_{\rho,L,i} \overline{b_j}\right| \le 3\delta\right] \\ &\le N_{\Omega_{\mathcal{M}}}\left(\frac{\delta}{R^2}\right) \left(3\delta f(3\delta, \frac{1}{L}, L, R)\right)^2 \cdot \left((3\delta)^2 g(3\delta, \frac{1}{L}, L, R)\right)^{\frac{n}{2}-1} \\ &= N_{\Omega_{\mathcal{M}}}\left(\frac{\delta}{R^2}\right) (3\delta)^n f(3\delta, \frac{1}{L}, L, R)^2 \cdot g(3\delta, \frac{1}{L}, L, R)^{\frac{n}{2}-1}. \end{aligned}$$

- When n is odd,

$$\begin{aligned} P_L &\le \sum_{i=1}^{N_{\Omega_{\mathcal{M},L}}(\rho)} \prod_{j=1}^{\frac{n+1}{2}} \mathbb{P}\left[\left|a_j^* M'_{\rho,L,i} \overline{b_j}\right| \le 3\delta\right] \\ &\le N_{\Omega_{\mathcal{M}}}\left(\frac{\delta}{R^2}\right) \left(3\delta f(3\delta, \frac{1}{L}, L, R)\right) \cdot \left((3\delta)^2 g(3\delta, \frac{1}{L}, L, R)\right)^{\frac{n-1}{2}} \\ &= N_{\Omega_{\mathcal{M}}}\left(\frac{\delta}{R^2}\right) (3\delta)^n f(3\delta, \frac{1}{L}, L, R) \cdot g(3\delta, \frac{1}{L}, L, R)^{\frac{n-1}{2}}. \end{aligned}$$

Whether n is even or odd, we have $P_L = O\big(N_{\Omega_{\mathcal{M}}}\left(\frac{\delta}{R^2}\right)\delta^n\big)$. By the same argument as in the proof of Lemma 2.3.3, the sample complexity is $n > \underline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}})$. □

2.4.3 Proof of Theorem 2.2.4

In this section, we establish the stability results in blind deconvolution. The measurement in (Noisy BD) can be rewritten in the frequency domain:

$$\tilde{z}^{(j)} := \frac{1}{\sqrt{n}}(Fz)^{(j)} = (FD)^{(j,:)} x_0 (FE)^{(j,:)} y_0 + \frac{1}{\sqrt{n}}(F\xi)^{(j)} = a_j^* M_0 \overline{b_j} + \tilde{\xi}^{(j)},$$

where $M_0 = x_0 y_0^T$, $a_j = (FD)^{(j,:)*}$, $b_j = (FE)^{(j,:)*}$, and $\tilde{\xi} = \frac{1}{\sqrt{n}} F\xi$. Define linear operator $\mathcal{A}(M)$ by $\mathcal{A}(M) = \left[a_1^* M \overline{b_1}, a_2^* M \overline{b_2}, \cdots, a_n^* M \overline{b_n}\right]^T$. We rewrite (Noisy BD) in the frequency domain:

$$\begin{aligned} (\text{Noisy BD})_f \qquad \min_M . \quad & \left\|\mathcal{A}(M) - \tilde{z}\right\|_2, \\ \text{s.t.} \quad & M \in \sigma\Omega_{\mathcal{B}}, \end{aligned}$$

where $\sigma\Omega_{\mathcal{B}} = \{xy^T : x \in \Omega_{\mathcal{X}}, y \in \Omega_{\mathcal{Y}}, \left\|xy^T\right\|_{\mathrm{F}} \leq \sigma\}$.

Note that

$$\mathcal{A}(M) = \frac{1}{\sqrt{n}} F\mathcal{G}_{DE}(M),$$
$$\left\|\mathcal{A}(M)\right\|_2 = \frac{1}{\sqrt{n}} \left\|\mathcal{G}_{DE}(M)\right\|_2 .$$

The single point stability result in the subspace constraints scenario in Theorem 2.2.4 follows from Lemma 2.4.3 , with every δ replaced by $\frac{\delta}{\sqrt{n}}$. All other cases can be proved using similar lemmas, which we omit here for brevity.

Lemma 2.4.3. *In (Noisy BD)$_f$ with subspace constraints, assume that the random vectors $\{a_j\}_{j=1}^n$ are i.i.d. following a uniform distribution on $R\mathcal{B}_{\mathbb{C}^{m_1}}$, and $\{b_j\}_{j=1}^n$ are i.i.d. following a uniform distribution on $R\mathcal{B}_{\mathbb{C}^{m_2}}$. Let the true matrix be $M_0 \in \Omega_{\mathcal{B}} = \Omega_{\mathcal{M}} \bigcap \mathcal{B}_{\mathbb{C}^{m_1 \times m_2}} = \{xy^T : x \in \mathcal{B}_{\mathbb{C}^{m_1}}, y \in \mathcal{B}_{\mathbb{C}^{m_2}}\}$. If $n > m_1 + m_2$ and $\delta \leq R^2$, then with probability at least*

$$1 - \left(648\ m_1 m_2 \left(1 + 2\ln\frac{2R^2}{3\delta}\right)\right)^n \left(\frac{\delta^2}{R^4}\right)^{n-m_1-m_2} \left(\frac{1}{\varepsilon^2}\right)^n,$$

for all $M \in \Omega_{\mathcal{B}}$ such that $\left\|\mathcal{A}(M) - \mathcal{A}(M_0)\right\|_2 \leq \delta$, we have $\left\|M - M_0\right\|_2 \leq \varepsilon$.

To ensure that the probability bound is nontrivial, we insist that

$$\epsilon > \sqrt{648\ m_1 m_2 \left(1 + 2\ln\frac{2R^2}{3\delta}\right)} \left(\frac{\delta}{R^2}\right)^{\frac{n-m_1-m_2}{n}}.$$

Since the right-hand size vanishes as δ approaches 0, the above lemma guarantees stable recovery in (Noisy BD)$_f$. Next, we prove this lemma, exploiting a key result in the proof of Lemma 2.3.3.

Proof of Lemma 2.4.3. We need to bound the following probability of stability:

$$\begin{aligned}
P_s :=& \mathbb{P}\Big[\forall M \in \Omega_{\mathcal{B}},\ \text{if}\ \left\|\mathcal{A}(M) - \mathcal{A}(M_0)\right\|_2 \leq \delta,\ \text{then}\ \left\|M - M_0\right\|_2 \leq \varepsilon\Big] \\
=& 1 - \mathbb{P}\Big[\exists M \in \Omega_{\mathcal{B}},\ \text{s.t.}\ \left\|\mathcal{A}(M) - \mathcal{A}(M_0)\right\|_2 \leq \delta,\ \text{and}\ \left\|M - M_0\right\|_2 > \varepsilon\Big] \\
=& 1 - \mathbb{P}\Big[\exists M \in \Omega_{\mathcal{B}} - M_0,\ \text{s.t.}\ \left\|M\right\|_2 > \varepsilon \text{ and } \left\|\mathcal{A}(M)\right\|_2 \leq \delta\Big] \\
=:& 1 - P_f,
\end{aligned}$$

where the probability of failure P_f satisfies:

$$
\begin{aligned}
P_f =& \mathbb{P}\Big[\exists M \in \Omega_{\mathcal{B}} - M_0, \text{ s.t. } \|M\|_2 > \varepsilon \text{ and } \|\mathcal{A}(M)\|_2 \leq \delta\Big] \\
\leq& \mathbb{P}\Big[\exists M \in \Omega_{\mathcal{B}} - M_0, \text{ s.t. } \|M\|_2 > \varepsilon \text{ and } |a_j^* M \overline{b_j}| \leq \delta, j = 1, 2, \cdots, n\Big] \\
\leq& N_{\Omega_{\mathcal{B}}}\left(\frac{\delta}{R^2}\right) (3\delta)^{2n} g(3\delta, \varepsilon, 2, R)^n \qquad (2.13) \\
\leq& \left(\frac{6\sqrt{2}R^2}{\delta}\right)^{2m_1+2m_2} (3\delta)^{2n} \\
& \cdot \left(\frac{\pi^2 \cdot V_{\mathbb{C}^{m_1-1}}(R) \cdot V_{\mathbb{C}^{m_2-1}}(R)}{\varepsilon^2 \cdot V_{\mathbb{C}^{m_1}}(R) \cdot V_{\mathbb{C}^{m_2}}(R)} \left(1 + 2\ln\frac{2R^2}{3\delta}\right)\right)^n \qquad (2.14) \\
=& \left(\frac{6\sqrt{2}R^2}{\delta}\right)^{2m_1+2m_2} (3\delta)^{2n} \cdot \left(\frac{m_1 m_2}{\varepsilon^2 R^4} \left(1 + 2\ln\frac{2R^2}{3\delta}\right)\right)^n \qquad (2.15) \\
\leq& \left(\frac{R^2}{\delta}\right)^{2m_1+2m_2} (6\sqrt{2})^{2n} (3\delta)^{2n} \cdot \left(\frac{m_1 m_2}{\varepsilon^2 R^4} \left(1 + 2\ln\frac{2R^2}{3\delta}\right)\right)^n \\
=& \left(\frac{\delta^2}{R^4}\right)^{n-m_1-m_2} \left(\frac{1}{\varepsilon^2}\right)^n \cdot \left(648\ m_1 m_2 \left(1 + 2\ln\frac{2R^2}{3\delta}\right)\right)^n.
\end{aligned}
$$

Inequality (2.13) follows from (2.8), with the norm bounds $\varepsilon < \|M\|_2 \leq 2$. In (2.14), the bound on the covering number of $\Omega_{\mathcal{B}} = \Omega_{\mathcal{M}} \bigcap \mathcal{B}_{\mathbb{C}^{m_1 \times m_2}} = \{xy^T : x \in \mathcal{B}_{\mathbb{C}^{m_1}}, y \in \mathcal{B}_{\mathbb{C}^{m_2}}\}$ is derived as follows:

$$
\begin{aligned}
& N_{\Omega_{\mathcal{B}}}\left(\frac{\delta}{R^2}\right) \leq N_{\mathcal{B}_{\mathbb{C}^{m_1}}}\left(\frac{\delta}{2R^2}\right) N_{\mathcal{B}_{\mathbb{C}^{m_2}}}\left(\frac{\delta}{2R^2}\right) \\
& \leq \left(N_{\mathcal{B}_{\mathbb{R}^{m_1}}}\left(\frac{\delta}{2\sqrt{2}R^2}\right)\right)^2 \left(N_{\mathcal{B}_{\mathbb{R}^{m_2}}}\left(\frac{\delta}{2\sqrt{2}R^2}\right)\right)^2 \\
& \leq \left(\left(\frac{6\sqrt{2}R^2}{\delta}\right)^{m_1}\right)^2 \left(\left(\frac{6\sqrt{2}R^2}{\delta}\right)^{m_2}\right)^2 = \left(\frac{6\sqrt{2}R^2}{\delta}\right)^{2m_1+2m_2},
\end{aligned}
$$

where the first two inequalities follow from (A.7), (A.8) in Appendix A.2, and the third inequality follows from (2.10) and the assumption $\delta \leq R^2$. The expression for $g(3\delta, \frac{1}{L}, 2, R)$ is given by (A.6) in Appendix A.1. Recall that $V_{\mathbb{C}^{m_1}}(R)$ denotes the volume of a ball of radius R in $\mathbb{C}^{m_1}$. Equation (2.15) follows from the fact that $V_{\mathbb{C}^m}(R) = V_{\mathbb{R}^{2m}}(R) = \frac{\pi^m R^{2m}}{m!}$. That completes the proof. □

CHAPTER 3

IDENTIFIABILITY IN BLIND GAIN AND PHASE CALIBRATION

3.1 Problem Statement

3.1.1 Notations

We use upper-case letters A, X and Y to denote matrices, and lower-case letters to denote vectors. The diagonal matrix with the elements of vector λ on the diagonal is denoted by $\mathrm{diag}(\lambda)$. The vector formed by a concatenation of the columns of X is denoted by $\mathrm{vec}(X)$. Matrices I_n and F_n denote the identity matrix and the discrete Fourier transform (DFT) matrix of size $n \times n$. Unless otherwise stated, all vectors are column vectors. The dimensions of all vectors and matrices are made clear in the context. The circular convolution is denoted by $\circledast$. The Kronecker product of two matrices is denoted by $\otimes$. The entrywise product is denoted by $\odot$. The range space of the conjugate transpose of a matrix D is denoted by $\mathcal{R}^*(D) = \mathcal{R}(D^*)$, and the nullspace of D is denoted by $\mathcal{N}(D)$. The orthogonal complement of a subspace $\mathcal{V}$ is denoted by $\mathcal{V}^\perp$. Given a vector $x \in \mathbb{C}^n$, $\mathrm{span}(x)$ denotes the one-dimensional subspace of $\mathbb{C}^n$ spanned by x, and $x^\perp$ denotes its orthogonal complement.

We use j, k to denote indices, and J, K to denote index sets. If a matrix or a vector has dimension n, then an index set J is a subset of $\{1, 2, \ldots, n\}$. The cardinality of J is denoted by $|J|$, and the complement of J is denoted by J^c. Superscript letters denote subvectors or submatrices. For example, the submatrix $A^{(J,K)}$ has size $|J| \times |K|$ and consists of the entries indexed by $J \times K$. Borrowing the colon notation from MATLAB, the vector $A^{(:,k)}$ represents the kth column of matrix A.

3.1.2 The BGPC Problem

Blind gain and phase calibration (BGPC) is the following constrained bilinear inverse problem given the measurement $Y = \mathrm{diag}(\lambda_0)\Phi_0$:

$$\begin{aligned} \text{Find} \quad & (\lambda, \Phi), \\ \text{s.t.} \quad & \mathrm{diag}(\lambda)\Phi = Y, \\ & \lambda \in \Omega_\Lambda, \ \Phi \in \Omega_\Phi, \end{aligned}$$

where $\lambda \in \Omega_\Lambda \subset \mathbb{C}^n$ is the unknown gain and phase vector, and $\Phi \in \Omega_\Phi \subset \mathbb{C}^{n\times N}$ is the signal matrix. In this chapter, we impose no constraints on λ, i.e., $\Omega_\Lambda = \mathbb{C}^n$. As for the matrix Φ, we impose subspace or joint sparsity constraints. In both scenarios, Φ can be represented in the factorized form $\Phi = AX$, where the columns of $A \in \mathbb{C}^{n\times m}$ form a basis or a frame (an overcomplete dictionary), and $X \in \Omega_\mathcal{X} \subset \mathbb{C}^{m\times N}$ is the matrix of coordinates. The constraint set becomes $\Omega_\Phi = \{\Phi = AX : X \in \Omega_\mathcal{X}\}$. Under some mild conditions[1] on A, the uniqueness of Φ is equivalent to the uniqueness of X. For simplicity, we treat the following problem as the BGPC problem from now on.

$$\begin{aligned} \text{(BGPC)} \qquad \text{Find} \quad & (\lambda, X), \\ \text{s.t.} \quad & \mathrm{diag}(\lambda)AX = Y, \\ & \lambda \in \mathbb{C}^n, \ X \in \Omega_\mathcal{X}. \end{aligned}$$

We consider two scenarios in this chapter:

(1) *Subspace constraints:* The signals represented by the columns of Φ reside in a low-dimensional subspace spanned by the columns of a known matrix A. The matrix A is tall ($n > m$) and has full column rank. The constraint set is $\Omega_\mathcal{X} = \mathbb{C}^{m\times N}$.

(2) *Joint sparsity constraints:* The columns of Φ are jointly sparse over a known dictionary A, where A is a square matrix ($n = m$) or a fat matrix ($n < m$). The constraint set $\Omega_\mathcal{X}$ is

$$\Omega_\mathcal{X} = \{X \in \mathbb{C}^{m\times N} : X \text{ has at most } s \text{ nonzero rows}\}.$$

[1] Under a subspace constraint, A is required to have full column rank. Under a joint sparsity constraint, A is required to satisfy the spark condition [86]. Both conditions are satisfied by a generic A.

In other words, the columns of X are jointly s-sparse.

In the rest of this chapter, we address the identifiability in the above BGPC problem. In BGPC, the constraint sets Ω_Λ and $\Omega_\mathcal{X}$ are cones – they are closed under scalar multiplication. For any nonzero scalar σ, the pairs (λ_0, X_0) and $(\sigma\lambda_0, \frac{1}{\sigma}X_0)$ map to the same Y and therefore are non-distinguishable. This problem is said to suffer from scaling ambiguity. The set $\{(\sigma\lambda_0, \frac{1}{\sigma}X_0) : \sigma \neq 0\}$ is an equivalence class of solutions generated by a group of scaling transformations. The solution (λ_0, X_0) is said to be identifiable up to scaling if every solution to BGPC is a scaled version of (λ_0, X_0) in that equivalent class. In this chapter, we answer the following question: Under what conditions is the solution (λ_0, X_0) unique up to scaling?

Our results are stated in terms of sample complexities, which are the numbers of data samples or measurements needed for unique recovery of the solutions. They are given by inequalities describing the conditions that need to be satisfied by the problem parameters, n, m, s, and N. The numbers n and m denote the length of the signals and the dimension of the subspace in which they are assume to reside, in the subspace constraint scenario. The sparsity level s is the number (out of m) of nonzero rows of X in the joint sparsity scenario. Finally, the number of signals captured (number of columns of Y and Φ) is denoted by N. Since it is often difficult to acquire a large number of signals, it is desirable to have sample complexities that requires small N. We defer the reader to Section 3.1.3 for a detailed discussion of these quantities in specific applications.

3.1.3 Applications

This section gives a detailed account of the applications of BGPC. Table 3.1 summarizes what parameters n, m, s, and N represent in these applications.

Subspace Constraint: In inverse rendering [29], the columns of $Y = \text{diag}(\lambda)\Phi$ represent images under different lighting conditions, where λ represents the unknown albedos,[2] and the columns of Φ represent the intensity maps of incident light under different conditions. The columns of A are the first several spherical harmonics extracted from the 3D model of the object. They form a basis of the low-dimensional subspace in which the intensity

[2]In inverse rendering, albedos are real and positive. We ignore this extra information here for simplicity.

maps reside.

Multichannel blind deconvolution with the circular convolution model also falls into this category. The measurement $Y^{(:,j)} = \text{diag}(\lambda)\Phi^{(:,j)}$ can be also written as:

$$F_n^{-1}Y^{(:,j)} = (F_n^{-1}\lambda) \circledast (F_n^{-1}\Phi^{(:,j)}).$$

The vector λ represents the DFT of the signal, the columns of Φ represent the DFTs of the impulse responses of the channels, and the columns of Y represent the DFTs of the channel outputs. The columns of $F_n^{-1}A$ form a basis for the low-dimensional subspace in which the impulse responses of the channels reside. For example, when the multiple channels are FIR filters that share the same support J, they reside in a low-dimensional subspace whose basis is $F_n^{-1}A = I^{(:,J)}$. By symmetry, the roles of signals and channels can be switched. In channel encoding, when multiple signals are encoded by the same tall matrix E, they reside in a low-dimensional subspace whose basis is $F_n^{-1}A = E$. In this case, the vector λ represents the DFT of the channel.

Joint Sparsity Constraint: In sensor array processing with uncalibrated sensors, the vector λ represents unknown gains and phases of the sensors, and the columns of Φ represent snapshots captured at different time instants, assuming unit gain and zero phase for all sensors. Consider a scene with radiating sources whose positions (directions of arrival in the far-field scenario) are discretized, using a grid of m positions. Then each column of $A \in \mathbb{C}^{n\times m}$ represents the array response to a single source at one position on the grid. With only $s < m$ unknown sources, each column of Φ is the superposition of the same s columns of A. It follows that the columns of the source matrix X have a common support determined by the source positions, and are jointly s-sparse.

If the impulse responses in multichannel blind deconvolution are jointly sparse over the dictionary $F_n^{-1}A$, then as argued in the subspace constraints case, the vector λ, the columns of Φ, and the columns of Y represent the DFTs of the signal, the impulse responses, and the channel outputs, respectively. By symmetry, the roles of signals and channels can be switched. For example, in hyperspectral imaging, image samples at different frequencies in the light spectrum are likely to share the same discontinuities, and be jointly sparse over the same dictionary. If all image samples are corrupted with the same blurring kernel, then the deblurring procedure is a BGPC problem with

Table 3.1: Physical meanings of the problem parameters in applications.

	Inverse Rendering	Sensor Array Processing	Multichannel Blind Deconvolution
n	# pixels	# sensors	Length of the signal
m	# spherical harmonics	# positions on the grid	Dimension of the channel subspace *(subspace constraint)*
s	————	# sources	Channel sparsity level *(joint sparsity constraint)*
N	# images	# snapshots	# channels

joint sparsity constraints.

SAR autofocus [38] is a special multichannel blind deconvolution problem, where X represents the SAR image and $A = F$ is the 1D DFT matrix. The entries in λ represent the phase error in the Fourier imaging data, which varies only along the cross-range dimension.[3] If the coverage of the image is extended by oversampling the Fourier domain in the cross-range dimension, the rows of the image X corresponding to the region that is not illuminated by the antenna beam will be composed of zeros. Thus, the SAR image X can be modeled as a matrix with jointly sparse columns.

3.2 Main Results

3.2.1 BGPC with a Subspace Constraint

We first consider identifiability in BGPC with a subspace constraint. The measurement in the following problem is $Y = \text{diag}(\lambda_0)AX_0$. The known matrix $A \in \mathbb{C}^{n\times m}$ is tall ($n > m$), and therefore the columns of $\Phi = AX$ reside in a low-dimensional subspace. The corresponding constraint sets are $\Omega_\Lambda = \mathbb{C}^n$ and $\Omega_\mathcal{X} = \mathbb{C}^{m\times N}$, hence the problem is unconstrained with respect to λ and X.

In a previous work [1], we showed that $N \geq m$ is sufficient to guarantee

[3] In SAR autofocus, the entries of the phase error λ have unit moduli. We ignore this extra information here for simplicity.

identifiability when A, λ_0, and X_0 are generic. However, numerical experiments show that when $\frac{n-1}{n-m} \leq N \leq m$, the solution can still be identifiable (see [1, Section 3.3]). In this section, we explore the regime where λ_0, X_0, and A are generic, and $\frac{n-1}{n-m} \leq N \leq m$. We prove the following sufficient condition for the identifiability of (λ_0, X_0) up to scaling.

Theorem 3.2.1. *In the BGPC problem with a subspace constraint, if $n > m$ and $\frac{n-1}{n-m} \leq N \leq m$, then for almost all $\lambda_0 \in \mathbb{C}^n$, almost all $X_0 \in \mathbb{C}^{m\times N}$, and almost all $A \in \mathbb{C}^{n\times m}$, the pair (λ_0, X_0) is identifiable up to an unknown scaling.*

The sample complexity required by this theorem, $N \geq \frac{n-1}{n-m}$, is much less demanding than the condition $N \geq m$ in our previous results [1, Theorem 3.3 and Corollary 3.4]. In fact, this sample complexity is optimal, since it matches the sample complexity in the necessary condition [1, Proposition 3.5]. It suggests that if $m \leq \frac{n}{2}$, i.e., the dimension of the subspace is less than half the ambient dimension, then $N = 2$ signal vectors are sufficient to recover (λ_0, X_0) uniquely. This result provides a favorable bound for real-world applications. For example, the typical dimension of the intensity map subspace in inverse rendering is $m = 9$, which is really small when compared to the size of the images (e.g., $n = 256 \times 256 = 2^{16}$). Therefore, having two images under different lighting conditions is sufficient for the uniqueness of the solution. We will prove this result in Section 3.3.1.

When the sample complexity is achieved, for almost all λ_0, X_0, and A, the solution (λ_0, X_0) is unique up to scaling. In other words, this result is violated only for (λ_0, X_0, A) on a subset of $\mathbb{C}^n \times \mathbb{C}^{m\times N} \times \mathbb{C}^{n\times m}$ that has Lebesgue measure zero. If (λ_0, X_0, A) is random, following a distribution that is absolutely continuous with respect to the Lebesgue measure (e.g., the entries of λ_0, X_0, and A are i.i.d. following a Gaussian distribution), then the solution to BGPC is identifiable up to scaling with probability 1. Moreover, the degenerate set of (λ_0, X_0, A) that fails the test, is an algebraic variety, which is not dense in the ambient space. In real-world applications, λ_0 and AX_0 represent natural signals, which are not likely to belong to the particular lower-dimensional manifold of degeneracy.

As shown later in the proof of Theorem 3.2.1, the identifiability hinges on the following conditions:

1. There are no zero rows in AX_0, and all the entries of λ_0 are nonzero.

2. The matrix in (3.2), which is a function of A and X_0, has full column rank.

For a given combination of λ_0, X_0, and A, one can test whether the above conditions are satisfied. Figure 3.1 shows the test results of random λ_0, X_0, and A, whose entries are generated as i.i.d. Gaussian random variables $N(0,1)$. We fix $n = 20$, and check Conditions 1 and 2 for different values of (m, N). Here, white (resp. black) means that the conditions are satisfied (resp. are not satisfied). The red line represents the boundary $N = \frac{n-1}{n-m}$. The test results remain the same for 20 independent random experiments, which is consistent with the fact that Conditions 1 and 2 are satisfied for random Gaussian vectors and matrices with probability 1 when the sample complexity $N \geq \frac{n-1}{n-m}$ is achieved.

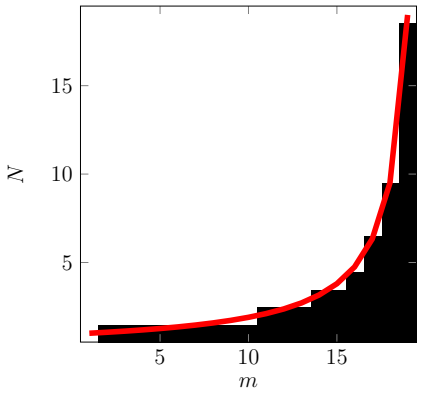


Figure 3.1: Verification of the identifiability conditions in 1 and 2 for random $(\lambda_0, X_0, A) \in \mathbb{C}^n \times \mathbb{C}^{m\times N} \times \mathbb{C}^{n\times m}$. The white (resp. black) region means that the conditions are satisfied (resp. are not satisfied) for all 20 trials. The red line represents the boundary $N = (n-1)/(n-m)$ corresponding to the necessary and sufficient condition in Theorem 3.2.1.

3.2.2 BGPC with a Joint Sparsity Constraint

Next, consider identifiability in BGPC with a joint sparsity constraint. The measurement is $Y = \mathrm{diag}(\lambda_0)AX_0$. The columns of $A \in \mathbb{C}^{n\times m}$ form a basis or frame for the signals. There are s nonzero rows in X_0, and the corresponding constraint set is

$$\Omega_{\mathcal{X}} = \{X \in \mathbb{C}^{m\times N} : X \text{ has at most } s \text{ nonzero rows}\}.$$

In a previous work [1], sufficient conditions for the uniqueness of the solution to the above problem were derived for some special cases (e.g., $A = F$). A sample complexity $N \geq s$ was established as sufficient for these special cases. However, when λ_0, X_0, and A are generic, a less demanding sufficient condition can be proved using essentially the same argument as in the proof of Theorem 3.2.1. The proof is presented in Section 3.3.3.

Theorem 3.2.2. *In the BGPC problem with a joint sparsity constraint, if* $n > 2s$ *and* $\frac{n-1}{n-2s} \leq N \leq s$*, then for almost all* $\lambda_0 \in \mathbb{C}^n$*, almost all* $X_0 \in \mathbb{C}^{m \times N}$ *with* s *nonzero rows, and almost all* $A \in \mathbb{C}^{n \times m}$*, the pair* (λ_0, X_0) *is identifiable up to an unknown scaling.*

The sample complexity $N \geq \frac{n-1}{n-2s}$ is far superior to the previous sufficient condition [1] of $N \geq s$, when the sparsity level s is much smaller than the ambient dimension n. For example, if $s \leq \frac{n}{4}$, then $N = 2$ is sufficient. In sensor array processing, the number of sources s is often much smaller than the number of sensors n. Therefore, only two snapshots are needed to recover the unknown gains and phases uniquely. This is especially significant when the working conditions of the sensor array and/or the source locations vary over time, and it needs to be re-calibrated continuously. One can achieve higher temporal resolution by solving BGPC using fewer snapshots.

Next, we compare the sample complexity in the sufficient condition of Theorem 3.2.2 to a necessary condition for this scenario. Suppose the support of X_0 is known, then the joint sparsity constraint reduces to a subspace constraint. By the necessary condition for the subspace scenario [1, Proposition 3.5], it follows that a necessary condition for the joint sparsity scenario is $N \geq \frac{n-1}{n-s}$. The sample complexity in the sufficient condition of Theorem 3.2.2, $N \geq \frac{n-1}{n-2s}$, nearly matches the necessary condition when the sparsity level s is small compared to signal length n. Indeed, when $n \geq 4s$, the above sufficient and necessary conditions both reduce to $N \geq 2$. When $n \geq 3s$, the necessary condition reduces to $N \geq 2$, and the sufficient condition reduces to $N \geq 3$, which is off by 1.

3.3 Proof of the Main Results

3.3.1 Proof of Theorem 3.2.1

First, BGPC is a bilinear inverse problem. Theorem 2.8 [1] stated equivalent conditions for identifiability in bilinear inverse problems up to some transformation groups. Specializing this result to the identifiability in BGPC up to scaling, we have the following lemma.

Lemma 3.3.1. *In BGPC, the pair* $(\lambda_0, X_0) \in \Omega_\Lambda \times \Omega_\mathcal{X}$ $(\lambda_0 \neq 0, X_0 \neq 0)$ *is identifiable up to scaling if and only if the following two conditions are met:*

1. *If* $\mathrm{diag}(\lambda_1)AX_1 = \mathrm{diag}(\lambda_0)AX_0$ *for some* $(\lambda_1, X_1) \in \Omega_\Lambda \times \Omega_\mathcal{X}$*, then* $X_1 = \sigma X_0$ *for some nonzero* σ*.*

2. *If* $\mathrm{diag}(\lambda_1)AX_0 = \mathrm{diag}(\lambda_0)AX_0$ *for some* $\lambda_1 \in \Omega_\Lambda$*, then* $\lambda_1 = \lambda_0$*.*

We first show that Condition 2 holds: that is, if X_0 is given, then the recovery of λ_0 is unique. Note that for almost all matrices $A \in \mathbb{C}^{n\times m}$ and $X_0 \in \mathbb{C}^{m\times N}$, there are no zero rows in the product AX_0. It follows that, if $\mathrm{diag}(\lambda_0)AX_0 = \mathrm{diag}(\lambda_1)AX_0$ for some $\lambda_1 \in \mathbb{C}^n$, then $\lambda_1 = \lambda_0$.

By Lemma 3.3.1, to complete the proof, we only need to show that Condition 1 also holds for generic λ_0, X_0, and A.[4] Suppose there exists (λ_1, X_1) such that $\mathrm{diag}(\lambda_0)AX_0 = \mathrm{diag}(\lambda_1)AX_1$. Consider the k-th row on both sides of the equation, which can be written as

$$(I_N \otimes A^{(k,:)})\mathrm{vec}(X_0)\lambda_0^{(k)} = (I_N \otimes A^{(k,:)})\mathrm{vec}(X_1)\lambda_1^{(k)}.$$

Now, for almost all λ_0, X_0, and A, the left-hand side is nonzero. Therefore λ_1 and X_1 are nonzero. It follows that

$$(I_N \otimes A^{(k,:)})\big(\mathrm{vec}(X_1) - \frac{\lambda_0^{(k)}}{\lambda_1^{(k)}}\mathrm{vec}(X_0)\big) = 0,$$

and hence,

$$\mathrm{vec}(X_1) \in \mathcal{N}(I_N \otimes A^{(k,:)}) + \mathrm{span}(\mathrm{vec}(X_0)).$$

[4] We use arguments similar to those used for the proof of [23, Theorem 4.2].

Next, we project $\mathrm{vec}(X_1)$ onto the orthogonal complement of $\mathrm{span}(\mathrm{vec}(X_0))$. It follows that

$$\begin{aligned} P_{\mathrm{vec}(X_0)^\perp}\mathrm{vec}(X_1) =& \mathrm{vec}(X_1) - P_{\mathrm{span}(\mathrm{vec}(X_0))}\mathrm{vec}(X_1) \\ & \in \mathcal{N}(I_N \otimes A^{(k,:)}) + \mathrm{span}(\mathrm{vec}(X_0)). \end{aligned}$$

For linear vector spaces $\mathcal{V}_1$ and $\mathcal{V}_2$, $\mathcal{V}_1 + \mathcal{V}_2 = (\mathcal{V}_1^\perp \bigcap \mathcal{V}_2^\perp)^\perp$. Using the fact that $\mathcal{N}(I_N \otimes A^{(k,:)})^\perp = \mathcal{R}^*(I_N \otimes A^{(k,:)})$, and $\mathrm{span}(\mathrm{vec}(X_0))^\perp = \mathrm{vec}(X_0)^\perp$, we have

$$P_{\mathrm{vec}(X_0)^\perp}\mathrm{vec}(X_1) \in \left(\mathcal{R}^*(I_N \otimes A^{(k,:)}) \bigcap \mathrm{vec}(X_0)^\perp\right)^\perp,$$

for $k = 1, 2, \ldots, n$. Taking note of the fact that

$$P_{\mathrm{vec}(X_0)^\perp}\mathrm{vec}(X_1) \in \mathrm{vec}(X_0)^\perp,$$

we have

$$P_{\mathrm{vec}(X_0)^\perp}\mathrm{vec}(X_1) \in \mathrm{vec}(X_0)^\perp \bigcap \left(\bigcap_{k=1,2,\ldots,n} \left(\mathcal{R}^*(I_N \otimes A^{(k,:)}) \bigcap \mathrm{vec}(X_0)^\perp\right)^\perp \right). \tag{3.1}$$

Since

$$I_N \otimes A^{(k,:)} = \begin{bmatrix} A^{(k,:)} & 0 & 0 & \ldots & 0 \\ 0 & A^{(k,:)} & 0 & \ldots & 0 \\ 0 & 0 & A^{(k,:)} & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & A^{(k,:)} \end{bmatrix},$$

$$\mathrm{vec}(X_0)^* = \begin{bmatrix} X_0^{(:,1)*} & X_0^{(:,2)*} & X_0^{(:,3)*} & \ldots & X_0^{(:,N)*} \end{bmatrix},$$

it is easy to verify that, for almost all A and X_0, the intersection of the row space of $I_N \otimes A^{(k,:)}$ and the orthocomplement of $\mathrm{vec}(X_0)$ is an $(N-1)$-dimensional subspace:

$$\mathcal{R}^*(I_N \otimes A^{(k,:)}) \bigcap \mathrm{vec}(X_0)^\perp = \mathcal{R}^*\left(D(A^{(k,:)}, X_0)\right),$$

where the matrix $D(A^{(k,:)}, X_0) \in \mathbb{C}^{(N-1)\times mN}$ is a function of $A^{(k,:)}$ and X_0:

$$D(A^{(k,:)}, X_0) = \begin{bmatrix} -\gamma_2 & \gamma_1 & 0 & \dots & 0 \\ -\gamma_3 & 0 & \gamma_1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\gamma_N & 0 & 0 & \dots & \gamma_1 \end{bmatrix} \otimes A^{(k,:)},$$

and $\gamma_j = A^{(k,:)} X_0^{(:,j)}$ for $j = 1, 2, \dots, N$. For generic matrices A and X_0, $D(A^{(k,:)}, X_0)$ has full row rank, which is $N - 1$. By (3.1),

$$P_{\mathrm{vec}(X_0)^\perp} \mathrm{vec}(X_1) \in \mathcal{N} \left(\begin{bmatrix} \mathrm{vec}(X_0)^* \\ D(A^{(1,:)}, X_0) \\ D(A^{(2,:)}, X_0) \\ \vdots \\ D(A^{(n,:)}, X_0) \end{bmatrix} \right). \tag{3.2}$$

We have the following claim, to be proved in Section 3.3.2.

Claim 3.3.2. *For almost all X_0 and A, if $n > m$ and $\frac{n-1}{n-m} \le N \le m$, then the matrix in* (3.2) *has full column rank, which is mN.*

Given this claim, for almost all X_0 and A,

$$P_{\mathrm{vec}(X_0)^\perp} \mathrm{vec}(X_1) = 0.$$

Therefore, X_1 resides in the 1-dimensional subspace in $\mathbb{C}^{m\times N}$ spanned by X_0, i.e., $X_1 = \sigma X_0$. Recall that X_1 is nonzero, hence $\sigma \neq 0$, establishing Condition 2 in Lemma 3.3.1, thus proving Theorem 3.2.1.

3.3.2 Proof of Claim 3.3.2

We prove that the matrix in (3.2) has full column rank for almost all X_0 and A that satisfy $n > m$ and $\frac{n-1}{n-m} \le N \le m$. By the definition of matrix $D(A^{(k,:)}, X_0)$, we have $D(A^{(k,:)}, X_0)\mathrm{vec}(X_0) = 0$. Hence the first row $\mathrm{vec}(X_0)^*$ is orthogonal to the rest of the rows in the matrix in (3.2). Therefore, we only need to show the rank of the following matrix is at least $mN - 1$ for

almost all X_0 and A:

$$D(A, X_0) = \begin{bmatrix} D(A^{(1,:)}, X_0) \\ D(A^{(2,:)}, X_0) \\ \vdots \\ D(A^{(n,:)}, X_0) \end{bmatrix} \in \mathbb{C}^{n(N-1) \times mN}.$$

Using the basic result in algebraic geometry that a polynomial function from $\mathbb{C}^n$ to $\mathbb{C}$ is either identically zero, or nonzero almost everywhere, it follows easily that the rank of $D(A, X_0)$ is at least $mN - 1$ for almost all A and X_0, if the rank is $mN - 1$ for at least one choice of A and X_0.[5] The rest of the proof is an explicit construction of A and X_0 that satisfies this rank condition.

The matrix $X_0 \in \mathbb{C}^{m \times N}$ is a tall matrix ($N \leq m$), hence one can choose X_0 as the first N columns of I_m. The matrix $A \in \mathbb{C}^{n \times m}$ is also tall ($n > m$), therefore one can choose A as a subset of m columns from F_n. The first N columns are $A^{(:,1:N)} = F_n^{(:,1:N)}$. We pick $m - N$ columns out of $F_n^{(:,N+1:n)}$ as $A^{(:,N+1:m)}$ in a manner such that there are no blocks of consecutive N columns except for the first N columns. To satisfy this condition, the columns $F_n^{(:,N+1)}$ and $F_n^{(:,n)}$ must not be picked.[6] This can be demonstrated by Figure 3.2. This can be done because $(n - m)N \geq n - 1$.

Given this choice of X_0 and A,

$$D(A^{(k,:)}, X_0) = \begin{bmatrix} -\alpha^{k-1} & 1 & 0 & \dots & 0 \\ -\alpha^{2(k-1)} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\alpha^{(N-1)(k-1)} & 0 & 0 & \dots & 1 \end{bmatrix} \otimes A^{(k,:)},$$

where $\alpha = e^{-\frac{2\pi\sqrt{-1}}{n}}$. One can view $D(A, X_0)$ as a block matrix with n blocks, one on top of the other. Each block itself is a block matrix with $(N-1) \times N$ blocks.

[5] Indeed, if $D(\widetilde{A}, \widetilde{X}_0)$ has rank $mN - 1$ for some $\widetilde{A}$ and $\widetilde{X}_0$, then there exists a subset K of $mN - 1$ columns such that $D(\widetilde{A}, \widetilde{X}_0)^{(:,K)}$ has full column rank. By [25, Lemma 1], for almost all A and X_0, we have $D(A, X_0)^{(:,K)}$ has full column rank, and hence $D(A, X_0)$ has rank at least $mN - 1$.

[6] Because of the circular nature of the DFT matrix, the first column and the last column of F_n are also considered "consecutive".

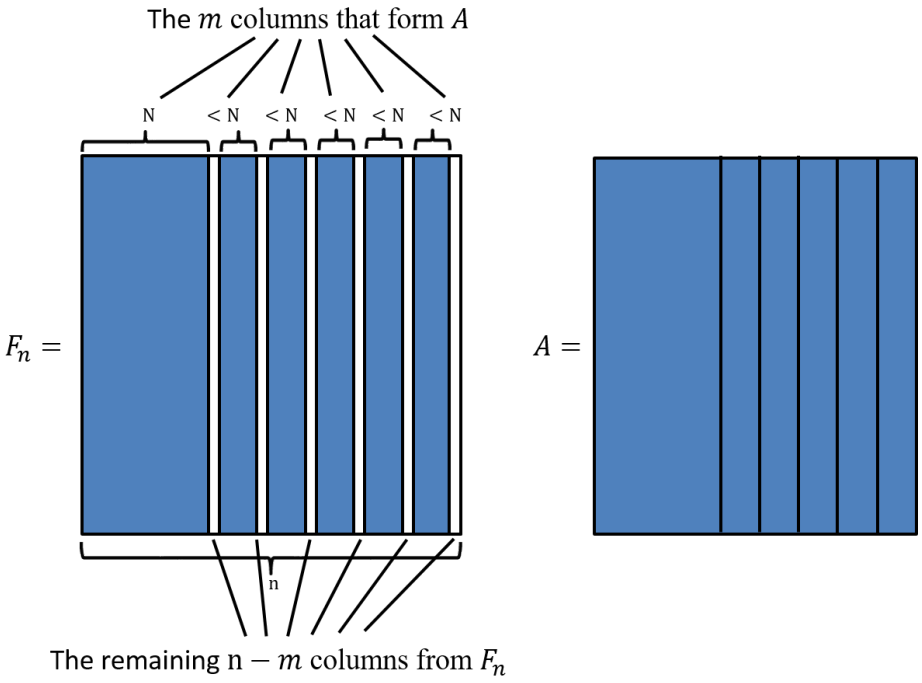


Figure 3.2: Construction of the matrix A from the DFT matrix F_n.

Consider the left null vector $w \in \mathbb{C}^{n(N-1)}$ of the matrix $D(A, X_0)$. Suppose

$$w = [w_{1,1}, w_{1,2}, \ldots, w_{1,N-1}, w_{2,1}, w_{2,2}, \ldots, w_{2,N-1}, \ldots, w_{n,1}, w_{n,2}, \ldots, w_{n,N-1}]^\top,$$

and $w^* D(A, X_0) = 0$. Then we have

$$\sum_{k=1}^{n} \left(\sum_{j=1}^{N-1} \alpha^{j(k-1)} \overline{w_{k,j}} \right) A^{(k,:)} = 0, \tag{3.3}$$

$$\sum_{k=1}^{n} \overline{w_{k,j}} A^{(k,:)} = 0, \text{ for } j = 1, 2, \ldots, N-1. \tag{3.4}$$

In order to show that $D(A, X_0)$ has rank $mN - 1$, we need to prove that there are exactly $M \coloneqq n(N-1) - (mN-1) = nN - mN - n + 1$ linearly independent left null vectors w. This number is greater than or equal to zero

because $N \geq \frac{n-1}{n-m}$. Consider the following matrix:

$$W = \begin{bmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,N-1} \\ w_{2,1} & w_{2,2} & \dots & w_{2,N-1} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n,1} & w_{n,2} & \dots & w_{n,N-1} \end{bmatrix}.$$

By (3.4), the columns of W are orthogonal to the columns of A. Recall that the columns of A are a subset of the columns of F_n. We use $A_\perp \in \mathbb{C}^{n\times(n-m)}$ to denote the matrix whose columns are the complement set of columns, i.e., the remaining $n-m$ columns in F_n that are not picked. Then $W = A_\perp Q$ for some $Q \in \mathbb{C}^{(n-m)\times(N-1)}$.

Next, we show that there are exactly M linearly independent matrices Q such that $W = A_\perp Q$ satisfies (3.3). Consider the following vector $v \in \mathbb{C}^n$ whose entries are the coefficients in (3.3):

$$\begin{aligned} v \coloneqq \begin{bmatrix} \sum_{j=1}^{N-1} \alpha^{-j\cdot 0} w_{1,j} \\ \sum_{j=1}^{N-1} \alpha^{-j\cdot 1} w_{2,j} \\ \vdots \\ \sum_{j=1}^{N-1} \alpha^{-j(n-1)} w_{n,j} \end{bmatrix} &= \sum_{j=1}^{N-1} F_n^{(:,n+1-j)} \odot W^{(:,j)} \\ = \sum_{i=1}^{n-m} \sum_{j=1}^{N-1} \big(F_n^{(:,n+1-j)} \odot A_\perp^{(:,i)}\big) Q^{(i,j)}. & \end{aligned} \tag{3.5}$$

By (3.3), v is also orthogonal to the columns in A. Therefore, there exists a vector $p \in \mathbb{C}^{n-m}$ such that

$$v = A_\perp p = \sum_{i=1}^{n-m} A_\perp^{(:,i)} p^{(i)}. \tag{3.6}$$

By (3.5) and (3.6), we have

$$\sum_{i=1}^{n-m} A_\perp^{(:,i)} p^{(i)} - \sum_{i=1}^{n-m} \sum_{j=1}^{N-1} \big(F_n^{(:,n+1-j)} \odot A_\perp^{(:,i)}\big) Q^{(i,j)} = v - v = 0. \tag{3.7}$$

The entrywise product of two columns in F_n is still a column in F_n. In particular, if $j_2 > j_1$, then $F_n^{(:,n+1-j_1)} \odot F_n^{(:,j_2)} = F_n^{(:,j_2-j_1)}$. Therefore, $\{F_n^{(:,n+1-j)} \odot A_\perp^{(:,i)}\}_{j=1}^{N-1}$ are $N-1$ consecutive columns of F_n on the left of

$A_\perp^{(:,i)}$. Hence every term in the sum of (3.7) contains a column of F_n.

Next, we investigate which columns of F_n are included in this sum. Based on the way we partition F_n into A and $A_\perp$, at least one column in any N consecutive columns in $F_n^{(:,N+1:n)}$ must belong to $A_\perp$ (see Figure 3.2). The only exception is that, between $F_n^{(:,n)}$ and $F_n^{(:,N+1)}$, which are adjacent columns in $A_\perp$,[7] there are N columns $F_n^{(:,1:N)}$. Therefore, when $i = 2, 3, \ldots, n-m$, the columns $\{F_n^{(:,n+1-j)} \odot A_\perp^{(:,i)}\}_{j=1}^{N-1}$ sweep to the left of $A_\perp^{(:,i)}$, and "fill the gap" by covering all the columns between $A_\perp^{(:,i)}$ and $A_\perp^{(:,i-1)}$. In general, since the gap between $A_\perp^{(:,i-1)}$ and $A_\perp^{(:,i)}$ could be smaller than $N-1$, there could be overlaps of columns. When $i = 1$, since $A_\perp^{(:,1)} = F_n^{(:,N+1)}$, the columns $\{F_n^{(:,n+1-j)} \odot A_\perp^{(:,1)}\}_{j=1}^{N-1}$ are $F_n^{(:,2:N)}$. Therefore $F_n^{(:,1)}$ is not included in this sum. In summary,

$$\left\{A_\perp^{(:,i)}\right\}_{i=1}^{n-m} \bigcup \left\{\left\{F_n^{(:,n+1-j)} \odot A_\perp^{(:,i)}\right\}_{j=1}^{N-1}\right\}_{i=1}^{n-m} = \left\{F_n^{(:,j)}\right\}_{j=2}^{n},$$

i.e., the $(n-m)+(n-m)(N-1)$ terms in the sum of (3.7) actually contain $n-1$ distinct columns of F_n. It follows that there are $(n-m)+(n-m)(N-1)-(n-1) = nN-mN-n+1 = M$ linearly independent choices of the coefficient vector $[\mathrm{vec}(Q)^\top, p^\top]^\top$. We denote these linearly independent vectors by $[\mathrm{vec}(Q_k)^\top, p_k^\top]^\top$, $k = 1, 2, \ldots, M$.

Next we prove that $Q_1, Q_2, \ldots, Q_M$ are linearly independent. We argue by contradiction. Suppose they are linearly dependent, and there exists $\beta_1, \beta_2, \ldots, \beta_M$ such that

$$\sum_{k=1}^{M} \beta_k Q_k = 0. \tag{3.8}$$

Then,

$$\begin{aligned} A_\perp\left(\sum_{k=1}^{M} \beta_k p_k\right) &= \sum_{k=1}^{M} \beta_k A_\perp p_k = \sum_{k=1}^{M} \beta_k \sum_{i=1}^{n-m} \sum_{j=1}^{N-1} \big(F_n^{(:,n+1-j)} \odot A_\perp^{(:,i)}\big) Q_k^{(i,j)} \\ &= \sum_{i=1}^{n-m} \sum_{j=1}^{N-1} \big(F_n^{(:,n+1-j)} \odot A_\perp^{(:,i)}\big)\left(\sum_{k=1}^{M} \beta_k Q_k^{(i,j)}\right) = 0. \end{aligned}$$

The second equation follows from (3.7), and the last equation follows from

[7] Recall that $F_n^{(:,n)}$ and $F_n^{(:,N+1)}$ are not picked for A, and are the last and the first columns of $A_\perp$.

(3.8). Since the matrix $A_\perp$ has full column rank, we have

$$\sum_{k=1}^{M} \beta_k p_k = 0. \tag{3.9}$$

Equations (3.8) and (3.9) suggest that $[\mathrm{vec}(Q_k)^\top, p_k^\top]^\top (k = 1, 2, \ldots, M)$ are linearly dependent, which causes a contradiction. Therefore, $Q_1, Q_2, \ldots, Q_M$ are linearly independent. There exist exactly M linearly independent left null vectors for $D(A, X_0)$. Therefore, $D(A, X_0)$ has rank $mN - 1$ for the special choice of A and X_0, which completes the proof.

3.3.3 Proof of Theorem 3.2.2

First, by the same argument as in the proof of Theorem 3.2.1, if X_0 is given, the recovery of λ_0 is unique. Again by Lemma 3.3.1, we only need to show that for generic λ_0, X_0, and A, if there exists (λ_1, X_1) such that $\mathrm{diag}(\lambda_0)AX_0 = \mathrm{diag}(\lambda_1)AX_1$, then $X_1 = \sigma X_0$ for some nonzero σ.

We start by fixing the supports of X_0 and X_1. Suppose $\mathrm{diag}(\lambda_0)AX_0 = \mathrm{diag}(\lambda_1)AX_1$, and J_0 and J_1 are the row supports (the index set on which the rows of a matrix are nonzero) of X_0 and X_1, respectively, and $|J_0| = |J_1| = s$. Then focus on the following equation, containing the nonzero rows of X_0 and X_1:

$$\mathrm{diag}(\lambda_0)A^{(:, J_0 \bigcup J_1)} X_0^{(J_0 \bigcup J_1, :)} = \mathrm{diag}(\lambda_1)A^{(:, J_0 \bigcup J_1)} X_1^{(J_0 \bigcup J_1, :)}.$$

Obviously, the cardinality of the set $J_0 \bigcup J_1$ is at most $2s$. Let $\ell = |J_0 \bigcup J_1| \leq 2s$. We can show that $X_1^{(J_0 \bigcup J_1, :)} = \sigma X_0^{(J_0 \bigcup J_1, :)}$ for some nonzero σ, following the same steps as in the proof of Theorem 3.2.1, with Claim 3.3.2 replaced by the following claim:

Claim 3.3.3. *For almost all X_0 with row support J_0 and almost all A, if $n > 2s \geq \ell$ and $\frac{n-1}{n-2s} \leq N \leq s$, then the following matrix has full column*

rank, which is ℓN:

$$\begin{bmatrix} \operatorname{vec}(X_0^{(J_0 \bigcup J_1,:)})^* \\ D(A^{(1,J_0 \bigcup J_1)}, X_0^{(J_0 \bigcup J_1,:)}) \\ D(A^{(2,J_0 \bigcup J_1)}, X_0^{(J_0 \bigcup J_1,:)}) \\ \vdots \\ D(A^{(n,J_0 \bigcup J_1)}, X_0^{(J_0 \bigcup J_1,:)}) \end{bmatrix}. \tag{3.10}$$

The proof of Claim 3.3.3 uses arguments similar to those in the proof of Claim 3.3.2: an explicit construction of $A^{(:,J_0 \bigcup J_1)}$ and $X_0^{(J_0 \bigcup J_1,:)}$ that satisfies a rank condition described below. Here, one cannot choose every entry of $X_0^{(J_0 \bigcup J_1,:)}$ freely, since it has only s nonzero rows. Let Q be an $\ell \times \ell$ permutation matrix, such that the first s rows of $QX_0^{(J_0 \bigcup J_1,:)}$ are nonzero. Then we apply the construction of A and X in the the proof of Claim 3.3.2, to $A^{(:,J_0 \bigcup J_1)}Q^{-1}$ and $QX_0^{(J_0 \bigcup J_1,:)}$. For example, we choose $X_0^{(J_0 \bigcup J_1,:)}$ such that $QX_0^{(J_0 \bigcup J_1,:)}$ is the first $N \leq s$ columns of I_ℓ. Then by the proof of Claim 3.3.2, the following matrix has full column rank ℓN:

$$\begin{bmatrix} \operatorname{vec}(QX_0^{(J_0 \bigcup J_1,:)})^* \\ D(A^{(1,J_0 \bigcup J_1)}Q^{-1}, QX_0^{(J_0 \bigcup J_1,:)}) \\ D(A^{(2,J_0 \bigcup J_1)}Q^{-1}, QX_0^{(J_0 \bigcup J_1,:)}) \\ \vdots \\ D(A^{(n,J_0 \bigcup J_1)}Q^{-1}, QX_0^{(J_0 \bigcup J_1,:)}) \end{bmatrix}. \tag{3.11}$$

We complete the proof of Claim 3.3.3 by making the following observation: (3.11) is a permutation of the columns of (3.10), and the two matrices have the same rank.

We continue the proof of Theorem 3.2.2. We have established that

$$X_1^{(J_0 \bigcup J_1,:)} = \sigma X_0^{(J_0 \bigcup J_1,:)}$$

for some nonzero σ. Recall that the other rows of X_0 and X_1 are zero. Hence $X_1 = \sigma X_0$. Therefore, for almost all λ_0 and A, and almost all X_0 whose row support is J_0, the solution (λ_1, X_1), for which the support of X_1 is J_1, satisfies that $X_1 = \sigma X_0$ and $\lambda_1 = \frac{1}{\sigma}\lambda_0$. There are a finite number of choices for the supports J_0 and J_1, $\binom{m}{s}^2$ choices to be exact. Therefore, we can complete the proof by enumerating over all possible choices for J_0 and J_1.

CHAPTER 4

BLIND GAIN AND PHASE CALIBRATION VIA SPARSE SPECTRAL METHODS

4.1 Introduction

4.1.1 Notations

We use $A^\top$, $\overline{A}$, and A^* to denote the transpose, the complex conjugate, and the conjugate transpose of a matrix A, respectively. The k-th entry of a vector λ is denoted by λ_k. The j-th column, the k-th row (in a column vector form), and the (k,j)-th entry of a matrix A are denoted by $a_{\cdot j}$, $a_{k\cdot}$, and a_{kj}, respectively. Upper script t in a vector $\eta^{(t)}$ denotes the iteration number in an iterative algorithm. We use I_n to denote the identity matrix of size $n \times n$, and $\mathbf{1}_{n,m}$ and $\mathbf{0}_{n,m}$ to denote the matrices of all ones and all zeros of size $n \times m$, respectively. The i-th standard basis vector is denoted by e_i, whose ambient dimension is clear in the context. The ℓ_p norm and ℓ_0 "norm" of a vector x are denoted by $\|x\|_p$ and $\|x\|_0$, respectively. The Frobenius norm and the spectral norm of a matrix A are denoted by $\|A\|_{\mathrm{F}}$ and $\|A\|$, respectively. The support of a sparse vector x is denoted by $\mathrm{supp}(x)$. The vector $\mathrm{vec}(X)$ denotes the concatenation of the columns of $X = [x_{\cdot 1}, x_{\cdot 2}, \ldots, x_{\cdot N}]$, i.e., $\mathrm{vec}(X) = [x_{\cdot 1}^\top, x_{\cdot 2}^\top, \ldots, x_{\cdot N}^\top]^\top$. A diagonal matrix with the entries of vector x on the diagonal is denoted by $\mathrm{diag}(x)$. The Kronecker product is denoted by $\otimes$. We use $\gtrsim$ to denote the relation greater than up to log factors. We use $[n]$ to denote the set $\{1, 2, \ldots, n\}$. For an index set T, the projection operator onto T is denoted by Π_T, and the operator that restricts onto T is denoted by Ω_T. We use these operator notations for different spaces, and the ambient dimensions will be clarified in the context.

4.1.2 Problem Formulation

In this section, we introduce the BGPC problem with a subspace constraint or a sparsity constraint. Suppose $A \in \mathbb{C}^{n\times m}$ is the known measurement matrix, and $\lambda \in \mathbb{C}^n$ is the vector of unknown gains and phases, the k-th entry of which is $\lambda_k = |\lambda_k| e^{\sqrt{-1}\varphi_k}$. Here, $|\lambda_k|$ and φ_k denote the gain and phase of the k-th sensor, respectively. The BGPC problem is the simultaneous recovery of λ and the unknown signal matrix $X \in \mathbb{C}^{m\times N}$ from the following measurement:

$$Y = \mathrm{diag}(\lambda) A X + W, \tag{4.1}$$

where $W \in \mathbb{C}^{n\times N}$ is the measurement noise. The (k,j)-th entry in the measurement y_{kj} has the following expression:

$$y_{kj} = \lambda_k \, a_{k\cdot}^\top \, x_{\cdot j} + w_{kj}.$$

Clearly, BGPC is a bilinear inverse problem. The solution (λ, X) suffers from scaling ambiguity, i.e., $(\lambda/\sigma, \sigma X)$ generates the same measurements as (λ, X), and therefore cannot be distinguished from it. Despite the fact that the solution can have other ambiguity issues, in this chapter, we consider the generic setting where the solution suffers only from scaling ambiguity [3].[1] Even in this setting, the solution is not unique, unless we exploit the structure of the signals. In this chapter, we solve the BGPC problem under two scenarios – BGPC with a subspace structure, and BGPC with sparsity.

(1) **Subspace case:** Suppose that the known matrix A is tall ($n > m$) and has full column rank. Then the columns of AX reside in the low-dimensional subspace spanned by the columns of A. The problem is effectively unconstrained with respect to X.

(2) **Sparsity case:** Suppose that A is a known dictionary with $m \geq n$, while the columns of X are s_0-sparse, i.e., $\|x_{\cdot j}\|_0 \leq s_0$ for all $j \in [N]$. A variation of this setting is that the columns of X are jointly s_0-sparse, i.e., there are at most s_0 nonzero rows in X. In this case, the subspace constraint on AX no longer applies, and one must solve the problem with a sparsity (or joint sparsity) constraint.

[1] An example of another ambiguity is a shift ambiguity when A is the discrete Fourier transform matrix [1, 41]. For a generic matrix A, the solution to BGPC does not suffer from shift ambiguity.

Table 4.1: Comparison of sample complexities with prior work.

	Subspace	Joint Sparsity	Sparsity
Unique Recovery [3]	$n > m$ $N \geq \frac{n-1}{n-m}$	$n > 2s_0$ $N \geq \frac{n-1}{n-2s_0}$	–
Least Squares [40]	$n \gtrsim m$ $N \gtrsim 1$	–	–
ℓ_1 Minimization [41]	–	–	$n \gtrsim s_0$ $N \gtrsim n$
This Paper	$n \gtrsim m$ $N \gtrsim 1$	$n \gtrsim s_0$ $N \gtrsim \sqrt{s_0}$	–

Note: n, N, m and s_0 represent the number of sensors, the number of snapshots, the subspace dimension, and the sparsity level, respectively.

The BGPC problem arises in applications including inverse rending, sensor array processing, multichannel blind deconvolution, and SAR autofocus. We refer the reader to our previous work [3, Section II.C] for a detailed account of applications of BGPC. For consistency, from now on, we use the convention in sensor array processing, and refer to n and N as the numbers of sensors and snapshots, respectively.

4.1.3 Our Contributions

We reformulate BGPC as the problem of finding the principal eigenvector of a matrix (or operator). In the subspace case, this can be solved using any eigen-solver, e.g., power iteration (Algorithm 1). In the sparsity case, we propose to solve this problem using truncated power iteration (Algorithm 2). Our main results can be summarized as follows.

Theorem 4.1.1. *Under certain assumptions on A, λ, X, and W, one can solve the BGPC problem with high probability using:*

(1) ***Subspace case:*** *algorithms that find the principal eigenvector of a certain matrix, e.g., power iteration, if $n \gtrsim m$ and $N \gtrsim 1$.*

(2) ***Joint sparsity case:*** *truncated power iteration with a good initialization, if $n \gtrsim s_0$ and $N \gtrsim \sqrt{s_0}$.*

In Table 4.1, we compare the above results with the sample complexities for unique recovery in BGPC [3], and previous guaranteed algorithms for

BGPC in the subspace and sparsity case [40, 41]. In the subspace case, power iteration solves BGPC using optimal (up to log factors) numbers of sensors and snapshots. These sample complexities are comparable to the least squares method in [40]. Moreover, we show that power iteration is empirically more robust against noise than least squares.

Truncated power iteration solves BGPC with a joint sparsity structure, with an optimal (up to log factors) number of sensors, and a slightly sub-optimal (within a factor of $\sqrt{s_0}$ and log factors) number of snapshots. In comparison, the ℓ_1 minimization method for the sparsity case of BGPC uses a similar number of sensors, but a much larger number of snapshots. Numerical experiments show that truncated power iteration empirically succeed, in both the joint sparsity case and the more general sparsity case, in the optimal regime.

The success of truncated power iteration relies on a good initial estimate of X and λ. We propose a simple initialization algorithm (Algorithm 3) with the following guarantee.

Theorem 4.1.2. *Under additional assumptions on the absolute values of the nonzero entries in X, our initialization algorithm produces a sufficiently good estimate of λ and X if $n \gtrsim s_0^2$. (We do not require any additional assumption on the number N of snapshots.)*

Despite the above scaling law predicted by theory, numerical experiments suggest that our initialization scheme is effective when $n \gtrsim s_0$.

4.1.4 Related Work

BGPC arises in many real-world scenarios, and previous solutions have mostly been tailored to specific applications such as sensor array processing [30, 89, 90], sensor network calibration [69, 91], synthetic aperture radar autofocus [38], and computational relighting [29]. However, the previous methods do *not* have theoretical guarantees in the forms of quantitative error bounds.

The idea of solving BGPC by reformulating it into a linear inverse problem, which is a key idea in this chapter, has been proposed by many prior works [69, 38, 29]. In particular, Bilen et al. [70] provided a solution to BGPC with high-dimensional but sparse signals using ℓ_1 minimization. However, such methods have not been carefully analyzed until recently. Ling and Strohmer

[40] derived an error bound for the least squares solution in the subspace case of BGPC. In this chapter, the power iteration method has sample complexities comparable to those of the least squares method [40], and is empirically more robust to noise than the latter. Wang and Chi [41] gave a theoretical guarantee for ℓ_1 minimization that solves BGPC in the sparsity case, where they assumed that A is the discrete Fourier transform (DFT) matrix and X is random following a Bernoulli–sub-Gaussian model. In this chapter, we give a guarantee for truncated power iteration under the assumption that A is a complex Gaussian random matrix, and X is *jointly* sparse, well-conditioned, and deterministic. In this sense, we consider an adversarial scenario for the signal X. Our sample complexity results require a near optimal number n of sensors, and a much smaller number N of snapshots. Moreover, truncated power iteration is more robust against noise and inaccurate initial estimate of phases. Very recently, Eldar et al. [92] proposed new methods for BGPC with signals whose sparse components may lie off the grid. Similar to earlier work on blind calibration of sensor arrays [30], these methods rely on empirical covariance matrices of the measurements and therefore need a relatively large number of snapshots.

A problem related to BGPC is multichannel blind deconvolution (MBD). Most previous works on MBD consider linear convolution with a finite impulse response (FIR) filter model (see [53, 54], and a recent stabilized method [57, 58]). In comparison, BGPC is equivalent to MBD with *circular* convolution and a subspace model or a sparsity model, akin to some recent studies [40, 41]. BGPC is more general in the sense that: (a) linear convolution can be rewritten as circular convolution via zero-padding the signal and the filter; (b) the FIR filter model is a special case of the subspace model.

To position BGPC in a more broad context, it is a special bilinear inverse problem [1], which in turn is a special case of low-rank matrix recovery from incomplete measurements [93, 94, 28, 24]. A resurgence of interest in bilinear inverse problems was pioneered by the recent studies in single-channel blind deconvolution of signals with subspace or sparsity structures, where both the signal and the filter are structured [14, 19, 21, 22, 26].

Another related bilinear inverse problem is blind calibration via repeated measurements from multiple different sensing operators [95, 96, 97, 98, 99, 100]. Since blind calibration with repeated measurements is in principle an easier problem than BGPC [40], we believe our methods for BGPC and our

theoretical analysis can be extended to this scenario.

Also related is the phase retrieval problem [101], where there only exists uncertainty in the phases (and not the gains) of the sensing system. An active line of work solves phase retrieval with guaranteed algorithms (see [15, 102, 103, 104, 105, 106, 107] and [108] for a recent review).

The error bounds of power iteration and truncated power iteration have been analyzed in general settings, e.g., in [42, Section 8.2.1] and [44]. These previous results hinge on spectral properties of matrices such as gaps between eigenvalues, which do not translate directly to sample complexity requirements. This chapter undertakes analysis specific to BGPC. We relate spectral properties in BGPC to some technical conditions on λ, A, X, and W, and derive recovery error under near optimal sample complexities. We also adapt the analysis of sparse PCA [44] to accommodate a structured sparsity constraint in BGPC.

BGPC and our proposed methods are non-convex in nature. In particular, our truncated power iteration algorithm can be interpreted as projected gradient descent for a non-convex optimization problem. There have been rapid developments in guaranteed non-convex methods [109] in a variety of domains such as matrix completion [110, 111, 112], dictionary learning [71, 74], blind deconvolution [21, 26], and phase retrieval [103, 102, 72]. It is a common theme that carefully crafted non-convex methods have better theoretical guarantees in terms of sample complexity than their convex counterparts, and often have faster implementations and better empirical performance. This chapter is a new example of such superiority of non-convex methods.

4.2 Power Iteration Algorithms for BGPC

Next, we describe the algorithms we use to solve BGPC. In Section 4.2.1, we introduce a simple trick that turns the bilinear inverse problem in BGPC to a linear inverse problem. In Sections 4.2.2 and 4.2.3, we introduce the power iteration algorithm we use to solve BGPC with a subspace structure, and the truncated (or sparse) power iteration algorithm we use to solve BGPC with sparsity, respectively.

4.2.1 From Bilinearity to Linearity

We use a simple trick to turn BGPC into a linear inverse problem [69]. Without loss of generality, assume that $\lambda_k \neq 0$ for $k \in [n]$. Indeed, if any sensor has zero gain, then the corresponding row in Y is all zero or contains only noise, and we can simply remove the corresponding row in (4.1). Let γ denote the entrywise inverse of λ, i.e., $\gamma_k = 1/\lambda_k$ for $k \in [n]$. We have

$$\operatorname{diag}(\gamma) Y_{\mathrm{s}} = AX, \tag{4.2}$$

where $Y_{\mathrm{s}} = \operatorname{diag}(\lambda) AX$ is the noiseless measurement. Equation (4.2) is linear in all the entries of γ and X. The bilinear inverse problem in (λ, X) now becomes a linear inverse problem in (γ, X). In practice, since only the noisy measurement Y is available, one can solve $\operatorname{diag}(\gamma) Y \approx AX$.

This technique was widely used to solve BGPC with a subspace structure, in applications such as sensor network calibration [69], synthetic aperture radar autofocus [38], and computational relighting [29]. Recently, Ling and Strohmer [40] analyzed the least squares solution to (4.2). Wang and Chi [41] considered a special case where A is the DFT matrix, and analyzed the solution of a sparse X by minimizing the ℓ_1 norm of $A^{-1}\operatorname{diag}(\gamma) Y$.

We use the same trick in our algorithms. Define

$$D \coloneqq \begin{bmatrix} I_N \otimes a_{1\cdot}^\top \\ \vdots \\ I_N \otimes a_{n\cdot}^\top \end{bmatrix}, \tag{4.3}$$

$$E \coloneqq \begin{bmatrix} y_{1\cdot} & & \\ & \ddots & \\ & & y_{n\cdot} \end{bmatrix}. \tag{4.4}$$

We can decompose E into $E = E_{\mathrm{s}} + E_{\mathrm{n}}$, where

$$E_{\mathrm{s}} \coloneqq \begin{bmatrix} \lambda_1 X^\top a_{1\cdot} & & \\ & \ddots & \\ & & \lambda_n X^\top a_{n\cdot} \end{bmatrix},$$

$$E_{\mathrm{n}} \coloneqq \begin{bmatrix} w_{1\cdot} & & \\ & \ddots & \\ & & w_{n\cdot} \end{bmatrix}.$$

Define also

$$B \coloneqq \begin{bmatrix} D^*D & \alpha D^*E \\ \alpha E^*D & \alpha^2 E^*E \end{bmatrix}, \tag{4.5}$$

$$B_{\mathrm{s}} \coloneqq \begin{bmatrix} D^*D & \alpha D^*E_{\mathrm{s}} \\ \alpha E_{\mathrm{s}}^*D & \alpha^2 E_{\mathrm{s}}^*E_{\mathrm{s}} \end{bmatrix},$$

where α is a nonzero constant specified later.

Clearly, (4.2) can be rewritten as

$$Dx - E_{\mathrm{s}}\gamma = 0,$$

where $x = \mathrm{vec}(X)$. Equivalently, $\eta = [x^\top, -\gamma^\top/\alpha]^\top$ is a null vector of B_{s}. When certain sufficient conditions are satisfied, η is the unique null vector of B_{s}. For example, if λ, A, and X are in general positions in $\mathbb{C}^n$, $\mathbb{C}^{n\times m}$, and $\mathbb{C}^{m\times N}$, respectively, then $N \geq \frac{n-1}{n-m}$ snapshots are sufficient to guarantee uniqueness of the solution to BGPC in the subspace case. We refer readers to our work on the identifiability in BGPC for more details [1, 3].

Since only the noisy matrix B is accessible in practice, one can instead find the minor eigenvector, i.e., the eigenvector corresponding to the smallest eigenvalue of B. The rest of this section focuses on algorithms that find such an eigenvector of B, with no constraint (in the subspace case), or with a sparsity constraint (in the sparsity case).

4.2.2 Power Iteration for BGPC with a Subspace Structure

In the subspace case ($n > m$), we solve for the minor eigenvector of the positive definite matrix B. In Section 4.3, we derive an upper bound on the error between this eigenvector and the true solution η.

The minor eigenvector of B can be computed by a variety of methods. Here, we propose an algorithm that remains computationally efficient for large scale problems. By eigenvalue decomposition, the null vector of B is

identical to the principal eigenvector of

$$G = \beta I_{mN+n} - B, \tag{4.6}$$

for a large enough constant β. This eigenvector can be computed using the power iteration algorithm (see Algorithm 1).

The size of G is $(Nm+n) \times (Nm+n)$. An advantage of Algorithm 1 over an eigen-solver that decomposes G, is that one does not need to explicitly compute the entries of G to iteratively apply it to a vector. Furthermore, rather than $O((Nm+n)^2)$, by the structure of D and E, the per iteration time complexity of applying the operator G to a vector is only $O(mnN)$. This can be further reduced if A and A^* are linear operators with implementations faster than $O(mn)$.

The rule of thumb for selecting parameter α is that the ℓ_2 norms of the columns of D be close to those of αE so that G in (4.6) exhibits good spectral properties for power iterations. A safe choice for β is $\|B\|$, which may be conservatively large in some cases, but works well in practice. In Section 4.3, we discuss our choice of parameters α, β under certain normalization assumptions (see Remark 4.3.6).

Algorithm 1 converges to the principal eigenvector of G, as long as the initial estimate $\eta^{(0)}$ is not orthogonal to that eigenvector. This insensitivity to initialization is a privilege not shared by the sparsity case (see Section 4.2.3).

Algorithm 1: Power Iteration for BGPC

Input: $A \in \mathbb{C}^{n\times m}$, $Y \in \mathbb{C}^{n\times N}$, initial estimate $\eta^{(0)} \in \mathbb{C}^{Nm+n}$
Output: $\eta^{(t)} \in \mathbb{C}^{Nm+n}$
Parameters: α, β
Compute operator $G : \mathbb{C}^{Nm+n} \to \mathbb{C}^{Nm+n}$ by (4.3), (4.4), (4.5), (4.6)
$t \leftarrow 1$
repeat
  Compute $\eta^{(t)} = G\eta^{(t-1)} / \left\|G\eta^{(t-1)}\right\|_2$
  $t \leftarrow t+1$
until *convergence criterion is met*

4.2.3 Truncated Power Iteration for BGPC with Sparsity

When $2 \le n \le m$, $[D, \alpha E] \in \mathbb{C}^{Nn \times (Nm+n)}$ is a fat matrix, and the null space of B has dimension at least 2. Therefore, there exist at least two linearly independent eigenvectors corresponding to the largest eigenvalue of G. To overcome the ill-posedness, one can leverage the sparsity structure in X to make the solution to the eigenvector problem unique.

Let $\Pi_s(x)$ denote the projection of a vector x onto the set of s-sparse vectors. It is computed by setting to zero all but the s entries of x of the largest absolute values. Let $\Pi'_s(X)$ denote the projection of a matrix X onto the set of matrices whose columns are jointly s-sparse. This projection is computed by setting to zero all but the s rows of X of the largest ℓ_2 norms. We define two projection operators on $\eta = [x^\top, -\gamma^\top/\alpha]^\top$ that will be used repeatedly in the rest of this chapter:

$$\begin{aligned}
\widetilde{\Pi}_s(\eta) &\coloneqq [\Pi_s(x_{\cdot 1})^\top, \Pi_s(x_{\cdot 2})^\top, \ldots, \Pi_s(x_{\cdot N})^\top, -\gamma^\top/\alpha]^\top, \\
\widetilde{\Pi}'_s(\eta) &\coloneqq [\mathrm{vec}\big(\Pi'_s(X)\big)^\top, -\gamma^\top/\alpha]^\top.
\end{aligned}$$

For the sparsity case of BGPC, we adapt the eigenvector problem in Section 4.2.2 by adding a sparsity constraint:

$$\begin{aligned}
\max_{\eta} \quad & \eta^* G \eta \\
\text{s.t.} \quad & \|\eta\|_2 = 1, \\
& \widetilde{\Pi}_{s_0}(\eta) = \eta.
\end{aligned} \tag{4.7}$$

This nonconvex optimization is very similar to the sparse PCA problem. The only difference lies in the structure of the sparsity constraint. In sparse PCA, the principal component is s_0-sparse. In (4.7), the vector η consists of s_0-sparse vectors $x_{\cdot 1}, x_{\cdot 2}, \ldots, x_{\cdot N}$, and a dense vector $-\gamma/\alpha$.

To solve (4.7), we adopt a sparse PCA algorithm called truncated power iteration [44], and revise it to adapt to the sparsity structure of BGPC (see Algorithm 2). One can choose parameters α and β using the same rules as in Section 4.2.2. Note that we use a sparsity level $s_1 \ge s_0$ in this algorithm, for two reasons: (a) in practice, it is easier to obtain an upper bound on the sparsity level instead of the exact number of nonzero entries in the signal; and (b) the ratio s_0/s_1 is an important constant in the main results, con-

trolling the trade-off between the number of measurements and the rate of convergence.

For the joint sparsity case, we use essentially the same algorithm, with $\widetilde{\Pi}_{s_1}$ replaced by $\widetilde{\Pi}'_{s_1}$.

Since (4.7) is a nonconvex optimization problem, a good initialization $\eta^{(0)}$ is crucial to the success of Algorithm 2. Algorithm 3 outlines one such initialization. We denote by Π_{T_x} the projection onto the support set T_x, which sets to zero all rows of D^*E but the s_1 rows of the largest ℓ_2 norms in each block. (The j-th block of D^*E consists of m contiguous rows indexed by $\{(j-1)m+\ell\}_{\ell\in[m]}$.) Then the normalized left and right singular vectors u and v of $\Pi_{T_x}D^*E$ are computed as initial estimates for x and λ. We use $1./v$ to denote the entrywise inverse of v except for zero entries, which are kept zero. In Section 4.3, we further comment on how to choose a proper initial estimate $\eta^{(0)}$ (see Remark 4.3.11).

Algorithm 2: Truncated Power Iteration for BGPC with Sparsity

Input: $A \in \mathbb{C}^{n\times m}$, $Y \in \mathbb{C}^{n\times N}$, initial estimate $\eta^{(0)} \in \mathbb{C}^{Nm+n}$
Output: $\eta^{(t)} \in \mathbb{C}^{Nm+n}$
Parameters: α, β, s_1
Compute operator $G : \mathbb{C}^{Nm+n} \to \mathbb{C}^{Nm+n}$ by (4.3), (4.4), (4.5), (4.6)
$t \leftarrow 1$
repeat
  Compute $\tilde{\eta}^{(t)} = G\eta^{(t-1)} / \left\|G\eta^{(t-1)}\right\|_2$
  Compute $\eta^{(t)} = \widetilde{\Pi}_{s_1}(\tilde{\eta}^{(t)}) / \left\|\widetilde{\Pi}_{s_1}(\tilde{\eta}^{(t)})\right\|_2$
  $t \leftarrow t+1$
until *convergence criterion is met*

4.2.4 Alternative Interpretation as Projected Gradient Descent

Algorithms 1 and 2 can be interpreted as gradient descent and projected gradient descent, respectively. Next, we explain such equivalence using the sparsity case as an example.

Recall that BGPC is linearized as $\begin{bmatrix} D & \alpha E \end{bmatrix} \eta = 0$. Relaxing the sparsity

Algorithm 3: Initialization for Truncated Power Iteration

Input: $A \in \mathbb{C}^{n\times m}$, $Y \in \mathbb{C}^{n\times N}$
Output: initial estimate $\eta^{(0)} \in \mathbb{C}^{Nm+n}$
Parameters: s_1
Compute matrix $D^*E \in \mathbb{C}^{Nm\times n}$ by (4.3), (4.4)
$T_x \leftarrow \emptyset$
for $j \in [N]$ **do**
  Compute the row norms $\left\|d^*_{\cdot((j-1)m+\ell)}E\right\|_2$ for $\ell \in [m]$
  Find subset $T_j \subset [m]$ ($|T_j| = s_1$) s.t. for $\ell \in T_j$ and $\ell' \in [m]\backslash T_j$:

$$\left\|d^*_{\cdot((j-1)m+\ell)}E\right\|_2 \geq \left\|d^*_{\cdot((j-1)m+\ell')}E\right\|_2$$

  Merge support $T_x \leftarrow T_x \bigcup (T_j + \{(j-1)m\})$
end
Compute the principal left and right singular vectors u, v of $\Pi_{T_x} D^*E$
$\eta^{(0)} \leftarrow [u^\top, -(1./v^\top)/n]^\top$
$\eta^{(0)} \leftarrow \eta^{(0)}/\left\|\eta^{(0)}\right\|_2$

level from s_0 to s_1, the optimization in (4.7) is equivalent to:

$$\begin{aligned} \min_{\eta} \quad & \frac{1}{2}\left\|\begin{bmatrix} D & \alpha E\end{bmatrix}\eta\right\|_2^2 \\ \text{s.t.} \quad & \|\eta\|_2 = 1, \\ & \widetilde{\Pi}_{s_1}(\eta) = \eta. \end{aligned}$$

The gradient of the objective function at $\eta^{(t-1)}$ is

$$\begin{bmatrix} D^* \\ \alpha E^*\end{bmatrix}\begin{bmatrix} D & \alpha E\end{bmatrix}\eta^{(t-1)} = B\eta^{(t-1)}.$$

Each iteration of projected gradient descent consists of two steps:
(i) **Gradient descent** with a step size of $1/\beta$:

$$\tilde{\eta}^{(t)} = \eta^{(t-1)} - \frac{1}{\beta}B\,\eta^{(t-1)} = \frac{1}{\beta}G\,\eta^{(t-1)}.$$

(ii) **Projection** onto the constraint set, i.e., the intersection of a cone ($\widetilde{\Pi}_{s_1}(\eta) =$

η) and a sphere ($\|\eta\|_2 = 1$):

$$\eta^{(t)} = \widetilde{\Pi}_{s_1}(\tilde{\eta}^{(t)}) / \left\| \widetilde{\Pi}_{s_1}(\tilde{\eta}^{(t)}) \right\|_2 .$$

Clearly, the two steps are identical to those in each truncated power iteration except for a different scaling in Step (i), which, due to the normalization in Step (ii), is insignificant.

4.3 Main Results

In this section, we give theoretical guarantees for Algorithms 1 and 2 in the subspace case and in the joint sparsity case, respectively. We also give a guarantee for the initialization by Algorithm 3.

4.3.1 Main Assumptions

We start by stating the assumptions on A, λ, X and W, which we use throughout this section.

Assumption 4.3.1. *A is a complex Gaussian random matrix, whose entries are i.i.d. following $\mathcal{CN}(0, \frac{1}{n})$. Equivalently, the vectors $\{a_{k\cdot}\}_{k=1}^n$ are i.i.d. following $\mathcal{CN}(\mathbf{0}_{m,1}, \frac{1}{n} I_m)$.*

Assumption 4.3.2. *The vector λ has "flat" gains in the sense that $1 - \delta \leq |\lambda_k|^2 \leq 1 + \delta$ for some $\delta \in (0, 1)$.*

Assumption 4.3.3. *The matrix $X \in \mathbb{C}^{m \times N}$ is normalized and has good conditioning, i.e., $\|X\|_{\mathrm{F}} = 1$, and for some $\theta \in (0, 1)$ we have:*

- ***Subspace case:***

$$\min\{\|NX^*X - I_N\|, \|mXX^* - I_m\|\} \leq \theta,$$

- ***Joint sparsity case:***

$$\min\{\|NX^*X - I_N\|, \left\|s_0 \Omega_{T_0} XX^* \Omega_{T_0}^* - I_{s_0}\right\|\} \leq \theta,$$

where Ω_T denotes the operator that restricts a matrix to the row support T, and $T_0 \coloneqq \{i \in [m] | \left\|e_i^\top X\right\|_2 > 0\}$ ($|T_0| = s_0$) is the row support of X.

Assumptions 4.3.1 – 4.3.3 can be relaxed in practice.

- The complex Gaussian distribution in Assumption 4.3.1 can be relaxed to $\mathcal{CN}(0, \sigma_A^2)$ for any $\sigma_A > 0$. We choose the particular scaling $\sigma_A^2 = 1/n$, because then A satisfies the restricted isometry property (RIP) [113], i.e., $(1 - \delta_s) \|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_s) \|x\|_2^2$ for some $\delta_s \in (0, 1)$, when n is large compared to the number s of nonzero entries in x.

- The gains can center around any $\sigma > 0$, i.e., $\sigma(1-\delta) \leq |\lambda_k|^2 \leq \sigma(1+\delta)$. Due to bilinearity, we may assume that λ_k's are centered around 1 without loss of generality by solving for $(\lambda/\sigma, \sigma X)$.

- The Frobenius norm $\|X\|_{\mathrm{F}}$ of matrix X can be any positive number. If $\|X\|_{\mathrm{F}}$ is known, one can scale X to have unit Frobenius norm before solving BGPC. In practice, the norm of X is generally unknown. However, due to Assumptions 4.3.1 (RIP) and 4.3.2 ("flat" gains), we have

  $$\sqrt{(1 - \delta_s)(1 - \delta)} \leq \frac{\|\mathrm{diag}(\lambda) A X\|_{\mathrm{F}}}{\|X\|_{\mathrm{F}}} \leq \sqrt{(1 + \delta_s)(1 + \delta)}.$$

  Hence $\|Y\|_{\mathrm{F}}$ is a good surrogate for $\|X\|_{\mathrm{F}}$ in noiseless or low noise settings, and one can scale X by $1/\|Y\|_{\mathrm{F}}$ to achieve the desired scaling. The slight deviation of $\|X\|_{\mathrm{F}} / \|Y\|_{\mathrm{F}}$ from 1 does *not* have any significant impact on our theoretical analysis. Therefore, we assume $\|X\|_{\mathrm{F}} = 1$ to simply the constants in our derivation.

- The conditioning of X can also be relaxed. When N is large, one can choose a subset of $N' < N$ columns in Y, such that the matrix formed from the corresponding columns of X has good conditioning. When noise amplification is not of concern (noiseless or low noise settings), one can choose a preconditioning matrix $H \in \mathbb{C}^{N \times N}$ such that $X' = XH$ is well conditioned, and then solve the BGPC with $Y' = YH$.

In summary, we can manipulate the BGPC problem and make it approxi-

mately satisfy our assumptions. For example, (4.1) can be rewritten as:

$$\frac{1}{\|YH\|_{\mathrm{F}}} YH = \operatorname{diag}\Big(\frac{\lambda}{\sigma}\Big)\Big(\frac{1}{\sqrt{n}\sigma_A} A\Big)\Big(\frac{\sqrt{n}\sigma\sigma_A}{\|YH\|_{\mathrm{F}}} XH\Big) + \frac{1}{\|YH\|_{\mathrm{F}}} WH.$$

We can run Algorithms 1 and 2 with input $\frac{1}{\sqrt{n}\sigma_A} A$ and $\frac{1}{\|YH\|_{\mathrm{F}}} YH$, and solve for $\frac{\lambda}{\sigma}$ and $\frac{\sqrt{n}\sigma\sigma_A}{\|YH\|_{\mathrm{F}}} XH$. The above manipulations do not have any significant impact on the solution, or on our theoretical analysis. However, by making these assumptions, we eliminate some tedious and unnecessary discussions.

We also need an assumption on the noise level.

Assumption 4.3.4. *The noise term W satisfies*

- ***Subspace case:*** $\max_{k\in[n],j\in[N]} |w_{kj}| \le \frac{C_W}{\sqrt{nN}}$
- ***Joint sparsity case:*** $\max_{k\in[n],j\in[N]} |w_{kj}| \le \frac{C_W}{\sqrt{n}N^2}$

for an absolute constant $C_W > 0$.

In the subspace case, the assumption on the noise level is very mild. Because under Assumptions 4.3.1 – 4.3.3, $\|\mathrm{diag}(\lambda)AX\|_{\mathrm{F}} \le \sqrt{(1+\delta_s)(1+\delta)}$, the noise term W, which satisfies $\|W\|_{\mathrm{F}} \le C_W$, can be on the same order in terms of Frobenius norm as the clean signal $\mathrm{diag}(\lambda)AX$.

Finally, the following assumption is required for a theoretical guarantee of the initialization.

Assumption 4.3.5. *For all $j \in [N]$, there exists $T_j' \subset \mathrm{supp}(x_{\cdot j}) \subset [m]$, such that for all $\ell \in T_j'$,*

$$\frac{|x_{\ell j}|^2}{\|x_{\cdot j}\|_2^2} \ge \frac{\omega}{s_0},$$

for some absolute constant ω, and

$$\frac{\sum_{\ell' \in [m]\setminus T_j'} |x_{\ell' j}|^2}{\|x_{\cdot j}\|_2^2} \le \delta_X,$$

for some small absolute constant $\delta_X \in (0,1)$.

Assumption 4.3.5 says that the support of $x_{\cdot j}$ can be partitioned into two subsets. The absolute values of the entries in the first subset T_j' are sufficiently large. Moreover, the total energy (sum of squares of the entries) in the second subset is small compared to the squared norm of $x_{\cdot j}$. For example,

the assumption is satisfied in the following special case: $T'_j = \mathrm{supp}(x_{\cdot j})$ (therefore $x_{\ell' j} = 0$ for $\ell' \in [m] \backslash T'_j$), and the absolute values of the nonzero entries are all comparable, e.g., $x_{\ell j} = \pm \frac{\|x_{\cdot j}\|}{\sqrt{s_0}}$.

Before introducing our main results, we disclose the choice of parameters α and β for our theoretical analysis of Algorithms 1 and 2.

Remark 4.3.6. *When Assumptions 4.3.1 – 4.3.3 are satisfied, we choose* $\alpha = \sqrt{n}$ *and* $\beta = 3/2$.

4.3.2 A Perturbation Bound for the Eigenvector Problem

Next, we introduce a key result, a perturbation bound for the eigenvector problem, which is used to derive error bounds for power iteration algorithms.

Let $\{T_j\}_{j=1}^N$ denote subsets of $[m]$, such that $|T_j| = s$ and $\mathrm{supp}(x_{\cdot j}) \subset T_j$. We define $T_x \subset [Nm]$ and $T_\eta \subset [Nm + n]$ as follows:

$$T_x \coloneqq \bigcup_{j \in [N]} \big(T_j + \{(j-1)m\}\big), \tag{4.8}$$

$$T_\eta \coloneqq T_x \bigcup \big([n] + \{Nm\}\big). \tag{4.9}$$

Recall that Ω_T restricts a vector to the support T, and hence $\Omega_T^* \Omega_T$ is the projection operator onto the support T. Clearly, we have $x = \Omega_{T_x}^* \Omega_{T_x} x$, and $\eta = \Omega_{T_\eta}^* \Omega_{T_\eta} \eta$. In the subspace case discussed in Theorem 4.3.7, we have $s = m$, $T_j = [m]$, $T_x = [Nm]$, and $T_\eta = [Nm + n]$. In the *joint* sparsity case, we have $T_1 = T_2 = \cdots = T_N$. We set $|T_j| = s = s_0 + 2s_1$, which we justify later in the analysis of truncated power iteration.

Let

$$\dot{\eta} \coloneqq \frac{\eta}{\|\eta\|_2}$$

denote the normalized version of η, which is the eigenvector of B_{s} and $\mathbb{E}B_{\mathrm{s}}$ corresponding to eigenvalue 0. Let $\hat{\eta}$ denote the principal eigenvector of G. In the joint sparsity case, let $\hat{\eta}_{T_\eta}$ denote the principal eigenvector of $\Omega_{T_\eta} G \Omega_{T_\eta}^*$, where $T = T_1 = \cdots = T_N$, $|T| = s$, and the support of η is a subset of T_η defined in (4.9).

In Algorithms 1 and 2 and in our analysis, vectors $\dot{\eta}$, $\hat{\eta}$, and $\eta^{(t)}$ are normalized to unit norm. However, multiplication by a scalar of unit modulus is a remaining ambiguity, i.e., the set $\{e^{\sqrt{-1}\varphi} \dot{\eta} : \varphi \in [0, 2\pi)\}$ is an equiva-

lence class for $\dot{\eta}$. Our main results use $d(\eta, \eta') \coloneqq \min_{\varphi} \left\| e^{\sqrt{-1}\varphi}\eta - \eta' \right\|_2$ to denote the distance between η and η', which is a metric on the set of such equivalence classes.

Theorem 4.3.7 (**Subspace Case**). *Let $\alpha = \sqrt{n}$, and suppose Assumptions 4.3.1 – 4.3.4 are satisfied with $\delta < 1/3$ and a sufficiently small absolute constant $C_W > 0$. Then there exist absolute constants $c, C, C' > 0$, such that if*

$$\max\Big\{ \frac{m \log^2(Nm+n)}{n}, \frac{\log(Nm+n)}{N}, \frac{\log(Nm+n)}{m} \Big\} \leq C, \tag{4.10}$$

then with probability at least $1 - 2n^{-c} - e^{-cm}$,

$$d(\hat{\eta},\ \dot{\eta}) \leq \Delta,$$

where

$$\Delta \coloneqq \frac{8C'}{1-3\delta} \max\{\nu,\ \nu^2\}, \tag{4.11}$$

and

$$\nu \coloneqq \sqrt{nN} \max_{k\in[n], j\in[N]} |w_{kj}|. \tag{4.12}$$

We defer the proof to Section 4.5, and summarize the mathematical tools we use here. By the Davis-Kahan $\sin\theta$ Theorem [114], the error $d(\hat{\eta},\ \dot{\eta})$ in the eigenvector is bounded if there exists a sufficiently large spectral gap between the two largest (in terms of absolute values) eigenvalues of $G = \beta I - B$. We divide this task into two parts: (1) show that there exists a large spectral gap in $\beta I - \mathbb{E}B$; (2) prove that $\|B - \mathbb{E}B\|$ is small using concentration of measure inequalities, e.g., the matrix Bernstein inequality [115, Theorem 1.6].

When m is large (e.g., $m \geq n$), (4.10) does not hold, hence the perturbation bound of the eigenvector $\hat{\eta}$ of G in Theorem 4.3.7 is no longer true. We can, however, bound the perturbation of the eigenvectors of submatrices of G uniformly.

Theorem 4.3.8 (**Joint Sparsity Case**). *Let $\alpha = \sqrt{n}$ and $s = s_0 + 2s_1$, and suppose Assumptions 4.3.1 – 4.3.4 are satisfied with $\delta < 1/3$ and a sufficiently small absolute constant $C_W > 0$. Then there exist absolute constants*

$c, C, C' > 0$, *such that if*

$$\max\Big\{\frac{(s+N)\log^8 n\log^2(sN+m)}{n}, \frac{\sqrt{s}\log^2 n\log(sN+m)}{N}, \frac{\log^4 n\log^2(sN+m)}{s_0}\Big\} \le C, \tag{4.13}$$

then with probability at least $1-2n^{-c}-m^{-cs}$,

$$d(\hat{\eta}_{T_\eta},\ \Omega_{T_\eta}\dot{\eta}) \le \widetilde{\Delta},$$

where

$$\widetilde{\Delta} := \frac{8C'}{1-3\delta}\max\{N^{3/2}\nu,\ \nu^2\}, \tag{4.14}$$

and ν *is defined in* (4.12).

The main challenge in the joint sparsity case is that, instead of bounding the spectral norm of $B-\mathbb{E}B$, one must bound the "sparse" norm of $B-\mathbb{E}B$, i.e., the maximum spectral norm of all principal submatrices whose row (and column) support is T_η defined by (4.9). Since $B-\mathbb{E}B$ can be broken down into the sum of several terms, we give a uniform bound over all submatrices on each term. For any given term, we adopt one of two approaches, whichever provides a tighter bound: (1) we bound the spectral norm of an individual submatrix, and apply a union bound over all submatrices; (2) we use a variational form of the sparse norm, and apply a bound on the suprema of second order chaos [116, Theorem 2.3].

The error bounds for Algorithms 1 and 2 in the next section rely on Theorems 4.3.7 and 4.3.8, and existing analysis of power iteration [42] and truncated power iteration [44]. Additionally, the perturbation bounds in this section are of independent interest. In particular, Theorem 4.3.7 shows that if the assumptions and the prescribed sample complexities in (4.10) are satisfied, then with high probability the principal eigenvector $\hat{\eta}$ of G is an accurate estimate of the vector $\dot{\eta}$ that concatenates the unknown variables. It gives an error bound for any algorithm that finds the principal eigenvector of G. On the other hand, while Theorem 4.3.8 does not directly guarantee the success of any particular algorithm, it can be used to analyze other algorithms that find the sparse principal component of G, similar to the analysis of Algorithm

2 in Theorem 4.3.10.

4.3.3 Error Bounds for the Power Iteration Algorithms

In this section, we give performance guarantees for Algorithms 1 and 2 under the assumptions stated in Section 4.3.1. Under the conditions in Theorem 4.3.9 (resp. Theorem 4.3.10), the iterates in Algorithm 1 (resp. Algorithm 2), in the noiseless case, converge linearly to the true solution. In the noisy case, the recovery error is proportional to the noise level.

Theorem 4.3.9 (**Subspace Case**). *Suppose Assumptions 4.3.1 – 4.3.4 are satisfied with $\delta < 1/4$ and a sufficiently small absolute constant $C_W > 0$. Let $\alpha = \sqrt{n}$, and $\beta = 3/2$. Assume that $\xi := |\hat{\eta}^* \eta^{(0)}| > 0$. Then there exist absolute constants $c, C, C' > 0$, such that if (4.10) is satisfied, then with probability at least $1 - 2n^{-c} - e^{-cm}$, the iterates in Algorithm 1 satisfy*

$$d(\eta^{(t)},\ \dot{\eta}) \leq \rho^t d(\eta^{(0)},\ \dot{\eta}) + 2\Delta,$$

where Δ is defined in (4.11), and

$$\rho := \Big\{1 - \frac{1}{2}\Big[1 - \Big(\frac{1+6\delta}{3-2\delta}\Big)^2\Big]\xi(1+\xi)\Big\}^{1/2}. \tag{4.15}$$

Theorem 4.3.9 shows that the power iteration algorithm requires $n = O(m \log^2(Nm + n))$ sensors and $N = O(\log(Nm + n))$ snapshots to successfully recover X and λ. This agrees, up to log factors, with the sample complexity required for the uniqueness of (λ, X) in the subspace case, which is $n > m$ and $N \geq \frac{n-1}{n-m}$ [3].

Next, we compare Theorem 4.3.9 with a similar error bound for the least squares approach by Ling and Strohmer [40, Theorem 3.5]. The sample complexity in Theorem 4.3.9 matches the numbers required by the least squares approach $n = O(m \log^2(Nm + n))$ and $N = O(\log^2(Nm + n))$ (up to one log factor). One caveat in the least squares approach is that, apart from the linear equation (4.2), it needs an extra linear constraint to avoid the trivial solution $\gamma = 0$, $X = 0$. Unfortunately, as revealed by [40, Theorem 3.5], in the noisy setting, the recovery error by the least squares approach is sensitive to this extra linear constraint. Our numerical experiments (Section 4.6) show that power iteration outperforms least squares in the noisy setting.

Theorem 4.3.10 (**Joint Sparsity Case**). *Suppose Assumptions 4.3.1 – 4.3.4 are satisfied with $\delta < 1/4$ and a sufficiently small absolute constant $C_W > 0$. Let $\alpha = \sqrt{n}$, $\beta = 3/2$, $s_1 \geq s_0$ in Algorithm 2, and define $s = s_0 + 2s_1$. Then there exist absolute constants $c, C, C' > 0$, such that if $|\dot{\eta}^* \eta^{(0)}| \geq \xi + \widetilde{\Delta}$ for some $\xi \in (0, 1)$, and (4.13) is satisfied, then with probability at least $1 - 2n^{-c} - m^{-cs}$, the iterates in Algorithm 2 for the* joint *sparsity case satisfy*

$$d(\eta^{(t)},\ \dot{\eta}) \leq \tilde{\rho}^t d(\eta^{(0)},\ \dot{\eta}) + \frac{2\sqrt{5}\widetilde{\Delta}}{1-\tilde{\rho}},$$

where $\widetilde{\Delta}$ is defined in (4.14)*, and $\tilde{\rho} < 1$ has the following expression:*

$$\tilde{\rho} \coloneqq \rho \cdot \left(1 + 2\sqrt{\frac{s_0}{s_1}} + \frac{2s_0}{s_1}\right)^{1/2}, \tag{4.16}$$

and ρ is defined in (4.15)*.*

Theorem 4.3.10 is only valid when $\tilde{\rho} < 1$. With the choice $s_1 = 2s_0$, when δ approaches 0, and ξ approaches 1, the convergence rate $\tilde{\rho}$ is roughly $\frac{1}{3}\sqrt{1+\sqrt{2}+2} \approx 0.62$. We discuss a more realistic scenario next.

Remark 4.3.11. *A good initialization for λ alone is usually sufficient. Suppose one has a good initial estimate for the gains and phases, i.e., λ satisfies $|\lambda_k - e^{\sqrt{-1}\varphi_k}| < \sqrt{1+\delta} - 1$ for known phase estimates $\{\varphi_k\}_{k=1}^n$. One can initialize Algorithm 2 with $\eta^{(0)} = [\mathbf{0}_{Nm,1}^\top, e^{-\sqrt{-1}\varphi_1}, \ldots, e^{-\sqrt{-1}\varphi_n}]^\top$, then when Δ is negligible (noiseless or low noise settings), ξ in Theorem 4.3.10 can be set to $1/\sqrt{(1+\delta)(2+\delta)}$. For example, if $\delta = 0.05$ and $s_1 \geq 10s_0$, then $\tilde{\rho} < 1$. Since we do not attempt to optimize the constants in this chapter, the constants in this exemplary scenario are conservative.*

Theorem 4.3.10 states that for Algorithm 2 to recover λ and a jointly sparse X, it is sufficient to have $n = O(s_0 \log^8 n \log^2(s_0 N + m))$ sensors and $N = O(\sqrt{s_0} \log^2 n \log(s_0 N + m))$ snapshots. In comparison, the (up to a factor of 2) optimal sample complexity for unique recovery in the joint sparsity case is $n > 2s_0$ and $N \geq \frac{n-1}{n-2s_0}$ [3]. Hence, the number of sensors required in Theorem 4.3.10 is (up to log factors) optimal, but the number of snapshots required is suboptimal. Another drawback is that these results apply only to the joint sparsity case, and not to the more general sparsity

case. However, we believe these drawbacks are due to artifacts of our analysis. For both the joint sparsity case and the sparsity case, we have Nn complex-valued measurements, and $Ns_0 + n - 1$ complex-valued unknowns. One may expect successful recovery when n and N are (up to log factors) on the order of s_0 and 1, respectively. In fact, numerical experiments in Section 4.6 confirms that truncated power iteration successfully recovers λ and X in this regime for the more general sparsity case.

Wang and Chi [41] analyzed the performance of ℓ_1 minimization for BGPC in the sparsity case, where they assumed that A is the DFT matrix, and X is a Bernoulli–sub-Gaussian random matrix. Their sample complexity for ℓ_1 minimization is $n = O(s)$ and $N = O(n \log^4 n)$. The success of their algorithm relies on a restrictive assumption that $\lambda_k \approx 1$, which is analogous to the dependence of our algorithm on a good initialization of λ_k. In the next section, we show that such dependence can be relaxed under some additional conditions using the initialization provided by Algorithm 3.

4.3.4 A Theoretical Guarantee of the Initialization

The next theorem shows that, under certain conditions, Algorithm 3 recovers the locations of the large entries in X correctly, and yields an initial estimate $\eta^{(0)}$ that satisfies $|\dot{\eta}^* \eta^{(0)}| > 1 - 2\delta$ (close to 1).

Theorem 4.3.12 (Initialization)**.** *Suppose Assumptions 4.3.1 – 4.3.5 are satisfied. Then there exist absolute constants $C'', c' > 0$, such that if*

$$n > C'' s_0^2 \log^6(nmN),$$

then with probability at least $1 - n^{-c''}$, for all $j \in [N]$ the set T_j' in Assumption 4.3.5 is a subset of T_j in Algorithm 3. Additionally, in the joint sparsity case, if sample complexity (4.13) is satisfied with a sufficiently large C, Assumption 4.3.4 is satisfied with a sufficiently small C_W, and Assumption 4.3.5 is satisfied with a sufficiently small δ_X, then η_0 produced by Algorithm 3 will satisfy that $|\dot{\eta}^* \eta^{(0)}|$ is arbitrarily close to

$$\frac{n^{3/2} + \|\lambda\|_2 \|\gamma\|_2^2}{\sqrt{n^2 + \|\lambda\|_2^2 \|\gamma\|_2^2} \sqrt{n + \|\gamma\|_2^2}} > 1 - 2\delta.$$

By Theorem 4.3.12, the constant ξ in Theorem 4.3.10 can be set to $1 - 2\delta$ in a low noise setting. For $\delta < 0.19$, this constant ξ is larger than the one in Remark 4.3.11, and allows $\tilde{\rho} < 1$ for more choices of s_1.

Our guarantee for the initialization requires that the number n of sensors scales quadratically (up to log factors) in the sparsity s_0, which seems suboptimal. Similar suboptimal sampling complexities show up in sparse PCA [117] and sparse phase retrieval [102, 104, 118].

In the joint sparsity case, instead of estimating the supports of $x_{\cdot 1}, x_{\cdot 2}, \ldots, x_{\cdot N}$ separately, one can estimate the row support of X directly by sorting

$$\sum_{j \in [N]} \left\| d^*_{\cdot((j-1)m+\ell)} E \right\|_2^2$$

for $\ell \in [m]$ and finding the s_1 largest. In this case, Assumption 4.3.5 can be changed to: There exists a subset T' of large rows (in terms of ℓ_2 norm), such that for all $\ell \in T'$,

$$\frac{\sum_{j \in [N]} |x_{\ell j}|^2}{\|X\|_{\mathrm{F}}^2} \geq \frac{\omega}{s_0},$$

and

$$\frac{\sum_{j \in [N], \ell' \in [m] \setminus T'} |x_{\ell' j}|^2}{\|X\|_{\mathrm{F}}^2} \leq \delta_X.$$

In this case, the subset T' can be identified and an initialization $\eta^{(0)}$ can be computed under the same conditions as in Theorem 4.3.12, which can be proved using the same arguments.

4.4 Fundamental Estimates

To prove the main results, we must first establish some fundamental estimates specific to BGPC. Proofs of some lemmas in this section can be found in Appendix B.

4.4.1 A Gap in Eigenvalues

A key component in establishing a perturbation bound for an eigenvector problem (e.g., Theorem 4.3.7) is bounding the gap between eigenvalues. Lemma 4.4.1 gives us such a bound.

Lemma 4.4.1. *Suppose Assumptions 4.3.1 – 4.3.3 are satisfied and* $\alpha = \sqrt{n}$. *Then the smallest eigenvalue of* $\mathbb{E}\Omega_{T_\eta} B_{\mathrm{s}} \Omega_{T_\eta}^*$ *is* 0, *and the rest of the eigenvalues reside in the interval* $[\frac{(1-\delta)^2}{1+\delta}, 2(1+\delta)]$.

4.4.2 Perturbation Due to Randomness in A

Next, we show that $\Omega_{T_\eta} B_{\mathrm{s}} \Omega_{T_\eta}^*$, whose randomness comes from A, is close to its mean $\mathbb{E}\Omega_{T_\eta} B_{\mathrm{s}} \Omega_{T_\eta}^*$ under certain conditions.

Lemma 4.4.2. *Suppose Assumptions 4.3.1 – 4.3.3 are satisfied, and* $\alpha = \sqrt{n}$. *For any constant* $\delta_B > 0$, *there exist absolute constants* $C, c > 0$, *such that:*

- ***Subspace case:*** *If* (4.10) *is satisfied with* C, *then*

$$\|B_{\mathrm{s}} - \mathbb{E}B_{\mathrm{s}}\| \leq \delta_B$$

  with probability at least $1 - n^{-c} - e^{-cm}$.

- ***Joint sparsity case:*** *If* (4.13) *is satisfied with* C, *then*

$$\left\|\Omega_{T_\eta} B_{\mathrm{s}} \Omega_{T_\eta}^* - \mathbb{E}\Omega_{T_\eta} B_{\mathrm{s}} \Omega_{T_\eta}^*\right\| \leq \delta_B$$

  for all $T_1 = \cdots = T_N$ *and* T_η *defined in* (4.9), *with probability at least* $1 - n^{-c} - m^{-cs}$.

Proof of Lemma 4.4.2. Recall that

$$\Omega_{T_\eta} B_{\mathrm{s}} \Omega_{T_\eta}^* = \begin{bmatrix} \Omega_{T_x} D^* D \Omega_{T_x}^* & \sqrt{n}\Omega_{T_x} D^* E_{\mathrm{s}} \\ \sqrt{n} E_{\mathrm{s}}^* D \Omega_{T_x}^* & n E_{\mathrm{s}}^* E_{\mathrm{s}} \end{bmatrix}.$$

It follows that

$$\begin{aligned} &\left\|\Omega_{T_\eta} B_{\mathrm{s}} \Omega_{T_\eta}^* - \mathbb{E}\Omega_{T_\eta} B_{\mathrm{s}} \Omega_{T_\eta}^*\right\| \\ &\leq \left\|\Omega_{T_x} D^* D \Omega_{T_x}^* - \mathbb{E}\Omega_{T_x} D^* D \Omega_{T_x}^*\right\| && (4.17) \\ &+ n \left\|E_{\mathrm{s}}^* E_{\mathrm{s}} - \mathbb{E}E_{\mathrm{s}}^* E_{\mathrm{s}}\right\| && (4.18) \\ &+ 2\sqrt{n} \left\|\Omega_{T_x} D^* E_{\mathrm{s}} - \mathbb{E}\Omega_{T_x} D^* E_{\mathrm{s}}\right\|. && (4.19) \end{aligned}$$

Lemma 4.4.2 follows from the bounds on the spectral norms in (4.17) – (4.19) in Lemmas 4.4.3 – 4.4.6, respectively. □

Lemma 4.4.3. *Suppose Assumption 4.3.1 is satisfied, then there exist absolute constants $C_1, c_1 > 0$, such that:*

- ***Subspace case:***

$$\|D^*D - \mathbb{E}D^*D\| \leq C_1\sqrt{\frac{m}{n}},$$

  with probability at least $1 - e^{-c_1 m}$.

- ***Joint sparsity case:*** *For any $\{T_j\}_{j=1}^N$ and T_x defined in (4.8),*

$$\left\|\Omega_{T_x}D^*D\Omega_{T_x}^* - \mathbb{E}\Omega_{T_x}D^*D\Omega_{T_x}^*\right\| \leq C_1\sqrt{\frac{s}{n}\log m},$$

  with probability at least $1 - m^{-c_1 s}$.

Lemma 4.4.4. *Suppose Assumptions 4.3.1 – 4.3.3 are satisfied, then there exist absolute constants $C_2, c_2 > 0$, such that*

- ***Subspace case:***

$$\|E_{\mathrm{s}}^*E_{\mathrm{s}} - \mathbb{E}E_{\mathrm{s}}^*E_{\mathrm{s}}\| \leq \frac{C_2}{n}\max\Big\{\sqrt{\frac{\log n}{N}}, \sqrt{\frac{\log n}{m}}, \frac{\log n}{N}, \frac{\log n}{m}\Big\},$$

- ***Joint sparsity case:***

$$\|E_{\mathrm{s}}^*E_{\mathrm{s}} - \mathbb{E}E_{\mathrm{s}}^*E_{\mathrm{s}}\| \leq \frac{C_2}{n}\max\Big\{\sqrt{\frac{\log n}{N}}, \sqrt{\frac{\log n}{s_0}}, \frac{\log n}{N}, \frac{\log n}{s_0}\Big\},$$

with probability at least $1 - n^{-c_2}$.

Lemma 4.4.5 (Subspace Case). *Suppose Assumptions 4.3.1 – 4.3.3 are satisfied, and $\min\{N, m\} > \log n$, then there exist absolute constants $C_3, c_3 > 0$, such that*

$$\|D^*E_{\mathrm{s}} - \mathbb{E}D^*E_{\mathrm{s}}\| \leq C_3\max\Big\{\sqrt{\frac{\log(Nm+n)}{nN}}, \sqrt{\frac{\log(Nm+n)}{nm}}, \frac{\sqrt{m}\log(Nm+n)}{n}\Big\},$$

with probability at least $1 - n^{-c_3}$.

Lemma 4.4.6 (**Joint Sparsity Case**). *Suppose Assumptions 4.3.1 – 4.3.3 are satisfied, then there exist absolute constants* $C_3, c_3 > 0$*, such that for all* $T_1 = \cdots = T_N$*,*

$$\begin{aligned}&\|\Omega_{T_x} D^* E_{\mathrm{s}} - \mathbb{E}\Omega_{T_x} D^* E_{\mathrm{s}}\| \\ &\leq \frac{C_3 s_0^{1/4}(s+N)^{1/4}(\sqrt{n}+\sqrt{s+N})^{1/2}}{n\min\{\sqrt{s_0},\sqrt{N}\}} \log^3 n \log(sN+m),\end{aligned}$$

with probability at least $1 - n^{-c_3}$.

4.4.3 Perturbation Due to Noise

We established some fundamental estimates regarding B_{s} in Sections 4.4.1 and 4.4.2. In this section, we turn to perturbation caused by noise. By the definitions of B, B_{s}, E, E_{s}, and E_{n}, we have

$$B = B_{\mathrm{s}} + B_{\mathrm{n}},$$

where

$$B_{\mathrm{n}} \coloneqq \begin{bmatrix} 0 & \alpha D^* E_{\mathrm{n}} \\ \alpha E_{\mathrm{n}}^* D & \alpha^2(E_{\mathrm{s}}^* E_{\mathrm{n}} + E_{\mathrm{n}}^* E_{\mathrm{s}} + E_{\mathrm{n}}^* E_{\mathrm{n}}) \end{bmatrix}.$$

Therefore,

$$\begin{aligned}&\Omega_{T_\eta} B_{\mathrm{n}} \Omega_{T_\eta}^* \\ &= \begin{bmatrix} 0 & \alpha \Omega_{T_x} D^* E_{\mathrm{n}} \\ \alpha E_{\mathrm{n}}^* D \Omega_{T_x}^* & \alpha^2(E_{\mathrm{s}}^* E_{\mathrm{n}} + E_{\mathrm{n}}^* E_{\mathrm{s}} + E_{\mathrm{n}}^* E_{\mathrm{n}}) \end{bmatrix}.\end{aligned}$$

Lemma 4.4.7 gives an upper bound on the spectral norm of the perturbation from noise.

Lemma 4.4.7. *Suppose Assumptions 4.3.1 – 4.3.3 are satisfied. Let* $\alpha = \sqrt{n}$ *and let* ν *be defined by* (4.12)*. Then there exist absolute constants* $c, C, C' > 0$ *such that:*

- ***Subspace case:*** *If* (4.10) *is satisfied, then with probability at least*

$1 - n^{-c}$

$$\|B_{\mathrm{n}}\| \leq C' \max\{\nu,\ \nu^2\}.$$

Additionally, for any constant $\delta_W > 0$, there exists an absolute constant $C_W > 0$, if Assumption 4.3.4 is satisfied with C_W, then the above bound becomes

$$\|B_{\mathrm{n}}\| \leq \delta_W.$$

- ***Joint sparsity case:*** *If* (4.13) *is satisfied, then with probability at least* $1 - n^{-c}$

$$\left\|\Omega_{T_\eta} B_{\mathrm{n}} \Omega_{T_\eta}^*\right\| \leq C' \max\{N^{3/2}\nu,\ \nu^2\}$$

for all $T_1 = \cdots = T_N$ and T_η defined in (4.9)*. Additionally, for any constant $\delta_W > 0$, there exists an absolute constant $C_W > 0$, if Assumption 4.3.4 is satisfied with C_W, then the above bound becomes*

$$\left\|\Omega_{T_\eta} B_{\mathrm{n}} \Omega_{T_\eta}^*\right\| \leq \delta_W.$$

Proof. To complete the proof, we bound the spectral norms of $\Omega_{T_x} D^* E_{\mathrm{n}}$, $E_{\mathrm{s}}^* E_{\mathrm{n}}$, and $E_{\mathrm{n}}^* E_{\mathrm{n}}$ in Lemmas 4.4.8, 4.4.10, and 4.4.11, respectively. □

Lemma 4.4.8 (**Subspace Case**)**.** *Suppose Assumption 4.3.1 is satisfied, and $m > \log n$, then there exist absolute constants $C_4, c_4 > 0$, such that*

$$\|D^* E_{\mathrm{n}}\| \leq C_4 \max\Big\{\sqrt{\log(Nm+n)}, \sqrt{\frac{Nm}{n}}\log(Nm+n)\Big\} \max_{k\in[n], j\in[N]} |w_{kj}|,$$

with probability at least $1 - n^{-c_4}$.

Lemma 4.4.9 (**Joint Sparsity Case**)**.** *Suppose Assumption 4.3.1 is satisfied, then there exist absolute constants $C_4, c_4 > 0$, such that for all $T_1 = \cdots = T_N$,*

$$\|\Omega_{T_x} D^* E_{\mathrm{n}}\| \leq C_4\big(\sqrt{s}N + \sqrt{sN\log m} + \sqrt{N\log^3 n}\big) \cdot \sqrt{\log n} \max_{k\in[n], j\in[N]} |w_{kj}|,$$

with probability at least $1 - n^{-c_4}$.

Lemma 4.4.10. *Suppose Assumptions 4.3.1 – 4.3.3 are satisfied, then there exist absolute constants $C_5, c_5 > 0$, such that*

- ***Subspace case:***

$$\|E_{\mathrm{s}}^* E_{\mathrm{n}}\| \le C_5 \sqrt{\frac{N}{n}} \max\Big\{1, \sqrt{\frac{\log n}{N}}, \sqrt{\frac{\log n}{m}}\Big\} \cdot \max_{k\in[n], j\in[N]} |w_{kj}|,$$

- ***Joint sparsity case:***

$$\|E_{\mathrm{s}}^* E_{\mathrm{n}}\| \le C_5 \sqrt{\frac{N}{n}} \max\Big\{1, \sqrt{\frac{\log n}{N}}, \sqrt{\frac{\log n}{s_0}}\Big\} \cdot \max_{k\in[n], j\in[N]} |w_{kj}|,$$

with probability at least $1 - n^{-c_5}$.

Lemma 4.4.11.

$$\|E_{\mathrm{n}}^* E_{\mathrm{n}}\| \le N \max_{k\in[n], j\in[N]} |w_{kj}|^2,$$

4.4.4 Scalar Concentration

We now introduce a few scalar concentration bounds that are useful in the proof of Theorem 4.3.12.

Lemma 4.4.12. *Suppose Assumptions 4.3.1 – 4.3.4 is satisfied, then there exist absolute constants $C_6, c_6 > 0$, such that for all $j \in [N]$ and $\ell \in [m]$, we have*

$$\left| \sum_{k\in[n]} \left(|\lambda_k \overline{a_{k\ell}} a_{k\cdot}^\top x_{\cdot j}|^2 - \mathbb{E}|\lambda_k \overline{a_{k\ell}} a_{k\cdot}^\top x_{\cdot j}|^2 \right) \right| \le \frac{C_6 \|x_{\cdot j}\|_2^2 \log^3(nmN)}{n^{3/2}}, \tag{4.20}$$

$$\begin{aligned} &\left| \sum_{k\in[n]} \lambda_k a_{k\ell} \overline{a_{k\ell}} a_{k\cdot}^\top x_{\cdot j} \overline{w_{kj}} \right| \le \frac{C_6 \|x_{\cdot j}\|_2 \log^2(nmN)}{n} \max_{k\in[n], j\in[N]} |w_{kj}| \\ &\le \frac{C_6 C_W \|x_{\cdot j}\|_2^2 \log^2(nmN)}{\sqrt{1-\theta} n^{3/2}}, \end{aligned} \tag{4.21}$$

and

$$\left|\sum_{k\in[n]}\left(|\overline{a_{k\ell}}w_{kj}|^2-\mathbb{E}|\overline{a_{k\ell}}w_{kj}|^2\right)\right|\le\frac{C_6\log^2(nmN)}{n^{1/2}}\max_{k\in[n],j\in[N]}|w_{kj}|^2$$
$$\le\frac{C_6C_W^2\left\|x_{\cdot j}\right\|_2^2\log^2(nmN)}{(1-\theta)n^{3/2}},\tag{4.22}$$

with probability at least $1-n^{-c_6}$.

4.5 Proofs of the Main Results

4.5.1 Proof of the Perturbation Bound for the Eigenvector Problem

In this section, we prove Theorem 4.3.7. Theorem 4.3.8 can be proved similarly.

Proof of Theorem 4.3.7. First,

$$G=\beta I_{Nm+n}-B=(\beta I_{Nm+n}-\mathbb{E}B_{\mathrm{s}})-(B_{\mathrm{s}}-\mathbb{E}B_{\mathrm{s}})-B_{\mathrm{n}}.\tag{4.23}$$

Lemma 4.4.1 establishes a gap in the eigenvalues of the matrix $\mathbb{E}B_{\mathrm{s}}$ – the smallest and the second-smallest eigenvalues of $\mathbb{E}B_{\mathrm{s}}$ are separated by a gap of at least

$$\frac{(1-\delta)^2}{1+\delta}\ge 1-3\delta>0.$$

Therefore, the gap between the largest and the second-largest eigenvalues of $\beta I_{Nm+n}-\mathbb{E}B_{\mathrm{s}}$ is at least $1-3\delta$. By Lemmas 4.4.2 and 4.4.7, there exist absolute constants $c,C,C',C_W>0$ such that if all the assumptions are satisfied, then with probability at least $1-2n^{-c}-e^{-cm}$,

$$\|(B_{\mathrm{s}}-\mathbb{E}B_{\mathrm{s}})+B_{\mathrm{n}}\|\le\|B_{\mathrm{s}}-\mathbb{E}B_{\mathrm{s}}\|+\|B_{\mathrm{n}}\|\le\frac{1-3\delta}{4},\tag{4.24}$$

$$\|B_{\mathrm{n}}\|\le C'\max\{\nu,\ \nu^2\}.\tag{4.25}$$

Recall that $\dot{\eta}$ is the principal eigenvector of $\beta I_{Nm+n}-\mathbb{E}B_{\mathrm{s}}$. By the Davis-

Kahan $\sin\theta$ Theorem ([114]; see also [42, Theorem 8.1.12]), (4.24) and (4.25) imply

$$\begin{aligned}\sin\angle(\dot{\eta},\hat{\eta}) &\leq \frac{4}{1-3\delta}\left\|(B_{\mathrm{s}}-\mathbb{E}B_{\mathrm{s}}+B_{\mathrm{n}})\dot{\eta}\right\|_2 \\ &\leq \frac{4}{1-3\delta}\left\|B_{\mathrm{n}}\right\| \leq \frac{4C'}{1-3\delta}\max\{\nu,\ \nu^2\},\end{aligned}$$

where the second inequality is due to $B_{\mathrm{s}}\dot{\eta}=\mathbb{E}B_{\mathrm{s}}\dot{\eta}=0$.

Theorem 4.3.7 follows from the above bound, and the fact that

$$d(\dot{\eta},\hat{\eta})=\sqrt{2-2\cos\angle(\dot{\eta},\hat{\eta})}=2\sin\frac{\angle(\dot{\eta},\hat{\eta})}{2}\leq 2\sin\angle(\dot{\eta},\hat{\eta}).$$

□

One can prove Theorem 4.3.8 using the same steps as in the proof of Theorem 4.3.7, by restricting rows and columns of matrices to the support T_η and applying the corresponding uniform bounds on submatrices.

4.5.2 Proof of the Error Bound for Algorithm 1

Proof of Theorem 4.3.9. Recall that the largest eigenvalue of $\beta I_{Nm+n}-\mathbb{E}B_{\mathrm{s}}$ is $\beta-0=\frac{3}{2}$, and all other eigenvalues reside in the interval $[\frac{3}{2}-2(1+\delta),\frac{3}{2}-\frac{(1-\delta)^2}{1+\delta}]$. By Lemmas 4.4.2 and 4.4.7, there exist constants $c,C,C_W>0$ such that

$$\left\|(B_{\mathrm{s}}-\mathbb{E}B_{\mathrm{s}})+B_{\mathrm{n}}\right\|\leq\left\|B_{\mathrm{s}}-\mathbb{E}B_{\mathrm{s}}\right\|+\left\|B_{\mathrm{n}}\right\|\leq\min\Big\{\delta,\frac{(1-\delta)^2}{1+\delta}+3\delta-1\Big\},$$

with probability at least $1-2n^{-c}-e^{-cm}$. By (4.23), the largest eigenvalue of G is $\|G\|\geq\frac{3}{2}-\delta$, the corresponding eigenvector is $\hat{\eta}$, and all the other eigenvalues of G reside in the interval $[-\frac{1}{2}-3\delta,\frac{1}{2}+3\delta]$.

Next, we establish the convergence rate of power iterations for BGPC. By the eigenvalue decomposition of G and the Pythagorean theorem,

$$G\hat{\eta}=\|G\|\,\hat{\eta},$$

$$\left\|G\eta^{(t-1)}\right\|\leq\sqrt{\|G\|^2\,|\hat{\eta}^*\eta^{(t-1)}|^2+\Big(\frac{1}{2}+3\delta\Big)^2(1-|\hat{\eta}^*\eta^{(t-1)}|^2)}.$$

Therefore,

$$
\begin{aligned}
|\hat{\eta}^* \eta^{(t)}| &= \frac{|\hat{\eta}^* G \eta^{(t-1)}|}{\|G\eta^{(t-1)}\|_2} \\
&\geq \frac{\|G\| \, |\hat{\eta}^* \eta^{(t-1)}|}{\sqrt{\|G\|^2 \, |\hat{\eta}^* \eta^{(t-1)}|^2 + (\frac{1}{2} + 3\delta)^2 (1 - |\hat{\eta}^* \eta^{(t-1)}|^2)}} \\
&\geq |\hat{\eta}^* \eta^{(t-1)}| \frac{1}{\sqrt{|\hat{\eta}^* \eta^{(t-1)}|^2 + (\frac{1+6\delta}{3-2\delta})^2 (1 - |\hat{\eta}^* \eta^{(t-1)}|^2)}} \\
&= |\hat{\eta}^* \eta^{(t-1)}| \frac{1}{\sqrt{1 - \big(1 - (\frac{1+6\delta}{3-2\delta})^2\big)(1 - |\hat{\eta}^* \eta^{(t-1)}|^2)}} \\
&\geq |\hat{\eta}^* \eta^{(t-1)}| \Big[1 + \frac{1}{2}\big(1 - (\frac{1+6\delta}{3-2\delta})^2\big)(1 - |\hat{\eta}^* \eta^{(t-1)}|^2)\Big],
\end{aligned}
$$

where the last inequality is due to $\frac{1}{\sqrt{1-z}} \geq 1 + \frac{1}{2}z$ for $z \in (0, 1)$. It follows that

$$
\begin{aligned}
&[1 - |\hat{\eta}^* \eta^{(t)}|] \leq [1 - |\hat{\eta}^* \eta^{(t-1)}|] \\
&\times \Big[1 - \frac{1}{2}\big(1 - (\frac{1+6\delta}{3-2\delta})^2\big)|\hat{\eta}^* \eta^{(t-1)}|(1 + |\hat{\eta}^* \eta^{(t-1)}|)\Big]. \qquad (4.26)
\end{aligned}
$$

Clearly, $\{|\hat{\eta}^* \eta^{(\tau)}|\}_{\tau=0}^{t}$ is monotonically increasing unless $|\hat{\eta}^* \eta^{(0)}| = 0$. By the definition $\xi \coloneqq |\hat{\eta}^* \eta^{(0)}|$, the convergence rate in (4.26) is bounded by $\rho^2 < 1$. It follows that

$$
[1 - |\hat{\eta}^* \eta^{(t)}|] \leq \rho^2 [1 - |\hat{\eta}^* \eta^{(t-1)}|] \leq \rho^{2t} \cdot [1 - |\hat{\eta}^* \eta^{(0)}|].
$$

Hence

$$
d(\hat{\eta}, \eta^{(t)}) \leq \rho^t \cdot d(\hat{\eta}, \eta^{(0)}).
$$

By Theorem 4.3.7, for $\tau = 0, \ldots, t$

$$
d(\dot{\eta}, \hat{\eta}) \leq \Delta.
$$

It follows from the triangle inequality that

$$
d(\dot{\eta}, \eta^{(t)}) \leq \rho^t \cdot d(\dot{\eta}, \eta^{(0)}) + 2\Delta.
$$

□

4.5.3 Proof of the Error Bound for Algorithm 2

Proof of Theorem 4.3.10. In the joint sparsity case, any iterate

$$\eta^{(\tau)} = [x^{(\tau)\top}, -\gamma^{(\tau)\top}/\alpha]^\top$$

satisfies that $x^{(\tau)}$ is the concatenation of jointly sparse $\{x_{\cdot j}^{(\tau)}\}_{j=1}^N$. In the t-th iteration, we define a support set $T^{(t)}$ that has cardinality $s = s_0 + 2s_1$, and satisfies

$$\mathrm{supp}(x_{\cdot j}) \bigcup \mathrm{supp}(x_{\cdot j}^{(t-1)}) \bigcup \mathrm{supp}(x_{\cdot j}^{(t)}) \subset T^{(t)},$$

for all $j \in [N]$. Define $T_\eta^{(t)}$ using (4.8) and (4.9) with $T_1 = \cdots = T_N = T^{(t)}$. Next, we focus on the submatrix $\Omega_{T_\eta^{(t)}} G \Omega_{T_\eta^{(t)}}^*$ and subvectors $\Omega_{T_\eta^{(t)}} \dot{\eta}$ and $\Omega_{T_\eta^{(t)}} \eta^{(t)}$, etc. Since the supports of $\eta^{(t)}$ and $\dot{\eta}$ are subsets of $T_\eta^{(t)}$, we have $|\dot{\eta}^* \Omega_{T_\eta^{(t)}}^* \Omega_{T_\eta^{(t)}} \eta^{(t)}| = |\dot{\eta}^* \eta^{(t)}|$.

We prove by induction that $\{|\dot{\eta}^* \eta^{(\tau)}|\}_{\tau=0}^t$ is monotonically increasing (until it crosses a threshold specified later in the proof). Suppose $\{|\dot{\eta}^* \eta^{(\tau)}|\}_{\tau=0}^{t-1}$ is monotonically increasing. Next, we prove

$$|\dot{\eta}^* \eta^{(t)}| > |\dot{\eta}^* \eta^{(t-1)}|.$$

By the assumption that $|\dot{\eta}^* \eta^{(0)}| \geq \xi + \widetilde{\Delta}$ and Theorem 4.3.8, we have

$$\begin{aligned} &|\hat{\eta}_{T_\eta^{(t)}}^* \Omega_{T_\eta^{(t)}} \eta^{(t-1)}| \geq |\dot{\eta}^* \eta^{(t-1)}| - d(\Omega_{T_\eta^{(t)}} \dot{\eta},\ \hat{\eta}_{T_\eta^{(t)}}) \\ &\geq \xi + \widetilde{\Delta} - \widetilde{\Delta} = \xi. \end{aligned}$$

Following the same steps in the proof of Theorem 4.3.9, we obtain a bound

for $\tilde{\eta}^{(t)}$ similar to (4.26):

$$\begin{aligned}
&[1 - |\hat{\eta}^*_{T_\eta^{(t)}} \Omega_{T_\eta^{(t)}} \tilde{\eta}^{(t)}|] \\
&\leq [1 - |\hat{\eta}^*_{T_\eta^{(t)}} \Omega_{T_\eta^{(t)}} \eta^{(t-1)}|] \Big[1 - \frac{1}{2}\big(1 - (\frac{1+6\delta}{3-2\delta})^2\big) \\
&\qquad |\hat{\eta}^*_{T_\eta^{(t)}} \Omega_{T_\eta^{(t)}} \eta^{(t-1)}|(1 + |\hat{\eta}^*_{T_\eta^{(t)}} \Omega_{T_\eta^{(t)}} \eta^{(t-1)}|)\Big] \\
&\leq [1 - |\hat{\eta}^*_{T_\eta^{(t)}} \Omega_{T_\eta^{(t)}} \eta^{(t-1)}|] \Big[1 - \frac{1}{2}\big(1 - (\frac{1+6\delta}{3-2\delta})^2\big)\xi(1+\xi)\Big] \\
&= \rho^2 [1 - |\hat{\eta}^*_{T_\eta^{(t)}} \Omega_{T_\eta^{(t)}} \eta^{(t-1)}|],
\end{aligned}$$

where ρ is defined in (4.15). It follows that

$$d(\hat{\eta}_{T_\eta^{(t)}},\ \Omega_{T_\eta^{(t)}} \tilde{\eta}^{(t)}) \leq \rho \cdot d(\hat{\eta}_{T_\eta^{(t)}},\ \Omega_{T_\eta^{(t)}} \eta^{(t-1)}).$$

We use the perturbation bound in Theorem 4.3.8 one more time:

$$d(\Omega_{T_\eta^{(t)}} \dot{\eta},\ \Omega_{T_\eta^{(t)}} \tilde{\eta}^{(t)}) \leq \rho \cdot d(\Omega_{T_\eta^{(t)}} \dot{\eta},\ \Omega_{T_\eta^{(t)}} \eta^{(t-1)}) + 2\widetilde{\Delta}.$$

Equivalently,

$$\sqrt{1 - |\dot{\eta}^* \tilde{\eta}^{(t)}|} \leq \rho \sqrt{1 - |\dot{\eta}^* \eta^{(t-1)}|} + \sqrt{2}\widetilde{\Delta}. \tag{4.27}$$

Next, we show that the truncation step amplifies the error only by a small factor. The vector $\widetilde{\Pi}_{s_1}(\tilde{\eta}^{(t)})$ is the projection of $\tilde{\eta}^{(t)}$ onto the set of structured sparse vectors, and $\eta^{(t)}$ is the normalized version. We define three index sets

$$\begin{aligned}
T_a &= \mathrm{supp}(\dot{\eta}) \backslash \mathrm{supp}(\eta^{(t)}), \\
T_b &= \mathrm{supp}(\dot{\eta}) \bigcap \mathrm{supp}(\eta^{(t)}), \\
T_c &= \mathrm{supp}(\eta^{(t)}) \backslash \mathrm{supp}(\dot{\eta}).
\end{aligned}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned}
|\dot{\eta}^* \tilde{\eta}^{(t)}|^2 &\leq \big\|\Omega_{T_a} \tilde{\eta}^{(t)}\big\|_2^2 + \big\|\Omega_{T_b} \tilde{\eta}^{(t)}\big\|_2^2 \\
&\leq 1 - \big\|\Omega_{T_c} \tilde{\eta}^{(t)}\big\|_2^2 \leq 1 - \frac{|T_c|}{|T_a|} \big\|\Omega_{T_a} \tilde{\eta}^{(t)}\big\|_2^2,
\end{aligned}$$

where the last inequality is due to projection rule, i.e., $\widetilde{\Pi}_{s_1}(\tilde{\eta}^{(t)})$ keeps the largest entries of $\tilde{\eta}^{(t)}$ (in the part corresponding to x). Since $|T_c|/|T_a| \geq s_1/s_0$,

we have

$$\left\|\Omega_{T_a}\tilde{\eta}^{(t)}\right\|_2 \le \sqrt{\frac{s_0}{s_1}(1-|\dot{\eta}^*\tilde{\eta}^{(t)}|^2)}. \tag{4.28}$$

Also by the Cauchy-Schwarz inequality,

$$\begin{aligned}
&|\dot{\eta}^*\tilde{\eta}^{(t)}|^2 \le (\left\|\Omega_{T_a}\tilde{\eta}^{(t)}\right\|_2 \|\Omega_{T_a}\dot{\eta}\|_2 + \left\|\Omega_{T_b}\tilde{\eta}^{(t)}\right\|_2 \|\Omega_{T_b}\dot{\eta}\|_2)^2 \\
&\le \Big(\left\|\Omega_{T_a}\tilde{\eta}^{(t)}\right\|_2 \|\Omega_{T_a}\dot{\eta}\|_2 + \sqrt{1-\|\Omega_{T_a}\tilde{\eta}^{(t)}\|_2^2}\sqrt{1-\|\Omega_{T_a}\dot{\eta}\|_2^2}\ \Big)^2 \\
&\le 1-(\left\|\Omega_{T_a}\tilde{\eta}^{(t)}\right\|_2 - \|\Omega_{T_a}\dot{\eta}\|_2)^2.
\end{aligned}$$

It follows that

$$\|\Omega_{T_a}\dot{\eta}\|_2 \le \left\|\Omega_{T_a}\tilde{\eta}^{(t)}\right\|_2 + \sqrt{1-|\dot{\eta}^*\tilde{\eta}^{(t)}|^2}. \tag{4.29}$$

By (4.28) and (4.29),

$$\begin{aligned}
&|\dot{\eta}^*\tilde{\eta}^{(t)}| - |\dot{\eta}^*\widetilde{\Pi}_{s_1}(\tilde{\eta}^{(t)})| \le |\dot{\eta}^*\big(\tilde{\eta}^{(t)} - \widetilde{\Pi}_{s_1}(\tilde{\eta}^{(t)})\big)| \\
&= \left\|\Omega_{T_a}\tilde{\eta}^{(t)}\right\|_2 \|\Omega_{T_a}\dot{\eta}\|_2 \le \Big(\sqrt{\frac{s_0}{s_1}} + \frac{s_0}{s_1}\Big)(1-|\dot{\eta}^*\tilde{\eta}^{(t)}|^2).
\end{aligned} \tag{4.30}$$

By (4.27) and (4.30),

$$\begin{aligned}
&\sqrt{1-|\dot{\eta}^*\eta^{(t)}|} \le \sqrt{1-|\dot{\eta}^*\widetilde{\Pi}_{s_1}(\tilde{\eta}^{(t)})|} \\
&\le \sqrt{1-|\dot{\eta}^*\tilde{\eta}^{(t)}|}\sqrt{1+\Big(\sqrt{\frac{s_0}{s_1}}+\frac{s_0}{s_1}\Big)(1+|\dot{\eta}^*\tilde{\eta}^{(t)}|)} \\
&\le \sqrt{1-|\dot{\eta}^*\tilde{\eta}^{(t)}|}\sqrt{1+2\Big(\sqrt{\frac{s_0}{s_1}}+\frac{s_0}{s_1}\Big)} \\
&\le \rho\sqrt{1+2\sqrt{\frac{s_0}{s_1}}+\frac{2s_0}{s_1}}\sqrt{1-|\dot{\eta}^*\eta^{(t-1)}|} + \sqrt{10}\widetilde{\Delta} \\
&\le \tilde{\rho}\sqrt{1-|\dot{\eta}^*\eta^{(t-1)}|} + \sqrt{10}\widetilde{\Delta}.
\end{aligned}$$

Therefore, $\{|\dot{\eta}^*\eta^{(\tau)}|\}_{\tau=0}^t$ indeed monotonically increases unless $\sqrt{1-|\dot{\eta}^*\eta^{(\tau)}|}$ reaches $\sqrt{10}\widetilde{\Delta}/(1-\tilde{\rho})$ for some τ. The proof by induction is complete.

It follows that

$$\sqrt{1-|\dot{\eta}^*\eta^{(t)}|} \le \tilde{\rho}^t\sqrt{1-|\dot{\eta}^*\eta^{(0)}|} + \frac{\sqrt{10}\widetilde{\Delta}}{1-\tilde{\rho}},$$

or equivalently

$$d(\dot{\eta}, \eta^{(t)}) \leq \tilde{\rho}^t d(\dot{\eta}, \eta^{(0)}) + \frac{2\sqrt{5}\widetilde{\Delta}}{1-\tilde{\rho}}.$$

□

4.5.4 Proof of the Guarantee for Algorithm 3

Proof of Theorem 4.3.12. We first show that, under the conditions in Theorem 4.3.12, the support T_j in Algorithm 3 contains $T_j' \subset \mathrm{supp}(x_{\cdot j})$ in Assumption 4.3.5. To this end, we prove that the norms of the rows of D^*E indexed by T_j' are larger than those outside $\mathrm{supp}(x_{\cdot j})$. For a fixed $j \in [N]$, the j-th block of D^*E is indexed by the set $(j-1)m + [m]$. Therefore, the goal is to show that

$$\min_{\ell \in T_j'} \left\| d^*_{\cdot((j-1)m+\ell)} E \right\|_2^2 > \max_{\ell' \in [m] \setminus \mathrm{supp}(x_{\cdot j})} \left\| d^*_{\cdot((j-1)m+\ell')} E \right\|_2^2,$$

or equivalently,

$$\min_{\ell \in T_j'} \sum_{k \in [n]} |\overline{a_{k\ell}} y_{kj}|^2 > \max_{\ell' \in [m] \setminus \mathrm{supp}(x_{\cdot j})} \sum_{k \in [n]} |\overline{a_{k\ell'}} y_{kj}|^2.$$

Since

$$\mathbb{E}|\overline{a_{k\ell}} y_{kj}|^2 = \frac{1}{n^2} |\lambda_k|^2 (\|x_{\cdot j}\|_2^2 + |x_{\ell j}|^2) + \frac{1}{n} |w_{kj}|^2,$$

it suffices to show that for all $\ell \in T_j'$ and $\ell'' \in [m]$,

$$\frac{1}{n^2} \sum_{k \in [n]} |\lambda_k|^2 |x_{\ell j}|^2 > 2 \Big| \sum_{k \in [n]} \left(|\overline{a_{k\ell''}} y_{kj}|^2 - \mathbb{E}|\overline{a_{k\ell''}} y_{kj}|^2 \right) \Big|. \tag{4.31}$$

Recall that

$$y_{kj} = \lambda_k a_{k\cdot}^\top x_{\cdot j} + w_{kj}.$$

By the triangle inequality and Lemma 4.4.12, for all $j \in [N]$ and $\ell \in [m]$,

$$\begin{aligned}
&\Big|\sum_{k\in[n]}\Big(|\overline{a_{k\ell}}y_{kj}|^2 - \mathbb{E}|\overline{a_{k\ell}}y_{kj}|^2\Big)\Big| \\
&\le \Big|\sum_{k\in[n]}\Big(|\lambda_k\overline{a_{k\ell}}a_{k\cdot}^\top x_{\cdot j}|^2 - \mathbb{E}|\lambda_k\overline{a_{k\ell}}a_{k\cdot}^\top x_{\cdot j}|^2\Big)\Big| \\
&\quad + 2\Big|\sum_{k\in[n]}\mathrm{Re}\Big(\lambda_k a_{k\ell}\overline{a_{k\ell}}a_{k\cdot}^\top x_{\cdot j}\overline{w_{kj}}\Big)\Big| \\
&\quad + \Big|\sum_{k\in[n]}\Big(|\overline{a_{k\ell}}w_{kj}|^2 - \mathbb{E}|\overline{a_{k\ell}}w_{kj}|^2\Big)\Big| \\
&\le C_6\left(1+\frac{C_W}{\sqrt{1-\theta}}\right)^2\frac{\|x_{\cdot j}\|_2^2\log^3(nmN)}{n^{3/2}},
\end{aligned}$$

with probability at least $1 - n^{-c_6}$.

By Assumptions 4.3.2 and 4.3.5, if we plug the above result into (4.31), then the following sample complexity is sufficient for Algorithm 3 to correctly identify the subsets T_j' ($j \in [N]$) with probability at least $1 - n^{-c_6}$:

$$n^{1/2} > \frac{2C_6}{\omega(1-\delta)}\left(1+\frac{C_W}{\sqrt{1-\theta}}\right)^2 s_0\log^3(nmN).$$

Thus the first half of Theorem 4.3.12 is proved.

Given that the support T_j covers the large entries indexed by T_j',

$$\begin{aligned}
&\left\|\mathbb{E}\Pi_{T_x}D^*E - \frac{1}{n}x\lambda^\top\right\| = \left\|\frac{1}{n}\Pi_{T_x}x\lambda^\top - \frac{1}{n}x\lambda^\top\right\| \\
&\le \sqrt{\frac{1+\delta}{n}\sum_{j\in[N],\ell'\in[m]\setminus T_j'}|x_{\ell' j}|^2} \le \sqrt{\frac{(1+\delta)\delta_X}{n}}.
\end{aligned} \tag{4.32}$$

We also have

$$\begin{aligned}
&\|\Pi_{T_x}D^*E - \mathbb{E}\Pi_{T_x}D^*E\| \\
&\le \|\Omega_{T_x}D^*E_{\mathrm{s}} - \mathbb{E}\Omega_{T_x}D^*E_{\mathrm{s}}\| + \|\Omega_{T_x}D^*E_{\mathrm{n}}\| \\
&\le \frac{1}{\alpha}\Big(\Big\|\Omega_{T_\eta}B_{\mathrm{s}}\Omega_{T_\eta}^* - \Omega_{T_\eta}\mathbb{E}B_{\mathrm{s}}\Omega_{T_\eta}^*\Big\| + \Big\|\Omega_{T_\eta}B_{\mathrm{n}}\Omega_{T_\eta}^*\Big\|\Big) \\
&\le \frac{1}{\sqrt{n}}(\delta_B + \delta_W),
\end{aligned} \tag{4.33}$$

where the last inequality follows from Lemmas 4.4.2 and 4.4.7, given that the

conditions of Theorem 4.3.10 are satisfied. By the triangle inequality, and (4.32) and (4.33),

$$\left\| \Pi_{T_x} D^* E - \frac{1}{n} x \lambda^\top \right\| \leq \frac{1}{\sqrt{n}} (\delta_B + \delta_W + \sqrt{(1+\delta)\delta_X}),$$

where δ_B can be made arbitrarily small by a sufficiently large C in (4.13), δ_W can be made arbitrarily small by a sufficiently small C_W in Assumption 4.3.4, and the last term can be made arbitrarily small by a sufficiently small δ_X in Assumption 4.3.5. Therefore, the first left and right singular vectors u and v can become arbitrarily close to x and to $\lambda / \|\lambda\|_2$ (up to a global phase factor, i.e., a constant of unit modulus), respectively, and $|\dot{\eta}^* \eta^{(0)}|$ approaches

$$\frac{n^{3/2} + \|\lambda\|_2 \|\gamma\|_2^2}{\sqrt{n^2 + \|\lambda\|_2^2 \|\gamma\|_2^2} \sqrt{n + \|\gamma\|_2^2}} > 1 - 2\delta.$$

The inequality follows from Assumption 4.3.2, i.e., $\sqrt{1-\delta} \leq |\lambda_k| \leq \sqrt{1+\delta}$, and $1/\sqrt{1+\delta} \leq |\gamma_k| = 1/|\lambda_k| \leq 1/\sqrt{1-\delta}$. □

4.6 Numerical Experiments

In this section, we test the empirical performance of Algorithm 1 and Algorithm 2.

4.6.1 Subspace Case: Power Iteration vs. Least Squares

In Algorithm 1, we choose $\alpha = \sqrt{n}$, and $\beta = \|B\|$ (computed using another power iteration on B). We compare Algorithm 1 with the least squares approach in [40, Section 3.3], where $\gamma_1 = 1$ is used to avoid the trivial solution.

We generate $A \in \mathbb{C}^{n \times m}$ as a complex Gaussian random matrix, whose entries are drawn independently from $\mathcal{CN}(0, \frac{1}{n})$, i.e., the real and imaginary part are drawn independently from $\mathcal{N}(0, \frac{1}{2n})$. The unknown gains and phases λ_k are generated as follows:

$$\lambda_k = e^{\sqrt{-1}\varphi_k} \left(1 + (\sqrt{1+\delta} - 1) e^{\sqrt{-1}\varphi'_k}\right), \quad \forall k \in [n], \tag{4.34}$$

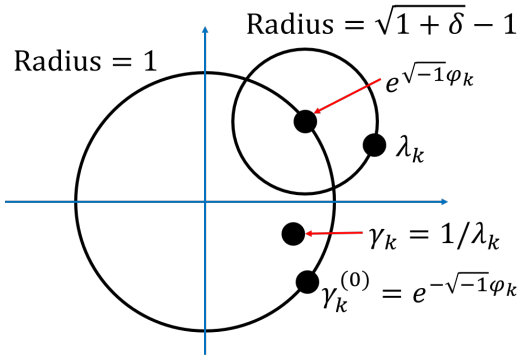


Figure 4.1: Illustration of λ_k in the complex plane.

such that λ_k is on a small circle of radius $\sqrt{1+\delta}-1$ centered at a point on the unit circle, and φ_k and φ_k' are drawn independently from a uniform distribution on $[0, 2\pi)$. Figure 4.1 visualizes one such synthesized λ_k in the complex plane. We set $\delta = 0.1$ in all the numerical experiments. The entries of $X \in \mathbb{C}^{m\times N}$ are drawn independently from $\mathcal{CN}(0, \frac{1}{Nm})$, so that the Frobenius norm of X is approximately 1. In the noisy setting, we generate complex white Gaussian noise $W \in \mathbb{C}^{n\times N}$, whose entries are drawn from $\mathcal{CN}(0, \frac{\sigma_W^2}{Nn})$. We define measurement signal-to-noise ratio (MSNR) and recovery signal-to-noise ratio (RSNR) as:

$$\text{MSNR} \coloneqq 20\log_{10}\frac{\|\text{diag}(\lambda)AX\|_{\text{F}}}{\|W\|_{\text{F}}},$$

$$\text{RSNR} \coloneqq -10\log_{10}(2-2|\dot{\eta}^*\eta^{(t)}|).$$

We test the two approaches at four noise levels: $\sigma_W = 0$, 0.1, 0.2, and 0.5, which roughly correspond to MSNR of ∞, 20 dB, 14 dB, and 6 dB. At these noise levels, we say the recovery is successful if the RSNR exceeds 30 dB, 20 dB, 14 dB, 6 dB, respectively. The success rates do not change dramatically as functions of these thresholds. In the experiments, we set $n = 128$, $N = 16$, and $m = 8, 16, 24, \ldots, 64$. For each m, we repeat the experiments 100 times and compute the empirical success rates, which are shown in Figure 4.2.

As seen in Figure 4.2(a), both power iteration and least squares achieve perfect recovery in the noiseless setting. However, as seen in Figures 4.2(b) – 4.2(d), power iteration is clearly more robust against noise than least squares, whose performance degrades more severely in the noisy settings.

The empirical phase transitions of power iteration are shown in Figure 4.3.

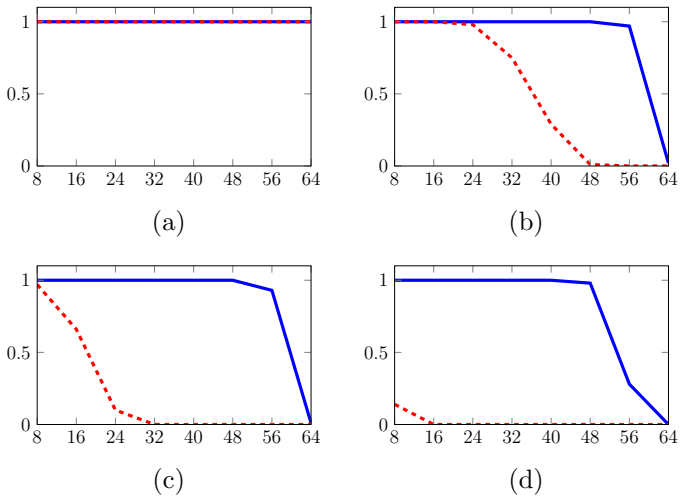


Figure 4.2: Subspace case: The empirical success rates of power iteration (blue solid line) and least squares (red dashed line). The x-axis represents m, and the y-axis represents the empirical success rate. (a) – (d) are the results with $\sigma_W = 0$, 0.1, 0.2, and 0.5, respectively.

We fix $N = 16$ and plot the phase transition with respect to n and m (Figure 4.3(a)); we then fix $n = 2m$ and plot the phase transition with respect to N and m (Figure 4.3(b)). Clearly, to achieve successful recovery, n must scale linearly with m, but N can be small compared to m and n. This confirms the sample complexity in Theorem 4.3.9, of $n \gtrsim m$ and $N \gtrsim 1$. Careful readers may notice in Figure 4.3(b) that for $N = 5$ the success rates at $m < 16$ are worse than those at $m \geq 16$. This seemingly peculiar phenomenon is caused by a small $n = 2m$, which does not belong to the large number regime associated with a high probability.

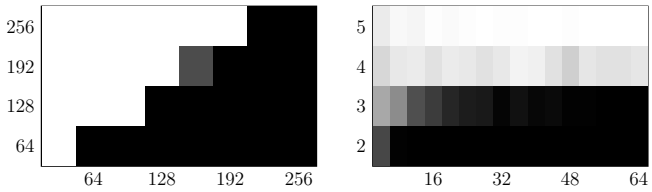


Figure 4.3: The empirical phase transition of power iteration. Grayscale represents success rates, where white equals 1, and black equals 0. (a) The x-axis represents m, and the y-axis represents n. (b) The x-axis represents m, and the y-axis represents N.

4.6.2 Sparsity Case: Truncated Power Iteration vs. ℓ_1 Minimization

In the sparsity case, we use the same setup described in the previous section, except for the signal X. The supports of the s_0-sparse columns of X are chosen uniformly at random, and the nonzero entries follow $\mathcal{CN}(0, \frac{1}{Ns_0})$. This unstructured sparsity case is more challenging than the joint sparsity case in Theorem 4.3.10.

In Algorithm 2, we choose $\alpha = \sqrt{n}$, and $\beta = \|B\|$. In all the experiments, we assume that the sparsity level s_0 is known, and set $s_1 = 2s_0$ for convenience. A more sophisticated scheme that decreases s_1 as the iteration number increases may lead to better empirical performance [44].

For the experiment we suppose that the phases $\{\varphi_k\}_{k=1}^n$ in (4.34) are available, and let

$$\gamma^{(0)} := [e^{-\sqrt{-1}\varphi_1}, \ldots, e^{-\sqrt{-1}\varphi_n}]^\top \tag{4.35}$$

denote the initial estimate of γ, which is close to but different from the true γ, i.e., the entrywise inverse of λ in (4.34). See Figure 4.1 for an illustration of λ_k, γ_k, and $\gamma_k^{(0)}$. Then we initialize Algorithm 2 with $\eta^{(0)} = [\mathbf{0}_{Nm,1}^\top, \gamma^{(0)\top}]^\top$.

We compare Algorithm 2 with an ℓ_1 minimization approach. Wang and Chi [41] adopted an approach tailored for the case where A is the DFT matrix and $\lambda_k \approx 1$. They use a linear constraint $\sum_{k\in[n]} \gamma_k = n$ to avoid the trivial solution of all zeros. For fair comparison, we revise their approach to accommodate arbitrary A and λ. The revised approach uses the alternating direction method of multipliers (ADMM) [119] to solve the following convex optimization problem:[2]

$$\begin{aligned} \min_{\gamma, X} \quad & \|\mathrm{vec}(X)\|_1 \\ \text{s.t.} \quad & \mathrm{diag}(\gamma) Y = AX, \\ & \gamma^{(0)*}\gamma = n. \end{aligned}$$

Here, $\gamma^{(0)}$ is the initial estimate of γ defined in (4.35), and used as initialization in our Algorithm 2 in this comparison.

[2]In the noisy setting, one could replace the linear constraint $\mathrm{diag}(\gamma)Y = AX$ with an ellipsoid constraint $\|\mathrm{diag}(\gamma)Y - AX\|_{\mathrm{F}} \le \epsilon$. However, the parameter ϵ needs to be adjusted with noise levels. For fair comparison of robustness to noise, we use the linear constrained ℓ_1 minimization in the noisy setting (similar to [41]).

We conduct numerical experiments with the same four noise levels and criterion for successful recovery as in Section 4.6.1. In the experiments, we set $n = 128$, $m = 256$, $N = 16$, and $s_0 = 8, 16, 24, \ldots, 64$. For each s_0, we repeat the experiments 100 times and compute the empirical success rates, which are shown in Figure 4.4. In the noiseless case (Figure 4.4(a)), ℓ_1 minimization achieves a slightly higher success rate near the phase transition. However, truncated power iteration is more robust against noise than ℓ_1 minimization, which breaks down completely at the higher noise levels (Figures. 4.4(b) – 4.4(d)).

Figure 4.4(a) clearly shows that truncated power iteration recovers η successfully when $n = 128$, $N = 16$, and $s_0 = 32$. This suggests that truncated power iteration may succeed when n and N are (up to log factors) on the order of s_0 and 1, respectively. However, while the scaling with the number of sensors n agrees with Theorem 4.3.10, success with such small number of snapshots N is not guaranteed by our current theoretical analysis.

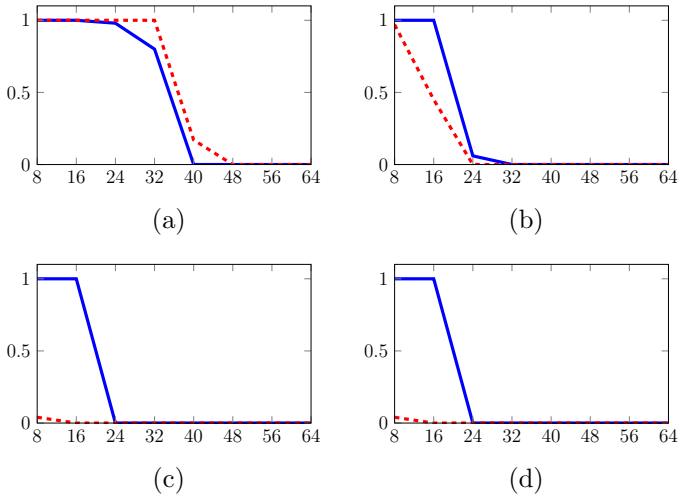
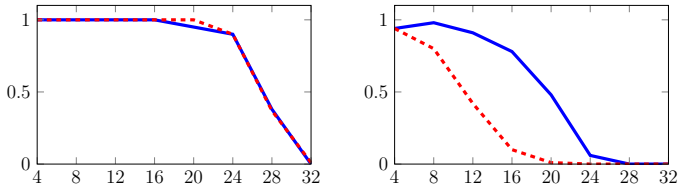


Figure 4.4: Sparsity case: The empirical success rates of truncated power iteration (blue solid line) and ℓ_1 minimization (red dashed line). The x-axis represents s_0, and the y-axis represents the empirical success rate. (a) – (d) are the results with $\sigma_W = 0$, 0.1, 0.2, and 0.5, respectively.

Next, we assume that only a subset of the phases $\{\varphi_k\}_{k=1}^n$ are available, and examine to what extent Algorithm 2 and ℓ_1 minimization depend on a good initial estimate of γ. In the numerical results shown in Figure 4.5, we consider only the noiseless setting of BGPC with sparsity, and set $s_0 = 4, 8, 12, \ldots, 32$. In Figures 4.5(a) and 4.5(b), we replace 1/2 and 3/4 of $\{\varphi_k\}_{k=1}^n$ with random

phases, respectively, and use the resulting bad estimate $\gamma^{(0)}$ in Algorithm 2 and ℓ_1 minimization. As seen in Figure 4.5, truncated power iteration is less dependent on accurate initial estimate of γ.



(a) (b)

Figure 4.5: Sparsity case: The empirical success rates of truncated power iteration (blue solid line) and ℓ_1 minimization (red dashed line), with bad initial estimate of the phases. The x-axis represents s_0, and the y-axis represents the empirical success rate. (a) and (b) are the results for which 1/2 and 3/4 of $\{\varphi_k\}_{k=1}^n$ are initialized with random phases.

We repeat the above experiments for the joint sparsity case, where we replace $\widetilde{\Pi}_{s_1}$ in Algorithm 2 with $\widetilde{\Pi}'_{s_1}$. We also replace the ℓ_1 norm $\|\mathrm{vec}(X)\|_1$ in the competing approach with a mixed norm:

$$\|X\|_{2,1} = \sum_{\ell\in[m]} \Big(\sum_{j\in[N]} |x_{\ell j}|^2 \Big)^{1/2},$$

which is a well-known convex method for the recovery of jointly sparse signals. The results for different noise levels and for inaccurate $\gamma^{(0)}$ are shown in Figures 4.6 and 4.7, respectively. In the joint sparsity case, truncated power iteration is robust against noise, but seems less robust against errors in the initial phase estimate. We conjecture that the failure of Algorithm 2 in the joint sparsity case is due to the restriction of $\widetilde{\Pi}'_{s_1}$. By projecting onto jointly sparse supports, the algorithm is likely to converge prematurely to an incorrect support. When compared to the results in Figures 4.7(a) and 4.7(b), Figures 4.7(c) and 4.7(d) show that using $\widetilde{\Pi}_{s_1}$ instead of $\widetilde{\Pi}'_{s_1}$ in the first half of the iterations indeed improves the performance of Algorithm 2 in the joint sparsity case. In the rest of the experiments, we use $\widetilde{\Pi}_{s_1}$ during the first half of the iterations in Algorithm 2 for the joint sparsity case.

Next, we plot the phase transitions for truncated power iteration. We fix $N = 16$ and $m = 2n$ and plot the empirical phase transition with respect to n and s_0 (sparsity case in Figure 4.8(a), and joint sparsity case in Figure 4.8(c)); we then fix $n = 4s_0$ and $m = 2n$ and plot the empirical phase transition with

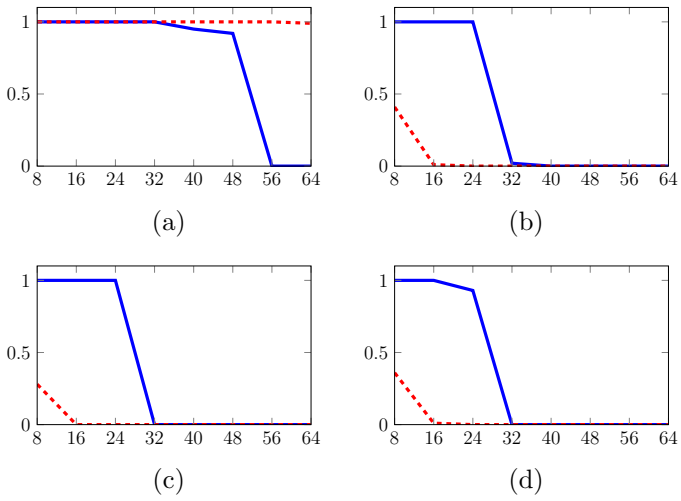


Figure 4.6: Joint sparsity case: The empirical success rates of truncated power iteration (blue solid line) and mixed minimization (red dashed line). The x-axis represents s_0, and the y-axis represents the empirical success rate. (a) – (d) are the results with $\sigma_W = 0$, 0.1, 0.2, and 0.5, respectively.

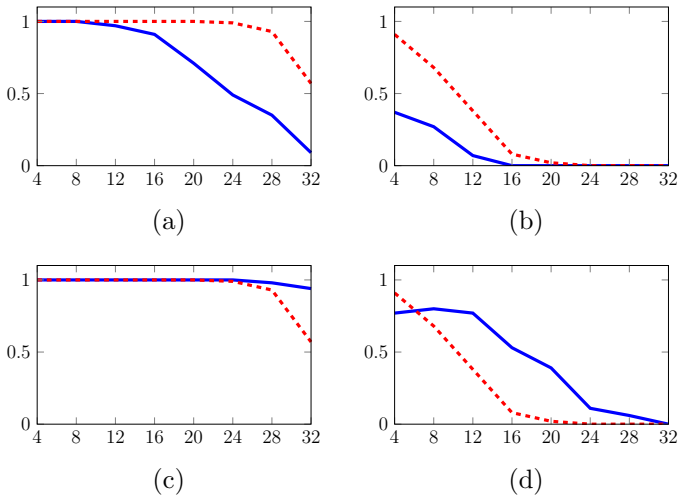


Figure 4.7: Joint sparsity case: The empirical success rates of truncated power iteration with $\widetilde{\Pi}'_{s_1}$ (blue solid line) and mixed minimization (red dashed line), with bad initial estimate of the phases. The x-axis represents s_0, and the y-axis represents the empirical success rate. (a) and (b) are the results for which 1/2 and 3/4 of $\{\varphi_k\}_{k=1}^n$ are initialized with random phases. In (c) and (d), we repeat the experiments, but use $\widetilde{\Pi}_{s_1}$ instead of $\widetilde{\Pi}'_{s_1}$ in the first half of the iterations.

respect to N and s_0 (sparsity case in Figure 4.8(b), and joint sparsity case in Figure 4.8(d)). It is seen that, to achieve successful recovery, n must scale linearly with s_0, but N can be small compared to s_0 and n. On the one hand, the scaling law $n \gtrsim s_0$ in Theorem 4.3.10 is confirmed by Figure 4.8; on the other hand, $N \gtrsim \sqrt{s_0}$ seems conservative and might be an artifact of our proof techniques. We have yet to come up with a theoretical guarantee that covers the more general sparsity case, or requires a less demanding sample complexity $N \gtrsim 1$. In Figures 4.8(b) and 4.8(d), the success rates at smaller s_0 are lower than those at a larger s_0, because the number of sensors $n = 4s_0$ is too small to yield a high probability.

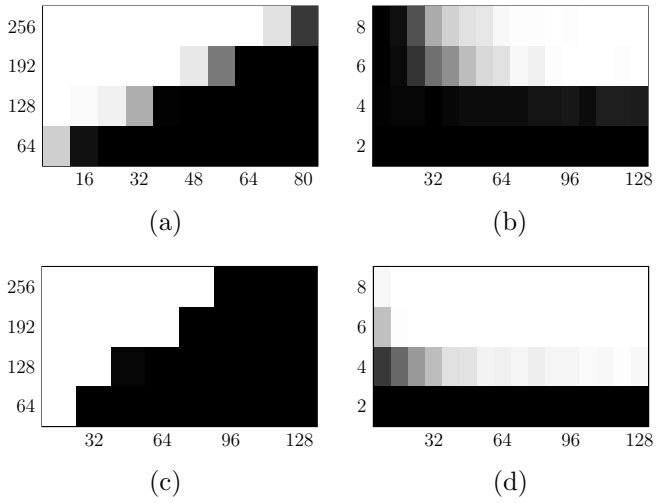


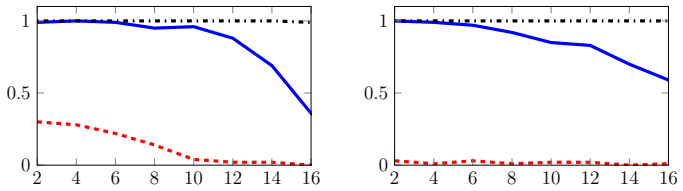
Figure 4.8: The empirical phase transition of truncated power iteration. Grayscale represents success rates, where white equals 1, and black equals 0. (a) Sparsity case: The x-axis represents s_0, and the y-axis represents n. (b) Sparsity case: The x-axis represents s_0, and the y-axis represents N. (c) Joint sparsity case: The x-axis represents s_0, and the y-axis represents n. (d) Joint sparsity case: The x-axis represents s_0, and the y-axis represents N.

4.6.3 Sparsity Case: Initialization

In this section, we examine the quality of the initialization produced by Algorithm 3 by comparing it with two different initializations: (i) the good initialization $\eta^{(0)} = [\mathbf{0}_{Nm,1}^\top, \gamma^{(0)\top}]^\top$ aided by side information on the phase in Section 4.6.2; and (ii) a baseline initialization $\eta^{(0)} = [\mathbf{0}_{Nm,1}^\top, \mathbf{1}_{n,1}^\top]^\top$. We

use the same setting as in Section 4.6.2, except that $N = 32$. We let $\sigma_W = 0.1$, and claim the recovery is successful if the RSNR exceeds 20 dB. In the experiment for the joint sparsity case, for the reason mentioned in Section 4.6.2, we ignore the joint sparsity structure and estimate the support of different columns of X independently in the initialization and during the first half of the iterations. Only in the second half of the iterations, we use the projection $\widetilde{\Pi}'_{s_1}$ onto jointly sparse supports.

Figure 4.9 shows that, although the initialization provided by Algorithm 3 is not as good as the accurate initialization with side information, it is far better than the baseline. Figure 4.10 shows the empirical phase transition with respect to n and s_0, when Algorithm 3 is used to initialize truncated power iteration (sparsity case in Figure 4.10(a), and joint sparsity case in Figure 4.10(b)). The results suggest that when n scales linearly with s_0, Algorithm 3 can provide a sufficiently good initialization for truncated power iteration. For example, in 4.10(a), the success rate is 1 when $n = 256$ and $s_0 = 20$. Therefore, the sample complexity $n \gtrsim s_0^2$ in Theorem 4.3.12 could be overly conservative and an artifact of our analysis.



(a) (b)

Figure 4.9: The empirical success rates of truncated power iteration with the initialization in Algorithm 3 (blue solid line), with a baseline initialization $\eta^{(0)} = [\mathbf{0}_{Nm,1}^\top, \mathbf{1}_{n,1}^\top]^\top$ (red dashed line), and with the accurate initialization $\eta^{(0)} = [\mathbf{0}_{Nm,1}^\top, \gamma^{(0)\top}]^\top$ with side information in Section 4.6.2 (black dash-dot line). The x-axis represents s_0, and the y-axis represents the empirical success rate. (a) is the result for the sparsity case, and (b) is the result for the joint sparsity case.

4.6.4 Application: Inverse Rendering

In this section, we apply the power iteration algorithm to the inverse rendering problem in computational relighting – given images of an object under different lighting conditions (Figure 4.11(a)), and the surface normals of the

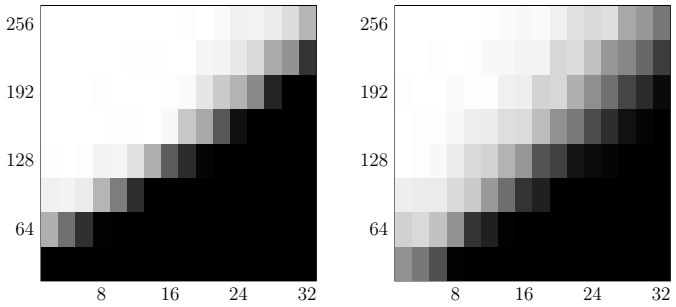


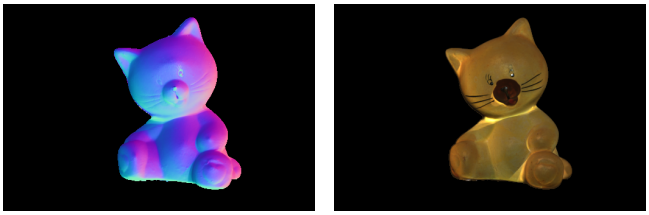
Figure 4.10: The empirical phase transition of truncated power iteration with the initialization in Algorithm 3. The x-axis represents s_0, and the y-axis represents n. (a) is the result for the sparsity case, and (b) is the result for the joint sparsity case.

object (Figure 4.11(b)), the goal is to recover the albedos (also known as reflection coefficients) of the object surface and the lighting conditions. In this problem, the columns of $Y = \text{diag}(\lambda)AX \in \mathbb{R}^{n\times N}$ represent images under different lighting conditions, which are the products of the unknown albedo map $\lambda \in \mathbb{R}^n$ and the intensity maps of incident light under different conditions AX. For Lambertian surfaces, it is reasonable to assume that the intensity of incident light resides in a subspace spanned by the first nine spherical harmonics computed from the surface normals [29], which we denote by the columns of $A \in \mathbb{R}^{n\times 9}$. Then the columns of X are the coordinates of the spherical harmonic expansion, which parameterize the lighting conditions. We can solve for λ and X using Algorithm 1. Our approach is similar to that of Nguyen et al. [29], which also formulates inverse rendering as an eigenvector problem. Despite the fact that the two approaches solve for the eigenvectors of different matrices, they yield identical solutions in the ideal scenario where the model is exact and the solution is unique.

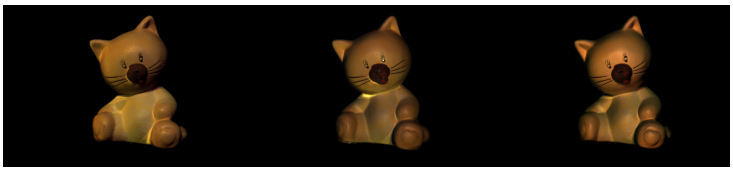
In our experiment, we obtain $N = 12$ color images and the surface normals of an object under different lighting conditions,[3] and we compute the first $m = 9$ spherical harmonics. We apply Algorithm 1 to each of the three color channels, and the albedo map recovered using 200 power iterations is shown in Figure 4.11(c). We also compute new images of the object under new lighting conditions (Figure 4.11(d)).

[3]The images are downloaded from *https://courses.cs.washington.edu/courses/csep576/05wi/projects/project3/project3.htm* on September 16, 2017. The surface normals are computed using the method described in the same webpage.

(a)



(b) (c)



(d)

Figure 4.11: Inverse rendering and relighting. (a) We use 12 images of the object under different lighting conditions. (b) The surface normals. The three dimensions of the normal vectors are represented by the RGB channels of the color image. (c) The recovered albedo map. (d) Computed images of the object under new lighting conditions.

CHAPTER 5

MULTICHANNEL SPARSE BLIND DECONVOLUTION VIA MANIFOLD GRADIENT DESCENT

5.1 MSBD on the Sphere

5.1.1 Notations

We use $[n]$ as a shorthand for the index set $\{1, 2, \ldots, n\}$. We use $x_{(j)}$ to denote the j-th entry of $x \in \mathbb{R}^n$, and $H_{(jk)}$ to denote the entry of $H \in \mathbb{R}^{n\times n}$ in the j-th row and k-th column. The superscript in $h^{(t)}$ denotes iteration number in an iterative algorithm. Throughout the chapter, if an index $j \notin [n]$, then the actual index is computed as modulo of n. The circulant matrix whose first column is x is denoted by C_x. We use δ_{jk} to denote the Kronecker delta ($\delta_{jk} = 0$ if $j \neq k$ and $\delta_{jk} = 1$ if $j = k$). The entrywise product between vectors x and y is denoted by $x \odot y$, and the entrywise k-th power of a vector x is denoted by $x^{\odot k}$. We use $\|\cdot\|$ to denote the ℓ_2 norm (for a vector), or the spectral norm (for a matrix). We use $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ to denote the real and imaginary parts of a complex vector or matrix.

5.1.2 Problem Statement

In MSBD, the measurements $y_1, y_2, \ldots, y_N \in \mathbb{R}^n$ are the circular convolutions of unknown sparse vectors $x_1, x_2, \ldots, x_N \in \mathbb{R}^n$ and an unknown vector $f \in \mathbb{R}^n$, i.e., $y_i = x_i \circledast f$. In this chapter, we solve for $\{x_i\}_{i=1}^n$ and f from $\{y_i\}_{i=1}^N$. One can rewrite the measurement as $Y = C_f X$, where $Y = [y_1, y_2, \ldots, y_N]$ and $X = [x_1, x_2, \ldots, x_N]$ are $n \times N$ matrices. Without structures, one can solve the problem by choosing any invertible circulant matrix C_f and compute $X = C_f^{-1} Y$. The fact that X is sparse narrows down the search space.

Even with sparsity, the problem suffers from inherent scale and shift ambiguities. Suppose $\mathcal{S}_j : \mathbb{R}^n \to \mathbb{R}^n$ denotes a circular shift by j positions,

i.e., $\mathcal{S}_j(x)_{(k)} = x_{(k-j)}$ for $j, k \in [n]$. Note that we have $y_i = x_i \circledast f = (\alpha \mathcal{S}_j(x_i)) \circledast (\alpha^{-1} \mathcal{S}_{-j}(f))$ for every nonzero $\alpha \in \mathbb{R}$ and $j \in [n]$. Therefore, MSBD has equivalent solutions generated by scaling and circularly shifting $\{x_i\}_{i=1}^n$ and f.

Throughout this chapter, we assume that the circular convolution with the signal f is invertible, i.e., there exists a filter g such that $f \circledast g = e_1$ (the first standard basis vector). Equivalently, C_f is an invertible matrix, and the DFT of f is nonzero everywhere. Since $y_i \circledast g = x_i \circledast f \circledast g = x_i$, one can find g by solving the following optimization problem:

$$\text{(P0)} \quad \min_{h \in \mathbb{R}^n} \quad \frac{1}{N} \sum_{i=1}^{N} \|C_{y_i} h\|_0 , \quad \text{s.t.} \quad h \neq 0.$$

The constraint eliminates the trivial solution that is 0. If the solution to MSBD is unique up to the aforementioned ambiguities, then the only minimizers of (P0) are $h = \alpha \mathcal{S}_j g$ ($\alpha \neq 0$, $j \in [n]$).

5.1.3 Smooth Formulation

Minimizing the non-smooth ℓ_0 "norm" is usually challenging. Instead, one can choose a smooth surrogate function for sparsity, which can be minimized using first-order or second-order optimization methods.

Here, we make two observations: (1) one can eliminate scaling ambiguity by restricting h to the unit sphere S^{n-1}; (2) sparse recovery can be achieved by maximizing the "spikiness" $\|\cdot\|_4^4$ [120]. Based on these observations, we adopt the following optimization problem:

$$\text{(P1)} \quad \min_{h \in \mathbb{R}^n} \quad -\frac{1}{4N} \sum_{i=1}^{N} \|C_{y_i} R h\|_4^4 , \quad \text{s.t.} \quad \|h\| = 1.$$

The matrix $R \coloneqq (\frac{1}{\theta n N} \sum_{i=1}^{N} C_{y_i}^\top C_{y_i})^{-1/2} \in \mathbb{R}^{n \times n}$ is a preconditioner, where θ is a parameter that is proportional to the sparsity level of $\{x_i\}_{i=1}^N$. In Section 5.2, under specific probabilistic assumptions on $\{x_i\}_{i=1}^N$, we explain how the preconditioner R works.

Problem (P1) can be solved using first-order or second-order optimization methods over Riemannian manifolds. The main result of this chapter pro-

vides a geometric view of the objective function over the sphere S^{n-1} (see Figure 5.2). We show that some off-the-shelf optimization methods can be used to obtain a solution $\hat{h}$ close to a scaled and circularly shifted version of the ground truth. Specifically, $\hat{h}$ satisfies $C_f R\hat{h} \approx \pm e_j$ for some $j \in [n]$, i.e., $R\hat{h}$ is approximately a signed and shifted version of the inverse of f. Given solution $\hat{h}$ to (P1), one can recover f and x_i $(i = 1, 2, \ldots, N)$ as follows:[1]

$$\hat{f} = \mathcal{F}^{-1}\big[\mathcal{F}(R\hat{h})^{\odot -1}\big], \tag{5.1}$$

$$\hat{x}_i = C_{y_i} R\hat{h}. \tag{5.2}$$

5.2 Global Geometric View

5.2.1 Main Result

In this chapter, we assume that $\{x_i\}_{i=1}^N$ are random sparse vectors, and f is invertible:

(A1) The channels $\{x_i\}_{i=1}^N$ follow a Bernoulli-Rademacher model. More precisely, $x_{i(j)} = A_{ij}B_{ij}$, where $\{A_{ij}, B_{ij}\}_{i\in[N],j\in[n]}$ are independent random variables, B_{ij}'s follow a Bernoulli distribution $\mathrm{Ber}(\theta)$, and A_{ij}'s follow a Rademacher distribution (taking values 1 and -1, each with probability 1/2).

(A2) The circular convolution with the signal f is invertible. We use κ to denote the condition number of f, which is defined as $\kappa := \frac{\max_j |(\mathcal{F}f)_{(j)}|}{\min_k |(\mathcal{F}f)_{(k)}|}$, i.e., the ratio of the largest and smallest magnitudes of the DFT. This is also the condition number of the circulant matrix C_f, i.e. $\kappa = \frac{\sigma_1(C_f)}{\sigma_n(C_f)}$.

The Bernoulli-Rademacher model is a special case of the Bernoulli–sub-Gaussian models. The derivation in this chapter can be repeated for other sub-Gaussian nonzero entries, with different tail bounds. We use the Rademacher distribution for simplicity.

[1] An alternative way to recover a sparse vector x_i given the recovered $\hat{f}$ and the measurement y_i, is to solve the non-blind deconvolution problem. For example, one can solve the sparse recovery problem $\min_x \frac{1}{2}\left\|C_{\hat{f}}x - y_i\right\|^2 + \lambda\|x\|_1$ using FISTA [121]. We omit the analysis of such a solution in this chapter, and focus on the simple reconstruction $\hat{x}_i = C_{y_i}R\hat{h}$.

Let $\phi(x) = -\frac{1}{4}\|x\|_4^4$. Its gradient and Hessian are defined by $\nabla_\phi(x)_{(j)} = -x_j^3$, and $H_\phi(x)_{(jk)} = -3x_j^2\delta_{jk}$. Then the objective function in (P1) is

$$L(h) = \frac{1}{N}\sum_{i=1}^{N}\phi(C_{y_i}Rh),$$

where $R = (\frac{1}{\theta nN}\sum_{i=1}^{N} C_{y_i}^\top C_{y_i})^{-1/2}$. The gradient and Hessian are

$$\nabla_L(h) = \frac{1}{N}\sum_{i=1}^{N} R^\top C_{y_i}^\top \nabla_\phi(C_{y_i}Rh),$$
$$H_L(h) = \frac{1}{N}\sum_{i=1}^{N} R^\top C_{y_i}^\top H_\phi(C_{y_i}Rh)C_{y_i}R.$$

Since $L(h)$ is to be minimized over S^{n-1}, we use optimization methods over Riemannian manifolds [122]. To this end, we define the tangent space at $h \in S^{n-1}$ as $\{z \in \mathbb{R}^n : z \perp h\}$ (see Figure 5.1). We study the Riemannian gradient and Riemannian Hessian of $L(h)$ (gradient and Hessian along the tangent space at $h \in S^{n-1}$):

$$\widehat{\nabla}_L(h) = P_{h^\perp}\nabla_L(h),$$
$$\widehat{H}_L(h) = P_{h^\perp}H_L(h)P_{h^\perp} - \langle\nabla_L(h), h\rangle P_{h^\perp},$$

where $P_{h^\perp} = I - hh^\top$ is the projection onto the tangent space at h. We refer the readers to [122] for a more comprehensive discussion of these concepts.

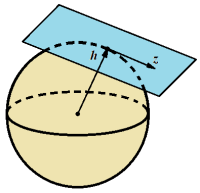
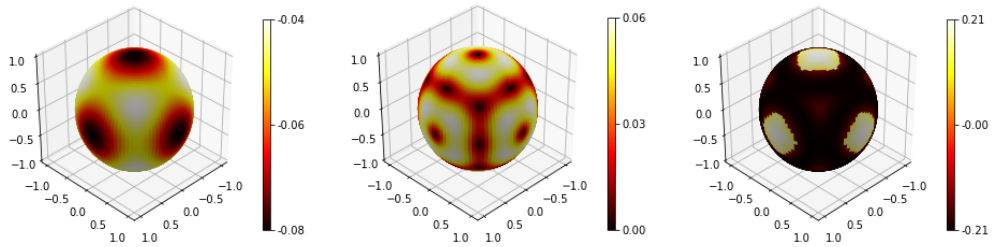


Figure 5.1: A demonstration of the tangent space of S^{n-1} at h, the origin of which is translated to h. The Riemannian gradient and Riemannian Hessian are defined on tangent spaces.

The toy example in Figure 5.2 demonstrates the geometric structure of the objective function on S^{n-1}. (As shown later, the quantity $\mathbb{E}L''(h)$ is, up to an unimportant rotation of the coordinate system, a good approximation to $L(h)$.) The local minima correspond to signed shifted versions of the ground



(a) (b) (c)

Figure 5.2: Geometric structure of the objective function over the sphere. For $n = 3$, we plot the following quantities on the sphere S^2: (a) $\mathbb{E}L''(h)$, (b) $\left\|\mathbb{E}\widehat{\nabla}_{L''}(h)\right\|$, and (c) $\min_{z\perp h,\|z\|=1} z^\top \mathbb{E}\widehat{H}_{L''}(h)z$.

truth (Figure 5.2(a)). The Riemannian gradient is zero at stationary points, including local minima, saddle points, and local maxima of the objective function when restricted to the sphere S^{n-1} (Figure 5.2(b)). The Riemannian Hessian is positive definite in the neighborhoods of local minima, and has at least one strictly negative eigenvalue in the neighborhoods of local maxima and saddle points (Figure 5.2(c)). We say that a stationary point is a "strict saddle point" if the Riemannian Hessian has at least one strictly negative eigenvalue. The Riemannian Hessian is negative definite in the neighborhood of a local maximum. Hence, local maxima are strict saddle points. Our main result Theorem 5.2.1 formalizes the observation that $L(h)$ only has two types of stationary points: (1) local minima, which are close to signed shifted versions of the ground truth, and (2) strict saddle points.

Theorem 5.2.1. *Suppose Assumptions (A1) and (A2) are satisfied, and the Bernoulli probability satisfies $\frac{1}{n} \le \theta < \frac{1}{3}$. Let κ be the condition number of f, let $\rho < 10^{-3}$ be a small tolerance constant. There exist constants $c_1, c_1' > 0$ (depending only on θ), such that: if $N > \max\{\frac{c_1 n^9}{\rho^4}\log\frac{n}{\rho}, \frac{c_1\kappa^8 n^8}{\rho^4}\log n\}$, then with probability at least $1 - n^{-c_1'}$, every local minimum h^* in (P1) is close to a signed shifted version of the ground truth. I.e., for some $j \in [n]$:*

$$\|C_f R h^* \pm e_j\| \le 2\sqrt{\rho}.$$

Moreover, one can partition S^{n-1} into three sets $\mathcal{H}_1$, $\mathcal{H}_2$, and $\mathcal{H}_3$ that satisfy (for some $c(n,\theta,\rho) > 0$):

- *$L(h)$ is strongly convex in $\mathcal{H}_1$, i.e., the Riemannian Hessian is positive*

definite:

$$\min_{\substack{z:\|z\|=1 \\ z\perp h}} z^\top \widehat{H}_L(h) z \geq c(n,\theta,\rho) > 0.$$

- *$L(h)$ has negative curvature in $\mathcal{H}_2$, i.e., the Riemannian Hessian has a strictly negative eigenvalue:*

$$\min_{\substack{z:\|z\|=1 \\ z\perp h}} z^\top \widehat{H}_L(h) z \leq -c(n,\theta,\rho) < 0.$$

- *$L(h)$ has a descent direction in $\mathcal{H}_3$, i.e., the Riemannian gradient is nonzero:*

$$\left\| \widehat{\nabla}_L(h) \right\| \geq c(n,\theta,\rho) > 0.$$

Clearly, all the stationary points of $L(h)$ on S^{n-1} belong to $\mathcal{H}_1$ or $\mathcal{H}_2$. The stationary points in $\mathcal{H}_1$ are local minima, and the stationary points in $\mathcal{H}_2$ are strict saddle points. The sets $\mathcal{H}_1$, $\mathcal{H}_2$, $\mathcal{H}_3$ are defined in (5.12)*, and the positive number $c(n,\theta,\rho)$ is defined in* (5.13)*.*

We only consider the noiseless case in Theorem 5.2.1. One can extend our analysis to noisy measurements by bounding the perturbation of the objective function caused by noise. In Section 5.5, we verify by numerical experiments that the formulation in this chapter is robust against noise.

5.2.2 Proof of the Main Result

Note that $R = (\frac{1}{\theta n N} \sum_{i=1}^N C_{y_i}^\top C_{y_i})^{-1/2}$ asymptotically converges to $(C_f^\top C_f)^{-1/2}$ as N increases. Therefore, $L(h)$ can be approximated by

$$L'(h) = \frac{1}{N} \sum_{i=1}^N \phi(C_{y_i} (C_f^\top C_f)^{-1/2} h) = \frac{1}{N} \sum_{i=1}^N \phi(C_{x_i} C_f (C_f^\top C_f)^{-1/2} h).$$

Since $C_f (C_f^\top C_f)^{-1/2}$ is an orthogonal matrix, one can study the following objective function by rotating on the sphere $h' = C_f (C_f^\top C_f)^{-1/2} h$:

$$L''(h') = \frac{1}{N} \sum_{i=1}^N \phi(C_{x_i} h').$$

Our analysis consists of three parts: (1) geometric structure of $\mathbb{E}L''$, (2) deviation of L'' (or its rotated version L') from its expectation $\mathbb{E}L''$, and (3) difference between L and L'.

Geometric structure of $\mathbb{E}L''$. By the Bernoulli-Rademacher model (A1), the Riemannian gradient for $h \in S^{n-1}$ is computed as

$$\mathbb{E}\widehat{\nabla}_{L''}(h) = P_{h^\perp}\mathbb{E}\nabla_{L''}(h) = n\theta(1-3\theta)(\|h\|_4^4 \cdot h - h^{\odot 3}). \tag{5.3}$$

The Riemannian Hessian is

$$\begin{aligned}\mathbb{E}\widehat{H}_{L''}(h) &= P_{h^\perp}\mathbb{E}H_{L''}(h)P_{h^\perp} - h^\top \mathbb{E}\nabla_{L''}(h)\cdot P_{h^\perp} \\ &= n\theta(1-3\theta)\big[\|h\|_4^4 \cdot I + 2\|h\|_4^4 \cdot hh^\top - 3\cdot \mathrm{diag}(h^{\odot 2})\big].\end{aligned} \tag{5.4}$$

Details of the derivation of (5.3) and (5.4) can be found in Appendix C.1.1.

At a stationary point of $\mathbb{E}L''(h)$ on S^{n-1}, the Riemannian gradient is zero. Since

$$\begin{aligned}\left\|\mathbb{E}\widehat{\nabla}_{L''}(h)\right\| &= n\theta(1-3\theta)\sqrt{\|h\|_6^6 - \|h\|_4^8} \\ &= n\theta(1-3\theta)\sqrt{\sum_{1\le j<k\le n} h_{(j)}^2 h_{(k)}^2 (h_{(j)}^2 - h_{(k)}^2)^2},\end{aligned} \tag{5.5}$$

all nonzero entries of a stationary point h_0 have the same absolute value. Equivalently, $h_{0(j)} = \pm 1/\sqrt{r}$ if $j \in \Omega$ and $h_{0(j)} = 0$ if $j \notin \Omega$, for some $r \in [n]$ and $\Omega \subset [n]$ such that $|\Omega| = r$. Without loss of generality (as justified below), we focus on stationary points that satisfy $h_{0(j)} = 1/\sqrt{r}$ if $j \in \{1, 2, \ldots, r\}$ and $h_{0(j)} = 0$ if $j \in \{r+1, \ldots, n\}$. The Riemannian Hessian at these stationary points is

$$\mathbb{E}\widehat{H}_{L''}(h_0) = \frac{n\theta(1-3\theta)}{r}\begin{bmatrix} \frac{2}{r}\mathbf{1}_{r\times r} - 2I_r & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(n-r)\times r} & I_{n-r}\end{bmatrix}. \tag{5.6}$$

When $r = 1$, $h_0 = [1, 0, 0, \ldots, 0]^\top$, we have $\mathbb{E}\widehat{H}_{L''}(h_0) = n\theta(1-3\theta)P_{h_0^\perp}$. This Riemannian Hessian is positive definite on the tangent space,

$$\min_{\substack{z:\|z\|=1 \\ z\perp h_0}} z^\top \mathbb{E}\widehat{H}_{L''}(h_0)z = n\theta(1-3\theta) > 0. \tag{5.7}$$

Therefore, stationary points with one nonzero entry are local minima.

When $r > 1$, the Riemannian Hessian has at least one strictly negative eigenvalue:

$$\min_{\substack{z:\|z\|=1 \\ z\perp h_0}} z^\top \mathbb{E}\widehat{H}_{L''}(h_0)z = -\frac{2n\theta(1-3\theta)}{r} < 0. \tag{5.8}$$

Therefore, stationary points with more than one nonzero entry are strict saddle points, which, by definition, have at least one negative curvature direction on S^{n-1}. One such negative curvature direction satisfies $z_{(1)} = (r-1)/\sqrt{r(r-1)}$, $z_{(j)} = -1/\sqrt{r(r-1)}$ for $j \in \{2, 3, \ldots, r\}$, and $z_{(j)} = 0$ for $j \in \{r+1, \ldots, n\}$.

The Riemannian Hessian at other stationary points (different from the above stationary points by permutations and sign changes) can be computed similarly. By (5.4), a permutation and sign changes of the entries in h_0 has no effect on the bounds in (5.7) and (5.8), because the eigenvector z that attains the minimum undergoes the same permutation and sign changes as h_0.

Next, in Lemma 5.2.3, we show that the properties of positive definiteness and negative curvature not only hold at the stationary points, but also hold in their neighborhoods defined as follows.

Definition 5.2.2. *We say that a point h is in the (ρ, r)-neighborhood of a stationary point h_0 of $\mathbb{E}L''(h)$ with r nonzero entries, if $\left\|h^{\odot 2} - h_0^{\odot 2}\right\|_\infty \le \frac{\rho}{r}$. We define three sets:*

$$\begin{aligned}
\mathcal{H}_1'' &\coloneqq \{\textit{Points in the } (\rho,1)\textit{-neighborhoods of stationary points} \\
&\qquad \textit{with } 1 \textit{ nonzero entry}\}, \\
\mathcal{H}_2'' &\coloneqq \{\textit{Points in the } (\rho,r)\textit{-neighborhoods of stationary points} \\
&\qquad \textit{with } r > 1 \textit{ nonzero entries}\}, \\
\mathcal{H}_3'' &\coloneqq S^{n-1} \backslash (\mathcal{H}_1'' \cup \mathcal{H}_2'').
\end{aligned}$$

Clearly, $\mathcal{H}_1 \cap \mathcal{H}_2 = \emptyset$ for $\rho < 1/3$, hence $\mathcal{H}_1''$, $\mathcal{H}_2''$, and $\mathcal{H}_3''$ form a partition of S^{n-1}.

Lemma 5.2.3. *Assume that positive constants $\theta < 1/3$, and $\rho < 10^{-3}$. Then*

- *For $h \in \mathcal{H}_1''$,*

$$\min_{\substack{z:\|z\|=1 \\ z\perp h}} z^\top \mathbb{E}\widehat{H}_{L''}(h)z \geq n\theta(1-3\theta)(1-24\sqrt{\rho}) > 0. \tag{5.9}$$

- *For $h \in \mathcal{H}_2''$,*

$$\min_{\substack{z:\|z\|=1 \\ z\perp h}} z^\top \mathbb{E}\widehat{H}_{L''}(h)z \leq -\frac{n\theta(1-3\theta)(2-24\sqrt{\rho})}{r} < 0. \tag{5.10}$$

- *For $h \in \mathcal{H}_3''$,*

$$\left\|\mathbb{E}\widehat{\nabla}_{L''}(h)\right\| \geq \frac{\theta(1-3\theta)\rho^2}{n} > 0. \tag{5.11}$$

Lemma 5.2.3, and all other lemmas, are proved in Appendix C.

Deviation of L'' from $\mathbb{E}L''$. As the number N of channels increases, the objective function L'' asymptotically converges to its expected value $\mathbb{E}L''$. Therefore, we can establish the geometric structure of L'' based on its similarity to $\mathbb{E}L''$. To this end, we give the following result.

Lemma 5.2.4. *Suppose that $\theta < 1/3$. There exist constants $c_2, c_2' > 0$ (depending only on θ), such that: if $N > \frac{c_2 n^9}{\rho^4}\log\frac{n}{\rho}$, then with probability at least $1 - e^{-c_2' n}$,*

$$\sup_{h\in S^{n-1}} \left\|\widehat{\nabla}_{L''}(h) - \mathbb{E}\widehat{\nabla}_{L''}(h)\right\| \leq \frac{\theta(1-3\theta)\rho^2}{4n},$$

$$\sup_{h\in S^{n-1}} \left\|\widehat{H}_{L''}(h) - \mathbb{E}\widehat{H}_{L''}(h)\right\| \leq \frac{\theta(1-3\theta)\rho^2}{n}.$$

By Lemma 5.2.4, the deviations from the corresponding expected values of the Riemannian gradient and Hessian due to a finite number of random x_i's are small compared to the bounds in Lemma 5.2.3. Therefore, the Rimannian Hessian of L'' is still positive definite in the neighborhood of local minima, and has at least one strictly negative eigenvalue in the neighborhood of strict saddle points; and the Riemannian gradient of L'' is nonzero for all other points on the sphere. Since L' and L'' differ only by an orthogonal matrix

transformation of their argument, the geometric structure of L' is identical to that of L'' up to a rotation on the sphere.

Difference between L and L'. Recall that L asymptotically converges to L' as N increases. The following result bounds the difference for a finite N.

Lemma 5.2.5. *Suppose that $\frac{1}{n} \le \theta < \frac{1}{3}$. There exist constants $c_3, c_3' > 0$ (depending only on θ), such that: if $N > \frac{c_3 \kappa^8 n^8}{\rho^4} \log n$, then with probability at least $1 - n^{-c_3'}$,*

$$\sup_{h \in S^{n-1}} \left\| \widehat{\nabla}_L(h) - \widehat{\nabla}_{L'}(h) \right\| \le \frac{\theta(1-3\theta)\rho^2}{4n},$$

$$\sup_{h \in S^{n-1}} \left\| \widehat{H}_L(h) - \widehat{H}_{L'}(h) \right\| \le \frac{\theta(1-3\theta)\rho^2}{n}.$$

We use $(C_f^\top C_f)^{1/2} C_f^{-1} \mathcal{H} = \{(C_f^\top C_f)^{1/2} C_f^{-1} h : h \in \mathcal{H}\}$ to denote the rotation of a set $\mathcal{H}$ by the orthogonal matrix $(C_f^\top C_f)^{1/2} C_f^{-1}$. Define the rotations of $\mathcal{H}_1''$, $\mathcal{H}_2''$, and $\mathcal{H}_3''$:

$$\begin{aligned} \mathcal{H}_1 &\coloneqq (C_f^\top C_f)^{1/2} C_f^{-1} \mathcal{H}_1'', \\ \mathcal{H}_2 &\coloneqq (C_f^\top C_f)^{1/2} C_f^{-1} \mathcal{H}_2'', \\ \mathcal{H}_3 &\coloneqq (C_f^\top C_f)^{1/2} C_f^{-1} \mathcal{H}_3''. \end{aligned} \tag{5.12}$$

Combining Lemmas 5.2.3, 5.2.4, and 5.2.5, and the rotation relation between L' and L'', we have:

- For $h \in \mathcal{H}_1$, the Riemannian Hessian is positive definite:

$$\min_{\substack{z : \|z\| = 1 \\ z \perp h}} z^\top \widehat{H}_L(h) z \ge n\theta(1-3\theta)(1 - 24\sqrt{\rho} - \frac{2\rho^2}{n^2}) > 0.$$

- For $h \in \mathcal{H}_2$, the Riemannian Hessian has a strictly negative eigenvalue:

$$\min_{\substack{z : \|z\| = 1 \\ z \perp h}} z^\top \widehat{H}_L(h) z \le -\frac{n\theta(1-3\theta)(2 - 24\sqrt{\rho} - 2r\rho^2/n^2)}{r} < 0.$$

- For $h \in \mathcal{H}_3$, the Riemannian gradient is nonzero:

$$\left\|\widehat{\nabla}_L(h)\right\| \geq \frac{\theta(1-3\theta)\rho^2}{2n} > 0.$$

Clearly, all the local minima of $L(h)$ on S^{n-1} belong to $\mathcal{H}_1$, and all the other stationary points are strict saddle points and belong to $\mathcal{H}_2$. The bounds in Theorem 5.2.1 on the Riemannian Hessian and the Riemannian gradient follows by setting

$$c(n,\theta,\rho) := \frac{\theta(1-3\theta)\rho^2}{2n}. \tag{5.13}$$

We complete the proof of Theorem 5.2.1 by giving the following result about $\mathcal{H}_1$.

Lemma 5.2.6. *If $h^* \in \mathcal{H}_1$, then for some $j \in [n]$,*

$$\|C_f R h^* \pm e_j\| \leq 2\sqrt{\rho}.$$

5.3 Optimization Method

5.3.1 Guaranteed First-Order Optimization Algorithm

Second-order methods over a Riemannian manifold are known to be able to escape saddle points, for example, the trust region method [75], and the negative curvature method [84]. Recent works proposed to solve dictionary learning [74], and phase retrieval [72] using these methods, without any special initialization schemes. Thanks to the geometric structure (Section 5.2) and the Lipschitz continuity of the objective function for our multichannel blind deconvolution formulation (Section 5.1), these second-order methods can recover signed shifted versions of the ground truth without special initialization.

Recently, first-order methods have been shown to escape strict saddle points with random initialization [77, 78]. In this chapter, we use the manifold gradient descent algorithm studied by Lee et al. [76]. One can initialize

the algorithm with a random $h^{(0)}$, and use the following iterative update:

$$h^{(t+1)} = \mathcal{A}(h^{(t)}) \coloneqq P_{S^{n-1}}\big(h^{(t)} - \gamma \widehat{\nabla}_L(h^{(t)})\big). \tag{5.14}$$

Each iteration takes a Riemannian gradient descent step in the tangent space, and does a retraction by normalizing the iterate (projecting onto S^{n-1}). Using the geometric structure introduced in Section 5.2, and some technical results in [75, 76], the following result gives a theoretical guarantee for manifold gradient descent for our formulation of MSBD: convergence to an accurate estimate (up to the inherent sign and shift ambiguity) of the true solution.

Theorem 5.3.1. *Suppose that the geometric structure in Theorem 5.2.1 is satisfied. If manifold gradient descent* (5.14) *is initialized with a random $h^{(0)}$ drawn from a uniform distribution on S^{n-1}, and the step size is chosen as $\gamma = \frac{1}{128n^3}$, then* (5.14) *converges to a local minimum of $L(h)$ on S^{n-1} almost surely. It particular, after at most $T = \frac{4096n^8}{\theta^2(1-3\theta)^2\rho^4}$ iterations, $h^{(T)} \in \mathcal{H}_1$. Moreover, for some $j \in [n]$*

$$\big\|C_f R h^{(T)} \pm e_j\big\| \leq 2\sqrt{\rho}.$$

One can further bound the recovery error of the signal and the channels as follows.

Corollary 5.3.2. *If the conditions of Theorem 5.3.1 are satisfied, then the recovered $\hat{f}$ and $\hat{x}_i$ ($i = 1, 2, \ldots, N$) in* (5.1) *and* (5.2)*, computed using the output of manifold gradient descent $\hat{h} = h^{(T)}$, satisfy*

$$\frac{\|\hat{x}_i \pm \mathcal{S}_j(x_i)\|}{\|x_i\|} \leq 2\sqrt{\rho n},$$
$$\frac{\big\|\hat{f} \pm S_{-j}(f)\big\|}{\|f\|} \leq \frac{2\sqrt{\rho n}}{1 - 2\sqrt{\rho n}},$$

for some $j \in [n]$.

Theorem 5.3.1 and Corollary 5.3.2 show that, with a random initialization and a fixed step size, manifold gradient descent outputs, in polynomial time, a solution that is close to a signed and shifted version of the ground truth.

5.3.2 Proof of the Algorithm Guarantee

We first establish that after T steps, the iterate $h^{(T)} \in \mathcal{H}_1 \cup \mathcal{H}_2$, by applying [75, Theorem 4]. To this end, one needs to show that **(C1)** $L(h)$ has a finite lower bound, and that **(C2)** the function $\widehat{L}(z) \coloneqq L(\frac{h+z}{\|h+z\|})$ (defined on $\{z : z \perp h\}$) is well approximated by its first-order Taylor expansion at $z = 0$. We verify conditions **(C1)** and **(C2)** in the following lemmas.

Lemma 5.3.3. *For all $h \in S^{n-1}$, $-4n^3 \le L(h) \le 0$, $\|\nabla_L(h)\| \le 16n^3$, $\|H_L(h)\| \le 48n^3$.*

Lemma 5.3.4. *Let $\widehat{L}(z) \coloneqq L(\frac{h+z}{\|h+z\|})$. Then for all $z \perp h$,*

$$\left|\widehat{L}(z) - \widehat{L}(0) - \langle z, \nabla_{\widehat{L}}(0)\rangle\right| \le 64n^3 \|z\|^2 .$$

By [75, Theorem 4] and Lemmas 5.3.3 and 5.3.4, manifold gradient decent (5.14) with a fixed step size $\gamma = 1/(2 \times 64n^3)$ achieves $\left\|\widehat{\nabla}_L(h^{(t)})\right\| < \tau$ after $t = 2[L(h^{(0)}) - \min_{h\in S^{n-1}} L(h)]/(\gamma\tau^2)$ iterations. Setting $\tau = \theta(1-3\theta)\rho^2/(2n)$ and $T = 4096n^8/[\theta^2(1-3\theta)^2\rho^4]$, it follows that

$$\left\|\widehat{\nabla}_L(h^{(t)})\right\| < \frac{\theta(1-3\theta)\rho^2}{2n} = c(n,\theta,\rho)$$

after $t \ge T$ iterations. By Theorem 5.2.1, we have $\{h^{(t)}\}_{t\ge T} \subset \mathcal{H}_1 \cup \mathcal{H}_2$. Since the distance between every pair of points $h_1 \in \mathcal{H}_1$ and $h_2 \in \mathcal{H}_2$ satisfies $\|h_1 - h_2\| \gg \gamma \left\|\widehat{\nabla}_L(h^{(t)})\right\|$, the iterates $\{h^{(t)}\}_{t\ge T}$ all belong to $\mathcal{H}_1$ or all belong to $\mathcal{H}_2$, and cannot jump from one set to the other.

Next, we show that if the initialization $h^{(0)}$ follows a random distribution on S^{n-1}, then $h^{(T)} \in \mathcal{H}_1$ almost surely, by applying [76, Theorem 2]. To this end, we verify that **(C3)** the strict saddle points are unstable fixed points of (5.14), and that **(C4)** the differential of $\mathcal{A}(\cdot)$ in (5.14) is invertible.

Let $h' \coloneqq \mathcal{A}(h) = P_{S^{n-1}}(h - \gamma\widehat{\nabla}_L(h))$. The differential $D\mathcal{A}(h)$ defined in [76, Definition 4] is

$$D\mathcal{A}(h) = P_{h'^{\perp}} P_{h^{\perp}} [I - \gamma \widehat{H}_L(h)] P_{h^{\perp}}. \tag{5.15}$$

At strict saddle points $\widehat{\nabla}_L(h) = 0$ and $h' = h$. Because, as we have shown, $\widehat{H}_L(h)$ has a strictly negative eigenvalue, it follows from [76, Proposition 8] that $D\mathcal{A}(h)$ has at least one eigenvalue larger than 1. Therefore, strict saddle

points are unstable fixed points of (5.14) (see [76, Definition 5]), i.e., **(C3)** is satisfied.

We verify **(C4)** in the following lemma.

Lemma 5.3.5. *For step size* $\gamma = \frac{1}{128n^3}$*, and all* $h \in S^{n-1}$*, we have* $\det(D\mathcal{A}(h)) \neq 0$.

Since conditions **(C3)** and **(C4)** are satisfied, by [76, Theorem 2], the set of initial points that converge to strict saddle points have measure 0. Therefore, a random $h^{(0)}$ uniformly distributed on S^{n-1} converges to a local minimum almost surely. Hence $\{h^{(t)}\}_{t\geq T} \subset \mathcal{H}_1$. By Lemma 5.2.6,

$$\left\|C_f R h^{(T)} \pm e_j\right\| \leq 2\sqrt{\rho},$$

for some $j \in [n]$.

5.4 Extensions

We believe that our formulation and/or analysis can be extended to other scenarios that are not covered by our theoretical guarantees.

- Bernoulli–sub-Gaussian channels. As stated at the beginning of Section 5.2, the Bernoulli-Rademacher assumption (A1) is a special case of the Bernoulli–sub-Gaussian distribution, which simplifies our analysis. Similar bounds can be established for general sub-Gaussian distributions.
- Jointly sparse channels. This is a special case where the supports of x_i $(i = 1, 2, \ldots, N)$ are identical. Due to the shared support, the x_i's are no longer independent. In this case, one needs a more careful analysis conditioned on the joint support.
- Complex signal and channels. We mainly consider real signals in this chapter. However, a similar approach can be derived and analyzed for complex signals. We discuss this extension in the rest of this section.

Empirical evidence that our method works in these scenarios is provided in Section 5.5.3.

For complex $f, x_i \in \mathbb{C}^n$, one can solve the following problem:

$$\min_{h\in\mathbb{C}^n} \quad \frac{1}{N}\sum_{i=1}^{N} \phi(\mathrm{Re}(C_{y_i}Rh)) + \phi(\mathrm{Im}(C_{y_i}Rh)), \quad \text{s.t.} \quad \|h\| = 1,$$

where $R \coloneqq (\frac{1}{\theta n N}\sum_{i=1}^{N} C_{y_i}^{\mathrm{H}} C_{y_i})^{-1/2} \in \mathbb{C}^{n\times n}$, and $(\cdot)^{\mathrm{H}}$ represents the Hermitian transpose. If one treats the real and imaginary parts of h separately, then this optimization in $\mathbb{C}^n$ can be recast into $\mathbb{R}^{2n}$, and the gradient with respect to $\mathrm{Re}(h)$ and $\mathrm{Im}(h)$ can be used in first-order methods. This is related to Wirtinger gradient descent algorithms (see the discussion in [103]). The Riemannian gradient with respect to h is

$$P_{(\mathbb{R}\cdot h)^\perp}\Big(\frac{1}{N}\sum_{i=1}^{N} R^{\mathrm{H}} C_{y_i}^{\mathrm{H}} w_i(h)\Big),$$

where $w_i(h)$ represents the following complex vector:

$$w_i(h) = \nabla_\phi(\mathrm{Re}(C_{y_i}Rh)) + \sqrt{-1}\nabla_\phi(\mathrm{Im}(C_{y_i}Rh)),$$

and $P_{(\mathbb{R}\cdot h)^\perp}$ represents the projection onto the tangent space at h in $S^{2n-1} \subset \mathbb{R}^{2n}$:

$$P_{(\mathbb{R}\cdot h)^\perp} z = z - \mathrm{Re}(h^{\mathrm{H}} z)\cdot h.$$

In the complex case, one can initialize the manifold gradient descent algorithm with a random $h^{(0)}$, for which $[\mathrm{Re}(h^{(0)})^\top, \mathrm{Im}(h^{(0)})^\top]^\top$ follows a uniform distribution on S^{2n-1}.

5.5 Numerical Experiments

5.5.1 Deconvolution with Synthetic Data

In this section, we examine the empirical performance of manifold gradient descent (5.14) in solving the multichannel sparse blind deconvolution problem (P1). We synthesize $\{x_i\}_{i=1}^N$ following the Bernoulli-Rademacher model, and synthesize f following a Gaussian distribution $N(\mathbf{0}_{n\times 1}, I_n)$. In all experiments, we run manifold gradient descent for $T = 100$ iterations, with a fixed step size of $\gamma = 0.1$.

Recall that the desired h is a signed shifted version of the ground truth, i.e., $C_f Rh = \pm e_j$ ($j \in [n]$) is a standard basis vector. Therefore, to evaluate the accuracy of the output $h^{(T)}$, we compute $C_f Rh^{(T)}$ with the true f, and declare successful recovery if

$$\frac{\left\|C_f Rh^{(T)}\right\|_\infty}{\|C_f Rh^{(T)}\|} > 0.95,$$

or equivalently, if

$$\max_{j\in[n]} \left|\cos\angle\left(C_f Rh^{(T)}, e_j\right)\right| > 0.95.$$

We compute the success rate based on 100 Monte Carlo instances.

In the first experiment, we fix $\theta = 0.1$ (sparsity level, mean of the Bernoulli distribution), and run experiments with $n = 32, 64, \ldots, 256$ and $N = 32, 64, \ldots, 256$ (see Figure 5.3(a)). In the second experiment, we fix $n = 256$, and run experiments with $\theta = 0.02, 0.04, \ldots, 0.16$ and $N = 32, 64, \ldots, 256$ (see Figure 5.3(d)). The empirical phase transitions suggest that, for sparsity level relatively small (e.g., $\theta < 0.12$), there exist a constant $c > 0$ such that manifold gradient descent can recover a signed shifted version of the ground truth with $N \geq cn\theta$.

In the third experiment, we examine the phase transition with respect to N and the condition number κ of f, which is the ratio of the largest and smallest magnitudes of its DFT. To synthesize f with specific κ, we generate the DFT $\tilde{f}$ of f that is random with the following distribution: (1) the DFT $\tilde{f}$ is symmetric, i.e., $\tilde{f}_{(j)} = \tilde{f}_{(n+2-j)}$, so that f is real; (2) the phase of $\tilde{f}_{(j)}$ follows a uniform distribution on $[0, 2\pi)$, except for the phases of $\tilde{f}_{(1)}$ and $\tilde{f}_{(n/2+1)}$ (if n is even), which are always 0, for symmetry; and (3) the gains of $\tilde{f}$ follows a uniform distribution on $[1, \kappa]$. We fix $n = 256$ and $\theta = 0.1$, and run experiments with $\kappa = 1, 2, 4, \ldots, 128$ and $N = 32, 64, \ldots, 256$ (see Figure 5.3(g)). The phase transition suggests that the number N for successful empirical recovery is not sensitive to the condition number κ.

Manifold gradient descent is robust against noise. We repeat the above experiments with noisy measurements: $y_i = x_i \circledast f + \sigma\varepsilon_i$, where ε_i follows a Gaussian distribution $N(\mathbf{0}_{n\times 1}, I_n)$. The phase transitions for $\sigma = 0.01\sqrt{n\theta}$ (SNR ≈ 40 dB) and $\sigma = 0.1\sqrt{n\theta}$ (SNR ≈ 20 dB) are shown in Figures 5.3(b), 5.3(e), 5.3(h), and Figures 5.3(c), 5.3(f), 5.3(i), respectively. For reasonable noise levels, the number N of noisy measurements we need to accurately

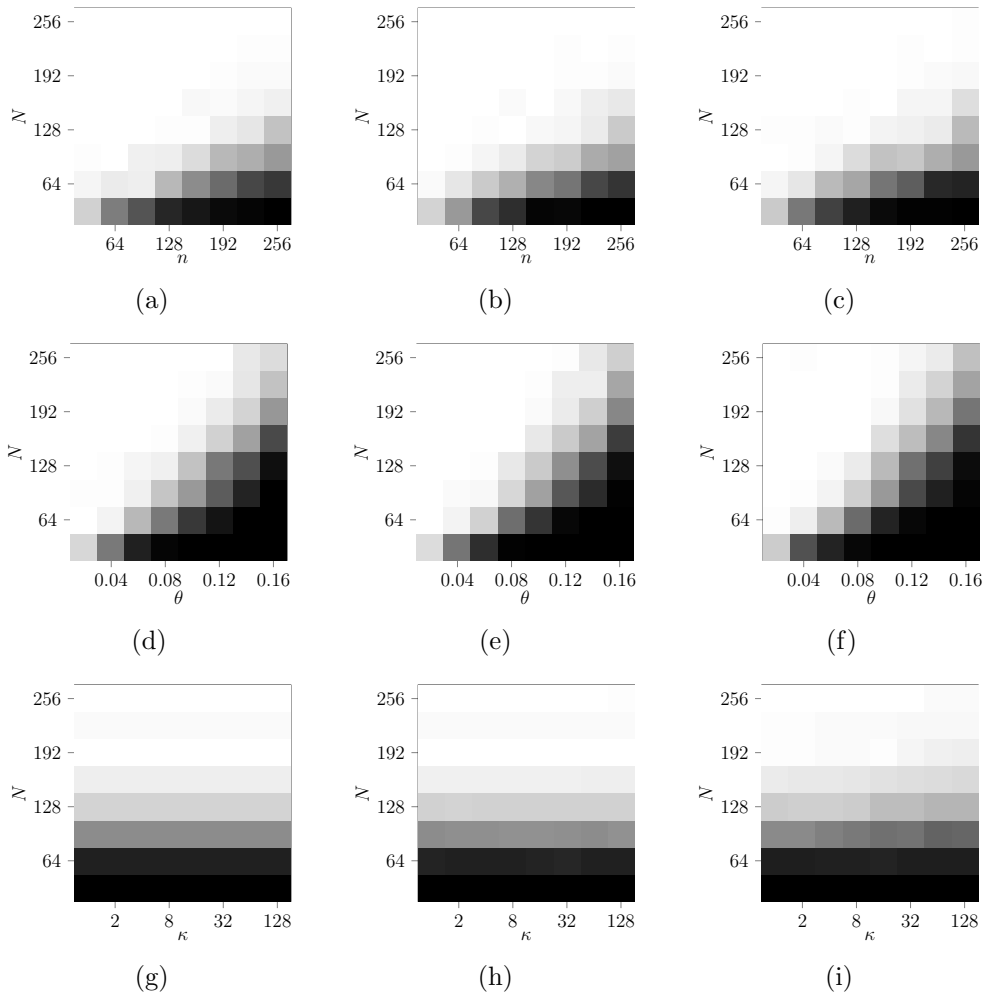


Figure 5.3: Empirical phase transition (grayscale values represent success rates). The first row shows the phase transitions of N versus n, given that $\theta = 0.1$. The second row shows the phase transitions of N versus θ, given that $n = 256$. The third row shows the phase transitions of N versus κ, given that $n = 256$ and $\theta = 0.1$. The first column shows the results for the noiseless case. The second column shows the results for SNR $\approx 40\,\mathrm{dB}$. The third column shows are the results for SNR $\approx 20\,\mathrm{dB}$.

recover a signed shifted version of the ground truth is roughly the same as with noiseless measurements.

5.5.2 2D Deconvolution

Next, we run a numerical experiment with blind image deconvolution. Suppose the circular convolutions $\{y_i\}_{i=1}^N$ (Figure 5.4(c)) of an unknown image f (Figure 5.4(a)) and unknown sparse channels $\{x_i\}_{i=1}^N$ (Figure 5.4(b)) are observed. The recovered image $\hat{f}$ (Figure 5.4(d)) is computed as follows:

$$\hat{f} = \mathcal{F}^{-1}\big[\mathcal{F}(Rh^{(T)})^{\odot -1}\big],$$

where $\mathcal{F}$ denotes the 2D DFT, and $h^{(T)}$ is the output of manifold gradient descent (5.14), with a random initialization $h^{(0)}$ that is uniformly distributed on the sphere.

Figure 5.4 shows that, although the sparse channels are completely unknown and the convolutional observations have corrupted the image beyond recognition, manifold gradient descent is capable of recovering a shifted version of the (negative) image, starting from a random point on the sphere (see the image recovered using a random initialization in Figure 5.4(d), and then corrected with the true sign and shift in Figure 5.4(e)). In this example, all images and channels are of size 64×64, the number of channels is $N = 256$, and the sparsity level is $\theta = 0.01$. We run $T = 100$ iterations of manifold gradient descent with a fixed step size $\gamma = 0.05$. The accuracy $\frac{\|C_f Rh^{(t)}\|_\infty}{\|C_f Rh^{(t)}\|}$ as a function of iteration number t is shown in Figure 5.4(f), and exhibits a sharp transition at a modest number (≈ 80) of iterations.

5.5.3 Jointly Sparse Complex Gaussian Channels

In this section, we examine the performance of manifold gradient descent when Assumption (A1) is not satisfied, and the channel model is extended as in Section 5.4. More specifically, we consider f following a Gaussian distribution $CN(\mathbf{0}_{n\times 1}, I_n)$, i.e., the real and imaginary parts are independent following $N(\mathbf{0}_{n\times 1}, I_n/2)$. And we consider $\{x_i\}_{i=1}^N$ that are:

- **Jointly s-sparse:** The joint support of $\{x_i\}_{i=1}^N$ is chosen uniformly at

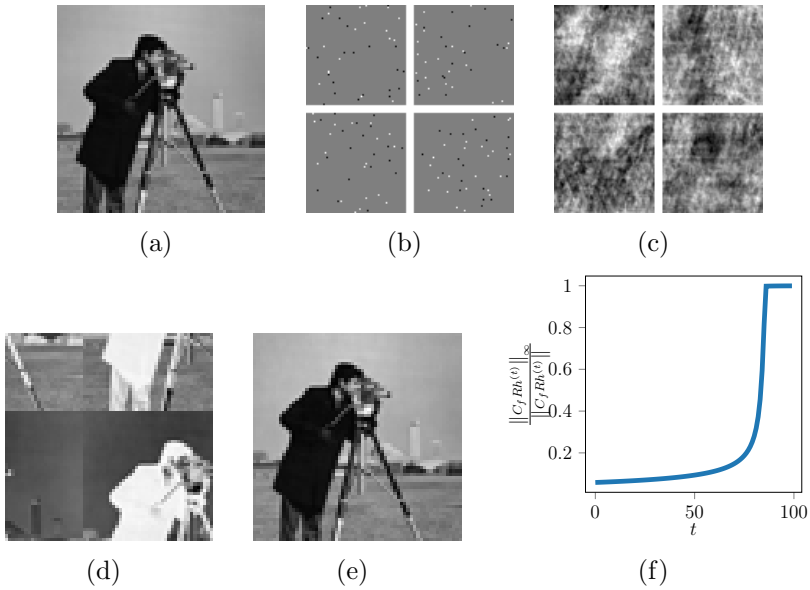


Figure 5.4: Multichannel blind image deconvolution. (a) True image. (b) Sparse channels. (c) Observations. (d) Recovered image using manifold gradient descent. (e) Recovered image with sign and shift correction. (f) The accuracy as a function of iteration number. All images and channels in this figure are of the same size (64×64).

random on $[n]$.

- **Complex Gaussian:** The nonzero entries of $\{x_i\}_{i=1}^N$ follow a complex Gaussian distribution $CN(0,1)$.

We compare manifold gradient descent (with random initialization) with three blind calibration algorithms that solve MSBD in the frequency domain: truncated power iteration [4] (initialized with $f^{(0)} = e_1$ and $x_i^{(0)} = 0$), an off-the-grid algebraic method [123] (simplified from [30]), and an off-the-grid optimization approach [92].

We fix $n = 128$, and run experiments for $N = 16, 32, 48, \cdots, 128$, and $s = 2, 4, 6, \ldots, 16$. We use f and $\hat{f}$ to denote the true signal and the recovered signal, respectively. We say the recovery is successful if[2]

$$\frac{\left\| \mathcal{F}^{-1}\left[\mathcal{F}(f) \odot \mathcal{F}(\hat{f})^{\odot -1}\right] \right\|_\infty}{\left\| \mathcal{F}^{-1}\left[\mathcal{F}(f) \odot \mathcal{F}(\hat{f})^{\odot -1}\right] \right\|} > 0.7. \tag{5.16}$$

[2]A perfect recovery $\hat{f}$ is a scaled shifted version of f, for which $\mathcal{F}^{-1}\left[\mathcal{F}(f) \odot \mathcal{F}(\hat{f})^{\odot -1}\right]$ is a scaled shifted Kronecker delta.

By the phase transitions in Figure 5.5, manifold gradient descent and truncated power iteration are successful when N is large and s is small. Although truncated power iteration achieves higher success rates when both N and s are small, it fails for $s > 8$ even with a large N. On the other hand, manifold gradient descent can recover channels with $s = 16$ when $N \geq 80$.[3] In comparison, the off-the-grid methods are based on the properties of the covariance matrix $\frac{1}{N}\sum_{i=1}^{N} y_i y_i^{\mathrm{H}}$, and require larger N (than the first two algorithms) to achieve high success rates.

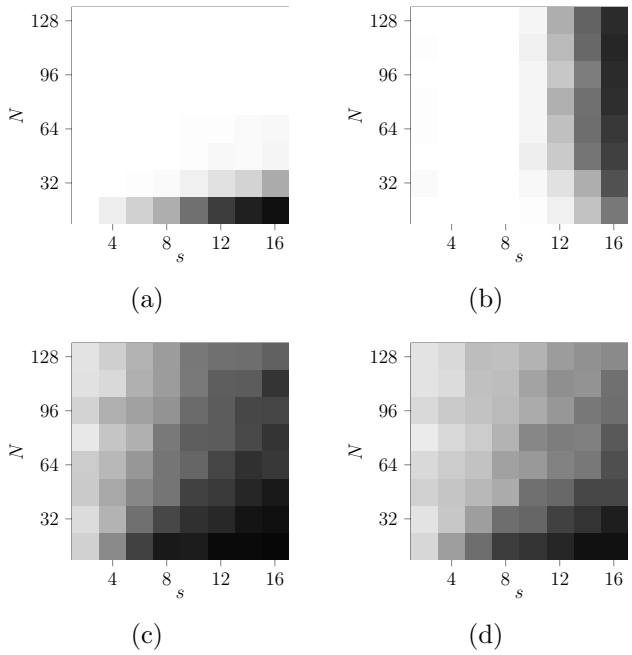


Figure 5.5: Empirical phase transition of N versus s, given that $n = 128$. (a) Manifold gradient descent. (b) Truncated power iteration [4]. (c) Off-the-grid algebraic method [123]. (d) Off-the-grid optimization approach [92].

5.5.4 MSBD with a Linear Convolution Model

In this section, we empirically study MSBD with a linear convolution model. Suppose the observations $y_i = x_i' * f' \in \mathbb{R}^n$ $(i = 1, 2, \ldots, N)$ are linear convolutions of s-sparse channels $x_i' \in \mathbb{R}^m$ and a signal $f' \in \mathbb{R}^{n-m+1}$. Let

[3] By our theoretical prediction, manifold gradient descent can succeed for $s = 40 < \frac{n}{3}$ provided that we have a sufficiently large N.

$x_i \in \mathbb{R}^n$ and $f \in \mathbb{R}^n$ denote the zero-padded versions of x_i' and f. Then

$$y_i = x_i' * f' = x_i \circledast f.$$

In this section, we show that one can solve for f and x_i using the optimization formulation (P1) and the manifold gradient descent algorithm, without knowledge of the length m of the channels.

We compare our approach to the subspace method based on cross convolution [55], which solves for the concatenation of the channels as a null vector of a structured matrix. For fairness, we also compare to an alternative method that takes advantage of the sparsity of the channels, and finds a sparse null vector of the same structured matrix as in [55], using truncated power iteration [44, 1].[4]

In our experiments, we synthesize f' using a random Gaussian vector following $N(\mathbf{0}_{(n-m+1)\times 1}, I_{n-m+1})$. We synthesize s-sparse channels x_i such that the support is chosen uniformly at random, and the nonzero entries are independent following $N(0,1)$. We denote the zero-padded versions of the true signal and the recovered signal by f and $\hat{f}$, and declare success if (5.16) is satisfied. We study the empirical success rates of our method and the competing methods in three experiments:

- N versus s, given that $n = 128$ and $m = 64$.
- N versus m, given that $n = 128$ and $s = 4$.
- N versus n, given that $m = 64$ and $s = 4$.

The phase transitions in Figure 5.6 show that our manifold gradient descent method consistently has higher success rates than the competing methods based on cross convolution. The subspace method and the truncated power iteration method are only successful when m is small compared to n, while our method is successful for a large range of m and n. The sparsity prior exploited by truncated power iteration improves the success rate over the subspace method, but only when the sparsity level s is small compared to m. In comparison, our method, given a sufficiently large number N of channels, can recover channels with a much larger s.

[4]For an example of finding sparse null vectors using truncated power iteration, we refer the readers to our previous paper [1, Section II].

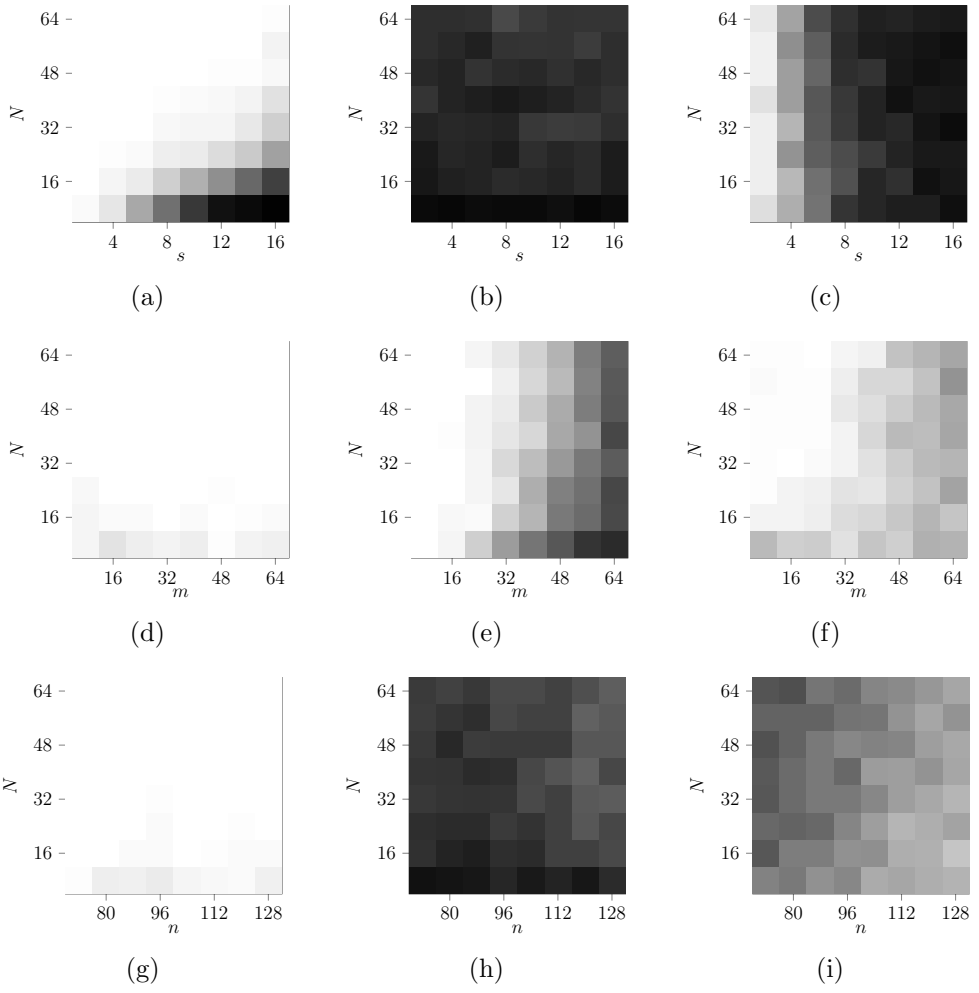


Figure 5.6: Empirical phase transition of MSBD with a linear convolution model. The first row shows the phase transitions of N versus s. The second row shows the phase transitions of N versus m. The third row shows the phase transitions of N versus n. The first column shows the results for manifold gradient descent. The second column shows the results for the subspace method [55]. The third column shows are the results for truncated power iteration.

5.5.5 Super-Resolution Fluorescence Microscopy

Manifold gradient descent can be applied to deconvolution of time resolved fluorescence microscopy images. The goal is to recover sharp images x_i's from observations y_i's that are blurred by an unknown PSF f.

We use a publicly available microtubule dataset [66], which contains $N = 626$ images (Figure 5.7(a)). Since fluorophores are are turned on and off stochastically, the images x_i's are random sparse samples of the 64×64 microtubule image (Figure 5.7(b)). The observations y_i's (Figures 5.7(c), 5.7(d)) are synthesized by circular convolutions with the PSF in Figure 5.7(i). The recovered images (Figures 5.7(e), 5.7(f)) and kernel (Figure 5.7(j)) clearly demonstrate the effectiveness of our approach in this setting.

Blind deconvolution is less sensitive to instrument calibration error than non-blind deconvolution. If the PSF used in a non-blind deconvolution method fails to account for certain optic aberration, the resulting images may suffer from spurious artifacts. For example, if we use a miscalibrated PSF (Figure 5.7(k)) in non-blind image reconstruction using FISTA [121], then the recovered images (Figures 5.7(g), 5.7(h)) suffer from serious spurious artifacts.

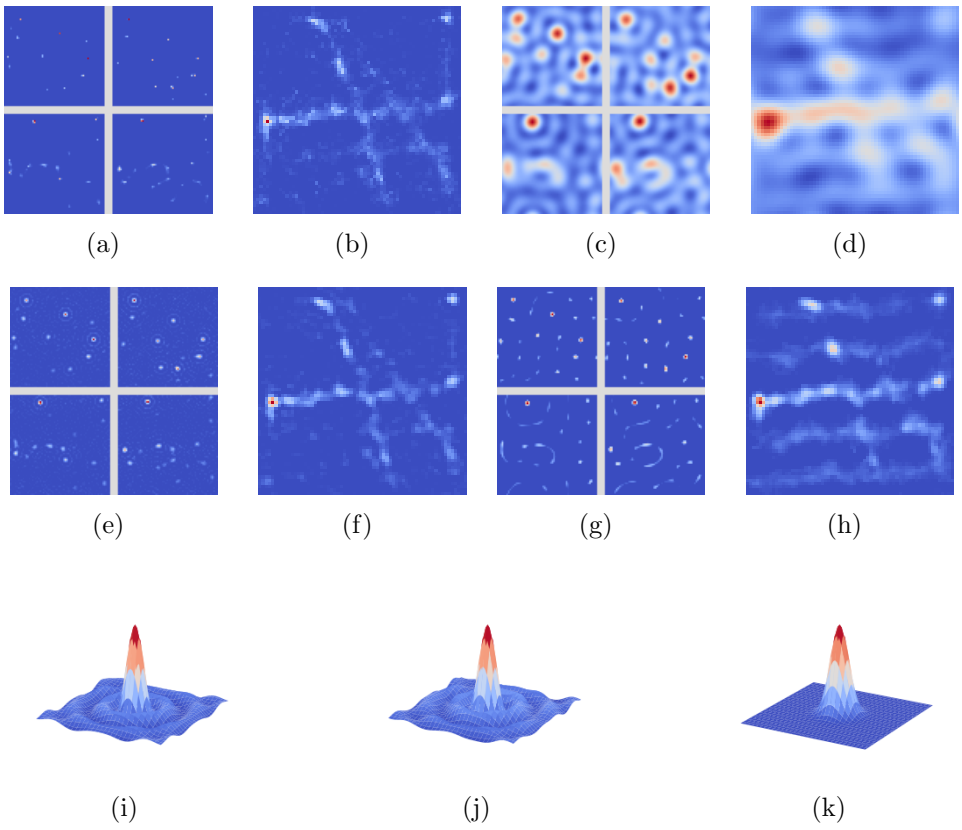


Figure 5.7: Super-resolution fluorescence microscopy experiment using manifold gradient descent. (a) True images. (b) Average of true images. (c) Observed images. (d) Average of observed images. (e) Recovered images using blind deconvolution. (f) Average of recovered images using blind deconvolution. (g) Recovered images using non-blind deconvolution and a miscalibrated PSF. (h) Average of recovered images using non-blind deconvolution and a miscalibrated PSF. (i) True PSF. (j) Recovered PSF using blind deconvolution. (k) Miscalibrated PSF used in non-blind deconvolution. All images in this figure are of the same size (64×64).

CHAPTER 6

CONCLUSION

In Chapter 2, we studied the identifiability of blind deconvolution problems with subspace or sparsity constraints. The sample complexity results in Section 2.2 are, to within a small additive term of at most five samples, optimal. Our results are derived with generic bases or frames, which means they are invalid only on a set of Lebesgue measure zero. If we assume that the bases or frames are drawn from any distribution that is absolutely continuous with respect to the Lebesgue measure on the space of bases or frames, then the results hold almost surely. Furthermore, if the bases or frames follow a distribution specified in Chapter 2, then under the same sample complexities, the recovery is not only unique with probability 1, but also stable with high probability against small perturbations in the measurements. These results provide the first *tight* sample complexity bounds, without large constants or log factors, for unique or stable recovery in blind deconvolution. They are fundamental to the blind deconvolution problem, independent of algorithms.

Despite the fact that, under the sufficient conditions in Chapter 2, the degenerate set of bases or frames has Lebesgue measure zero, it is unclear whether commonly used bases and frames (e.g., standard basis, wavelets) belong to the degenerate set. Therefore, it is an interesting open problem to show optimal sample complexity results for these bases and frames.

In Chapter 3, we addressed the identifiability of the BGPC problem with subspace or joint sparsity constraint, up to scaling. We provided sufficient conditions for identifiability that feature optimal (resp. near optimal) sample complexities for the subspace constraint case (resp. the joint sparsity constraint case). These results are for generic vectors or matrices, and are violated only for a set of Lebesgue measure zero. We did not address the stability of BGPC in Chapter 3. The regime under which the problem can be solved stably is an interesting open problem.

In Chapter 4, we formulated the BGPC problem as an eigenvector problem,

and proposed to solve BGPC using power iteration, and solve BGPC with a sparsity structure using truncated power iteration. We give theoretical guarantees for the subspace case with a near optimal sample complexity, and for the joint sparsity case with a suboptimal sample complexity. Numerical experiments show that both power iteration and truncated power iteration can recover the unknown gain and phase, and the unknown signal, using a near optimal number of samples. It is an open problem to obtain theoretical guarantees with optimal sample complexities, for truncated power iteration that solves BGPC with joint sparsity or sparsity constraints.

In Chapter 5, we studied the geometric structure of multichannel sparse blind deconvolution over the unit sphere. Our theoretical analysis reveals that local minima of a sparsity promoting smooth objective function correspond to signed shifted version of the ground truth, and saddle points have strictly negative curvatures. Thanks to the favorable geometric properties of the objective, we can simultaneously recover the unknown signal and unknown channels from convolutional measurements using manifold gradient descent with a random initialization. In practice, many convolutional measurement models are subsampled in the spatial domain (e.g., image super-resolution) or in the frequency domain (e.g., radio astronomy). Studying the effect of subsampling on the geometric structure of multichannel sparse blind deconvolution is an interesting problem for future work.

APPENDIX A

LEMMAS FOR CHAPTER 2

A.1 Concentration of Measure

Lemma A.1.1. *Suppose $a \in \mathbb{R}^{m_1}$ and $b \in \mathbb{R}^{m_2}$ are independent random vectors, following uniform distributions on $R\mathcal{B}_{\mathbb{R}^{m_1}}$ and $R\mathcal{B}_{\mathbb{R}^{m_2}}$, respectively. If a matrix $M \in \mathbb{R}^{m_1 \times m_2}$ satisfies $\ell \leq \|M\|_2 \leq L$, then*

$$\mathbb{P}\left[\left|a^T M b\right| \leq \rho\right] \leq \rho f(\rho, \ell, L, R),$$

where $f(\rho, \ell, L, R)$ satisfies $\lim_{\rho \to 0} \frac{\log f(\rho,\ell,L,R)}{\log \frac{1}{\rho}} = 0$.

Proof. Suppose the singular value decomposition (SVD) of M is

$$M = U \Sigma V^T,$$

where $U \in \mathbb{R}^{m_1 \times m_1}$ and $V \in \mathbb{R}^{m_2 \times m_2}$ are orthogonal matrices, and $\Sigma \in \mathbb{R}^{m_1 \times m_2}$ satisfies $\ell < \Sigma^{(1,1)} = \|M\|_2 < L$. Let $\tilde{a} \coloneqq U^T a$, and $\tilde{b} \coloneqq V^T b$, then $\tilde{a}$ and $\tilde{b}$ are also independent random vectors, following uniform distributions on $R\mathcal{B}_{\mathbb{R}^{m_1}}$ and $R\mathcal{B}_{\mathbb{R}^{m_2}}$, respectively.

Therefore,

$$\begin{aligned}
&\mathbb{P}\left[\left|a^T M b\right| \leq \rho\right] \\
&= \mathbb{P}\left[\left|\tilde{a}^T \Sigma \tilde{b}\right| \leq \rho\right] \\
&= \frac{\int_{R\mathcal{B}_{\mathbb{R}^{m_1}}} \mathrm{d}\tilde{a} \int_{R\mathcal{B}_{\mathbb{R}^{m_2}}} \mathrm{d}\tilde{b}\ \mathbf{1}\left(|\tilde{a}^T \Sigma \tilde{b}| \leq \rho\right)}{\int_{R\mathcal{B}_{\mathbb{R}^{m_1}}} \mathrm{d}\tilde{a} \int_{R\mathcal{B}_{\mathbb{R}^{m_2}}} \mathrm{d}\tilde{b}} \\
&= \frac{1}{V_{\mathbb{R}^{m_1}}(R) \cdot V_{\mathbb{R}^{m_2}}(R)} \int\limits_{R\mathcal{B}_{\mathbb{R}^{m_1-1}}} \mathrm{d}\tilde{a}^{(2:m_1)} \int\limits_{R\mathcal{B}_{\mathbb{R}^{m_2-1}}} \mathrm{d}\tilde{b}^{(2:m_2)}\ \phi(\tilde{a}, \tilde{b}), \qquad \text{(A.1)}
\end{aligned}$$

where $V_{\mathbb{R}^{m_1}}(R) = \int_{R\mathcal{B}_{\mathbb{R}^{m_1}}} \mathrm{d}\tilde{a}$ denotes the volume of a ball of radius R in $\mathbb{R}^{m_1}$,

and

$$
\begin{aligned}
&\phi(\tilde{a},\tilde{b}) && \text{(A.2)}\\
&= \int_{-R}^{R} \mathrm{d}\tilde{a}^{(1)} \int_{-R}^{R} \mathrm{d}\tilde{b}^{(1)}\ \mathbf{1}\left(|\tilde{a}^T\Sigma\tilde{b}| \le \rho\right)\\
&\qquad \cdot \mathbf{1}\left(|\tilde{a}^{(1)}|^2 \le R^2 - \left\|\tilde{a}^{(2:m_1)}\right\|_2^2\right) \cdot \mathbf{1}\left(|\tilde{b}^{(1)}|^2 \le R^2 - \left\|\tilde{b}^{(2:m_2)}\right\|_2^2\right)\\
&\le \int_{-R}^{R} \mathrm{d}\tilde{a}^{(1)} \int_{-R}^{R} \mathrm{d}\tilde{b}^{(1)} \mathbf{1}\Bigg(|\tilde{a}^{(1)}\tilde{b}^{(1)}\\
&\qquad + \frac{1}{\|M\|_2}\tilde{a}^{(2:m_1)T}\Sigma^{(2:m_1,2:m_2)}\tilde{b}^{(2:m_2)}| \le \frac{\rho}{\|M\|_2}\Bigg)\\
&\le \int_{-R}^{R} \mathrm{d}\tilde{a}^{(1)} \min\Bigg\{\int_{-\infty}^{\infty} \mathrm{d}\tilde{b}^{(1)} \mathbf{1}\left(|\tilde{b}^{(1)} + \theta(\tilde{a},\tilde{b}^{(2:m_2)})| \le \frac{\rho}{\|M\|_2 |\tilde{a}^{(1)}|}\right),\\
&\qquad \int_{-R}^{R} 1\ \mathrm{d}\tilde{b}^{(1)}\Bigg\} && \text{(A.3)}\\
&\le \int_{-R}^{R} \mathrm{d}\tilde{a}^{(1)} \min\left(\frac{2\rho}{\|M\|_2 |\tilde{a}^{(1)}|}, 2R\right) && \text{(A.4)}\\
&= \frac{4\rho}{\|M\|_2}\left(1 + \ln\frac{\|M\|_2 R^2}{\rho}\right)\\
&\le \frac{4\rho}{\ell}\left(1 + \ln\frac{LR^2}{\rho}\right). && \text{(A.5)}
\end{aligned}
$$

In (A.3), $\theta(\tilde{a},\tilde{b}^{(2:m_2)}) = \frac{1}{\|M\|_2\tilde{a}^{(1)}}\tilde{a}^{(2:m_1)T}\Sigma^{(2:m_1,2:m_2)}\tilde{b}^{(2:m_2)}$ does not affect the integral. Substituting (A.5) into (A.1), we obtain

$$
\mathbb{P}\left[\left|a^T M b\right| \le \rho\right] \le \frac{4\rho \cdot V_{\mathbb{R}^{m_1-1}}(R) \cdot V_{\mathbb{R}^{m_2-1}}(R)}{\ell \cdot V_{\mathbb{R}^{m_1}}(R) \cdot V_{\mathbb{R}^{m_2}}(R)}\left(1 + \ln\frac{LR^2}{\rho}\right).
$$

Define

$$
f(\rho,\ell,L,R) \coloneqq \frac{4 \cdot V_{\mathbb{R}^{m_1-1}}(R) \cdot V_{\mathbb{R}^{m_2-1}}(R)}{\ell \cdot V_{\mathbb{R}^{m_1}}(R) \cdot V_{\mathbb{R}^{m_2}}(R)}\left(1 + \ln\frac{LR^2}{\rho}\right).
$$

Clearly, $\lim_{\rho\to 0}\frac{\log f(\rho,\ell,L,R)}{\log\frac{1}{\rho}} = 0$. □

Lemma A.1.1 is a simplified version of [27, Lemma 4], with improved constants. Although [27, Lemma 4] has a better bound in terms of its dependence on ρ when the rank of M is larger than 1, our simplified bound does not affect

our proof of identifiability in any negative way. As a bonus, we can deduce stability results directly from the simplified proof of identifiability.

Next, we derive a similar concentration of measure bounds for the complex case. Despite the similarity between the proofs of Lemmas A.1.1 and A.1.2, the latter is not a direct consequence of the former.

Lemma A.1.2. *Suppose $a \in \mathbb{C}^{m_1}$ and $b \in \mathbb{C}^{m_2}$ are independent random vectors, following uniform distributions on $R\mathcal{B}_{\mathbb{C}^{m_1}}$ and $R\mathcal{B}_{\mathbb{C}^{m_2}}$, respectively. If a matrix $M \in \mathbb{C}^{m_1 \times m_2}$ satisfies $\ell \leq \|M\|_2 \leq L$, then*

$$\mathbb{P}\left[\left|a^* M \bar{b}\right| \leq \rho\right] \leq \rho^2 g(\rho, \ell, L, R),$$

where $g(\rho, \ell, L, R)$ satisfies $\lim_{\rho \to 0} \frac{\log g(\rho,\ell,L,R)}{\log \frac{1}{\rho}} = 0$.

Proof. The proof follows steps mostly analogous to those in the proof of Lemma A.1.1 by replacing the real field by the complex field. Here, we define $\tilde{a} \coloneqq U^T \overline{a}$, and $\tilde{b} \coloneqq V^* \overline{b}$. It follows that (A.1) – (A.3) apply, with the real field replaced by the complex field, and the interval of integration $[-R, R]$ replaced by the disk in the complex plane $R\mathcal{B}_{\mathbb{C}^1}$. Then (A.4) – (A.5) are replaced by

$$\begin{aligned}
\phi(\tilde{a}, \tilde{b}) &\leq \int_{R\mathcal{B}_{\mathbb{C}^1}} \mathrm{d}\tilde{a}^{(1)} \min\left(\frac{\pi\rho^2}{\|M\|_2^2 \, |\tilde{a}^{(1)}|^2}, \pi R^2\right) \\
&= \frac{\pi^2 \rho^2}{\|M\|_2^2} \left(1 + 2\ln \frac{\|M\|_2 R^2}{\rho}\right) \leq \frac{\pi^2 \rho^2}{\ell^2} \left(1 + 2\ln \frac{LR^2}{\rho}\right).
\end{aligned}$$

In a manner analogous to the proof of Lemma A.1.1, it follows that

$$\mathbb{P}\left[\left|a^* M \bar{b}\right| \leq \rho\right] \leq \frac{\pi^2 \rho^2 \cdot V_{\mathbb{C}^{m_1-1}}(R) \cdot V_{\mathbb{C}^{m_2-1}}(R)}{\ell^2 \cdot V_{\mathbb{C}^{m_1}}(R) \cdot V_{\mathbb{C}^{m_2}}(R)} \left(1 + 2\ln \frac{LR^2}{\rho}\right).$$

Here we use $V_{\mathbb{C}^{m_1}}(R) = \int_{R\mathcal{B}_{\mathbb{C}^{m_1}}} \mathrm{d}\tilde{a}$ to denote the volume of a ball of radius R in $\mathbb{C}^{m_1}$. Define

$$g(\rho, \ell, L, R) \coloneqq \frac{\pi^2 \cdot V_{\mathbb{C}^{m_1-1}}(R) \cdot V_{\mathbb{C}^{m_2-1}}(R)}{\ell^2 \cdot V_{\mathbb{C}^{m_1}}(R) \cdot V_{\mathbb{C}^{m_2}}(R)} \left(1 + 2\ln \frac{LR^2}{\rho}\right). \tag{A.6}$$

Clearly, $\lim_{\rho \to 0} \frac{\log g(\rho,\ell,L,R)}{\log \frac{1}{\rho}} = 0$. □

A.2 Useful Lemmas about Minkowski Dimension

Lemma A.2.1. *Let $\Omega_{\mathcal{X}}$ and $\Omega_{\mathcal{Y}}$ be nonempty bounded subsets of a normed vector space. Then $\overline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{X}} - \Omega_{\mathcal{Y}}) \leq \overline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{X}}) + \overline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{Y}})$.*

Proof. We cover $\Omega_{\mathcal{X}}$ and $\Omega_{\mathcal{Y}}$ with balls of radius ρ centered at $\{x_i\}_{i=1}^{N_{\Omega_{\mathcal{X}}}(\rho)}$ and $\{y_i\}_{i=1}^{N_{\Omega_{\mathcal{Y}}}(\rho)}$, respectively. Given any point $x - y \in \Omega_{\mathcal{X}} - \Omega_{\mathcal{Y}}$, we can find centers of the above covering, x_{i_1} and y_{i_2}, such that

$$\|x - x_{i_1}\| \leq \rho, \quad \|y - y_{i_2}\| \leq \rho.$$

Hence,

$$\|(x - y) - (x_{i_1} - y_{i_2})\| \leq \|x - x_{i_1}\| + \|y - y_{i_2}\| \leq 2\rho.$$

Therefore, the set $\Omega_{\mathcal{X}} - \Omega_{\mathcal{Y}}$ can be covered by $N_{\Omega_{\mathcal{X}}}(\rho)N_{\Omega_{\mathcal{Y}}}(\rho)$ balls of radius 2ρ centered at points (like $x_{i_1} - y_{i_2}$) generated by the centers $\{x_i\}_{i=1}^{N_{\Omega_{\mathcal{X}}}(\rho)}$ and $\{y_i\}_{i=1}^{N_{\Omega_{\mathcal{Y}}}(\rho)}$. It follows that

$$N_{(\Omega_{\mathcal{X}} - \Omega_{\mathcal{Y}})}(2\rho) \leq N_{\Omega_{\mathcal{X}}}(\rho)N_{\Omega_{\mathcal{Y}}}(\rho).$$

We then bound the Minkowski dimension:

$$\begin{aligned}
&\overline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{X}} - \Omega_{\mathcal{Y}}) = \limsup_{\rho \to 0} \frac{\log N_{(\Omega_{\mathcal{X}} - \Omega_{\mathcal{Y}})}(2\rho)}{\log \frac{1}{2\rho}} \\
&\leq \limsup_{\rho \to 0} \frac{\log N_{\Omega_{\mathcal{X}}}(\rho)N_{\Omega_{\mathcal{Y}}}(\rho)}{\log \frac{1}{2\rho}} \leq \limsup_{\rho \to 0} \frac{\log N_{\Omega_{\mathcal{X}}}(\rho)}{\log \frac{1}{2\rho}} + \limsup_{\rho \to 0} \frac{\log N_{\Omega_{\mathcal{Y}}}(\rho)}{\log \frac{1}{2\rho}} \\
&= \overline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{X}}) + \overline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{Y}}).
\end{aligned}$$

□

Lemma A.2.2. *Let $\Omega_{\mathcal{X}}$ and $\Omega_{\mathcal{Y}}$ be nonempty bounded subsets of $\mathbb{C}^{m_1}$ and $\mathbb{C}^{m_2}$, respectively. Let $\Omega_{\mathcal{M}} = \{xy^T : x \in \Omega_{\mathcal{X}}, y \in \Omega_{\mathcal{Y}}\} \subset \mathbb{C}^{m_1 \times m_2}$. Then $\overline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}}) \leq \overline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{X}}) + \overline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{Y}})$.*

Proof. Since $\Omega_{\mathcal{X}}$ and $\Omega_{\mathcal{Y}}$ are bounded, there exists a large enough constant L such that

$$\Omega_{\mathcal{X}} \subset L\mathcal{B}_{\mathbb{C}^{m_1}}, \qquad \Omega_{\mathcal{Y}} \subset L\mathcal{B}_{\mathbb{C}^{m_2}}.$$

We cover $\Omega_{\mathcal{X}}$ and $\Omega_{\mathcal{Y}}$ with balls of radius ρ centered at the following two sets

of points, respectively:

$$\{x_i\}_{i=1}^{N_{\Omega_{\mathcal{X}}}(\rho)} \subset L\mathcal{B}_{\mathbb{C}^{m_1}}, \qquad \{y_i\}_{i=1}^{N_{\Omega_{\mathcal{Y}}}(\rho)} \subset L\mathcal{B}_{\mathbb{C}^{m_2}}.$$

Given any point $xy^T \in \Omega_{\mathcal{M}}$, we can find centers of the above coverings, x_{i_1} and y_{i_2}, such that

$$\|x - x_{i_1}\|_2 \leq \rho, \qquad \|y - y_{i_2}\|_2 \leq \rho.$$

Then

$$\begin{aligned}&\left\|xy^T - x_{i_1}y_{i_2}^T\right\|_{\mathrm{F}} = \left\|xy^T - x_{i_1}y^T + x_{i_1}y^T - x_{i_1}y_{i_2}^T\right\|_{\mathrm{F}} \\ &\leq \|x - x_{i_1}\|_2 \|y\|_2 + \|y - y_{i_2}\|_2 \|x_{i_1}\|_2 \leq 2L\rho.\end{aligned}$$

Therefore, the set $\Omega_{\mathcal{M}}$ can be covered by $N_{\Omega_{\mathcal{X}}}(\rho)N_{\Omega_{\mathcal{Y}}}(\rho)$ balls in $\mathbb{C}^{m_1 \times m_2}$ of radius $2L\rho$, centered at the rank-1 matrices (like $x_{i_1}y_{i_2}^T$) generated by the centers of the coverings of $\Omega_{\mathcal{X}}$ and $\Omega_{\mathcal{Y}}$. It follows that

$$N_{\Omega_{\mathcal{M}}}(2L\rho) \leq N_{\Omega_{\mathcal{X}}}(\rho)N_{\Omega_{\mathcal{Y}}}(\rho). \tag{A.7}$$

Therefore,

$$\begin{aligned}&\overline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{M}}) = \limsup_{\rho\to 0} \frac{\log N_{\Omega_{\mathcal{M}}}(2L\rho)}{\log \frac{1}{2L\rho}} \\ &\leq \limsup_{\rho\to 0} \frac{\log N_{\Omega_{\mathcal{X}}}(\rho)N_{\Omega_{\mathcal{Y}}}(\rho)}{\log \frac{1}{2L\rho}} \leq \limsup_{\rho\to 0} \frac{\log N_{\Omega_{\mathcal{X}}}(\rho)}{\log \frac{1}{2L\rho}} + \limsup_{\rho\to 0} \frac{\log N_{\Omega_{\mathcal{Y}}}(\rho)}{\log \frac{1}{2L\rho}} \\ &= \overline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{X}}) + \overline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{Y}}).\end{aligned}$$

□

Lemma A.2.3. *Let $\Omega_{\mathcal{X}}$ be a nonempty bounded subset of $\mathbb{C}^m$. Let $\mathrm{Re}(\Omega_{\mathcal{X}}) = \{\mathrm{Re}(x) : x \in \Omega_{\mathcal{X}}\}$, and $\mathrm{Im}(\Omega_{\mathcal{X}}) = \{\mathrm{Im}(x) : x \in \Omega_{\mathcal{X}}\}$. Then $\overline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{X}}) \leq \overline{\dim}_{\mathrm{B}}(\mathrm{Re}(\Omega_{\mathcal{X}})) + \overline{\dim}_{\mathrm{B}}(\mathrm{Im}(\Omega_{\mathcal{X}}))$.*

Proof. The real and imaginary parts $\mathrm{Re}(\Omega_{\mathcal{X}})$ and $\mathrm{Im}(\Omega_{\mathcal{X}})$ are bounded subsets of $\mathbb{R}^m$. There exists a large enough constant L such that

$$\mathrm{Re}(\Omega_{\mathcal{X}}), \mathrm{Im}(\Omega_{\mathcal{X}}) \subset L\mathcal{B}_{\mathbb{R}^m}.$$

We cover $\mathrm{Re}(\Omega_{\mathcal{X}})$ and $\mathrm{Im}(\Omega_{\mathcal{X}})$ with balls of radius ρ centered at the following two sets of points, respectively:

$$\left\{x_i^{\mathrm{Re}}\right\}_{i=1}^{N_{\mathrm{Re}(\Omega_{\mathcal{X}})}(\rho)},\ \left\{x_i^{\mathrm{Im}}\right\}_{i=1}^{N_{\mathrm{Im}(\Omega_{\mathcal{X}})}(\rho)} \subset L\mathcal{B}_{\mathbb{R}^m}.$$

Given any point $x \in \Omega_{\mathcal{X}}$, we can find centers of the above coverings, $x_{i_1}^{\mathrm{Re}}$ and $x_{i_2}^{\mathrm{Im}}$, such that

$$\left\|\mathrm{Re}(x) - x_{i_1}^{\mathrm{Re}}\right\|_2 \leq \rho, \qquad \left\|\mathrm{Im}(x) - x_{i_2}^{\mathrm{Im}}\right\|_2 \leq \rho.$$

Let $x_c = x_{i_1}^{\mathrm{Re}} + \sqrt{-1}x_{i_2}^{\mathrm{Im}}$. Then

$$\|x - x_c\|_2 = \sqrt{\left\|\mathrm{Re}(x) - x_{i_1}^{\mathrm{Re}}\right\|_2^2 + \left\|\mathrm{Im}(x) - x_{i_2}^{\mathrm{Im}}\right\|_2^2} \leq \sqrt{2}\rho.$$

Therefore, the set $\Omega_{\mathcal{X}}$ can be covered by $N_{\mathrm{Re}(\Omega_{\mathcal{X}})}(\rho)N_{\mathrm{Im}(\Omega_{\mathcal{X}})}(\rho)$ balls in $\mathbb{C}^m$ of radius $\sqrt{2}\rho$, centered at the complex vectors (like x_c) generated by the centers of the coverings of $\mathrm{Re}(\Omega_{\mathcal{X}})$ and $\mathrm{Im}(\Omega_{\mathcal{X}})$. It follows that

$$N_{\Omega_{\mathcal{X}}}(\sqrt{2}\rho) \leq N_{\mathrm{Re}(\Omega_{\mathcal{X}})}(\rho)N_{\mathrm{Im}(\Omega_{\mathcal{X}})}(\rho). \tag{A.8}$$

Therefore,

$$\begin{aligned}
\overline{\dim}_{\mathrm{B}}(\Omega_{\mathcal{X}}) &= \limsup_{\rho\to 0}\frac{\log N_{\Omega_{\mathcal{X}}}(\sqrt{2}\rho)}{\log\frac{1}{\sqrt{2}\rho}} \\
&\leq \limsup_{\rho\to 0}\frac{\log N_{\mathrm{Re}(\Omega_{\mathcal{X}})}(\rho)N_{\mathrm{Im}(\Omega_{\mathcal{X}})}(\rho)}{\log\frac{1}{\sqrt{2}\rho}} \\
&\leq \limsup_{\rho\to 0}\frac{\log N_{\mathrm{Re}(\Omega_{\mathcal{X}})}(\rho)}{\log\frac{1}{\sqrt{2}\rho}} + \limsup_{\rho\to 0}\frac{\log N_{\mathrm{Im}(\Omega_{\mathcal{X}})}(\rho)}{\log\frac{1}{\sqrt{2}\rho}} \\
&= \overline{\dim}_{\mathrm{B}}(\mathrm{Re}(\Omega_{\mathcal{X}})) + \overline{\dim}_{\mathrm{B}}(\mathrm{Im}(\Omega_{\mathcal{X}})).
\end{aligned}$$

□

APPENDIX B

PROOFS OF LEMMAS IN CHAPTER 4

B.1 Gap in Eigenvalues

Proof of Lemma 4.4.1. We have

$$D^*D = I_N \otimes (A^*A), \tag{B.1}$$

$$D^*E_{\mathrm{s}} = \begin{bmatrix} \lambda_1\overline{a_{1\cdot}}a_{1\cdot}^\top x_{\cdot 1} & \cdots & \lambda_n\overline{a_{n\cdot}}a_{n\cdot}^\top x_{\cdot 1} \\ \vdots & \ddots & \vdots \\ \lambda_1\overline{a_{1\cdot}}a_{1\cdot}^\top x_{\cdot N} & \cdots & \lambda_n\overline{a_{n\cdot}}a_{n\cdot}^\top x_{\cdot N} \end{bmatrix}, \tag{B.2}$$

$$E_{\mathrm{s}}^*E_{\mathrm{s}} = \begin{bmatrix} |\lambda_1|^2 a_{1\cdot}^\top XX^*\overline{a_{1\cdot}} & & \\ & \ddots & \\ & & |\lambda_n|^2 a_{n\cdot}^\top XX^*\overline{a_{n\cdot}} \end{bmatrix}. \tag{B.3}$$

Under Assumptions 4.3.1 and 4.3.3, we have

$$\mathbb{E}D^*D = I_{Nm}, \tag{B.4}$$

$$\mathbb{E}D^*E_{\mathrm{s}} = \frac{1}{n}x\lambda^\top, \tag{B.5}$$

$$\mathbb{E}E_{\mathrm{s}}^*E_{\mathrm{s}} = \frac{1}{n}\left\|X\right\|_{\mathrm{F}}^2 \operatorname{diag}([|\lambda_1|^2,\ldots,|\lambda_n|^2]) = \frac{1}{n}\operatorname{diag}([|\lambda_1|^2,\ldots,|\lambda_n|^2]). \tag{B.6}$$

Set $\alpha = \sqrt{n}$, we have

$$\mathbb{E}B_{\mathrm{s}} = \begin{bmatrix} I_{Nm} & \frac{1}{\sqrt{n}}x\lambda^\top \\ \frac{1}{\sqrt{n}}\overline{\lambda}x^* & \operatorname{diag}([|\lambda_1|^2,\ldots,|\lambda_n|^2]) \end{bmatrix},$$

and

$$\mathbb{E}\Omega_{T_\eta} B_{\mathrm{s}} \Omega_{T_\eta}^* = \begin{bmatrix} I_{Ns} & \frac{1}{\sqrt{n}}\Omega_{T_x} x \lambda^\top \\ \frac{1}{\sqrt{n}}\overline{\lambda} x^* \Omega_{T_x}^* & \mathrm{diag}([|\lambda_1|^2, \ldots, |\lambda_n|^2]) \end{bmatrix} = P^* Q P,$$

where

$$P = \mathrm{diag}([\mathbf{1}_{1,Ns}, \lambda^\top]),$$

$$Q = \begin{bmatrix} I_{Ns} & \frac{1}{\sqrt{n}}\Omega_{T_x} x \mathbf{1}_{n,1}^\top \\ \frac{1}{\sqrt{n}}\mathbf{1}_{n,1} x^* \Omega_{T_x}^* & I_n \end{bmatrix}.$$

The matrix Q has eigenvalues $0, 1, 1, \ldots, 1, 2$. The eigenvectors corresponding to 0 and 2 are $\mu = [(\Omega_{T_x} x)^\top, -\mathbf{1}_{n,1}^\top/\sqrt{n}]^\top/\sqrt{2}$ and $[(\Omega_{T_x} x)^\top, \mathbf{1}_{n,1}^\top/\sqrt{n}]^\top/\sqrt{2}$, respectively. Any vector orthogonal to these two vectors is an eigenvector of Q corresponding to 1. It follows that $Q + \mu\mu^* - I_{Ns+n}$ is positive semidefinite.

Since μ is a null vector of Q, we have $P^{-1}\mu$ is a null vector of P^*QP (note that $\Omega_{T_\eta}\eta = \sqrt{2}P^{-1}\mu$). Therefore, the smallest eigenvalue of the positive semidefinite matrix P^*QP is 0.

Next, we bound the largest eigenvalue of P^*QP, which satisfies

$$\begin{aligned} &\max_{\|z\|_2 \le 1} \|P^*QPz\|_2 \le \sqrt{1+\delta} \max_{\|Pz\|_2 \le \sqrt{1+\delta}} \|QPz\|_2 \\ &= (1+\delta) \max_{\|z\|_2 \le 1} \|Qz\|_2 \le 2(1+\delta), \end{aligned} \tag{B.7}$$

where the first inequality follows from Assumption 4.3.2, and the second inequality follows from the largest eigenvalue of Q.

Next, we bound the second smallest eigenvalue of P^*QP, which satisfies

$$\begin{aligned} &\min_{z \perp P^{-1}\mu,\ \|z\|_2 \ge 1} \|P^*QPz\|_2 \\ &\ge \sqrt{1-\delta} \min_{Pz \perp (PP^*)^{-1}\mu,\ \|Pz\|_2 \ge \sqrt{1-\delta}} \|QPz\|_2 \\ &= (1-\delta) \min_{z \perp (PP^*)^{-1}\mu,\ \|z\|_2 \ge 1} \|Qz\|_2 \\ &\ge (1-\delta) \min_{z \perp (PP^*)^{-1}\mu,\ \|z\|_2 = 1} \|(I_{Ns+n} - \mu\mu^*)z\|_2 \\ &= (1-\delta) \min_{z \perp (PP^*)^{-1}\mu,\ \|z\|_2 = 1} \sqrt{1 - |\mu^* z|^2} \\ &= (1-\delta) \frac{|\mu^* (PP^*)^{-1}\mu|}{\|(PP^*)^{-1}\mu\|_2} \ge \frac{(1-\delta)^2}{1+\delta}, \end{aligned} \tag{B.8}$$

where the first and third inequalities follow from Assumption 4.3.2, and the second inequality is due to the fact that $Q + \mu\mu^* - I_{Ns+n}$ is positive semidefinite.

By (B.7) and (B.8), all nonzero eigenvalues of $\mathbb{E}\Omega_{T_\eta} B_{\mathrm{s}} \Omega_{T_\eta}^*$ reside in the interval $[\frac{(1-\delta)^2}{1+\delta},\, 2(1+\delta)]$. □

B.2 Bounds of Perturbation Due to Randomness in A

Proof of Lemma 4.4.3. We prove only the joint sparsity case. One can prove the subspace case by replacing s with m and getting rid of the union bound.

It is well-known that, for sufficiently large n, a Gaussian random matrix satisfies RIP [113]. Here, we use a bound for real Gaussian random matrices [124], and present its extension to complex Gaussian random matrices. Let $T \subset [m]$ denote an index set of cardinality s, i.e., $|T| = s < n$. Let $\widehat{A} \coloneqq [\mathrm{Re}(A)\Omega_T^*, \mathrm{Im}(A)\Omega_T^*]$. By [124, Theorem 2.13],

$$\mathbb{P}\Big[\Big\|2\widehat{A}^*\widehat{A} - I_{2s}\Big\| \le 3\Big(\sqrt{\frac{2s}{n}} + \varepsilon\Big)\Big] \ge 1 - 2\exp\Big(-\frac{n\varepsilon^2}{2}\Big).$$

Note also that

$$\begin{aligned}
\Omega_T A^* A \Omega_T^* = {} & \Omega_T \mathrm{Re}(A)^\top \mathrm{Re}(A) \Omega_T^* \\
& + \sqrt{-1}\Omega_T \mathrm{Re}(A)^\top \mathrm{Im}(A) \Omega_T^* \\
& - \sqrt{-1}\Omega_T \mathrm{Im}(A)^\top \mathrm{Re}(A) \Omega_T^* \\
& + \Omega_T \mathrm{Im}(A)^\top \mathrm{Im}(A) \Omega_T^*.
\end{aligned}$$

$$\begin{aligned}
\|\Omega_T A^* A \Omega_T^* - I_s\| \le {} & \big\|\Omega_T \mathrm{Re}(A)^\top \mathrm{Re}(A) \Omega_T^* - I_s/2\big\| \\
& + \big\|\Omega_T \mathrm{Re}(A)^\top \mathrm{Im}(A) \Omega_T^*\big\| \\
& + \big\|\Omega_T \mathrm{Im}(A)^\top \mathrm{Re}(A) \Omega_T^*\big\| \\
& + \big\|\Omega_T \mathrm{Im}(A)^\top \mathrm{Im}(A) \Omega_T^* - I_s/2\big\| \\
\le {} & 4\left\|\widehat{A}^*\widehat{A} - I_{2s}/2\right\|.
\end{aligned}$$

It follows that

$$\mathbb{P}\Big[\|\Omega_T A^* A \Omega_T^* - I_s\| \le 6\Big(\sqrt{\frac{s}{n}} + \varepsilon\Big)\Big] \ge 1 - 2\exp\Big(-\frac{n\varepsilon^2}{2}\Big).$$

Therefore, there exist constants $C_1, c_1 > 0$, such that

$$\begin{aligned}
&\mathbb{P}\Big[\|\Omega_T A^* A \Omega_T^* - I_s\| \le C_1\sqrt{\frac{s}{n}\log m},\ \forall T \text{ s.t. } |T| = s\Big] \\
&\ge 1 - 2\binom{m}{s}\exp\Big(-\Big(\frac{C_1}{6} - 1\Big)^2\frac{s}{2}\log m\Big) \\
&\ge 1 - m^{-c_1 s},
\end{aligned}$$

where the first inequality follows from a union bound, and setting $\varepsilon = (\frac{C_1}{6} - 1)\sqrt{\frac{s}{n}\log m}$; the second inequality follows from Stirling's approximation $\binom{m}{s} \le \big(\frac{em}{s}\big)^s$.

We obtain Lemma 4.4.3 by applying the above bound to every diagonal block of the block diagonal matrix $\Omega_{T_x} D^* D \Omega_{T_x}^*$. □

Proof of Lemma 4.4.4. By a consequence of the Hanson-Wright inequality (see [125, Theorem 2.1], and its complexification in [125, Section 3.1]), there exists an absolute constant c_2' such that

$$\mathbb{P}\Big[\big|\sqrt{n}\,\big\|X^\top a_{k\cdot}\big\|_2 - 1\big| \le \varepsilon\Big] \ge 1 - 2\exp\Big(-\frac{c_2'\varepsilon^2}{\|X\|^2}\Big). \tag{B.9}$$

Set $\varepsilon = C_2' \|X\| \sqrt{\log n}$ for some $C_2' > 0$, then by a union bound, there exists an absolute constant $c_2 > 0$ such that

$$\mathbb{P}\Big[\big|\sqrt{n}\,\big\|X^\top a_{k\cdot}\big\|_2 - 1\big| \le C_2' \|X\| \sqrt{\log n},\ \forall k \in [n]\Big] \ge 1 - n^{-c_2}. \tag{B.10}$$

By Assumption 4.3.2,

$$\begin{aligned}
&\mathbb{P}\Big[|\lambda_k|^2 \big|a_{k\cdot}^\top X X^* \overline{a_{k\cdot}} - \frac{1}{n}\big| \le \frac{(2C_2' + C_2'^2)(1+\delta)}{n} \\
&\cdot \max\Big\{\|X\|\sqrt{\log n}, \|X\|^2 \log n\Big\},\ \forall k \in [n]\Big] \\
&\ge 1 - n^{-c_2}.
\end{aligned} \tag{B.11}$$

The spectral norm $\|X\|$ is bounded in Assumption 4.3.3:

$$\textbf{Subspace case:} \quad \|X\|^2 \leq (1+\theta)\max\{\frac{1}{N}, \frac{1}{m}\},$$
$$\textbf{Joint sparsity case:} \quad \|X\|^2 \leq (1+\theta)\max\{\frac{1}{N}, \frac{1}{s_0}\}.$$

Therefore, Lemma 4.4.4 follows from (B.3), (B.6), and (B.11). □

Proof of Lemma 4.4.5. By (B.2), the columns of D^*E_{s} are independent random vectors. Define

$$\phi_k \coloneqq \begin{bmatrix} \overline{a_{k\cdot}} a_{k\cdot}^\top x_{\cdot 1} \\ \overline{a_{k\cdot}} a_{k\cdot}^\top x_{\cdot 2} \\ \vdots \\ \overline{a_{k\cdot}} a_{k\cdot}^\top x_{\cdot N} \end{bmatrix}.$$

Then $D^*E_{\mathrm{s}} = [\phi_1, \phi_2, \ldots, \phi_n]\mathrm{diag}(\lambda)$. Next, we bound the spectral norm of the random matrix $\Phi - \mathbb{E}\Phi$, where $\Phi := [\phi_1, \phi_2, \ldots, \phi_n]$, using matrix Bernstein inequality [115, Theorem 1.6]. We need the following bounds to proceed:

(1) A bound on $\|\phi_k - \mathbb{E}\phi_k\|_2$.

First, by [125, Theorem 2.1 and Section 3.1], there exists a constant c_3'

$$\mathbb{P}\Big[\big|\sqrt{n}\,\|a_{k\cdot}\|_2 - \sqrt{m}\big| \leq \varepsilon\Big] \geq 1 - 2\exp(-c_3'\varepsilon^2).$$

By a union bound over all $k \in [n]$, there exists a constant C_3' such that

$$\begin{aligned} &\mathbb{P}\Big[\big|\sqrt{n}\,\|a_{k\cdot}\|_2 - \sqrt{m}\big| \leq C_3'\sqrt{\log n},\ \forall k \in [n]\Big] \\ &\geq 1 - 2n\exp\big(-c_3' C_3'^2 \log n\big) \\ &\geq 1 - n^{-c_2}. \end{aligned} \tag{B.12}$$

Note that

$$\|\mathbb{E}\phi_k\|_2 = \frac{1}{n}\|X\|_{\mathrm{F}} = \frac{1}{n},$$
$$\|\phi_k\|_2 \leq \|a_{k\cdot}\|_2 \left\|X^\top a_{k\cdot}\right\|_2.$$

By (B.10) and (B.12), there exists a constant C_3'', such that with probability

at least $1 - 2n^{-c_2}$,

$$\begin{aligned}
&\|\phi_k - \mathbb{E}\phi_k\|_2 \\
&\leq \frac{C_3''}{n} \max\left\{\sqrt{m}, \sqrt{\log n}\right\} \max\left\{1, \sqrt{\frac{\log n}{N}}, \sqrt{\frac{\log n}{m}}\right\} \\
&\leq \frac{C_3''\sqrt{m}}{n},
\end{aligned}$$

for all $k \in [n]$, where the second inequality uses the assumption that $\min\{N, m\} > \log n$.

(2) A bound on $\|\mathbb{E}[(\Phi - \mathbb{E}\Phi)^*(\Phi - \mathbb{E}\Phi)]\|$.

One should observe that

$$\mathbb{E}[(\phi_k - \mathbb{E}\phi_k)^*(\phi_k - \mathbb{E}\phi_k)] = \frac{m}{n^2},$$

$$\mathbb{E}[(\phi_k - \mathbb{E}\phi_k)^*(\phi_{k'} - \mathbb{E}\phi_{k'})] = 0,$$

for $k \neq k'$. Therefore,

$$\mathbb{E}[(\Phi - \mathbb{E}\Phi)^*(\Phi - \mathbb{E}\Phi)] = \frac{m}{n^2} I_n,$$

$$\|\mathbb{E}[(\Phi - \mathbb{E}\Phi)^*(\Phi - \mathbb{E}\Phi)]\| = \frac{m}{n^2}.$$

(3) A bound on $\|\mathbb{E}[(\Phi - \mathbb{E}\Phi)(\Phi - \mathbb{E}\Phi)^*]\|$.

Since $\{\phi_k\}_{k=1}^n$ are i.i.d. random vectors,

$$\begin{aligned}
&\mathbb{E}[(\Phi - \mathbb{E}\Phi)(\Phi - \mathbb{E}\Phi)^*] \\
&= \sum_{k=1}^{n} \mathbb{E}[(\phi_k - \mathbb{E}\phi_k)(\phi_k - \mathbb{E}\phi_k)^*] \\
&= n\mathbb{E}[(\phi_1 - \mathbb{E}\phi_1)(\phi_1 - \mathbb{E}\phi_1)^*] \\
&= n[\mathbb{E}(\phi_1\phi_1^*) - (\mathbb{E}\phi_1)(\mathbb{E}\phi_1)^*] \\
&= \frac{1}{n}(X^\top \overline{X} \otimes I_m).
\end{aligned}$$

By Assumption 4.3.3, in the subspace case,

$$\|\mathbb{E}[(\Phi - \mathbb{E}\Phi)(\Phi - \mathbb{E}\Phi)^*]\| = \frac{1}{n}\left\|X^\top \overline{X}\right\| \leq \frac{1+\theta}{n} \max\{\frac{1}{N}, \frac{1}{m}\}.$$

Given the above bounds, we apply the matrix Bernstein inequality [115,

Theorem 1.6] as follows:

$$\mathbb{P}\Big[\|\Phi - \mathbb{E}\Phi\| \le \varepsilon \Big| \, \|\phi_k - \mathbb{E}\phi_k\|_2 \le R, \ \forall k \in [n]\Big] \\ \ge 1 - (Nm+n)\exp\Big(-\frac{\varepsilon^2/2}{\sigma^2 + R\varepsilon/3}\Big),$$

where

$$\sigma^2 = \max\Big\{\frac{m}{n^2}, \frac{1+\theta}{nN}, \frac{1+\theta}{nm}\Big\},$$
$$R = \frac{C_3''\sqrt{m}}{n}.$$

It follows that

$$\mathbb{P}\Big[\|\Phi - \mathbb{E}\Phi\| \le \varepsilon\Big] \ge 1 - (Nm+n)\exp\Big(-\frac{\varepsilon^2/2}{\sigma^2 + R\varepsilon/3}\Big) - 2n^{-c_2},$$

where the last term $2n^{-c_2}$ bounds the probability that $\|\phi_k - \mathbb{E}\phi_k\|_2 > R$ for some k. Hence there exist constants $C_3, c_3 > 0$ such that

$$\mathbb{P}\Big[\|\Phi - \mathbb{E}\Phi\| \le \frac{C_3}{\sqrt{1+\delta}} \max\Big\{\sqrt{\frac{\log(Nm+n)}{nN}}, \\ \sqrt{\frac{\log(Nm+n)}{nm}}, \frac{\sqrt{m}\log(Nm+n)}{n}\Big\}\Big] \ge 1 - n^{-c_3}.$$

Lemma 4.4.4 follows from the above bound, and

$$\|\Omega_{T_x} D^* E_{\mathrm{s}} - \mathbb{E}\Omega_{T_x} D^* E_{\mathrm{s}}\| = \|\Phi - \mathbb{E}\Phi\| \, \|\mathrm{diag}(\lambda)\| \le \sqrt{1+\delta} \, \|\Phi - \mathbb{E}\Phi\| .$$

□

Proof of Lemma 4.4.6. We introduce some notations for this proof. We use B_p^n and $B_{S_p^{m,n}}$ to denote unit balls in $\mathbb{C}^n$ with ℓ_p norm, and in $\mathbb{C}^{m\times n}$ with Schatten p norm, respectively. The projection on the support set T is denoted by Π_T. For a set $\mathcal{A}$ of matrices, $d_{\mathrm{F}}(\mathcal{A})$ and $d_{\mathrm{op}}(\mathcal{A})$ denote the radii of $\mathcal{A}$ in the Frobenius norm and in the spectral norm, respectively. We use $\gamma_2(\mathcal{A}, \|\cdot\|)$ the γ_2 functional of $\mathcal{A}$, which is another way to quantify the size of $\mathcal{A}$ [116, Section 2.2]. These are key quantities in the upper bound of the supremum of an asymmetric second-order process [116, Theorem 2.3], which we use to prove Lemma 4.4.6.

Note that

$$\max_{\substack{T\subset[m]\\|T|=s}} \|\Omega_{T_x} D^* E_{\mathrm{s}} - \mathbb{E}\Omega_{T_x} D^* E_{\mathrm{s}}\| = \max_{\substack{T\subset[m]\\|T|=s}} \max_{\substack{v\in B_2^{mN}\\(I_N\otimes\Pi_T)v=v}} \max_{u\in B_2^n} |v^*\Phi u - \mathbb{E}v^*\Phi u|, \tag{B.13}$$

where $\Phi = D^* E_{\mathrm{s}}$. Let $z = \sqrt{n}[a_{1\cdot}^*, \dots, a_{n\cdot}^*]^\top$. Then z follows $\mathcal{CN}(\mathbf{0}_{mn,1}, I_{mn})$ and $v^*\Phi u$ is written as a quadratic form in z as follows:

$$v^*\Phi u = \sum_{k=1}^{n}\sum_{j=1}^{N} u_k a_{k\cdot}^\top x_{\cdot j} v_{\cdot j}^* \overline{a_{k\cdot}} = z^*(\mathrm{diag}(u)\otimes\Pi_{T_0})\Big(\frac{1}{n} I_n \otimes XV^*\Big) z, \tag{B.14}$$

where $u = [u_1, \dots, u_n]^\top$, $v = [v_{\cdot 1}^\top, \dots, v_{\cdot N}^\top]^\top$, $V = [v_{\cdot 1}, \dots, v_{\cdot N}]$, and $T_0 = \{i \in [m] | \left\|e_i^\top X\right\|_2 > 0\}$ denotes the row support of $X = [x_{\cdot 1}, \dots, x_{\cdot N}]$.

Let

$$\mathcal{A} = \{A_u | u \in B_2^n\},$$

and

$$\mathcal{B} = \{B_v | v \in B_2^{mN},\ (I_N \otimes \Pi_T)v = v\},$$

where A_u and B_v are left and right factors in the quadratic form in (B.14), i.e.,

$$A_u = \mathrm{diag}(u) \otimes \Pi_{T_0},$$

and

$$B_v = \frac{1}{n} I_n \otimes XV^*.$$

Then (B.13) is equivalent to

$$\sup_{A_u\in\mathcal{A}} \sup_{B_v\in\mathcal{B}} |z^* A_u B_v z - \mathbb{E} z^* A_u B_v z|,$$

which is a supremum of an asymmetric second-order process. We use the result on suprema of asymmetric second-order chaos processes by Lee and Junge [116, Theorem 2.3], which extends the original result by Krahmer et al. [126] to asymmetric cases.

Next, we compute the key quantities, given as functions of $\mathcal{A}$ and $\mathcal{B}$, which we need to apply [116, Theorem 2.3]. Let $A_u \in \mathcal{A}$. Since $|T_0| \le s_0$, we have

$$\|A_u\|_{\mathrm{F}} = \sqrt{s_0}\,\|u\|_2 \le \sqrt{s_0}$$

and the radius of $\mathcal{A}$ in the Frobenius norm satisfies

$$d_{\mathrm{F}}(\mathcal{A}) \le \sqrt{s_0}.$$

On the other hand,

$$\|A_u\| = \|u\|_\infty \le 1,$$

which implies that the radius of $\mathcal{A}$ in the spectral norm satisfies

$$d_{\mathrm{op}}(\mathcal{A}) \le 1.$$

Moreover, for $A_u, A_u' \in \mathcal{A}$, we have

$$\|A_u - A_{u'}\| = \|u - u'\|_\infty .$$

Therefore, by the Dudley's inequality [127],

$$\begin{aligned}
\gamma_2(\mathcal{A}, \|\cdot\|) &\lesssim \int_0^\infty \sqrt{\log N(\mathcal{A}, \|\cdot\|\,; t)}dt \\
&\le \int_0^\infty \sqrt{\log N(B_2^n, \|\cdot\|_\infty\,; t)}dt \\
&\lesssim \int_0^\infty \sqrt{\log N(B_1^n, \|\cdot\|_2\,; t)}dt \\
&\lesssim \log^{3/2} n,
\end{aligned}$$

where the third step follows from the entropy duality result by Artstein et al. [128] and the last step follows from Maurey's empirical method [129] (also see [130, Lemma 3.1]). Collecting the above estimates shows that the relevant quantities are given by

$$\begin{aligned}
&\gamma_2(\mathcal{A}, \|\cdot\|)(d_{\mathrm{F}}(\mathcal{A}) + \gamma_2(\mathcal{A}, \|\cdot\|)) + d_{\mathrm{F}}(\mathcal{A}) d_{\mathrm{op}}(\mathcal{A}) \\
&\qquad \lesssim \max\{\sqrt{s_0}\log^{3/2} n,\ \log^3 n\}, \\
&d_{\mathrm{op}}(\mathcal{A})(\gamma_2(\mathcal{A}, \|\cdot\|) + d_{\mathrm{F}}(\mathcal{A})) \lesssim \max\{\sqrt{s_0},\ \log^{3/2} n\}, \\
&d_{\mathrm{op}}(\mathcal{A})^2 \le 1.
\end{aligned}$$

Next we consider the other set $\mathcal{B}$. Let $B_v \in \mathcal{B}$. Then

$$\|B_v\|_{\mathrm{F}} = \frac{1}{\sqrt{n}} \|XV^*\|_{\mathrm{F}} \le \frac{1}{\sqrt{n}} \|X\| \, \|V\|_{\mathrm{F}} = \frac{1}{\sqrt{n}} \|X\| .$$

Therefore

$$d_{\mathrm{F}}(\mathcal{B}) \le \frac{1}{\sqrt{n}} \|X\| .$$

On the other hand,

$$\|B_v\| = \frac{1}{n} \|XV^*\| \le \frac{1}{n} \|X\| \, \|V\| ,$$

which implies

$$d_{\mathrm{op}}(\mathcal{B}) \le \frac{1}{n} \|X\| .$$

Moreover, for $B_v, B_{v'} \in \mathcal{B}$, we have

$$\|B_v - B_{v'}\| \le \frac{1}{n} \|X\| \, \|V - V'\| ,$$

where $V' = [v'_{\cdot 1}, \dots, v'_{\cdot N}]$ and $v' = [v'^{\top}_{\cdot 1}, \dots, v'^{\top}_{\cdot N}]^{\top}$. Therefore,

$$\begin{aligned}
&\gamma_2(\mathcal{B}, \|\cdot\|) \\
&\lesssim \frac{1}{n} \|X\| \int_0^{\infty} \sqrt{\log N(\cup_{|T|=s} \Pi_T B_{S_2^{m,N}}, \|\cdot\|_{S_\infty^{m,N}} ; t)} dt \\
&\le \frac{1}{n} \|X\| \int_0^{1} \sqrt{\log N(\cup_{|T|=s} \Pi_T B_{S_2^{m,N}}, \|\cdot\|_{S_\infty^{m,N}} ; t)} dt \\
&\le \frac{1}{n} \|X\| \int_0^{1} \sqrt{\log \sum_{|T|=s} N(\Pi_T B_{S_2^{m,N}}, \|\cdot\|_{S_\infty^{m,N}} ; t)} dt \\
&\le \frac{1}{n} \|X\| \int_0^{1} \sqrt{s \log m + \log N(B_{S_2^{s,N}}, \|\cdot\|_{S_\infty^{s,N}} ; t)} dt \\
&\le \frac{1}{n} \|X\| \left(\sqrt{s \log m} + \int_0^{1} \sqrt{\log N(B_{S_2^{s,N}}, \|\cdot\|_{S_\infty^{s,N}} ; t)} dt \right) \\
&\lesssim \frac{1}{n} \|X\| \sqrt{s+N} \log(sN+m),
\end{aligned}$$

where the last step follows from Lemma B.2.1. Therefore, the parameters for $\mathcal{B}$ are estimated as

$$\begin{aligned}
&\gamma_2(\mathcal{B}, \|\cdot\|)(d_{\mathrm{F}}(\mathcal{B}) + \gamma_2(\mathcal{B}, \|\cdot\|)) + d_{\mathrm{F}}(\mathcal{B}) d_{\mathrm{op}}(\mathcal{B}) \\
&\quad \lesssim \frac{1}{n^2} \|X\|^2 \left((s+N) \log^2(sN+m) + \sqrt{s+N}\sqrt{n} \log(sN+m) \right), \\
&d_{\mathrm{op}}(\mathcal{B})(\gamma_2(\mathcal{B}, \|\cdot\|) + d_{\mathrm{F}}(\mathcal{B})) \lesssim \frac{1}{n^2} \|X\|^2 \left(\sqrt{s+N} \log(sN+m) + \sqrt{n} \right), \\
&d_{\mathrm{op}}(\mathcal{B})^2 \le \frac{1}{n^2} \|X\|^2 .
\end{aligned}$$

According to [116, Theorem 2.3], the optimal upper bound is obtained as the geometric mean of the dominant parameters for the two sets. More precisely, the suprema is (up to an absolute constant) no larger than

$$\frac{s_0^{1/4}(s+N)^{1/4}(\sqrt{n}+\sqrt{s+N})^{1/2}}{n} \cdot \|X\| \log^3 n \log(sN+m)$$

with probability $1 - n^{-c_3}$. By Assumptions 4.3.2 and 4.3.3,

$$|\lambda_k| \leq \sqrt{1+\delta},$$

$$\|X\| \leq \max\Big\{\sqrt{\frac{1+\theta}{N}},\ \sqrt{\frac{1+\theta}{s_0}}\Big\},$$

which completes the proof. □

Lemma B.2.1.

$$\int_0^\infty \sqrt{\log N(B_{S_2^{m,N}}, tB_{S_\infty^{m,N}})}dt \lesssim \sqrt{m+N}\log(mN).$$

Proof of Lemma B.2.1. First, by the dual entropy result by Artstein et al. [128], we have

$$\log N(B_{S_2^{m,N}}, tB_{S_\infty^{m,N}}) \lesssim \log N(B_{S_1^{m,N}}, tB_{S_2^{m,N}}).$$

Then we approximate the S_1 ball as a polytope using a trick proposed by Junge and Lee [130]. Let R be the set of all rank-1 matrices in the unit sphere of $S_2^{m,N}$. Then $B_{S_1^{m,N}}$ is the absolute convex hull of R. We construct an ϵ-net Δ_m of the sphere S^{m-1}. Then

$$|\Delta_m| \leq \Big(1+\frac{2}{\epsilon}\Big)^m.$$

For an arbitrary $f \in S^{m-1}$, we have a sequence $\{f_l\}_{l=1}^\infty \subset \Delta_m$ such that

$$f = \sum_{l=1}^\infty \alpha_l f_l,$$

and

$$\sum_{l=1}^\infty |\alpha_l| \leq \frac{1}{1-\epsilon}.$$

The existence of such a sequence follows from the optimality of the construction of the net. Similarly we construct an ϵ-net $\Delta_N \subset S^{N-1}$ of S^{N-1}. Then

$$|\Delta_N| \leq \left(1 + \frac{2}{\epsilon}\right)^N.$$

For an arbitrary $g \in S^{N-1}$, we have a sequence $\{g_k\}_{k=1}^{\infty} \subset \Delta_N$ such that

$$g = \sum_{k=1}^{\infty} \beta_k g_k$$

and

$$\sum_{k=1}^{\infty} |\beta_k| \leq \frac{1}{1-\epsilon}.$$

Therefore,

$$fg^* = \sum_{l,k=1}^{\infty} \alpha_l \beta_k f_l g_k^*$$

and

$$\sum_{l,k=1}^{\infty} |\alpha_l||\beta_k| \leq \left(\frac{1}{1-\epsilon}\right)^2.$$

We can choose ϵ so that

$$\left(\frac{1}{1-\epsilon}\right)^2 \leq 2$$

and

$$1 + \frac{2}{\epsilon} \leq 8.$$

Let $\Delta_{m,N} = \Delta_m \times \Delta_N$. Then

$$\log(|\Delta_{m,N}|) \leq (m+N)\log 8$$

and

$$B_{S_1^{m,N}} \subset 2\text{absconv}(\Delta_{m,N}).$$

Now, it suffices to compute

$$\int_0^{\infty} \sqrt{\log N(2\text{absconv}(\Delta_{m,N}), tB_{S_2^{m,N}})}dt.$$

Then use a change of variable and get

$$\begin{aligned}&\int_0^\infty \sqrt{\log N(2\mathrm{absconv}(\Delta_{m,N}), tB_{S_2^{m,N}})}dt\\&= 2\int_0^\infty \sqrt{\log N(\mathrm{absconv}(\Delta_{m,N}), tB_{S_2^{m,N}})}dt.\end{aligned}$$

Let $\Delta_{m,N} = \{q_1, \ldots, q_M\}$, where $M = |\Delta_{m,N}|$. Define linear mapping $Q : \ell_1^M \to \ell_2^{mN}$ by $Q(e_i) = \mathrm{vec}(q_i)$ for $i = 1, \ldots, M$. Since $\|\mathrm{vec}(q_i)\|_2 = \|q_i\|_{S_2} = 1$ for all i, we have

$$\left\|Q : \ell_1^M \to \ell_2^{mN}\right\| = 1.$$

Note

$$\int_0^\infty \sqrt{\log N(\mathrm{absconv}(\Delta_{m,N}), tB_{S_2^{m,N}})}dt = \int_0^\infty \sqrt{\log N(Q(B_1^M), tB_{\ell_2^{mN}})}dt.$$

By a version of Maurey's empirical method (see for example [130, Proposition 3.2]), we have

$$\int_0^\infty \sqrt{\log N(Q(B_1^M), tB_{\ell_2^{mN}})}dt \lesssim \sqrt{\log M}\log(mN) \lesssim \sqrt{m+N}\log(mN).$$

This completes the proof. □

B.3 Bounds of Perturbation Due to Noise

Proof of Lemma 4.4.8. Bear in mind that the columns of $\Psi \coloneqq D^* E_{\mathrm{n}}$, which we denote by $\{\psi_k\}_{k=1}^n$, are independent random vectors with zero mean:

$$\psi_k \coloneqq \begin{bmatrix}\overline{a_{k\cdot}} w_{k1}\\ \overline{a_{k\cdot}} w_{k2}\\ \vdots\\ \overline{a_{k\cdot}} w_{kN}\end{bmatrix}.$$

We bound $\|D^* E_{\mathrm{n}}\|$ using the matrix Bernstein inequality [115, Theorem 1.6]. We need the following bounds:

(1) A bound on $\|\psi_k\|_2$.

Since

$$\|\psi_k\|_2 \le \|a_{k\cdot}\|_2 \|w_{k\cdot}\|_2$$

By (B.12), and $m > \log n$,

$$\|\psi_k\|_2 \leq (C_3' + 1)\sqrt{\frac{m}{n}} \times \sqrt{N} \max_{k\in[n], j\in[N]} |w_{kj}|,$$

with probability at least $1 - n^{-c_2}$.

(2) A bound on $\|\mathbb{E}\Psi^*\Psi\|$.

Since

$$\mathbb{E}\Psi^*\Psi = \frac{m}{n}\mathrm{diag}([\|w_{1\cdot}\|_2^2, \|w_{2\cdot}\|_2^2, \ldots, \|w_{k\cdot}\|_2^2]),$$

we have

$$\|\mathbb{E}\Psi^*\Psi\| = \frac{m}{n}\max_{k\in[n]}\|w_{k\cdot}\|_2^2 \leq \frac{mN}{n}\max_{k\in[n],j\in[N]}|w_{kj}|^2.$$

(3) A bound on $\|\mathbb{E}\Psi\Psi^*\|$.

Since

$$\mathbb{E}\Psi\Psi^* = \sum_{k\in[n]}\frac{1}{n}\mathrm{diag}([|w_{k1}|^2, |w_{k2}|^2, \ldots, |w_{kN}|^2]) \otimes I_m,$$

we have

$$\|\mathbb{E}\Psi\Psi^*\| = \frac{1}{n}\max_{j\in[N]}\sum_{k\in[n]}|w_{kj}|^2 \leq \max_{k\in[n],j\in[N]}|w_{kj}|^2.$$

Given the above bounds, we completes the proof using the matrix Bernstein inequality (similar to the proof of Lemma 4.4.5). There exist constants $C_4, c_4 > 0$ such that

$$\|D^*E_{\mathrm{n}}\| = \|\Psi\| \leq C_4 \max\Big\{\sqrt{\log(Nm+n)}, \sqrt{\frac{Nm}{n}}\log(Nm+n)\Big\}\max_{k\in[n],j\in[N]}|w_{kj}|,$$

with probability at least $1 - n^{-c_4}$. □

Proof of Lemma 4.4.9. Note that

$$\max_{\substack{T\subset[m]\\|T|=s}}\|\Omega_{T_x}D^*E_{\mathrm{n}}\| = \max_{\substack{T\subset[m]\\|T|=s}}\max_{\substack{v\in B_2^{mN}\\(I_N\otimes\Pi_T)v=v}}\max_{u\in B_2^n}|v^*\Psi u|,$$

where

$$\Psi = D^* E_{\mathrm{n}} = \begin{bmatrix} I_N \otimes \overline{a_{1\cdot}} & \dots & I_N \otimes \overline{a_{n\cdot}} \end{bmatrix} \begin{bmatrix} w_{1\cdot} & & \\ & \ddots & \\ & & w_{n\cdot} \end{bmatrix}.$$

Let $z = \sqrt{n}[a_{1\cdot}^*, \dots, a_{n\cdot}^*]^\top$. Then z is a standard Gaussian vector, and

$$v^* \Psi u = \frac{1}{\sqrt{n}} (\mathbf{1}_{1,n} \otimes v^*)(E_{\mathrm{n}} \otimes I_m)(\mathrm{diag}(u) \otimes I_m) z.$$

Let

$$q_{u,v} \coloneqq \frac{1}{\sqrt{n}} (\mathrm{diag}(u)^* \otimes I_m)(E_{\mathrm{n}}^* \otimes I_m)(\mathbf{1}_{n,1} \otimes v).$$

The L_2 metric is given by

$$d((u,v),(u',v')) = \sqrt{\mathbb{E}(q_{u,v}^* z - q_{u',v'}^* z)^2} = \|q_{u,v} - q_{u',v'}\|_2 .$$

Indeed,

$$\begin{aligned} & d((u,v),(u',v')) \\ & \le d((u,v),(u,v')) + d((u,v'),(u',v')) \\ & \le \|\mathrm{diag}(u-u')\|_\infty \|E_{\mathrm{n}}\| \|v\|_2 + \|\mathrm{diag}(u')\|_\infty \|E_{\mathrm{n}}\| \|v - v'\|_2 \\ & \le \|\mathrm{diag}(u-u')\|_\infty \|E_{\mathrm{n}}\| + \|E_{\mathrm{n}}\| \|v - v'\|_2 . \end{aligned}$$

Let $\Gamma_s = \{v \in B_2^{mN} : \ T \subset [m], \ |T| = s, \ (I_N \otimes \Pi_T) v = v\}$. By Dudley's theorem (see e.g., [127, Theorem 11.17]), we have

$$\begin{aligned} & \mathbb{E} \sup_{\substack{T \subset [m] \\ |T| = s}} \sup_{\substack{v \in B_2^{mN} \\ (I_N \otimes \Pi_T) v = v}} \sup_{u \in B_2^n} v^* \Psi u \\ & \le 24 \int_0^\infty \sqrt{\log N(\Gamma_s \times B_2^n, d(\cdot); \epsilon)} d\epsilon \\ & \le 24 \|E_{\mathrm{n}}\| \left(\int_0^\infty \sqrt{\log N(\Gamma_s, \|\cdot\|_2 ; \epsilon)} d\epsilon + \int_0^\infty \sqrt{\log N(B_2^n, \|\cdot\|_\infty ; \epsilon)} d\epsilon \right) \\ & \le 24 \|E_{\mathrm{n}}\| \left(\int_0^\infty \sqrt{\log N(\Gamma_s, \|\cdot\|_2 ; \epsilon)} d\epsilon + \int_0^\infty \sqrt{\log N(B_1^n, \|\cdot\|_2 ; \epsilon)} d\epsilon \right) \\ & \lesssim \|E_{\mathrm{n}}\| (\sqrt{s \log m} + \sqrt{Ns} + \log^{3/2} n). \end{aligned}$$

By an extension of Dudley's inequality to moments [131, Section 8.9, page

263],

$$\Big(\mathbb{E}\sup_{\substack{T\subset[m]\\|T|=s}}\sup_{\substack{v\in B_2^{mN}\\(I_N\otimes\Pi_T)v=v}}\sup_{u\in B_2^n}|v^*\Psi u|^p\Big)^{1/p}$$
$$\lesssim \|E_{\mathrm{n}}\|\,(\sqrt{s\log m}+\sqrt{Ns}+\log^{3/2}n)\sqrt{p}.$$

By a variation of Markov's inequality [131, Proposition 7.11], there exist absolute constants $C_4, c_4 > 0$ such that

$$\sup_{\substack{T\subset[m]\\|T|=s}}\sup_{\substack{v\in B_2^{mN}\\(I_N\otimes\Pi_T)v=v}}\sup_{u\in B_2^n}|v^*\Psi u|$$
$$\le C_4\,\|E_{\mathrm{n}}\|\,(\sqrt{s\log m}+\sqrt{Ns}+\log^{3/2}n)\sqrt{\log n},$$

with probability at least $1-n^{-c_4}$.

Therefore, Lemma 4.4.9 follows from

$$\|E_{\mathrm{n}}\| = \max_{k\in[n]}\|w_{k\cdot}\|_2 \le \sqrt{N}\max_{k\in[n],j\in[N]}|w_{kj}|.$$

□

Proof of Lemma 4.4.10. If assumptions 4.3.1 – 4.3.3 are satisfied, then by (B.10),

$$\|y_{k\cdot}\|_2 \le \frac{(C_2'+1)\sqrt{1+\delta}}{\sqrt{n}}\max\Big\{1,\|X\|\sqrt{\log n}\Big\}$$

for all $k\in[n]$, with probability at least $1-n^{-c_2}$.

Since

$$E_{\mathrm{s}}^*E_{\mathrm{n}} = \mathrm{diag}([y_{1\cdot}^*w_{1\cdot}, y_{2\cdot}^*w_{2\cdot},\dots,y_{n\cdot}^*w_{n\cdot}]),$$

there exist constants $C_5 = (C_2'+1)\sqrt{(1+\delta)(1+\theta)} > 0$ such that

$$\|E_{\mathrm{s}}^*E_{\mathrm{n}}\| \le \max_k\|y_{k\cdot}\|_2 \times \sqrt{N}\max_{k\in[n],j\in[N]}|w_{kj}|$$
$$\le \frac{C_5}{\sqrt{1+\theta}}\sqrt{\frac{N}{n}}\max\Big\{1,\|X\|\sqrt{\log n}\Big\}\max_{k\in[n],j\in[N]}|w_{kj}|,$$

with probability at least $1-n^{-c_2}$. Therefore, Lemma 4.4.10 follows from Assumption 4.3.3. □

Proof of Lemma 4.4.11. Lemma 4.4.11 follows from

$$E_{\mathrm{n}}^* E_{\mathrm{n}} = \mathrm{diag}([\|w_{1\cdot}\|_2^2, \|w_{2\cdot}\|_2^2, \ldots, \|w_{n\cdot}\|_2^2]).$$

□

B.4 Scalar Concentration Bounds

Proof of Lemma 4.4.12. We prove these inequalities using the Hoeffding's inequality.

For all $j \in [N]$, $\ell \in [m]$, and $k \in [n]$,

$$\begin{aligned}
&\left||\overline{a_{k\ell}} a_{k\cdot}^\top x_{\cdot j}|^2 - \mathbb{E}|\overline{a_{k\ell}} a_{k\cdot}^\top x_{\cdot j}|^2\right| \\
&\leq |a_{k\ell}|^2 |a_{k\cdot}^\top x_{\cdot j}|^2 + \frac{1}{n^2}(\|x_{\cdot j}\|_2^2 + |x_{\ell j}|^2) \\
&\leq C_6' \frac{\log(nm)}{n} \cdot \frac{\|x_{\cdot j}\|_2^2 \log(nN)}{n} + \frac{2\|x_{\cdot j}\|_2^2}{n^2} \\
&\leq \frac{(C_6'+2)\|x_{\cdot j}\|_2^2 \log^2(nmN)}{n^2},
\end{aligned}$$

where the third line is true with probability at least $1 - n^{-c_6'}$ for some absolute constant c_6'. We show this by applying a Chernoff bound and a union bound to $|a_{k\ell}|^2$, and applying the Hanson-Wright inequality (B.9) and a union bound to $|a_{k\cdot}^\top x_{\cdot j}|^2$. Then it follows from the Hoeffding's inequality and a union bound, that there exist absolute constants $C_6, c_6 > 0$ such that for all $j \in [N]$ and $\ell \in [m]$ we have (4.20).

Similarly, for all $j \in [N]$, $\ell \in [m]$, and $k \in [n]$,

$$|a_{k\ell}|^2 |a_{k\cdot}^\top x_{\cdot j}| \leq C_6' \frac{\log(nm)}{n} \cdot \frac{\|x_{\cdot j}\|_2 \sqrt{\log(nN)}}{\sqrt{n}},$$

with probability at least $1 - n^{-c_6'}$. By the Hoeffding's inequality and a union bound, we have (4.21). Here we use the following facts: By Assumption 4.3.3, $\|x_{\cdot j}\| \geq \sqrt{\frac{1-\theta}{N}}$, and by Assumption 4.3.4, $\max_{k\in[n], j\in[N]} |w_{kj}| \leq \frac{C_W}{\sqrt{nN}}$.

For $\ell \in [m]$ and $k \in [n]$,

$$|a_{k\ell}|^2 - \mathbb{E}|a_{k\ell}|^2 \leq C_6' \frac{\log(nm)}{n},$$

with probability at least $1 - n^{-c_6'}$. By the Hoeffding's inequality and a union bound, we have (4.22). $\square$

APPENDIX C

PROOFS FOR CHAPTER 5

C.1 Proofs for Section 5.2

C.1.1 Derivation of (5.3) and (5.4)

Recall that

$$\nabla_{L''}(h) = \frac{1}{N}\sum_{i=1}^{N} \nabla_i'',$$
$$H_{L''}(h) = \frac{1}{N}\sum_{i=1}^{N} H_i'',$$

where $\nabla_i'' \coloneqq C_{x_i}^\top \nabla_\phi(C_{x_i}h)$, and $H_i'' = C_{x_i}^\top H_\phi(C_{x_i}h)C_{x_i}$.

For the Bernoulli-Rademacher model in (A1), we have

$$\begin{aligned}
\mathbb{E}\nabla_{i(j)}'' &= -\mathbb{E}\sum_{s=1}^{n} x_{i(1+s-j)}\Big(\sum_{t=1}^{n} x_{i(1+s-t)}h_{(t)}\Big)^3 \\
&= -n\Big(\theta h_{(j)}^3 + 3\theta^2 h_{(j)}\sum_{\ell\neq j} h_{(\ell)}^2\Big) \\
&= -n\theta(1-3\theta)h_{(j)}^3 - 3n\theta^2 h_{(j)},
\end{aligned}$$

where the last line uses the fact that $\sum_{j=1}^n h_{(j)}^2 = \|h\| = 1$. Therefore, the gradient and the Riemannian gradient are

$$\begin{aligned}
&\mathbb{E}\nabla_{L''}(h) = -n\theta(1-3\theta)h^{\odot 3} - 3n\theta^2 h, \\
&\mathbb{E}\widehat{\nabla}_{L''}(h) = P_{h^\perp}\mathbb{E}\nabla_{L''}(h) = n\theta(1-3\theta)(\|h\|_4^4 \cdot h - h^{\odot 3}).
\end{aligned}$$

Similarly, we have

$$\begin{aligned}\mathbb{E}H''_{i(jk)} &= -3\mathbb{E}\sum_{s=1}^{n} x_{i(1+s-j)}x_{i(1+s-k)}\Big(\sum_{t=1}^{n} x_{i(1+s-t)}h_{(t)}\Big)^2 \\ &= -3n \times \begin{cases} \theta h_{(j)}^2 + \theta^2 \sum_{\ell\neq j} h_{(\ell)}^2 & \text{if } j = k \\ 2\theta^2 h_{(j)}h_{(k)} & \text{if } j \neq k \end{cases} \\ &= -3n\big[\theta^2\delta_{jk} + \theta(1-3\theta)h_{(j)}^2\delta_{jk} + 2\theta^2 h_{(j)}h_{(k)}\big].\end{aligned}$$

The Hessian and the Riemannian Hessian are

$$\begin{aligned}\mathbb{E}H_{L''}(h) &= -3n\big[\theta^2 I + \theta(1-3\theta)\mathrm{diag}(h^{\odot 2}) + 2\theta^2 hh^\top\big], \\ \mathbb{E}\widehat{H}_{L''}(h) &= P_{h^\perp}\mathbb{E}H_{L''}(h)P_{h^\perp} - h^\top \mathbb{E}\nabla_{L''}(h)\cdot P_{h^\perp} \\ &= n\theta(1-3\theta)\big[\|h\|_4^4 \cdot I + 2\|h\|_4^4 \cdot hh^\top - 3\cdot \mathrm{diag}(h^{\odot 2})\big].\end{aligned}$$

C.1.2 Proofs of Lemmas in Section 5.2

Proof of Lemma 5.2.3. We first investigate the Riemannian Hessian at points in $\mathcal{H}_1''$ and $\mathcal{H}_2''$. Without loss of generality, we consider points close to the representative stationary point $h_0 = [1/\sqrt{r}, \ldots, 1/\sqrt{r}, 0, \ldots, 0]$. We have

$$\begin{aligned}&|h_{(j)}^2 - 1/r| \le \rho/r, \qquad \forall j \in \{1, 2, \ldots, r\}, \\ &h_{(j)}^2 \le \rho/r, \qquad \forall j \in \{r+1, \ldots, n\}, \\ &\sum_{j=r+1}^{n} h_{(j)}^2 = 1 - \sum_{j=1}^{r} h_{(j)}^2 \le \rho.\end{aligned}$$

Therefore,

$$\|h - h_0\| \le \sqrt{r \times \big(\frac{1-\sqrt{1-\rho}}{\sqrt{r}}\big)^2 + \rho} \le \sqrt{2\rho}, \tag{C.1}$$

$$\left\|\mathrm{diag}(h^{\odot 2}) - \frac{1}{r}\begin{bmatrix} I_r & \\ & \mathbf{0}_{(n-r)\times(n-r)} \end{bmatrix}\right\| \le \frac{\rho}{r}, \tag{C.2}$$

and

$$\left\| hh^\top - \frac{1}{r} \begin{bmatrix} \mathbf{1}_{r\times r} & \\ & \mathbf{0}_{(n-r)\times(n-r)} \end{bmatrix} \right\| \le 2 \left\| h - h_0 \right\| \le 2\sqrt{2\rho}. \tag{C.3}$$

We also bound $\|h\|_4^4$ as follows:

$$\|h\|_4^4 \le r \times \frac{(1+\rho)^2}{r^2} + \min\Big\{ (n-r) \times \frac{\rho^2}{r^2},\ \frac{\rho^2}{n-r} \Big\} \le \frac{1 + 2\rho + 2\rho^2}{r},$$

$$\|h\|_4^4 \ge r \times \frac{(1-\rho)^2}{r^2} \ge \frac{1 - 2\rho + \rho^2}{r}.$$

Since $\rho < 10^{-3} < 1/2$,

$$\left| \|h\|_4^4 - \frac{1}{r} \right| \le \frac{3\rho}{r}. \tag{C.4}$$

Next we obtain bounds on the Riemannian curvature of $\mathbb{E}L''$ at points $h \in \mathcal{H}_1''$ or $h \in \mathcal{H}_2''$ by bounding its deviation from the Riemannian curvature at a corresponding stationary point h_0. By (C.2), (C.3), (C.4), and the expressions in (5.4), (5.6):

$$\begin{aligned} &\left\| \mathbb{E}\widehat{H}_{L''}(h) - \mathbb{E}\widehat{H}_{L''}(h_0) \right\| \\ &\le n\theta(1-3\theta)\Big[\frac{3\rho}{r} + 2 \times \frac{3\rho + 2\sqrt{2\rho}}{r} + 3 \times \frac{\rho}{r} \Big] \\ &= \frac{n\theta(1-3\theta)}{r}(12\rho + 4\sqrt{2\rho}). \end{aligned} \tag{C.5}$$

It follows that

$$
\begin{aligned}
&\Big|\min_{\substack{z:\|z\|=1\\ z\perp h}} z^\top \mathbb{E}\widehat{H}_{L''}(h)z - \min_{\substack{z:\|z\|=1\\ z\perp h_0}} z^\top \mathbb{E}\widehat{H}_{L''}(h_0)z\Big| \\
&\le \Big|\min_{\substack{z:\|z\|=1\\ z\perp h}} z^\top \mathbb{E}\widehat{H}_{L''}(h)z - \min_{\substack{z:\|z\|=1\\ z\perp h}} z^\top \mathbb{E}\widehat{H}_{L''}(h_0)z\Big| \\
&\quad + \Big|\min_{\substack{z:\|z\|=1\\ z\perp h}} z^\top \mathbb{E}\widehat{H}_{L''}(h_0)z - \min_{\substack{z:\|z\|=1\\ z\perp h_0}} z^\top \mathbb{E}\widehat{H}_{L''}(h_0)z\Big| \\
&\le \Big\|V^\top \mathbb{E}\widehat{H}_{L''}(h)V - V^\top \mathbb{E}\widehat{H}_{L''}(h_0)V\Big\| + \Big\|V^\top \mathbb{E}\widehat{H}_{L''}(h_0)V - V_0^\top \mathbb{E}\widehat{H}_{L''}(h_0)V_0\Big\| \\
&\le \Big\|\mathbb{E}\widehat{H}_{L''}(h) - \mathbb{E}\widehat{H}_{L''}(h_0)\Big\| + 2\Big\|\mathbb{E}\widehat{H}_{L''}(h_0)\Big\| \cdot \|V - V_0\| \\
&\le \frac{n\theta(1-3\theta)}{r}(12\rho + 4\sqrt{2\rho}) + 2 \times \frac{2n\theta(1-3\theta)}{r} \times \sqrt{2\rho} \\
&= \frac{n\theta(1-3\theta)}{r}(12\rho + 8\sqrt{2\rho}) \\
&\le \frac{n\theta(1-3\theta)(24\sqrt{\rho})}{r}, \qquad \text{(C.6)}
\end{aligned}
$$

where $V, V_0 \in \mathbb{R}^{n\times(n-1)}$ satisfy: (I) the columns of V (resp. V_0) form an orthonormal basis for the tangent space at h (resp. h_0); (II) $\|V - V_0\| \le \sqrt{2\rho}$. We construct V and V_0 as follows, for the non-trivial case where $h \ne h_0$. Suppose the columns of $V_\cap \in \mathbb{R}^{n\times(n-2)}$ form an orthonormal basis for the intersection of the tangent spaces at h and at h_0. Let $c \coloneqq \langle h, h_0\rangle < 1$, and let $h' \coloneqq \frac{1}{\sqrt{1-c^2}}(h_0 - ch)$ and $h_0' \coloneqq \frac{1}{\sqrt{1-c^2}}(ch_0 - h)$. It is easy to verify that $V \coloneqq [V_\cap, h']$ and $V_0 \coloneqq [V_\cap, h_0']$ satisfy (I). To verify (II), we have $\|V - V_0\| = \|h' - h_0'\| = \frac{1-c}{\sqrt{1-c^2}}\|h + h_0\| = \|h - h_0\| \le \sqrt{2\rho}$.

Positive definiteness (5.9) follows from (5.7) and (C.6). Negative curvature (5.10) follows from (5.8) and (C.6).

Next, we prove contrapositive of (5.11), i.e., suppose

$$\Big\|\mathbb{E}\widehat{\nabla}_{L''}(h)\Big\| < \theta(1-3\theta)\rho^2/n$$

for some $h \in S^{n-1}$, then we show $h \in \mathcal{H}_1'' \cup \mathcal{H}_2''$. First, it follows from $\Big\|\mathbb{E}\widehat{\nabla}_{L''}(h)\Big\| < \theta(1-3\theta)\rho^2/n$, and the expression in (5.5), that for all $j, k \in [n]$,

$$h_{(j)}^2 h_{(k)}^2 (h_{(j)}^2 - h_{(k)}^2)^2 < \frac{\rho^4}{n^4}.$$

As a result, $|h_{(j)}^2 - h_{(k)}^2| < \rho/n$ if $h_{(j)}^2 \ge \rho/n$ and $h_{(k)}^2 \ge \rho/n$.

Let $\Omega \coloneqq \{j : h_{(j)}^2 \geq \rho/n\} \subset [n]$, and $r \coloneqq |\Omega|$. Then

$$h_{(j)}^2 < \rho/n \leq \rho/r, \qquad \forall j \notin \Omega, \tag{C.7}$$

and

$$1 - (n-r) \cdot \frac{\rho}{n} < \sum_{j \in \Omega} h_{(j)}^2 \leq 1. \tag{C.8}$$

In addition, $|h_{(j)}^2 - h_{(k)}^2| < \rho/n$ for $j, k \in \Omega$. Therefore, for $k \in \Omega$, $h_{(k)}^2$ is close to the average $\frac{1}{r}\sum_{j\in\Omega} h_{(j)}^2$:

$$\left| h_{(k)}^2 - \frac{1}{r}\sum_{j \in \Omega} h_{(j)}^2 \right| < \rho/n, \qquad \forall k \in \Omega. \tag{C.9}$$

By (C.8) and (C.9), for $k \in \Omega$:

$$h_{(k)}^2 \leq \frac{1}{r} + \frac{\rho}{n} \leq \frac{1+\rho}{r},$$

$$h_{(k)}^2 \geq \frac{1 - (n-r)\cdot\frac{\rho}{n}}{r} - \frac{\rho}{n} = \frac{1-\rho}{r}.$$

Therefore,

$$\left| h_{(k)}^2 - \frac{1}{r} \right| \leq \frac{\rho}{r}, \qquad \forall k \in \Omega. \tag{C.10}$$

It follows from (C.7) and (C.10) that h is in the (ρ, r)-neighborhood of a stationary point h_0, where $h_{0(j)} = 1/\sqrt{r}$ if $j \in \Omega$ and $h_{0(j)} = 0$ if $j \notin \Omega$. Clearly, such an h belongs to $\mathcal{H}_1'' \cup \mathcal{H}_2''$. By contraposition, any point $h \in \mathcal{H}_3'' = S^{n-1} \backslash (\mathcal{H}_1'' \cup \mathcal{H}_2'')$ satisfies (5.11). □

Proof of Lemma 5.2.4. For any given $h \in S^{n-1}$ one can bound the deviation of the gradient (or Hessian) from its mean using matrix Bernstein inequality [115]. Let S_ϵ be an ϵ-net of S^{n-1}. Then $|S_\epsilon| \leq (3/\epsilon)^n$ [127, Lemma 9.5]. We can then bound the deviation over S^{n-1} by a union bound over S_ϵ.

Define $\nabla_i'' \coloneqq C_{x_i}^\top \nabla_\phi(C_{x_i} h)$, and $H_i'' = C_{x_i}^\top H_\phi(C_{x_i} h) C_{x_i}$. For the Bernoulli-

Rademacher model in (A1), we have $|x_{i(j)}| \le 1$. Therefore,

$$\begin{aligned}\big|\nabla''_{i(j)}\big| &= \Big|\sum_{s=1}^{n} x_{i(1+s-j)}\Big(\sum_{t=1}^{n} x_{i(1+s-t)}h_{(t)}\Big)^3\Big| \\ &\le n\Big(\sum_{t=1}^{n}|h_{(t)}|\Big)^3 \\ &\le n^2\sqrt{n},\end{aligned}$$

$$\begin{aligned}\big|H''_{i(jk)}\big| &= \Big|3\sum_{s=1}^{n} x_{i(1+s-j)}x_{i(1+s-k)}\Big(\sum_{t=1}^{n} x_{i(1+s-t)}h_{(t)}\Big)^2\Big| \\ &\le 3n\Big(\sum_{t=1}^{n}|h_{(t)}|\Big)^2 \\ &\le 3n^2.\end{aligned}$$

It follows that $\|\nabla''_i\| \le n^3$, and $\|H''_i\| \le \|H''_i\|_{\mathrm{F}} \le 3n^3$.

Our goal is to bound the following average of independent random terms with zero mean:

$$\nabla_{L''}(h) - \mathbb{E}\nabla_{L''}(h) = \frac{1}{N}\sum_{i=1}^{N}\big(\nabla''_i - \mathbb{E}\nabla''_i\big).$$

$$H_{L''}(h) - \mathbb{E}H_{L''}(h) = \frac{1}{N}\sum_{i=1}^{N}\big(H''_i - \mathbb{E}H''_i\big).$$

Since $\|\nabla''_i\| \le n^3$, we have

$$\begin{aligned}&\|\nabla''_i - \mathbb{E}\nabla''_i\| \le 2n^3, \\ &\sum_{i=1}^{N}\mathbb{E}\,\|\nabla''_i - \mathbb{E}\nabla''_i\|^2 = N(\mathbb{E}\,\|\nabla''_i\|^2 - \|\mathbb{E}\nabla''_i\|^2) \le Nn^6, \\ &\Big\|\sum_{i=1}^{N}\mathbb{E}(\nabla''_i - \mathbb{E}\nabla''_i)(\nabla''_i - \mathbb{E}\nabla''_i)^\top\Big\| \le N(\mathbb{E}\,\|\nabla''_i\|^2 + \|\mathbb{E}\nabla''_i\|^2) \le 2Nn^6.\end{aligned}$$

By the rectangular version of the matrix Bernstein inequality [115, Theorem

1.6], and a union bound over S_ϵ,

$$\begin{aligned}&\mathbb{P}\Big[\sup_{h\in S_\epsilon}\|\nabla_{L''}(h)-\mathbb{E}\nabla_{L''}(h)\|\le\tau\Big]\\&\ge 1-\Big(\frac{3}{\epsilon}\Big)^n(n+1)\exp\Big(\frac{-N^2\tau^2/2}{2Nn^6+2n^3N\tau/3}\Big).\end{aligned}\tag{C.11}$$

Similarly, since $\|H_i''\|\le 3n^3$, we have

$$\begin{aligned}&\|H_i''-\mathbb{E}H_i''\|\le 6n^3,\\&\Big\|\sum_{i=1}^N\mathbb{E}(H_i''-\mathbb{E}H_i'')^2\Big\|\le N\left\|\mathbb{E}H_i''^2-(\mathbb{E}H_i'')^2\right\|\le 2N(3n^3)^2=18Nn^6.\end{aligned}$$

By the symmetric version of the matrix Bernstein inequality [115, Theorem 1.4], and a union bound over S_ϵ,

$$\begin{aligned}&\mathbb{P}\Big[\sup_{h\in S_\epsilon}\|H_{L''}(h)-\mathbb{E}H_{L''}(h)\|\le\tau\Big]\\&\ge 1-\Big(\frac{3}{\epsilon}\Big)^n(2n)\exp\Big(\frac{-N^2\tau^2/2}{18Nn^6+6n^3N\tau/3}\Big).\end{aligned}\tag{C.12}$$

Choose $\tau=\frac{\theta(1-3\theta)\rho^2}{8n}$, and $\epsilon=\frac{\tau}{6n^3}=\frac{\theta(1-3\theta)\rho^2}{48n^4}$. By (C.11) and (C.12), there exist constants c_2, $c_2'>0$ (depending only on θ), such that: if $N>\frac{c_2n^9}{\rho^4}\log\frac{n}{\rho}$, then with probability at least $1-e^{-c_2'n}$,

$$\sup_{h\in S_\epsilon}\|\nabla_{L''}(h)-\mathbb{E}\nabla_{L''}(h)\|\le\tau=\frac{\theta(1-3\theta)\rho^2}{8n},$$

$$\sup_{h\in S_\epsilon}\|H_{L''}(h)-\mathbb{E}H_{L''}(h)\|\le\tau=\frac{\theta(1-3\theta)\rho^2}{8n}.$$

To finish the proof, we extrapolate the concentration bounds over S_ϵ to all points in S^{n-1}. For any $h\in S^{n-1}$, there exists $h'\in S_\epsilon$ such that $\|h-h'\|\le\epsilon$. Furthermore, thanks to the Lipschitz continuity of the gradient and the Hessian,

$$\begin{aligned}&\|\nabla_i''(h)-\nabla_i''(h')\|\\&\le\|C_{x_i}\|\cdot\sqrt{n}(3\|x_i\|^2)\cdot\|x_i\|\,\|h-h'\|\\&\le 3n^3\epsilon,\end{aligned}$$

$$
\begin{aligned}
&\|H_i''(h) - H_i''(h')\| \\
&\leq \|C_{x_i}\|^2 \cdot (6\|x_i\|) \cdot \|x_i\| \, \|h - h'\| \\
&\leq 6n^3\epsilon,
\end{aligned}
$$

where $3\|x_i\|^2$ and $6\|x_i\|$ are the Lipschitz constants of $(\cdot)^3$ and $3(\cdot)^2$ on the interval $[-\|x_i\|, \|x_i\|]$. We also use the fact that $|x_{i(j)}| < 1$, hence $\|x_i\| \leq \sqrt{n}$ and $\|C_{x_i}\| \leq n$. As a consequence,

$$
\begin{aligned}
&\sup_{h \in S^{n-1}} \|\nabla_{L''}(h) - \mathbb{E}\nabla_{L''}(h)\| \\
&\leq \sup_{h \in S_\epsilon} \|\nabla_{L''}(h) - \mathbb{E}\nabla_{L''}(h)\| + 2 \max_{i \in [n]} \sup_{\|h - h'\| \leq \epsilon} \|\nabla_i''(h) - \nabla_i''(h')\| \\
&\leq \tau + 6n^3\epsilon = 2\tau = \frac{\theta(1 - 3\theta)\rho^2}{4n},
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{h \in S^{n-1}} \left\| \widehat{\nabla}_{L''}(h) - \mathbb{E}\widehat{\nabla}_{L''}(h) \right\| \\
&\leq \sup_{h \in S^{n-1}} \|\nabla_{L''}(h) - \mathbb{E}\nabla_{L''}(h)\| \\
&\leq \frac{\theta(1 - 3\theta)\rho^2}{4n}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\sup_{h \in S^{n-1}} \|H_{L''}(h) - \mathbb{E}H_{L''}(h)\| \\
&\leq \sup_{h \in S_\epsilon} \|H_{L''}(h) - \mathbb{E}H_{L''}(h)\| + 2 \max_{i \in [n]} \sup_{\|h - h'\| \leq \epsilon} \|H_i''(h) - H_i''(h')\| \\
&\leq \tau + 12n^3\epsilon = 3\tau = \frac{3\theta(1 - 3\theta)\rho^2}{8n},
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{h \in S^{n-1}} \left\| \widehat{H}_{L''}(h) - \mathbb{E}\widehat{H}_{L''}(h) \right\| \\
&\leq \sup_{h \in S^{n-1}} \|H_{L''}(h) - \mathbb{E}H_{L''}(h)\| + \sup_{h \in S^{n-1}} \|\nabla_{L''}(h) - \mathbb{E}\nabla_{L''}(h)\| \\
&\leq \frac{\theta(1 - 3\theta)\rho^2}{n}.
\end{aligned}
$$

□

Proof of Lemma 5.2.5. We have $\mathbb{E}\frac{1}{\theta n N}\sum_{i=1}^N C_{x_i}^\top C_{x_i} = I$. We first bound

$\left\|\frac{1}{\theta nN}\sum_{i=1}^{N} C_{x_i}^\top C_{x_i} - I\right\|$ using the matrix Bernstein inequality. To this end, we bound the spectral norm of $\mathbb{E}(C_{x_i}^\top C_{x_i})^2$, the eigenvalues of which can be computed using the DFT of x_i. The eigenvalue corresponding to the t-th frequency satisfies

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\sum_{k=1}^{n} e^{-\sqrt{-1}(k-1)t/n} x_{i(k)}\Big)\Big(\sum_{k=1}^{n} e^{\sqrt{-1}(k-1)t/n} x_{i(k)}\Big)\Big]^2 \\
&= \mathbb{E}\Big(\sum_{k=1}^{n} x_{i(k)}^2 + \sum_{1\le k<j\le n} 2\cos((j-k)t/n) x_{i(j)} x_{i(k)}\Big)^2 \\
&\le n\theta + \frac{n(n-1)}{2}\times 4\theta^2 + n(n-1)\theta^2 \\
&= n\theta + 3n(n-1)\theta^2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\Big\|\sum_{i=1}^{N} \mathbb{E}\Big(\frac{1}{\theta n} C_{x_i}^\top C_{x_i} - I\Big)^2\Big\| \\
&= N\Big\|\frac{1}{\theta^2 n^2}\mathbb{E}(C_{x_i}^\top C_{x_i})^2 - I\Big\| \\
&\le \frac{N}{\theta^2 n^2}\left\|\mathbb{E}(C_{x_i}^\top C_{x_i})^2\right\| + N \\
&\le \frac{N}{\theta^2 n^2}(n\theta + 3n(n-1)\theta^2) + N \\
&\le \frac{N}{\theta n} + 3N + N \\
&\le 5N.
\end{aligned}$$

We also have

$$\left\|\frac{1}{\theta n} C_{x_i}^\top C_{x_i} - I\right\| \le \frac{1}{\theta n}\left\|C_{x_i}\right\|^2 + 1 \le \frac{n^2}{\theta n} + 1 \le n^2 + 1.$$

By the matrix Bernstein inequality [115, Theorem 1.4],

$$\mathbb{P}\Big[\Big\|\frac{1}{\theta nN}\sum_{i=1}^{N} C_{x_i}^\top C_{x_i} - I\Big\| \le \tau\Big] \ge 1 - 2n\exp\Big(\frac{-N^2\tau^2/2}{5N + (n^2+1)N\tau/3}\Big).$$

Set $\tau = \frac{\theta(1-3\theta)\rho^2}{200n^4\kappa^4}$. Then there exist constants c_3, $c_3' > 0$ (depending only on

θ) such that: if $N > \frac{c_3 n^8 \kappa^8}{\rho^4} \log n$, then with probability at least $1 - n^{-c_3'}$,

$$\Big\| \frac{1}{\theta n N} \sum_{i=1}^{N} C_{x_i}^\top C_{x_i} - I \Big\| \le \frac{\theta(1-3\theta)\rho^2}{200 n^4 \kappa^4}. \tag{C.13}$$

Next, we bound $\big\| C_f R - C_f (C_f^\top C_f)^{-1/2} \big\|$ (similar to the proofs of [71, Lemma 15] and [132, Lemma B.2]). Define $Q \coloneqq \frac{1}{\theta n N} \sum_{i=1}^{N} C_{x_i}^\top C_{x_i}$. Then

$$\begin{aligned}
&\big\| C_f R - C_f (C_f^\top C_f)^{-1/2} \big\| \\
&= \big\| C_f (C_f^\top Q C_f)^{-1/2} - C_f (C_f^\top C_f)^{-1/2} \big\| \\
&\le \sigma_1(C_f) \cdot \big\| (C_f^\top Q C_f)^{-1/2} - (C_f^\top C_f)^{-1/2} \big\| \\
&\le \sigma_1(C_f) \frac{\big\| (C_f^\top Q C_f)^{-1} - (C_f^\top C_f)^{-1} \big\|}{\sigma_n\big((C_f^\top C_f)^{-1/2}\big)} \qquad \text{(C.14)} \\
&= \sigma_1^2(C_f) \big\| (C_f^\top Q C_f)^{-1} - (C_f^\top C_f)^{-1} \big\| \\
&\le \frac{\sigma_1^2(C_f)}{\sigma_n^2(C_f)} \big\| (C_f^\top C_f)(C_f^\top Q C_f)^{-1} - I \big\| \\
&= \kappa^2 \Big\| \Big[I + (C_f^\top (Q-I) C_f)(C_f^\top C_f)^{-1} \Big]^{-1} - I \Big\| \\
&\le \kappa^2 \frac{\big\| C_f^\top (Q-I) C_f \big\| \, \big\| (C_f^\top C_f)^{-1} \big\|}{1 - \big\| C_f^\top (Q-I) C_f \big\| \, \big\| (C_f^\top C_f)^{-1} \big\|} \qquad \text{(C.15)} \\
&\le \kappa^4 \frac{\|Q-I\|}{1-1/2} \le \frac{\theta(1-3\theta)\rho^2}{100 n^4}. \qquad \text{(C.16)}
\end{aligned}$$

The inequality (C.14) follows from the fact ([133, Theorem 6.2]) that, for positive definite A and B,

$$\big\| A^{-1/2} - B^{-1/2} \big\| \le \frac{\|A^{-1} - B^{-1}\|}{\sigma_n(A^{-1/2} + B^{-1/2})} \le \frac{\|A^{-1} - B^{-1}\|}{\sigma_n(B^{-1/2})},$$

which in turn follows from the identity

$$(A^{-1/2} - B^{-1/2})(A^{-1/2} + B^{-1/2}) + (A^{-1/2} + B^{-1/2})(A^{-1/2} - B^{-1/2}) = 2(A^{-1} - B^{-1}).$$

The inequality (C.15) is due to the fact that $\|(I+A)^{-1} - I\| \le \|(I+A)^{-1}\| \, \|A\| \le \frac{\|A\|}{1 - \|A\|}$ for $\|A\| < 1$. The last line (C.16) follows from (C.13) and

$$\big\| C_f^\top (Q-I) C_f \big\| \, \big\| (C_f^\top C_f)^{-1} \big\| \le \kappa^2 \, \|Q - I\| < \frac{1}{2}.$$

The rest of Lemma 5.2.5 follows from the Lipschitz continuity of the objective function. Define $U \coloneqq C_f R$, and $U' \coloneqq C_f (C_f^\top C_f)^{-1/2}$, which is an orthogonal matrix. We have

$$\|C_f R\| = \|U\| \le \|U'\| + \|U - U'\| < 2. \tag{C.17}$$

Recall that for the Bernoulli-Rademacher model, $\|x_i\| \le \sqrt{n}$ and $\|C_{x_i}\| \le n$. Then the difference of the gradients of $L(h) = \frac{1}{N}\sum_{i=1}^N \phi(C_{x_i} U h)$ and $L'(h) = \frac{1}{N}\sum_{i=1}^N \phi(C_{x_i} U' h)$ can be bounded as follows:

$$\begin{aligned}
&\|\nabla_L(h) - \nabla_{L'}(h)\| \\
&\le \max_{i\in[n]} \left\| U^\top C_{x_i}^\top \nabla_\phi(C_{x_i} U h) - U'^\top C_{x_i}^\top \nabla_\phi(C_{x_i} U' h) \right\| \\
&\le \max_{i\in[n]} \left\| U^\top C_{x_i}^\top \nabla_\phi(C_{x_i} U h) - U^\top C_{x_i}^\top \nabla_\phi(C_{x_i} U' h) \right\| \\
&\quad + \max_{i\in[n]} \left\| U^\top C_{x_i}^\top \nabla_\phi(C_{x_i} U' h) - U'^\top C_{x_i}^\top \nabla_\phi(C_{x_i} U' h) \right\| \\
&\le \max_{i\in[n]} \|U\| \, \|C_{x_i}\| \cdot \sqrt{n}[3(\|U\| \, \|x_i\|)^2] \cdot \|U - U'\| \, \|x_i\| \\
&\quad + \max_{i\in[n]} \|U - U'\| \, \|C_{x_i}\| \cdot \sqrt{n} \, \|x_i\|^3 \\
&\le 25\sqrt{n} \cdot \max_{i\in[n]} \|C_{x_i}\| \, \|x_i\|^3 \, \|U - U'\| \\
&\le 25 n^3 \, \|U - U'\| \, ,
\end{aligned}$$

where the third inequality follows from the fact that $\nabla_\phi(\cdot)$ is Lipschitz continous and bounded on compact sets – the Lipschitz constant of $(\cdot)^3$ on the interval $[-\|U\| \, \|x_i\|)^2, \|U\| \, \|x_i\|)^2]$ is $3(\|U\| \, \|x_i\|)^2$, and the upper bound of $|(\cdot)^3|$ on the interval $[-\|x_i\|, \|x_i\|]$ is $\|x_i\|^3$. Similarly the difference of the

Hessians can be bounded as follows:

$$
\begin{aligned}
&\|H_L(h) - H_{L'}(h)\| \\
&\leq \max_{i\in[n]} \left\|U^\top C_{x_i}^\top H_\phi(C_{x_i}Uh)C_{x_i}U - U'^\top C_{x_i}^\top H_\phi(C_{x_i}U'h)C_{x_i}U'\right\| \\
&\leq \max_{i\in[n]} \left\|U^\top C_{x_i}^\top H_\phi(C_{x_i}Uh)C_{x_i}U - U^\top C_{x_i}^\top H_\phi(C_{x_i}U'h)C_{x_i}U\right\| \\
&\quad + \max_{i\in[n]} \left\|U^\top C_{x_i}^\top H_\phi(C_{x_i}U'h)C_{x_i}U - U'^\top C_{x_i}^\top H_\phi(C_{x_i}U'h)C_{x_i}U\right\| \\
&\quad + \max_{i\in[n]} \left\|U'^\top C_{x_i}^\top H_\phi(C_{x_i}U'h)C_{x_i}U - U'^\top C_{x_i}^\top H_\phi(C_{x_i}U'h)C_{x_i}U'\right\| \\
&\leq \max_{i\in[n]} \|U\|^2 \|C_{x_i}\|^2 \cdot [6(\|U\| \|x_i\|)] \cdot \|U - U'\| \|x_i\| \\
&\quad + \max_{i\in[n]} \|U - U'\| \|U\| \|C_{x_i}\|^2 \cdot [3\|x_i\|^2] \\
&\quad + \max_{i\in[n]} \|U - U'\| \|C_{x_i}\|^2 \cdot [3\|x_i\|^2] \\
&\leq 57 \cdot \max_{i\in[n]} \|C_{x_i}\|^2 \|x_i\|^2 \|U - U'\| \\
&\leq 57n^3 \|U - U'\|,
\end{aligned}
$$

where the third inequality uses the Lipschitz constant and upper bound of $3(\cdot)^2$.

It follows from (C.16) and the above bounds that

$$
\begin{aligned}
&\sup_{h\in S^{n-1}} \left\|\widehat{\nabla}_L(h) - \widehat{\nabla}_{L'}(h)\right\| \\
&\leq \sup_{h\in S^{n-1}} \|\nabla_L(h) - \nabla_{L'}(h)\| \\
&\leq 25n^3 \|U - U'\| \leq \frac{\theta(1-3\theta)\rho^2}{4n}.
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{h\in S^{n-1}} \left\|\widehat{H}_L(h) - \widehat{H}_{L'}(h)\right\| \\
&\leq \sup_{h\in S^{n-1}} \|H_L(h) - H_{L'}(h)\| + \sup_{h\in S^{n-1}} \|\nabla_L(h) - \nabla_{L'}(h)\| \\
&\leq 100n^3 \|U - U'\| \leq \frac{\theta(1-3\theta)\rho^2}{n}.
\end{aligned}
$$

□

Proof of Lemma 5.2.6. The set $\mathcal{H}_1''$ equals the union of $(\rho, 1)$-neighborhoods

of $\{\pm e_j\}_{j=1}^n$, and the columns of $C_f^{-1} = C_g$ are the shifted versions of the inverse filter g. Therefore, by (C.1), every point $h^* \in (C_f^\top C_f)^{1/2} C_f^{-1} \mathcal{H}_1''$ satisfies

$$\left\| C_f (C_f^\top C_f)^{-1/2} h^* \pm e_j \right\| \le \sqrt{2\rho},$$

for some $j \in [n]$. It follows that

$$\begin{aligned}
&\|C_f R h^* \pm e_j\| \\
&\le \left\| C_f R h^* - C_f (C_f^\top C_f)^{-1/2} h^* \right\| + \left\| C_f (C_f^\top C_f)^{-1/2} h^* \pm e_j \right\| \\
&\le \frac{\theta(1-3\theta)\rho^2}{100n^4} + \sqrt{2\rho} \\
&\le 2\sqrt{\rho},
\end{aligned}$$

where the second to last line follows from (C.16), and the last line follows from $\theta(1-3\theta)\rho^2/(100n^4) < (2-\sqrt{2})\sqrt{\rho}$. □

C.2 Proofs for Section 5.3

Proof of Lemma 5.3.3. Clearly, $L(h) \le 0$ for all $h \in S^{n-1}$. For the Bernoulli-Rademacher model in (A1), we have $\|x_i\| \le \sqrt{n}$ and $\|C_{x_i}\| \le n$. Therefore,

$$\begin{aligned}
\phi(C_{y_i} R h) &= -\frac{1}{4} \|C_{x_i} C_f R h\|_4^4 \\
&\ge -\frac{n}{4} (\|x_i\| \, \|C_f R h\|)^4 \\
&\ge -4n^3,
\end{aligned}$$

where the first inequality follows from the Cauchy-Schwarz inequality, and the second inequality follows from $\|C_f R h\| \le \|C_f R\| \le 2$ (see (C.17)). Then $L(h) = \frac{1}{N} \sum_{i=1}^N L_i \ge -4n^3$.

We can bound the the norm of $\nabla_L(h)$ and $H_L(h)$ similarly. To bound $\|\nabla_L(h)\|$, we observe that

$$\begin{aligned}
\left| (C_{x_i}^\top \nabla_\phi(C_{y_i} R h))_{(j)} \right| &\le \|x_i\| \, \|\nabla_\phi(C_{y_i} R h)\| \\
&\le \|x_i\| \times \sqrt{n} (\|x_i\| \, \|C_f R h\|)^3 \\
&\le \sqrt{n} \, \|x_i\|^4 \, \|C_f R\|^3 \\
&\le 8n^2 \sqrt{n},
\end{aligned}$$

and hence

$$\begin{aligned}\|\nabla_L(h)\| &= \left\| \frac{1}{N} \sum_{i=1}^{N} R^\top C_{y_i}^\top \nabla_\phi(C_{y_i} R h) \right\| \\ &\le \left\| R^\top C_f^\top \right\| \left\| \frac{1}{N} \sum_{i=1}^{N} C_{x_i}^\top \nabla_\phi(C_{y_i} R h) \right\| \\ &\le \left\| R^\top C_f^\top \right\| \times \sqrt{n} \max_{i \in [N],\, j \in [n]} \left| (C_{x_i}^\top \nabla_\phi(C_{y_i} R h))_{(j)} \right| \\ &\le 16 n^3.\end{aligned}$$

To bound $\|H_L(h)\|$, we have

$$\begin{aligned}\left| (C_{x_i}^\top H_\phi(C_{y_i} R h) C_{x_i})_{(jk)} \right| &\le \|x_i\|^2 \, \|H_\phi(C_{y_i} R h)\| \\ &\le \|x_i\|^2 \times 3(\|x_i\| \, \|C_f R h\|)^2 \\ &\le 3 \|x_i\|^4 \, \|C_f R\|^2 \\ &\le 12 n^2,\end{aligned}$$

and hence

$$\begin{aligned}\|H_L(h)\| &= \left\| \frac{1}{N} \sum_{i=1}^{N} R^\top C_{y_i}^\top H_\phi(C_{y_i} R h) C_{y_i} R \right\| \\ &\le \left\| R^\top C_f^\top \right\| \left\| \frac{1}{N} \sum_{i=1}^{N} C_{x_i}^\top H_\phi(C_{y_i} R h) C_{x_i} \right\| \|C_f R\| \\ &\le \|C_f R\|^2 \times n \max_{i \in [N],\, j \in [n],\, k \in [n]} \left| (C_{x_i}^\top H_\phi(C_{y_i} R h) C_{x_i})_{(jk)} \right| \\ &\le 48 n^3.\end{aligned}$$

□

Proof of Lemma 5.3.4. For $z \perp h$, and $h' = \frac{h+z}{\|h+z\|} = \frac{h+z}{\sqrt{1+\|z\|^2}}$, $\widehat{L}(z) = L(h')$, $\widehat{L}(0) = L(h)$, and $\nabla_{\widehat{L}}(0) = \widehat{\nabla}_L(h)$. By the mean value theorem, there exists a convex combination h'' of h and h' such that $L(h') - L(h) = \langle h' - h, \nabla_L(h'') \rangle$, and a convex combination of h''' of h and h'' such that $\nabla_L(h'') - \nabla_L(h) =$

$H_L(h''')(h''-h)$. It follows that

$$\begin{aligned}
&\left|L(h')-L(h)-\langle z,\widehat{\nabla}_L(h)\rangle\right| \\
&=\left|\langle h'-h,\nabla_L(h'')\rangle-\langle z,\nabla_L(h)\rangle\right| \\
&\le\left|\langle h'-h-z,\nabla_L(h'')\rangle\right|+\left|\langle z,\nabla_L(h'')-\nabla_L(h)\rangle\right| \\
&\le\frac{\|z\|^2}{1+\sqrt{1+\|z\|^2}}\|\nabla_L(h'')\|+\|z\|\,\|H_L(h''')\|\,\|h''-h\| \\
&\le\frac{\|z\|^2}{2}\times 16n^3+48n^3\,\|z\|\,\|h-h'\| \\
&\le 64n^3\,\|z\|^2,
\end{aligned}$$

where the third inequality follows from Lemma 5.3.3, and the last inequality follows from the fact that $\|h-h'\|\le\|z\|$. □

Proof of Lemma 5.3.5. Suppose the columns of matrix $V\in\mathbb{R}^{n\times(n-1)}$ (resp. $V'\in\mathbb{R}^{n\times(n-1)}$) form a orthonormal basis for the tangent subspace at h (resp. h'). Then a matrix representation of $D\mathcal{A}(h)$ in (5.15) as a mapping from the tangent space of h to the tangent space at h' with respect to the bases of these spaces is $V'^{\top}V(I_{n-1}-\gamma V^{\top}\widehat{H}_L(h)V)$.

Note that $|\det(V'^{\top}V)|$ does not depend on the specific choice of orthogonal bases V and V' (multiplication by an orthonormal matrix does not change $|\det(\cdot)|$). Therefore, we consider the following construction of V and V'. Suppose the columns of $V_\cap\in\mathbb{R}^{n\times(n-2)}$ form an orthonormal basis for the intersection of the tangent spaces at h and at h'. Let $c\coloneqq\langle h,h'\rangle<1$, then it is easy to verify that $V\coloneqq[V_\cap,\,\frac{1}{\sqrt{1-c^2}}(h'-ch)]$ and $V'\coloneqq[V_\cap,\,\frac{1}{\sqrt{1-c^2}}(ch'-h)]$ are valid orthonormal bases. It follows that

$$|\det(V'^{\top}V)|=\begin{vmatrix} I_{n-2} & \mathbf{0}_{(n-2)\times 1}\\ \mathbf{0}_{1\times(n-2)} & c\end{vmatrix}=|c|.$$

Since $\langle h,h'\rangle=\langle h,\,h-\gamma\widehat{\nabla}_L(h)\rangle/\left\|h-\gamma\widehat{\nabla}_L(h)\right\|=\|h\|^2/\left\|h-\gamma\widehat{\nabla}_L(h)\right\|=1/\left\|h-\gamma\widehat{\nabla}_L(h)\right\|>0$, we have $|\det(V'^{\top}V)|=|\langle h,h'\rangle|>0$.

By Lemma 5.3.3, for all $h\in S^{n-1}$,

$$\begin{aligned}
\left\|\widehat{H}_L(h)\right\|&\le\|H_L(h)\|+\|\nabla_L(h)\|\\
&\le 48n^3+16n^3=64n^3.
\end{aligned}$$

Therefore $I_{n-1} - \gamma V^\top \widehat{H}_L(h) V$ is strictly positive definite for $\gamma < 1/(64n^3)$.

It follows that

$$|\det(D\mathcal{A}(h))| = |\det(V'^\top V)| \cdot |\det\big(I_{n-1} - \gamma V^\top H_L(h) V\big)| > 0.$$

□

Proof of Corollary 5.3.2. Since $\left\|C_f R\hat{h} \pm e_j\right\| \le 2\sqrt{\rho}$ for some $j \in [n]$, by the Cauchy-Schwarz inequality

$$\left\|\mathcal{F}(f) \odot \mathcal{F}(R\hat{h}) - \mathcal{F}(\mp e_j)\right\|_\infty \le \sqrt{n}\left\|C_f R\hat{h} \pm e_j\right\| \le 2\sqrt{\rho n}. \tag{C.18}$$

Equivalently, the circular convolution operators satisfy

$$\left\|C_f C_{R\hat{h}} - C_{\mp e_j}\right\| \le 2\sqrt{\rho n}.$$

It follows that

$$\begin{aligned}
&\|\hat{x}_i \pm \mathcal{S}_j(x_i)\| = \left\|C_{y_i} R\hat{h} \pm \mathcal{S}_j(x_i)\right\| \\
&= \left\|C_f C_{R\hat{h}} x_i - C_{\mp e_j} x_i\right\| \le \left\|C_f C_{R\hat{h}} - C_{\mp e_j}\right\| \cdot \|x_i\| \\
&\le 2\sqrt{\rho n} \cdot \|x_i\| .
\end{aligned}$$

It follows from (C.18) that

$$|\mathcal{F}(f)_{(k)} \times \mathcal{F}(R\hat{h})_{(k)} - e^{\frac{\pm\sqrt{-1}(j-1)(k-1)}{n}}| \le 2\sqrt{\rho n},$$

for all $k \in [n]$. Therefore,

$$|\mathcal{F}(R\hat{h})_{(k)}| \ge \frac{1 - 2\sqrt{\rho n}}{|\mathcal{F}(f)_{(k)}|}.$$

Since $\min_{k\in[n]} |\mathcal{F}(R\hat{h})_{(k)}| = \sigma_n(C_{R\hat{h}})$, and $\max_{k\in[n]} |\mathcal{F}(f)_{(k)}| = \|\mathcal{F}(f)\|_\infty \le \sqrt{n}\,\|f\|$, we have

$$\sigma_n(C_{R\hat{h}}) \ge \frac{1 - 2\sqrt{\rho n}}{\sqrt{n}\,\|f\|}.$$

Combining the above bound with the following

$$\left\|C_{R\hat{h}}(f \pm \mathcal{S}_j(\hat{f}))\right\| = \left\|C_f R\hat{h} \pm e_j\right\| \le 2\sqrt{\rho},$$

we have

$$\begin{aligned}
&\left\|\hat{f} \pm \mathcal{S}_{-j}(f)\right\| = \left\|f \pm \mathcal{S}_j(\hat{f})\right\| \\
&\leq \frac{\left\|C_{R\hat{h}}(f \pm \mathcal{S}_j(\hat{f}))\right\|}{\sigma_n(C_{R\hat{h}})} \leq 2\sqrt{\rho} \times \frac{\sqrt{n}\,\|f\|}{1 - 2\sqrt{\rho n}} \\
&= \frac{2\sqrt{\rho n}}{1 - 2\sqrt{\rho n}} \cdot \|f\| .
\end{aligned}$$

□

REFERENCES

[1] Y. Li, K. Lee, and Y. Bresler, "Identifiability in bilinear inverse problems with applications to subspace or sparsity-constrained blind gain and phase calibration," *IEEE Transactions on Information Theory*, vol. 63, no. 2, pp. 822–842, Feb 2017.

[2] Y. Li, K. Lee, and Y. Bresler, "Identifiability and stability in blind deconvolution under minimal assumptions," *IEEE Transactions on Information Theory*, vol. 63, no. 7, pp. 4619–4633, July 2017.

[3] Y. Li, K. Lee, and Y. Bresler, "Optimal sample complexity for blind gain and phase calibration," *IEEE Transactions on Signal Processing*, vol. 64, no. 21, pp. 5549–5556, Nov 2016.

[4] Y. Li, K. Lee, and Y. Bresler, "Blind gain and phase calibration via sparse spectral methods," *arXiv preprint arXiv:1712.00111*, 2017.

[5] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Processing Magazine*, vol. 13, no. 3, pp. 43–64, May 1996.

[6] L. R. Litwin, "Blind channel equalization," *IEEE Potentials*, vol. 18, no. 4, pp. 9–12, Oct 1999.

[7] P. A. Naylor and N. D. Gaubitch, Eds., *Speech Dereverberation*. Springer London, 2010.

[8] O. Yilmaz, *Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data*. Society of Exploration Geophysicists, 2001. [Online]. Available: http://library.seg.org/doi/book/10.1190/1.9781560801580

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

[10] K. Herrity, R. Raich, and A. O. Hero, "Blind reconstruction of sparse images with unknown point spread function," in *Computational Imaging VI*, vol. 6814. International Society for Optics and Photonics, 2008, p. 68140K.

[11] M. S. Asif, W. Mantzel, and J. Romberg, "Random channel coding and blind deconvolution," in *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on.* IEEE, 2009, pp. 1021–1025.

[12] D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *CVPR 2011*, June 2011, pp. 233–240.

[13] A. Repetti, M. Q. Pham, L. Duval, É. Chouzenoux, and J. C. Pesquet, "Euclid in a taxicab: Sparse blind deconvolution with smoothed ell_1/ell_2 regularization," *IEEE Signal Processing Letters*, vol. 22, no. 5, pp. 539–543, May 2015.

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

[15] E. J. Candès, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1241–1274, 2013.

[16] S. Choudhary and U. Mitra, "Identifiability scaling laws in bilinear inverse problems," *arXiv preprint arXiv:1402.2637*, 2014.

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

[18] Y. C. Eldar, D. Needell, and Y. Plan, "Uniqueness conditions for low-rank matrix recovery," *Applied and Computational Harmonic Analysis*, vol. 33, no. 2, pp. 309–314, 2012.

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

[20] K. Lee, Y. Li, M. Junge, and Y. Bresler, "Stability in blind deconvolution of sparse signals and reconstruction by alternating minimization," in *2015 International Conference on Sampling Theory and Applications (SampTA)*, May 2015, pp. 158–162.

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

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

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

[24] K. Lee, Y. Wu, and Y. Bresler, "Near optimal compressed sensing of a class of sparse low-rank matrices via sparse power factorization," *arXiv preprint arXiv:1312.0525*, 2013.

[25] G. Harikumar and Y. Bresler, "FIR perfect signal reconstruction from multiple convolutions: Minimum deconvolver orders," *IEEE Transactions on Signal Processing*, vol. 46, no. 1, pp. 215–218, Jan 1998.

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

[27] E. Riegler, D. Stotz, and H. Bölcskei, "Information-theoretic limits of matrix completion," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 1836–1840.

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

[29] H. Q. Nguyen, S. Liu, and M. N. Do, "Subspace methods for computational relighting," in *Computational Imaging XI*, vol. 8657. International Society for Optics and Photonics, 2013, p. 865703.

[30] A. Paulraj and T. Kailath, "Direction of arrival estimation by eigenstructure methods with unknown sensor gain and phase," in *ICASSP '85. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 10, Apr 1985, pp. 640–643.

[31] B. Friedlander and A. J. Weiss, "Eigenstructure methods for direction finding with sensor gain and phase uncertainties," in *ICASSP-88., International Conference on Acoustics, Speech, and Signal Processing*, Apr 1988, pp. 2681–2684 vol.5.

[32] J. Kim, H. J. Yang, B. W. Jung, and J. Chun, "Blind calibration for a linear array with gain and phase error using independent component analysis," *IEEE Antennas and Wireless Propagation Letters*, vol. 9, pp. 1259–1262, 2010.

[33] A. Liu, G. Liao, C. Zeng, Z. Yang, and Q. Xu, "An eigenstructure method for estimating DOA and sensor gain-phase errors," *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 5944–5956, Dec 2011.

[34] S. Cao, Z. Ye, D. Xu, and X. Xu, "A Hadamard product based method for DOA estimation and gain-phase error calibration," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 49, no. 2, pp. 1224–1233, Apr 2013.

[35] K. Han, P. Yang, and A. Nehorai, "Calibrating nested sensor arrays with model errors," *IEEE Transactions on Antennas and Propagation*, vol. 63, no. 11, pp. 4739–4748, Nov 2015.

[36] W. Ye, T. S. Yeo, and Z. Bao, "Weighted least-squares estimation of phase errors for SAR/ISAR autofocus," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 5, pp. 2487–2494, Sep 1999.

[37] G. Farquharson, P. Lopez-Dekker, and S. J. Frasier, "Contrast-based phase calibration for remote sensing systems with digital beamforming antennas," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 3, pp. 1744–1754, March 2013.

[38] R. L. Morrison, M. N. Do, and D. C. Munson, "MCA: A multichannel approach to SAR autofocus," *IEEE Transactions on Image Processing*, vol. 18, no. 4, pp. 840–853, April 2009.

[39] Y. Li, K. Lee, and Y. Bresler, "Uniqueness in bilinear inverse problems with applications to subspace and joint sparsity models," in *2015 International Conference on Sampling Theory and Applications (SampTA)*, May 2015, pp. 568–572.

[40] S. Ling and T. Strohmer, "Self-calibration via linear least squares," *arXiv preprint*, 2016.

[41] L. Wang and Y. Chi, "Blind deconvolution from multiple sparse inputs," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1384–1388, Oct 2016.

[42] G. H. Golub and C. F. Van Loan, *Matrix Computations (third edition)*. JHU Press, 1996.

[43] B. Moghaddam, Y. Weiss, and S. Avidan, "Spectral bounds for sparse PCA: Exact and greedy algorithms," in *Advances in Neural Information Processing Systems*, 2006, pp. 915–922.

[44] X.-T. Yuan and T. Zhang, "Truncated power method for sparse eigenvalue problems," *Journal of Machine Learning Research*, vol. 14, no. Apr, pp. 899–925, 2013.

[45] S. Cho and S. Lee, "Fast motion deblurring," in *ACM Transactions on Graphics (TOG)*, vol. 28, no. 5. ACM, 2009, p. 145.

[46] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding blind deconvolution algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2354–2367, Dec 2011.

[47] L. Xu, S. Zheng, and J. Jia, "Unnatural l0 sparse representation for natural image deblurring," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 1107–1114.

[48] W. Huang and P. Hand, "Blind deconvolution by a steepest descent algorithm on a quotient manifold," *arXiv preprint arXiv:1710.03309*, 2017.

[49] Y. Zhang, Y. Lau, H.-w. Kuo, S. Cheung, A. Pasupathy, and J. Wright, "On the global geometry of sphere-constrained sparse blind deconvolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4894–4902.

[50] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 1951–1968, Oct 1998.

[51] H. She, R.-R. Chen, D. Liang, Y. Chang, and L. Ying, "Image reconstruction from phased-array data based on multichannel blind deconvolution," *Magnetic Resonance Imaging*, vol. 33, no. 9, pp. 1106–1113, 2015.

[52] H. Zhang, D. Wipf, and Y. Zhang, "Multi-image blind deblurring using a coupled adaptive sparse prior," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 1051–1058.

[53] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channels," in *[1991] Conference Record of the Twenty-Fifth Asilomar Conference on Signals, Systems Computers*, vol. 2, Nov 1991, pp. 856–860.

[54] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 516–525, Feb 1995.

[55] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Transactions on Signal Processing*, vol. 43, no. 12, pp. 2982–2993, Dec 1995.

[56] M. I. Gurelli and C. L. Nikias, "EVAM: An eigenvector-based algorithm for multichannel blind deconvolution of input colored signals," *IEEE Transactions on Signal Processing*, vol. 43, no. 1, pp. 134–149, Jan 1995.

[57] K. Lee, F. Krahmer, and J. Romberg, "Spectral methods for passive imaging: Non-asymptotic performance and robustness," *arXiv preprint arXiv:1708.04343*, 2017.

[58] K. Lee, N. Tian, and J. Romberg, "Fast and guaranteed blind multichannel deconvolution under a bilinear system model," *arXiv preprint arXiv:1610.06469*, 2016.

[59] C. R. Berger, S. Zhou, J. C. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1708–1721, March 2010.

[60] K. G. Sabra and D. R. Dowling, "Blind deconvolution in ocean waveguides using artificial time reversal," *The Journal of the Acoustical Society of America*, vol. 116, no. 1, pp. 262–271, 2004.

[61] N. Tian, S.-H. Byun, K. Sabra, and J. Romberg, "Multichannel myopic deconvolution in underwater acoustic channels via low-rank recovery," *The Journal of the Acoustical Society of America*, vol. 141, no. 5, pp. 3337–3348, 2017.

[62] K. F. Kaaresen and T. Taxt, "Multichannel blind deconvolution of seismic signals," *Geophysics*, vol. 63, no. 6, pp. 2093–2107, 1998.

[63] D. R. Gitelman, W. D. Penny, J. Ashburner, and K. J. Friston, "Modeling regional and psychophysiologic interactions in fMRI: The importance of hemodynamic deconvolution," *Neuroimage*, vol. 19, no. 1, pp. 200–207, 2003.

[64] M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," *Nature Methods*, vol. 3, no. 10, p. 793, 2006.

[65] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution," *Science*, vol. 313, no. 5793, pp. 1642–1645, 2006.

[66] E. A. Mukamel, H. Babcock, and X. Zhuang, "Statistical deconvolution for superresolution fluorescence microscopy," *Biophysical Journal*, vol. 102, no. 10, pp. 2391–2400, 2012.

[67] P. Sarder and A. Nehorai, "Deconvolution methods for 3-D fluorescence microscopy images," *IEEE Signal Processing Magazine*, vol. 23, no. 3, pp. 32–45, May 2006.

[68] T. Strohmer, "Four short stories about toeplitz matrix calculations," *Linear Algebra and its Applications*, vol. 343, pp. 321–344, 2002.

[69] L. Balzano and R. Nowak, "Blind calibration of sensor networks," in *Proceedings of the 6th International Conference on Information Processing in Sensor Networks*. ACM, 2007, pp. 79–88.

[70] C. Bilen, G. Puy, R. Gribonval, and L. Daudet, "Convex optimization approaches for blind sensor calibration using sparsity," *IEEE Transactions on Signal Processing*, vol. 62, no. 18, pp. 4847–4856, Sept 2014.

[71] J. Sun, Q. Qu, and J. Wright, "Complete dictionary recovery over the sphere i: Overview and the geometric picture," *IEEE Transactions on Information Theory*, vol. 63, no. 2, pp. 853–884, Feb 2017.

[72] J. Sun, Q. Qu, and J. Wright, "A geometric analysis of phase retrieval," *Foundations of Computational Mathematics*, Aug 2017. [Online]. Available: https://doi.org/10.1007/s10208-017-9365-9

[73] S. Mei, Y. Bai, and A. Montanari, "The landscape of empirical risk for non-convex losses," *arXiv preprint arXiv:1607.06534*, 2016.

[74] J. Sun, Q. Qu, and J. Wright, "Complete dictionary recovery over the sphere ii: Recovery by Riemannian trust-region method," *IEEE Transactions on Information Theory*, vol. 63, no. 2, pp. 885–914, Feb 2017.

[75] N. Boumal, P.-A. Absil, and C. Cartis, "Global rates of convergence for nonconvex optimization on manifolds," *arXiv preprint arXiv:1605.08101*, 2016.

[76] J. D. Lee, I. Panageas, G. Piliouras, M. Simchowitz, M. I. Jordan, and B. Recht, "First-order methods almost always avoid saddle points," *arXiv preprint arXiv:1710.07406*, 2017.

[77] J. D. Lee, M. Simchowitz, M. I. Jordan, and B. Recht, "Gradient descent only converges to minimizers," in *Conference on Learning Theory*, 2016, pp. 1246–1257.

[78] I. Panageas and G. Piliouras, "Gradient descent only converges to minimizers: Non-isolated critical points and invariant regions," *arXiv preprint arXiv:1605.00405*, 2016.

[79] R. Ge, F. Huang, C. Jin, and Y. Yuan, "Escaping from saddle points – Online stochastic gradient for tensor decomposition," in *Conference on Learning Theory*, 2015, pp. 797–842.

[80] C. Jin, R. Ge, P. Netrapalli, S. M. Kakade, and M. I. Jordan, "How to escape saddle points efficiently," in *International Conference on Machine Learning*, 2017, pp. 1724–1732.

[81] Z. Allen-Zhu, "Natasha: Faster stochastic non-convex optimization via strongly non-convex parameter," *arXiv preprint arXiv:1702.00763*, 2017.

[82] Z. Allen-Zhu, "Natasha 2: Faster non-convex optimization than SGD," *arXiv preprint arXiv:1708.08694*, 2017.

[83] N. Agarwal, Z. Allen-Zhu, B. Bullins, E. Hazan, and T. Ma, "Finding approximate local minima faster than gradient descent," in *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*. ACM, 2017, pp. 1195–1199.

[84] D. Goldfarb, C. Mu, J. Wright, and C. Zhou, "Using negative curvature in solving nonlinear programs," *Computational Optimization and Applications*, vol. 68, no. 3, pp. 479–502, Dec 2017. [Online]. Available: https://doi.org/10.1007/s10589-017-9925-6

[85] W. Mantzel and J. Romberg, "Compressed subspace matching on the continuum," *Information and Inference: A Journal of the IMA*, vol. 4, no. 2, pp. 79–107, 2015.

[86] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[87] G. Reeves, "The fundamental limits of stable recovery in compressed sensing," in *2014 IEEE International Symposium on Information Theory*, June 2014, pp. 3017–3021.

[88] D. Pollard, "Empirical processes: Theory and applications," in *NSF-CBMS Regional Conference Series in Probability and Statistics*. JSTOR, 1990, pp. i–86.

[89] Y. Rockah, H. Messer, and P. M. Schultheiss, "Localization performance of arrays subject to phase errors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 24, no. 4, pp. 402–410, July 1988.

[90] A. J. Weiss and B. Friedlander, "Eigenstructure methods for direction finding with sensor gain and phase uncertainties," *Circuits, Systems and Signal Processing*, vol. 9, no. 3, pp. 271–300, 1990.

[91] J. Lipor and L. Balzano, "Robust blind calibration via total least squares," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 4244–4248.

[92] Y. C. Eldar, W. Liao, and S. Tang, "Sensor calibration for off-the-grid spectral estimation," *arXiv preprint arXiv:1707.03378*, 2017.

[93] M. A. Davenport and J. Romberg, "An overview of low-rank matrix recovery from incomplete observations," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 608–622, June 2016.

[94] Y. Li, K. Lee, and Y. Bresler, "Optimal sample complexity for stable matrix recovery," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 81–85.

[95] S. Bahmani and J. Romberg, "Lifting for blind deconvolution in random mask imaging: Identifiability and convex relaxation," *SIAM Journal on Imaging Sciences*, vol. 8, no. 4, pp. 2203–2238, 2015.

[96] G. Tang and B. Recht, "Convex blind deconvolution with random masks," in *Computational Optical Sensing and Imaging*. Optical Society of America, 2014, pp. CW4C–1.

[97] V. Cambareri and L. Jacques, "A non-convex blind calibration method for randomised sensing strategies," in *2016 4th International Workshop on Compressed Sensing Theory and Its Applications to Radar, Sonar and Remote Sensing (CoSeRa)*, Sept 2016, pp. 16–20.

[98] V. Cambareri, A. Moshtaghpour, and L. Jacques, "A greedy blind calibration method for compressed sensing with unknown sensor gains," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 1132–1136.

[99] A. Ahmed and L. Demanet, "Leveraging diversity and sparsity in blind deconvolution," *IEEE Transactions on Information Theory*, vol. 64, no. 6, pp. 3975–4000, June 2018.

[100] A. Cosse, "From blind deconvolution to blind super-resolution through convex programming," *arXiv preprint arXiv:1709.09279*, 2017.

[101] J. R. Fienup, "Phase retrieval algorithms: A comparison," *Applied Optics*, vol. 21, no. 15, pp. 2758–2769, 1982.

[102] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," in *Advances in Neural Information Processing Systems*, 2013, pp. 2796–2804.

[103] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1985–2007, April 2015.

[104] T. T. Cai, X. Li, Z. Ma et al., "Optimal rates of convergence for noisy sparse phase retrieval via thresholded Wirtinger flow," *The Annals of Statistics*, vol. 44, no. 5, pp. 2221–2251, 2016.

[105] S. Bahmani and J. Romberg, "Phase retrieval meets statistical learning theory: A flexible convex relaxation," *arXiv preprint arXiv:1610.04210*, 2016.

[106] T. Goldstein and C. Studer, "Phasemax: Convex phase retrieval via basis pursuit," *IEEE Transactions on Information Theory*, vol. 64, no. 4, pp. 2675–2689, April 2018.

[107] Y. Chen, Y. Chi, J. Fan, and C. Ma, "Gradient descent with random initialization: Fast global convergence for nonconvex phase retrieval," *arXiv preprint arXiv:1803.07726*, 2018.

[108] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: A contemporary overview," *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 87–109, May 2015.

[109] J. Sun, Q. Qu, and J. Wright, "When are nonconvex problems not scary?" *arXiv preprint arXiv:1510.06096*, 2015.

[110] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from a few entries," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2980–2998, June 2010.

[111] P. Jain, P. Netrapalli, and S. Sanghavi, "Low-rank matrix completion using alternating minimization," in *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*. ACM, 2013, pp. 665–674.

[112] R. Sun and Z. Q. Luo, "Guaranteed matrix completion via non-convex factorization," *IEEE Transactions on Information Theory*, vol. 62, no. 11, pp. 6535–6579, Nov 2016.

[113] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, Dec 2005.

[114] C. Davis and W. M. Kahan, "The rotation of eigenvectors by a perturbation. iii," *SIAM Journal on Numerical Analysis*, vol. 7, no. 1, pp. 1–46, 1970.

[115] J. A. Tropp, "User-friendly tail bounds for sums of random matrices," *Foundations of Computational Mathematics*, vol. 12, no. 4, pp. 389–434, aug 2011.

[116] K. Lee and M. Junge, "RIP-like properties in subsampled blind deconvolution," *arXiv preprint arXiv:1511.06146*, 2015.

[117] Q. Berthet and P. Rigollet, "Computational lower bounds for sparse PCA," *arXiv preprint arXiv:1304.0828*, 2013.

[118] K. Jaganathan, S. Oymak, and B. Hassibi, "Sparse phase retrieval: Uniqueness guarantees and recovery algorithms," *IEEE Transactions on Signal Processing*, vol. 65, no. 9, pp. 2402–2410, May 2017.

[119] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein et al., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[120] Y. Zhang, H.-W. Kuo, and J. Wright, "Structured local optima in sparse blind deconvolution," in *Proceedings of the 10th NIPS Workshop on Optimization for Machine Learning (OPTML)*, 2017.

[121] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[122] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2009.

[123] M. P. Wylie, S. Roy, and R. F. Schmitt, "Self-calibration of linear equi-spaced (LES) arrays," in *1993 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, April 1993, pp. 281–284 vol.1.

[124] K. R. Davidson and S. J. Szarek, "Local operator theory, random matrices and Banach spaces," *Handbook of the Geometry of Banach Spaces*, vol. 1, no. 317-366, p. 131, 2001.

[125] M. Rudelson, R. Vershynin et al., "Hanson-Wright inequality and sub-Gaussian concentration," *Electronic Communications in Probability*, vol. 18, 2013.

[126] F. Krahmer, S. Mendelson, and H. Rauhut, "Suprema of chaos processes and the restricted isometry property," *Communications on Pure and Applied Mathematics*, vol. 67, no. 11, pp. 1877–1904, 2014.

[127] M. Ledoux and M. Talagrand, *Probability in Banach Spaces: Isoperimetry and Processes*. Springer Science & Business Media, 2013.

[128] S. Artstein, V. Milman, and S. J. Szarek, "Duality of metric entropy," *Annals of Mathematics*, pp. 1313–1328, 2004.

[129] B. Carl, "Inequalities of Bernstein-Jackson-type and the degree of compactness of operators in Banach spaces," *Ann. Inst. Fourier (Grenoble)*, vol. 35, no. 3, pp. 79–118, 1985.

[130] M. Junge and K. Lee, "Generalized notions of sparsity and restricted isometry property. part i: A unified framework," *arXiv preprint arXiv:1706.09410*, 2017.

[131] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Birkhäuser Basel, 2013, vol. 1, no. 3.

[132] J. Sun, Q. Qu, and J. Wright, "Complete dictionary recovery over the sphere," *arXiv preprint arXiv:1504.06785*, 2015.

[133] N. J. Higham, *Functions of Matrices: Theory and Computation*. SIAM, 2008, vol. 104.