A UNIFIED FRAMEWORK FOR IDENTIFIABILITY ANALYSIS IN BILINEAR INVERSE PROBLEMS

BY

YANJUN LI

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Adviser:

Professor Yoram Bresler

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, properties of natural signals are exploited to solve BIPs. For example, subspace constraints or sparsity constraints are imposed to reduce the search space. These approaches have shown some success in practice. However, there are few results on uniqueness in BIPs. For most BIPs, the fundamental question of under what condition the problem admits a unique solution, is yet to be answered. As an effort to address the question, we propose a unified framework for identifiability analysis in BIPs. We define identifiability of a BIP up to a group of transformations. Then we derive necessary and sufficient conditions for such identifiability, i.e., the conditions under which the solutions can be uniquely determined up to the transformation group.

Blind gain and phase calibration (BGPC) is a structured bilinear inverse problem, which arises in many applications, including inverse rendering in computational relighting (albedo estimation with unknown lighting), blind phase and gain calibration in sensor array processing, and multichannel blind deconvolution (MBD). Applying our unified framework to BGPC, we derive sufficient conditions for unique recovery under several scenarios, including subspace, joint sparsity, and sparsity models. For BGPC with joint sparsity or sparsity constraints, we develop a procedure to compute the transformation groups corresponding to inherent ambiguities. We also give necessary conditions in the form of tight lower bounds on sample complexities, and demonstrate the tightness of these bounds by numerical experiments.

Blind deconvolution (BD), the resolution of a signal and a filter given their convolution, is another bilinear inverse problem routinely encountered in signal processing and communications. Existing theoretical analysis on uniqueness in BD is rather limited. We derive sufficient conditions under which two vectors can be uniquely identified from their circular convolution, subject to subspace or sparsity constraints. These sufficient conditions provide the first algebraic sample complexities for BD. We first derive a sufficient condition that applies to almost all bases or frames. Then we impose a sub-band structure on one basis, and derive a less demanding sufficient condition, which is essentially optimal, using our unified framework. We present the extensions of these results to BD with sparsity constraints or mixed constraints, with the sparsity level replacing the subspace dimension. The cost for the unknown support in this case is an extra factor of 2 in the sample complexity.

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 first three years here. He shared with me his knowledge, experience, insight and philosophy. He showed me what is good research and how to do good research. Moreover, his advice went beyond study and research, and directed me through difficult times. I owe what I learned in graduate school mostly to him.

My thanks go to my colleagues Kiryung Lee, Saiprasad Ravishankar, Luke Pfister, and Bihan Wen. I learned a lot from them and was inspired by them every day. They made graduate school so much more fun.

I would also like to thank my parents. Their support and love have always been and will always be my source of confidence and motivation. I love you, mom and dad.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Whereas linear inverse problems are well-understood and the literature on them is vast, much less is known about bilinear inverse problems (BIPs). BIPs, i.e., recovering two variables x and y given a bilinear measurement $z = \mathcal{F}(x, y)$, have attracted considerable attention recently. However, in spite of recent progress, the question of identifiability - or uniqueness of the solutions in BIPs under a variety of realistic conditions - has been largely open. BIPs arise in many important applications, such as blind deconvolution [1, 2], phase retrieval [3, 4], dictionary learning [5], etc. These problems usually involve recovering the inputs of an under-determined bilinear system. They also suffer from scaling ambiguity among other possible ambiguities (e.g., shift ambiguity of blind deconvolution, multiplication by a permutation matrix in dictionary learning, multiplication by an arbitrary invertible matrix in matrix factorization problems, etc.). Therefore, these problems are ill-posed and do not yield unique solutions. By introducing further constraints that exploit the properties of natural signals, one can reduce the search space, which may help identifiability. For example, cone constraints, such as positivity constraints, subspace constraints, and union of subspaces constraints (e.g., sparsity or joint sparsity), are very common in BIPs. However, even with a reduced feasible set, a BIP often still exhibits some ambiguities, such as scaling [6].

In this thesis, we study the identifiability in bilinear inverse problems. We expand the notion of identifiability and propose a unified framework, namely identifiability up to transformation groups. We also study the identifiability in two special BIPs, blind gain and phase calibration (BGPC) and blind deconvolution (BD), within the unified framework.

1.1 Bilinear Inverse Problem

1.1.1 Bilinear Inverse Problem

We formally state the bilinear inverse problem (BIP) in this section. First, a bilinear mapping is defined as follows.

Definition 1.1.1. *Let $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ be three linear vector spaces. A bilinear mapping is a function $\mathcal{F} : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ such that for any $y \in \mathcal{Y}$ the mapping $x \mapsto \mathcal{F}(x, y)$ is a linear mapping from $\mathcal{X}$ to $\mathcal{Z}$ and for any $x \in \mathcal{X}$ the mapping $y \mapsto \mathcal{F}(x, y)$ is a linear mapping from $\mathcal{Y}$ to $\mathcal{Z}$.*

Given the measurement $z = \mathcal{F}(x_0, y_0)$, the following feasibility problem is called the unconstrained bilinear inverse problem:

$$\begin{aligned} \text{(Unconstrained BIP)} \qquad \text{find} \;\; & (x, y) \in \mathcal{X} \times \mathcal{Y}, \\ \text{s.t.} \;\; & \mathcal{F}(x, y) = z. \end{aligned}$$

Bilinear inverse problems are usually underdetermined, and hence do not yield unique solutions. A variety of constraints $x \in \Omega_\mathcal{X} \subset \mathcal{X}, y \in \Omega_\mathcal{Y} \subset \mathcal{Y}$ can be imposed to reduce the search space and make the problem better-posed. The constrained bilinear inverse problem is:

$$\begin{aligned} \text{(Constrained BIP)} \qquad \text{find} \;\; & (x, y), \\ \text{s.t.} \;\; & \mathcal{F}(x, y) = z, \\ & x \in \Omega_\mathcal{X}, \;\; y \in \Omega_\mathcal{Y}. \end{aligned} \tag{1.1}$$

For any nonzero scalar σ, the pairs (x_0, y_0) and $(\sigma x_0, \frac{1}{\sigma} y_0)$ map to the same z and hence are non-distinguishable. If the constraint sets $\Omega_\mathcal{X}$ and $\Omega_\mathcal{Y}$ contain such scaled versions of (x_0, y_0), we say that this problem suffers from scaling ambiguity. Suppose $\Omega_\mathcal{X}$ and $\Omega_\mathcal{Y}$ are closed under scalar multiplication, then $\{(\sigma x_0, \frac{1}{\sigma} y_0) : \sigma \neq 0\}$ is an equivalence class of solutions generated by a group of scaling transformations. More complex ambiguities and equivalence classes will be analyzed later. In Chapter 2, to address the issues of ambiguity, we expand the notion of identifiability of BIPs. We resolve the ambiguity issues by allowing uniqueness up to a group of transformations, which define equivalence classes of solutions. We then derive necessary and sufficient conditions for identifiability in BIPs up to the transformation group.

1.1.2 Related Work

A standard method for solving bilinear inverse problems is the Gauss-Newton method, if $\mathcal{F}(x, y)$ is Fréchet differentiable with respect to (x, y). The Gauss-Newton method is applied to minimize $\|r(x, y)\|_2^2$, where $r(x, y) = \mathcal{F}(x, y) - z$ is the residual. After initializing with a guess of (x, y), in each step, the algorithm linearizes the residual, and solves the normal equation that arises in the minimization of $\|r(x, y)\|_2^2$ for the linearized residual. A related approach, instead of solving the bilinear equation in x and y, is to solve a nonlinear equation in one of the variables. If we assume that the BIP is uniquely solvable for y given x, then the solution $y = y(x)$ is a function of x [7, 8]. If $\mathcal{F}(x, y(x))$ is Fréchet differentiable with respect to x, then the Gauss-Newton algorithm can be applied to minimize $\|\mathcal{F}(x, y(x)) - z\|_2^2$. To avoid inverting the normal operator in the Gauss-Newton algorithm in large scale problems, one can apply more computationally efficient iterative methods, such as the conjugate gradient method or the Kaczmarz method [8]. These algorithms usually take advantage of simple regularizers and constraints to resolve ill-posedness. For example, a Tikhonov regularizer of (x, y) or a linear constraint $\mathbf{1}^* x = 1$ can eliminate the scaling ambiguity [9].

Another approach for attacking bilinear inverse problems is the Bayesian approach. The measurement z is assumed to follow a probability distribution (e.g., Gaussian distribution) with mean $\mathcal{F}(x, y)$. The conditional distribution is called the likelihood. Instead of using deterministic models, the Bayesian approach uses probabilistic models for x and y whose distributions are called prior distributions. The posterior distribution of x and y given z can be computed using Bayes' rule. The maximum a posteriori (MAP) estimator and the minimum mean square error (MMSE) estimator of (x, y) are the mode and the expectation of the posterior distribution, respectively. Markov chain Monte Carlo (MCMC) or variational methods can be deployed to overcome computational challenges. Examples of applying Bayesian or variational Bayesian methods to bilinear inverse problems include color constancy in vision systems [10], blood-oxygen-level dependent (BOLD) signal analysis in functional magnetic resonance imaging (fMRI) [11], and blind image deconvolution [12, 13, 14, 15].

Recently, solving bilinear or quadratic inverse problems with the methodology of "lifting" has attracted much attention. Examples include recent works on blind deconvolution [16] and phase retrieval [17, 18, 19]. The lifting framework is based on the fact that for any bilinear mapping $\mathcal{F} : \mathbb{C}^m \times \mathbb{C}^n \to \mathcal{Z}$, there exists a linear operator $\mathcal{G} : \mathbb{C}^{m\times n} \to \mathcal{Z}$ such that $\mathcal{G}(xy^{\mathrm{T}}) = \mathcal{F}(x, y)$. Given the measurement $z = \mathcal{G}(x_0 y_0^{\mathrm{T}}) = \mathcal{F}(x_0, y_0)$, one can recast the BIP as the recovery of the rank-1 matrix $x_0 y_0^{\mathrm{T}} \in \Omega_{\mathcal{M}} = \{xy^{\mathrm{T}} : x \in \Omega_{\mathcal{X}}, y \in \Omega_{\mathcal{Y}}\}$.

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

Choudhary and Mitra [6] adopted this framework, and showed that the lifted BIP has a unique solution $M_0 = x_0 y_0^{\mathrm{T}}$ if the null space of $\mathcal{G}$ does not contain the difference of M_0 and any other matrix in $\Omega_{\mathcal{M}}$, i.e.,

$$
\mathcal{N}(\mathcal{G}) \bigcap \{M_0 - M : M \in \Omega_{\mathcal{M}}\} = \{0\}.
$$

The identifiability analysis hinges on finding the set of rank-2 matrices in the null space of $\mathcal{G}$. They addressed the question of identifiability in an abstract BIP under the assumptions that the set of rank-2 matrices in $\mathcal{N}(\mathcal{G})$ has low complexity (e.g., finite cardinality or small covering number). Using this framework, they showed that blind deconvolution with a canonical sparsity prior is *not* identifiable [20].

In contrast, we propose a more general framework in Chapter 2. Our framework deals with bilinear mappings defined on general vector spaces (not just Euclidean spaces). Besides scaling ambiguity, our framework allows other ambiguities. We extend the notion of identifiability to identifiability up to transformation groups. Our framework is amenable to BIPs with matrix multiplications, such as dictionary learning [21, 22, 23, 24, 25] and blind gain and phase calibration (cf. Chapter 3).

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 λ and signal vectors $\phi_1, \phi_2, \cdots, \phi_N$ given the entrywise product $Y = \mathrm{diag}(\lambda)\Phi$, where $\Phi = [\phi_1, \phi_2, \cdots, \phi_N]$. Given the measurement $Y = \mathrm{diag}(\lambda_0)\Phi_0$, BGPC is the following constrained BIP:

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

In inverse rendering [26], when the surface profile (3D model) of the object is known, the joint recovery of the albedo[1] and the lighting conditions is a BGPC problem. In sensor array processing [27], if the directions of arrival of source signals are properly discretized using a grid, and the sensors have unknown gain and phase, the joint recovery of the source signals and the gain and phase of the sensors is a BGPC problem. In multichannel blind deconvolution (MBD) with the circular convolution model, the joint recovery of the signal and multiple channels is a BGPC problem. In all these problems, it is common to impose subspace, joint sparsity, or sparsity constraints on the signals represented by the columns of Φ.

After deriving general necessary and sufficient conditions for identifiability in a BIP up to the transformation group in Chapter 2, we apply these to BGPC and give identifiability results under several scenarios in Chapter 3. We first consider a subspace constraint and provide an alternative proof for the result in inverse rendering [26]. Then we consider a joint sparsity constraint. We develop a procedure to determine the relevant equivalence classes and transformation groups for different bases. Then we give sufficient conditions for the identifiability of jointly sparse signals (1D or 2D), or piecewise constant signals.

For BGPC with subspace or joint sparsity constraints, we also give necessary conditions in the form of tight lower bounds on sample complexities. We show that the sufficient conditions and the necessary conditions coincide in some cases. We design algorithms to check the identifiability of given signals and demonstrate the tightness of our sample complexity bounds. We analyze the gaps and present conjectures about how to bridge them.

Then we derive a universal sufficient condition for BGPC with a sparsity constraint. This condition is the most stringent, but applies to all bases and

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

all equivalence classes of solutions. Once the condition is met, the solution of the BGPC problem can be recovered uniquely up to an unknown generalized permutation, regardless of the basis.

The structure of the BGPC problem arises in many signal processing applications. In each of these, the problem formulation and treatment were tailored to the application. Instead, we address the identifiability of all these problems within the one common framework. Nguyen et al. [26] showed a sufficient condition for unique inverse rendering, which falls into the category of BGPC problems with subspace constraints. By examining the problem in our framework, we are able to replicate Nguyen's result and provide an alternative proof. In addition, we give a new necessary condition that features a tight lower bound.

Morrison et al. [28] proposed an algorithm for SAR autofocus and showed a necessary condition for their algorithm. If the support is unknown, the SAR autofocus problem falls into the category of BGPC problems with joint sparsity constraints. Using our notion of identifiability up to a transformation group, we provide a sufficient condition for unique recovery up to an unknown scaling and a circular shift.

Most works on the identifiability of MBD considered the linear convolution model [29, 2]. These traditional works used finite impulse response (FIR) models, and never incorporated joint sparsity, or sparsity. In contrast, we consider the circular convolution model, which is more challenging in that the circular convolution with a vector can be non-injective, while the linear convolution with a vector is always injective. On the other hand, the circular convolution model is more general. By zero padding the signal and the channels (equivalent to Fourier domain oversampling), linear convolutions can be rewritten as circular convolutions with a support constraint. That falls into the category of BGPC with a subspace constraint. As an important extension of the theory of MBD, we study in Chapter 3 MBD with subspace, joint-sparsity, and sparsity constraints.

1.3 Blind Deconvolution

Blind deconvolution (BD) is the bilinear inverse problem of recovering the signal and the filter simultaneously given the their convolution or circular convolution. It arises in many applications, including blind image deblurring [1], blind channel equalization [30], speech dereverberation [31], and seismic data analysis [32]. Without further constraints, BD is an ill-posed problem, and does not yield a unique solution. A variety of constraints have been introduced to exploit the properties of natural signals and reduce the search space. Examples of such constraints include positivity (the signals are non-negative), subspace constraint (the signals reside in a lower-dimensional subspace) and sparsity (the signals are sparse over some dictionary). In Chapter 4 of this thesis, we focus on subspace or sparsity constraints, which can be imposed on both the signal and the filter.

Consider the example of blind image deblurring: a natural image can be considered sparse over a wavelet dictionary or the discrete cosine transform (DCT) dictionary. The support of the point spread function (PSF) modeling the blur is usually significantly smaller than the image itself. Therefore the filter resides in a lower-dimensional subspace. These priors serve as constraints or regularizers [33, 34, 35, 36, 16]. With a reduced search space, BD can be better-posed. However, despite the success in practice, the theoretical results on the uniqueness in BD with a subspace or sparsity constraint are limited.

Early works on the identifiability in blind deconvolution studied multichannel blind deconvolution with finite impulse response (FIR) models [29, 2], in which sparsity was not considered. For single channel blind deconvolution, sparsity was imposed as a prior without theoretical justification [33, 34, 36, 35, 37].

As mentioned in Section 1.1.2, Choudhary and Mitra [6] adopted the lifting framework and showed that the identifiability in BD (or 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 blind deconvolution with the linear convolution model and a canonical sparsity prior, that is, sparsity over the natural basis, is *not* identifiable [20]. However, the identifiability of blind deconvolution with the circular convolution model or with signals that are sparse over other dictionaries has not been analyzed.

Using the lifting framework, Ahmed et al. [16] showed that BD with subspace constraints is identifiable up to scaling. More specifically, if the signal subspace follows a random Gaussian model, and the filter subspace satisfies some coherence conditions, convex programming was shown to recover the signal and the filter up to scaling with high probability, when the dimensions of the subspaces m_1 and m_2 are in a near optimal regime $m_1 + m_2 = O(n)$, where n denotes the length of the signal. Ling and Strohmer [38] extended the model in [16] to blind deconvolution with mixed constraints: the signal is sparse over a random Gaussian dictionary or a randomly subsampled partial Fourier matrix, and the filter resides in a subspace that satisfies some coherence condition. They showed that the signal and the filter can be simultaneously identified with high probability using ℓ_1 norm minimization (instead of nuclear norm minimization as in [16]) when the sparsity level s_1 and the subspace dimension m_2 satisfy $s_1 m_2 = O(n)$. Lee et al. [39] further extended the model to blind deconvolution with sparsity constraints on both the signal and the filter, and showed successful recovery with high probability using alternating minimization when the sparsity levels s_1 and s_2 satisfy $s_1 + s_2 = O(n)$. A common drawback of these works is that the probabilistic assumptions on the bases or frames are very limiting in practice. On the positive side, these identifiability results are constructive, being demonstrated by establishing performance guarantees of algorithms. However, these guarantees too are shown only in some probabilistic sense.

In Chapter 3, we study multichannel blind deconvolution as a special case of BGPC. Using the unified framework of identifiability up to transformation groups, we derive identifiability results under subspace, joint sparsity or sparsity constraints.

In Chapter 4, we address the identifiability in single channel blind deconvolution up to scaling under subspace or sparsity constraints. We present the first algebraic sample complexities for BD with fully deterministic signal models. First, we derive sufficient conditions for BD with generic bases or frames, using the lifting framework. Then, we derive much less demanding sufficient conditions for BD with a sub-band structured basis, using the unified framework in Chapter 2. Notably, the sample complexities of the sufficient conditions in this case match those of corresponding necessary conditions, and hence are optimal.

CHAPTER 2

IDENTIFIABILITY IN BILINEAR INVERSE PROBLEMS

2.1 Notations

We use $\Omega_\mathcal{X}, \Omega_\mathcal{Y}$ to denote subsets of vector spaces $\mathcal{X}, \mathcal{Y}$. The Cartesian product of two sets is denoted by $\Omega_\mathcal{X} \times \Omega_\mathcal{Y}$. An element of $\Omega_\mathcal{X} \times \Omega_\mathcal{Y}$ is denoted by (x, y), where $x \in \Omega_\mathcal{X}$ and $y \in \Omega_\mathcal{Y}$. We use $\mathscr{T}_\mathcal{X}$ and $\mathscr{T}_\mathcal{Y}$ to denote transformation groups (to be defined in Section 2.2). The Cartesian product of two transformation groups $\mathscr{T}_\mathcal{X}, \mathscr{T}_\mathcal{Y}$ (also known as direct product in group theory terminology) is denoted by $\mathscr{T}_\mathcal{X} \times \mathscr{T}_\mathcal{Y}$. Elements of the transformation groups are denoted by $\mathcal{T}_\mathcal{X} \in \mathscr{T}_\mathcal{X}$, $\mathcal{T}_\mathcal{Y} \in \mathscr{T}_\mathcal{Y}$ and $(\mathcal{T}_\mathcal{X}, \mathcal{T}_\mathcal{Y}) \in \mathscr{T}_\mathcal{X} \times \mathscr{T}_\mathcal{Y}$.

2.2 Transformation Groups and Equivalence Classes

An important question concerning a bilinear inverse problem is to determine when it admits a unique solution. To formulate a good answer, we need to be able to handle the ambiguities of a bilinear inverse problem. For any nonzero scalar σ such that $\sigma x_0 \in \Omega_\mathcal{X}$ and $\frac{1}{\sigma} y_0 \in \Omega_\mathcal{Y}$, by bilinearity, $\mathcal{F}(\sigma x_0, \frac{1}{\sigma} y_0) = \mathcal{F}(x_0, y_0) = z$. Therefore, the constrained BIP does not yield a unique solution if $\Omega_\mathcal{X}, \Omega_\mathcal{Y}$ contain such scaled versions of x_0, y_0. This is called scaling ambiguity.

When $\Omega_\mathcal{X}, \Omega_\mathcal{Y}$ are closed under scalar multiplication (e.g., subspaces or unions of subspaces), the set $[(x_0, y_0)] = \{(\sigma x_0, \frac{1}{\sigma} y_0) : \sigma \neq 0\}$ is an equivalence class with an exemplar (x_0, y_0). The transformation $\mathcal{T} : \Omega_\mathcal{X} \times \Omega_\mathcal{Y} \to \Omega_\mathcal{X} \times \Omega_\mathcal{Y}$ such that $\mathcal{T}(x, y) = (\sigma x, \frac{1}{\sigma} y)$ is an equivalence transformation. The set of all such transformations

$$\mathscr{T} = \{\mathcal{T} : \mathcal{T}(x, y) = (\sigma x, \frac{1}{\sigma} y), \text{ for some nonzero } \sigma \in \mathbb{C}\} \tag{2.1}$$

forms a transformation group. In group theory terminology, the equivalence class $[(x_0, y_0)]$ is the orbit of (x_0, y_0) under the action of $\mathscr{T}$ [40].

Any valid definition of unique recovery must include uniqueness up to scaling, i.e., the equivalence class $[(x_0, y_0)]$ can be uniquely identified. There can be other ambiguities for a particular bilinear inverse problem (e.g., shift ambiguity of blind deconvolution).

We need formal definitions of transformation groups and equivalence classes before proceeding towards identifiability.

Definition 2.2.1. *A set $\mathscr{T}_\mathcal{X}$ of transformations from $\Omega_\mathcal{X}$ to itself is said to be a transformation group on $\Omega_\mathcal{X}$, if the following properties hold:*

1. *For any $\mathcal{T}_{\mathcal{X},1}, \mathcal{T}_{\mathcal{X},2} \in \mathscr{T}_\mathcal{X}$, the composition of the two transformations $\mathcal{T}_{\mathcal{X},2} \circ \mathcal{T}_{\mathcal{X},1}$ belongs to $\mathscr{T}_\mathcal{X}$.*

2. *$\mathscr{T}_\mathcal{X}$ contains identity transformation $\mathbf{1}_\mathcal{X}(x) = x$ for all $x \in \Omega_\mathcal{X}$.*

3. *For any $\mathcal{T}_\mathcal{X} \in \mathscr{T}_\mathcal{X}$, there exists $\mathcal{T}_\mathcal{X}^{-1} \in \mathscr{T}_\mathcal{X}$ such that $\mathcal{T}_\mathcal{X}^{-1} \circ \mathcal{T}_\mathcal{X} = \mathcal{T}_\mathcal{X} \circ \mathcal{T}_\mathcal{X}^{-1} = \mathbf{1}_\mathcal{X}$.*

If $\mathscr{T}_\mathcal{X}, \mathscr{T}_\mathcal{Y}$ are transformation groups on $\Omega_\mathcal{X}, \Omega_\mathcal{Y}$ respectively, then their direct product $\mathscr{T}_\mathcal{X} \times \mathscr{T}_\mathcal{Y}$ is a transformation group on $\Omega_\mathcal{X} \times \Omega_\mathcal{Y}$. The action of $(\mathcal{T}_\mathcal{X}, \mathcal{T}_\mathcal{Y}) \in \mathscr{T}_\mathcal{X} \times \mathscr{T}_\mathcal{Y}$ on $(x, y) \in \Omega_\mathcal{X} \times \Omega_\mathcal{Y}$ is $(\mathcal{T}_\mathcal{X}(x), \mathcal{T}_\mathcal{Y}(y))$. If there exists $\mathcal{T} = (\mathcal{T}_\mathcal{X}, \mathcal{T}_\mathcal{Y}) \in \mathscr{T}_\mathcal{X} \times \mathscr{T}_\mathcal{Y}$, such that

$$\mathcal{F}(\mathcal{T}(x, y)) = \mathcal{F}(\mathcal{T}_\mathcal{X}(x), \mathcal{T}_\mathcal{Y}(y)) = \mathcal{F}(x, y),$$

for all $(x, y) \in \Omega_\mathcal{X} \times \Omega_\mathcal{Y}$, then $\mathcal{T}$ maps a pair (x, y) to another pair $(\mathcal{T}_\mathcal{X}(x), \mathcal{T}_\mathcal{Y}(y))$ so that the two pairs cannot be distinguished by their images under $\mathcal{F}$. If a set of such $\mathcal{T}$'s form a subgroup of $\mathscr{T}_\mathcal{X} \times \mathscr{T}_\mathcal{Y}$, we have a *transformation group associated with the bilinear mapping $\mathcal{F}$*.

Definition 2.2.2. *A transformation group $\mathscr{T}$ on $\Omega_\mathcal{X} \times \Omega_\mathcal{Y}$ is said to be a transformation group associated with the bilinear mapping $\mathcal{F}$ if:*

1. *$\mathscr{T} \subset \mathscr{T}_\mathcal{X} \times \mathscr{T}_\mathcal{Y}$ is a subgroup of the direct product of two transformation groups $\mathscr{T}_\mathcal{X}$ and $\mathscr{T}_\mathcal{Y}$, on $\Omega_\mathcal{X}$ and $\Omega_\mathcal{Y}$, respectively.*

2. *For all $(x, y) \in \Omega_\mathcal{X} \times \Omega_\mathcal{Y}$ and for all $\mathcal{T} \in \mathscr{T}$, $\mathcal{F}(x, y) = \mathcal{F}(\mathcal{T}(x, y))$. Or equivalently, $\mathcal{F} = \mathcal{F} \circ \mathcal{T}$ for all $\mathcal{T} \in \mathscr{T}$.*

To enable an identifiability result up to a transformation group (see Section 2.3), the transformation group must capture *all* inherent ambiguities of the BIP. This motivates the following definition of the *ambiguity transformation group of the bilinear mapping.*

Definition 2.2.3. *A transformation group $\mathscr{T}$ on $\Omega_{\mathcal{X}} \times \Omega_{\mathcal{Y}}$ is said to be the ambiguity transformation group of the bilinear mapping $\mathcal{F}$ if $\mathscr{T}$ is the largest transformation group associated with $\mathcal{F}$, i.e., if $\mathscr{T}$ contains all transformation groups associated with $\mathcal{F}$. A transformation $\mathcal{T}$ in the ambiguity transformation group $\mathscr{T}$ of the bilinear mapping $\mathcal{F}$ is said to be an equivalence transformation associated with $\mathcal{F}$.*

Next, we define an *equivalence class associated with the bilinear inverse problem.*

Definition 2.2.4. *Given the ambiguity transformation group $\mathscr{T}$ of the bilinear mapping $\mathcal{F}$ on $\Omega_{\mathcal{X}} \times \Omega_{\mathcal{Y}}$, and $(x_0, y_0) \in \Omega_{\mathcal{X}} \times \Omega_{\mathcal{Y}}$, the set*

$$[(x_0, y_0)]_{\mathscr{T}} = \{(x, y) \in \Omega_{\mathcal{X}} \times \Omega_{\mathcal{Y}} : (x, y) = \mathcal{T}(x_0, y_0) \textit{ for some } \mathcal{T} \in \mathscr{T}\}$$

is called the equivalence class of (x_0, y_0) associated with the bilinear inverse problem in (1.1)*. In group theory terminology, $[(x_0, y_0)]_{\mathscr{T}}$ is called the orbit of (x_0, y_0) under the action of $\mathscr{T}$.*

Definition 2.2.5. *Given the ambiguity transformation group $\mathscr{T}$ of the bilinear mapping $\mathcal{F}$ on $\Omega_{\mathcal{X}} \times \Omega_{\mathcal{Y}}$, and $x_0 \in \Omega_{\mathcal{X}}$, the set*

$$[x_0]^L_{\mathscr{T}} = \{x \in \Omega_{\mathcal{X}} : \exists y_0, y \in \Omega_{\mathcal{Y}}, \ \textit{s.t.} \ (x, y) \in [(x_0, y_0)]_{\mathscr{T}}\}$$

is called the left equivalence class of x_0.

Similarly, given the ambiguity transformation group $\mathscr{T}$ of the bilinear mapping $\mathcal{F}$ on $\Omega_{\mathcal{X}} \times \Omega_{\mathcal{Y}}$, and $y_0 \in \Omega_{\mathcal{Y}}$, the set

$$[y_0]^R_{\mathscr{T}} = \{y \in \Omega_{\mathcal{Y}} : \exists x_0, x \in \Omega_{\mathcal{X}}, \ \textit{s.t.} \ (x, y) \in [(x_0, y_0)]_{\mathscr{T}}\}$$

is called the right equivalence class of y_0.

The definition of a transformation group guarantees that the relation between elements in an orbit satisfies reflexivity, transitivity and symmetry.

Therefore, an orbit is an equivalence class. If $\mathscr{T}$ is the ambiguity transformation group of the bilinear mapping $\mathcal{F}$, then all the elements in the equivalence class $[(x_0, y_0)]_{\mathscr{T}}$ share the same image under $\mathcal{F}$. Therefore, they are equivalent solutions to the bilinear inverse problem in (1.1). In fact, under some mild conditions on the bilinear mapping, Definitions 2.2.2 and 2.2.3 have additional implications.

Proposition 2.2.6. *Assume that the bilinear mapping $\mathcal{F}$ has no non-trivial left annihilator of $\Omega_{\mathcal{Y}}$, i.e., if $\mathcal{F}(x_0, y) = 0$ for all $y \in \Omega_{\mathcal{Y}}$, then $x_0 = 0$. Then every equivalence transformation $\mathcal{T} = (\mathcal{T}_{\mathcal{X}}, \mathcal{T}_{\mathcal{Y}}) \in \mathscr{T}$ satisfies the following:*

- *If $0 \in \Omega_{\mathcal{X}}$, then $\mathcal{T}_{\mathcal{X}}(0) = 0$.*
- *For $x_1, x_2 \in \Omega_{\mathcal{X}}$ and scalars a_1, a_2, if $a_1 x_1 + a_2 x_2 \in \Omega_{\mathcal{X}}$, then*

$$\mathcal{T}_{\mathcal{X}}(a_1 x_1 + a_2 x_2) = a_1 \mathcal{T}_{\mathcal{X}}(x_1) + a_2 \mathcal{T}_{\mathcal{X}}(x_2).$$

  If $\Omega_{\mathcal{X}}$ is a linear vector space, then $\mathcal{T}_{\mathcal{X}}$ is a linear transformation.

Similarly, assume that the bilinear mapping $\mathcal{F}$ has no non-trivial right annihilator of $\Omega_{\mathcal{X}}$, i.e., if $\mathcal{F}(x, y_0) = 0$ for all $x \in \Omega_{\mathcal{X}}$, then $y_0 = 0$. Then every equivalence transformation $\mathcal{T} = (\mathcal{T}_{\mathcal{X}}, \mathcal{T}_{\mathcal{Y}}) \in \mathscr{T}$ satisfies the following:

- *If $0 \in \Omega_{\mathcal{Y}}$, then $\mathcal{T}_{\mathcal{Y}}(0) = 0$.*
- *For $y_1, y_2 \in \Omega_{\mathcal{Y}}$ and scalars b_1, b_2, if $b_1 y_1 + b_2 y_2 \in \Omega_{\mathcal{Y}}$, then*

$$\mathcal{T}_{\mathcal{Y}}(b_1 y_1 + b_2 y_2) = b_1 \mathcal{T}_{\mathcal{Y}}(y_1) + b_2 \mathcal{T}_{\mathcal{Y}}(y_2).$$

  If $\Omega_{\mathcal{Y}}$ is a linear vector space, then $\mathcal{T}_{\mathcal{Y}}$ is a linear transformation.

Proof. Due to the symmetry, we only need to prove the results for $\mathcal{T}_{\mathcal{X}}$.

If $0 \in \Omega_{\mathcal{X}}$, then $\mathcal{F}(\mathcal{T}_{\mathcal{X}}(0), y) = \mathcal{F}(\mathcal{T}(0, \mathcal{T}_{\mathcal{Y}}^{-1}(y))) = \mathcal{F}(0, \mathcal{T}_{\mathcal{Y}}^{-1}(y)) = 0$ for all $y \in \Omega_{\mathcal{Y}}$. By assumption, there is no non-trivial left annihilator of $\Omega_{\mathcal{Y}}$. Therefore, $\mathcal{T}_{\mathcal{X}}(0) = 0$.

If $a_1x_1 + a_2x_2 \in \Omega_{\mathcal{X}}$, then

$$\begin{aligned}
&\mathcal{F}(\mathcal{T}_{\mathcal{X}}(a_1x_1 + a_2x_2), y)\\
=&\mathcal{F}(\mathcal{T}(a_1x_1 + a_2x_2, \mathcal{T}_{\mathcal{Y}}^{-1}(y)))\\
=&\mathcal{F}(a_1x_1 + a_2x_2, \mathcal{T}_{\mathcal{Y}}^{-1}(y))\\
=&a_1\mathcal{F}(x_1, \mathcal{T}_{\mathcal{Y}}^{-1}(y)) + a_2\mathcal{F}(x_2, \mathcal{T}_{\mathcal{Y}}^{-1}(y))\\
=&a_1\mathcal{F}(\mathcal{T}_{\mathcal{X}}(x_1), y) + a_2\mathcal{F}(\mathcal{T}_{\mathcal{X}}(x_2), y)\\
=&\mathcal{F}(a_1\mathcal{T}_{\mathcal{X}}(x_1) + a_2\mathcal{T}_{\mathcal{X}}(x_2), y).
\end{aligned}$$

Then $\mathcal{F}(\mathcal{T}_{\mathcal{X}}(a_1x_1 + a_2x_2) - (a_1\mathcal{T}_{\mathcal{X}}(x_1) + a_2\mathcal{T}_{\mathcal{X}}(x_2)), y) = 0$ for all $y \in \Omega_{\mathcal{Y}}$. There is no non-trivial left annihilator of $\Omega_{\mathcal{Y}}$. Hence $\mathcal{T}_{\mathcal{X}}(a_1x_1 + a_2x_2) = a_1\mathcal{T}_{\mathcal{X}}(x_1) + a_2\mathcal{T}_{\mathcal{X}}(x_2)$, and $\mathcal{T}_{\mathcal{X}}$ is a linear transformation if $\Omega_{\mathcal{X}}$ is a linear vector space. □

Bilinear mappings that arise in applications usually have no non-trivial left or right annihilators. Therefore, common equivalence transformations, such as scaling and shift, are linear transformations. However, there are examples where equivalence transformations are nonlinear (cf. Appendix A.1).

Before proceeding to identifiability, let us consider the following blind deconvolution problem as a concrete example. The measurement $z = x_0 \circledast y_0 \in \mathbb{C}^n$ is the circular convolution of two vectors.

$$\begin{aligned}
\text{find}\quad & (x, y),\\
\text{s.t.}\quad & x \circledast y = z,\\
& x \in \mathbb{C}^n,\ y \in \mathbb{C}^n.
\end{aligned}$$

Define transformation groups $\mathscr{T}_{\mathcal{X}}, \mathscr{T}_{\mathcal{Y}}$ on $\mathcal{X} = \mathcal{Y} = \mathbb{C}^n$:

$$\mathscr{T}_{\mathcal{X}} = \mathscr{T}_{\mathcal{Y}} = \{\mathcal{T}_{\mathbb{C}^n} : \mathcal{T}_{\mathbb{C}^n}(x) = \sigma S_\ell(x),\ \text{for some } \sigma \neq 0 \text{ and } \ell \in \mathbb{Z}\},$$

where the linear transformation S_ℓ is the circular shift by ℓ, defined as follows. If $x = S_\ell(x_0)$, then $x^{(j)} = x_0^{(k)}$ for all $1 \leq j, k \leq n$ where $j - k = \ell$ (modulo n). Then the following subgroup $\mathscr{T} \subset \mathscr{T}_{\mathcal{X}} \times \mathscr{T}_{\mathcal{Y}}$ is a transformation group associated with circular convolution:

$$\mathscr{T} = \left\{ \mathcal{T} : \mathcal{T}(x,y) = \left(\sigma S_\ell(x), \frac{1}{\sigma} S_{-\ell}(y) \right),\ \text{for some } \sigma \neq 0 \text{ and } \ell \in \mathbb{Z} \right\}. \tag{2.2}$$

Note that $\mathscr{T}$ is a transformation group associated with circular convolution, and a subgroup of $\mathscr{T}_{\mathcal{X}} \times \mathscr{T}_{\mathcal{Y}}$. However, it is not separable, i.e., it cannot be written as the direct product of two transformation groups. Furthermore, $\mathscr{T}$ is not the ambiguity transformation group, because it does not capture all the ambiguities of the above blind deconvolution problem. For example, there exist non-trivial vectors $u, v \in \mathbb{C}^n$ such that $u \circledast v$ is the Kronecker delta. Thus, $(x \circledast u, y \circledast v)$ is an equivalent pair of (x, y). The set of such transformations is not contained in $\mathscr{T}$.

2.3 Identifiability up to a Transformation Group

The concept of identifiability should be generalized to allow unique recovery up to the ambiguity transformation group. If the equivalence class containing the solution can be uniquely identified, the solution is considered identifiable.

Definition 2.3.1. *In the constrained BIP, the solution* (x_0, y_0) *in which* $x_0 \neq 0, y_0 \neq 0$ *is said to be identifiable up to a transformation group* $\mathscr{T}$*, if every solution* (x, y) *satisfies that* $(x, y) = \mathcal{T}(x_0, y_0)$ *for some* $\mathcal{T} \in \mathscr{T}$*, or equivalently,* $(x, y) \in [(x_0, y_0)]_{\mathscr{T}}$.

In general, the ambiguity transformation group for a certain BIP may not be known a priori. It may require some insight to capture all the ambiguities inherent in the problem. However, we can tell whether or not a given transformation group is the ambiguity transformation group by checking the identifiability. If there exists an identifiability result up to this transformation group, it has to be the largest. If the constraint sets $\Omega_{\mathcal{X}}$ and $\Omega_{\mathcal{Y}}$ are closed under scalar multiplication, then one can start by checking the group of scaling transformations defined in (2.1). For some BIPs, the ambiguities go beyond scaling ambiguity. Hence we have to choose larger transformation groups. An example is BGPC with a joint sparsity constraint (Section 3.4.1).

We derive a necessary and sufficient condition for identifiability in Theorem 2.3.2, and a more intuitive sufficient condition in Corollary 2.3.3. Here is how we interpret these results: In order to prove that certain conditions are sufficient to guarantee identifiability up to a transformation group, it suffices to first show that x_0 can be identified up to the transformation group; and then show that once x_0 is identified and substituted in the problem, y_0 can be identified. By the symmetry of the problem, we can derive another sufficient condition by switching the roles of x_0 and y_0.

Theorem 2.3.2. *In the constrained BIP, the pair (x_0, y_0) $(x_0 \neq 0, y_0 \neq 0)$ is identifiable up to $\mathscr{T}$ if and only if the following two conditions are met:*

1. *If $\mathcal{F}(x, y) = \mathcal{F}(x_0, y_0)$, then $x \in [x_0]^L_{\mathscr{T}}$.*
2. *If $\mathcal{F}(x_0, y) = \mathcal{F}(x_0, y_0)$, then $(x_0, y) \in [(x_0, y_0)]_{\mathscr{T}}$.*

Proof. To prove sufficiency, we suppose Conditions 1 and 2 are met. Let $\mathcal{F}(x, y) = \mathcal{F}(x_0, y_0)$ for nonzero x_0, y_0. Then, by Condition 1, $x \in [x_0]^L_{\mathscr{T}}$. Hence, there exists $\mathcal{T}_1 = (\mathcal{T}_{\mathcal{X},1}, \mathcal{T}_{\mathcal{Y},1}) \in \mathscr{T}$ such that $x = \mathcal{T}_{\mathcal{X},1}(x_0)$. Therefore $\mathcal{F}(x_0, y_0) = \mathcal{F}(x, y) = \mathcal{F}(\mathcal{T}_1^{-1}(x, y)) = \mathcal{F}(x_0, \mathcal{T}_{\mathcal{Y},1}^{-1}(y))$. By Condition 2, there exists $\mathcal{T}_2 \in \mathscr{T}$ such that $(x_0, \mathcal{T}_{\mathcal{Y},1}^{-1}(y)) = \mathcal{T}_2(x_0, y_0)$. Hence $(x, y) = \mathcal{T}_1(x_0, \mathcal{T}_{\mathcal{Y},1}^{-1}(y)) = \mathcal{T}_1 \circ \mathcal{T}_2(x_0, y_0)$, and (x_0, y_0) is identifiable up to $\mathscr{T}$.

Next we prove necessity. Given that (x_0, y_0) $(x_0 \neq 0, y_0 \neq 0)$ is identifiable up to $\mathscr{T}$, by Definition 2.3.1, if $\mathcal{F}(x, y) = \mathcal{F}(x_0, y_0)$, then $(x, y) \in [(x_0, y_0)]_{\mathscr{T}}$. The necessity of Conditions 1 and 2 follows. □

Corollary 2.3.3. *In the constrained BIP, the pair (x_0, y_0) $(x_0 \neq 0, y_0 \neq 0)$ is identifiable up to $\mathscr{T}$ if the following two conditions are met:*

1. *If $\mathcal{F}(x, y) = \mathcal{F}(x_0, y_0)$, then $x \in [x_0]^L_{\mathscr{T}}$.*
2. *If $\mathcal{F}(x_0, y) = \mathcal{F}(x_0, y_0)$, then $y = y_0$.*

Furthermore, if $\mathcal{F}$ has no non-trivial right annihilator of $\Omega_{\mathcal{X}}$, and for $(\mathcal{T}_{\mathcal{X}}, \mathcal{T}_{\mathcal{Y}}) \in \mathscr{T}$, $\mathcal{T}_{\mathcal{X}}(x_0) = x_0$ only if $\mathcal{T}_{\mathcal{X}} = \mathbf{1}_{\mathcal{X}}$, then the sufficient conditions above are also necessary.

Proof. Given that $y = y_0$, we have that $(x_0, y) = \mathbf{1}(x_0, y_0)$ and hence $(x_0, y) \in [(x_0, y_0)]_{\mathscr{T}}$. Therefore, condition 2 in Corollary 2.3.3 is more demanding than that of Theorem 2.3.2. Sufficiency follows.

The necessity of condition 1 also follows from Theorem 2.3.2. Next we show that with the extra assumptions, condition 2 is also necessary. Given that (x_0, y_0) $(x_0 \neq 0, y_0 \neq 0)$ is identifiable up to $\mathscr{T}$, by Theorem 2.3.2, if $\mathcal{F}(x_0, y) = \mathcal{F}(x_0, y_0)$, then there exists $\mathcal{T} = (\mathcal{T}_\mathcal{X}, \mathcal{T}_\mathcal{Y}) \in \mathscr{T}$ such that $(x_0, y) = \mathcal{T}(x_0, y_0)$. The first argument $\mathcal{T}_\mathcal{X}(x_0) = x_0$, by the extra assumption, $\mathcal{T}_\mathcal{X} = \mathbf{1}_\mathcal{X}$. Now, for all $(x_1, y_1) \in \Omega_\mathcal{X} \times \Omega_\mathcal{Y}$, $\mathcal{F}(x_1, y_1) = \mathcal{F}(\mathcal{T}(x_1, y_1)) = \mathcal{F}(\mathbf{1}_\mathcal{X}(x_1), \mathcal{T}_\mathcal{Y}(y_1)) = \mathcal{F}(x_1, \mathcal{T}_\mathcal{Y}(y_1))$, or equivalently, $\mathcal{F}(x_1, y_1 - \mathcal{T}_\mathcal{Y}(y_1)) = 0$. By the extra assumption that $\mathcal{F}$ has no non-trivial right annihilator of $\Omega_\mathcal{X}$, $y_1 - \mathcal{T}_\mathcal{Y}(y_1) = 0$ for all $y_1 \in \Omega_\mathcal{Y}$, or equivalently, $\mathcal{T}_\mathcal{Y} = \mathbf{1}_\mathcal{Y}$. Therefore, $y = \mathcal{T}_\mathcal{Y}(y_0) = y_0$, and condition 2 is necessary. □

The extra assumptions in Corollary 2.3.3 are usually satisfied, which means that Condition 2 is usually also necessary. Indeed, most bilinear mappings that arise in applications have no non-trivial annihilators. The assumption that "$\mathcal{T}_\mathcal{X}(x_0) = x_0$ only if $\mathcal{T}_\mathcal{X} = \mathbf{1}_\mathcal{X}$" is also true in many scenarios. For example, if $\mathcal{T}_\mathcal{X}$ is scaling by a nonzero complex number and $\mathcal{T}_\mathcal{X}(x_0) = x_0$ for some nonzero x_0, then $\mathcal{T}_\mathcal{X}$ has to be identity. However, there are examples for which Corollary 2.3.3 is not necessary (cf. Appendix A.1).

Later in this thesis, we repeatedly apply Corollary 2.3.3 to various scenarios of the blind gain and phase calibration problem and derive sufficient conditions for identifiability up to transformation groups.

CHAPTER 3

IDENTIFIABILITY IN BLIND GAIN AND PHASE CALIBRATION

3.1 Notations

We state the notations that will be used throughout the chapter. We use upper-case letters A, X and Y to denote matrices, and lower-case letters to denote vectors. The diagonal matrix whose diagonal entries are the entries of vector λ is denoted by $\mathrm{diag}(\lambda)$. We use I to denote the identity matrix and F to denote the normalized 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. A vector is said to be non-vanishing if *all* its entries are nonzero.

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 $\{1, 2, \cdots, n\}$. We use $|J|$ to denote the cardinality of J, and J^c to denote its complement. We use superscript letters to denote subvectors or submatrices. Thus, $x^{(J)}$ represents the subvector of x consisting of the entries indexed by J. The scalar $x^{(j)}$ represents the jth entry of x. The submatrix $A^{(J,K)}$ has size $|J| \times |K|$ and consists of the entries indexed by $J \times K$. The vector $A^{(:,k)}$ represents the kth column of the matrix A. The colon notation is inherited from MATLAB.

We use $./$ and $\odot$ to denote entrywise division and entrywise product, respectively. Circular convolution is denoted by $\circledast$. The direct sum of two subspaces is denoted by $\oplus$. The Kronecker product of two matrices is denoted by $\otimes$. The row space and column space of a matrix are denoted by $\mathcal{R}(\cdot)$ and $\mathcal{C}(\cdot)$, respectively.

3.2 Problem Statement

Blind gain and phase calibration (BGPC) is the following constrained BIP 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, $\Phi \in \Omega_\Phi \subset \mathbb{C}^{n\times N}$ 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, joint sparsity, or sparsity constraints. In all three 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.

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

Next, we elaborate on the three scenarios considered in this chapter:

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

In inverse rendering [26], the columns of $Y = \mathrm{diag}(\lambda)\Phi$ 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. The columns of A are the first several spherical harmonics extracted from the 3D

[1] Under a subspace constraint, A is required to have full column rank. Under a joint sparsity or sparsity constraint, A is required to satisfy the spark condition [41].

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

model of the object. They form a basis of the low-dimensional subspace in which the intensity maps reside.

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

$$F^* Y^{(:,j)} = \frac{1}{\sqrt{n}}(F^*\lambda) \circledast (F^*\Phi^{(:,j)}).$$

The vector λ represents the DFT of the signal, and columns of Φ represent the DFT of the channels. The columns of F^*A form a basis for the low-dimensional subspace in which 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 $F^*A = I^{(:,J)}$. 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 $F^*A = E$. In this case, the vector λ represents the DFT of the channel.

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

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

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

In sensor array processing with uncalibrated sensors, the vector λ represents unknown gain and phase for the sensors, and the columns of Φ represent array snapshots captured at different time instants. If the direction of arrival (DOA) is discretized using a grid, then each column of A represents the array response of one direction on the grid. With only s unknown sources, each column of Φ is the superposition of the same s columns of A. Hence the columns of the source matrix X are jointly s-sparse.

In synthetic aperture radar (SAR) autofocus [28], which is a special multichannel blind deconvolution problem, 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 we extend the coverage of the image 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 are zeros. Thus, the SAR image X can be modeled as a matrix with jointly sparse columns.

(III) *Sparsity constraints.* The matrix Φ is sparse over a dictionary A, where A is a square matrix ($n = m$) or a fat matrix ($n < m$). The constraint set $\Omega_{\mathcal{X}}$ is

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

A matrix X with sparse columns can be considered as a special case of this scenario.

Consider the following multichannel blind deconvolution problem. An acoustic signal is transmitted under reverberant conditions and recorded by a microphone array. The DFT of the signal is λ, $A = F$ is the DFT matrix, each column of $\Phi = AX$ is the DFT of the channel of a corresponding microphone, and the corresponding column of X is a sparse multipath channel that contains nonzero values at a few locations.

3.3 BGPC with a Subspace Constraint

In this section, we consider the identifiability of the BGPC problem with a subspace constraint. The measurement in the following problem is $Y = \mathrm{diag}(\lambda_0) A X_0$. The known matrix $A \in \mathbb{C}^{n \times m}$ is tall ($n > m$). The columns of $\Phi = AX$ reside in a low-dimensional subspace. The constraint sets are $\Omega_\Lambda = \mathbb{C}^n$ and $\Omega_{\mathcal{X}} = \mathbb{C}^{m \times N}$, hence the problem in unconstrained with respect to λ and X.

$$\begin{aligned} \text{find} \quad & (\lambda, X), \\ \text{s.t.} \quad & \mathrm{diag}(\lambda) A X = Y, \\ & \lambda \in \mathbb{C}^n, \ X \in \mathbb{C}^{m \times N}. \end{aligned}$$

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

3.3.1 Sufficient Condition

As was mentioned earlier, the BGPC problem suffers from scaling ambiguity. The ambiguity transformation group is defined as follows:

$$\mathscr{T} = \{\mathcal{T} : \mathcal{T}(\lambda, X) = (\sigma\lambda, \frac{1}{\sigma}X),\ \text{for some nonzero } \sigma \in \mathbb{C}\}. \tag{3.1}$$

Next, we investigate identifiability up to scaling within the framework of Section 2. By applying Corollary 2.3.3, we provide an alternative proof for the results by Nguyen et al. [26]. We need the following definition and lemma (see Appendix B.1 for the proof).

Definition 3.3.1. *The row space of a matrix $A \in \mathbb{C}^{n\times m}$ is said to be decomposable if there exists a non-empty proper subset (neither the empty set nor the universal set) $J \subset \{1, 2, \cdots, n\}$ and its complement J^c such that $\mathcal{R}(A) = \mathcal{R}(A^{(J,:)}) \oplus \mathcal{R}(A^{(J^c,:)})$.*

Lemma 3.3.2.

1. *If A has full row rank, then the row space of A is decomposable.*
2. *If $A \in \mathbb{C}^{n\times m}$ has full column rank and its row space is not decomposable, then $n > m$.*
3. *The row space of A is not decomposable if and only if $\dim(\mathcal{R}(A)) < \dim(\mathcal{R}(A^{(J,:)})) + \dim(\mathcal{R}(A^{(J^c,:)}))$ for all non-empty proper subsets $J \subset \{1, 2, \cdots, n\}$.*

Nguyen et al. [26] referred to the property that "A has full column rank and its row space is not decomposable" as "nonseparable full rank". Here is our restatement of the identifiability result followed by an alternative proof.

Theorem 3.3.3. *In the BGPC problem with a subspace constraint, the pair $(\lambda_0, X_0) \in \mathbb{C}^n \times \mathbb{C}^{m\times N}$ is identifiable up to an unknown scaling if the following conditions are met:*

1. *Vector λ_0 is non-vanishing, i.e., all the entries of λ_0 are nonzero.*
2. *Matrix X_0 has full row rank.*
3. *Matrix A has full column rank and its row space is not decomposable.*

Proof. We apply Corollary 2.3.3 to the BGPC problem, and verify that the two conditions in the corollary are satisfied. First, since the vector λ_0 is non-vanishing and the matrix A has full column rank, $\mathrm{diag}(\lambda_0)A$ has full column rank. It follows that if $\mathrm{diag}(\lambda_0)AX_0 = \mathrm{diag}(\lambda_0)AX_1$, then $X_1 = X_0$. Hence, given λ_0, the recovery of X_0 is unique. This verifies Condition 2 in Corollary 2.3.3. To verify Condition 1, we only need to show that λ_0 is identifiable up to scaling.

We prove by contradiction. Suppose the opposite, that there exists (λ_1, X_1) such that $\mathrm{diag}(\lambda_0)AX_0 = \mathrm{diag}(\lambda_1)AX_1$ but $\lambda_1 \notin [\lambda_0]^{L}_{\mathscr{T}}$. Recall that all the entries of λ_0 are nonzero, A has full column rank and X_0 has full row rank. Therefore, $\mathrm{rank}(\mathrm{diag}(\lambda_0)AX_0) = \mathrm{rank}(\mathrm{diag}(\lambda_1)AX_1) = m$, and X_1 too has full row rank. Since the row space of A is not decomposable, there are no zero rows in A. Because X_0 and X_1 have full row rank, it follows that there are no zero rows in AX_0 or AX_1. The vector λ_0 is non-vanishing, hence λ_1 too is non-vanishing. Let $\gamma = \lambda_1./\lambda_0$ denote the entrywise ratio of λ_1 over λ_0, where $\gamma^{(j)} = \lambda_1^{(j)}/\lambda_0^{(j)} \neq 0, j = 1, 2, \cdots, n$. By the assumption that $\lambda_1 \notin [\lambda_0]^{L}_{\mathscr{T}}$, the entrywise ratio is not the repetition of the same number, i.e., there exist j_1, j_2 such that $\gamma^{(j_1)} \neq \gamma^{(j_2)}$. Let T denote the number of distinct values of $\gamma^{(j)}$. Create a partition of the index set $\{1, 2, \cdots, n\}$, denoted by $J_1, J_2, \cdots, J_T$, such that $\gamma^{(j)} = \gamma_t$ for all $j \in J_t$, $t = 1, 2, \cdots, T$. Note that $\gamma_1, \gamma_2, \cdots, \gamma_T$ are the distinct values of $\gamma^{(j)}$.

Consider the row spaces of A:

$$\mathcal{R}(A) = \sum_{t=1}^{T} \mathcal{R}(A^{(J_t,:)}). \tag{3.2}$$

Denote the dimension of $\mathcal{R}(A^{(J_t,:)})$ by m_t. Then there exists a subset $J_t^b \subset J_t$ such that $|J_t^b| = m_t$ and the rows of $A^{(J_t^b,:)}$ form a basis of $\mathcal{R}(A^{(J_t,:)})$. By the condition that the row space of A is not decomposable, the sum in (3.2) is not a direct sum, hence $m = \mathrm{rank}(A) < \sum_{t=1}^{T} m_t$. Furthermore, by (3.2), there exists a subset

$$J^b = \{j_1, j_2, \cdots, j_m\} \subset \bigcup_{t=1}^{T} J_t^b \subset \bigcup_{t=1}^{T} J_t = \{1, 2, \cdots, n\}$$

such that $|J^b| = m$ and the rows of $A^{(J^b,:)}$ form a basis of $\mathcal{R}(A)$. The set $(\bigcup_{t=1}^{T} J_t^b) \setminus J^b$ is not empty because $m < \sum_{t=1}^{T} m_t$. Without loss of generality, we may assume that there exists $j_0 \in J_1^b \setminus J^b$. The row $A^{(j_0,:)}$ can be written as a linear combination of the rows of $A^{(J^b,:)}$ and the representation is unique. We denote the representation by:

$$A^{(j_0,:)} = \alpha_{j_1} A^{(j_1,:)} + \alpha_{j_2} A^{(j_2,:)} + \cdots + \alpha_{j_m} A^{(j_m,:)}. \tag{3.3}$$

The rows of $A^{(J_1^b,:)}$ are linearly independent, and $j_0 \notin J_1^b \bigcap J^b$, hence $A^{(j_0,:)}$ cannot be written as a linear combination of the rows of $A^{(J_1^b \bigcap J^b,:)}$; there exists at least one nonzero term in the representation (3.3) corresponding to one of the rows of $A^{(J^b \setminus J_1^b,:)}$. Thus, without loss of generality, there exists $j_1 \in J^b \bigcap J_2^b$, such that $\alpha_{j_1} \neq 0$.

Recall that $\mathrm{rank}(\mathrm{diag}(\lambda_0)AX_0) = \mathrm{rank}(\mathrm{diag}(\lambda_1)AX_1) = m$, and X_0 and X_1 have full row rank m. Therefore, the column spaces satisfy:

$$\mathcal{C}(\mathrm{diag}(\lambda_0)A) = \mathcal{C}(\mathrm{diag}(\lambda_0)AX_0) = \mathcal{C}(\mathrm{diag}(\lambda_1)AX_1) = \mathcal{C}(\mathrm{diag}(\lambda_1)A).$$

Hence $\mathrm{rank}([\mathrm{diag}(\lambda_0)A, \mathrm{diag}(\lambda_1)A]) = m$. Defining matrix

$$B \coloneqq [A, \mathrm{diag}(\gamma)A] = [\mathrm{diag}(\lambda_0)]^{-1}[\mathrm{diag}(\lambda_0)A, \mathrm{diag}(\lambda_1)A].$$

We have that

$$\mathrm{rank}(B) = \mathrm{rank}([\mathrm{diag}(\lambda_0)A, \mathrm{diag}(\lambda_1)A]) = m. \tag{3.4}$$

Then we consider the row spaces of B. The dimension of the row space $\mathcal{R}(B^{(J_t,:)}) = \mathcal{R}([A^{(J_t,:)}, \gamma_t A^{(J_t,:)}])$ is also m_t, and the rows of $B^{(J_t^b,:)}$ form a basis of the above row space. The rows of $B^{(J^b,:)}$ form a linearly independent set of cardinality m. By (3.4), the rows of $B^{(J^b,:)}$ form a basis of $\mathcal{R}(B)$. The row $B^{(j_0,:)}$ can be can be written as a linear combination of the rows of $B^{(J^b,:)}$. We denote the representation by:

$$B^{(j_0,:)} = \beta_{j_1} B^{(j_1,:)} + \beta_{j_2} B^{(j_2,:)} + \cdots + \beta_{j_m} B^{(j_m,:)}. \tag{3.5}$$

Recall that $j_0 \in J_1^b$, $j_1 \in J_2^b$, hence $\gamma^{(j_0)} = \gamma_1$, and $\gamma^{(j_1)} = \gamma_2$. Using the definition of B, we rewrite (3.5) as:

$$A^{(j_0,:)} = \beta_{j_1} A^{(j_1,:)} + \beta_{j_2} A^{(j_2,:)} + \cdots + \beta_{j_m} A^{(j_m,:)}, \tag{3.6}$$
$$\gamma_1 A^{(j_0,:)} = \beta_{j_1} \gamma_2 A^{(j_1,:)} + \beta_{j_2} \gamma^{(j_2)} A^{(j_2,:)} + \cdots + \beta_{j_m} \gamma^{(j_m)} A^{(j_m,:)}. \tag{3.7}$$

Since the representation in (3.3) is unique, the representations in (3.6) and (3.7) must satisfy:

$$\beta_{j_1} = \beta_{j_1} \frac{\gamma_2}{\gamma_1} = \alpha_{j_1} \neq 0.$$

It follows that $\gamma_1 = \gamma_2$, which contradicts the assumption that γ_1 and γ_2 are distinct. Hence the assumption that $\lambda_1 \notin [\lambda_0]^L_{\mathscr{T}}$ is false, and λ_0 is identifiable up to an unknown scaling. □

For generic signals, we can show that Theorem 3.3.3 reduces to a simple condition (Corollary 3.3.4) on the dimensions n, m and N. We say that a property holds for almost all signals if the property holds for all signals but a set of measure zero.

Corollary 3.3.4. *In the BGPC problem with a subspace constraint, if $n > m$ and $N \geq m$, then (λ_0, X_0) is identifiable up to an unknown scaling for almost all $\lambda_0 \in \mathbb{C}^n$, almost all $X_0 \in \mathbb{C}^{m \times N}$ and almost all $A \in \mathbb{C}^{n \times m}$.*

Proof. Almost all $\lambda_0 \in \mathbb{C}^n$ are non-vanishing. If $N \geq m$, almost all $X_0 \in \mathbb{C}^{m \times N}$ have full row rank. If $n > m$, almost all $A \in \mathbb{C}^{n \times m}$ have full column rank. Next we show that the row spaces of almost all A are not decomposable. For almost all A, the submatrices $A^{(J,:)}$ and $A^{(J^c,:)}$ have full rank for every non-empty proper subset $J \subset \{1, 2, \cdots, n\}$. Therefore, one of the following cases has to be true.

1. If $|J| < m$ and $|J^c| < m$, then for almost all A, $\dim(\mathcal{R}(A)) = m$, $\dim(\mathcal{R}(A^{(J,:)})) = |J|$, $\dim(\mathcal{R}(A^{(J^c,:)})) = |J^c|$. Hence for almost all A,

$$\dim(\mathcal{R}(A)) = m < n = |J| + |J^c| = \dim(\mathcal{R}(A^{(J,:)})) + \dim(\mathcal{R}(A^{(J^c,:)})).$$

2. If $|J| \geq m$, then for almost all A, $\dim(\mathcal{R}(A^{(J,:)})) = m$. Hence for almost all A,

$$\dim(\mathcal{R}(A)) = m < m + 1 \leq \dim(\mathcal{R}(A^{(J,:)})) + \dim(\mathcal{R}(A^{(J^c,:)})).$$

3. If $|J^c| \geq m$, then for almost all A, $\dim(\mathcal{R}(A^{(J^c,:)})) = m$. Hence for almost all A,

$$\dim(\mathcal{R}(A)) = m < 1 + m \leq \dim(\mathcal{R}(A^{(J,:)})) + \dim(\mathcal{R}(A^{(J^c,:)})).$$

Therefore, $\dim(\mathcal{R}(A)) < \dim(\mathcal{R}(A^{(J,:)})) + \dim(\mathcal{R}(A^{(J^c,:)}))$ for every non-empty proper subset $J \subset \{1, 2, \cdots, n\}$, establishing that the row spaces of almost all A are not decomposable. By Theorem 3.3.3, given that $N \geq m$ and $n > m$, the pair (λ_0, X_0) is identifiable up to an unknown scaling for almost all λ_0, X_0 and A. □

Corollary 3.3.4 shows that, in the BGPC problem with a subspace constraint, for almost all vectors λ_0, almost all tall matrices A and almost all fat matrices X_0, the solution (λ_0, X_0) is identifiable up to an unknown scaling.

3.3.2 Necessary Condition

Given that λ_0 is non-vanishing, Nguyen et al. [26] showed that "the row space of A is not decomposable" is necessary. Lacking, however, is a necessary condition for the sample complexity.

As we demonstrate in the next subsection by construction of counter-examples, the sample complexity $N \geq m$, as required by Theorem 3.3.3 implicitly and Corollary 3.3.4 explicitly, is not necessary. Instead, a necessary condition is suggested by heuristically counting the number of degrees of freedom and the number of measurements in $Y = \operatorname{diag}(\lambda)AX$. The numbers of free variables in λ and X are n and mN, respectively. The unknown scaling of λ and X is counted twice, hence 1 is subtracted yielding $n + mN - 1$ for the total number of degrees of freedom. The total number of measurements is nN. Heuristically, to achieve uniqueness, nN must be greater than or equal to $n + mN - 1$, which implies $N \geq \frac{n-1}{n-m}$. This turns out to be a valid necessary condition, as we now state and prove rigorously.

Proposition 3.3.5. *In the BGPC problem with a subspace constraint, if A has full column rank, and (λ_0, X_0) (with a non-vanishing λ_0) is identifiable up to scaling, then $N \geq \frac{n-1}{n-m}$.*

Proof. We show that if $N < \frac{n-1}{n-m}$, then the recovery cannot be unique. Let $A_\perp \in \mathbb{C}^{n\times(n-m)}$ denote a matrix whose columns form a basis for the ortho-complement of the column space of A. Hence $A_\perp^*$ is an annihilator of the column space of A. Consider the linear operator $\mathcal{G} : \mathbb{C}^n \to \mathbb{C}^{(n-m)\times N}$ defined by

$$\mathcal{G}(x) := A_\perp^* \operatorname{diag}(x) Y = A_\perp^* \operatorname{diag}(x) \operatorname{diag}(\lambda_0) A X_0.$$

We claim that every non-vanishing null vector of $\mathcal{G}$ produces a solution to the BGPC problem. Indeed, if $x \in \mathcal{N}(\mathcal{G})$, then

$$A_\perp^* \operatorname{diag}(x) \operatorname{diag}(\lambda_0) A X_0 = 0,$$

hence the columns in $\operatorname{diag}(x)\operatorname{diag}(\lambda_0)AX_0$ must reside in the column space of A. Let

$$\operatorname{diag}(x) \operatorname{diag}(\lambda_0) A X_0 = A X_1.$$

If x is non-vanishing, then (λ_1, X_1) is a solution, where λ_1 is the entrywise inverse of x.

Let x_0 denote the entrywise inverse of λ_0, then $x_0 \in \mathcal{N}(\mathcal{G})$. There are $N(n-m)$ equations in $\mathcal{G}(x) = 0$. If $N < \frac{n-1}{n-m}$, i.e., $N(n-m) \le n-2$, the dimension of the null space $\mathcal{N}(\mathcal{G})$ is at least 2. Hence, there exists another vector $x_1 \in \mathcal{N}(\mathcal{G})$ such that x_0, x_1 are linearly independent. Let α be a complex number such that $0 < |\alpha| < \frac{1}{\|\lambda_0\|_\infty \|x_1\|_\infty}$. Then $x_0 + \alpha x_1 \in \mathcal{N}(\mathcal{G})$ is non-vanishing, because the entries of $x_0 + \alpha x_1$ satisfy that

$$\left|x_0^{(j)} + \alpha x_1^{(j)}\right| \ge \left|x_0^{(j)}\right| - |\alpha| \left|x_1^{(j)}\right| \ge \frac{1}{\|\lambda_0\|_\infty} - |\alpha| \, \|x_1\|_\infty > 0, \quad \forall j \in \{1, 2, \cdots, n\}.$$

This null vector is not a scaled version of x_0. Hence there exists a solution that does not belong to the equivalence class $[(\lambda_0, X_0)]_{\mathscr{T}}$. Therefore, $N \ge \frac{n-1}{n-m}$ is necessary. □

The two sample complexities $N \ge m$ and $N \ge \frac{n-1}{n-m}$ coincide when $m = 1$ or $m = n-1$. The gap between these two sample complexities when $2 \le m \le n-2$ is analyzed next.

3.3.3 Gap Between the Sufficient and the Necessary Conditions

The sample complexity in the sufficient condition is $N \geq m$, which can be represented by the region above a line segment. The sample complexity in the necessary condition is $N \geq \frac{n-1}{n-m}$, which can be represented by the region above part of a hyperbola. The gap between the two sample complexities is the region between the line segment and the hyperbola (cf. Figure 3.1).

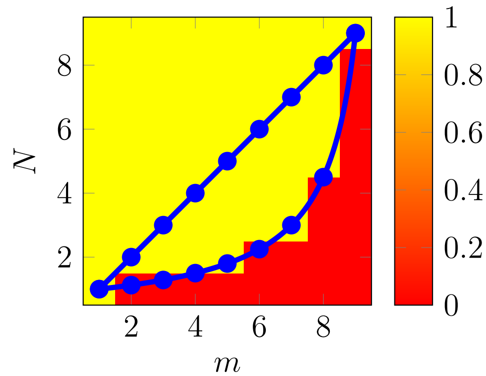


Figure 3.1: The sample complexities for BGPC with a subspace constraint, and the ratio of identifiable pairs generated randomly.

To explore this gap, we wish to determine whether (λ_0, X_0), in BGPC with a subspace constraint, is identifiable up to scaling. We now show that this can be done by Algorithm 1. Given A that has full column rank and $Y = \text{diag}(\lambda_0)AX_0$ that has no zero rows, Algorithm 1 returns a Boolean value indicating whether (λ_0, X_0) is identifiable up to scaling.

Algorithm 1 Identifiability of the BGPC problem with a subspace constraint

---

input: A,Y **output**: identifiability of (λ_0, X_0)
$[Q, R] = \texttt{qr}(A)$ {`QR decomposition of` A}
$A_\perp \leftarrow Q^{(:,m+1:n)}$
$G \leftarrow \Big[[\text{diag}(Y^{(:,1)})]^* A_\perp \quad [\text{diag}(Y^{(:,2)})]^* A_\perp \quad \cdots \quad [\text{diag}(Y^{(:,N)})]^* A_\perp\Big]^*$
if $\text{rank}(G) \leq n - 2$ **then**
  return False
else
  return True
end if

Proposition 3.3.6. *Given A that has full column rank and $Y = \operatorname{diag}(\lambda_0)AX_0$ that has no zero rows, the pair (λ_0, X_0) is identifiable up to scaling if Algorithm 1 returns True, and not identifiable up to scaling if Algorithm 1 returns False.*

Proof. The columns of $A_\perp$ form a basis for the ortho-complement of the column space of A, hence $A_\perp^*$ is an annihilator of the column space of A. The matrix $G \in \mathbb{C}^{N(n-m)\times n}$ satisfies that $Gx = \operatorname{vec}(A_\perp^* \operatorname{diag}(x)Y)$. Given Y that has no zero rows, any solution to the BGPC problem (λ, X) satisfies that λ is non-vanishing, and that the entrywise inverse of λ is a null vector of G. On the other hand, as argued in the proof of Proposition 3.3.5, any non-vanishing null vector of G produces a solution (λ, X).

If Algorithm 1 returns True, then $\operatorname{rank}(G) \geq n-1$. Given a solution (λ_0, X_0), G has at least one null vector x_0, which is the entrywise inverse of λ_0. Hence $\operatorname{rank}(G) = n-1$. All the null vectors of G reside in the one-dimensional subspace spanned by x_0. Therefore λ in any solution is a scaled version of λ_0, or $\lambda \in [\lambda_0]^L_{\mathscr{T}}$. Given non-vanishing λ_0 and A with full column rank, $\operatorname{diag}(\lambda_0)A$ has full column rank and the recovery of X_0 has to be unique. By Corollary 2.3.3, (λ_0, X_0) is identifiable up to scaling.

If Algorithm 1 returns False, then $\operatorname{rank}(G) \leq n-2$. By the proof of Proposition 3.3.5, (λ_0, X_0) is not identifiable. □

We now use Algorithm 1 to construct counter-examples demonstrating that the sufficient condition in Theorem 3.3.3 is not necessary. Let $n = 10$, $1 \leq m \leq 9$, and $1 \leq N \leq 9$. The entries of $\lambda_0 \in \mathbb{R}^n$ and $X_0 \in \mathbb{R}^{m\times N}$ are generated as iid Gaussian random variables $N(0, 1)$. The matrix $A \in \mathbb{R}^{n\times m}$ is the first m columns from an $n \times n$ random orthogonal matrix. Then $A_\perp$ comprises the last $(n-m)$ columns from the same random orthogonal matrix. We use Algorithm 1 to determine whether or not (λ_0, X_0) is identifiable up to scaling. For every value of m and N, the numerical experiment is repeated 100 times independently. The ratio of identifiable pairs as a function of (m, N) is shown in Figure 3.1. As is expected, the solution (λ_0, X_0) is identifiable when $N \geq m$, and is not identifiable when $N < \frac{n-1}{n-m}$. Meanwhile, when $\frac{n-1}{n-m} \leq N < m$, the ratio of identifiable pairs is 1. Therefore, $N \geq m$ is not necessary.

On the other hand, the necessary condition in Proposition 3.3.5 is not sufficient. For example, if $n = 8$, $m = 4$ and $\frac{n-1}{n-m} < N = 2 < m$, let A be the structured matrix

$$A = \begin{bmatrix} A_1 & \mathrm{diag}(\gamma)A_1 \end{bmatrix},$$

where $A_1 \in \mathbb{C}^{8\times 2}$, $\gamma \in \mathbb{C}^8$. There exists an A_1 and a γ such that the matrix A has full column rank and the row space of A is not decomposable. For example, let $A_1 = 2\sqrt{2}F^{(:,1:2)}$ and $\gamma = 2\sqrt{2}F^{(:,3)}$, then $A = 2\sqrt{2}F^{(:,1:4)}$. However, (λ_0, X_0) is not identifiable and $\lambda_0 A X_0 = \lambda_1 A X_1$, if

$$X_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad X_1 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix},$$

$$\lambda_0 = \gamma \odot \lambda_1.$$

However, according to the ratio of identifiable pairs shown in Figure 3.1, the unidentifiable case does not occur even once in 100 random trials. We have the following conjecture:

Conjecture 3.3.7. *In the BGPC problem with a subspace constraint, if $n > m$ and $N \geq \frac{n-1}{n-m}$, then (λ_0, X_0) is identifiable up to an unknown scaling for almost all $\lambda_0 \in \mathbb{C}^n$, almost all $X_0 \in \mathbb{C}^{m\times N}$ and almost all $A \in \mathbb{C}^{n\times m}$.*

If the above conjecture is true, the necessary condition $N \geq \frac{n-1}{n-m}$ is tight except for a set of measure zero.

3.4 BGPC with a Joint Sparsity Constraint

Here we consider the identifiability in the BGPC problem with a joint sparsity constraint:

$$\begin{aligned} \text{(P1)} \quad \text{find} \quad & (\lambda, X), \\ \text{s.t.} \quad & \mathrm{diag}(\lambda)AX = Y, \\ & \lambda \in \mathbb{C}^n,\ X \in \Omega_{\mathcal{X}} = \{X \in \mathbb{C}^{n\times N} : \text{the columns of } X \text{ are jointly } s\text{-sparse}\}. \end{aligned}$$

The measurement in the above problem is $Y = \mathrm{diag}(\lambda_0)AX_0$. We only consider the case where $A \in \mathbb{C}^{n\times n}$ is an invertible square matrix. The vector $\lambda_0 \in \mathbb{C}^n$ is non-vanishing. The columns of $X_0 \in \mathbb{C}^{n\times N}$ are jointly s-sparse

(X_0 has at most s nonzero rows). Unless otherwise stated, we assume that the sparsity level s is known a priori. However, if s is unknown, one can solve the following optimization problem instead:

$$\begin{aligned}
\text{(P2)} \quad \min_{(\lambda,X)}. \quad & \text{row-sparsity}(X), \\
\text{s.t.} \quad & \operatorname{diag}(\lambda) A X = Y, \\
& \lambda \in \mathbb{C}^n, \ X \in \mathbb{C}^{n \times N}.
\end{aligned}$$

In this section, we define ambiguities and transformation groups that depend on the matrix A. For two special cases of A, we give sufficient conditions for identifiability up to the ambiguity transformation groups.

3.4.1 Ambiguities and Transformation Groups

Geometrically, a joint sparsity constraint corresponds to a union of subspaces; hence, it is less restrictive than the previously discussed subspace constraint. This results in greater ambiguity in identifying a solution to BGPC with a joint sparsity constraint, than just the scaling ambiguity. In this case, to obtain identifiability results, we must choose the largest transformation group associated with the BIP, which captures *all* ambiguities inherent in the problem. In this section, we develop a procedure to do so.

A generalized permutation matrix is an invertible square matrix with exactly one nonzero entry in each row and each column. It preserves the joint sparsity structure. That is, if the columns of X_0 are jointly s-sparse and P is a generalized permutation matrix, then the columns of $X_1 = PX_0$ are also jointly s-sparse. Suppose there exists a vector $\gamma \in \mathbb{C}^n$ such that $P = A^{-1} \operatorname{diag}(\gamma) A$ is a generalized permutation matrix; then clearly γ has to be non-vanishing. Now, given a solution (λ_0, X_0) to the BGPC problem, there exist $\lambda_1 = \lambda_0./\gamma$ and $X_1 = PX_0 \in \Omega_{\mathcal{X}}$ such that

$$\operatorname{diag}(\lambda_1) A X_1 = \operatorname{diag}(\lambda_0) [\operatorname{diag}(\gamma)]^{-1} A A^{-1} \operatorname{diag}(\gamma) A X_0 = \operatorname{diag}(\lambda_0) A X_0.$$

This ambiguity is inevitable. To address this ambiguity, we define the set

$$\Gamma(A) = \{\gamma \in \mathbb{C}^n : A^{-1} \operatorname{diag}(\gamma) A \text{ is a generalized permutation matrix}\}, \tag{3.8}$$

and the ambiguity transformation group

$$\mathscr{T} = \{\mathcal{T} : \mathcal{T}(\lambda, X) = (\lambda./\gamma, A^{-1}\operatorname{diag}(\gamma)AX) \text{ for some } \gamma \in \Gamma(A)\}. \quad (3.9)$$

Then (λ_1, X_1) is in the equivalence class $[(\lambda_0, X_0)]_{\mathscr{T}}$.

Note that the set $\Gamma(A)$ depends on A. In particular, when A is the normalized DFT matrix $A = F \in \mathbb{C}^{n\times n}$, the matrix $F^* \operatorname{diag}(\gamma)F$ is a circulant matrix whose first column is $\frac{1}{\sqrt{n}}F^*\gamma$. The matrix $F^* \operatorname{diag}(\gamma)F$ is a generalized permutation matrix if and only if there is exactly one nonzero entry in $\frac{1}{\sqrt{n}}F^*\gamma$, which means that the circulant matrix $F^* \operatorname{diag}(\gamma)F$ is a scaled circular shift. Therefore,

$$\Gamma(F) = \left\{\gamma = \sigma\sqrt{n}F^{(:,k)} : \sigma \in \mathbb{C} \text{ is nonzero}, k \in \{1, 2, \cdots, n\}\right\}. \quad (3.10)$$

$$\mathscr{T} = \{\mathcal{T} : \mathcal{T}(\lambda, X) = (\lambda./\gamma, F^*\operatorname{diag}(\gamma)FX) \text{ for some } \gamma \in \Gamma(F)\}. \quad (3.11)$$

An equivalence transformation $\mathcal{T} \in \mathscr{T}$ defined in (3.11) is a complex exponential modulation of λ scaled by $\frac{1}{\sigma}$ and a circular shift of X scaled by σ. In MBD, if we shift the signal by $1-k$ and scale it by $\frac{1}{\sigma}$, and shift the channels by $k-1$ and scale them by σ, the outputs of the channels remain unchanged.

The ambiguity transformation groups for other choices of A can be figured out in a similar fashion. For more examples, please refer to Section 3.4.3 and to Appendix A.2.

3.4.2 Identifiability of Jointly Sparse Signals

In this section, we assume that $A = F$ is the DFT matrix and the columns of X are jointly s-sparse. In multichannel blind deconvolution, the non-vanishing vector λ_0 is the DFT of the signal and the jointly sparse columns of X_0 are the multiple channels. We derive a sufficient condition and a necessary condition for (λ_0, X_0) to be identifiable up to the transformation group defined in (3.11).

Sufficient Condition

We can prove a sufficient condition for identifiability up to the transformation group in (3.11) within the framework of Section 2 by again invoking Corollary

2.3.3. We need the following definition to state this sufficient condition.

Definition 3.4.1. *The index set $J = \{j_1, j_2, \cdots, j_s\} \subset \{1, 2, \cdots, n\}$ is said to be periodic with period ℓ (ℓ being an integer such that $0 < \ell < n$), if $J = \{j_1 + \ell, j_2 + \ell, \cdots, j_s + \ell\}$ (modulo n). The smallest integer ℓ with this property is called the fundamental period.*

The universal set $\{1, 2, \cdots, n\}$ is always periodic with period ℓ (ℓ being any integer from 1 to $n - 1$). The fundamental period is 1. For $n = 10$ and $s = 4$, the set $J = \{1, 2, 6, 7\}$ is periodic with fundamental period 5. Periodicity has the following property.

Remark 3.4.2. *If the set $J = \{j_1, j_2, \cdots, j_s\}$ is periodic with period ℓ, then the complement J^c, the flipped version $J^f = \{-j_1, -j_2, \cdots, -j_s\}$ (modulo n) and the shifted version $\{j_1 + k, j_2 + k, \cdots, j_s + k\}$ (modulo n) are all periodic with period ℓ.*

Here is the sufficient condition for the identifiability of the BGPC problem with DFT matrix and a joint sparsity constraint.

Theorem 3.4.3. *In the BGPC problem with DFT matrix and a joint sparsity constraint at sparsity level s, the pair $(\lambda_0, X_0) \in \mathbb{C}^n \times \Omega_{\mathcal{X}}$ is identifiable up to the transformation group $\mathscr{T}$ defined in (3.11) if the following conditions are met:*

1. *Vector λ_0 is non-vanishing.*
2. *Matrix X_0 has exactly s nonzero rows and rank s.*
3. *The joint support of the columns of X_0 is not periodic.*

Proof. First, given non-vanishing λ_0 and the DFT matrix F, the matrix $\mathrm{diag}(\lambda_0)F$ has full rank. If $\mathrm{diag}(\lambda_0)FX_0 = \mathrm{diag}(\lambda_0)FX_1$, then $X_1 = X_0$. Hence, given λ_0, the recovery of X_0 is unique. By Corollary 2.3.3, to complete the proof, we only need to show that λ_0 is identifiable up to the transformation group.

By assumption, the matrix X_0 has rank s and the joint support of the columns of X_0, denoted by $J = \{j_1, j_2, \cdots, j_s\}$, is not periodic. Given that $\mathrm{diag}(\lambda_0)FX_0 = \mathrm{diag}(\lambda_1)FX_1$, we show that $\lambda_1 \in [\lambda_0]^L_{\mathscr{T}}$. Now, the matrix X_0 has s linearly independent columns, $\mathrm{diag}(\lambda_0)F$ has full rank, hence the corresponding columns of X_1 are also linearly independent. Without loss of generality, we may assume that X_0 and X_1 only have s columns, which are linearly independent, by removing redundant columns at the same locations in both matrices. Then $X_0, X_1 \in \mathbb{C}^{n\times s}$ have full column rank s and exactly s nonzero rows. Because F has no zero entries, it follows that there are no zero rows in FX_0 or FX_1. The vector λ_0 is non-vanishing, hence λ_1 is also non-vanishing. We know that

$$P = F^*[\operatorname{diag}(\lambda_1)]^{-1}\operatorname{diag}(\lambda_0)F \tag{3.12}$$

is a circulant matrix and that $X_1 = PX_0$. Let $X_0^\dagger \in \mathbb{C}^{s\times n}$ denote the pseudo-inverse (also the left inverse) of X_0, and $X_{0\perp} \in \mathbb{C}^{n\times(n-s)}$ denote a matrix whose columns form a basis for the ortho-complement of the column space of X_0. Since X_0 has full column rank s and exactly s nonzero rows indexed by J, we may choose $X_0^\dagger$ such that its nonzero columns are indexed by J, and choose the columns of $X_{0\perp}$ to be the standard basis vectors $\{I^{(:,k)} : k \in J^c\}$. The matrix P as in $X_1 = PX_0$ satisfies

$$P = X_1X_0^\dagger + QX_{0\perp}^*, \tag{3.13}$$

where $Q \in \mathbb{C}^{n\times(n-s)}$ is a free matrix. Note that the nonzero columns of $QX_{0\perp}^*$ are indexed by J^c and the nonzero columns of $X_1X_0^\dagger$ are indexed by J. Hence $P^{(:,J)} = X_1X_0^{\dagger(:,J)}$. The submatrix $P^{(:,J)}$ has no more than s nonzero rows because X_1 has s nonzero rows.

We prove $\lambda_1 \in [\lambda_0]_{\mathscr{T}}^L$ by contradiction. Suppose that $\lambda_1 \notin [\lambda_0]_{\mathscr{T}}^L$. By (3.10) and (3.11), the entrywise ratio $\gamma = \lambda_0./\lambda_1 \notin \Gamma(F)$, which means that $\frac{1}{\sqrt{n}}F^*\gamma$, the first column of the circulant matrix P (as in (3.12)), has more than one nonzero entry. Denote the indices of the first two nonzero entries of $P^{(:,1)}$ by k_1 and k_2. By the structure of circulant matrices, the rows of $P^{(:,J)}$ indexed by the following two sets (interpreted modulo n) are nonzero:

$$\begin{aligned} K_1 &= \{k_1 + j_1 - 1, k_1 + j_2 - 1, \cdots, k_1 + j_s - 1\}, \\ K_2 &= \{k_2 + j_1 - 1, k_2 + j_2 - 1, \cdots, k_2 + j_s - 1\}. \end{aligned}$$

Note that $|K_1| = |K_2| = s$. Recall that $P^{(:,J)}$ has no more than s nonzero rows, hence $K_1 = K_2$. It follows that set K_1 is periodic with period $\ell =$

$|k_2 - k_1|$. By the property in Remark 3.4.2, the set J is also periodic with the same period, and we reach a contradiction. Therefore, the assumption that $\lambda_1 \notin [\lambda_0]^L_{\mathscr{T}}$ is false, and Condition 1 of Corollary 2.3.3 is satisfied - the vector λ_0 is identifiable up to the transformation group. $\square$

Corollary 3.4.4. *If $N \geq s$, then the conclusion of Theorem 3.4.3 holds for almost all $\lambda_0 \in \mathbb{C}^n$, and almost all $X_0 \in \mathbb{C}^{n\times N}$ that has s nonzero rows and non-periodic joint support.*

Proof. Almost all $\lambda_0 \in \mathbb{C}^n$ are non-vanishing. If $N \geq s$, then almost all $X_0 \in \mathbb{C}^{n\times N}$ with s nonzero rows have rank s. In addition, the joint support of X_0 is not periodic. Therefore, the conditions in Theorem 3.4.3 are met, and (λ_0, X_0) is identifiable up to the transformation group $\mathscr{T}$ defined in (3.11). $\square$

Corollary 3.4.4 shows that, in the BGPC problem with DFT matrix and a joint sparsity constraint, given that $N \geq s$, the identifiability of generic signals (λ_0, X_0) hinges on the joint support of X_0. If the joint support is non-periodic, (λ_0, X_0) is almost always identifiable. Other priors may imply non-periodicity. For example, if the joint support is a contiguous block, or if n and s are coprime, the joint support has to be non-periodic.

Corollary 3.4.5. *If $N \geq s$, then the conclusion of Theorem 3.4.3 holds for almost all $\lambda_0 \in \mathbb{C}^n$, and almost all $X_0 \in \mathbb{C}^{n\times N}$ that has s nonzero rows that are contiguous.*

Corollary 3.4.6. *If $N \geq s$, and n and s are coprime, then the conclusion of Theorem 3.4.3 holds for almost all $\lambda_0 \in \mathbb{C}^n$, and almost all $X_0 \in \mathbb{C}^{n\times N}$ that has s nonzero rows.*

Clearly, the coprimeness condition in Corollary 3.4.6 is satisfied for all $s < n$ if n is a prime number.

The above results are under the assumption that the sparsity level s is known a priori. If s is unknown, instead of solving the feasibility problem (P1), one can solve the optimization problem (P2). We have the following corollary, whose proof is almost identical to that of Theorem 3.4.3.

Corollary 3.4.7. *In the BGPC problem with DFT matrix and unknown sparsity level, the pair $(\lambda_0, X_0) \in \mathbb{C}^n \times \Omega_{\mathcal{X}}$ is the unique minimizer of (P2) up* *to the transformation group $\mathscr{T}$ defined in* (3.11)*, if the following conditions are met:*

1. *Vector λ_0 is non-vanishing.*

2. *Matrix X_0 has rank equal to the number of nonzero rows.*

3. *The joint support of the columns of X_0 is not periodic.*

We can derive row sparsity minimization analogs of Corollaries 3.4.4, 3.4.5 and 3.4.6 in a similar fashion. These results are omitted for the sake of brevity.

Necessary Condition

Given that λ_0 is non-vanishing, "the joint support of the columns of X_0 is not periodic" is necessary. We prove this by contraposition. We assume that the joint support of the columns of X_0 is periodic with period ℓ, and next show that (λ_0, X_0) is not identifiable up to the transformation group in (3.11). Let P be a circulant matrix whose first column has two nonzero entries $P^{(1,1)} = 1$ and $P^{(\ell+1,1)} = 2$. Thus, the DFT $\gamma = \sqrt{n}FP^{(:,1)}$ of the first column of P is non-vanishing. Let $\lambda_1 = \lambda_0./\gamma$ and $X_1 = PX_0$. Then P satisfies (3.12), and $\mathrm{diag}(\lambda_1)FX_1 = \mathrm{diag}(\lambda_0)FX_0$. Since P is not a generalized permutation matrix, X_1 is not a scaled and circularly shifted version of X_0. Hence (λ_0, X_0) is not identifiable up to the transformation group in (3.11).

The above necessary condition does not address the sample complexity. Like Proposition 3.3.5, we have the following necessary condition for the sample complexity.

Proposition 3.4.8. *In the BGPC problem with DFT matrix and a joint sparsity constraint, if (λ_0, X_0) (λ_0 is non-vanishing, X_0 has at most s nonzero rows) is identifiable up to the transformation group in* (3.11)*, then $N \geq \frac{n-1}{n-s}$.*

Proof. The matrix X_0 has at least $n-s$ zero rows. If we know the locations of $n-s$ zero rows, the problem becomes a BGPC problem with a subspace constraint. The columns of AX_0 reside in an s-dimensional subspace. If $N < \frac{n-1}{n-s}$, the pair (λ_0, X_0) is not identifiable up to scaling and circular shift. The proof is almost identical to that of Proposition 3.3.5.

The pair (λ_0, X_0) cannot be identified even if we know the locations of $n - s$ zero rows. Hence it is not identifiable without knowing the locations of zero rows. $\square$

The above necessary condition gives a tight lower bound on sample complexity. Morrison et al. [28] showed the same necessary condition for SAR autofocus (in the case of known row support of X_0). The two sample complexities, $N \geq s$, as is required by Theorem 3.4.3 implicitly and Corollary 3.4.4 explicitly, and $N \geq \frac{n-1}{n-s}$, coincide when $s = 1$ or $s = n - 1$. The gap between the sufficient condition and the necessary condition is analyzed next.

Gap Between the Sufficient and the Necessary Conditions

The sample complexity $N \geq s$ in the sufficient condition and the sample complexity $N \geq \frac{n-1}{n-s}$ in the necessary condition can be represented by the regions above the line segment and the hyperbola, respectively (cf. Figure 3.2).

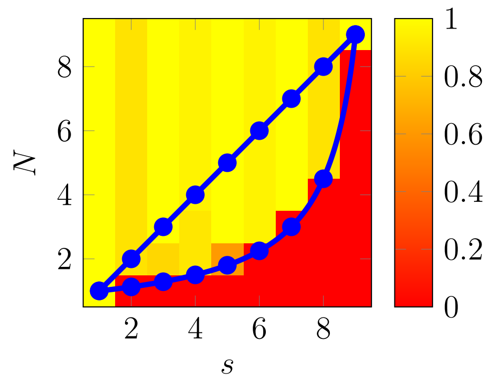


Figure 3.2: The sample complexities for BGPC with DFT matrix and a joint sparsity constraint, and the ratio of identifiable pairs generated randomly.

Algorithm 2 can be used to check the identifiability of BGPC with DFT matrix and a joint-sparsity constraint. Given $Y = \mathrm{diag}(\lambda_0) F X_0$ that has no zero rows and joint support of X_0 that has cardinality s, Algorithm 2 returns a Boolean value indicating whether or not (λ_0, X_0) is identifiable up to the transformation group in (3.11). The procedure enumerates all joint supports of cardinality s.

Algorithm 2 Identifiability of the BGPC problem with DFT matrix and a joint sparsity constraint

input: Y, the joint support J **output**: identifiability of (λ_0, X_0)

for all support J' such that $|J'| = s$ **do**

  $G_{J'} \leftarrow \Big[[\mathrm{diag}(Y^{(:,1)})]^* F^{(:,J'^c)} \quad [\mathrm{diag}(Y^{(:,2)})]^* F^{(:,J'^c)} \quad \cdots \quad [\mathrm{diag}(Y^{(:,N)})]^* F^{(:,J'^c)}\Big]^*$

  if $\mathrm{rank}(G_{J'}) \le n-2$ **then**

    return False

  end if

  if $\mathrm{rank}(G_{J'}) = n-1$ **and** J' is not a shifted version of J **then**

    return False

  end if

end for

return True

Proposition 3.4.9. *Given $Y = \mathrm{diag}(\lambda_0) F X_0$ that has no zero rows and the joint support of X_0 that has cardinality s, the pair (λ_0, X_0) is identifiable (up to the transformation group in* (3.11)*) if Algorithm 2 returns True, and not identifiable otherwise.*

Proof. The matrix $G_{J'} \in \mathbb{C}^{N(n-s)\times n}$ satisfies that $G_{J'}x = \mathrm{vec}(F^{(:,J'^c)*}\,\mathrm{diag}(x) Y)$, where $F^{(:,J'^c)*}$ is an annihilator of the column space of $F^{(:,J')}$. Given Y that has no zero rows, any solution to the BGPC problem (λ, X) satisfies that λ is non-vanishing, and that the entrywise inverse of λ is a null vector of $G_{J'}$, where J' is the joint support of X. On the other hand, any null vector of $G_{J'}$ produces a solution (λ, X), where X is supported on J'.

If Algorithm 2 returns False, then at least one of the following two cases happens:

1. $\mathrm{rank}(G_{J'}) \le n-2$ for some $|J'| = s$. By the proof of Proposition 3.3.5, the solution is not identifiable even if the support J' is known.

2. $\mathrm{rank}(G_{J'}) = n-1$ for some J' that is not a shifted version of J. There exists a solution (λ, X), for which $X \notin [X_0]^R_{\mathscr{T}}$. Therefore (λ_0, X_0) is not identifiable.

In either case, (λ_0, X_0) is not identifiable up to the transformation group in (3.11).

If Algorithm 2 returns True, then $\mathrm{rank}(G_{J'}) \ge n-1$ for all J' of cardinality s, and $\mathrm{rank}(G_{J'}) = n-1$ only if J' is a shifted version of J. Hence any solution (λ, X) must satisfy that the joint support J' is a shifted version of J. Now, given any shifted joint support J', there exists a solution $(\lambda_{J'}, X_{J'}) \in [(\lambda_0, X_0)]_{\mathscr{T}}$. Therefore $G_{J'}$ has at least one null vector $x_{J'}$, which is the entrywise inverse of $\lambda_{J'}$. Hence $\mathrm{rank}(G_{J'}) = n-1$, and the null vectors of $G_{J'}$ reside in the one-dimensional subspace spanned by $x_{J'}$. It follows that given the joint support J', λ in any solution must be a scaled version of $\lambda_{J'}$. Therefore $\lambda \in [\lambda_{J'}]^L_{\mathscr{T}} = [\lambda_0]^L_{\mathscr{T}}$. On the other hand, given non-vanishing λ_0, $\mathrm{diag}(\lambda_0)F$ has full rank and the recovery of X_0 has to be unique. Hence, by Corollary 2.3.3, (λ_0, X_0) is identifiable up to the transformation group in (3.11). $\square$

The sufficient condition in Theorem 3.4.3 is not necessary, as shown by the following numerically constructed counter-examples. Let $n = 10$, $1 \leq s \leq 9$, and $1 \leq N \leq 9$. The joint support J of the columns of $X_0 \in \mathbb{R}^{n\times N}$ is chosen uniformly at random. The entries of $\lambda_0 \in \mathbb{R}^n$ and the nonzero entries of X_0 are generated as iid Gaussian random variables $N(0,1)$. We use Algorithm 2 to determine whether (λ_0, X_0) is identifiable up to the transformation group in (3.11). For every value of s and N, and every support J of cardinality s, the numerical experiment is repeated independently. The ratio of identifiable pairs as a function of (s, N) is shown in Figure 3.2. When $\frac{n-1}{n-s} \leq N < s$ (between the line and the hyperbola), the ratio of identifiable pairs is nonzero. Therefore, $N \geq s$ is not necessary.

The necessary condition in Proposition 3.4.8 is not sufficient. This too can be demonstrated by Figure 3.2. The ratio of identifiable pairs is less than 1 in some regions above the hyperbola. Unidentifiable examples of (λ_0, X_0) that satisfy the necessary condition can be found in Appendix A.3.

As shown by Figure 3.2, when $N < \frac{n-1}{n-s}$ (below the hyperpola), the pairs are not identifiable. When $N \geq s$ (above the line segment), the identifiability hinges on the joint support of the columns of X_0. The "stripes" above the line segment where the ratios of identifiable pairs are slightly less than 1 are due to periodic supports. Most supports are not periodic, hence most pairs are identifiable. When $\frac{n-1}{n-s} \leq N < s$ (between the line and the hyperbola), the situation is more complicated. Besides periodic supports, other joint supports of X_0 can also cause non-identifiability. However, given some "good" joint support of X_0 that depends on both s and N, a randomly chosen (λ_0, X_0) is identifiable almost surely. Recall that non-periodicity of the joint support is necessary, hence "good" supports are a subset of non-periodic supports when $\frac{n-1}{n-s} \leq N < s$. For example, when $s = 5$ and $N = 2$, about 60% of the non-periodic supports are "good". When $s = 7$ and $N = 3$, there is no "good" support. When $s = 7$ and $N = 4$, all non-periodic supports are "good". We have the following conjecture:

Conjecture 3.4.10. *In the BGPC problem with DFT matrix and a joint sparsity constraint, if $N \geq \frac{n-1}{n-s}$, then for almost all $\lambda_0 \in \mathbb{C}^n$ and almost all $X_0 \in \mathbb{C}^{n \times N}$ that has s nonzero rows and some "good" joint support, the pair (λ_0, X_0) is identifiable up to the transformation group $\mathscr{T}$ defined in* (3.11).

Extensions of the Model

The results in Section 3.4.2 apply to $A = F$. This corresponds to MBD where the multiple channels are jointly sparse in the standard basis. Since the product of two circulant matrices is still a circulant matrix, we can easily show that the above results also apply to $A = FC$, where C is a known invertible circulant matrix. This corresponds to MBD where the multiple channels are jointly sparse in the basis formed by the columns of C. In fact, results such as Theorem 3.4.3 can also be derived for other matrices. In Section 3.4.3, we derive a sufficient condition for the identifiability of piecewise constant signals.

Although the results in Section 3.4.2 deal with 1D circular convolutions, extensions to higher-dimensional circular convolutions are straightforward. Let us consider a 2D MBD problem with a joint sparsity constraint as an example, and present a sufficient condition analogous to Theorem 3.4.3. Here $A = F \otimes F \in \mathbb{C}^{n \times n}$ is the 2D DFT matrix, where $F \in \mathbb{C}^{\sqrt{n} \times \sqrt{n}}$ is the 1D DFT matrix. In the 2D problem, the row index of X can be represented by a pair of vertical and horizontal indices. For example, the j-th row of X corresponds to the following index pair:

$$(j^v, j^h) = \left(j - \sqrt{n}\lfloor\frac{j-1}{\sqrt{n}}\rfloor, \lfloor\frac{j-1}{\sqrt{n}}\rfloor + 1\right),$$

where $\lfloor\cdot\rfloor$ denotes the floor operation. Repeating the procedure in Section

3.4.1, the transformation group for the 2D problem is defined by:

$$\Gamma(F \otimes F) = \left\{\gamma = \sigma\sqrt{n}(F \otimes F)^{(:,k)} : \sigma \in \mathbb{C} \text{ is nonzero}, k \in \{1, 2, \cdots, n\}\right\}. \tag{3.14}$$

$$\mathscr{T} = \{\mathcal{T} : \mathcal{T}(\lambda, X) = (\lambda./\gamma, (F \otimes F)^* \operatorname{diag}(\gamma)(F \otimes F)X) \text{ for some } \gamma \in \Gamma(F \otimes F)\}. \tag{3.15}$$

An equivalence transformation $\mathcal{T} \in \mathscr{T}$ maps X into a scaled 2D circular shift version of itself. The periodicity is defined as follows:

Definition 3.4.11. *The index set* $J = \{(j_1^v, j_1^h), (j_2^v, j_2^h), \cdots, (j_s^v, j_s^h)\} \subset \{1, 2, \cdots, \sqrt{n}\}^2$ *is said to be periodic with period* (ℓ^v, ℓ^h) *(*ℓ^v *and* ℓ^h *being integers such that* $0 \leq \ell^v, \ell^h < \sqrt{n}$ *and at least one of the two integers is nonzero), if* $J = \{(j_1^v + \ell^v, j_1^h + \ell^h), (j_2^v + \ell^v, j_2^h + \ell^h), \cdots, (j_s^v + \ell^v, j_s^h + \ell^h)\}$ *(modulo* $(\sqrt{n}, \sqrt{n})$*).*

For example, if $\sqrt{n} = 6$, then the index set $\{(1, 1), (1, 4)\}$ is periodic with period $(0, 3)$. The index set $\{(1, 1), (4, 4)\}$ is periodic with period $(3, 3)$. The index set $\{(1, 1), (4, 1), (1, 4), (4, 4)\}$ is periodic with period $(3, 0)$, $(0, 3)$, or $(3, 3)$. The index set $\{(1, 1), (5, 3), (3, 5)\}$ is periodic with period $(4, 2)$ or $(2, 4)$. The last two examples are shown in Figure 3.3.

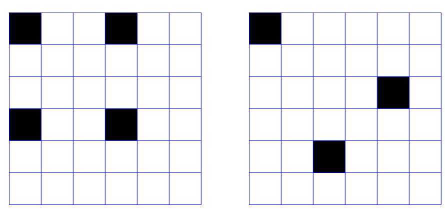


Figure 3.3: Examples of 2D periodic index sets.

Here is the sufficient condition for the 2D problem, whose proof is almost identical to that of Theorem 3.4.3.

Theorem 3.4.12. *In the BGPC problem with 2D DFT matrix* $F \otimes F \in \mathbb{C}^n$ *and a joint sparsity constraint at sparsity level* s*, the pair* $(\lambda_0, X_0) \in \mathbb{C}^n \times \Omega_{\mathcal{X}}$ *is identifiable up to the transformation group* $\mathscr{T}$ *defined in* (3.15) *if the following conditions are met:*

1. *Vector* λ_0 *is non-vanishing.*

2. *Matrix* X_0 *has exactly* s *nonzero rows and rank* s*.*

3. The joint support of the columns of X_0, represented in the index pair form, is not periodic.

3.4.3 Identifiability of Piecewise Constant Signals

Define the finite difference matrix $D \in \mathbb{C}^{n\times n}$ and its inverse as:

$$D = \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}, \quad D^{-1} = \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ \vdots & \vdots & \ddots & \\ 1 & 1 & \cdots & 1 \end{bmatrix}.$$

A piecewise constant signal u can be sparsified by the finite difference operator D. Equivalently, u has the representation $u = D^{-1}x$ in which x is sparse. If $U = D^{-1}X$ in which the columns of X are jointly sparse, then the columns of U are piecewise constant and the discontinuities are at the same locations.

In this section, we consider the following blind deconvolution problem. The observation model is $Y = \operatorname{diag}(\lambda_0)FD^{-1}X_0$, where the matrix X_0 has at most s nonzero rows. The non-vanishing vector λ_0 is the DFT of the filter. The columns of $D^{-1}X_0$ are the signals, which are piecewise constant and share the same discontinuities. An example is deblurring of hyperspectral images. The recovery of (λ_0, X_0) is the BGPC problem with $A = FD^{-1}$ and a joint sparsity constraint.

First, we need to figure out the ambiguity transformation group. The structured matrix $P = A^{-1}\operatorname{diag}(\gamma)A = DF^*\operatorname{diag}(\gamma)FD^{-1} = DCD^{-1}$ is

$$P = \begin{bmatrix} \sum_{j=1}^{n} c^{(j)} & \sum_{j=2}^{n} c^{(j)} & \sum_{j=2}^{n-1} c^{(j)} & \sum_{j=2}^{n-2} c^{(j)} & \cdots & c^{(2)} \\ 0 & c^{(1)} - c^{(2)} & c^{(n)} - c^{(2)} & c^{(n-1)} - c^{(2)} & \cdots & c^{(3)} - c^{(2)} \\ 0 & c^{(2)} - c^{(3)} & c^{(1)} - c^{(3)} & c^{(n)} - c^{(3)} & \cdots & c^{(4)} - c^{(3)} \\ 0 & c^{(3)} - c^{(4)} & c^{(2)} - c^{(4)} & c^{(1)} - c^{(4)} & \cdots & c^{(5)} - c^{(4)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & c^{(n-1)} - c^{(n)} & c^{(n-2)} - c^{(n)} & c^{(n-3)} - c^{(n)} & \cdots & c^{(1)} - c^{(n)} \end{bmatrix}, \tag{3.16}$$

where $C = F^*\operatorname{diag}(\gamma)F$ is a circulant matrix whose first column is

$$\frac{1}{\sqrt{n}}F^*\gamma = c = [c^{(1)}, c^{(2)}, \cdots, c^{(n)}]^{\mathrm{T}}.$$

For P to be a generalized permutation matrix, we must have $c^{(2)} = c^{(3)} = \cdots = c^{(n)} = 0$, and $c^{(1)} \neq 0$. Hence $\gamma = \sqrt{n}Fc = c^{(1)}[1, 1, \cdots, 1]^{\mathrm{T}}$. The ambiguity transformation group in (3.9) becomes (3.1). We only allow an unknown scaling in the recovery.

Next we investigate identifiability up to scaling within the framework of Section 2 and derive a sufficient condition. As in Theorem 3.4.3, one of the requirements is in terms of the joint support of the columns of X_0. We need the following definitions to state this sufficient condition.

Definition 3.4.13. *Let the index sets $J_1, J_2, \cdots, J_T$ be the nodes of an undirected graph. There is an edge between J_{t_1} and J_{t_2} ($1 \leq t_1 < t_2 \leq T$) if $J_{t_1} \bigcap J_{t_2} \neq \varnothing$. The index sets $J_1, J_2, \cdots, J_T$ are said to be connected if the above graph is connected.*

Definition 3.4.14. *The index set $J = \{j_1, j_2, \cdots, j_s\} \subset \{1, 2, \cdots, n\}$ is said to be "friendly" if for any $0 \leq k_1 < k_2 < \cdots < k_{n-s} \leq n-1$, the circularly shifted index sets $J_1, J_2, \cdots, J_{n-s}$, defined by $J_t = \{j_1+k_t, j_2+k_t, \cdots, j_s+k_t\}$ (modulo n), satisfy that*

1. $|\bigcup_{t=1}^{n-s} J_t| \geq n-1$.
2. $J_1, J_2, \cdots, J_{n-s}$ *are connected.*

We make the convention that $\{1, 2, \cdots, n\}$ is friendly.

If the index set J is friendly, and the entries indexed by its circularly shifted version J_t ($1 \leq t \leq n-s$) are equivalent in some sense, then due to transitivity of the equivalence relation, and the connectivity of the circularly shifted index sets, at least $n-1$ out of n entries are equivalent. This property is used in the proof of Theorem 3.4.20.

Remark 3.4.15. *If the index set J is friendly, then its flipped and shifted versions are also friendly.*

We have the following propositions regarding the "friendliness" of an index set. Proposition 3.4.16 shows that, for a non-trivial problem, a friendly index set must have cardinality at least 3, which helps to avoid degeneracy in the proof of Theorem 3.4.20. Propositions 3.4.17 and 3.4.18 give two sufficient conditions for friendliness, which makes the property more readily interpretable. Corollary 3.4.19 gives an alternative characterization of Condition 1 in Definition 3.4.14. See Appendix B.2 for the proofs.

Proposition 3.4.16. *If $n \geq 4$ and the index set J is friendly, then $|J| \geq 3$.*

Proposition 3.4.17. *The index set J is friendly if $|J| \geq 3$ and J is contiguous.*[4]

Proposition 3.4.18. *The index set J is friendly if $|J| > \frac{n}{2}$ and J is not periodic.*

Corollary 3.4.19. *Let $|J| = s < n$. Then $|\bigcup_{t=1}^{n-s} J_t| \geq n-1$ for all choices of $n-s$ shifted index sets J_t if and only if J is not periodic.*

Here is the sufficient condition for identifiability of piecewise constant signals.

Theorem 3.4.20. *Consider the BGPC problem with $A = FD^{-1}$ and two constraints: λ is non-vanishing, and the columns of X are jointly s-sparse. The pair $(\lambda_0, X_0) \in \mathbb{C}^n \times \Omega_{\mathcal{X}}$ is identifiable up to an unknown scaling, if the following conditions are met (assume that $n \geq 4$ and $J = \{j_1, j_2, \cdots, j_s\}$ denotes the joint support of the columns of X_0):*

1. *The vector λ_0 is non-vanishing.*
2. *The matrix X_0 has exactly s nonzero rows, and has rank s.*
3. *$1 \notin J$.*
4. *$\{1\} \bigcup J$ is friendly.*

Proof. First, given non-vanishing λ_0 and $A = FD^{-1}$, the matrix $\operatorname{diag}(\lambda_0)FD^{-1}$ has full rank. If $\operatorname{diag}(\lambda_0)FD^{-1}X_0 = \operatorname{diag}(\lambda_0)FD^{-1}X_1$, then $X_1 = X_0$. Hence, given λ_0, the recovery of X_0 is unique. By Corollary 2.3.3, to establish the result, we only need to show that λ_0 is identifiable up to an unknown scaling.

Assuming that Conditions 1-4 of the theorem are satisfied, we show that λ_1 is a scaled version of λ_0, if $\operatorname{diag}(\lambda_0)FD^{-1}X_0 = \operatorname{diag}(\lambda_1)FD^{-1}X_1$ for (λ_1, X_1) that satisfies the two constraints. The matrix $\operatorname{diag}(\lambda_0)FD^{-1}$ has full rank, hence both X_0 and X_1 have rank s. Without loss of generality, we may assume that X_0 and X_1 only have s columns, which are linearly independent, by removing redundant columns at the same locations in both matrices. They

[4] Index sets like $\{n, 1, 2\}$ are considered contiguous due to the circularity.

both have full column rank s and exactly s nonzero rows. By assumption, the vectors λ_0 and λ_1 are non-vanishing. Write X_1 in terms of X_0, $X_1 = PX_0$, where

$$P = DF^*[\operatorname{diag}(\lambda_1)]^{-1}\operatorname{diag}(\lambda_0)FD^{-1} = DF^*\operatorname{diag}(\gamma)FD^{-1}.$$

The matrix P has the structure in (3.16) where $c = \frac{1}{\sqrt{n}}F^*\gamma = \frac{1}{\sqrt{n}}F^*(\lambda_0./\lambda_1)$. Furthermore, P satisfies (3.13) in the proof of Theorem 3.4.3. The submatrix $P^{(:,J)} = X_1X_0^{\dagger(:,J)}$ has at most s nonzero rows and at least $n-s$ zero rows. The submatrix $P^{(2:n,J)}$ has at least $n-s-1$ zero rows. We denote the corresponding index set by $K = \{k_1, k_2, \cdots, k_{n-s-1}\}$. By (3.16), the row $P^{(k,J)}$ $(k \in K)$ is:

$$P^{(k,J)} = \left[c^{(k+1-j_1)} - c^{(k)}, c^{(k+1-j_2)} - c^{(k)}, \cdots, c^{(k+1-j_s)} - c^{(k)}\right].$$

The index set $J_k = \{k, k+1-j_1, k+1-j_2, \cdots, k+1-j_s\}$ is a flipped and shifted version of $\{1\}\bigcup J = \{1, j_1, j_2, \cdots, j_s\}$. The above row $P^{(k,J)}$ is zero, which means all the entries of the subvector $c^{(J_k)}$ are equal. By the assumption that $\{1\}\bigcup J$ is friendly, the index sets $J_{k_1}, J_{k_2}, \cdots, J_{k_{n-s-1}}$ are connected. That means all the entries of c indexed by $\bigcup_{t=1}^{n-s-1} J_{k_t}$ are equal. Besides, $|\bigcup_{t=1}^{n-s-1} J_{k_t}| \geq n-1$. That means either all the entries of c are equal or there is one entry with a different value. There are three different cases:

1. All the entries of c are equal. Then the vector $\lambda_0./\lambda_1 = \sqrt{n}Fc$ has $n-1$ zeros, which contradicts the assumption that λ_0, λ_1 are non-vanishing.

2. All but the k_0-th entry of c are equal, where $k_0 \neq 1$. Then all the entries of $P^{(2:n,J)}$ that do not contain $c^{(k_0)}$ are zeros, and all the entries that contain $c^{(k_0)}$ are nonzeros. The rows indexed by K are zeros, hence they do not contain $c^{(k_0)}$. The row indexed by k_0 is shown in (3.17), and is nonzero. The rows that contain any of the s entries in (3.18) are also nonzeros.

$$c^{(k_0-j_1+1)} - c^{(k_0)}, c^{(k_0-j_2+1)} - c^{(k_0)}, \cdots, c^{(k_0-j_s+1)} - c^{(k_0)} \tag{3.17}$$

$$c^{(k_0)} - c^{(k_0+j_1-1)}, c^{(k_0)} - c^{(k_0+j_2-1)}, \cdots, c^{(k_0)} - c^{(k_0+j_s-1)} \tag{3.18}$$

Note that no two entries in (3.18) can belong to the same row; no entry in (3.18) belongs to the row in (3.17). If every entry in (3.18) belonged to a row in $P^{(2:n,J)}$, there would be $s+1$ nonzero rows in $P^{(2:n,J)}$. The number of nonzero rows in $P^{(:,J)}$ is at most s. Hence, one of the s entries in (3.18) is not in any row of $P^{(2:n,J)}$. By observation, the only entry that could be missing is $c^{(k_0)} - c^{(1)}$. Assume that, without loss of generality, $c^{(k_0)} - c^{(k_0+j_1-1)}$ is not in any row of $P^{(2:n,J)}$. That implies $k_0 + j_1 - 1 = 1$ (modulo n). Hence there exists an entry in the first row $P^{(1,j_1)} = \sum_{j=2}^{n+2-j_1} c^{(j)} = \sum_{j=2}^{k_0} c^{(j)}$. Since $n \geq 4$ and $\{1\} \bigcup J$ is friendly, by Proposition 3.4.16, $|J| \geq 2$. Hence there exists another entry in the first row $P^{(1,j_2)} = \sum_{j=2}^{n+2-j_2} c^{(j)}$. Since there are s nonzero rows in $P^{(2:n,J)}$, the first row $P^{(1,J)}$ must be zero. Hence,

$$\sum_{j=2}^{k_0} c^{(j)} = \sum_{j=2}^{n+2-j_2} c^{(j)} = 0.$$

Recall that all the entries of c are equal except for $c^{(k_0)}$. It follows that $c^{(1)} = c^{(2)} = \cdots = c^{(n)} = 0$, resulting in a contradiction.

3. All but the first entry of c are equal. Then all the entries of $P^{(2:n,J)}$ that do not contain $c^{(1)}$ are zeros, and all the entries that contain $c^{(1)}$ are nonzeros. In particular, the entries $c^{(1)} - c^{(j_1)}, c^{(1)} - c^{(j_2)}, \cdots, c^{(1)} - c^{(j_s)}$ in the rows indexed by $j_1, j_2, \cdots, j_s$ are nonzeros. Hence the first row $P^{(1,J)}$ must be zero. Therefore, $c^{(2)} = c^{(3)} = \cdots = c^{(n)} = 0$, and $c^{(1)} \neq 0$.

The only case that does not cause a contradiction is the third, which leads to $c = [c^{(1)}, 0, 0, \cdots, 0]^{\mathrm{T}}$ and $\lambda_0./\lambda_1 = \sqrt{n}Fc = c^{(1)}[1, 1, \cdots, 1]^{\mathrm{T}}$. Therefore, $\lambda_1 = \frac{1}{c^{(1)}}\lambda_0$ is a scaled version of λ_0. □

A result for generic signals, analogous to Corollary 3.4.4, follows immediately.

The requirement $N \geq s$, implied by Theorem 3.4.20, is not necessary. We have the following necessary condition, which can be proved similarly to Proposition 3.4.8.

Proposition 3.4.21. *In the BGPC problem with $A = FD^{-1}$ and a joint sparsity constraint, if (λ_0, X_0) (λ_0 is non-vanishing, X_0 has at most s nonzero rows) is identifiable up to scaling, then $N \geq \frac{n-1}{n-s}$.*

An analysis of the gap between the sufficient and the necessary conditions, similar to Section 3.4.2, can be carried out for these results too. It is omitted for brevity.

3.5 Universal Sufficient Condition for BGPC with a Sparsity Constraint

In this section, we consider the BGPC problem with a sparsity constraint on the total number of nonzero entries in the matrix X, denoted by $\|X\|_0$. Consider the following problem:

$$\begin{aligned} \text{(P3)} \quad \text{find} \quad & (\lambda, X), \\ \text{s.t.} \quad & \operatorname{diag}(\lambda) A X = Y, \\ & \lambda \in \mathbb{C}^n,\ X \in \Omega_{\mathcal{X}} = \{X \in \mathbb{C}^{n \times N} : \|X\|_0 \leq s\}. \end{aligned}$$

The measurement is $Y = \operatorname{diag}(\lambda_0) A X_0$. We only consider the case where $A \in \mathbb{C}^{n \times n}$ is an invertible square matrix. The vector $\lambda_0 \in \mathbb{C}^n$ is non-vanishing. The matrix $X_0 \in \mathbb{C}^{n \times N}$ has at most s nonzero entries.

The ambiguity transformation group $\mathscr{T}$ associated with the matrix A is the same as in Section 3.4.1. In Theorem 3.5.1, we show that X_0 is identifiable up to a generalized permutation in the ambiguity transformation group associated with A if the rows of X_0 form the most sparse basis of its row space. This is a universal sufficient condition for BGPC with a sparsity constraint, which applies to every invertible square matrix A. This universal result is derived using the general framework in Section 2.

Theorem 3.5.1. *In the BGPC problem with a sparsity constraint at sparsity level s, the pair (λ_0, X_0) is identifiable up to the ambiguity transformation group $\mathscr{T}$ associated with A, if the following conditions are met:*

1. *Vector λ_0 is non-vanishing.*

2. *If an invertible matrix $P \in \mathbb{C}^{n \times n}$ satisfies that $\|PX_0\|_0 \leq \|X_0\|_0$, then P is a generalized permutation matrix.*

3. *$\|X_0\|_0 = s$.*

Proof. Given non-vanishing λ_0 and invertible A, the matrix $\operatorname{diag}(\lambda_0)A$ is invertible. Hence given λ_0, the matrix X_0 is identifiable. By Corollary 2.3.3, we only need to show that λ_0 is identifiable. Suppose that $\operatorname{diag}(\lambda_0)AX_0 = \operatorname{diag}(\lambda_1)AX_1$ and $\|X_1\|_0 \leq s = \|X_0\|_0$. By the above Condition 2, X_0 has full row rank n. Otherwise, there exists an invertible matrix P that is not a permutation matrix and satisfies $PX_0 = X_0$, which clearly violates Condition 2. The matrix $\operatorname{diag}(\lambda_0)A$ is invertible, hence $\operatorname{rank}(X_1) = \operatorname{rank}(X_0) = n$. There are no zero rows in AX_0 or AX_1. Hence λ_1 is also non-vanishing. Write X_1 in terms of X_0, $X_1 = PX_0$, where $P = A^{-1}[\operatorname{diag}(\lambda_1)]^{-1}\operatorname{diag}(\lambda_0)A$. By the above Condition 2, P has to be a generalized permutation matrix. By (3.8) and (3.9), $\gamma = \lambda_0./\lambda_1 \in \Gamma(A)$ and $\lambda_1 \in [\lambda_0]^{L}_{\mathscr{T}}$. Therefore, λ_0 is identifiable. □

If the sparsity level is not known a priori, we can solve the following optimization problem (P4). Under the above Conditions 1 and 2, the minimizer in (P4) is unique up to the same transformation group. If the minimizer to (P4) has sparsity s, then it is the solution to (P3) as well.

$$
\begin{aligned}
\text{(P4)} \quad \min_{(\lambda, X)}. \quad & \|X\|_0, \\
\text{s.t.} \quad & \operatorname{diag}(\lambda)AX = Y, \\
& \lambda \in \mathbb{C}^n, \; X \in \mathbb{C}^{n \times N}.
\end{aligned}
$$

The following universal sufficient condition follows by combining Theorem 3.5.1 with results about the distribution of non-zero elements in random matrices and in the products of such matrices with vectors [21].

Theorem 3.5.2. *Suppose that the vector λ_0 is non-vanishing, the matrix $X_0 \in \mathbb{C}^{n \times N}$ is Bernoulli-Gaussian random matrix, where $X_0 = B \odot G$, the entries of B are iid Bernoulli random variables $B(1, \theta)$, and the entries of G are iid Gaussian random variables $N(0, 1)$. If $\frac{1}{n} < \theta < \frac{1}{4}$ and $N > Cn \log n$ for a sufficiently large absolute constant C, then the pair (λ_0, X_0) is identifiable in (P4), up to the ambiguity transformation group $\mathscr{T}$ associated with A, with probability at least $1 - \exp(-c\theta N)$ for some absolute constant c.*

Proof. We prove the identifiability by showing that Condition 2 in Theorem 3.5.1 is satisfied with probability at least $1 - \exp(-c\theta N)$ given the above Bernoulli-Gaussian model. Assume that $P \in \mathbb{C}^{n \times n}$ is an invertible matrix but not a generalized permutation matrix. Since P is invertible, there exists a permutation of $1, 2, \cdots, n$, denoted by $j_1, j_2, \cdots, j_n$, such that the support of the kth row $P^{(k,:)}$ contain the index j_k, i.e., $P^{(k,j_k)} \neq 0$, for $1 \leq k \leq n$. Since P is not a generalized permutation matrix, there exists at least one row with more than one nonzero entries. If the row $P^{(k,:)}$ has only one nonzero entry $P^{(k,j_k)}$, then $\left\|(PX_0)^{(k,:)}\right\|_0 = \left\|P^{(k,:)}X_0\right\|_0 = \left\|X_0^{(j_k,:)}\right\|_0$. Next, we show that if $P^{(k,:)}$ has more than one nonzero entries, then $\left\|(PX_0)^{(k,:)}\right\|_0 > \left\|X_0^{(j_k,:)}\right\|_0$ with high probability.

By Lemma 17 in [21], if the Bernoulli-Gaussian matrix X_0 satisfies that $\frac{1}{n} < \theta < \frac{1}{4}$ and $N > Cn\log n$ for a sufficiently large constant C, then the probability that there exists a vector $v \in \mathbb{C}^n$ with more than one nonzero entries such that $\|v^* X_0\|_0 \leq \frac{11}{9}\theta N$ is at most $\exp(-c_1\theta N)$, for some absolute constant c_1. Therefore, with probability at least $1 - \exp(-c_1\theta N)$,

$$\left\|(PX_0)^{(k,:)}\right\|_0 > \frac{11}{9}\theta N \tag{3.19}$$

for every index k such that $P^{(k,:)}$ has more than one nonzero entries.

By Lemma 18 in [21], the probability that any row of the matrix X_0 has more than $\frac{10}{9}\theta N$ nonzero entries is at most $n\exp(-\theta N/243)$. Since $N > Cn\log n$ for a sufficiently large constant C, the probability $n\exp(-\theta N/243) \leq \exp(-c_2\theta N)$ for some absolute constant c_2. Therefore, with probability at least $1 - \exp(-c_2\theta N)$,

$$\left\|X_0^{(j_k,:)}\right\|_0 \leq \frac{10}{9}\theta N \tag{3.20}$$

for every k.

Combining (3.19) and (3.20), $\left\|(PX_0)^{(k,:)}\right\|_0 > \left\|X_0^{(j_k,:)}\right\|_0$ for every index k such that $P^{(k,:)}$ has more than one nonzero entries, with probability at least $1 - \exp(-c\theta N)$ for some absolute constant c. Therefore, with the same probability,

$$\|PX_0\|_0 = \sum_{k=1}^{n} \left\|(PX_0)^{(k,:)}\right\|_0 > \sum_{k=1}^{n} \left\|X_0^{(j_k,:)}\right\|_0 = \|X_0\|_0 .$$

Equivalently, Condition 2 in Theorem 3.5.1 is satisfied with probability at least $1 - \exp(-c\theta N)$. $\square$

CHAPTER 4

IDENTIFIABILITY IN BLIND DECONVOLUTION

4.1 Notations

We state the notations that will be used throughout the chapter. We use lower-case letters x, y, z to denote vectors, and upper-case letters D and E to denote matrices. We use I_n to denote the identity matrix and F_n to denote the normalized discrete Fourier transform (DFT) matrix. The DFT of the vector $x \in \mathbb{C}^n$ is denoted by $\widetilde{x} = F_n x$. We use $\mathbf{1}_{m,n}$ to denote a matrix whose entries are all ones and $\mathbf{0}_{m,n}$ to denote a matrix whose entries are all zeros. The subscripts stand for the dimensions of these matrices. We say that a vector is *non-vanishing* if all its entries are nonzero. Unless otherwise stated, all vectors are column vectors. The dimensions of all vectors and matrices are made clear in the context.

The projection operator onto a subspace $\mathcal{V}$ is denoted by $P_{\mathcal{V}}$. The nullspace and the range space of a linear operator are denoted by $\mathcal{N}(\cdot)$ and $\mathcal{R}(\cdot)$, respectively. We use $\Omega_{\mathcal{X}}, \Omega_{\mathcal{Y}}$ to denote constraint sets. The Cartesian product of two sets are denoted by $\Omega_{\mathcal{X}} \times \Omega_{\mathcal{Y}}$. The pair $(x, y) \in \Omega_{\mathcal{X}} \times \Omega_{\mathcal{Y}}$ represents an element of the Cartesian product. We use $./$ and $\odot$ to denote entrywise division and entrywise product, respectively. Circular convolution is denoted by $\circledast$. Kronecker product is denoted by $\otimes$. The direct sum of two subspaces is denoted by $\oplus$.

We use j, k to denote indices, and J, K to denote index sets. If the universal index set is $\{1, 2, \cdots, n\}$, then J, K are subsets. We use $|J|$ to denote the cardinality of J. We use J^c to denote the complement of J. We use superscript letters to denote subvectors or submatrices. For example, $x^{(J)}$ represents the subvector of x consisting of the entries indexed by J. The scalar $x^{(j)}$ represents the jth entry of x. The submatrix $D^{(J,K)}$ has size $|J| \times |K|$ and consists of the entries indexed by $J \times K$. The vector $D^{(:,k)}$

represents the kth column of the matrix D. The colon notation is borrowed from MATLAB.

We say a property holds for almost all signals (generic signals) if the property holds for all signals but a set of measure zero.

4.2 Problem Statement

4.2.1 Blind Deconvolution

In this chapter, we study the blind deconvolution 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 $\mathbb{C}^n$.

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

We consider the following scenarios:

1. *(Subspace Constraints)* The signal u and the filter v reside in lower-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. The matrices D and E have full column ranks. Therefore,

$$\begin{aligned} \Omega_{\mathcal{U}} &= \{u \in \mathbb{C}^n : u = Dx \text{ for some } x \in \mathbb{C}^{m_1}\}, \\ \Omega_{\mathcal{V}} &= \{v \in \mathbb{C}^n : v = Ey \text{ for some } y \in \mathbb{C}^{m_2}\}. \end{aligned}$$

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. The matrices D and E are bases or frames that satisfy the spark condition [41]: the spark, namely the smallest number of columns that are linearly dependent, of D (resp.

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

E) is greater than $2s_1$ (resp. $2s_2$). Therefore,

$$\begin{aligned}
\Omega_{\mathcal{U}} &= \{u \in \mathbb{C}^n : u = Dx \text{ for some } x \in \mathbb{C}^{m_1} \text{ s.t. } \|x\|_0 \leq s_1\}, \\
\Omega_{\mathcal{V}} &= \{v \in \mathbb{C}^n : v = Ey \text{ for some } y \in \mathbb{C}^{m_2} \text{ s.t. } \|y\|_0 \leq s_2\}.
\end{aligned}$$

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 $E \in \mathbb{C}^{n \times m_2}$. The matrix D satisfies the spark condition, and E has full column rank. Therefore,

$$\begin{aligned}
\Omega_{\mathcal{U}} &= \{u \in \mathbb{C}^n : u = Dx \text{ for some } x \in \mathbb{C}^{m_1} \text{ s.t. } \|x\|_0 \leq s_1\}, \\
\Omega_{\mathcal{V}} &= \{v \in \mathbb{C}^n : v = Ey \text{ for some } y \in \mathbb{C}^{m_2}\}.
\end{aligned}$$

In all three scenarios, the vectors x, y, and z reside in Euclidean spaces $\mathbb{C}^{m_1}$, $\mathbb{C}^{m_2}$ and $\mathbb{C}^n$. With the representations $u = Dx$ and $v = Ey$, it is easy to verify that $z = u \circledast v = (Dx) \circledast (Ey)$ is a bilinear function of x and y. 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 spark condition, then the uniqueness of (u, v) is equivalent to the uniqueness of (x, y). 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, $\Omega_{\mathcal{X}} = \mathbb{C}^{m_1}$, $\Omega_{\mathcal{Y}} = \mathbb{C}^{m_2}$. For sparsity constraints, $\Omega_{\mathcal{X}} = \{x \in \mathbb{C}^{m_1} : \|x\|_0 \leq s_1\}$, $\Omega_{\mathcal{Y}} = \{y \in \mathbb{C}^{m_2} : \|y\|_0 \leq s_2\}$.

4.2.2 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. Any valid definition of unique recovery in BD must address this issue. If every solution (x, y) is a scaled version of (x_0, y_0), then we must say (x_0, y_0) can be uniquely identified up to scaling. We define identifiability as follows.

Definition 4.2.1. *For the constrained BD problem, the solution (x_0, y_0), in which $x_0 \neq 0$ and $y_0 \neq 0$, is said to be 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 blind deconvolution, there exists a linear operator $\mathcal{G}_{DE} : \mathbb{C}^{m_1 \times m_2} \to \mathbb{C}^n$ such that $\mathcal{G}_{DE}(xy^T) = (Dx) \circledast (Ey)$. 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 $M_0 = x_0 y_0^T \in \Omega_{\mathcal{M}} = \{xy^T : x \in \Omega_{\mathcal{X}}, y \in \Omega_{\mathcal{Y}}\}$. The uniqueness of M_0 is equivalent to the identifiability of (x_0, y_0) up to scaling. This procedure is called "lifting".

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

It was shown [6] that the lifted BD has a unique solution for every $M_0 \in \Omega_{\mathcal{M}}$ if the nullspace of $\mathcal{G}_{DE}$ does not contain the difference of two different matrices in $\Omega_{\mathcal{M}}$.

Proposition 4.2.2. *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 in (BD), or equivalently, the solution $M_0 = x_0 y_0^T \in \Omega_{\mathcal{M}}$ is unique in (Lifted BD), if and only if*

$$\mathcal{N}(\mathcal{G}_{DE}) \bigcap \{M_0 - M : M \in \Omega_{\mathcal{M}}\} = \{0\}.$$

Proposition 4.2.2 is difficult to apply because it is not clear how to find the nullspace of the structured linear operator $\mathcal{G}_{DE}$. To overcome this limitation, in Chapter 2 (see Theorem 2.3.2), we derived a necessary and sufficient condition for the identifiability in a bilinear inverse problem up to a transformation group. As a special case, we have the following necessary and sufficient condition for the identifiability in BD up to scaling, which holds for any $\Omega_{\mathcal{X}}$ and $\Omega_{\mathcal{Y}}$.

Proposition 4.2.3. *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 in (BD) if and only if the following two conditions are met:*

1. *If there exists $(x, y) \in \Omega_{\mathcal{X}} \times \Omega_{\mathcal{Y}}$ such that $(Dx) \circledast (Ey) = (Dx_0) \circledast (Ey_0)$, then $x = \sigma x_0$ for some nonzero $\sigma \in \mathbb{C}$.*

2. *If there exists $y \in \Omega_{\mathcal{Y}}$ such that $(Dx_0) \circledast (Ey) = (Dx_0) \circledast (Ey_0)$, then $y = y_0$.*

Propositions 4.2.2 and 4.2.3 are two equivalent conditions for the identifiability in blind deconvolution. Proposition 4.2.2 shows how the identifiability of (x, y) is connected to that of the lifted variable xy^T. Proposition 4.2.3 shows how the identifiability of (x, y) can be divided into the identifiability of x and y individually. In this chapter, we derive more readily interpretable conditions for the uniqueness of solution to BD with subspace or sparsity constraints. We first derive a sufficient condition for the case where the bases or frames are generic, using the lifting framework. We also apply 4.2.3 and derive a sufficient condition for the case where one of the bases has a sub-band structure.

4.3 Blind Deconvolution with Generic Bases or Frames

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 [33, 34, 36, 35, 37], or impose probabilistic models and show successful recovery with high probability [16, 38, 39]. In this section, we derive sufficient conditions for the identifiability of blind deconvolution under subspace or sparsity constraints. These conditions are fully deterministic and provide uniform upper bounds for the sample complexities of blind deconvolution with almost all bases or frames.

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). The linear operator $\mathcal{G}_{DE}$ can also be represented by a matrix $G_{DE} \in \mathbb{C}^{n \times m_1 m_2}$ such that $\mathcal{G}_{DE}(M_0) = G_{DE}\,\mathrm{vec}(M_0)$, where $\mathrm{vec}(M_0)$ stacks the columns of $M_0 \in \mathbb{C}^{m_1 \times m_2}$ on top of one another and forms a vector in $\mathbb{C}^{m_1 m_2}$. The columns of G_{DE} have the form $D^{(:,j)} \circledast E^{(:,k)} = \sqrt{n} F_n^* (\widetilde{D}^{(:,j)} \odot \widetilde{E}^{(:,k)})$, where $\widetilde{D} = F_n D$ and $\widetilde{E} = F_n E$. Clearly, the matrix

G_{DE} is a function of the matrices D and E. It has the following properties (see Appendix C.1 for the proofs).

Lemma 4.3.1. *If $n \geq m_1 m_2$, then for almost all $D \in \mathbb{C}^{n\times m_1}$ and $E \in \mathbb{C}^{n\times m_2}$, the matrix G_{DE} has full column rank.*

Lemma 4.3.2. *If $n \geq 2s_1 m_2$, then for any $0 \leq t_1 \leq s_1$, for almost all $D_0 \in \mathbb{C}^{n\times t_1}$, $D_1 \in \mathbb{C}^{n\times (s_1-t_1)}$, $D_2 \in \mathbb{C}^{n\times (s_1-t_1)}$, and $E \in \mathbb{C}^{n\times m_2}$, the columns of G_{D_0E}, G_{D_1E}, G_{D_2E} together form a linearly independent set.*

Lemma 4.3.3. *If $n \geq 2s_1 s_2$, then for any $0 \leq t_1 \leq s_1$, $0 \leq t_2 \leq s_2$, for almost all $D_0 \in \mathbb{C}^{n\times t_1}$, $D_1 \in \mathbb{C}^{n\times (s_1-t_1)}$, $D_2 \in \mathbb{C}^{n\times (s_1-t_1)}$, $E_0 \in \mathbb{C}^{n\times t_2}$, $E_1 \in \mathbb{C}^{n\times (s_2-t_2)}$, and $E_2 \in \mathbb{C}^{n\times (s_2-t_2)}$, the columns of $G_{D_0E_0}$, $G_{D_1E_0}$, $G_{D_2E_0}$, $G_{D_0E_1}$, $G_{D_1E_1}$, $G_{D_0E_2}$, $G_{D_2E_2}$ together form a linearly independent set.*

Next, we state and prove sufficient conditions for identifiability of blind deconvolution with generic bases or frames.

Theorem 4.3.4 (Subspace Constraints)**.** *In (BD) with subspace constraints, $(x_0, y_0) \in \mathbb{C}^{m_1} \times \mathbb{C}^{m_2}$ ($x_0 \neq 0$, $y_0 \neq 0$) is identifiable up to scaling, for almost all $D \in \mathbb{C}^{n\times m_1}$ and $E \in \mathbb{C}^{n\times m_2}$, if $n \geq m_1 m_2$.*

Proof. By Lemma 4.3.1, if $n \geq m_1 m_2$, for almost all $D \in \mathbb{C}^{n\times m_1}$ and $E \in \mathbb{C}^{n\times m_2}$, the matrix G_{DE} has full column rank. Therefore, $\mathcal{N}(\mathcal{G}_{DE}) = \{0\}$, and the lifted problem has a unique solution. It follows that every pair (x_0, y_0) is identifiable up to scaling. □

Theorem 4.3.5 (Mixed Constraints)**.** *In (BD) with mixed constraints, $(x_0, y_0) \in \mathbb{C}^{m_1} \times \mathbb{C}^{m_2}$ ($\|x_0\|_0 \leq s_1$, $x_0 \neq 0$, $y_0 \neq 0$) is identifiable up to scaling, for almost all $D \in \mathbb{C}^{n\times m_1}$ and $E \in \mathbb{C}^{n\times m_2}$, if $n \geq 2s_1 m_2$.*

Proof. Fix index sets $J_0, J \subset \{1, 2, \cdots, m_1\}$, for which $|J_0| = |J| = s_1$ and $|J_0 \bigcap J| = t_1$. Let

$$D_0 = D^{(:,J_0 \bigcap J)} \in \mathbb{C}^{n\times t_1}, \quad D_1 = D^{(:,J_0\backslash J)} \in \mathbb{C}^{n\times (s_1-t_1)}, \quad D_2 = D^{(:,J\backslash J_0)} \in \mathbb{C}^{n\times (s_1-t_1)}.$$

By Lemma 4.3.2, if $n \geq 2s_1 m_2$, then for almost all D and E, the columns of G_{D_0E}, G_{D_1E}, G_{D_2E} together form a linearly independent set. For every (x_0, y_0) and (x, y) such that the s_1-sparse x_0 and x are supported on J_0 and J respectively, if $\mathcal{G}_{DE}(x_0 y_0^T) = \mathcal{G}_{DE}(xy^T)$, then

$$G_{D_0E}v_0 + G_{D_1E}v_1 + G_{D_2E}v_2 = G_{DE}v = \mathcal{G}_{DE}(x_0 y_0^T) - \mathcal{G}_{DE}(xy^T) = 0,$$

where $v = \mathrm{vec}(x_0 y_0^T - xy^T)$, $v_0 = \mathrm{vec}(x_0^{(J_0 \bigcap J)} y_0^T - x^{(J_0 \bigcap J)} y^T)$, $v_1 = \mathrm{vec}(x_0^{(J_0 \backslash J)} y_0^T)$ and $v_2 = \mathrm{vec}(-x^{(J \backslash J_0)} y^T)$. By linear independence, the vectors v_0, v_1, v_2 are all zero vectors, and so is v. Hence for almost all D and E, and all pairs (x_0, y_0) and (x, y) such that x_0 and x are supported on J_0 and J respectively, if $\mathcal{G}_{DE}(x_0 y_0^T) = \mathcal{G}_{DE}(xy^T)$, then $x_0 y_0^T = xy^T$. Note that J_0 and J are arbitrary, and there is only a finite number $(\binom{m_1}{s_1}^2)$ of combinations of J_0, J. Therefore, for almost all D and E, every pair (x_0, y_0) ($\|x_0\|_0 \leq s_1$, $x_0 \neq 0$, $y_0 \neq 0$) is identifiable up to scaling. $\square$

Theorem 4.3.6 (Sparsity Constraints). *In (BD) with sparsity constraints, $(x_0, y_0) \in \mathbb{C}^{m_1} \times \mathbb{C}^{m_2}$ ($\|x_0\|_0 \leq s_1$, $\|y_0\|_0 \leq s_2$, $x_0 \neq 0$, $y_0 \neq 0$) is identifiable up to scaling, for almost all $D \in \mathbb{C}^{n \times m_1}$ and $E \in \mathbb{C}^{n \times m_2}$, if $n \geq 2 s_1 s_2$.*

Proof. Fix index sets $J_0, J \subset \{1, 2, \cdots, m_1\}$, for which $|J_0| = |J| = s_1$ and $|J_0 \bigcap J| = t_1$, and index sets $K_0, K \subset \{1, 2, \cdots, m_2\}$, for which $|K_0| = |K| = s_2$ and $|K_0 \bigcap K| = t_2$. Let

$$
\begin{aligned}
&D_0 = D^{(:, J_0 \bigcap J)} \in \mathbb{C}^{n \times t_1}, \quad D_1 = D^{(:, J_0 \backslash J)} \in \mathbb{C}^{n \times (s_1 - t_1)}, \\
&D_2 = D^{(:, J \backslash J_0)} \in \mathbb{C}^{n \times (s_1 - t_1)}, \quad E_0 = E^{(:, K_0 \bigcap K)} \in \mathbb{C}^{n \times t_2}, \\
&E_1 = E^{(:, K_0 \backslash K)} \in \mathbb{C}^{n \times (s_2 - t_2)}, \quad E_2 = E^{(:, K \backslash K_0)} \in \mathbb{C}^{n \times (s_2 - t_2)}.
\end{aligned}
$$

By Lemma 4.3.3, if $n \geq 2 s_1 s_2$, then for almost all D and E, the columns of $G_{D_0 E_0}$, $G_{D_1 E_0}$, $G_{D_2 E_0}$, $G_{D_0 E_1}$, $G_{D_1 E_1}$, $G_{D_0 E_2}$, $G_{D_2 E_2}$ together form a linearly independent set. For every (x_0, y_0) and (x, y) such that the s_1-sparse x_0 and x are and supported on J_0 and J respectively, and the s_2-sparse y_0 and y are supported on K_0 and K respectively, if $\mathcal{G}_{DE}(x_0 y_0^T) = \mathcal{G}_{DE}(xy^T)$, then

$$
\begin{aligned}
&G_{D_0 E_0} v_{00} + G_{D_1 E_0} v_{10} + G_{D_2 E_0} v_{20} + G_{D_0 E_1} v_{01} \\
&\quad + G_{D_1 E_1} v_{11} + G_{D_0 E_2} v_{02} + G_{D_2 E_2} v_{22} \\
=& G_{DE} v = \mathcal{G}_{DE}(x_0 y_0^T) - \mathcal{G}_{DE}(xy^T) = 0,
\end{aligned}
$$

where

$$
\begin{aligned}
&v = \mathrm{vec}(x_0 y_0^T - xy^T), \quad v_{00} = \mathrm{vec}(x_0^{(J_0 \bigcap J)} y_0^{(K_0 \bigcap K)T} - x^{(J_0 \bigcap J)} y^{(K_0 \bigcap K)T}), \\
&v_{10} = \mathrm{vec}(x_0^{(J_0 \backslash J)} y_0^{(K_0 \bigcap K)T}), \quad v_{20} = \mathrm{vec}(-x^{(J \backslash J_0)} y^{(K_0 \bigcap K)T}), \\
&v_{01} = \mathrm{vec}(x_0^{(J_0 \bigcap J)} y_0^{(K_0 \backslash K)T}), \quad v_{11} = \mathrm{vec}(x_0^{(J_0 \backslash J)} y_0^{(K_0 \backslash K)T}), \\
&v_{02} = \mathrm{vec}(-x^{(J_0 \bigcap J)} y^{(K \backslash K_0)T}), \quad v_{22} = \mathrm{vec}(-x^{(J \backslash J_0)} y^{(K \backslash K_0)T}).
\end{aligned}
$$

By linear independence, the vectors $v_{00}, v_{10}, v_{20}, v_{01}, v_{11}, v_{02}, v_{22}$ are all zero vectors, and so is v. Hence for almost all D and E, and all pairs (x_0, y_0) and (x, y) such that x_0 and x are supported on J_0 and J respectively, and y_0 and y are supported on K_0 and K respectively, if $\mathcal{G}_{DE}(x_0 y_0^T) = \mathcal{G}_{DE}(xy^T)$, then $x_0 y_0^T = xy^T$. Note that the supports J_0, J, K_0, K are arbitrary, and there is only a finite number $(\binom{m_1}{s_1}^2 \binom{m_2}{s_2}^2)$ of combinations of supports. Therefore, for almost all D and E, every pair (x_0, y_0) ($\|x_0\|_0 \leq s_1$, $\|y_0\|_0 \leq s_2$, $x_0 \neq 0$, $y_0 \neq 0$) is identifiable up to scaling. $\square$

Due to symmetry, we can derive another sufficient condition for the scenario where $u = Dx$ resides in a m_1-dimensional subspace spanned by the columns of D, and $v = Ey$ is s_2-sparse over E.

For generic bases or frames, the above sample complexities $n \geq m_1 m_2$, $n \geq 2s_1 m_2$ or $n \geq 2s_1 s_2$ are sufficient. These sampling complexities are not optimal, since they are in terms of the number of nonzero entries in $x_0 y_0^T$, instead of the number of degrees of freedom in x_0 and y_0. For example, in the scenario with subspace constraints, Theorem 4.3.4 requires $n \geq m_1 m_2$ samples, versus the number of degrees of freedom, which is $m_1 + m_2 - 1$. However, these results hold with essentially no assumptions on D or E. They are the first algebraic sample complexities for blind deconvolution.

4.4 Blind Deconvolution with a Sub-band Structured Basis

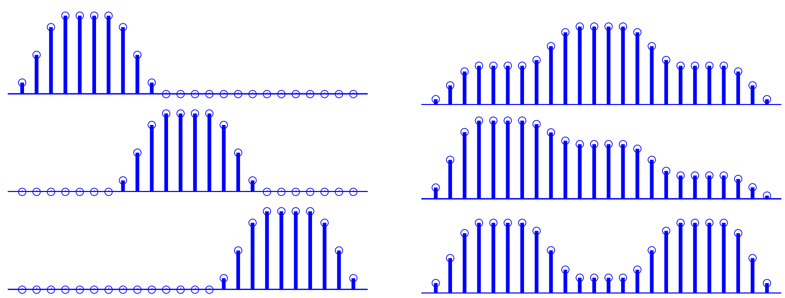
In this section, we consider the BD problem where the filter resides in a subspace spanned by a sub-band structured basis. For this setup, using the general framework for bilinear inverse problems we introduced in Chapter 2, and Proposition 4.2.3 above, we derive much stronger, essentially optimal sample complexity results.

Definition 4.4.1. *Let $\widetilde{E} = F_n E$, $E \in \mathbb{C}^{n \times m_2}$, and let J_k denote the support of $\widetilde{E}^{(:,k)}$ ($1 \leq k \leq m_2$). If*

$$\widehat{J}_k = J_k \setminus \left(\bigcup_{k' \neq k} J_{k'} \right) \neq \varnothing \quad \text{for} \quad 1 \leq k \leq m_2,$$

then we say E forms a sub-band structured basis. The nonempty index set $\widehat{J}_k$ and its cardinality $\ell_k := |\widehat{J}_k|$ are called the passband and the bandwidth of $E^{(:,k)}$, respectively.

Like filters in a filter bank, the basis vectors of a sub-band structured basis are supported on different sub-bands in the Fourier domain (Figure 4.1(a)). By Definition 4.4.1, the sub-bands may overlap partially. For each sub-band, its passband consists of the frequency components (which need not be contiguous) that are not present in any other sub-band. For example, in acoustic signal processing or communications, an equalizer that adjusts the relative gains $y^{(k)}$ of different frequency components can be considered as a filter $v = Ey$ that resides in a subspace with a sub-band structured basis. See Figure 4.1(b) for the DFTs of three different equalizers, and Figure 4.2 for the filter bank implementation of an equalizer.



(a) (b)

Figure 4.1: A sub-band structured basis. (a) DFTs of basis vectors. (b) Examples of frequency responses of filters in the span of the sub-band structured basis.

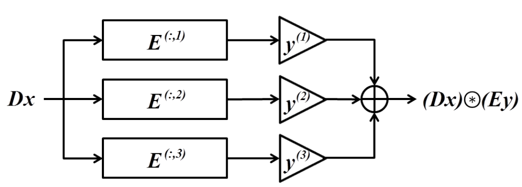


Figure 4.2: Filter bank implementation of an equalizer.

Next, we address the identifiability of the blind deconvolution problem where the filter resides in a subspace with a sub-band structured basis, and the signal resides in another subspace, or is sparse over some given dictionary. For example, consider the following blind deconvolution problem in channel encoding. An unknown source string x is encoded by a given tall-and-skinny matrix D and then transmitted through a channel whose gains in different sub-bands are unknown. Then the encoded string Dx resides in a subspace spanned by D, and the channel resides in a subspace with a sub-band structured basis. Simultaneous recovery of the channel and the encoded string from measurements of the channel output corresponds to blind deconvolution with a sub-band structured basis. Another example is the channel identification problem where the acoustic channel can be modeled as the serial concatenation of an equalizer and a multipath channel. The equalizer has known sub-bands but unknown gains. The multipath channel can be modeled as a sparse filter. Then the simultaneous recovery of the sparse multipath channel and the equalizer from the given input and measured output of the channel corresponds to blind deconvolution with a sub-band structured basis.

We consider first the case of subspace constraints, with one of the bases having a sub-band structure.

Theorem 4.4.2. *In (BD) with subspace constraints, suppose E forms a sub-band structured basis, $x_0 \in \mathbb{C}^{m_1}$ is nonzero and $y_0 \in \mathbb{C}^{m_2}$ is non-vanishing. If the sum of all the bandwidths $\sum_{k=1}^{m_2} \ell_k \geq m_1 + m_2 - 1$, then for almost all $D \in \mathbb{C}^{n \times m_1}$, the pair $(x_0, y_0) \in \Omega_{\mathcal{X}} \times \Omega_{\mathcal{Y}}$ is identifiable up to scaling.*

Proof. Let $\tilde{D} = F_n D$, $\tilde{E} = F_n E$. By the sub-band structure assumption, $\tilde{E}$ has full column rank. For nonzero x_0 and for almost all D, all the entries of $\widetilde{D}x_0$ are nonzero and the matrix $\mathrm{diag}(\widetilde{D}x_0)\widetilde{E}$ has full column rank. If there exists $y \in \Omega_{\mathcal{Y}}$ such that $(Dx_0) \circledast (Ey) = (Dx_0) \circledast (Ey_0)$, then

$$\mathrm{diag}(\widetilde{D}x_0)\widetilde{E}y = (\widetilde{D}x_0) \odot (\widetilde{E}y) = (\widetilde{D}x_0) \odot (\widetilde{E}y_0) = \mathrm{diag}(\widetilde{D}x_0)\widetilde{E}y_0.$$

It follows that $y = y_0$. By Proposition 4.2.3, to complete the proof, we only need to show that if there exists $(x, y) \in \Omega_{\mathcal{X}} \times \Omega_{\mathcal{Y}}$ such that $(Dx) \circledast (Ey) = (Dx_0) \circledast (Ey_0)$ then $x = \sigma x_0$ for some nonzero σ.

If there exists $(x, y) \in \Omega_{\mathcal{X}} \times \Omega_{\mathcal{Y}}$ such that $(Dx) \circledast (Ey) = (Dx_0) \circledast (Ey_0)$, we have

$$\mathrm{diag}(\widetilde{E}y)\widetilde{D}x = \mathrm{diag}(\widetilde{E}y_0)\widetilde{D}x_0.$$

Considering the passband $\widehat{J}_k$, we have

$$\operatorname{diag}(\widetilde{E}^{(\widehat{J}_k,:)}y)\widetilde{D}^{(\widehat{J}_k,:)}x = \operatorname{diag}(\widetilde{E}^{(\widehat{J}_k,:)}y_0)\widetilde{D}^{(\widehat{J}_k,:)}x_0,$$

or equivalently

$$\begin{aligned}\operatorname{diag}(\widetilde{E}^{(\widehat{J}_k,k)})\widetilde{D}^{(\widehat{J}_k,:)}xy^{(k)} &= \operatorname{diag}(\widetilde{E}^{(\widehat{J}_k,k)})\widetilde{D}^{(\widehat{J}_k,:)}x_0y_0^{(k)},\\ \widetilde{D}^{(\widehat{J}_k,:)}xy^{(k)} &= \widetilde{D}^{(\widehat{J}_k,:)}x_0y_0^{(k)}.\end{aligned}$$

By assumption, $y_0^{(k)} \neq 0$. For almost all D, $\widetilde{D}^{(\widehat{J}_k,:)}x_0 \neq 0$. Hence $y^{(k)} \neq 0$, $x \neq 0$. It follows that

$$\widetilde{D}^{(\widehat{J}_k,:)}(x - \frac{y_0^{(k)}}{y^{(k)}}x_0) = 0.$$

Hence

$$x \in \mathcal{N}(\widetilde{D}^{(\widehat{J}_k,:)}) + \operatorname{span}(x_0). \tag{4.1}$$

Let $x_0^\perp$ denote the orthogonal complement of $\operatorname{span}(x_0)$. Then

$$\begin{aligned}&P_{x_0^\perp}x \in x_0^\perp,\\ &P_{x_0^\perp}x = x - P_{\operatorname{span}(x_0)}x \in \mathcal{N}(\widetilde{D}^{(\widehat{J}_k,:)}) + \operatorname{span}(x_0).\end{aligned}$$

Hence

$$P_{x_0^\perp}x \in x_0^\perp \bigcap \Big(\mathcal{N}(\widetilde{D}^{(\widehat{J}_k,:)}) + \operatorname{span}(x_0)\Big) = x_0^\perp \bigcap \Big(\mathcal{R}(\widetilde{D}^{(\widehat{J}_k,:)*}) \bigcap x_0^\perp\Big)^\perp, \quad \forall k \in \{1, 2, \cdots, m_2\}, \tag{4.2}$$

where $(\cdot)^*$ denotes the conjugate transpose. The equation holds due to the fact that, for linear vector spaces $\mathcal{V}_1$ and $\mathcal{V}_2$, $\mathcal{V}_1 + \mathcal{V}_2 = (\mathcal{V}_1^\perp \bigcap \mathcal{V}_2^\perp)^\perp$.

Now, for almost all D, $\mathcal{R}(\widetilde{D}^{(\widehat{J}_k,:)*})$ is a generic ℓ_k-dimensional subspace of $\mathbb{C}^{m_1}$, and $\mathcal{R}(\widetilde{D}^{(\widehat{J}_k,:)*}) \not\subset x_0^\perp$. Hence there exists a generic $(\ell_k - 1)$-dimensional subspace $\mathcal{V}_k \subset x_0^\perp$ such that

$$\begin{aligned}&\mathcal{R}(\widetilde{D}^{(\widehat{J}_k,:)*}) = \mathcal{V}_k \oplus \operatorname{span}(P_{\mathcal{R}(\widetilde{D}^{(\widehat{J}_k,:)*})}x_0),\\ &\mathcal{R}(\widetilde{D}^{(\widehat{J}_k,:)*}) \bigcap x_0^\perp = \mathcal{V}_k.\end{aligned}$$

Therefore, (4.2) is equivalent to

$$P_{x_0^\perp} x \in x_0^\perp \bigcap \mathcal{V}_1^\perp \bigcap \mathcal{V}_2^\perp \bigcap \cdots \bigcap \mathcal{V}_{m_2}^\perp = (\mathrm{span}(x_0) + \sum_{k=1}^{m_2} \mathcal{V}_k)^\perp,$$

where $\mathcal{V}_1, \mathcal{V}_2, \cdots, \mathcal{V}_{m_2}$ are generic subspaces of $x_0^\perp$, the dimensions of which are $\ell_1 - 1, \ell_2 - 1, \cdots, \ell_{m_2} - 1$. For any such generic subspaces of $x_0^\perp$, if $\sum_{k=1}^{m_2} \ell_k \geq m_1 + m_2 - 1$, i.e., $\sum_{k=1}^{m_2} (\ell_k - 1) \geq m_1 - 1$, then

$$\sum_{k=1}^{m_2} \mathcal{V}_k = x_0^\perp.$$

Hence

$$\mathrm{span}(x_0) + \sum_{k=1}^{m_2} \mathcal{V}_k = \mathbb{C}^{m_1},$$

$$P_{x_0^\perp} x \in (\mathrm{span}(x_0) + \sum_{k=1}^{m_2} \mathcal{V}_k)^\perp = \{0\}.$$

Therefore, $P_{x_0^\perp} x = 0$, or $x \in \mathrm{span}(x_0)$. We have shown that $x \neq 0$, hence there exists a nonzero $\sigma \in \mathbb{C}$ such that $x = \sigma x_0$. The proof is complete. $\square$

We turn next to the case of blind deconvolution with mixed constraints, where the signal lives in a subspace spanned by a sub-band structured basis, and the filter is sparse.

Theorem 4.4.3. *In (BD) with mixed constraints, suppose E forms a sub-band structured basis, $x_0 \in \mathbb{C}^{m_1}$ satisfies that $\|x_0\|_0 \leq s_1$ and $x_0 \neq 0$, and $y_0 \in \mathbb{C}^{m_2}$ is non-vanishing. If the sum of all the bandwidths $\sum_{k=1}^{m_2} \ell_k \geq 2s_1 + m_2 - 1$, then for almost all $D \in \mathbb{C}^{n \times m_1}$, the pair $(x_0, y_0) \in \Omega_\mathcal{X} \times \Omega_\mathcal{Y}$ is identifiable up to scaling.*

Proof. The proof is very similar to that of Theorem 4.4.2. For nonzero x_0 and almost all D, if there exists $y \in \Omega_\mathcal{Y}$ such that $(Dx_0) \circledast (Ey) = (Dx_0) \circledast (Ey_0)$, then $y = y_0$. By Proposition 4.2.3, to complete the proof, we only need to show that if there exists $(x, y) \in \Omega_\mathcal{X} \times \Omega_\mathcal{Y}$ such that $\|x\|_0 \leq s_1$ and $(Dx) \circledast (Ey) = (Dx_0) \circledast (Ey_0)$, then $x = \sigma x_0$ for some nonzero σ.

Denote the support of x_0 by K_0, $|K_0| = s_1$. If there exists $(x, y) \in \Omega_\mathcal{X} \times \Omega_\mathcal{Y}$ such that x is supported on K, $|K| = s_1$, and $(Dx) \circledast (Ey) = (Dx_0) \circledast (Ey_0)$, then

$$\operatorname{diag}(\widetilde{E}y)\widetilde{D}^{(:,K_0\bigcup K)}x^{(K_0\bigcup K)} = \operatorname{diag}(\widetilde{E}y_0)\widetilde{D}^{(:,K_0\bigcup K)}x_0^{(K_0\bigcup K)}.$$

In this case, (4.1) and (4.2) in the proof of Theorem 4.4.2 become

$$\begin{aligned} &x^{(K_0\bigcup K)} \in \mathcal{N}(\widetilde{D}^{(\widehat{J}_k,K_0\bigcup K)}) + \operatorname{span}(x_0^{(K_0\bigcup K)}),\\ &P_{x_0^{(K_0\bigcup K)\perp}}x^{(K_0\bigcup K)} \in x_0^{(K_0\bigcup K)\perp}\bigcap\left(\mathcal{R}(\widetilde{D}^{(\widehat{J}_k,K_0\bigcup K)*})\bigcap x_0^{(K_0\bigcup K)\perp}\right)^{\perp},\\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall k\in\{1,2,\cdots,m_2\}. \end{aligned}$$

Since $|K_0| = |K| = s_1$, we have $|K_0\bigcup K| \leq 2s_1$. If $\sum_{k=1}^{m_2}\ell_k \geq 2s_1 + m_2 - 1$, then by an argument analogous to that in the proof of Theorem 4.4.2, we have that for almost all D, $P_{x_0^{(K_0\bigcup K)\perp}}x^{(K_0\bigcup K)}$ must be 0. Therefore, there exists a nonzero $\sigma\in\mathbb{C}$ such that $x = \sigma x_0$.

We complete the proof by enumerating all supports K of cardinality s_1. Since there is only a finite number ($\binom{m_1}{s_1}$) of such supports, for almost all D, if there exists (x,y) such that x is s_1-sparse and $(Dx)\circledast(Ey) = (Dx_0)\circledast(Ey_0)$, then $x = \sigma x_0$ for some nonzero σ. □

How do the sufficient conditions of Theorems 4.4.2 and 4.4.3 compare to the minimal required sample complexities? We address this question for the following scenario. Suppose that the supports J_k ($1\leq k\leq m_2$) form a partition of the frequency range, i.e.,

$$\begin{aligned} &J_{k_1}\bigcap J_{k_2} = \varnothing \quad \text{for all } k_1 \text{ and } k_2 \text{ such that } k_1\neq k_2,\\ &\bigcup_{1\leq k\leq m_2} J_k = \{1,2,\cdots,n\}. \end{aligned}$$

In this case the passbands are $\widehat{J}_k = J_k$ and $n = \sum_{k=1}^{m_2}\ell_k$. For example, this scenario applies when the filter bank is an array of ideal bandpass filters whose passbands partition the DFT frequency range (see Figure 4.3). Consider first (BD) with subspace constraints. Under the above scenario, the sufficient condition in Theorem 4.4.2 implies $n\geq m_1+m_2-1$. Next, we show that this sample complexity is also necessary.

Proposition 4.4.4. *In (BD) with subspace constraints, suppose E forms a sub-band structured basis, for which the supports J_k ($1\leq k\leq m_2$) are disjoint*

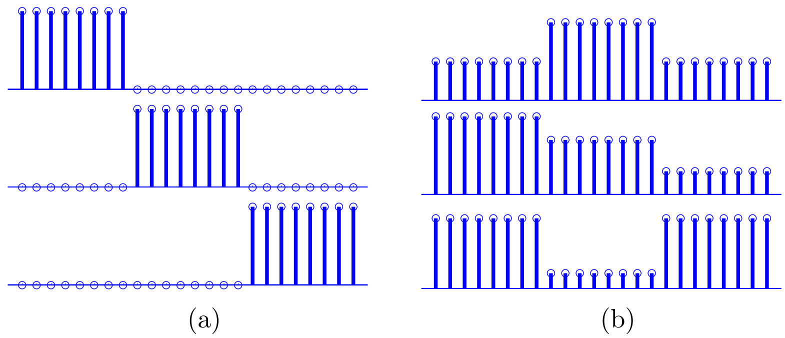


Figure 4.3: A sub-band structured basis with supports J_k that partition the DFT frequency range. (a) DFTs of basis vectors. (b) Examples of frequency responses of filters in the span of the basis.

and cover all the frequency components. If (x_0, y_0) *(*y_0 *is non-vanishing) is identifiable up to scaling, then* $n \geq m_1 + m_2 - 1$*.*

We turn next to (BD) with mixed constraints. Under the assumption that the passbands partition the DFT frequency range, the sufficient condition in Theorem 4.4.3 implies $n \geq 2s_1 + m_2 - 1$. Next, we show that this is almost necessary.

Corollary 4.4.5. *In (BD) with mixed constraints, suppose* E *forms a sub-band structured basis, for which the supports* J_k *(*$1 \leq k \leq m_2$*) are disjoint and cover all the frequency components. If* (x_0, y_0) *(*x_0 *is* s_1*-sparse,* y_0 *is non-vanishing) is identifiable up to scaling, then* $n \geq s_1 + m_2 - 1$*.*

The sample complexities in the sufficient conditions match (exactly for (BD) with subspace constraints and almost for (BD) with mixed constraints) those in the necessary conditions, hence they are optimal. The sample complexities are also optimal in the sense that the number of degrees of freedom is roughly equal to the number of measurements. We give the proofs of Proposition 4.4.4 and Corollary 4.4.5 in Appendix C.2.

CHAPTER 5

CONCLUSION

Previous results on the identifiability in bilinear inverse problems (BIPs) are limited. In this thesis, we defined identifiability of a BIP up to transformation groups. A general framework for proving identifiability was proposed, and was later applied to blind gain and phase calibration (BGPC) and blind deconvolution (BD).

In Chapter 3, we showed sufficient conditions for the unique recovery up to transformation groups in BGPC under three scenarios, with a subspace constraint, with a joint sparsity constraint, and with a sparsity constraint, respectively. We also provided necessary conditions for the scenarios with a subspace constraint or a joint sparsity constraint. We developed a procedure to determine the ambiguity transformation groups for BGPC with joint sparsity or with sparsity constraints. We also designed algorithms that can check the identifiability for BGPC with subspace or with joint sparsity constraints, and demonstrated the tightness of our sample complexity bounds by numerical experiments.

The analysis in Chapter 3 is not always optimal. In certain cases, there exist gaps between the sufficient conditions and the necessary conditions. For example, in the scenario with DFT matrix and a joint sparsity constraint, the gap between the sample complexities in the sufficient and the necessary conditions is $N \geq s$ versus $N \geq \frac{n-1}{n-s}$. However, we believe that it would be possible to bridge these gaps by introducing more stringent assumptions (e.g., generic vectors and matrices).

In Chapter 4, we studied the identifiability of blind deconvolution problems with subspace or sparsity constraints. We derived two algebraic conditions on blind deconvolution with subspace constraints. We first showed using the lifting framework that blind deconvolution from n observations with generic bases of dimensions m_1 and m_2 is identifiable up to scaling given that $n \geq m_1 m_2$. Then we applied the general framework in Chapter 2 to show that blind deconvolution with a sub-band structured basis is identifiable up to scaling given that $n \geq m_1 + m_2 - 1$. The second result was shown to be tight. These results are also generalized to blind deconvolution with sparsity constraints or mixed constraints, with sparsity level(s) replacing the subspace dimension(s). The extra cost for the unknown support in the case of sparsity constraints is an extra factor of 2 in the sample complexity.

We acknowledge that the results in Section 4.3 for generic bases may not be optimal. But they provide the first algebraic conditions for feasibility of blind deconvolution with subspace or sparsity priors. Furthermore, taking advantage of the interesting sub-band structure of some bases (such as filters in a filter bank implementation of equalizers), we achieved sample complexities that are essentially optimal. Our results are derived with generic bases or frames, which means they are violated on a set of Lebesgue measure zero.

One goal of this thesis is to motivate more research into the identifiability of bilinear inverse problems. For BGPC, additional identifiability results can be obtained for different bases A and different constraint sets $\Omega_\Lambda, \Omega_\mathcal{X}$. For example, exploiting the extra information regarding λ (positivity in inverse rendering, unit-modulus entries in SAR autofocus), is expected to provide less demanding conditions for identifiability. For BD, an interesting question is, without the sub-band structure, whether or not it is possible to provide an algebraic analysis of blind deconvolution that achieves optimal sample complexities. Furthermore, identifiability analysis of blind deconvolution with specific bases or frames that arise in applications is still an open problem. The merit of the framework in Chapter 2 for identifiability in bilinear inverse problems is not restricted to the demonstrated exemplary applications. It will be useful for analyzing a wider class of practical applications, including blind deconvolution with the linear convolution model, phase retrieval, dictionary learning, etc.

APPENDIX A

EXAMPLES FOR CHAPTERS 2 AND 3

A.1 Example of a Non-trivial Annihilator

Most bilinear mappings that arise in applications do not have non-trivial left or right annihilators, however this is not universally true. Here is an example in which the bilinear mapping does have a non-trivial right annihilator. Assume that $z = x_0 y_0^{(1)} \in \mathbb{C}^2$ in the following BIP:

$$\begin{aligned} \text{find} \quad & (x, y), \\ \text{s.t.} \quad & x y^{(1)} = z, \\ & x \in \mathbb{C}^2, \ y \in \mathbb{C}^2. \end{aligned}$$

Then (x_0, y_0) is identifiable up to the following transformation group:

$$\mathscr{T} = \left\{ \mathcal{T} : \mathcal{T}(x, y) = \left(\frac{1}{\sigma} x, [\sigma y^{(1)}, y^{(2)} + \tau]^{\mathrm{T}} \right) \text{ for some } \sigma \neq 0 \text{ and } \tau \in \mathbb{C} \right\}.$$

Let $\mathcal{T} = (\mathcal{T}_{\mathcal{X}}, \mathcal{T}_{\mathcal{Y}})$, where $\mathcal{T}_{\mathcal{X}}(x) = \frac{1}{\sigma} x$, $\mathcal{T}_{\mathcal{Y}}(y) = [\sigma y^{(1)}, y^{(2)} + \tau]^{\mathrm{T}}$. Note that $\mathcal{T}_{\mathcal{Y}}$ is not a linear transformation if $\tau \neq 0$. In addition, Condition 2 in Corollary 2.3.3 is not necessary. Given $\mathcal{F}(x_0, y_0) = \mathcal{F}(x_0, y)$, i.e., $x_0 y_0^{(1)} = x_0 y^{(1)}$, it is not necessary that $y = y_0$. The reason is that the bilinear mapping $\mathcal{F}$ has a non-trivial right annihilator $y = [0, 1]^{\mathrm{T}}$.

A.2 Examples of Ambiguity Transformation Groups

In the BGPC problem with a joint sparsity constraint, the ambiguity transformation groups for A can be figured out with the method in Section 3.4.1. The ambiguity transformation groups associated with $A = F$ and $A = FD^{-1}$

are shown in Section 3.4.1 and Section 3.4.3 respectively. We give more examples here.

The matrix A introduces some "mixing" to the rows of X. If $A = I$, there is no mixing. The structured matrix $I^{-1}\operatorname{diag}(\gamma)I = \operatorname{diag}(\gamma)$ is a diagonal matrix. It is a generalized permutation matrix provided that γ is non-vanishing. The set of γ which produces a generalized permutation matrix is $\Gamma(I) = \{\gamma \in \mathbb{C}^n : \gamma \text{ is non-vanishing}\}$. The ambiguity transformation group is

$$\mathscr{T} = \{\mathcal{T} : \mathcal{T}(\lambda, X) = (\lambda./\gamma, \operatorname{diag}(\gamma)X) \text{ for some non-vanishing } \gamma\}.$$

In this case, any non-vanishing λ is considered equivalent to λ_0. The identifiability of (λ_0, X_0) with this transformation group is not an interesting problem.

For some A, the structured matrix $A^{-1}\operatorname{diag}(\gamma)A$ is already studied in the literature. For example, if A is a DFT matrix, $A^{-1}\operatorname{diag}(\gamma)A$ is a circulant matrix. If A is the discrete cosine transform (DCT) matrix, $A^{-1}\operatorname{diag}(\gamma)A$ is the sum of a symmetric Toeplitz matrix and a Hankel matrix [42]. For other matrices, the structure of $A^{-1}\operatorname{diag}(\gamma)A$ can be figured out by symbolic computation. The matrix $A = FD^{-1}$ in Section 3.4.3 is an example. Another example is the Haar matrix H_n, corresponding to a wavelet transform. The matrix H_4 and the structured matrix $H_4^{-1}\operatorname{diag}(\gamma)H_4$ are

$$H_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix},$$

$$H_4^{-1}\operatorname{diag}(\gamma)H_4 = \frac{1}{4}\begin{bmatrix} \gamma_{\mathrm{III}} & \gamma_{\mathrm{II}} & \gamma_{\mathrm{I}} & \gamma_{\mathrm{I}} \\ \gamma_{\mathrm{II}} & \gamma_{\mathrm{III}} & \gamma_{\mathrm{I}} & \gamma_{\mathrm{I}} \\ \gamma_{\mathrm{I}} & \gamma_{\mathrm{I}} & \gamma_{\mathrm{V}} & \gamma_{\mathrm{IV}} \\ \gamma_{\mathrm{I}} & \gamma_{\mathrm{I}} & \gamma_{\mathrm{IV}} & \gamma_{\mathrm{V}} \end{bmatrix},$$

where $\gamma_{\mathrm{I}} = \gamma^{(1)} - \gamma^{(2)}$, $\gamma_{\mathrm{II}} = \gamma^{(1)} + \gamma^{(2)} - 2\gamma^{(3)}$, $\gamma_{\mathrm{III}} = \gamma^{(1)} + \gamma^{(2)} + 2\gamma^{(3)}$, $\gamma_{\mathrm{IV}} = \gamma^{(1)} + \gamma^{(2)} - 2\gamma^{(4)}$, and $\gamma_{\mathrm{V}} = \gamma^{(1)} + \gamma^{(2)} + 2\gamma^{(4)}$. The structured matrix $H_4^{-1}\operatorname{diag}(\gamma)H_4$ is a generalized permutation matrix if and only if $\gamma^{(2)} = \gamma^{(1)}$, $\gamma^{(3)} = \pm\gamma^{(1)}$ and $\gamma^{(4)} = \pm\gamma^{(1)}$. The set $\Gamma(H_4)$ and the ambiguity transformation group $\mathscr{T}$ are

$$\Gamma(H_4) = \{\gamma : \gamma^{(1)} = \gamma^{(2)} = \sigma, \gamma^{(3)} = \pm\sigma, \gamma^{(4)} = \pm\sigma, \text{ for some nonzero } \sigma \in \mathbb{C}\},$$

$$\mathscr{T} = \{\mathcal{T} : \mathcal{T}(\lambda, X) = (\lambda./\gamma, H_4^{-1}\operatorname{diag}(\gamma)H_4 X) \text{ for some } \gamma \in \Gamma(H_4)\}.$$

A.3 Insufficiency of the Condition in Proposition 3.4.8

The necessary condition in Proposition 3.4.8 is not sufficient, even when the locations of the zero rows are known a priori. For example, when $n = 7$, $s = 4$, $\lambda_0 \in \mathbb{C}^7$ is non-vanishing and

$$X_0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

the pair (λ_0, X_0) is not identifiable, even if we know that the last three rows of X_0 are zeros. There exists a circulant matrix P whose first column is $[1, 2, 0, 0, 0, 0, 0]^{\mathrm{T}}$ such that

$$X_1 = PX_0 = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

and $\lambda_1 = \lambda_0./\gamma$, where $\gamma = \sqrt{n}F[1, 2, 0, 0, 0, 0, 0]^{\mathrm{T}}$ is non-vanishing.

The above example is a degenerate case where the actual joint sparsity of X_0 is less than $s = 4$. A non-degenerate X_0 may also not be identifiable, if there is no extra knowledge of the locations of the zero rows. For example,

$$X_0 = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 1 & 3 \\ 3 & 2 & 1 \\ -29 & -28.5 & -17.5 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

There exists a circulant matrix P whose first column is $[2, 16, 1, 8, 0.5, 4, 32]^{\mathrm{T}}$, such that

$$X_1 = PX_0 = \begin{bmatrix} 63.5 & 31.75 & 95.25 \\ 0 & 0 & 0 \\ -889 & -889 & -508 \\ 0 & 0 & 0 \\ -444.5 & -444.5 & -254 \\ 0 & 0 & 0 \\ -190.5 & -127 & -63.5 \end{bmatrix},$$

and $\lambda_1 = \lambda_0./\gamma$, where $\gamma = \sqrt{n}F[2, 16, 1, 8, 0.5, 4, 32]^{\mathrm{T}}$ is non-vanishing.

The above pathological examples reside in a set of measure zero. Next, we show that when $\mathrm{rank}(X_0) = s$ but the joint support of the columns of X_0 is periodic, the pair (λ_0, X_0) is not identifiable. This set of unidentifiable X_0 has nonzero measure. Recall the proof of Theorem 3.4.3. Assume that the joint support of the columns of X_0 is periodic with period ℓ. There exists a circulant matrix P with two nonzero entries in the first column, indexed by k_1 and k_2, such that $k_2 - k_1 = \ell$ and $\gamma = \sqrt{n}FP^{(:,1)}$ is non-vanishing. Hence there exists $X_1 = PX_0$ and $\lambda_1 = \lambda_0./\gamma$ such that $\mathrm{diag}(\lambda_0)FX_0 = \mathrm{diag}(\lambda_1)FX_1$ and $\lambda_1 \notin [\lambda_0]^L_{\mathscr{T}}$. Therefore, (λ_0, X_0) is not identifiable.

APPENDIX B

PROOFS FOR CHAPTER 3

B.1 Proof of Lemma 3.3.2

1. If $A \in \mathbb{C}^{n\times m}$ has full row rank, then the rows of A form a basis for $\mathcal{R}(A)$ whose dimension is n. For every non-empty proper subset J and its complement J^c, $\mathcal{R}(A^{(J,:)})$ and $\mathcal{R}(A^{(J^c,:)})$ are two subspaces whose dimensions are $|J|$ and $|J^c|$ respectively. Therefore,

$$\begin{aligned}\mathcal{R}(A) &= \mathcal{R}(A^{(J,:)}) + \mathcal{R}(A^{(J^c,:)}),\\ \dim(\mathcal{R}(A)) &= n = |J| + |J^c| = \dim(\mathcal{R}(A^{(J,:)})) + \dim(\mathcal{R}(A^{(J^c,:)})).\end{aligned}$$

   Therefore, the sum of two subspaces is a direct sum, and the row space of A is decomposable.

2. If the row space of A is not decomposable, then A does not have full row rank. If the matrix A has full column rank, then $n \geq m$.

   Next, we prove $n > m$ by contradiction. Suppose that $n = m$. Since square matrix A has full column rank, it must have full row rank, which causes a contradiction. Therefore, the assumption is false, and n has to be greater than m.

3. The row space of A is not decomposable, if and only if the sum $\mathcal{R}(A) = \mathcal{R}(A^{(J,:)}) + \mathcal{R}(A^{(J^c,:)})$ is not a direct sum for any non-empty proper subset $J \subset \{1, 2, \cdots, n\}$, or equivalently,

$$\dim(\mathcal{R}(A)) < \dim(\mathcal{R}(A^{(J,:)})) + \dim(\mathcal{R}(A^{(J^c,:)})),$$

   for all non-empty proper subsets $J \subset \{1, 2, \cdots, n\}$.

B.2 Proofs of the Propositions Regarding "Friendliness"

Proof of Proposition 3.4.16. We prove by contraposition, i.e., if $n \geq 4$ and $|J| \leq 2$, then J is not friendly. First, if $J = \varnothing$ or $|J| = 1$, then the circularly shifted index sets $J_1, J_2, \cdots, J_{n-s}$ are not connected.

Next, we show that if $n \geq 4$ and $|J| = 2$, then $J_1, J_2, \cdots, J_{n-s}$ are not connected. Since all the circularly shifted index sets are equivalent, without loss of generality, we may assume that $J = \{1, r\}$, where $2 \leq r \leq \frac{n}{2} + 1$. Then all the sets $\{r_1, r_2\}$ such that $r_1 - r_2 = r - 1$ (modulo n) or $r_2 - r_1 = r - 1$ (modulo n) are circularly shifted versions of J. There are a total of n circularly shifted index sets.

If $r = \frac{n}{2} + 1$, then J is periodic. The sets like $\{r_1, r_1 + \frac{n}{2}\}$ $(1 \leq r_1 \leq \frac{n}{2})$ are counted twice because $\frac{n}{2} = -\frac{n}{2}$ (modulo n). And these index sets are not connected.

If $n \geq 4$ and $r < \frac{n}{2} + 1$, the n index sets are $\{1, r\}, \{2, r+1\}, \cdots, \{n-r+1, n\}, \{n-r+2, 1\}, \cdots, \{n, r-1\}$. By removing $\{r, 2r-1\}$ and $\{n-r+2, 1\}$, there are $n - 2 \geq 2$ index sets left. These circularly shifted versions of J are not connected because $J = \{1, r\}$ is not connected to the rest. □

Proof of Proposition 3.4.17. First, if J is contiguous and $|J| = s$, then $n - s$ shifted contiguous index sets cover at least $s + (n - s - 1) = n - 1$ indices. Therefore, $|\bigcup_{t=1}^{n-s} J_t| \geq n - 1$.

Next, we prove that the shifted index sets $J_1, J_2, \cdots, J_{n-s}$ are connected by showing that they form a cycle or a path in the graph. To this end, we show that between the $n - s$ pairs $(J_1, J_2), (J_2, J_3), \cdots, (J_{n-s}, J_1)$, there are at least $n - s - 1$ edges. Suppose the opposite, that there are fewer edges, for example two edges are missing in the above cycle. Then $n - s$ shifted contiguous index sets cover at least $s + s + (n - s - 2) = n + s - 2 \geq n + 1$ indices, a contradiction. □

Proof of Proposition 3.4.18. We first show that if J is not periodic, then $|\bigcup_{t=1}^{n-s} J_t| \geq n - 1$, or equivalently $|\bigcap_{t=1}^{n-s} J_t^c| \leq 1$. We prove the contrapositive, if there are two distinct indices $k', k'' \in \bigcap_{t=1}^{n-s} J_t^c$ then J is periodic. Note that $J_1^c, J_2^c, \cdots, J_{n-s}^c$ are all circularly shifted versions of the same index set

$J^c = \{j_1^c, j_2^c, \cdots, j_{n-s}^c\}$. Therefore,

$$\begin{aligned} J^c = \{j_1^c, j_2^c, \cdots, j_{n-s}^c\} &= \{k' - k_1, k' - k_2, \cdots, k' - k_{n-s}\} \\ &= \{k'' - k_1, k'' - k_2, \cdots, k'' - k_{n-s}\} \quad (\text{modulo } n), \end{aligned}$$

Hence J^c is periodic with period $\ell = |k'' - k'|$, so is J.

Next we show that the shifted index sets are connected. If $|J| > \frac{n}{2}$, then $J_{t_1} \bigcap J_{t_2} \neq \varnothing$ for any t_1, t_2. There is an edge between every pair of nodes, hence the graph is a complete graph, which is connected. □

Proof of Corollary 3.4.19. The sufficiency is shown in the proof of Proposition 3.4.18.

Next we prove necessity. If J is periodic with period ℓ and $|J| = s < n$, then for any k', k'' such that $k'' - k' = \ell$, we can always apply the proper shifts $k_1 = k' - j_1^c, k_2 = k' - j_2^c, \cdots, k_{n-s} = k' - j_{n-s}^c$ such that $k', k'' \in \bigcap_{t=1}^{n-s} J_t^c$. Hence we can pick $n - s$ shifted index sets such that $|\bigcup_{t=1}^{n-s} J_t| \leq n - 2$. □

APPENDIX C

PROOFS FOR CHAPTER 4

C.1 Proofs of Lemma 4.3.1, 4.3.2 and 4.3.3

Proof of Lemma 4.3.1. The entries of G_{DE} are multivariate polynomials in the entries of D and E, or to be more specific, quadratic forms in the entries of D and E. By Lemma 1 from [43], the matrix G_{DE} has full column rank for almost all D and E if it has full column rank for at least one choice of D and E.

We complete the proof by showing that G_{DE} has full column rank for the following choice of D and E. Let $D = F_n^{-1}\widetilde{D}$, with $\widetilde{D} \in \mathbb{C}^{n\times m_1}$ chosen such that all its submatrices have full rank. (For example, this will hold with probability 1 for a random matrix with iid Gaussian entries.) Let $E = F_n^{-1}\widetilde{E}$, with $\widetilde{E} \in \mathbb{C}^{n\times m_2}$ chosen such that the first m_1m_2 rows are the kronecker product:

$$\widetilde{E}^{(1:m_1m_2,:)} = I_{m_2} \otimes \mathbf{1}_{m_1,1}.$$

Let $\widetilde{G}_{DE} = F_nG_{DE}$, then the submatrix containing the first m_1m_2 rows of $\widetilde{G}_{DE}/\sqrt{n}$ is

$$\frac{1}{\sqrt{n}}\widetilde{G}_{DE}^{(1:m_1m_2,:)} = \begin{bmatrix} \widetilde{D}^{(1:m_1,:)} & & & \\ & \widetilde{D}^{(m_1+1:2m_1,:)} & & \\ & & \ddots & \\ & & & \widetilde{D}^{(m_1m_2-m_1+1:m_1m_2,:)} \end{bmatrix}.$$

By the assumption that all submatrices of $\widetilde{D}$ have full rank, it follows that $\widetilde{G}_{DE}^{(1:m_1m_2,:)}/\sqrt{n}$ has full column rank m_1m_2. Therefore, G_{DE} has full column rank. $\square$

Proof of Lemma 4.3.2. Let $D = [D_0, D_1, D_2]$, then G_{DE} is a permutation of the columns of $[G_{D_0E}, G_{D_1E}, G_{D_2E}]$. It is sufficient to prove that G_{DE} has full column rank, which follows from Lemma 4.3.1 because the number of columns in D is $m_1 = t_1 + 2 \times (s_1 - t_1) = 2s_1 - t_1 \le 2s_1$ and $n \ge 2s_1 m_2 \ge m_1 m_2$. $\square$

Proof of Lemma 4.3.3. Let $D = [D_0, D_1, D_2]$, $D' = [D_0, D_1]$ and $D'' = [D_0, D_2]$, then $[G_{DE_0}, G_{D'E_1}, G_{D''E_2}]$ is a permutation of all the columns of $G_{D_0E_0}$, $G_{D_1E_0}$, $G_{D_2E_0}$, $G_{D_0E_1}$, $G_{D_1E_1}$, $G_{D_0E_2}$, $G_{D_2E_2}$. By Lemma 1 from [43], it is sufficient to show that $[G_{DE_0}, G_{D'E_1}, G_{D''E_2}]$ has full column rank for at least one choice of $D_0, D_1, D_2, E_0, E_1, E_2$.

We complete the proof by showing that $[G_{DE_0}, G_{D'E_1}, G_{D''E_2}]$ has full column rank for the following choice. Let D_0, D_1, D_2 be chosen such that all submatrices of $\widetilde{D}_0, \widetilde{D}_1, \widetilde{D}_2$ have full rank. Let E_0, E_1, E_2 be chosen such that the first $2s_1s_2$ rows of $\widetilde{E}_0, \widetilde{E}_1, \widetilde{E}_2$ are

$$\widetilde{E}_0^{(1:2s_1s_2,:)} = \begin{bmatrix} I_{t_2} \otimes \mathbf{1}_{2s_1,1} \\ \hdashline \mathbf{0}_{s_1(s_2-t_2),t_2} \\ \hdashline \mathbf{0}_{s_1(s_2-t_2),t_2} \end{bmatrix}, \quad \widetilde{E}_1^{(1:2s_1s_2,:)} = \begin{bmatrix} \mathbf{0}_{2s_1t_2,s_2-t_2} \\ \hdashline I_{s_2-t_2} \otimes \mathbf{1}_{s_1,1} \\ \hdashline \mathbf{0}_{s_1(s_2-t_2),s_2-t_2} \end{bmatrix},$$

$$\widetilde{E}_2^{(1:2s_1s_2,:)} = \begin{bmatrix} \mathbf{0}_{2s_1t_2,s_2-t_2} \\ \hdashline \mathbf{0}_{s_1(s_2-t_2),s_2-t_2} \\ \hdashline I_{s_2-t_2} \otimes \mathbf{1}_{s_1,1} \end{bmatrix}.$$

By the proofs of Lemmas 4.3.1 and 4.3.2, $\widetilde{G}_{DE_0}^{(1:2s_1s_2,:)}$, $\widetilde{G}_{D'E_1}^{(1:2s_1s_2,:)}$ and $\widetilde{G}_{D''E_2}^{(1:2s_1s_2,:)}$ all have full column rank, and their nonzero entries are located in three disjoint row blocks. Hence $[\widetilde{G}_{DE_0}, \widetilde{G}_{D'E_1}, \widetilde{G}_{D''E_2}]^{(1:2s_1s_2,:)}$ has full column rank. Therefore, $[G_{DE_0}, G_{D'E_1}, G_{D''E_2}]$ has full column rank. $\square$

C.2 Proofs of the Necessary Conditions

Proof of Proposition 4.4.4. We show that if $n < m_1 + m_2 - 1$, then (x_0, y_0) is not identifiable up to scaling. Let $\widetilde{D}_\perp \in \mathbb{C}^{n \times (n-m_1)}$ denote a matrix whose columns form a basis for the orthogonal complement of the column space of $\widetilde{D}$. Then $\widetilde{D}_\perp^*$ is an annihilator of the column space of $\widetilde{D}$, i.e., $\widetilde{D}_\perp^* \widetilde{D} = 0$. Let $\widetilde{E}_{\text{inv}} \in \mathbb{C}^{n \times m_2}$ denote the entrywise inverse of $\widetilde{E}$:

$$\widetilde{E}_{\text{inv}}^{(j,k)} = \begin{cases} \frac{1}{\widetilde{E}^{(i,j)}} & \text{if } \widetilde{E}^{(i,j)} \neq 0, \\ 0 & \text{if } \widetilde{E}^{(i,j)} = 0. \end{cases}$$

Consider the linear operator $\mathcal{G} : \mathbb{C}^{m_2} \to \mathbb{C}^{n-m_1}$ defined by

$$\mathcal{G}(w) = \widetilde{D}_\perp^* \operatorname{diag}(\widetilde{E}_{\text{inv}} w) \operatorname{diag}(\widetilde{E} y_0) \widetilde{D} x_0$$

for $w \in \mathbb{C}^{m_2}$. We claim that every non-vanishing null vector of $\mathcal{G}$ produces a solution to the BD problem. Indeed, if $w_1 \in \mathcal{N}(\mathcal{G})$ is non-vanishing, then $\operatorname{diag}(\widetilde{E}_{\text{inv}} w_1) \operatorname{diag}(\widetilde{E} y_0) \widetilde{D} x_0$ is annihilated by $\widetilde{D}_\perp^*$ and therefore must reside in the column space of $\widetilde{D}$. Hence, there exists $x_1 \in \mathbb{C}^{m_1}$ such that

$$\operatorname{diag}(\widetilde{E}_{\text{inv}} w_1) \operatorname{diag}(\widetilde{E} y_0) \widetilde{D} x_0 = \widetilde{D} x_1. \tag{C.1}$$

Now, let y_1 denote the entrywise inverse of w_1. Recall that the supports of the columns of $\widetilde{E}$ are disjoint, hence $\widetilde{E} y_1$ is the entrywise inverse of $\widetilde{E}_{\text{inv}} w_1$. By Equation (C.1),

$$\operatorname{diag}(\widetilde{E} y_0) \widetilde{D} x_0 = \operatorname{diag}(\widetilde{E} y_1) \widetilde{D} x_1,$$
$$(D x_0) \circledast (E y_0) = (D x_1) \circledast (E y_1).$$

Hence (x_1, y_1) is a solution to the BD problem where $z = (D x_0) \circledast (E y_0)$. This establishes the claim.

It remains to show that $\mathcal{G}$ does have a non-vanishing null vector, and that the solution it produces does not coincide, up to scaling, with (x_0, y_0). Let w_0 denote the entrywise inverse of y_0, then $w_0 \in \mathcal{N}(\mathcal{G})$. There are $(n - m_1)$ equations in $\mathcal{G}(w) = 0$. If $n < m_1 + m_2 - 1$, then $n - m_1 \leq m_2 - 2$ and the dimension of $\mathcal{N}(\mathcal{G})$ is at least 2. Hence, there exists a vector $w_1 \in \mathcal{N}(\mathcal{G})$ such that w_0, w_1 are linearly independent. Let α be a complex number such that $0 < |\alpha| < \frac{1}{\|y_0\|_\infty \|w_1\|_\infty}$. Then $w_0 + \alpha w_1 \in \mathcal{N}(\mathcal{G})$ is non-vanishing, because the entries of $w_0 + \alpha w_1$ satisfy that

$$\left|w_0^{(j)} + \alpha w_1^{(j)}\right| \geq \left|w_0^{(j)}\right| - |\alpha| \left|w_1^{(j)}\right| \geq \frac{1}{\|y_0\|_\infty} - |\alpha| \|w_1\|_\infty > 0, \quad \text{for } j = 1, 2, \cdots, m_2.$$

Since $\alpha \neq 0$, the null vector $w_0 + \alpha w_1$ is not a scaled version of w_0. It produces a solution (x_2, y_2) in which y_2 is the entrywise inverse of $w_0 + \alpha w_1$ and is not a scaled version of y_0. Therefore, (x_0, y_0) is not identifiable up to scaling. □

Proof of Corollary 4.4.5. The vector x_0 is s_1-sparse. If we know the support of s_1, then the signal $u = Dx$ resides in a subspace spanned by s_1 columns of D and the problem reduces to BD with subspace constraints. By Proposition 4.4.4, if $n < s_1 + m_2 - 1$, then (x_0, y_0) cannot be identified up to scaling even if the support of x_0 is given. Hence (x_0, y_0) is not identifiable without knowing the support. Therefore, it is necessary that $n \geq s_1 + m_2 - 1$. $\square$

REFERENCES

[1] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Process. Mag.*, vol. 13, no. 3, pp. 43–64, May 1996.

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

[3] L. Taylor, "The phase retrieval problem," *IEEE Trans. Antennas Propag.*, vol. 29, no. 2, pp. 386–391, Mar 1981.

[4] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Appl. Opt.*, vol. 21, no. 15, pp. 2758–2769, Aug 1982.

[5] R. Rubinstein, A. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proc. IEEE*, vol. 98, no. 6, pp. 1045–1057, Jun 2010.

[6] S. Choudhary and U. Mitra, "Identifiability scaling laws in bilinear inverse problems," *arXiv preprint arXiv:1402.2637*, 2014.

[7] G. H. Golub and V. Pereyra, "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate," *SIAM J. Numerical Anal.*, vol. 10, no. 2, pp. 413–432, 1973.

[8] F. Natterer, "Numerical solution of bilinear inverse problems," *Preprints Angewandte Mathematik und Informatik*, pp. 1–25, Nov 1995.

[9] Z. Mou-yan and R. Unbehauen, "New algorithms of two-dimensional blind deconvolution," *Optical Eng.*, vol. 34, no. 10, pp. 2945–2956, 1995.

[10] D. H. Brainard and W. T. Freeman, "Bayesian color constancy," *J. Opt. Soc. Am. A*, vol. 14, no. 7, pp. 1393–1411, Jul 1997.

[11] S. Makni, C. Beckmann, S. Smith, and M. Woolrich, "Bayesian deconvolution fMRI data using bilinear dynamical systems," *NeuroImage*, vol. 42, no. 4, pp. 1381 – 1396, 2008.

[12] C. Likas and N. Galatsanos, “A variational approach for Bayesian blind image deconvolution,” *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2222–2233, Aug 2004.

[13] R. Molina, J. Mateos, and A. Katsaggelos, “Blind deconvolution using a variational approach to parameter, image, and blur estimation,” *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3715–3727, Dec 2006.

[14] S. Babacan, R. Molina, and A. Katsaggelos, “Variational Bayesian blind deconvolution using a total variation prior,” *IEEE Trans. Image Process.*, vol. 18, no. 1, pp. 12–26, Jan 2009.

[15] A. Levin, Y. Weiss, F. Durand, and W. Freeman, “Understanding blind deconvolution algorithms,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2354–2367, Dec 2011.

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

[17] E. J. Candès, T. Strohmer, and V. Voroninski, “Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming,” *Commun. Pure Appl. Math.*, vol. 66, no. 8, pp. 1241–1274, 2013.

[18] E. J. Candès and X. Li, “Solving quadratic equations via phaselift when there are about as many equations as unknowns,” *Found. Comput. Math.*, vol. 14, no. 5, pp. 1017–1026, 2014.

[19] E. J. Candès, Y. Eldar, T. Strohmer, and V. Voroninski, “Phase retrieval via matrix completion,” *SIAM J. Imaging Sci.*, vol. 6, no. 1, pp. 199–225, 2013.

[20] S. Choudhary and U. Mitra, “Sparse blind deconvolution: What cannot be done,” in *Proc. Int. Symp. Inform. Theory (ISIT)*. IEEE, June 2014, pp. 3002–3006.

[21] D. A. Spielman, H. Wang, and J. Wright, “Exact recovery of sparsely-used dictionaries,” in *Proc. 25th Annu. Conf. Learning Theory (COLT)*, vol. 23. JMLR, 2012, pp. 37.1–37.18.

[22] A. Agarwal, A. Anandkumar, and P. Netrapalli, “A clustering approach to learn sparsely-used overcomplete dictionaries,” *arXiv preprint arXiv:1309.1952*, 2013.

[23] A. Agarwal, A. Anandkumar, P. Jain, P. Netrapalli, and R. Tandon, “Learning sparsely used overcomplete dictionaries via alternating minimization,” *arXiv preprint arXiv:1310.7991*, 2013.

[24] S. Arora, R. Ge, and A. Moitra, "New algorithms for learning incoherent and overcomplete dictionaries," in *Proc. 27th Conf. Learning Theory (COLT)*, vol. 35. JMLR, 2014, pp. 1–28.

[25] S. Arora, A. Bhaskara, R. Ge, and T. Ma, "More algorithms for provable dictionary learning," *arXiv preprint arXiv:1401.0579*, 2014.

[26] H. Q. Nguyen, S. Liu, and M. N. Do, "Subspace methods for computational relighting," *Proc. SPIE*, vol. 8657, pp. 865 703.1–865 703.10, 2013.

[27] A. Paulraj and T. Kailath, "Direction of arrival estimation by eigenstructure methods with unknown sensor gain and phase," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 10. IEEE, Apr 1985, pp. 640–643.

[28] R. Morrison, M. Do, and J. Munson, D.C., "MCA: A multichannel approach to SAR autofocus," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 840–853, Apr 2009.

[29] E. Moulines, P. Duhamel, J. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel fir filters," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 516–525, Feb 1995.

[30] J. Litwin, L.R., "Blind channel equalization," *IEEE Potentials*, vol. 18, no. 4, pp. 9–12, Oct 1999.

[31] P. A. Naylor and N. D. Gaubitch, *Speech Dereverberation*, 1st ed. Springer, 2010.

[32] O. Yilmaz, *Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data.* Society of Exploration Geophysicists, 2001.

[33] T. Chan and C.-K. Wong, "Total variation blind deconvolution," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 370–375, Mar 1998.

[34] K. Herrity, R. Raich, and A. O. Hero III, "Blind reconstruction of sparse images with unknown point spread function," *Proc. SPIE*, vol. 6814, pp. 68 140K.1–68 140K.11, 2008.

[35] D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *Proc. Conf. Comput. Vision and Pattern Recognition (CVPR).* IEEE, June 2011, pp. 233–240.

[36] M. Salman Asif, W. Mantzel, and J. Romberg, "Random channel coding and blind deconvolution," in *Proc. 47th Annu. Allerton Conf. Commun., Control, and Computing*, Sept 2009, pp. 1021–1025.

[37] A. Repetti, M. Pham, L. Duval, E. Chouzenoux, and J.-C. Pesquet, "Euclid in a taxicab: Sparse blind deconvolution with smoothed ℓ_1/ℓ_2 regularization," *IEEE Signal Process. Lett.*, vol. 22, no. 5, pp. 539–543, May 2015.

[38] S. Ling and T. Strohmer, "Self-calibration and biconvex compressive sensing," *arXiv preprint arXiv:1501.06864*, 2015.

[39] K. Lee, Y. Li, M. Junge, and Y. Bresler, "Stability in blind deconvolution of sparse signals and reconstruction by alternating minimization," *Int. Conf. Sampling Theory and Applications (SampTA)*, 2015.

[40] G. E. Bredon, *Introduction to compact transformation groups.* Academic press, 1972, vol. 46.

[41] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proc. Natl. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, 2003.

[42] V. Sanchez, P. Garcia, A. Peinado, J. Segura, and A. Rubio, "Diagonalizing properties of the discrete cosine transforms," *IEEE Trans. Signal Process.*, vol. 43, no. 11, pp. 2631–2641, Nov 1995.

[43] G. Harikumar and Y. Bresler, "FIR perfect signal reconstruction from multiple convolutions: minimum deconvolver orders," *IEEE Trans. Signal Process.*, vol. 46, no. 1, pp. 215–218, Jan 1998.