SUBSPACE LEARNING BY RANDOMIZED SKETCHING

A Dissertation
Presented to
The Academic Faculty

By

Rakshith Sharma Srinivasa

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
Department of Electrical and Computer engineering

Georgia Institute of Technology

December 2020

SUBSPACE LEARNING BY RANDOMIZED SKETCHING

Thesis committee:

Dr. Justin Romberg
School of Electrical and Computer engineering
Georgia Institute of Technology

Dr. Kiryung Lee
Department of Electrical and Computer engineering
Ohio State University

Dr. Mark Davenport
School of Electrical and Computer engineering
Georgia Institute of Technology

Dr. Vladimir Koltchinskii
Department of Mathematics
Georgia Institute of Technology

Dr. Santosh Vempala
College of Computing
Georgia Institute of Technology

Date approved: December, 2020

To my kind family and friends.

ACKNOWLEDGMENTS

My time as a graduate student at Georgia Institute of Technology has been pivotal and truly formative. First and foremost, I would like to thank my advisor, Dr. Justin Romberg, for his constant support and guidance. He has been and will be a constant source of inspiration. His advice and direction have shaped my technical outlook and communication skills. I truly admire his work ethic and his sense of kindness. I am grateful to have had the opportunity to work with him.

I would like to thank Dr. Mark Davenport, whose advice was instrumental in my decision to pursue a Ph.D. Collaborating with Dr. Davenport has been truly enjoyable and formative. I thank him immensely for his support and encouragement. I have always cherished our conversations and I am constantly inspired by his intellectual curiosity and vast knowledge.

I would like to thank Dr. Kiryung Lee for being a very valuable mentor, collaborator, and friend. He has been instrumental in shaping my research interests and I am very grateful for his support. He has also on numerous occasions, provided me with valuable advice regarding my career.

I would also like to thank Dr. Vladimir Koltchinskii and Dr. Santosh Vempala for being part of my thesis committee. Their research and teaching have influenced me greatly and have shaped my own interests. I thank them for their time and valuable feedback on my thesis.

I would like to thank Santhosh Karnik and Coleman Delude for the wonderful collaboration we have had. I would also like to thank Sohail Bahamani, with whom I have had many insightful discussions.

I would like to thank Matt, Marissa, Greg, Nick, John, Michael, Andy, Tomer, Kyle, Pavel, Liangbei, Namrata, Andrew, Jihui, Sihan, Nishant, Nauman, Adam, Thinh, and the rest of the *children-of-the-norm* family. I feel very fortunate to have been part of this

group. I have enjoyed the many discussions, paper writing “challenges” and the white-board brainstorming sessions. I will also always cherish the many cooking and potluck-ing sessions and I look forward to many more in the future.

I would like to thank all of my friends from outside the lab. I owe them a debt of gratitude for helping me fix broken bones and hearts, and for bearing with me in general.

One of the best outcomes of my time at Georgia Tech is having met Soup a.k.a Madam Speaker. Her companionship, support, and the constant reminding of my aging have been truly uplifting. I am very grateful to her and I am most excited about our future together.

Finally, I would like to express my sincerest gratitude to my parents. They have been immensely open-minded, supportive, and kind to me. Their love and care give me inspiration and strength. I would not be the person I am without the upbringing they provided and their encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

High dimensional data is often accompanied by inherent low dimensionality that can be leveraged to design scalable machine learning and signal processing algorithms. Developing efficient computational frameworks that take advantage of the underlying structure in the data is crucial. In this thesis, we consider a particular form of inherent low dimensionality in data: subspace models. In many applications, data is known to lie close to a low dimensional subspace. The underlying subspace itself may or may not be known a priori. Incorporating this structure into data acquisition systems and algorithms can aid in scalability.

We first consider two specific applications in the field of array signal processing where subspace priors on the data are commonly used. For both these applications, we develop algorithms that require a number of measurements that scale with only the dimension of the underlying subspace. This is achieved by using linear dimensionality reduction maps on the signals received by the antenna array. In doing so, we show that such applications impose constraints on the way dimensionality reduction may be performed. In particular, arrays demand dimensionality reduction maps that can operate on individual subsets or blocks of data at a time, without having access to other blocks. We show how such a constraint can be incorporated into designing data acquisition systems that can still scale with the inherent low dimensionality.

Inspired by the block constraints imposed by an antenna array, we consider more general problems in numerical linear algebra where the data has a natural partition into blocks. This is common in applications with distributed or decentralized data. We study the problems of sketched ridge regression and sketched matrix multiplication under this constraint and give sample optimal theoretical guarantees on block diagonal sketching matrices.

Extending the block model to low-rank matrices, we then study the problem of recovering a low-rank matrix from compressed observations of each column. While each column

itself is compressed to a point that is beyond recovery, we leverage their joint structure to recover the matrix as a whole. To do so, we establish a new framework to design estimators of low-rank matrices that obey the constraints imposed by different observation models. Using this framework, we develop an algorithm to recover low-rank matrices from column-wise compressed observations. We also give favorable theoretical guarantees on our algorithm and provide a fast implementation using the Alternating direction method of multipliers (ADMM) framework.

Finally, we extend our framework to design low-rank matrix estimators for the application of blind deconvolution. Since it is a special instance of the general low-rank matrix recovery problem, we apply our previously established framework to design a low-rank estimator to this problem. This yields a novel estimation algorithm for which we provide uniform recovery guarantees that are sample optimal.

CHAPTER 1
INTRODUCTION AND BACKGROUND

1.1 Introduction

The famous curse of dimensionality states that as the dimension of data grows, statistically reliable inference requires an amount of data that grows exponentially in the dimension. A crucial observation that aids in circumventing such an unrealistic demand for data is that high dimensional data is very often highly structured, thus resulting in an *inherent low dimensionality*. Sparsity, subspaces, low dimensional manifolds, low-rank matrices are among the many ways in which low-dimensionality can be effectively modeled. Inherent low dimensional structure when incorporated into algorithms directly obviates the need for an exponential amount of data in the original dimension.

In this thesis, we focus on one particular form of structure found in high dimensional data, namely the subspace model. Under such a model, data in n dimensions is assumed to lie close to a subspace of dimension d such that $d \ll n$. The subspace itself may or may not be known *a priori*. In either scenario, a natural question is how to incorporate the knowledge of the existence of such a structure into both data acquisition systems and inference algorithms. The hope in doing so is to reduce the requirement on data acquisition, data storage and algorithm run-time from $O(f(n))$ to $O(f(d))$ where $f(\cdot)$ is any suitable function.

A clear motivating example arises in application from array signal processing. In a standard setting, an array of antenna elements sample and collect incoming signals. Some of the common goals are to identify the direction of the incoming signal or to reconstruct an image of the signal source. Let us consider an antenna array with 64x64 elements. If the elements are sampled at 25GHz, a completely reasonable rate in many applications,

the rate of acquisition is 12.8TB per second, more than 100 times the rate at which data is uploaded to YouTube worldwide! Another related example is the Square Kilometer Array (SKA) [1], where antenna elements are spread across the continents of Africa and Australia collecting data at rates of about 157TB/s! It is clear that even basic algorithms might fail at such a scale. However, as we shall see in the sequel, for many array processing tasks, the subspace-based prior provides a very realistic model and directly helps reduce the scale of the required operations.

In the rest of the thesis, we will present several problems where the subspace-based modeling is relevant and design data acquisition methods and algorithms to take advantage of the inherent low-dimensionality. In all of the problems we discuss, the method of choice will be linear dimensionality reduction of the data to a dimension that depends only on the subspace dimension and a tolerance parameter. We then develop algorithmic solutions that solve the original problem and attain the desired accuracy level. In designing such a linear map, we take inspiration from a range of methods loosely termed *sketching*, developed by the theoretical computer science community.

1.2 Sketching for numerical linear algebra

Large scale linear algebraic computations form the backbone of most signal processing and machine learning algorithms. Accelerating such computations is key to scale them to large real-world datasets. With this goal in mind, the *sketching* paradigm aims to solve problems in linear algebra only approximately, but with much smaller time and space complexities. They employ randomization of the data as a tool to reduce the dimensionality of the data while preserving the information required by downstream tasks. A fundamental result in theoretical computer science that captures this spirit is the famous Johnson-Lindenstrauss lemma, a version of which is stated below.

Theorem 1 *(Johnson-Lindenstrauss Lemma [2]) Let* $\mathbf{u}_i \in \mathbb{R}^n,\ i \in \{1, \cdots, m\}$ *and let* $\mathbf{R} \in \mathbb{R}^{k \times n}$ *such that* $\mathbf{R}_{ij} \sim \mathcal{N}(0,1)$ *and consider* $\mathbf{v}_i = \frac{1}{\sqrt{k}}\mathbf{R}\mathbf{u}\ \forall i$. *Then, if* $k =$

$O(\log(m)/\epsilon^2)$, *we have*

$$(1-\epsilon)\left\|\mathbf{u}_i-\mathbf{u}_j\right\|^2 \leqslant \left\|\mathbf{v}_i-\mathbf{v}_j\right\|^2 \leqslant (1+\epsilon)\left\|\mathbf{u}_i-\mathbf{u}_j\right\|^2 \tag{1.1}$$

for all i, j *with probability at least 0.9.*

Algorithms that depend on the Euclidean distance between the data points can then be directly designed in the lower dimensional space. The Johnson-Lindenstrauss lemma caters to any discrete set of m data points. Generalizing such dimensionality reduction maps to more complex sets of points (including subspaces) forms the core of sketching based methods. In this section, we briefly review sketching methods in the context of linear regression and matrix multiplication. They have a direct relationship with the methods developed in the rest of the thesis.

Let $\mathbf{A} \in \mathbb{R}^{n\times d}$ be a matrix with $n \gg d$, $\mathbf{y} \in \mathbb{R}^n$ be a vector of observations. The linear regression problem then tries to find the best fit for $\mathbf{y}$ in the column space of $\mathbf{A}$ by solving the problem

$$\arg\min_{\mathbf{x}} \left\|\mathbf{y}-\mathbf{A}\mathbf{x}\right\|^2. \tag{1.2}$$

Solving (Equation 1.2) requires $O(nd^2)$ time, which can be slow for large datasets. Now let $\mathbf{S} \in \mathbb{R}^{m\times n}$ with $d < m \ll n$ and consider the optimization problem

$$\arg\min_{\mathbf{x}} \left\|(\mathbf{S}\mathbf{y})-(\mathbf{S}\mathbf{A})\mathbf{x}\right\|^2. \tag{1.3}$$

This optimization problem takes only $O(md^2)$ time. If $\mathbf{S}_{ij} \sim \mathcal{N}(0, 1/m)$, then the following result guarantees that the solutions of the two optimization programs are close in terms of the prediction error.

Theorem 2 *[3] Let* $0 < \epsilon < 1$ *and* $\delta < 1$*, and* $\mathbf{S}_{ij} \sim \mathcal{N}(0, 1/m)$*. Let* $\mathbf{x}_0, \widehat{\mathbf{x}}$ *be the solutions*

to (Equation 1.2) *and* (Equation 1.3) *respectively and* $m = O((d + \log(1/\delta))\epsilon^{-2})$. *Then,*

$$(1-\epsilon)\left\|\mathbf{A}\mathbf{x}_0 - \mathbf{y}\right\|^2 \leqslant \left\|\mathbf{A}\widehat{\mathbf{x}} - \mathbf{y}\right\|^2 \leqslant (1+\epsilon)\left\|\mathbf{A}\mathbf{x}_0 - \mathbf{y}\right\|^2 \tag{1.4}$$

with probability at least $1 - \delta$.

This can be restated as a **subspace embedding** guarantee that preserves all pairwise distances between points in a subspace:

$$(1-\epsilon)\left\|\mathbf{z}\right\|^2 \leqslant \left\|\mathbf{S}\mathbf{z}\right\|^2 \leqslant (1+\epsilon)\left\|\mathbf{z}\right\|^2 \tag{1.5}$$

with probability at least $1 - \delta$ for all vectors $\mathbf{z}$ in the subspace spanned by the columns of the matrix $[\mathbf{A}\ \mathbf{y}]$. Choosing $\mathbf{S}$ as in Theorem 2 will not actually result in an overall algorithm that is faster than $O(nd^2)$, since computing $\mathbf{SA}$ itself takes $O(mdn) > O(nd^2)$. However, it captures the core idea of sketching. When $\mathbf{S}$ is such that the product $\mathbf{SA}$ can be computed quickly, the algorithmic contributions are more tangible.

More recent results show that $\mathbf{S}$ can be drawn from a suite of sparse and structured distributions that reduce the overall of complexity of computing $\mathbf{SA}$, resulting in a program that solves (Equation 1.3) in $O(\mathrm{nnz}(\mathbf{A})) + \widetilde{O}(d^3\epsilon^{-2})$ time where nnz() denotes the number of non-zeros. One category of such distributions on $\mathbf{S}$are the *sampling* matrices, where each row in $\mathbf{S}$ samples a single row from the matrix $\mathbf{A}$ in a randomized fashion. For such sampling methods, the distribution from which the rows are sampled determines if the corresponding sketch generates a subspace embedding. One particular distribution that is guaranteed to do so is given by the leverage score. Let $\mathbf{U}$ be an orthobasis for the column space of the matrix $\mathbf{A}$. Then the leverage score of the k-th row is defined as $\left\|\mathbf{U}(k,:)\right\|^2$. If the sampling distribution is proportional to the leverage scores, then the sketching process will generate a subspace embedding with high probability for $O(d \log d/\epsilon^2)$ number of samples. In chapter 3, we will encounter a similar notion of block leverage scores that determine how effective block diagonal sketching matrices are in solving the regression

problem.

Other distributions include sparse (but not sampling) matrices that take sparse linear combinations and Subsampled Randomized Hadamard Transforms (SRHT). A detailed account of these families of sketching matrices and their use in regression is provided in [3]. Further, when $\text{rank}(\mathbf{A}) < d$, similar sketching based methods can be used to solve the more stable ridge regression problem. In this case, [4] shows that the reduced dimension m can depend only on the *effective rank* of $\mathbf{A}$ (a notion that captures spectral decay), which can be much smaller than d.

The second problem of interest is that of computing approximate matrix products. Given two matrices $\mathbf{A} \in \mathbb{R}^{n \times p}$, $\mathbf{B} \in \mathbb{R}^{n \times q}$ with $n \gg p, q$, computing the product $\mathbf{A}^\top \mathbf{B}$ takes $O(npq)$ time. Taking the sketching route as before, we seek a matrix $\mathbf{S} \in \mathbb{R}^{m \times n}$ with $m \ll n$ such that

$$\left\| (\mathbf{SA})^\top (\mathbf{SB})^\top - \mathbf{A}^\top \mathbf{B} \right\|_X \leqslant \epsilon \left\| \mathbf{A} \right\|_X \left\| \mathbf{B} \right\|_X \tag{1.6}$$

where $\left\| \cdot \right\|_X$ can denote any matrix norm. A special case is where $\mathbf{B} = \mathbf{A}$ and the goal is to compute a Grammian matrix of the form $\mathbf{A}^\top \mathbf{A}$. Computing the product $(\mathbf{SA})^\top (\mathbf{SB})^\top$ takes only $O(mpq)$ time, not accounting for the time taken to compute the individual sketched matrices $\mathbf{SA}$ and $\mathbf{SB}$.

Sampling methods view the product $\mathbf{A}^\top \mathbf{B}$ as a sum of outer products of the rows of $\mathbf{A}$ and $\mathbf{B}$:

$$\mathbf{AB} = \sum_{k=1}^{n} \mathbf{A}(k,:)^\top \mathbf{B}(k,:). \tag{1.7}$$

The idea of sampling then aims to sample and rescale m rows of $\mathbf{A}$ and $\mathbf{B}$ and approximate the above sum by the outer product of just the m sampled rows. By using an appropriate distribution such as length-square distribution it can be shown that (Equation 1.6) hold in a Frobenius norm sense when $m = \Omega(1/\epsilon^2)$ [5, 6]. One particular distribution that has been shown to be effective is the so-called length-squared distribution: each row is sampled with a probability $p_k = \left\| \mathbf{A}(k,:) \right\|^2 / \left\| \mathbf{A} \right\|_F^2$.

A similar guarantee in the spectral norm sense can be achieved by drawing $\mathbf{S}$ from any of distributions with subspace embedding guarantees (as for regression) and $m = O((r + \log(1/\delta))/\epsilon^2)$. Further, when the matrices $\mathbf{A}$ and $\mathbf{B}$ are only approximately low-rank, [4] provides a bound on m that depends only on the *stable rank* of the matrices $\mathbf{A}$ and $\mathbf{B}$. Here, stable rank is defined as

$$\text{stable rank}(\mathbf{A}) = \frac{\|\mathbf{A}\|^2}{\|\mathbf{A}\|_F^2} \tag{1.8}$$

and is always upper bounded by the rank. This provides a tighter upper bound on the required sample complexity.

Finally, sketching methods are also highly useful in many related problems such as low-rank approximations of matrices, tensor regression and general ℓ_p-regression. [3].

1.3 Subspace learning for array signal processing

Subspace models are highly relevant to the field of array signal processing. In this thesis, we address two particular applications: i) Active imaging of range-limited targets and ii) Broadband source localization. For both applications, we will use sketching-inspired methods to design data acquisition systems that scale with the inherent dimensionality of the data rather than the ambient dimension.

Active array imaging

In active array imaging, a transmitter emits an excitation signal and an image of the target is reconstructed using reflections collected at an array of sensors. This is illustrated in Figure 1.1, where a target at coordinates (r_0, θ_0) is reflecting the signals emitted by the array. Let $\mathbf{x}_0 \in \mathbb{R}^N$ denote the target scene to be imaged. For an array with M elements, the measurements obtained with an excitation signal of wavelength λ_i follow a linear model:

$$\mathbf{y}_\lambda = \mathbf{A}_\lambda \mathbf{x}_0 \tag{1.9}$$

where $\mathbf{A}_\lambda \in \mathbb{C}^{M \times N}$ models the signal propagation and depends on the physical location

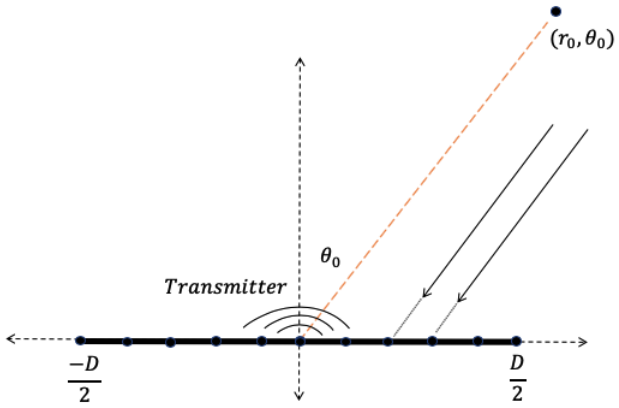


Figure 1.1: *The array imaging setup in the 1D case. R represents a point on the target image. The antenna array, illustrated using the set of dots, lies on the X-axis in the region $[\frac{-D}{2}, \frac{D}{2}]$. A single point target at (r_0, θ_0) reflects the signals emitted by the array.*

and extent of the target scene $\mathbf{x}_0$. Under broadband imaging, the excitation wavelength ranges over $\{\lambda_{\min} = \lambda_1, \cdots, \lambda_K = \lambda_{\max}\}$ and we obtain a total MK measurements of the target scene. The linear model for this ensemble of measurements is given as

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_K \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \vdots \\ \mathbf{A}_K \end{bmatrix} \mathbf{x}_0 = \mathbf{A}\mathbf{x}_0. \tag{1.10}$$

Let us consider a specific signal model, namely *range-limited* target scenes. Range-limited implies that the extent or the depth of the scene along a radial direction from the center of the array is limited. For such a signal model, the induced linear model $\mathbf{A}$ is of a much lower rank than MK. Owing to this structure, a common sketching operator across all the wavelengths can reduce the number of measurements to be collected, thus reducing the cost of operating the array. The resulting linear model follows that of (Equation 1.3),

but the sketching matrix has a very unique structure:

$$\Phi\mathbf{y} = \begin{bmatrix} \phi & 0 & \cdots & 0 \\ 0 & \phi & \cdots & 0 \\ . & . & . & . \\ 0 & 0 & \cdots & \phi \end{bmatrix} \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \vdots \\ \mathbf{A}_K \end{bmatrix} \mathbf{x}_0 \tag{1.11}$$

where $\phi \in \mathbb{R}^{\ell \times M}$ with $\ell \ll M$ and each row in ϕ contains the weights to obtain a single linear combination of the array outputs. The goal is then to design Φ such that the minimum-norm solutions to the two optimization problems below match:

$$\widehat{\mathbf{x}}_{LS} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 ; \;\; \widehat{\mathbf{x}}_{SLS} = \arg\min_{\mathbf{x}} \|\Phi\mathbf{y} - \Phi\mathbf{A}\mathbf{x}\|^2 . \tag{1.12}$$

It is worth noting here that the minimum norm solutions to the two optimization problems coincide when the row spaces of the matrices $\mathbf{A}$ and $\Phi\mathbf{A}$ coincide. It is equivalent to *learning the row space* of the matrix $\mathbf{A}$ using *structured low dimensional projections*. Hence, active array imaging using coded measurements is equivalent to solving a *sketched least squares* problem. We address this problem in chapter 2 and provide theoretical guarantees on the sample complexity ℓ for which the row spaces of the matrices $\mathbf{A}$ and $\Phi\mathbf{A}$ coincide.

Broadband source localization

Let a signal $x(t)$ be bandlimited to $[-\Delta, \Delta]$. Further, let it be sampled at a frequency $f_s > 2\Delta$ and consider N consecutive samples. Such a sampled signal has a 'digital bandwidth' $W = 2\Delta/f_s < 0.5$ and lies in a subspace of dimension approximately $2NW$. The basis for such a signal is given by the eigenvectors of the so-called Prolate matrix, defined as

$$\mathbf{B}_{N,W}(m,n) = \frac{2\pi W \sin(m-n)}{\pi(m-n)}. \tag{1.13}$$

Most real-world broadband signals bandlimited to $W < 0.5$ lie in the subspace spanned by the top roughly $2NW$ eigenvectors of the above matrix. If the signal $x(t)$ is modulated

to a frequency f_c, the resulting subspace also gets modulated by the corresponding digital sinusoid, given by $\mathbf{e}_{f_c}[n] = \exp -j2\pi n f_c/f_s$.

Given a superposition of a number of broadband signals, broadband source localization seeks to identify the frequency support of each of the signals. For sampled signals, we can model the superposition as

$$\mathbf{x}[n] = \sum_{\ell=1}^{L} \mathbf{x}_i[n],\ n = 1,\ \cdots,\ N. \tag{1.14}$$

where $\mathbf{x}_i[n]$'s are uniform samples of a broadband signal $x_i(t)$. Although there are N available samples of $\mathbf{x}$, the inherent dimensionality of such a signal can be much lower, depending on the bandwidth of the signals x_i. Since each of the vectors $\mathbf{x}_i$ lies approximately in a subspace of dimension $2NW$, their superposition will lie in a union of subspaces with a number of degrees of freedom upper bounded by $2NLW$.

Unlike in active array imaging, the underlying subspaces are unknown in the source localization problem. We only have knowledge of the set of candidate subspaces characterized by the center frequencies of each of the signals and the eigenvectors of the Prolate matrix. In chapter 2, we develop methods to *learn the active subspaces* using compressed measurements.

1.4 Sketched regression for decentralized data

The dimensionality reduction maps considered in the standard sketching literature usually require access to all or arbitrary subsets of the data matrices, as shown in Figure 1.2. However, in many applications, the data matrices are not available as a whole at a single location. For example, sketching for active imaging (section 1.3) required a block diagonal sketching matrix with repeated blocks on the diagonal.

We extend our previous results on sketching with block diagonal matrices: we assume that the blocks are no longer constrained to be the same or even have the same size and

analyze the Tikhonov regularized least squares problem. Also known as ridge regression, it is a commonly used method in regularizing linear regression problems, especially when the linear model has a decaying spectrum.

Formally, if $\mathbf{A} \in \mathbb{R}^{\tilde{N} \times d}$ is the matrix of covariates and $\mathbf{b} \in \mathbb{R}^{\tilde{N}}$ contains labels, the ridge regression problem is

$$\min_{\mathbf{x} \in \mathbb{R}^d} f(\mathbf{x}) := \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \lambda \|\mathbf{x}\|^2 . \tag{1.15}$$

The parameter λ is chosen to provide stability by overriding small singular values of $\mathbf{A}$. As before, the main idea is to expedite solving for $\mathbf{x}$ by applying a compression matrix $\mathbf{S}$ and then solving

$$\min_{\mathbf{x} \in \mathbb{R}^d} f_{\mathbf{S}}(\mathbf{x}) := \|\mathbf{S}\mathbf{A}\mathbf{x} - \mathbf{S}\mathbf{b}\|^2 + \lambda \|\mathbf{x}\|^2 , \tag{1.16}$$

where $\mathbf{S} \in \mathbb{R}^{\widetilde{M} \times \tilde{N}}$ with $\widetilde{M} \ll \tilde{N}$. Of course, the solution $\mathbf{x}_*$ to (Equation 1.15) and the solution $\hat{\mathbf{x}}$ to (Equation 1.16) will not be exactly the same. But if $\mathbf{A}$ has effective rank (defined as $\sum_i \frac{\sigma_i^2}{\sigma_i^2 + \lambda}$) r, then simply drawing the entries of $\mathbf{S}$ independently from a subgaussian distribution with $\widetilde{M} = \Omega(r/\epsilon)$ results in [4]$f(\hat{\mathbf{x}}) \leqslant (1 + \epsilon) f(\mathbf{x}_*)$ for small ϵ. Guarantees for sketching matrices that are sparse (and so can be applied efficiently) can also be found in [4], although with slightly worse sample complexities.

W are interested in a more structured sketching matrix. We consider an $\mathbf{S}$ that is block diagonal and under mild assumptions on $\mathbf{A}$, show that it can be as effective as standard sketching matrices. In particular, we let $\mathbf{S}$ assume the following structure:

$$\mathbf{S} = \mathbf{S}_D = \begin{bmatrix} \mathbf{S}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{S}_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \mathbf{S}_J \end{bmatrix} \tag{1.17}$$

with $\mathbf{S}_j$ of size $M_j \times N$ and $JN = \tilde{N}$. Each $\mathbf{S}_j$ provides M_j random projections that are *localized*, to the data block $\mathbf{A}_j$, with a total sample complexity of $\widetilde{M} = \sum_j M_j$.

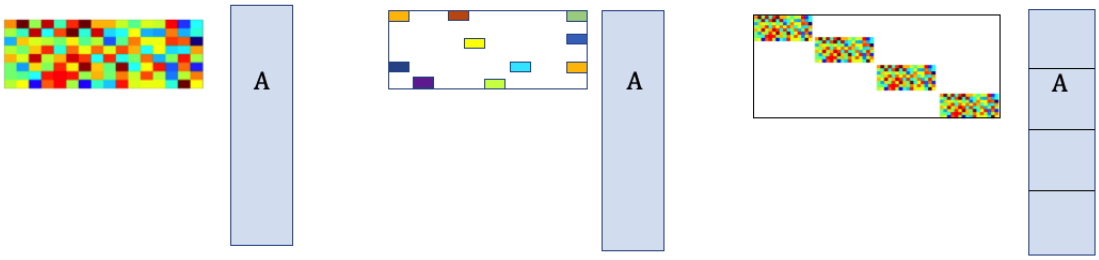


Figure 1.2: *Existing sketching strategies such as dense sub-Gaussian, SRHT matrices (left) and sparse sketching matrices (center) assume access to all or a few arbitrarily placed rows of* $\mathbf{A}$*. However, our localized model (right) needs access to only well-separated parts of the data matrix.*

Our motivation to study this problem is multifold. First, computing the product $\mathbf{S}_D\mathbf{A}$ can be much cheaper when compared with an unstructured random projection. For generic $\mathbf{S}_j$, the sketch $\mathbf{S}_D\mathbf{A}$ can be computed in time ($O(Nd\widetilde{M})$), as compared to the $O(\widetilde{N}d\widetilde{M})$ required for a dense, unstructured sketch. Second, the sketching computation is trivial to parallelize into J blocks, each requiring $O(NdM_j)$ time. Along with these computational advantages, the fact that the subsketches $\mathbf{S}_j\mathbf{A}_j$ can be computed independently gives us a method for sketching when the data is acquired in a distributed fashion. In fact, our block diagonal sketching framework can be thought of as a generalization of row sampling methods that randomly select rows of $\mathbf{A}$ based on their norms (i.e., leverage scores). The $\mathbf{S}_j$ for row sampling would be diagonal, with only 1's and 0's on the diagonals (modulo scaling).

En route to establishing theoretical guarantees on block diagonal sketching matrices for ridge regression, we establish guarantees on their utility in the approximate matrix product problem. For any two matrices $\mathbf{W} \in \mathbb{R}^{n\times m}$ and $\mathbf{Y} \in \mathbb{R}^{n\times p}$, we consider approximating the product $\mathbf{W}^\top\mathbf{Y}$ with $(\mathbf{S}_D\mathbf{W})^T(\mathbf{S}_D\mathbf{Y})$ and give bounds on the error as below.

$$\left\|(\mathbf{S}_D\mathbf{W})^T(\mathbf{S}_D\mathbf{Y}) - \mathbf{W}^T\mathbf{Y}\right\| \leqslant \epsilon \left\|\mathbf{W}\right\| \left\|\mathbf{Y}\right\|, \tag{1.18}$$

with high probability. The state of the art results show that the sample complexity (number of rows in $\mathbf{S}$) depends only on the stable ranks of the matrices $\mathbf{W}$ and $\mathbf{Y}$. Stable rank of a

matrix is defined as

$$\mathbf{sr}(\mathbf{M}) = \frac{\|\mathbf{M}\|_F^2}{\|\mathbf{M}\|_2^2}. \tag{1.19}$$

Note that $\mathbf{sr}(\mathbf{M}) \leqslant \text{rank}(\mathbf{M})$. We are able to provide similar bounds for block diagonal matrices.

To demonstrate that block diagonal matrices can have optimal sample complexity, we show simulation results on synthetic data: $N = 2000$, $J = 10$, $d = 50$. We design the singular values such that for $\lambda = 0.15$, $\mathbf{sd}_\lambda = 8.5$, but $\text{rank}(\mathbf{A}) = 50$. For each trial, we generate $\mathbf{S}$ with entries drawn from $\mathcal{N}(0, 1/\sqrt{\widetilde{M}})$ and $\mathbf{S}_D$ with the entries of $\mathbf{S}_j$ drawn from $\mathcal{N}(0, 1/\sqrt{M_j})$. In Figure 3.2, we plot the quantity $f(\hat{x})/f(x^*)$ averaged over 10 trials for different values of $\widetilde{M}$.

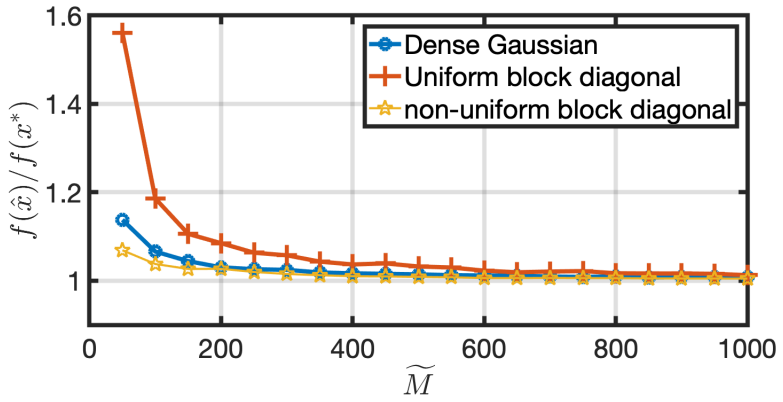


Figure 1.3: *$f(\hat{x})/f(x^*)$ for three sketching matrices: a dense matrix with standard Gaussian entries, a block diagonal matrix with equal sized blocks (uniform diagonal matrix) and a block diagonal matrix with block sizes chosen as $M_j = M_0\Gamma(U_j)$. A ratio close to 1 indicates that the sketching matrix is effective in solving* (Equation 3.3)*. When M_j's are chosen appropriately, block diagonal matrices can be as effective as a general matrix.*

1.5 Decentralized sketching for low-rank matrices

In this problem, we are interested in recovering data points that belong to a low dimensional subspace from compressed observations of each data point. We assume that the subspace itself is unknown a priori. By themselves, each of the vectors is compressed past the point where it can be recovered, but by recovering them jointly, their mutual structure

can be taken advantage of. The compression itself takes the form of a randomized linear dimensionality reducing map, similar to sketching.

Let $\mathbf{x}_i \in \mathbb{R}^{d_1}$ for $i = 1, \cdots, d_2$ be vectors from a subspace of dimension r and let $\mathbf{X}_0 = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_{d_2}]$. We observe $L < d_1$ noisy random linear measurements of each column of $\mathbf{X}_0$:

$$y_{l,i} = \langle \mathbf{A}_{l,i}, \mathbf{X}_0 \rangle + z_{l,i}, \tag{1.20}$$

$$\mathbf{A}_{l,i} = \frac{1}{\sqrt{L}} \mathbf{b}_{l,i} \mathbf{e}_i^\top \tag{1.21}$$

for $l = 1, \cdots L$ and $i = 1, \cdots d_2$, where $(b_{l,i})$ vectors of length d_1 and $\mathbf{e}_i \in \mathbb{R}^{d_2}$ is the l^{th} standard basis vector of length d_2. We assume that $z_{l,i}$ is noise drawn from $\mathcal{N}(0, \sigma^2)$. Our main motivation for this work is to address problems of distributed data acquisition and high dimensional PCA. For the sequel, we define $d = d_1 + d_2$.

Mixed-norm as a novel convex relaxation

Since the matrix $\mathbf{X}_0$ is of rank-r, we can cast the problem of recovering the vectors as a low-rank matrix recovery problem. We propose a novel relaxation for the set of low-rank matrices, using which $\mathbf{X}_0$ can be recovered from $y_{l,i}$ in (Equation 1.20) with almost optimal sample complexity L.

For a matrix $\mathbf{X} \in \mathbb{R}^{d_1 \times d_2}$ we define the maximum ℓ_2-norm of the columns of $\mathbf{X}$ as

$$\|\mathbf{X}\|_{1\to 2} = \max_{j=1\cdots d_2} \|\mathbf{X}\mathbf{e}_j\|_2 , \tag{1.22}$$

where $\mathbf{e}_j$ is the j^{th} standard basis vector. We also define the "mixed-norm" of a matrix $\mathbf{X}$ as

$$\|\mathbf{X}\|_{\text{mixed}} = \inf_{\mathbf{U},\mathbf{V}:\mathbf{U}\mathbf{V}^\top = X} \|\mathbf{U}\|_{\text{F}} \left\|\mathbf{V}^\top\right\|_{1\to 2} . \tag{1.23}$$

The above pair of norms arise as a result of interpreting a matrix $\mathbf{X} \in \mathbb{R}^{d_1 \times d_2}$ as an operator between the Banach spaces $\ell_1^{d_2}$ and $\ell_2^{d_1}$. Their utility in characterizing the set of low-rank

matrices follows from the following property:

Lemma 1 *Let* $\mathbf{X} \in \mathbb{R}^{d_1 \times d_2 a}$ *satisfy* $\operatorname{rank}(\mathbf{X}) \leqslant r$. *Then*

$$\|\mathbf{X}\|_{1\to 2} \leqslant \|\mathbf{X}\|_{\text{mixed}} \leqslant \sqrt{r}\,\|\mathbf{X}\|_{1\to 2}\,. \tag{1.24}$$

For matrices whose column norms are bounded uniformly by a constant α, their mixed-norm norm is bounded by $\sqrt{r}\alpha$. Hence, the pair of norms $\|\cdot\|_{1\to 2}$ and $\|\cdot\|_{\text{mixed}}$ together characterize the set of low-rank matrices with bounded column norms. To recover $\mathbf{X}_0$, we optimize a suitable loss function over the following set:

$$\kappa(\alpha, R) = \{\mathbf{X}\colon\ \|\mathbf{X}\|_{1\to 2} \leqslant \alpha,\ \|\mathbf{X}\|_{\text{mixed}} \leqslant R\}\,. \tag{1.25}$$

Note that the set of rank-r matrices with column norms bounded by α is a subset of $\kappa(\alpha, R)$ when $R \geqslant \alpha\sqrt{r}$. We use the following optimization to recover $\mathbf{X}_0$:

$$\begin{aligned} \underset{\mathbf{X}}{\text{minimize}} \quad & \sum_{l,i} |y_{l,i} - \langle \mathbf{A}_{l,i}, \mathbf{X}\rangle|^2 \\ \text{subject to} \quad & \mathbf{X} \in \kappa(\alpha, R). \end{aligned} \tag{1.26}$$

Our main result, stated in chapter 4, provides an upper bound on the Frobenius norm of the error between the estimate $\hat{\mathbf{X}}$ obtained from solving (Equation 1.26) and the ground truth matrix $\mathbf{X}_0$ that holds simultaneously for all matrices $\mathbf{X} \in \kappa(\alpha, R)$. Enroute to proving our guarantee, we indeed show that $\sum_{l,i}\langle \mathbf{A}_{l,i}, \mathbf{M}\rangle^2$ is well concentrated around its expected value of $\|\mathbf{M}\|_{\text{F}}^2$ for all $\mathbf{M} \in \kappa(\alpha, R)$. Hence, the measurements result in an embedding of the set $\kappa(\alpha, R)$.

Scalable algorithm for mixed-norm based optimization

The mixed-norm of any matrix $\mathbf{X}$ can be computed in polynomial time as

$$\begin{aligned} \underset{\mathbf{W}}{\text{minimize}} \quad & R \\ \text{subject to} \quad & \operatorname{trace}(\mathbf{W}_{11}) \leqslant R,\ \left\|\operatorname{diag}(\mathbf{W}_{22})\right\|_{\infty} \leqslant R. \end{aligned} \tag{1.27}$$

where $\operatorname{diag}(\cdot)$ denotes the vector of the diagonal elements of a matrix. Any optimization routine of the form

$$\begin{aligned} \underset{\mathbf{X}}{\text{minimize}} \quad & f(\mathbf{X}) \\ \text{subject to} \quad & \mathbf{X} \in \kappa(\alpha, R) \end{aligned} \tag{1.28}$$

can be reformulated as

$$\begin{aligned} \underset{\mathbf{W}_{11}, \mathbf{W}_{22}, \mathbf{X}}{\text{minimize}} \quad & f(\mathbf{X}) \\ \text{subject to} \quad & \operatorname{trace}(\mathbf{W}_{11}) \leqslant R,\ \left\|\operatorname{diag}(\mathbf{W}_{22})\right\|_{\infty} \leqslant R, \\ \left\|\mathbf{W}_{12}^{\top}\right\|_{2,\infty} \leqslant \alpha, \quad & \hat{\mathbf{W}} \succeq \mathbf{0} \end{aligned} \tag{1.29}$$

where $\hat{\mathbf{W}} = \left[\begin{smallmatrix} \mathbf{W}_{11} & \mathbf{X} \\ \mathbf{X}^{\top} & \mathbf{W}_{22} \end{smallmatrix}\right]$. The program in (Equation 1.29) is now a constrained convex optimization problem over the set of positive semidefinite matrices when $f(\cdot)$ is convex.

Experimental comparison of mixed nuclear norm based methods

We performed simulations to compare our proposed mixed-norm based optimization in (Equation 4.26) with the standard nuclear norm based estimator, where a LASSO type formulation minimizes the least square loss along with the nuclear norm of the matrix. We performed simulations on matrices of 1000×1000, at different noise levels and measured the relative Frobenius norm error between the estimate and the ground truth. As a pre-processing step, we normalize each column of the ground truth matrix to have the same norm. We perform Monte-Carlo simulations with randomly generated data, with 5 trials for each combination of noise level and sample complexity L. The results are summarized in Table 1.1.

Table 1.1: *Each box contains two entries: the left is the relative Frobenius norm between the estimate obtained using the mixed-norm optimization and the ground truth, the right is that for estimate obtained using nuclear norm minimization. The size of the test matrices was* 1000×1000 *and their rank was* 5*. Each column was pre-processed to have the same* ℓ_2*-norm. The error values shows are average values obtained over 5 trials.*

SNR $\backslash L$	50	150	200
0 dB	0.5180; 0.7518	0.2812; 0.6866	0.2527; 0.7157
10dB	0.1665; 0.2705	0.0876; 0.1978	0.0779; 0.1997
20dB	0.0532; 0.0746	0.0284; 0.0519	0.0248; 0.0482

1.6 Subspace based blind deconvolution

The column-wise sketching model developed in section 1.5 can also be used in a very specialized application commonly found in signal processing: blind deconvolution. Let $\mathbf{w} \in \mathbb{R}^L$ and $\mathbf{x} \in \mathbb{R}^L$ be two signals lying in known lower dimensional subspaces $\mathbf{D}$ and $\mathbf{C}$ respectively:

$$\mathbf{w} = \mathbf{D}\mathbf{h} \tag{1.30}$$

$$\mathbf{x} = \mathbf{C}\mathbf{m} \tag{1.31}$$

where $\mathbf{w}, \mathbf{x} \in \mathbb{R}^L$, $\mathbf{D} \in \mathbb{R}^{L\times K}$, $\mathbf{C} \in \mathbb{R}^{L\times N}$, $\mathbf{h} \in \mathbb{R}^K$ and $\mathbf{m} \in \mathbb{R}^N$. $\mathbf{D}$ and $\mathbf{C}$ are known a priori. The standard blind deconvolution observation model is given as:

$$\mathbf{y} = \mathbf{w} * \mathbf{x} + \nu, \tag{1.32}$$

where ν is additive noise, $\mathbf{y}$ are the measurements and $*$ denotes circular convolution defined as

$$(\mathbf{w} * \mathbf{x})[i] = \sum_{l=1}^{L} \mathbf{w}[l]\mathbf{x}[i-l+1].$$

where $i = 1, \cdots, L$ and the index $i-l+1$ is assumed to be modulo $\{1, \cdots, L\}$. Given $\mathbf{y}$, we are interested in recovering the signals $\mathbf{h}$ and $\mathbf{m}$. After applying the Fourier transform on both sides and after further simplification, it can be shown that the problem of recovering $\mathbf{h}$

and $\mathbf{m}$ can be restated as a rank-1 matrix recovery problem. Let $\mathbf{Z}_0 = \mathbf{m}\mathbf{h}^*$ and $\widehat{\mathbf{D}} = \mathbf{F}\mathbf{D}$, where $\mathbf{F}$ is the unitary discrete Fourier transform matrix. Then, the blind deconvolution problem can be recast as recovering $\mathbf{Z}_0$ from measurements given as

$$\mathbf{y}[\ell] = \langle \widehat{\mathbf{c}}_l \mathbf{e}_l^\top, \mathbf{Z}_0 \widehat{\mathbf{D}} \rangle + \nu \tag{1.33}$$

where $\widehat{\mathbf{c}}_l^\top$ is the lth row of the matrix $\mathbf{FC}$.

We need a few more assumptions on the subspaces $\mathbf{D}$ and $\mathbf{C}$ before we can proceed. In particular, we assume that the subspace $\mathbf{D}$ is time-limited to K: the last $L - K$ rows od $\mathbf{D}$ are zeros. Further, we assume $\mathbf{C}$ is *generic*. In other words, we assume $\mathbf{C}$ is drawn randomly from r-dimensional subspaces in $\mathbb{R}^L$. So we let the entries of the matrix $\mathbf{C}$ be drawn from the standard Gaussian distribution.

Under these assumptions, (Equation 1.33) is equivalent to the decentralized sketching model developed in (Equation 1.20), except for the additional subspace constraint on the right factor of the ground truth matrix. Similar to section 1.5, we develop a new convex relaxation for the set of low-rank matrices to account for the very specialized observation model in (Equation 1.33). Using such a relaxation, we provide sample optimal and uniform guarantees on the recovery error of our proposed estimator.

Both the column-wise sketching of low-rank matrices studied in section 1.5 and the blind deconvolution model in (Equation 1.33) are specific instances of the more general low-rank matrix recovery problem. Although standard methods to perform low-rank recovery use the nuclear norm as a convex proxy for the rank of a matrix, we note that this not the only choice. In many cases, the observation model can dictate the design of the recovery algorithm. Using the models in (Equation 1.20) and (Equation 1.33) as specific examples, we develop a general framework by which knowledge of the observation model can be used to tailor specific recovery algorithms for the set of low-rank matrices.

1.7 Organization of the thesis

This thesis studies and usage of subspace models in large scale signal processing and machine learning algorithms. It explores techniques to incorporate subspace priors into data acquisition systems and algorithms and aims to reduce the time and space complexity of the systems/algorithms to being proportional to the subspace dimension rather than the ambient dimension of the data.

In chapter 2, we consider two common problem in the field of array signal processing, namely active array imaging and source localization. For active array imaging, we first identify the underlying subspace model and then use structured random projections to reduce the number of spatial measurements required to image range-limited target scenes. With regards to source localization, we identify that existing methods do not address the broadband nature of real-world signals. By using a subspace model to describe band-limited signals, we develop algorithms to perform broadband source localization.

In chapter 3, we identify that existing matrix sketching techniques to generate subspace embeddings require access to all or arbitrary subsets of data to perform dimensionality reduction. We consider a block diagonal model on randomized sketching matrices and show theroetically and experimentally that they can be as efficient as standard sketching matrices, while also being amenable to applications where data is only available in parts.

In chapter 4, we consider extend the application of randomized sketching to applications where the underlying subspace is unknown a priori. We consider linear compression and recovery of a series of data points from a low dimensional subspace. By designing a convex relaxation of the set of low-rank matrices that is tailored for our particular observation model, we obtain theoretical guarantees and experimental performance that is better than standard nuclear-norm based low-rank estimators.

In chapter 5, we extend the framework of tailoring low-rank estimators that depend on the observation model developed in chapter 4 to the problem of blind deconvolution. By

posing the it as a low-rank matrix recovery problem, we develop a recovery algorithm that caters to the specifics of the blind deconvolution observation model. We provide uniform theoretical guarantees for our proposed estimator, thus improving upon existing methods that are able to provide only instance based guarantees.

CHAPTER 2
SKETCHED REGRESSION AND SUBSPACE LEARNING IN ARRAY SIGNAL PROCESSING

2.1 Active imaging of range-limited targets

2.1.1 Introduction

In active array imaging, a transmitter emits an excitation signal and then forms an image using the reflections collected at an array of sensors. This technique has been employed in a multitude of fields ranging from medicine to security and surveillance, among many others. Array imaging offers an observation window outside of the visible spectrum, which can prove crucial in applications where visible light cannot penetrate. However, a high cost barrier has prevented widespread adoption of array imaging in commercial products [7]. Nevertheless, applications such as autonomous vehicles, depth sensing, gesture recognition [8], and others have caused an increase of interest in commercializing active imaging modalities such as LiDAR and RADAR, leading to efforts aimed at reducing the cost and increasing the efficiency of array imaging systems. In such endeavors, developing a good signal model is key in designing efficient reconstruction algorithms. For example, sparsity based signal models naturally yield themselves to algorithms from the *compressed sensing* literature.

We are interested in a different model on the target scene where we assume that the targets are "range-limited": targets that have limited range or depth along each angle from the antenna array. When such targets are subjected to active imaging using illumination with a finite bandwidth, the reflected signals captured by the antenna array are nothing but the band-limited Fourier domain measurements of the range-limited target. Such a concentration in both the spectral and spatial domains imposes low dimensionality in the

received signals, leading to a limited number of degrees of freedom. This naturally leads to the question of whether spatially undersampled arrays can yield reconstructions as good as those obtained with full arrays. While this is reminiscent of *compressed sensing*, an important distinction between our work and the compressed sensing paradigm is that we do not require the scene to be sparse, but only to be range-limited.

Consider an antenna array system with M elements, exciting a scene with a narrowband signal of wavelength λ. Mathematically, the measurement process can be expressed as a linear model (we develop this model in more detail in subsection 2.1.2):

$$\mathbf{y}_\lambda = \mathbf{A}_\lambda \mathbf{x}_0 \tag{2.1}$$

where $\mathbf{x}_0 \in \mathbb{R}^N$ denotes the sampled target reflectivity and $\mathbf{A}_\lambda$ is the linear operator ($M \times N$ matrix) mapping the scene to the measurements and $\mathbf{y}_\lambda \in \mathbb{C}^M$ denotes the set of array outputs. Similar set of measurements collected across a discrete set of excitation wavelengths $\lambda_1 \ \cdots \lambda_K$ can be denoted as $\mathbf{y} = [\mathbf{y}_{\lambda_1}^\top \ \cdots \ \mathbf{y}_{\lambda_K}^\top]^\top$. While these measurements can be obtained by sequentially exciting the scene with different wavelengths, a standard approach is to use a single broadband excitation signal, collect the output of the M array elements and then obtain $\mathbf{y}$ by computing a temporal Fourier transform of the output of each element.

In order to take advantage of the low-dimensionality of the array outputs, we rely on the standard practice of obtaining linear combinations of array outputs as measurements in lieu of obtaining direct array element read-outs. This can be modeled as:

$$\mathbf{z}_\lambda = \psi_\lambda \mathbf{y}_\lambda = \psi_\lambda \mathbf{A}_\lambda \mathbf{x}_0 \tag{2.2}$$

where ψ_λ is typically an invertible $M \times M$ matrix. For example, a standard array processing method called *beamforming* collects linear combinations of the array elements' outputs to induce spatial directivity. In traditional beamforming the weights are chosen to induce spatial selectivity where each of the M measurements collects reflections from distinct spatial

sectors. Beamforming can be considered as a special case of *aperture coding*: collecting linear combinations of the array outputs across the aperture. Henceforth, we refer to any linear combination of array outputs as a spatial measurement or beam.

Our motivation in this work is based on the observation that sampling and storing all of the array elements' outputs, or obtaining all of the M independent linear combinations required for traditional beamforming can be challenging and is in fact wasteful when the target scene is range-limited. We show that by taking fewer *generic* linear combinations (or *aperture codes*) or even by spatial subsampling, one can obtain reconstructions of the same quality as that of using full measurements. Our measurement model can be mathematically described as obtaining the following set of measurements at each of the K wavelengths: $\mathbf{y}_\lambda = \phi \mathbf{A}_\lambda \mathbf{x}_0$ where ϕ is a compressive $L \times M$ matrix ($L < M$).

Since we do not impose any additional structure on the target scene such as sparsity, when the model $\mathbf{A}_\lambda$ is designed to incorporate the knowledge of the range-limit, the target scene can be reconstructed from the full set of measurements $\mathbf{y}$ using the ordinary least squares estimator:

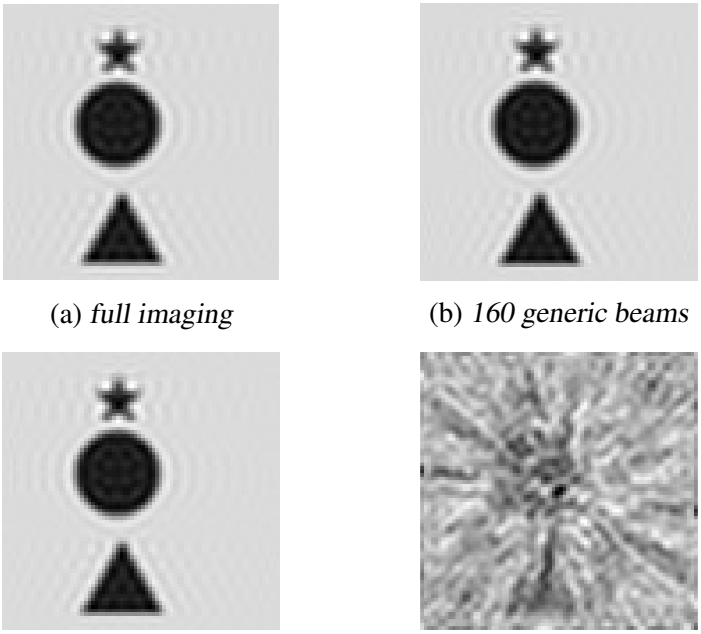
$$\min_x \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2$$

where $\mathbf{A} = [\mathbf{A}_{\lambda_1}^* \ \cdots \mathbf{A}_{\lambda_K}^*]^*$ and $*$ denotes conjugate transpose of a matrix. Our goal is to show that a similar reconstruction performance can be obtained using a subsampled set of measurements. We take inspiration sketching techniques and provide theoretical justification for the proposed signal acquisition method. We show that image reconstruction with a few generic aperture codes is equivalent to a *sketched least squares* problem of the form

$$\min_{\mathbf{x}} \|\Phi\mathbf{y} - \Phi\mathbf{A}\mathbf{x}\|^2 \tag{2.3}$$

where Φ is a highly structured compressive matrix and then establish equivalence between the solutions of the ordinary least squares method and the sketched least squares method.

As an example, consider an imaging setup where the array is two dimensional and has



(a) *full imaging* (b) *160 generic beams*

(c) *80 generic beams* (d) *50 generic beams*

Figure 2.1: *Aperture coded imaging for images at a constant range. (a) represents the conventional method, which uses 1100 beams. (b),(c) and (d) show the reconstruction results using only* 160, 80 *and* 50 *generic linear combinations of the antenna array outputs. Aperture coding can be used to reduce the number of spatial measurements to as low as 80. Reducing the number of measurements further to 50 leads to poor reconstruction.*

40×40 sensors with sensors placed $\lambda/2 = 3.75$cm apart and a target scene having a span of $[-45^\circ, 45^\circ]$ in both elevation and azimuthal angles. Let the scene have *delta thickness*: only one reflector per each angle, present at a constant known depth. Standard ways of imaging such a scene would require around 1100 beams at wavelength $\lambda = 7.5$cm. By introducing bandwidth in the excitation signal, we show that the scene can be imaged with as few as 80 spatial linear combinations. This is illustrated in Figure 2.1. For target scenes with higher range limits with multiple reflectors per angle, standard imaging methods need bandwidth for imaging [9]. We however show that this bandwidth can be used to obtain similar gains in the number of spatial linear combinations and provide theoretical justification for the gains that aperture coding can provide [10].

2.1.2 Propagation model and Fourier domain samples

We develop the broadband array imaging model for a one-dimensional (1D) antenna array, which can be easily extended to 2D arrays. Consider a uniform linear array of aperture length D placed on the x-axis, spanning $[-D/2, D/2]$. For now, we assume that the aperture is continuous. The time delay of arrival is measured with respect to the origin. We assume that the scene to be imaged lies in the X-Y plane and is in the far-field region of the array. Imaging the scene is equivalent to reconstructing its reflectivity map, which is a function of the distance of the target from the array center and the angle from the broadside, to be denoted as $p(r, \theta)$. The system consists of a single transmitting element which is co-located with the receiver at the array center. Let t be the continuous time index. For an excitation signal $s(t)$, the signal received at the array location d is $s(t-r_0/c-d\sin\theta_0/c)$ for a unit-strength reflector at (r_0, θ_0). For a general reflectivity map, the narrowband response at this location for the excitation signal $s(t) = e^{j2\pi ct/\lambda}$ is

$$\widetilde{y}_{d,\lambda}(t) = \mathrm{e}^{\mathrm{j}2\pi ct/\lambda} \int_{-\pi/2}^{\pi/2} \int p(r,\theta) \mathrm{e}^{-\mathrm{j}2\pi(r+2d\sin\theta)/\lambda}\, dr\, d\theta. \tag{2.4}$$

By making the substitution $\tau = (\sin\theta)/2$, the complex amplitude of the signal received at location $d \in [-D/2, D/2]$ for excitation wavelength λ can be written as

$$y_{d,\lambda} = \int_{-\pi/2}^{\pi/2} \int p(r,\theta) \mathrm{e}^{-\mathrm{j}2\pi(r+2d\tau)/\lambda}\, dr\, d\theta = \widehat{x}_c\left(\omega_r,\, \omega_\tau\right). \tag{2.5}$$

where $\omega_r = \frac{1}{\lambda},\ \omega_\tau = \frac{2d}{\lambda}$, and $\widehat{x}_c$ denotes the Fourier transform of $x_c(r,\tau) = \frac{p(r,\sin^{-1}(2\tau))}{\sqrt{1-4\tau^2}}$. This shows that the antenna aperture measures the Fourier transform of the target scene (after a coordinate transformation). $y_{d,\lambda}$ can be obtained by computing the Fourier transform of the temporal signal $\widetilde{y}_{d,\lambda}(t)$ received at the array after sampling it at a suitable rate, or by measuring the complex amplitude of the received signal. We do implicitly assume

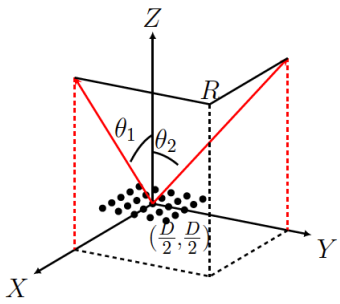
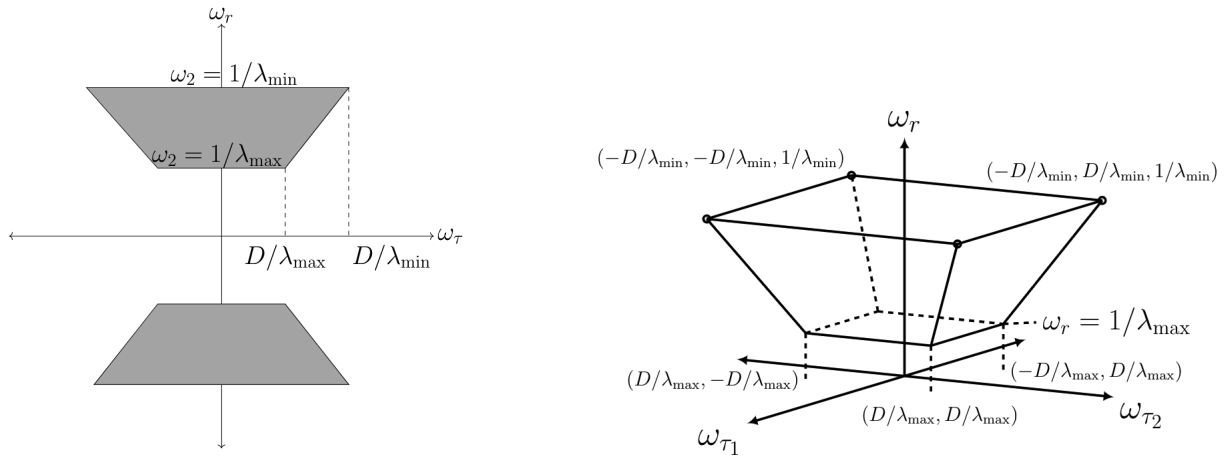


Figure 2.2: *The coordinate system used for the 2D array setup. R represents a point on the target image. The antenna array, illustrated using the set of dots, lies in the X-Y plane in the region $[\frac{-D}{2}, \frac{D}{2}] \times [\frac{-D}{2}, \frac{D}{2}]$.*

that we are sampling these time-domain signals fast enough that the Fourier transform of the received signal can be computed at the required frequencies. This implicit assumption is there in both the coded and full imaging scenarios. As a direct consequence of the finiteness of the aperture, we have that at any excitation wavelength λ, the accessible interval in the Fourier domain for ω_τ is limited to $[-D/\lambda, D/\lambda]$.

For an imaging system with a 2D array in the X-Y plane and a 3D scene, the extension of the setup is straightforward to derive. The coordinates in 3D can be denoted as (r, θ_1, θ_2), where r is the roundtrip distance to the array center, θ_1 is the angle with respect to the Y-Z plane, and θ_2 is the angle with respect to the X-Z plane, as shown in Figure 2.2. The scene reflectivity is denoted as $x_c(r, \tau_1, \tau_2)$ where $\tau_1 = (\sin\theta_1)/2$ and $\tau_2 = (\sin\theta_2)/2$. At excitation wavelength λ, the 2D array outputs $y_{d_1,d_2,\lambda} = \widehat{x}_c(1/\lambda,\ 2d_1/\lambda,\ 2d_2/\lambda)$ are samples of the Fourier transform of the scene sampled in the region bounded by $\left(\frac{\pm D_1}{\lambda}, \frac{\pm D_2}{\lambda}\right)$, where D_1 and D_2 are the dimensions of the 2D array. From now on, we use the 1D array to discuss our model for the sake of notational brevity. However, all our discussion and results hold for both cases and all simulations use 2D arrays.

We are mainly interested in a broadband excitation scenario. We assume that the excitation signal is a broadband pulse bandlimited to $[\lambda_{\min}, \lambda_{\max}]$. If the broadband signal used



(a) (b)

Figure 2.3: *(a) shows the region in the Fourier domain of a target scene acquired by using broadband excitation and a finite 1D aperture. (b) shows the same for a 2D array imaging a 3D scene.*

is $s_b(t)$, then the received signal at location d is

$$\widetilde{y}_d(t) = \int_{-\pi/2}^{\pi/2} \int p(r,\theta) s_b(t - r/c - d\sin\theta)\, dr\, d\theta. \tag{2.6}$$

The complex amplitudes at different wavelengths, $y_{d,\lambda}$ can then be obtained as the Fourier coefficients of $\widehat{y}_d$ after taking its temporal Fourier transform. Since the lateral frequency support increases with decrease in the excitation wavelength, the accessible Fourier domain has a trapezoidal shape as illustrated in Figure 2.3a. Similarly, for a 2D antenna, the region in the Fourier domain measured is shown in Figure 2.3b. Broadband imaging is hence the task of collecting Fourier measurements in a bandlimited region and inferring the target reflectivity profile using these measurements.

Array measurement model Our discussion above describes what can be observed through a finite but continuous aperture. In practice, we must measure this signal using a discrete array of sensors. This limits the Fourier domain measurements considered in the previous section to only a discrete set of samples. Let the 1D antenna considered in the previous section be an array of M discrete antenna elements placed uniformly at coordinates

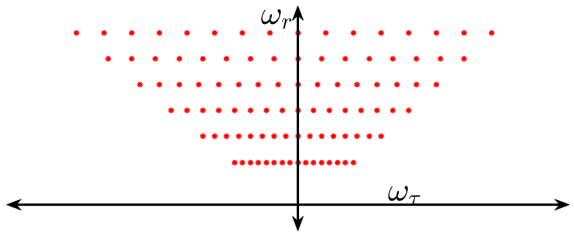


Figure 2.4: *With a discrete array, we can only collect discrete samples in the Fourier domain of the target scene. These samples are placed uniformly along the lateral frequency axis and and at regular intervals on the vertical frequency axis. Each sample denotes the output at a single array element, at a single excitation frequency. The limits for ω_τ at each ω_r depends on the aperture size and spans the region $[-\omega_r D,\ \omega_r D]$. A discrete set of temporal wavelengths is assumed instead of a continuous support.*

$d_{\frac{-M}{2}}, \cdots, d_{\frac{M}{2}} \in [-D/2, D/2]$. We also consider a discrete set of K excitation wavelengths $\{\lambda_1 = \lambda_{\min}, \lambda_2, \cdots, \lambda_K = \lambda_{\max}\}$. When the complex amplitudes are measured at these wavelengths, the Fourier samples obtained are located on a pseudopolar grid, as shown in Figure 2.4. Measurements at these wavelengths can be obtained by using a single broadband excitation, as explained earlier. The set of all measurements $\{y_{m,\lambda}\}$ where $m = 1, \cdots, M$, $\lambda \in \{\lambda_1, \cdots, \lambda_K\}$ can be denoted by a vector $\mathbf{y} \in \mathbb{C}^{MK}$. The set of measurements at each wavelength λ is denoted by $\mathbf{y}_\lambda \in \mathbb{C}^M$.

Traditional imaging involves collecting the samples shown in Figure 2.4 to reconstruct the target reflectivity profile. This can be achieved in many ways. One way is to directly read out the output of each antenna element. This amounts to measuring $\mathbf{y}$ directly. Physically, this can be realized by using a broadband pulse as before and then taking a Fourier transform. Alternatively, one could cycle through a set of narrowband excitation signals (stepped frequency excitation) and collect array measurements at each wavelength. In either case, this method would require M array element readouts.

Another standard method for acquiring the measurements is to collect linear combinations of the array outputs. In this method, the output of each array element is weighted and then added to the output of other elements. In particular, the procedure known as *beamforming* obtains specific linear combinations of the array outputs that induce spatial directivity. In narrowband beamforming, the weights are chosen such that the time delays

for signals coming from a particular spatial direction are compensated for, hence "focusing" the array in that physical direction, as outlined in [9]. Instead of acquiring M direct read-outs, M linear combinations are acquired. Beamfoming is also use in wideband array signal processing and examples of such practice can be found in [11, 12, 13, 14, 15].

Acquiring linear combinations at each excitation wavelength separately is equivalent to acquiring linear combinations of samples along each row of points in Figure 2.4. A different set of linear combinations may be used for each excitation wavelength. Mathematically, if the excitation wavelength is λ, the measurements made in time domain are

$$\widetilde{z}_{i,\lambda}(t) = \sum_{m=1}^{M} \phi_\lambda(i,m)\widehat{y}_{m,\lambda}(t) \tag{2.7}$$

for $i = 1, \cdots, M$. The vector of complex amplitudes $z_\lambda \in \mathbb{C}^M$ is then given by

$$\mathbf{z}_\lambda = \phi_\lambda \mathbf{y}_\lambda \tag{2.8}$$

where ϕ_λ is in general an $M \times M$ well-conditioned matrix whose $(i,m)^{\text{th}}$ element is $\phi_\lambda(i,m)$. If a single broadband pulse is used for excitation, then the set of weights for the linear combinations at different wavelengths are constrained to be the same and the vector of complex amplitudes at different wavelengths are given by:

$$\mathbf{z}_\lambda = \phi \mathbf{y}_\lambda \tag{2.9}$$

where ϕ is now common across all wavelengths. In general, we refer to acquiring linear combinations of the array elements as *coded aperture acquisition*. A variant of aperture coding was considered in [16] where the authors propose using different subarrays at different times and then using interpolation techniques to acquire all the samples shown in Figure 2.4. Subsampling the array is equivalent to using binary codes on the aperture. However, their signal model is different from ours and they do not consider imaging with

fewer than M measurements.

2.1.3 Degrees of freedom of range-limited targets

A finite bandwidth and aperture allow us to observe only a part of the Fourier transform of the image. A finite range restricts the number of degrees of freedom of this observed region of the Fourier transform. We intend to take advantage of this to achieve a more efficient sampling of the bandlimited spectrum of the image and achieve hence faster imaging. We now demonstrate the effect of range-limitedness using a target scene at a constant known distance, with *delta* thickness: where the scene has only one reflector per angle, with each reflector present at a constant known distance from the array. The underlying effect on the Fourier domain samples extends to scenes with a more general range limit.

Consider a scene with delta thickness at a constant range R_0 from the antenna array. Considering just the amplitude as before, we have

$$y_m(\lambda) = \int_{-1/2}^{1/2} p(R_0, \tau) \mathrm{e}^{-\mathrm{j}2\pi 2 d_m \tau/\lambda} \, \mathrm{d}\tau. \tag{2.10}$$

Now define

$$g(\omega_\tau) = \int_{-1/2}^{1/2} p(R_0, \tau) \mathrm{e}^{-\mathrm{j}2\pi \omega_\tau \tau} \, \mathrm{d}\tau. \tag{2.11}$$

The array outputs $y_m(\lambda)$ are then just **samples of the same function** $g(\omega_\tau)$ sampled uniformly in $[-D/\lambda, D/\lambda]$ (modulo known scaling factors). This is illustrated in Figure 2.5. In the 2D array case, the "slices" of the trapezoid corresponding to different excitation wavelengths sample a common function. As the range limit increases but remains finite, the functions sampled at different wavelengths start to differ, but still have limited degrees of freedom.

From the above discussion, it is clear that for scenes with delta thickness, collecting the full set of samples at the lowest excitation wavelength provides all available informa-

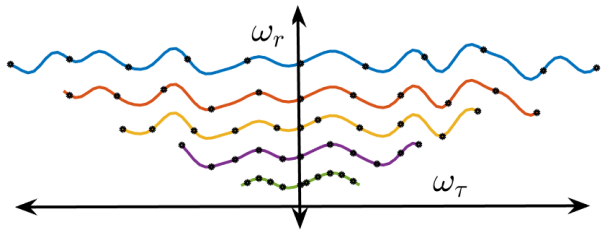


Figure 2.5: *For an image at a constant range, the array measurements at different wavelengths are samples of a common function*

tion. Collecting further samples at higher wavelengths offers no advantage. At the smallest wavelength (largest ω_τ), even though the samples are less dense, array element spacing of $\lambda_{\min}/2$ guarantees that only the samples at $\lambda_{\min}$ are enough for reconstruction. This redundancy can be used to collect fewer spatial samples but with broadband excitation. For scenes with larger range-limits, the outputs at higher wavelengths **do** offer extra information since they no longer sample the same function. However, the ensemble of array outputs across all the wavelengths lie in a low-dimensional subspace. This redundancy again implies that collecting the full set of broadband measurements for scenes with higher but finite range limits results in a number of measurements greater than the number of degrees of freedom. It is thus natural to expect that the target scene can be reconstructed with a number of measurements $l \ll M$, owing to the limited number of degrees of freedom.

For computational purposes, we can discretize the target scene and the array imaging operator. For a scene with delta thickness at a depth R_0, let $\mathbf{x}_{R_0} \in \mathbb{R}^N$ denote the target scene $\mathbf{x}_c$ sampled uniformly with $N \geqslant M$. Similarly, the integral mapping the target scene to the array measurements at wavelength λ can be discretized as a matrix $\mathbf{A}_{R_0,\lambda} \in \mathbb{C}^{M\times N}$ operating on $\mathbf{x}_{R_0}$:

$$\mathbf{A}_{R_0,\lambda}(m,n) = \exp^{-j2\pi R_0/\lambda} \exp^{-j2\pi 2d_m(-0.5+n/N)/\lambda}. \tag{2.12}$$

With this notation in place, the array measurements can be expressed as

$$\mathbf{y}_{R_0,\lambda} = \mathbf{A}_{R_0,\lambda}\mathbf{x}_{R_0}. \tag{2.13}$$

Similarly, a more general scene with a range limit R can be discretized as a vector of reflectivities as $\mathbf{x}_R \in \mathbb{R}^{N\times D}$ where N represents the number of discrete samples along the τ axis and D the number of samples along the r axis. Let the scene lie between the range limits $R_{\min}$ and $R_{\max}$. Define $d_r = \lfloor n/N + 1 \rfloor,\ n_\tau = \mod(n, N)$. Then the discretized array operator can be expressed as a matrix $\mathbf{A}_R \in \mathbb{C}^{M\times ND}$ as

$$\mathbf{A}_{R,\lambda}(m,n) = \exp^{-j2\pi(2d_m(-0.5+n_\tau/N)+(R_{\min}+d_rR/D))/\lambda} \tag{2.14}$$

and the array outputs can be expressed as

$$\mathbf{y}_{R,\lambda} = \mathbf{A}_{R,\lambda}\mathbf{x}_R. \tag{2.15}$$

Since our signal model considers only range-limited scenes, we drop the subscripts R and R_0 for further discussion. We will denote antenna array measurements at excitation wavelength λ as $y_\lambda \in \mathbb{C}^M$, the discretized target scene as vector $\mathbf{x}_0 \in \mathbb{R}^{ND}$ (or $\mathbb{R}^N$ for scenes with delta thickness) and the array imaging operator as $\mathbf{A}_\lambda \in \mathbb{C}^{M\times ND}$ (or $\mathbb{C}^{M\times N}$ for scenes with delta thickness). $\mathbf{A}_\lambda$ will incorporate knowledge of the target range profile, which is assumed to be known a priori. This helps us focus on the advantage of range-limitedness in itself. We later show how unknown range profiles can be handled algorithmically.

When the measurements at all the K wavelengths are considered, we obtain the linear

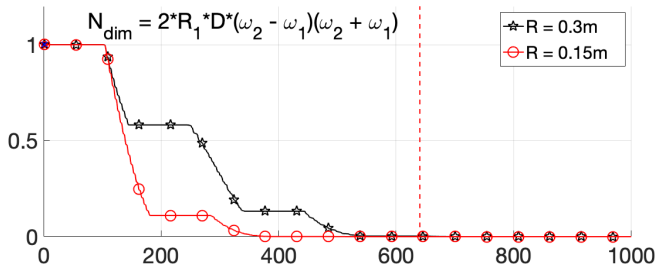


Figure 2.6: *Eigenvalue decomposition of operator* $\mathbf{A}$ *in the 1D array case for a range limit of* R*, aperture size* D *and excitation frequency range of* $[\omega_1/c,\ \omega_2/c]$.

system

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_{\lambda_1} \\ \mathbf{y}_{\lambda_2} \\ \vdots \\ \mathbf{y}_{\lambda_K} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{\lambda_1} \\ \mathbf{A}_{\lambda_2} \\ \vdots \\ \mathbf{A}_{\lambda_K} \end{bmatrix} \mathbf{x}_0 = \mathbf{A}\mathbf{x}_0. \tag{2.16}$$

The collection of array outputs at all the wavelengths lie in the column space of $\mathbf{A}$. The effective dimension of this subspace determines the number of degrees of freedom in $\mathbf{y}$. When the measurements are acquired with a coded aperture, we can model the measurements as

$$\mathbf{z} = \Phi\mathbf{y} = \Phi\mathbf{A}\mathbf{x}_0. \tag{2.17}$$

where

$$\Phi = \begin{bmatrix} \phi & 0 & \cdots & 0 \\ 0 & \phi & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi \end{bmatrix} \tag{2.18}$$

is a block diagonal matrix with repeating diagonal blocks. We refer to such a matrix as a repeated block diagonal (RBD) matrix. Our main goal is to show that the target scene can be constructed using highly underdetermined matrices Φ with no loss in resolution compared to the reconstruction obtained using the full set of measurements $\mathbf{y}$.

The reduction in the number of degrees of freedom in the case of range-limited targets described above is a consequence of the simultaneous concentration of energy in both spatial and spectral domains. Signals with such a property can be well approximated by a number of basis functions that is proportional to the product of the area/volume of the spatial and spectral supports. This has been studied in a set of seminal papers by Slepian, Landau and Pollock [17, 18, 19, 20, 21, 22] and by Simons et.al., in [23, 24]. In essence, the range space of space-limiting, band-limiting operators is approximately finite dimensional. This approximate dimension is the number of effective degrees of freedom of signals well concentrated in spatial and spectral domains. An efficient basis for the representation of such signals is the *prolate* basis [17]. The Smaller the spatial and spectral supports, the smaller is the number of degrees of freedom. See [25, 26, 27] for a quantitative non-asymptotic characterization of these properties in the discrete case.

Since the array imaging operator obtains band-limited measurements of range-limited target scenes, its range space has a low dimensional structure. To illustrate this, we present the singular values of $\mathbf{A}$ for target images with a finite range limit in Figure 2.6, where we consider two different range limits of 0.3m and 0.1m. The product of spatial and spectral supports in this case is approximately $2RD(\omega_2 - \omega_1)(\omega_2 + \omega_1)$. For these plots, an array with 213 elements was used and samples were collected at 25 wavelengths placed uniformly between 2GHz and 4GHz. The total number of samples collected is hence 5325, but these samples lie in a subspace of dimension approximately only 640 or 320, for the two range limits considered. For the same antenna array, we show the singular values of the operators associated with target scenes that have delta thickness in Figure 2.7. As expected, such scenes lie subspaces of even smaller dimensions.

A special case of range-limitedness is when the target has delta thickness. Figure 2.7 shows the singular value decomposition of the array operator for two examples of such scenes: one where all the reflectors are at a constant known distance from the array center; the other where each reflector is at a known but different distance from the array center. We

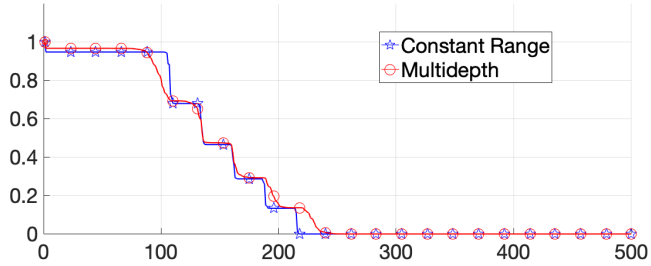


Figure 2.7: *Spectra of the operators for scenes with delta thickness. 'Constant range' describes a scene where the reflectors at all angles are at a constant depth from the array center and 'Multi-depth' describes a scene in which each reflector is at a different depth from the array. In each case, the infinite dimensional continuous domain image can be efficiently represented using a subspace of relatively small dimension*

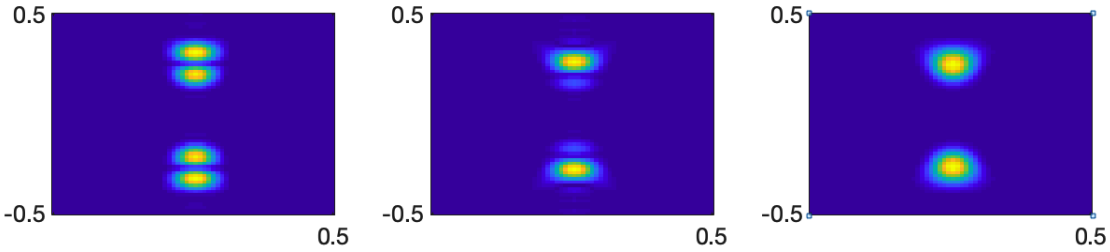


Figure 2.8: *Spectra of the top eigenvectors of $A^T A$. The eigenvectors have a frequency support that is mostly concentrated in the trapezoidal area and a spatial support limited to R as defined by the model.*

also show the Fourier transforms of the eigenvectors of $\mathbf{A}^T\mathbf{A}$ associated with the most and least significant eigenvalues in Figure 2.8.

The low dimensional structure in the array measurements forms the basis of our proposed imaging method. We show that this structure enables imaging with very few array measurements. In the following sections, we set up the reconstruction problem and provide theoretical guarantees for image reconstruction.

The limited number of degrees of freedom of broadband measurements leads us to ask the following question: what is the number of broadband measurements required to image range-limited target scenes? In order to answer that question, we first set up our reconstruction method. We do not assume any structure on the target scene such as sparsity, low total variation norm, or other structure apart from it being range-limited. Hence, with the

measurement model used in (Equation 2.17), we solve the ordinary least squares problem

$$\arg\min_{x} \|\Phi\mathbf{y} - \Phi\mathbf{A}\mathbf{x}\|_2^2. \tag{2.19}$$

Our standard of comparison is the ordinary least squares estimate obtained using the full data:

$$\arg\min_{x} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2. \tag{2.20}$$

The theoretical question we answer is: If each diagonal block ϕ in (Equation 2.17) is of size $l \times M$, what is the sample complexity l that can achieve the same reconstruction result as (Equation 2.20) using (Equation 2.19)?

2.1.4 Linear algebraic interpretation of range-limitedness

The least squares program in (Equation 2.20) searches for an image that best explains the measurements obtained, and lies in the row space of $\mathbf{A}$. The row space of $\mathbf{A}$ is not only approximately low dimensional for range-limited images but has a second tier of structure that makes aperture coding a very efficient way of obtaining fewer measurements. In particular, the relationship between the subspaces associated with the linear operators at various excitation wavelengths for range-limited images determines the number of aperture codes required. Consider array imaging at a single wavelength $\mathbf{A}_{\lambda_i}$. The least squares estimate in this case looks for a target in the row space of $\mathbf{A}_{\lambda_i}$ that best explains the measurements. We denote this subspace as $\mathcal{S}_i$. Similarly, imaging using wideband excitation results in an estimate that lies in the union of subspaces $\mathcal{S}_1, \mathcal{S}_2, \cdots, \mathcal{S}_k$. The sample complexity of aperture coding depends highly on the relationship between these subspaces.

Let $\mathbf{A}$ now denote a general $kM \times N$ matrix of a finite rank r such that

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1^T & \mathbf{A}_2^T & \cdots & \mathbf{A}_k^T \end{bmatrix}^T$$

where each $\mathbf{A}_i$ is $M \times N$. Let the rank of $\mathbf{A}_i$ be r_i. Without loss of generality, we can assume that $r_i \geqslant r_j$ for $i > j$, since the ordering of the row groups does not matter. We can then obtain the following factorization:

$$\begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \cdot \\ \mathbf{A}_k \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{11} & 0 & \cdots & 0 \\ \mathbf{C}_{21} & \mathbf{C}_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}_{k1} & \mathbf{C}_{k2} & \cdots & \mathbf{C}_{kk} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \\ \vdots \\ \mathbf{V}_k^T \end{bmatrix} = \mathbf{C}\mathbf{V}^T \tag{2.21}$$

where $\mathbf{C}_{ij} \in \mathbb{R}^{Mm \times d_j}$, each $\mathbf{C}_{ii}$ is full column rank when $d_i \neq 0$, and $\mathbf{V}$ is an $r \times N$ orthonormal matrix. The factorization is such that the row space of $\mathbf{A}_1$ is the span of the orthobasis $\mathbf{V}_1$, the row space of $\mathbf{A}_2$ is included in the span of $\mathbf{V}_1$ and $\mathbf{V}_2$. In general, $[\mathbf{V}_1\ \mathbf{V}_2\ \cdots \mathbf{V}_i]$ includes an orthobasis for the row space of $\mathbf{A}_i$. This factorization is equivalent to a block QR factorization of $\mathbf{A}^T$ and can be obtained for any general matrix A. We are particularly interested in the d_i's, as they capture the relationship between various subspaces. The diagonal blocks $\mathbf{C}_{ii}$'s represent the energy in the subspace orthogonal to the union of the row spaces of the previous blocks $\mathbf{A}_1, \cdots, \mathbf{A}_{i-1}$. Hence, smaller values of d_i indicate that the subspaces have a high degree of overlap.

A special case of high overlap is when the row spaces have a nested structure: $\mathrm{row}(\mathbf{A}_{i-1}) \subset \mathrm{row}(\mathbf{A}_i)\ \forall i = 2, \cdots, K$. In this case, the off-diagonal blocks $\mathbf{C}_{ij}, i < j$ and the orthogonal blocks $\mathbf{V}_1, \cdots, \mathbf{V}_{i-1}$ capture a significant part of $\mathrm{row}(\mathbf{A}_i)$. For low-rank systems, this naturally leads to smaller values of d_i. In contrast, when the row spaces are all almost orthogonal, the d_i's are all large and the off-diagonal blocks $\mathbf{C}_{ij}, i \neq j \approx 0$. We will later show that the broadband array imaging operator has the nested subspace structure for certain range-limited scenes.

Let us now relate the above factorization to the context of imaging. To begin, suppose that we use wavelengths up to λ_i. Then, d_{i+1} represents the rank of the update required to incorporate information from a new, lower wavelength λ_{i+1}. It represents the **innovation**

added by the measurements at the new wavelength.

As the range limit decreases, the overlap between the subspaces $\mathcal{S}_i$ increases, increasing the redundancy across wavelengths. This leads to the possibility of higher subsampling rates in the physical array domain. In the limiting case of only one reflector per angle, the subspaces have a nested structure. As we will observe in subsection 2.1.5, this plays a crucial role in determining the number of aperture codes required for successful imaging.

Our theoretical results formally state the effect of the relationship between the subspaces S_i on the number of aperture codes required for imaging. Theorem 3 provides a non-trivial estimate of the number of measurements needed for a given excitation bandwidth. Theorem 5 provides conditions under which a given set of K excitation wavelengths allow the number of spatial measurements to be reduced by a factor of K. Using these conditions and a given bandwidth of excitation, one can choose the set of excitation wavelengths and very few coded measurements to achieve imaging with no loss in resolution.

2.1.5 Random projections in nested subspaces

In the previous section, we set up the aperture coding problem as a sketched least squares problem that has a particular structure dictated by the physical problem of array imaging. In this section, we derive mathematical guarantees for such sketched systems and provide estimates of the required sample complexity. Let

$$\mathbf{x}_{\text{LS}} = \text{argmin}_{\mathbf{x}} \left\| \mathbf{y} - \mathbf{A}\mathbf{x} \right\|^2 \tag{2.22}$$

$$\mathbf{x}_{\text{SLS}} = \text{argmin}_{\mathbf{x}} \left\| \Phi\mathbf{y} - \Phi\mathbf{A}\mathbf{x} \right\|^2, \tag{2.23}$$

where $\mathbf{x}_{\text{LS}}$ and $\mathbf{x}_{\text{SLS}}$ are the solutions to the full and the sketched linear systems of equations erspectively. We start by reviewing the conditions that any general sketching operator has to satisfy in order for the solution to the sketched least squares problem x_{SLS} to be close to

the solution of the original ordinary least squares solution x_{LS}. In the noiseless case

$$\mathbf{x}_{\text{LS}} = \mathbf{A}^{\dagger}\mathbf{y} = \mathbf{A}^{\dagger}\mathbf{A}\mathbf{x}_0 = \mathbf{V}\mathbf{V}^{*}\mathbf{x}_0 \tag{2.24}$$

where $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^{*}$ is the SVD of the linear operator and $\mathbf{A}^{\dagger}$ denotes the pseudoinverse of $\mathbf{A}$. (Equation 2.24) shows that the least squares solution is a projection of the true solution onto the row space of $\mathbf{A}$. Hence, any sketching operator Φ should preserve the row space of $\mathbf{A}$. It has been well established in literature that a number of random projections of the rows of $\mathbf{A}$ greater than or equal to its rank capture the row space in case of exactly low-rank matrices [28]. When $\mathbf{A}$ of size $KM \times N$ and $\text{rank}(\mathbf{A}) = r \ll \min(M, N)$, if Φ is a $L \times KM$ dense standard normal random matrix with $L \geqslant r$, then $\text{row}(\mathbf{A}) \subset \text{row}(\Phi\mathbf{A})$. Since by construction we also have $\text{row}(\Phi\mathbf{A}) \subset \text{row}(\mathbf{A})$, we have

$$\mathbf{V}_{\Phi\mathbf{A}}\mathbf{V}_{\Phi\mathbf{A}}^{*} - \mathbf{V}\mathbf{V}^{*} = 0.$$

Let

$$\mathbf{x}_{\text{SLS}} = (\Phi\mathbf{A})^{\dagger}\phi\mathbf{y} = \mathbf{V}_{\Phi\mathbf{A}}\mathbf{V}_{\Phi\mathbf{A}}^{*}\mathbf{x}_0.$$

The least squares estimate from the sketched measurements $\phi\mathbf{y}$ is same as that from the full observation $\mathbf{y}$, since

$$\|\mathbf{x}_{\text{LS}} - \mathbf{x}_{\text{SLS}}\| \leqslant \|\mathbf{V}\mathbf{V}^{*} - \mathbf{V}_{\Phi\mathbf{A}}\mathbf{V}_{\Phi\mathbf{A}}^{*}\|\|\mathbf{x}_0\| = 0.$$

This idea forms the basis of using sketched measurements to solve a least squares problem. For any sketching matrix ϕ, a necessary and sufficient condition in the noiseless case is $\|(\mathbf{I} - \mathbf{P}_{(\phi\mathbf{A})^{*}})\mathbf{A}^{*}\| = 0$.

Our goal is to replicate the above result for an RBD sketching matrix. For such a matrix (shown in (Equation 2.18)), if l random projections are obtained per block, the equivalent result would be that a total number of measurements $Kl \geqslant r$ suffice to capture the row

space of the matrix $\mathbf{A}$. However, due to the highly structured nature of a block diagonal matrix, such a result does not hold uniformly for all matrices $\mathbf{A}$. We analyze the conditions on $\mathbf{A}$ under which such a result holds and show that array imaging matrices do obey these conditions, thus allowing for spatial subsampling.

RBD matrices obtain *localized* random projections: they take linear combinations of only a subset of the rows. In this section, we provide guarantees on the the error $\|(\mathbf{I} - \mathbf{P}_{(\Phi\mathbf{A})^*})\mathbf{A}^*\|$ when Φ is an RBD matrix. The focus will be on the sample complexity l required to drive this error to 0 with high probability.

It is immediately clear how to achieve this when we take $l \geqslant \max_i \text{rank}(\mathbf{A}_i)$. Let $\mathbf{A} = [\mathbf{A}_1^T \ \mathbf{A}_2^T \cdots \ \mathbf{A}_K^T]^T$ where each $\mathbf{A}_i$ is of size $M \times N$ and has rank r_i. Let $\mathbf{Y} = \Phi\mathbf{A}$. Hence

$$\mathbf{Y} = \begin{bmatrix} \phi\mathbf{A}_K \\ \phi\mathbf{A}_{K-1} \\ \vdots \\ \phi\mathbf{A}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_k \\ \mathbf{Y}_{k-1} \\ \vdots \\ \mathbf{Y}_1 \end{bmatrix}.$$

Since

$$\|(\mathbf{I} - \mathbf{P}_{\mathbf{Y}*})\mathbf{A}^*\| \leqslant \sum_{i=1}^{k} \|(\mathbf{I} - \mathbf{P}_{\mathbf{Y}_i^*})\mathbf{A}_i^*\|, \tag{2.25}$$

and $\|(\mathbf{I} - \mathbf{P}_{\mathbf{Y}_i^*})\mathbf{A}_i^*\| = 0$ for $l \geqslant r_i$, we obtain

$$\|(\mathbf{I} - \mathbf{P}_{\mathbf{Y}*})\mathbf{A}^*\| = 0 \tag{2.26}$$

for $l \geqslant \max_i r_i$.

Compared to using a dense random matrix, this can be worse by a factor of K. Intuitively, this straightforward application of results from [28] leads to capturing of the subspace spanned by each group of rows $\mathbf{A}_i$ individually, without considering the overlap between the subspaces. This is addressed in our first analytical result (Theorem 3), which provides a simple but non-trivial estimate of the number of random projections required.

We then improve this result in Theorem 4 and Theorem 5.

Theorem 3 *For a given matrix* $\mathbf{A}$ *of size* $KM \times N$*, let the* d_i *be defined as in* (Equation 2.21)*. Let* Φ *be a block diagonal matrix with repeated diagonal block* ϕ *of size* $l \times M$ *and whose entries are chosen i.i.d. from the standard normal distribution. Let* $\mathbf{Y} = \Phi\mathbf{A}$*. Define* $d_0 = \max_i d_i$*. For* $l \geqslant d_0$*,* $\|(\mathbf{I} - \mathbf{P}_{\mathbf{Y}*})\mathbf{A}^*\| = 0$ *with probability* 1*.*

Proof We have

$$\Phi\mathbf{A} = \Phi\mathbf{C}\mathbf{V}^T = \begin{bmatrix} \phi\mathbf{C}_{11} & 0 & \cdots & 0 \\ \phi\mathbf{C}_{21} & \phi\mathbf{C}_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \phi\mathbf{C}_{k1} & \phi\mathbf{C}_{k2} & \cdots & \phi\mathbf{C}_{kk} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \\ \vdots \\ \mathbf{V}_k^T \end{bmatrix}.$$

Then, if each diagonal block is full rank, the matrix $\Phi\mathbf{C}$ and hence $\Phi\mathbf{C}\mathbf{V}^T$ is full rank, since $\mathbf{V}$ is just an orthonormal matrix. Consider each diagonal block of $\Phi\mathbf{C}$, $\phi\mathbf{C}_{ii}$. Since $l \geqslant d_0$, $\mathrm{rank}(\phi\mathbf{C}_{ii}) = d_i$ with probability 1. Since the rank of any block triangular matrix is at least the sum of ranks of the diagonal blocks, $\mathrm{rank}(\Phi\mathbf{C}) = \mathrm{rank}(\Phi\mathbf{C}\mathbf{V}^T) = \sum_i d_i = r$. Since $\mathrm{row}(\Phi\mathbf{C}\mathbf{V}^T) \subset \mathrm{row}(\mathbf{C}\mathbf{V}^T)$ and $\mathrm{rank}(\Phi\mathbf{C}\mathbf{V}^T) = \mathrm{rank}(\mathbf{C}\mathbf{V}^T)$, $\mathrm{row}(\Phi\mathbf{C}\mathbf{V}^T) = \mathrm{row}(\mathbf{C}\mathbf{V}^T)$. Hence, $\mathbf{P}_{\mathbf{Y}*} = \mathbf{P}_{\mathbf{A}*}$ and the conclusion follows. ■

To understand this theorem, we can first consider the intuition behind random projections: owing to the randomness of the projections, a linearly independent set of vectors in the row/column space of the original matrix is obtained and hence the subspace is captured. With an RBD sketching matrix, one can only obtain random projections within the subspace of each group of rows. However, if the subspaces spanned by these row groups overlap, we may obtain a linearly independent set of vectors that capture the row space of the whole matrix by using a few random projections of each row group. The factorization in (Equation 2.21) captures this dependence between the subspaces spanned by the row groups.

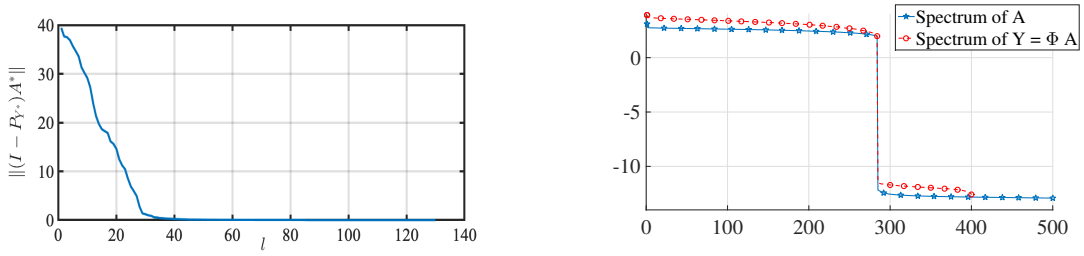


Figure 2.9: *(a) shows the error* $\|(\mathbf{I} - \mathbf{P}_{\mathbf{Y}*})\mathbf{A}^*\|$ *and (b) shows logarithm of the spectrum of* $\mathbf{A}$ *and* $\mathbf{Y}$ *for* $l = 40$*. The number of significant singular values of* $\mathbf{A}$ *and* Y *can be seen to be approximately the same and hence they have approximately the same rank and row space.*

The bound in Theorem 3 can be tight or loose depending on the candidate matrix $\mathbf{A}$. The bound is tight when the row spaces span non-overlapping subspaces, since in this case, random projections within a row group do not provide any information about the other row groups. When the subspaces do overlap the bound can be significantly improved, even though it is already better than the trivial bound $l \geqslant \max_i r_i$. For example, consider a scenario where the subspaces spanned by the $\mathbf{A}_i$'s are nested in the subspaces spanned by the $\mathbf{A}_j$'s for all $i < j$. In such a case, by obtaining random projections within the span of $\mathbf{A}_i$, we are also already obtaining random projections in the subspace spanned by $\mathbf{A}_j$ for all $i < j$. This allows for l to be much smaller, even if some of the d_i are large, contrary to what Theorem 3 predicts. As we note in the next subsection, this is precisely the case for the linear operator associated with range-limited images.

Although our result concerns exactly low-rank matrices, it can be generalized to matrices that are approximately low-rank using perturbation theory for projection matrices [29]. In general, for matrices with numerical ranks much smaller than the ambient dimensions, the subspaces spanned by the $\mathbf{A}_i$'s will overlap. This overlap reduces the number of random vectors needed per block. Figure 2.9 shows the error $\|(\mathbf{I} - \mathbf{P}_{\mathbf{Y}*})\mathbf{A}^*\|$ and the spectra of the matrices $\mathbf{A}$ and $\Phi\mathbf{A}$ for a randomly generated test matrix of size 2000×1000, with $\max_i(r_i) \approx 300$ and $d_0 = 110$. We observe empirically that the row space was captured with $l \approx 40$.

Consider a 1D array and a 2D image. For an image with delta thickness with reflectors

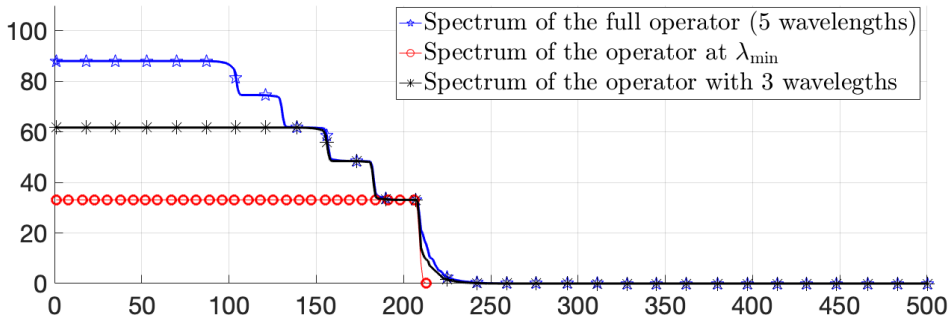


Figure 2.10: *Normalized spectra of the full operator with multiple excitation wavelengths and of the operator at just the highest frequency. It is clear that the row spaces of the operators at the higher wavelengths are nested in that of the operator at the lowest wavelength. The same relationship is observed between the operator at any other excitation wavelength and higher wavelengths.*

at a known fixed depth R_0, the rows of $\mathbf{A}_{\lambda_i}$ are unit-length sinusoids in the frequency range $[-D/\lambda_i,\ D/\lambda_i]$, with a modulation term $e^{-j2\pi R_0/\lambda_i}$. The row space of $\mathbf{A}_i$ is therefore well approximated by the first $2D/\lambda_i$ discrete prolate spheroidal sequences (DPSS) [18]. As the wavelength progresses from the highest $\lambda_{\max}$ to the lowest $\lambda_{\min}$, the subspace spanned by the rows of $\mathbf{A}_{\lambda_i}$ is nested in the subspace spanned by $\mathbf{A}_{\lambda_j}$ for all $j > i$. This is shown in Figure 2.10 where the full operator with all wavelengths has a row space of the same dimension as the operator at only the highest frequency.

Although Theorem 3 is appealing because of its simplicity, it can lead to suboptimal estimates of l. For example, if $\lambda_{\min} = 7.5$cm and $\lambda_{\max} = 15$cm, one could only hope for a reduction in the number of measurements by a factor of 2. But as we have already seen in Figure 2.1, we can easily achieve better subsampling.

The following two theorems provide much stronger results: they directly address the question of when $l \geqslant r/K$ random projections suffice to achieve $\|(\mathbf{I} - \mathbf{P}_{\mathbf{Y}^T})\mathbf{A}^T\| = 0$. In the context of array imaging, they address the question of when using K different excitation wavelengths can offer the luxury of imaging with roughly only M/K measurements. Their proofs are deferred to the appendix.

Theorem 4 *Let* $\mathbf{A} = \left[\begin{smallmatrix}\mathbf{A}_1\\ \mathbf{A}_2\end{smallmatrix}\right]$ *and* $\mathbf{H} = \left[\begin{smallmatrix}\phi\mathbf{A}_1\\ \phi\mathbf{A}_2\end{smallmatrix}\right]$, *with* $\mathrm{row}(\mathbf{A}_1) \subseteq \mathrm{row}(A_2)$. *Assume* $\mathbf{A}_2$ *is full row rank and define* $\mathbf{U} = \mathbf{A}_1\mathbf{A}_2^T$. *If the entries of* ϕ *are drawn from a continuous*

distribution, then with probability 1, $\|(\mathbf{I}-\mathbf{P}_{\mathbf{H}^T})\mathbf{A}^T = 0\|$ *for* $l \geqslant M/2$*, if no real eigenvalue of* $\mathbf{U}$ *has an algebraic multiplicity greater than* $M/2$*.*

Theorem 4 provides the necessary and sufficient conditions on an ensemble of two matrices with a nested subspace structure under which only $r/2$ projections of each block are sufficient to capture the row space. The condition prohibits the existence of an invariant subspace of dimension greater the $M/2$ that is common to both $\mathbf{A}_1$ and $\mathbf{A}_2$.

The eigenvalue distribution for the array imaging matrices is shown in Figure 2.11a when $\lambda_{\min} = 7.5$cm and $\lambda_{\max} = 10$cm. It is clear that the matrices meet the required condition and hence only $M/2$ measurements suffice, in contrast to what is predicted by Theorem 3, which would be at least $3M/4$.

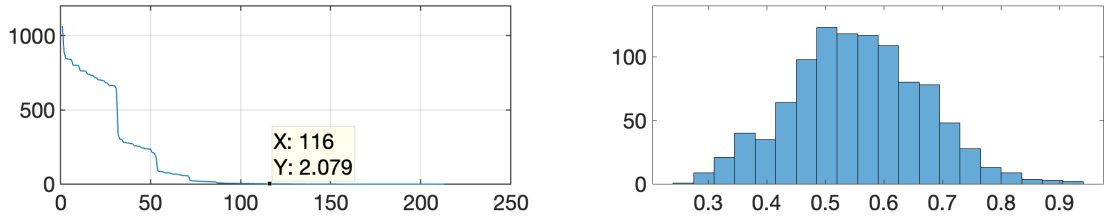
In our next result, we extend the result to the case with more than two blocks and provide a sufficient condition on the ensemble of K matrices $\{\mathbf{A}_i\}$ under which only $l \geqslant r/K$ random projections per block can capture the full row space. To do this, we define the following matrices:

$$\mathbf{H} = \begin{bmatrix} \phi\mathbf{A}_1 \\ \phi\mathbf{A}_2 \\ \vdots \\ \phi\mathbf{A}_K \end{bmatrix} \text{ and } \widehat{\mathbf{H}} = \begin{bmatrix} \mathbf{V}_{\mathcal{S}}\mathbf{A}_1 \\ \mathbf{V}_{\mathcal{S}}\mathbf{A}_2 \\ \vdots \\ \mathbf{V}_{\mathcal{S}}\mathbf{A}_K \end{bmatrix} \tag{2.27}$$

where $\mathbf{V}_{\mathcal{S}}$ is any orthonormal matrix of size $l \times M$.

Theorem 5 *Given an ensemble of* K *matrices* $\{\mathbf{A}_i\}$ *for* $i = 1, \cdots, K$*, each of size* $M \times N$ *and a matrix* $\phi \in \mathbb{R}^{l\times M}$ *with entries drawn from the standard normal distribution, with probability* 1*,* $\|(\mathbf{I} - \mathbf{P}_{\mathbf{H}^T})\mathbf{A}^T\| = 0$ *for* $l \geqslant r/K$*, if there exists an orthonormal basis* $V_{\mathcal{S}} \in \mathbb{R}^{l\times M}$ *such that* $\widehat{H}$ *has full row rank.*

Intuitively, Theorem 5 requires that there is at least one subspace of dimension $l = r/K$, which when projected onto the matrices $\mathbf{A}_i$ results in a set of K linearly independent subspaces. In Figure 2.11b, we show the histogram of the smallest singular values of the



(a) (b)

Figure 2.11: *(a) The eigenvalue distribution of $A_1 A_2^T$ for two exctiation wavelengths 7.5cm and 10cm. The eigenvalue distribution ensures that $l = M/2$ random projections suffice (b) Histogram of the smallest singular value of $\widehat{M}$ over 1000 realizations of randomly generated orthobases $V_{\mathcal{S}}$, for the array imaging operator with 8 excitation wavelengths were used between 7.5cm and 10cm. There exist many orthobases such that the sufficient condition of Theorem 5 holds. In this case, hence, $l = M/8$ random aperture codes are enough for imaging.*

matrix $\widehat{\mathbf{H}}$ for the array imaging operator with $K = 8$ excitation wavelengths placed uniformly between $\lambda_{\min} = 7.5$cm and $\lambda_{\max} = 15$cm, for 1000 realizations of randomly chosen orthonormal basis $\mathbf{V}_{\mathcal{S}}$. In this case, the number of array elements was $M = 213$ and the scene considered had delta thickness. Hence, approximately only 30 spatial measurements suffice in imaging any such scene. As the range extent of the images increases, the nested structure in the row spaces ceases to exist. However, the subspaces still have a high degree of overlap and a number of aperture codes much smaller than the number of conventional beams used suffice for imaging.

The conditions stated in Theorem 4 and Theorem 5 ensure that the lack of diversity among the diagonal blocks of the RBD matrix is compensated for by the data matrix $\mathbf{A}$ itself. As we show in further sections, the array imaging operator satisfies these conditions, thus lending itself to spatial subsampling.

2.1.6 Imaging experiments using coded aperture arrays

We now provide experimental results to show the effectiveness of aperture coding. Various experiments were conducted: aperture coding simulations were conducted with images at a constant depth from the antenna array, for flat images parallel to the 2D array and delta thickness images that are multi-depth with known and unknown range profiles. The next

set of experiments deal with subsampling the array for the same class of images. These are followed by simulations in the presence of noise. We then provide simulation results for objects whose range limits are not thin and span a range of 15cm. The following array parameters were used in all the simulations: an array of 40×40 elements were used with 15 excitation wavelengths placed regularly in the bandwidth of 7.5cm and 15cm. The elements were placed at half the smallest wavelength. The scene was assumed to be within the angular span of $[-\pi/4,\ \pi/4]$ in both directions. For this configuration the number of beams/ measurements needed at each frequency for conventional imaging is about 1100. Quantitative error values for all the experiments are given in Table 2.1. [1]

Constant range and multi-depth images

Figure 2.1 shows reconstruction results for images at a constant, known range. Conventional beamforming requires about 1100 beams in this case to scan over the entire image. It is clear that similar reconstruction performance can be obtained using as few as 80 beams with wideband excitation. The relative reconstruction error is almost negligible up to 80 aperture codes. (shown in Table 2.1).The row space of $\mathbf{A}$ in this case is 1100. Hence, with only 50 codes and 15 excitation frequencies, the coded aperture system ΦA has a maximum rank of only 750 and does not capture the full row space of $\mathbf{A}$. This results in a poor reconstruction, as seen in Figure 2.1d.

Aperture coding is also effective for multi-depth images. A scene with three segments, each at a different depth was considered. The depth map of the scene used in the simulations is shown in Figure 2.12. In Figure 2.13, we present simulation results when the depth profile was assumed to be known a priori. Again only about 80 beams are sufficient to get good quality reconstructions.

We further explore the performance of coded aperture imaging when the depth profile is unknown. Boufounos in [30] describes a method to infer the depth profile of an image using CoSaMP algorithm. The algorithm uses full array measurements in the least squares

[1]The simulations were performed in MATLAB and our code can be found at the following link: https://github.com/rsharma20/Aperture_coded_imaging

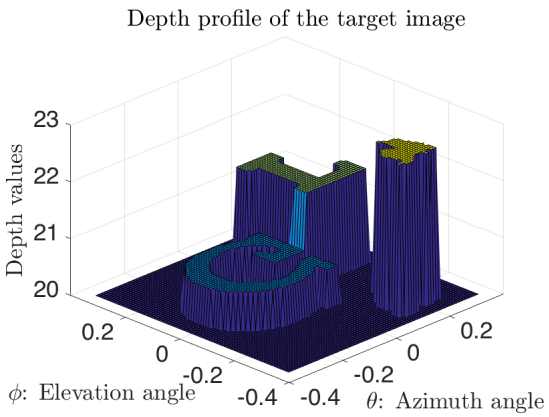


Figure 2.12: *Depth map of the multi-depth image used in simulations.*

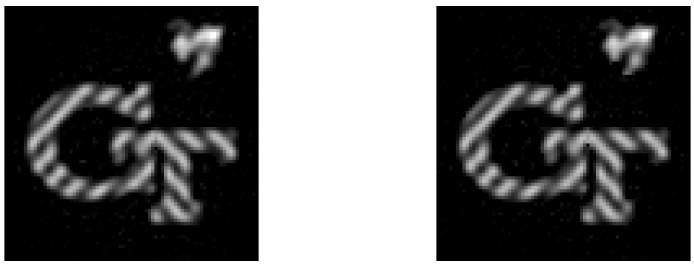
step. We replace this step with sketched least squares and coded measurements. Simulation results are shown in Figure 2.14. We note that aperture coding still performs well, with the relative error being negligible even for 160 generic beams.

Imaging with noise

We simulated an imaging scenario at 20dB input SNR level. In Figure 2.15, we show reconstruction results of these simulations. The corresponding signal reconstruction error values and output SNR are given in Table 2.1. The impact of noise on an aperture coded system is similar to its impact on the conventional system, except that the characterization depends on the coding matrix Φ as well. However, we have shown that this dependence can be predicted a priori once a Φ has been instantiated. The same analysis holds for any input SNR level. It is clear from Table 2.1 that aperture coded measurements can be used under noisy regimes as well and a desired level of performance can be achieved by studying the system ΦA a priori and tuning the system parameters.

Imaging scenes with higher range limits

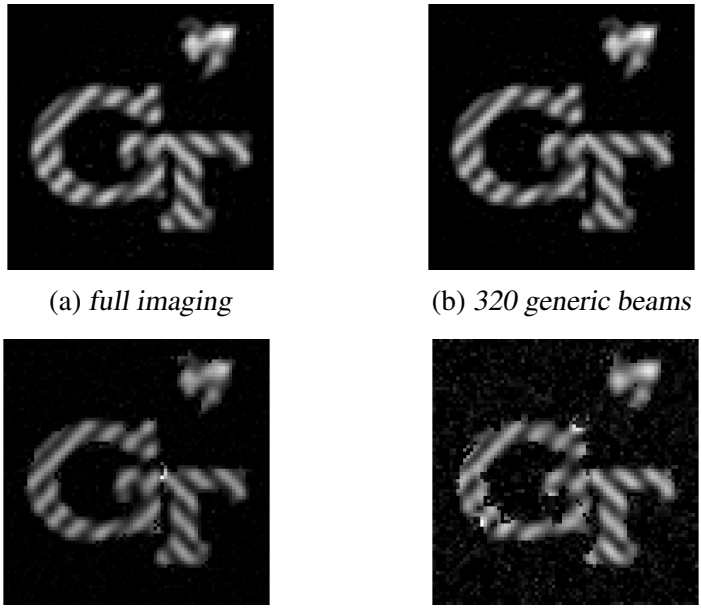
Aperture coding can be highly effective even when imaging scenes with higher range limits. As predicted by our theorems, when the range limit increases, more measurements are required, but it can still less be than that used in conventional imaging. We demonstrate this in our next set of simulations. We consider an object made up of five layers as shown in Figure 2.16a. These layers lie in a region of width 15cm, in the far-field of a 2D array



(a) *full imaging* (b) *320 generic beams*

(c) *160 generic beams* (d) *80 generic beams*

Figure 2.13: *Aperture coded imaging for a multi-depth image. (a) represents the conventional method, which uses about 1100 beams. (b), (c) and (d) show reconstruction results with* 320, 160 *and* 80 *generic beams.*



(a) *full imaging* (b) *320 generic beams*

(c) *160 generic beams* (d) *80 generic beams*

Figure 2.14: *Aperture coded imaging for a multi-depth image with an unknown range profile. (a) represents the conventional method, which uses about 1100 beams.*

at a depth of 20m. We again use 15 excitation wavelengths between 7.5cm and 15cm. In Figure 2.16, we present reconstruction results by using 640, 480, and 320 aperture codes

Table 2.1: *Relative reconstruction error values for different classes of images. Aperture codes with weights chosen from standard normal distribution were used.*

Imaging mode	(CR)	(FS)	(MD)	(MDUR)	Constant range noisy	
					Signal error	OP SNR
Full (1100 beams)	NA	NA	NA	NA	NA	16.09
320 codes	2.7e-5	4.4e-4	3e-4	1.1e-2	2.5e-5	16.0117
160 codes	7.4e-5	3.3e-4	1.2e-3	3.1e-2	6.3e-5	15.9112
80 codes	4.2e-4	5.1e-4	4.8e-2	1.6e-1	4.9e-4	15.3206

Table 2.2: *Relative reconstruction errors with subsampled array: The array was randomly subsampled to have 320, 160 and 80 elements. Abbreviations are as in Table 2.1.*

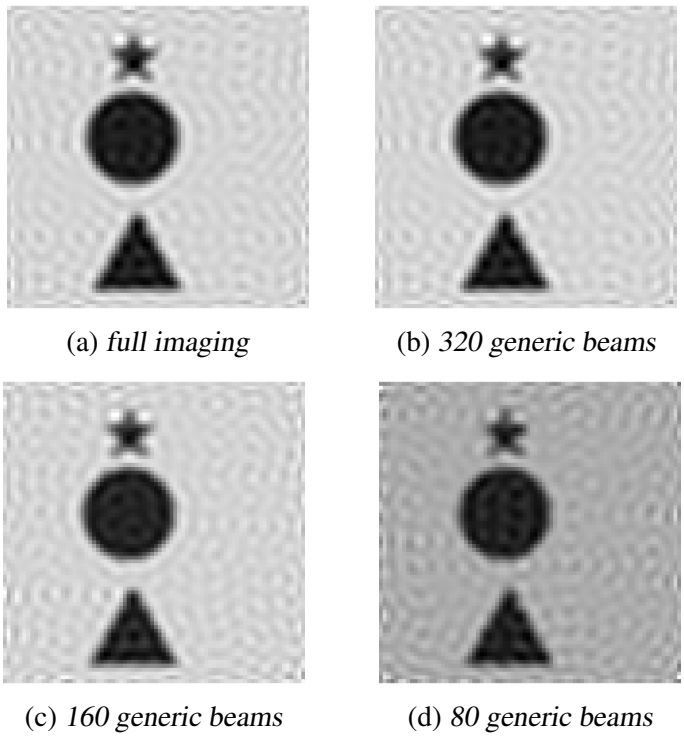
Imaging mode	CR	FS	MD	MDUR
320 elements	2e-4	1e-4	1e-3	3.8e-2
160 elements	9e-4	4e-4	3.1e-2	6.8e-2
80 elements	1.6e-3	1.1e-3	5.8e-3	18.4e-2

in place of 1100 beams used in conventional imaging.

2.2 Broadband source localization

2.2.1 Introduction

One of the main applications of antenna arrays is source localization. A typical application is that there are a number of signal sources present in the far-field region of an antenna array, each at a particular angle in space θ with respect to the array. Each source transmits a signal centered at some freqeuncy f_c (also known as the channel frequency), and the elements of the antenna array receive noisy, delayed versions of the superposition of these signals. Under the assumption that the pair (f_c^i, θ^i) uniquely identifies the i^{th} source, the task is to estimate the parameters f_c^i and θ^i for each source. We henceforth refer to this problem as the source localization problem. When the source signals are narrowband and the f_c^i's and θ^i's are on the continuum, this problem is the same as the line spectral estimation or the super-resolution problem [31, 32]. To develop a formal model of the problem, let us



(a) *full imaging* (b) *320 generic beams*

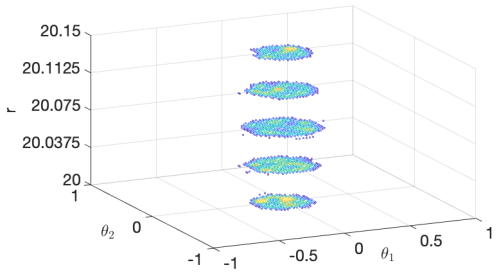
(c) *160 generic beams* (d) *80 generic beams*

Figure 2.15: *Aperture coding in the presence of noise. The regularization parameters was varied in each case to match the SNR of the full imaging scenario. The noise performance was preserved without compromising on the signal reconstruction quality, as seen in Table 2.2.*

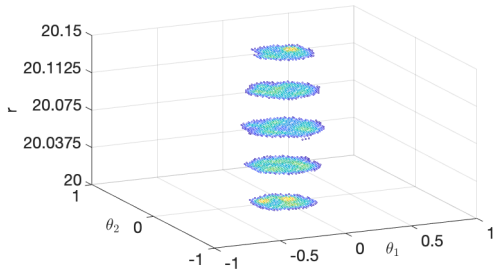
consider a linear array with M elements placed uniformly at intervals of length d. Let us further restrict the problem to identifying only the Direction Of Arrival (DOA) θ_i, by fixing the center frequencies of all the sources to a known frequency f_0. Assuming that each source emits a pure tone at frequency f_0, the signal received at the m^{th} element can be written as

$$\begin{aligned} x_m(t) &= \sum_{i=1}^{L} \exp(j2\pi f(t - \frac{md\sin\theta_i}{c})) \\ &= \exp(j2\pi ft)\sum_{i=1}^{L} \exp(-j2\pi m\frac{fd\sin\theta_i}{c}) \\ &= \exp(j2\pi ft)\sum_{i=1}^{L} \exp(-j2\pi m\frac{fd\sin\theta_i}{c}) \end{aligned}$$

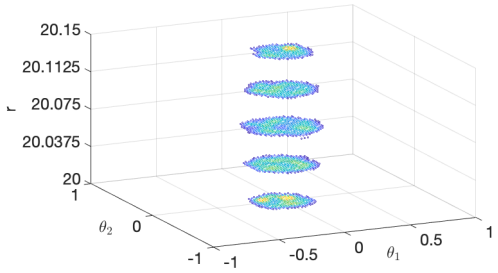
where c is the speed of light an we assume L sources are present. Note that the complex amplitude of the M-length vector of outputs at any time instant t is *spatial* sinusoid



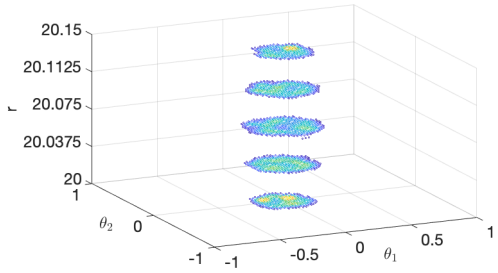
(a) *Full reconstruction* $\sim$ *1100 measurements*



(b) *640 generic beams, relative error: 0.0168*



(c) *480 generic beams, relative error: 0.0298*



(d) *320 generic beams, relative error: 0.0466*

Figure 2.16: *Reconstruction results for a target object with (at* 20m*) with aperture coded acquisition. The target scene consists of five discs one behind the other. In the figure above, the 2D array would face the discs from below. Using a set of* 15 *excitation wavelengths between* 7.5cm *and* 15cm *such a scene can be reconstructed using only about 320 spatial measurements, unlike conventional beamforming which would require about* 1100 *measurements.*

of frequency $\frac{fd\sin\theta_i}{c}$. Hence, both DOA estimation and channel frequency estimation are instances of the line spectral estimation problem: given finite number of samples of a superposition of sinusoids, the task is to estimate the active frequencies.

Although a well studied problem, the assumption that sources emit narrowband signals seldom holds true. When the source signals have a non-trivial bandwidth, existing algorithms discussed above often tend to fail in correctly identifying the sources. Figure 2.17 illustrates this phenomenon, with the percentage of sources detected decaying with an increase in the signal bandwidth. Our goal in this section is to develop algorithms that address this gap between the state-of-the-art algorithms and the common practical scenario of broadband sources.

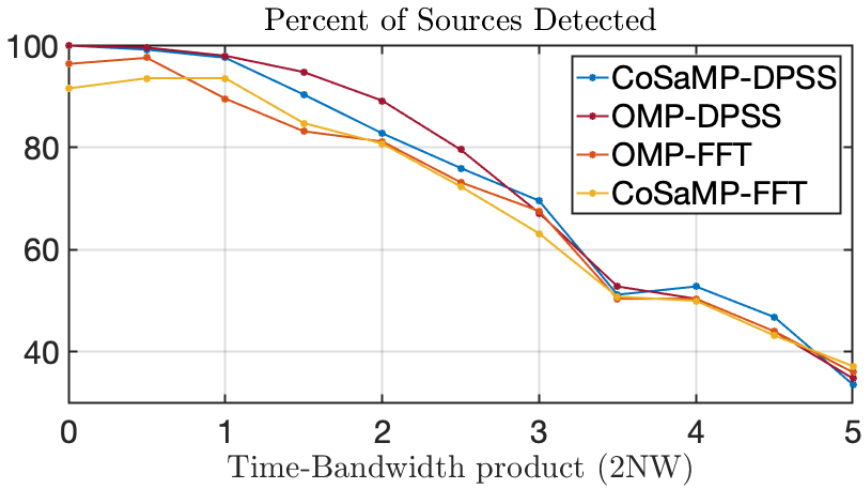


Figure 2.17: *An illustration of how iterative narrowband source localization algorithms fail when the underlying sources are broadband. In the figure, a bandwidth of greater than* 1 *indicates broadband signals. The percentage of sources detected by the narrowband algorithms decays for broadband sources.*

2.2.2 Broadband source localization as a subspace estimation problem

Our approach to perform broadband source localization relies on the fact that any bandlimited signal observed over a finite time window *lies approximately* in a low dimensional subspace. These subspaces are characterized by the center frequency f_c^i of the active band the signal occupies, and the source localization process is then a search over the set of candidate subspaces.

It is well known that a signal with finite temporal support must have infinite spectral support. In other words, a finite length signal cannot be bandlimited. Since most real-world signals are of finite time, how is it then possible for any of them to be bandlimited? The answer to this dilemma was the subject of a series of seminal papers by Slepian, Landau and Pollock [17, 20, 21, 22, 18, 19]. The answer lies in the fact that signal can be *approximately* bandlimited and time limited simultaneously. A natural question then is regarding the best representation of such signals. In other words, how can the simultaneous structure be best captured?

Signals that are "well supported" on compact intervals both temporally and spectrally

can in fact be shown to lie close to a low-dimensional subspace. Let $\mathbf{x}[n],\ n = 0,\ \cdots, N-1$ be the vector of N uniformly spaced samples of a signal bandlimited to $W < 0.5$. Then, any such $\mathbf{x}$ can be "well-approximated" by $2NW$ number of the so-called Discrete Prolate Spheroidal Sequences (DPSS's) when they are time limited to N samples [33]. We now give a brief overview of the DPSS's and their ability to act as low dimensional models of simultaneously bandlimited and time-limited signals.

Definition 1 *[21] Given N and $W < 0.5$, the Discrete Prolate Spheroidal Sequences (DPSS's) are a sequence of N real-values discrete sequences $s_{N,W}^{(0)}, s_{N,W}^{(1)}, \cdots, s_{N,W}^{(N-1)}$ that along with their associated scalar eigenvalues*

$$1 > \lambda_{N,W}^{(0)}, \lambda_{N,W}^{(1)}, \cdots \lambda_{N,W}^{(N-1)} > 0$$

satisfy

$$\mathcal{B}_W \left(\mathcal{T}_N \left(s_{N,W}^{(\ell)} \right) \right) = \lambda_{N,W}^{(\ell)} s_{N,W}^{(\ell)},$$

where $\mathcal{B}_W(\cdot)$ denotes the bandlimiting operator to W and $\mathcal{T}_N(\cdot)$ is the time-limiting operator that restricts a discrete sequence to length N.

Note that each sequence $s_{N,W}^{(\ell)}$ has infinite-support. The N-dimensional DPSS vectors are defined as follows.

Definition 2 *[33] Given N, W, the DPSS vectors are defined by time restricting the discrete prolate spheroidal sequences to the first N indices.*

$$\mathbf{s}_{N,W}^{(\ell)}[n] = \mathcal{T}_N \left(s_{N,W}^{(\ell)} \right)[n] \tag{2.28}$$

for $\ell, n = 0,\ \cdots, N-1$.

The N DPSS vectors generated as above are orthogonal and form an orthobasis for $\mathbb{R}^N$. The DPSS vectors provide an efficient way to represent a finite numbet of samples obtained frmo bandlimited signals. In particular, given N and W, the first $k = 2NW$

DPSS vectors span the "best" k-dimensional subspace that can uniformly approximate signals bandlimited to W and time limited to N samples. This notion of uniform approximation of bandlimited signals is captured in the following result provided in [33]. Let $\mathbf{E}_f = \mathrm{diag}(\mathrm{e}^{j2\pi f_c n}),\ = 0, 1\ \cdots,\ N-1$ for any digital frequency $0 \leqslant f_c \leqslant 0.5$ and let $\mathbf{S}_{N,W} = [\mathbf{s}_{N,W}^{(0)}\ \mathbf{s}_{N,W}^{(1)}\ \cdots \mathbf{s}_{N,W}^{(N-1)}]$ be the $N \times N$ matrix containing the N DPSS vectors.

Theorem 6 *[33] Let $x(t)$ denote a continuous-time, zero-mean, wide sense stationary random process with power spectrum*

$$P_x(F) = \begin{cases} \frac{1}{B_{band}},\ F \in [F_c - \frac{B_{band}}{2}, F_c + \frac{B_{band}}{2}], \\ 0,\ \text{otherwise} \end{cases}$$

and let $\mathbf{x} = [x(0)\ x(T_s)\ \cdots\ x((N-1)T_s)]^\top \in \mathbb{C}^N$ denote a finite vector of samples acquired from $x(t)$ with a sampling interval $T_s \leqslant 1/(2\max\{|F_c \pm \frac{B_{band}}{2}|\})$. Then, the space spanned by the first k columns of the matrix $\mathbf{Q} = \mathbf{E}_{f_c}\mathbf{S}_{N,W}$ best approximates $\mathbf{x}$ among all k-dimensional subspaces in an MSE sense and the MSE is given by:

$$\mathrm{E}[\|\mathbf{x} - P_{\mathbf{Q}}\mathbf{x}\|_2^2] = \frac{1}{2NW}\sum_{\ell=k}^{N-1}\lambda_{N,W}^{(\ell)}, \tag{2.29}$$

where $\mathrm{E}[\|\mathbf{x}\|_2^2] = N$ due to suitable normalization.

Note that this result only suggests in a probabilistic sense that most bandpass signals that are sampled and time-limited can be well-approximated by the k DPSS vectors. There will always be a relatively small number of bandpass signals that cannot be well approximated. Considering that the fraction of such signals is small, we will henceforth assume that bandlimited signals that are sampled and time-limited can be well approximated by the top $k \approx 2NW$ eigenvectors modulated to the center frequency of the band.

Let us denote the k-dimensional subspace spanned by the first k columns of the matrix $\mathbf{E}_f\mathbf{S}_{N,W}$ by $\mathcal{S}_f$. Then, the problem of broadband source localization is to identify the

collection of subspaces that make up the received signal. This is highly reminiscent of the standard compressed sensing problem, where signals are modeled to lie in the union of a number of 1-dimensional subspaces. In our case however, the union is of k-dimensional subspaces with each subspace being a member of a candidate set of subspaces, given by the span of roughly the first k columns of the matrix $\mathbf{E}_f\mathbf{S}_{N,W}$ where $f \in [0, 0.5]$. As we will see shortly, our proposed algorithm is a natural extension of the iterative algorithms proposed in the compressed sensing literature. These iterative algorithms rely on fast computational tools that allow fast projections onto the candidate subspaces and fast solution to linear systems involving the DPSS vectors that were recently developed in [34].

2.2.3 Computational tools for Slepian subspace coefficients

In order to identify the active subspaces, a basic operation that is necessary is the projection of the observations onto the candidate subspaces. One may note that for any given band of frequencies $[f_c - W, f_c + W]$, one can use a set of uniformly spaced sinusoids as a proxy for the basis of the space of bandlimited signals. In particular, consider the matrix formed by columns of the form

$$\mathbf{e}_m = [1 \ \mathrm{e}^{j2\pi m1/N} \ \cdots \ \mathrm{e}^{j2\pi m(N-1)/N}]^* \tag{2.30}$$

where $m/N \in [f_c - W, f_c + W]$ and $m \in \{0, 1, \cdots, N-1\}$. By leveraging the Fast Fourier Transform (FFT), one could compute projections onto the subspaces in $O(N \log N)$ time. However, as we shall see in subsection 2.2.5, the Fourier basis is easily outperformed by the DPSS vectors. We will now briefly review two main results from [34] that address the important problem of designing fast computational tools to work with the DPSS vectors.

Let us define the matrix $\mathbf{S}_K = [\mathbf{s}^{(0)}_{N,W} \ \mathbf{s}^{(1)}_{N,W} \ \cdots \ \mathbf{s}^{(K)}_{N,W}]$. The first result considers projecting a given N-dimensional signal $\mathbf{x}$ onto the subspace spanned by the columns of $\mathbf{S}_K$. Naive matrix multiplication to compute $\mathbf{S}_K\mathbf{x}$ requires $O(N^2W)$ computations, which may

be prohibitive for large scale applications including array signal processing. However, as shown in [34], the projection matrix $\mathbf{S}_K\mathbf{S}_K^*$ can be re-written as

$$\mathbf{S}_K\mathbf{S}_K^* = \mathbf{B}_{N,W} + \mathbf{U}_1\mathbf{U}_2^* + \mathbf{E} \tag{2.31}$$

where

$$\mathbf{B}_{N,W}(m,n) = \frac{\sin 2\pi W(m-n)}{\pi(m-n)}, \tag{2.32}$$

is the $N\times N$ *prolate* matrix, $\mathbf{U}_1, \mathbf{U}_2$ are $N\times r_2$ matrices with $r_2 \leqslant \left(\frac{8}{\pi^2}log(8N) + 12\right)\log\left(\frac{15}{\epsilon}\right)$ and $\|\mathbf{E}\| \leqslant \epsilon$. Note that $\mathbf{B}_{N,W}$ is a Toeplitz matrix and computing $\mathbf{B}_{N,W}\mathbf{x}$ can be done in $O(N\log N)$ time due to the fact that it can be diagonalized by the FFT matrix. Since $\mathbf{U}_1$ and $\mathbf{U}_2$ are $N\times O(\log N\log\frac{1}{\epsilon})$ matrices, they can be applied to a vector in $O(N\log N\log\frac{1}{\epsilon})$ time. Hence, by computing $\mathbf{B}_{N,W}\mathbf{x} + \mathbf{U}_1\mathbf{U}_2^*\mathbf{x}$, we can compute an ϵ-approximation to $\mathbf{S}_K\mathbf{S}_K^*\mathbf{x}$ in only $O(N\log N\log\frac{1}{\epsilon})$ time. This directly gives us a fast ($O(N\log N\log\frac{1}{\epsilon})$) algorithm to project onto any candidate subspace given by the matrix $\mathbf{Q} = \mathbf{E}_{f_c}\mathbf{S}_K$, since $\mathbf{E}_{f_c}$ is a diagonal matrix.

The second computational tool that we will leverage is an analog of the Fast Fourier Transform itself. Let us define the *tapered spectral estimate* of an N-length vector $\mathbf{x}$:

$$\widehat{S}_k(f) = |\sum_{n=0}^{N-1}\mathbf{s}_k[n]\mathbf{x}[n]\mathrm{e}^{-j2\pi fn}|^2. \tag{2.33}$$

We can then define the following *multitaper spectral estimate*:

$$\widehat{S}_K(f) = \frac{1}{K}\sum_{k=0}^{K-1}\widehat{S}_k(f). \tag{2.34}$$

We also have the following equivalence between $\widehat{S}_K(f)$ and the projection of $\mathbf{x}$ onto the K-dimensional Slepian subspace centered at f:

$$\widehat{S}_K(f) = \left\|\mathbf{S}_K^*\mathbf{E}_f^*\mathbf{x}\right\|^2. \tag{2.35}$$

Computing $\widehat{S}_K(f)$ at all the grid frequencies $f = m/N,\ m \in \{0, 1, \cdots,\ N-1\}$ then gives an alternate spectral estimate of the signal $\mathbf{x}$ and lets us evaluate the projection of $\mathbf{x}$ onto different candidate subspaces. We will show in subsection 2.2.5 that the multitaper spectral estimate forms a crucial tool in accurate broadband source localization. However, computing the spectral estimate $\widehat{S}_K(f), f \in \{0, 1/N,\ \cdots,\ N - 1/N\}$ requires $O(KN \log N)$ operations. But, as shown in [35], it is possible to compute an approximation $\widetilde{S_k}(f)$ at all grid frequencies such that

$$|\widehat{S}_K(f) - \widetilde{S_k}(f)| \leqslant \frac{\epsilon}{K} \|\mathbf{x}\|^2 \tag{2.36}$$

in $O(N \log^2 \log \frac{1}{\epsilon})$ time. Armed with both of the above tools for fast computations with the DPSS vectors, we are now ready to develop our algorithms for broadband source localization.

2.2.4 Iterative algorithms for broadband DOA

The problem of broadband source localization has many similarities with the classical compressed sensing problem setup. For a k-sparse signal $\mathbf{z} \in \mathbb{R}^n$ and a suitable "sensing" matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, consider noisy linear measurements of the form

$$\mathbf{x} = \mathbf{A}\mathbf{z} + \eta = \sum_{i \in \{\mathrm{j}:\mathbf{z}[j] \neq 0\}}^{k} \mathbf{A}_i \mathbf{z}_i + \eta \tag{2.37}$$

where η is additive noise. The vector $\mathbf{y}$ is then a sparse linear combination of the "dictionary element" given by the columns of $\mathbf{A}$. Similarly, we can model a multiband signal as a sparse linear combination of subspaces:

$$\mathbf{x} = \sum_{\ell=1}^{L} \mathbf{E}_{f_\ell} \mathbf{S}_K \alpha_\ell \tag{2.38}$$

where $\mathbf{E}_{f_\ell}\mathbf{S}_K$ denotes the basis for the ℓth band and α_ℓ denotes the corresponding set of coefficients. Note that we do not make recommendation on the value of K to be used, but (Equation 2.38) instead emphasizes that multiband signals are sparse linear combinations of subspaces.

Owing to the resemblance compressed sensing, we propose to adapt two algorithms from the compressed sensing literature, Orthogonal Matching Pursuit (OMP) [36] and Compressed Sampling Matching Pursuit (CoSaMP) [37]. Both these algorithms make use of the sparse linear combination signal model. In order to develop the algorithms for multiband signals, we use the following notation.

- $\mathbf{\Psi}_L$ represents a dictionary of L subspaces and takes the form

$$\mathbf{\Psi}_L = \begin{bmatrix} \mathcal{S}_1 & \mathcal{S}_2 & \cdots & \mathcal{S}_L. \end{bmatrix} \tag{2.39}$$

  $\mathcal{S}_i$ denotes the ith subspace and contains a basis for the subspace. In the sequel, we make use of two particular choices for the basis: one given by Fourier vectors, the other by DPSS vectors.

- $\mathrm{Spec}(\cdot)$ computes a spectral estimate of the input signal. For the purpose of this article, it could be either the Fourier transform or the multitaper spectral estimate.

- $P_\Delta(\widehat{S}, K)$ represents a peak-picking operator. Given a spectral estimate $\widehat{S}$, it picks K peaks, with a minimium spacing of Δ

- $\mathcal{P}(\Phi)$ denotes the projection operator that projects a given signal onto the union of the subspaces in the dictionary Φ. This involves solving a least squares problem to determine the coefficients of the active bands that generate the given signal.

- Given an estimate of a band located at f, $G_{\Delta f}(S, f)$ estimates the bandwidth of that source, if unknown a priori and also refines the estimate of f.

- $D_\epsilon(\mathbf{f}, \mathbf{W})$ creates the dictionary element corresponding to the band centered at $\mathbf{f}$ with a bandwidth of $\mathbf{W}$. The dictionary can either be a set of Fourier vectors from the band $[\mathbf{f} - \mathbf{W}, \mathbf{f} + \mathbf{W}]$ or be the set of DPSS vectors spanning the subspace of N-dimensional signals that are most concentrated in the band $[\mathbf{f} - \mathbf{W}, \mathbf{f} + \mathbf{W}]$.

- $\mathrm{Prune}(\mathbf{x}, \mathbf{f}, \mathbf{W}, L)$ is a pruning operator that selects the L sources with the most energy out of the identified support set $\mathbf{f}, \mathbf{W}$.

OMP for broadband source localization

Our first algorithm follows the general the general structure of the OMP algorithm. It is an iterative algorithm that operates on the spectrum of the signal. In each iteration, the frequency band centered around the largest peak in the spectrum selected and along with all such previously selected bands, is "nulled out" using a projection operation. Fast spectral estimation is possible due to the fast multitaper spectral estimation algorithm and fast nulling out is possible due to the fast projection algorithm discussed in the previous section. The algorithm is presented in detail in Algorithm 1.

Algorithm 1 OMP for Broadband Signals

$\mathbf{x} \leftarrow \text{multiband}(N, L)$ ▷ Assume L Sources
$\mathbf{r}^{(0)} \leftarrow \mathbf{x}$
$\mathbf{\Psi}^{(0)} \leftarrow \varnothing$
$i \leftarrow 0$
while Not Converged **do**
  $i \leftarrow i + 1$
  $\mathbf{v} = \mathbf{r}^{(i-1)}$ ▷ Form Proxy
  $\mathbf{S}(f) \leftarrow$Spec(v) ▷ Spectral Estimate
  $\hat{\mathbf{f}}_1 \leftarrow P_\Delta(\mathbf{S}_f, 1)$ ▷ Identify Largest Peak
  $\mathbf{f}_1, \mathbf{W}_1 \leftarrow G_{\Delta f}(\mathbf{S}_f, \hat{\mathbf{f}}_1)$ ▷ Estimate Bandwidth/Refine Center Frequency
  $\mathbf{\Psi}_1 \leftarrow D_\epsilon(\mathbf{f}_1, \mathbf{W}_1)$
  $\mathbf{\Psi}^{(i)} \leftarrow [\mathbf{\Psi}^{(i-1)}\, \mathbf{\Psi}_1]$ ▷ Merge Support
  $\mathbf{f}^{(i)} \leftarrow [\mathbf{f}^{(i-1)}\, \mathbf{f}_1]$
  $\mathbf{W}^{(i)} \leftarrow [\mathbf{W}^{(i-1)}\, \mathbf{W}_1]$
  $\mathbf{x}^{(i)} \leftarrow \mathcal{P}_{\mathbf{\Psi}^{(i)}}(\mathbf{x})$ ▷ Estimation
  $\mathbf{r}^{(i)} \leftarrow \mathbf{x} - \mathbf{x}^{(i)}$ ▷ Update Residual
$\mathbf{f}, \mathbf{W} \leftarrow \mathrm{Prune}(\mathbf{x}^{(i)}), \mathbf{f}^{(i)}, \mathbf{W}^{(i)}, L)$ ▷ Pruning

CoSaMP for broadband source localization

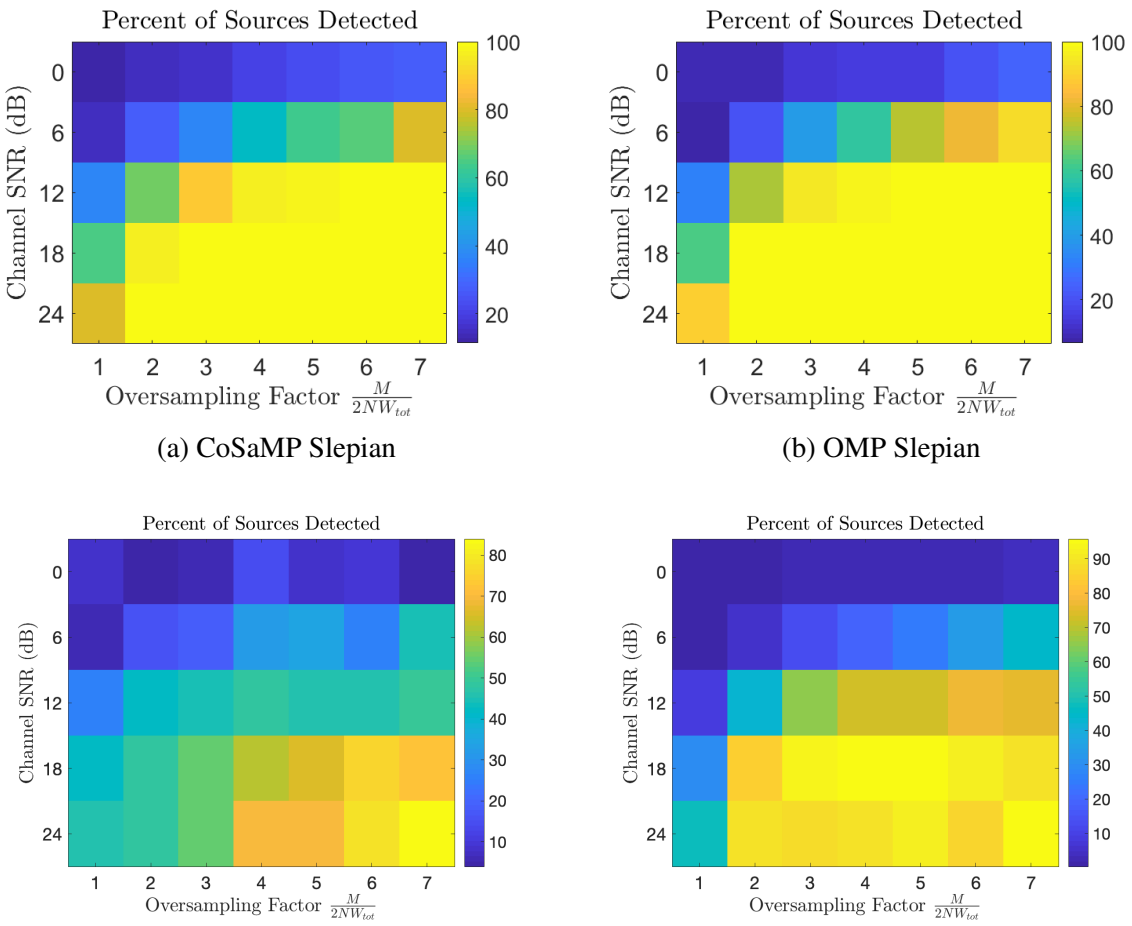
The next algorithm follows the general the general structure of the CoSaMP algorithm. As opposed to the OMP algorithm, CoSaMP chooses $2L$ peaks in every iteration, where L is the number of sources being localized. The $2L$ candidates are then merged with a previously identified set of L candidates. This is followed by a projection onto the set of the $3L$ candidates, followed by a pruning step that chooses the L sources with the most energies. The CoSaMP algorithm has a higher computational complexity per iteration, but generally requires fewer iterations than the OMP algorithm.

Algorithm 2 CoSaMP for Broadband Signals

$\mathbf{x} \leftarrow$ Multiband(N, L) ▷ Assume L Sources
$\mathbf{r}^{(0)} \leftarrow \mathbf{x}$
$\mathbf{\Psi}^{(0)} \leftarrow \varnothing$
$i \leftarrow 0$
while Not Converged **do**
    $i \leftarrow i + 1$
    $\mathbf{v} = \mathbf{r}^{(i-1)}$ ▷ Form Proxy
    $\mathbf{S}(f) \leftarrow$Spec(v) ▷ Spectral Estimate
    $\hat{\mathbf{f}}_{2L} \leftarrow P_{\Delta f}(\mathbf{S}_f, 2L)$ ▷ Identify $2L$ Largest Peaks
    $\mathbf{f}_{2L}, \mathbf{W}_{2L} \leftarrow G_{\Delta f}(\mathbf{S}_f, \hat{\mathbf{f}}_{2L})$ ▷ Estimate Bandwidth/Refine Center Frequency
    $\mathbf{\Psi}_{2L} \leftarrow D_\epsilon(\mathbf{f}_{2L}, \mathbf{W}_{2L})$
    $\hat{\mathbf{\Psi}} \leftarrow [\mathbf{\Psi}^{(i-1)}\, \mathbf{\Psi}_{2L}]$ ▷ Merge Support
    $\hat{\mathbf{f}} \leftarrow [\mathbf{f}^{(i-1)}\, \mathbf{f}_{2L}]$
    $\hat{\mathbf{W}} \leftarrow [\mathbf{W}^{(i-1)}\, \mathbf{W}_{2L}]$
    $\mathbf{x}^{(i)} \leftarrow \mathcal{P}_{\hat{\mathbf{\Psi}}}(\mathbf{x})$ ▷ Estimation
    $\mathbf{f}^{(i)}, \mathbf{W}^{(i)} \leftarrow \mathcal{S}_{\hat{\mathbf{\Psi}}}(\hat{\mathbf{x}}, \hat{\mathbf{f}}, \hat{\mathbf{W}}, L)$ ▷ Pruning
    $\mathbf{\Psi}^{(i)} \leftarrow D_\epsilon(\mathbf{f}^{(i)}, \mathbf{W}^{(i)})$
    $\mathbf{r}^{(i)} \leftarrow \mathbf{x} - \mathcal{P}_{\mathbf{\Psi}^{(i)}}(\mathbf{x})$ ▷ Update Residual

2.2.5 Source localization performance under noise and subsampling

We present a series of experiments to show the efficacy of our proposed broadband source localization algorithms. We compare algorithms based on the percentage of sources detected correctly and accuracy of the center frequencies estimated. The number of sources L was chosen to 10, and we set the total occupied bandwidth to be $\sum_{\ell=1}^{L} W_\ell = 0.04$, thus



(a) CoSaMP Slepian

(b) OMP Slepian

(c) CoSaMP Fourier

(d) OMP Fourier

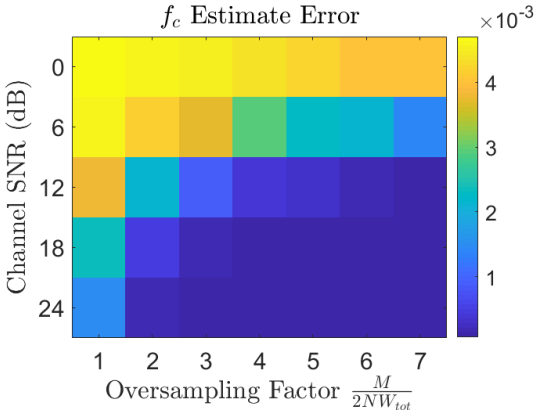
Figure 2.18: *Comparison of broadband OMP and CoSaMP algorithm's detection percentages when using a Fourier and Slepian dictionary. The vertical axis represents the inputs channel SNR and the horizontal axis represents the amount of oversampling w.r.t. to the inputs degrees of freedom.*

setting $2NW \ll N$.

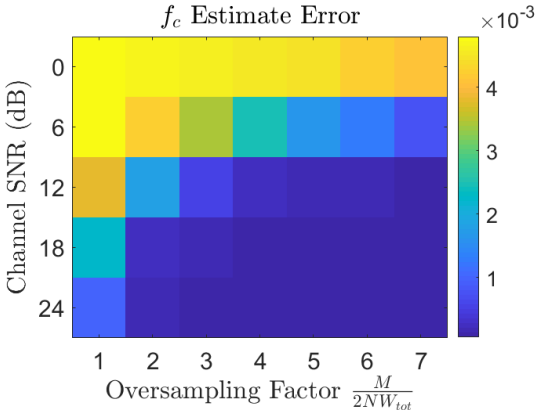
2.2.6 Detection Results

The detection results in Figure (Figure 2.18) depict the results of detection percentages for the CoSaMP and OMP algorithm when using either a Fourier or Slepian Dictionary to perform the estimation and nulling step. It is clear from these results that the Slepian dictionary yields far better detection performance across a wider variety of channel SNRs and oversampling factors. This is expected, since the DPSS vectors are more aptly suited for capturing the energy generally lost due to spectral leakage.

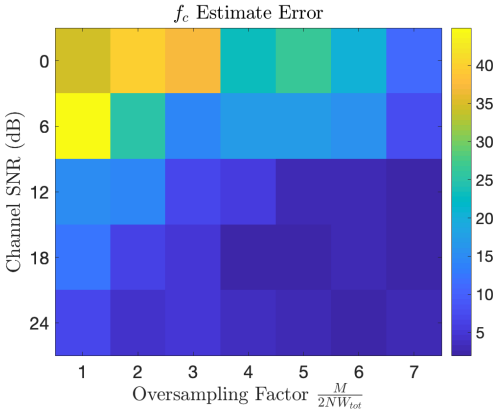
Figure 2.19 shows the center frequency estimation performace of the algorithms under both DPSS and Fourier dictionaries. For OMP and CoSaMP the average normalized f_c error is lower bounded by $\frac{1.03}{N}$ which is very close to the approximate Rayleigh limit of $\approx \frac{1}{N}$. For the Fourier dictionary however, the resolution is well outside the Rayleigh limits.



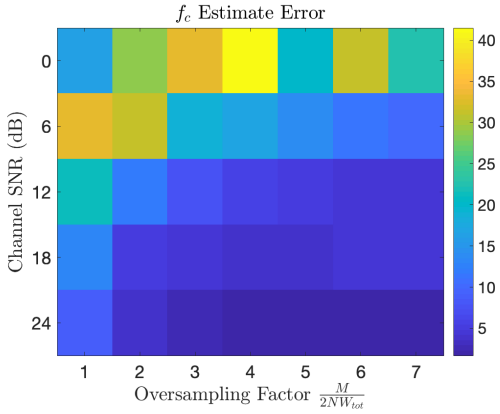
(a) CoSaMP Slepian



(b) OMP Slepian



(c) CoSaMP Fourier



(d) OMP Fourier

Figure 2.19: *Comparison of broadband OMP and CoSaMP algorithm's center frequnecy estimates $\hat{f}_c$ when using a Fourier and Slepian dictionary. The vertical axis represents the inputs channel SNR and the horizontal axis represents the amount of oversampling w.r.t. to the inputs degrees of freedom. The error units are in normalized frequency.*

CHAPTER 3

SKETCHED REGRESSION FOR DECENTRALIZED DATA

3.1 Randomization as an algorithmic tool

Efficient linear algebraic computations are of fundamental importance in machine learning and signal processing applications. The problems in array signal processing described in chapter 2 are among the many examples of real-world optimization problems that rely on linear-alegbraic routines. However, a massive rise in the amount of data calls for a rethinking of many classical algorithms. For example, solving a linear system of equations $\mathbf{y} = \mathbf{A}\mathbf{x}$ using classical methods takes $O(\widetilde{N}d^2)$ time, where $\mathbf{A} \in \mathbb{R}^{\widetilde{N}\times d}$ and this is already prohibitive for many large scale applications.

One particular algorithmic paradigm that aims to address this problem is the so-called "sketch and solve" method. This usually entails obtaining an approximation of the solution, while improving the time complexity of algorithms. These set of methods utilize randomization as a powerful tool to reduce the dimensionality of data and then apply the classical algorithms in the reduced dimension [38, 39, 40, 41]. In this work, we consider two specific examples: sketched matrix multiplication [5] and ridge regression [4] under a scenario where the data matrices have a natural partition, making it difficult to store them at a single location before applying dimensionality reduction.

Formally, if $\mathbf{W} \in \mathbb{R}^{\widetilde{N}\times m}$ and $\mathbf{Y} \in \mathbb{R}^{\widetilde{N}\times p}$, computing the product $\mathbf{W}^T\mathbf{Y}$ takes $O(mp\widetilde{N})$ time, which can be prohibitive for large $\widetilde{N}$. The sketched version then aims to find matrices $\mathbf{S} \in \mathbb{R}^{\widetilde{M}\times\widetilde{N}}$ such that

$$\left\|(\mathbf{SW})^T(\mathbf{SY}) - \mathbf{W}^T\mathbf{Y}\right\| \leqslant \epsilon \left\|\mathbf{W}\right\| \left\|\mathbf{Y}\right\|. \tag{3.1}$$

Computing the sketched matrix product $(\mathbf{SW})^T(\mathbf{SY})$ then takes only $O(mp\widetilde{M})$ time (not

accounting the time to compute $\mathbf{SW}$ and $\mathbf{SY}$ themselves). State-of-the-art bounds show that $\widetilde{M} = O(\max(\mathbf{sr}(\mathbf{W}), \mathbf{sr}(\mathbf{Y}))/\epsilon^2)$ suffices, where $\mathbf{sr}(\cdot)$ is the stable rank of a matrix (defined in subsection 3.2.1 and is a stable alternative for the rank). Similarly, given $\mathbf{A} \in \mathbb{R}^{\tilde{N} \times d}$ with $\widetilde{N} \gg d$ and $\mathbf{b} \in \mathbb{R}^{\tilde{N}}$, the ridge regression problem is

$$\mathbf{x}_* = \underset{\mathbf{x} \in \mathbb{R}^d}{\arg\min}\, f(\mathbf{x}) := \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \lambda \|\mathbf{x}\|^2 \tag{3.2}$$

and can be solved in $O(\widetilde{N}d^2)$ time. The sketched problem instead seeks to find matrices $\mathbf{S} \in \mathbb{R}^{\widetilde{M} \times \tilde{N}}$ such that solving

$$\hat{\mathbf{x}} = \underset{\mathbf{x} \in \mathbb{R}^d}{\arg\min}\, f_{\mathbf{S}}(\mathbf{x}) := \|\mathbf{S}\mathbf{A}\mathbf{x} - \mathbf{S}\mathbf{b}\|^2 + \lambda \|\mathbf{x}\|^2 \tag{3.3}$$

yields

$$f(\hat{\mathbf{x}}) \leqslant (1 + \epsilon) f(\mathbf{x}_*). \tag{3.4}$$

The state-of-the-art bounds show that for small ϵ, $\widetilde{M} = O(\mathbf{sd}_\lambda/\epsilon)$ suffices, where $\mathbf{sd}_\lambda$ is the statistical dimension and is again a more stable alternative to the rank of $\mathbf{A}$ (We define the statistical dimension in subsection 3.2.1).

With this background in place, let us consider a scenario where the data matrix $\mathbf{A}$ is naturally divided into J blocks that are not all available at a single location. Let each block then be of size $N \times d$, where $\tilde{N} = JN$. Such partitioning of data into different blocks occurs naturally in many applications. For example, dynamic systems produce data that evolve over time. To store the entire data before sketching it would require large amounts of memory [9]. It would be of use to sketch the system as it evolves, leading to a natural partition. In yet another application, consider the square kilometer array [11]. This array consists of antennas distributed across the continents of Australia and Africa. To handle the massive data rates (157 TB/s), it is desirable to sketch the data locally at each antenna and then transmit to the central processing location. In distributed systems that use edge-

cloud architecture, edge nodes collect data that needs to be communicated to the cloud for inference. The communication requirements can be made smaller if the data at each edge node is compressed to an "optimal" dimension.

A feature of existing sketching methods (including fast Johnson-Lindenstrauss matrices such as Subsampled Randomized Hadamard Transform (SRHT) [42] and sparse sketching matrices [43]) is that they need access to all or an arbitrary subset of the rows of $\mathbf{A}$ (See Figure 1.2). Clearly, this is unsuitable for an application where high dimensional data is available only in a decentralized manner, making it impossible to apply the standard dimensionality reducing techniques. This leads us to ask the following questions: Is there a way to adapt sketching techniques to such applications? What is the best way to model dimensionality reduction for such applications? Two naïve ways are readily available: i) Since each block is of size $N \times d$, its rank is upper bounded by d. One could obtain a subspace embedding for each block and communicate these sketched blocks to the central node. The resulting dimension of the aggregated data is then $O(Jd/\epsilon^2)$, since each block needs to be sketched to $O(d/\epsilon^2)$, ii) Sketch each data block separately, and **add** the resulting sketches at the central node instead of aggregating them. In fact, this results in a sketch of the entire data matrix $\mathbf{A}$. Using existing bounds, one can conclude that the final sketch needs to be $O(d/\epsilon^2)$, which again requires each data block also to be sketched to $O(d/\epsilon^2)$.

A major drawback of both of the above approaches is that they do not take advantage of the inherent low dimensionality of the entire matrix $\mathbf{A}$, resulting in a sketch size of $O(d/\epsilon^2)$ for each data block. Our observation is that it should be possible to lose information locally, while still retaining all the information about $\mathbf{A}$ globally. We show theoretically that it is possible for each of the blocks to be sketched to $O(d/J\epsilon^2)$. This implies that the sketch obtained from a single block may not be big enough to provide a subspace embedding for that block. Yet, an embedding of the entire matrix $\mathbf{A}$ can be obtained, once the sketches from the individual blocks are aggregated. Hence, our work aims to initiate a study of how to extend sketching methods to distributed data acquisition scenarios [44].

Our proposal is to impose a block diagonal structure on the sketching matrix $\mathbf{S}$. We denote such a sketching matrix as $\mathbf{S}_D$. We then partition the data matrices $\mathbf{W}$, $\mathbf{Y}$ and $\mathbf{A}$ analogously. This results in sketches of the form

$$\mathbf{S}_D\mathbf{A} = \begin{bmatrix} \mathbf{S}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{S}_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \mathbf{S}_J \end{bmatrix} \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \vdots \\ \mathbf{A}_J \end{bmatrix} = \begin{bmatrix} \mathbf{S}_1\mathbf{A}_1 \\ \mathbf{S}_2\mathbf{A}_2 \\ \vdots \\ \mathbf{S}_J\mathbf{A}_J \end{bmatrix}. \tag{3.5}$$

We assume that $\mathbf{A}_j \in \mathbb{R}^{N\times d}$ where $\widetilde{N} = JN$ and $\mathbf{S}_j \in \mathbb{R}^{M_j\times N}$ such that $\sum_j M_j = \widetilde{M}$, although our results extend to the case where the $\mathbf{A}_j$'s are of different sizes. We will also assume that the non-zero entries of the matrix $\mathbf{S}_D$ are drawn from the Gaussian distribution. Our goal is to derive upper bounds on the sample complexities $\widetilde{M}_j$ required to achieve similar guarantees as those in (Equation 3.1) and (Equation 3.4) when the sketching matrix $\mathbf{S}$ is of the form of $\mathbf{S}_D$ shown in (Equation 3.5).

Apart from the structural advantages described above, computing the product $\mathbf{S}_D\mathbf{A}$ can also be much cheaper when compared to an unstructured random projection. For generic $\mathbf{S}_j$, the sketch $\mathbf{S}_D\mathbf{A}$ can be computed in time $O(Nd\widetilde{M})$, as compared to the $O(\widetilde{N}d\widetilde{M})$ required for a dense, unstructured sketch. Second, the computation is trivial to parallelize into J blocks, each requiring $O(NdM_j)$ time. For large problems with low effective rank, when we can take $M_j = O(\log N)$, this gives us a sketch with structured randomness competitive with methods that use SRHT and sparse embedding matrices [38]. Furthermore, the blocks themselves could be designed to be fast transforms. Owing to these computational advantages, blocking could be a strategy by itself.

3.1.1 Related work

There is a vast and growing literature on sketching techniques. Here we briefly review some of the work most relevant to ours in the context of our setting. Note that while sketching can also be used as a pre-conditioning method [40], here we will only address "sketch

and solve" methods where the original problem is (approximately) solved in a reduced dimension.

Sketching methods for solving ordinary least squares problems are well summarized in [38]. However, as noted in [4], solutions for sketched ridge regression problems are more relevant in practice since regularization is often necessary. Similar to [4], we address this problem but in the setting where the sketching matrix is block diagonal. We provide conditions on the matrix $\begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix}$ under which such structured matrices can have the same sample complexity as [4].

Our work is closely related to that of [45] which studies the restricted isometry property (RIP) of block diagonal matrices. These results can be used to directly obtain subspace embedding guarantees for block diagonal matrices. However, this approach requires a sample complexity dependent on the rank of $\mathbf{A}$ and not its approximate rank. For large matrices with fast spectral decay, this dependency can lead to sub-optimal sample complexity. Another difference is that we consider block diagonal matrices that have different sized blocks, while [45] assumes that all the blocks are of the same size. One of the main conclusions of our paper is that choosing the block sizes in a data dependent fashion leads to improved (optimal) sample complexity.

A statistical analysis of sketched ridge regression in a distributed setting is provided in [39]. This work considers the ridge regression problem in the multivariate setting (where $\mathbf{b}$ and $\mathbf{x}$ are matrices) and analyzes model averaging in the case of distributed computation of the sketched ridge regression solution. In this setting, various processors each solve the problem with a part of the data and the estimators are then communicated to a central agent. In contrast, we consider a scenario where the estimate is computed by the central agent with only sketched data sent from various nodes.

Another work that is similar in spirit to ours and addresses sketched regression in a distributed setting is [46]. The setting considered in this work lies somewhere between that of [39] and ours. It considers multiple processors solving the ridge regression problem

with different parts of the data similar to [39], but also assumes that the data used by each processor is available to all other processors in a sketched form. In contrast, in our work, the sketched data from all the nodes is available to only a central computing agent.

A complimentary line of work focuses on the same problem but where $\widetilde{N} \ll d$. In [47], a sketching based algorithm is proposed that achieves a relative error guarantee for the solution vector. This result is further improved in [41]. Sketching has also been applied in the context of kernel ridge regression, where the data points are mapped to higher dimensional feature space before solving the regression problem. Sketching is used to reduce the number of such high dimensional features in [48] and [49]. Sampling and rescaling of features is considered in [48]. Random feature maps are also used to construct pre-conditioners in [49] to solve kernel ridge regression, where it is shown that a number of random feature maps proportional to the effective rank of the kernel matrix suffices to obtain a high quality pre-conditioner. While our work targets a different setting (where $\widetilde{N} \gg d$) and requires a different set of analytical tools, it is noteworthy that our guarantees involve a similar dependence on the stable rank of the underlying data matrix.

3.2 Localized sketching for regression and matrix multiplication

Our main contribution is theoretical analysis of the block model described in (Equation 3.5). A naïve strategy to analyze block diagonal matrices is to treat each block $\mathbf{A}_j$ separately and use a number of random projections proportional to its effective rank. But this would not take advantage of the low dimensional structure of the full matrix $\mathbf{A}$, resulting in a highly suboptimal sample complexity. Instead, we show that under mild assumptions on $\mathbf{A}$, the total sample complexity of $\widetilde{M}$ of the matrix $\mathbf{S}_D$ can match the existing bounds mentioned above.

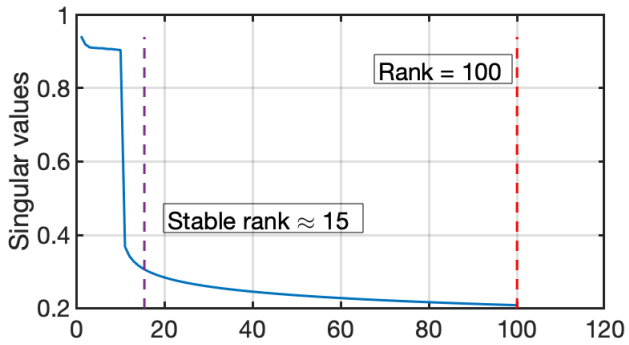


Figure 3.1: *The stable rank of a matrix captures the inherent low dimensionalty of approximately low-rank matrices. In this figure, we show the singular values of a* 100×100 *matrix that is technically full rank. However, the singular values decay, resulting in an effective rank of about 15.*

3.2.1 Stable rank, statistical dimension and incoherence

Before we can state our main results, we first define a few quantities that characterize the complexity of matrix multiplication and ridge regression problems.

Stable rank of a matrix:

The stable rank of a matrix $\mathbf{W}$ is defined as

$$\mathbf{sr}(\mathbf{W}) = \frac{\|\mathbf{W}\|_F^2}{\|\mathbf{W}\|^2}. \tag{3.6}$$

Note that $\mathbf{sr}(\mathbf{W}) \leqslant \text{rank}(\mathbf{W})$. For matrices with a flat spectrum, the stable rank equals the rank of the matrix. However, if the singular values decay, then the stable rank captures the effective low dimensionality of the matrix, even when it is technically full rank.

Statistical dimension of the ridge regression problem:

The ridge regression problem defined in (Equation 3.2) can be reformulated as

$$\min_{\mathbf{x}\in\mathbb{R}^d} \left\| \begin{bmatrix} \mathbf{A} \\ \sqrt{\lambda}\mathbf{I}_d \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \right\|^2 \Leftrightarrow \min_{\mathbf{x}\in\mathbb{R}^d} \left\| \widetilde{\mathbf{A}}\mathbf{x} - \widetilde{\mathbf{b}} \right\|^2.$$

The scalar multiple of the identity on the bottom of $\widetilde{\mathbf{A}}$ means it will technically be rank d. But in some sense, a more nuanced notion of rank would count dimensions in the column space of $\widetilde{\mathbf{A}}$ that have singular values greater than $\sqrt{\lambda}$ differently that those with singular values less than $\sqrt{\lambda}$. One way to bring about this distinction through the *statistical*

dimension

$$\mathbf{sd}_\lambda = \sum_i \frac{\sigma_i^2}{\sigma_i^2 + \lambda}.$$

In the sum above, if $\sigma_i^2 \gg \lambda$, then the contribution for that term is approximately one, while if $\sigma^2 \ll \lambda$, it is essentially zero. This allows us to interpret $\mathbf{sd}_\lambda$ as a kind of "effective rank". Note that $\mathbf{sd}_\lambda \leqslant \text{rank}(\mathbf{A})$ and can be much lower than $\text{rank}(\mathbf{A})$. While making λ very large can of course make $\mathbf{sd}_\lambda$ very small, this also introduces a larger bias in the estimates provided by (Equation 3.2) and (Equation 3.3), driving both of their solutions to zero. Choosing the λ that balances this bias-variance trade-off is equally important in sketched and non-sketched ridge regression.

Incoherence of the data matrices:

In randomized sampling schemes, the sampling probability of each row depends on the corresponding *leverage score*, which is the ℓ_2 norm of the corresponding row of an orthobasis $\mathbf{U}$ for $\mathbf{A}$. Leverage scores highlight the relative importance of each row of $\mathbf{A}$. Block diagonal matrices can be thought of as a generalization of sampling matrices. Instead of a single row, each block now accesses a submatrix of $\mathbf{A}$. Instead of using uniformly sized diagonal blocks $\mathbf{S}_j$, we show that a relative importance term associated with each block $\mathbf{A}_j$ similar to leverage scores dictates the number of random projections M_j required to attain optimal sample complexity. Let $\mathbf{U}$ be an orthobasis for the column space of the matrix $\mathbf{A}$. Let $\mathbf{U} = [\mathbf{U}_1^T \ \mathbf{U}_2^T \ \cdots \ \mathbf{U}_J^T]^T$, where $\mathbf{U}_j \in \mathbb{R}^{N \times d}$. We will show that the corresponding relative importance parameter, which we term as *coherence* of $\mathbf{U}_j$, is

$$\Gamma(\mathbf{U}_j) = \min\left(\|\mathbf{U}_j\|_\infty^2 N, \|\mathbf{U}_j\|_2^2\right).$$

Here, $\|\mathbf{U}_j\|_\infty$ denotes the element-wise infinity norm and $\|\mathbf{U}_j\|_2$ denotes the spectral norm. We can observe that

$$\frac{1}{J} \leqslant \max_j \Gamma(\mathbf{U}_j) \leqslant 1. \tag{3.7}$$

When the $\Gamma(\mathbf{U}_j)$'s are all close to $1/J$, the columns of $\mathbf{U}$ are incoherent, or not too aligned

with respect to the standard basis vectors. On the contrary, when they are close to 1, then there are vectors in the column space of $\mathbf{U}$ which are close (in an inner product sense) to the standard basis vectors. We describe bases $\mathbf{U}$ that have small coherence parameters as being *incoherent*. We will show that as long as the coherence is not too high, the sample complexity of block diagonal matrices can match that of generic sketching matrices.

3.2.2 Sample complexity bounds for localized sketching

Low values of the coherence parameter (highly incoherent bases) indicate relative uniformity in the importance of the blocks. For such subspaces, it would be reasonable to expect that roughly the same number of random projections can be drawn from each data block $\mathbf{A}_j$. On the other hand, when the coherence parameters $\Gamma(\mathbf{U}_j)$ have are uneven, it can be expected that the number of random projections from each block should be proportional to the corresponding $\Gamma(\mathbf{U}_j)$. This is precisely our proposed strategy to design the number of random projections M_j. We propose that M_j can be chosen as

$$M_j = M_0 \Gamma(\mathbf{U}_j) \tag{3.8}$$

for some constant M_0 that we will determine later. This results in a **total sample complexity** of $M_0 \sum_j \Gamma(\mathbf{U}_j)$. Our theoretical results state that block diagonal sketching matrices can achieve optimal sample complexity when M_j's are designed as in (Equation 3.8). This is also reminiscent of sampling algorithms, where the sampling probability of each row is proportional to the corresponding leverage score.

Localized sketching for matrix multiplication

Since matrices with high ranks can still be approximately low dimensional, we will characterize the sample complexity in terms of the stable rank of the multiplicands. When the sketching matrices are dense and contains entries drawn from the Gaussian random distribution, [5] shows that the sample complexity of $\mathbf{S}$ in (Equation 3.1) (under certain distributions) depends only on the stable ranks of the matrices. For optimal sample com-

plexity, the distributions $\mathcal{D}$ from which the sketching matrices $\mathbf{S}$ are drawn need to satisfy

$$\mathop{\mathbb{P}}_{\mathbf{S}\sim\mathcal{D}}\left(\left\|(\mathbf{SW})^T(\mathbf{SY})-\mathbf{W}^T\mathbf{Y}\right\|>\epsilon\left\|\mathbf{W}\right\|\left\|\mathbf{Y}\right\|(1+\mathbf{sr}(\mathbf{W})/k)^{1/2}(1+\mathbf{sr}(\mathbf{Y})/k)^{1/2}\right)<\delta \tag{3.9}$$

for any desired k and a suitable $\widetilde{M}$. When $\mathbf{S}$ is a dense matrix with sub-Gaussian entries, this holds for $\widetilde{M}=\Omega(\frac{k+\log(1/\delta)}{\epsilon^2})$. Then, for $k=\max(\mathbf{sr}(\mathbf{W}),\ \mathbf{sr}(\mathbf{Y}))$, $\mathbf{S}$ satisfies (Equation 3.1). Hence, to achieve a relative error in the spectral norm, $\mathbf{S}$ only needs to have a number of rows that is proportional to the stable ranks of $\mathbf{W}$ and $\mathbf{Y}$.

Our first main result is a similar guarantee for block diagonal sketching matrices. Unlike the distributions proposed in [5], block diagonal distributions cannot be both oblivious to the data matrices and have optimal sample complexity. A naïve way to achieve (Equation 3.9) when $\mathbf{S}$ is block diagonal is to use triangle inequality:

$$\left\|(\mathbf{S}_D\mathbf{W})^T(\mathbf{S}_D\mathbf{Y})-\mathbf{W}^T\mathbf{Y}\right\|\leqslant\sum_j\left\|(\mathbf{S}_j\mathbf{W}_j)^T(\mathbf{S}_j\mathbf{Y}_j)-\mathbf{W}_j^T\mathbf{Y}_j\right\| \tag{3.10}$$

where $\mathbf{W}_j$ and $\mathbf{Y}_j$ are corresponding blocks as in (Equation 3.5). However, this requires that $M_j=\Omega\left(\frac{\mathbf{sr}(\mathbf{W}_j)+\mathbf{sr}(\mathbf{Y}_j)}{\epsilon^2}\right)$ for each j. This can lead to suboptimal sample complexities, as $\mathbf{sr}(\mathbf{W}_j)$ and $\mathbf{sr}(\mathbf{Y}_j)$ can be as high as $\mathbf{sr}(\mathbf{W})$ and $\mathbf{sr}(\mathbf{Y})$ themselves. We show in our analysis that we can in fact achieve

$$\widetilde{M}=\sum_j M_j=\Omega\left(\frac{\mathbf{sr}(\mathbf{W})+\mathbf{sr}(\mathbf{Y})}{\epsilon^2}\right)$$

for incoherent matrices. With M_j designed as in (Equation 3.8), we have the following result for computing approximate matrix products:

Theorem 7 *Fix matrices* $\mathbf{W}$ *and* $\mathbf{Y}$ *and let* $\mathbf{S}_D$ *be a block diagonal matrix as in* (Equation 3.5) *with the entries of* $\mathbf{S}_j$ *are drawn from the distribution* $\mathcal{N}(0,1/M_j)$. *Let* $\mathbf{U}$ *be an orthobasis for the matrix* $[\mathbf{W}\ \mathbf{Y}]$ *and* $\Gamma(\mathbf{U}_j)$ *be the corresponding incoherence terms. Then the tail*

bound (Equation 3.9) *holds with* $\mathbf{S} = \mathbf{S}_D$ *when* M_j *are taken as in* (Equation 3.8) *with*

$$M_0 = \Omega\left(\frac{k\log(2/\delta)}{\epsilon^2}\right). \tag{3.11}$$

We can examine the total sample complexity of $\mathbf{S}_D$. Consider a highly incoherent basis $\mathbf{U}$: each entry of such a basis is bounded away from 1. Examples of such bases include orthobases of matrices with entries drawn from the Gaussian distribution and any subset of the Fourier basis. Since each column of $\mathbf{U}$ has an ℓ_2-norm of 1, for such bases, $\|\mathbf{U}_j\|_\infty \approx 1/\sqrt{\widetilde{N}}$. Then we have $M_j \approx \frac{M_0}{J}$ and $\widetilde{M} = \Omega\left(\frac{\max(\mathbf{sr}(\mathbf{W}),\mathbf{sr}(\mathbf{Y}))\log(2/\delta)}{\epsilon^2}\right)$. We see that even though $\mathbf{S}_D$ has a block diagonal structure, it can still have an optimal sample complexity.

Block diagonal sketching for ridge regression

Let us now consider the sketched ridge regression problem shown in (Equation 3.3). Let $\mathbf{U}_1 \in \mathbb{R}^{\widetilde{M}\times d}$ comprise the first n rows of an orthobasis for the matrix $\left[\begin{smallmatrix}\mathbf{A}\\ \sqrt{\lambda}\mathbf{I}_d\end{smallmatrix}\right]$. Then, (Equation 3.4) holds with constant probability, if $\mathbf{S}$ satisfies the following two conditions:

$$\left\|\mathbf{U}_1^T\mathbf{S}^T\mathbf{S}\mathbf{U}_1 - \mathbf{U}_1^T\mathbf{U}_1\right\| \leqslant \frac{1}{4}, \tag{3.12}$$

$$\left\|\mathbf{U}_1^T\mathbf{S}^T\mathbf{S}\mathbf{r}^* - \mathbf{U}_1^T\mathbf{r}^*\right\| \leqslant \sqrt{\frac{\epsilon f(\mathbf{x}^*)}{2}}, \tag{3.13}$$

where $\mathbf{r}_* = \mathbf{b} - \mathbf{A}\mathbf{x}^*$ and we recall that $f(\mathbf{x}^*) = \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|^2 + \lambda\|\mathbf{x}^*\|^2$. These conditions are well known in the randomized linear algebra community. (See [4] Lemma 9.) Both of the above conditions on $\mathbf{S}$ can be re-expressed as approximate matrix product guarantees by choosing the pair of matrices as $\mathbf{W} = \mathbf{Y} = \mathbf{U}_1$ for (Equation 3.12) and $\mathbf{W} = \mathbf{U}_1$ and $\mathbf{Y} = (\mathbf{b} - \mathbf{A}\mathbf{x}^*)$ for (Equation 3.13). We now state our main result for block diagonal sketching of ridge regression problems. Let $\mathbf{A}$ and $\mathbf{b}$ be as defined above and let $\mathbf{U}$ be an orthobasis for a basis for the range of $[\mathbf{A}\ \mathbf{b}]$ of size at most $\widetilde{N}\times(d+1)$ with $\Gamma(\mathbf{U}_j)$'s being the corresponding incoherence terms.

Theorem 8 *Let* $\mathbf{U}$ *be an orthobasis for the matrix* $[\mathbf{A}\ \mathbf{b}]$ *and* $\Gamma(\mathbf{U}_j)$ *be the corresponding incoherence terms. Let* $\mathbf{S}_D$ *be a block diagonal matrix as in* (Equation 3.5) *with the entries*

of $\mathbf{S}_j$ *are drawn from the distribution* $\mathcal{N}(0, 1/M_j)$. *Let* $\mathbf{x}_*$ *be the solution to* (Equation 3.2)*, and* $\hat{\mathbf{x}}$ *be the solution to* (Equation 3.3). *Then*

$$f(\hat{\mathbf{x}}) \leqslant (1+\epsilon) f(\mathbf{x}_*),$$

with constant probability when M_j *obeys* (Equation 3.8) *with* $M_0 = \Omega\left(\frac{\mathbf{sd}_\lambda}{\epsilon}\right)$.

As before, if $\mathbf{A}$ and $\mathbf{b}$ are such that the basis $\mathbf{U}$ is incoherent, then the total sample complexity $\widetilde{M} = \sum_j M_j = O(\frac{\mathbf{sd}_\lambda}{\epsilon})$. We are hence able to establish that though highly structured, block diagonal random matrices can in fact have optimal sample complexities.

Estimating the incoherence terms

An important question is about how the coherence parameters $\Gamma(\mathbf{U}_j)$'s can be estimated. Note that the main challenge is in computing an orthobasis for the data matrix $\mathbf{A}$. We develop an algorithm to empirically estimate the $\Gamma(\mathbf{U}_j)$'s to within a constant factor of the true values using a sketching based algorithm. The algorithm uses $O(d)$ fast localized random projections of the blocks $\mathbf{A}_j$'s and computes an estimate of the QR factorization of $\mathbf{A}$ at a central processing unit. Using the approximate R factor, the blocks $\mathbf{U}_j$'s are estimated locally. The algorithm has a worst case time complexity of $O(\widetilde{N} d \log N)$. Note that this is less than the sketch compute time $O(\widetilde{N} d \widetilde{M}/J)$ for N not too large. In Figure 3.3, we show the estimated incoherence parameters and the true parameters for a test matrix with $J = 100$, $\widetilde{N} = 10000$. We can see that the estimated values are within a constant factor of the true $\Gamma(\mathbf{U}_j)$'s. An important note here is that in many applications, an estimate of the $\Gamma(\mathbf{U}_j)$'s may be obtained using a priori domain knowledge. Yet another insight is that if distributional assumptions on the data can be made, as common in machine learning, then $\Gamma(\mathbf{U}_j)$'s can be very reliably estimated a priori [45]. Any such prior information will lead to better sample complexities as compared to the naïve techniques described in the introduction.

Algorithm for estimation of the incoherence parameters $\Gamma(\mathbf{U}_j)$

Our algorithm for estimating the block incoherence parameters is inspired by the algorithms

Algorithm 3 Estimation of incoherence parameters up to constant factor error

Input: Blocks $\mathbf{A}_j$.
Initialize $\Omega \in \mathbb{R}^{O(1)\times N}, \mathbf{Q} = 0, \mathbf{R} = 0$, $\hat{\mathbf{A}} = 0$ where Ω is drawn from any subsampled randomized FJLT.
While rank(R) not converged
Compute $\hat{\mathbf{A}}_j = \Omega\mathbf{A}$.
Aggregate $\hat{\mathbf{A}} = [\hat{\mathbf{A}}_1^\top\ \hat{\mathbf{A}}_2^\top \cdots \hat{\mathbf{A}}_J^\top]^\top$ at the central processing unit with previous estimate
Update $\mathbf{QR}$ = qr($\hat{\mathbf{A}}$)
Draw a new independent realization of Ω
Compute $\hat{\Gamma}(\mathbf{U}_j) = \|\mathbf{A}\mathbf{R}^{-1}\|_F^2$

Output: Normalized estimates $\hat{\Gamma}(\mathbf{U}_j)/\sum_j \hat{\Gamma}(\mathbf{U}_j)$

for leverage score estimation in the row sampling literature [50, 38] and from randomized SVD algorithms [51]. The main idea is the following: suppose we had access to the QR factorization of the data matrix $\mathbf{A} \in \widetilde{N} \times d$:

$$\mathbf{A} = \mathbf{QR}. \tag{3.14}$$

Then, an orthobasis can be obtained by computing $\mathbf{Q} = \mathbf{A}\mathbf{R}^{-1}$. However, computing the QR-factorization is as expensive as the matrix multiplication or ridge regression problems. We use a similar approach, but we only aim to capture the row space of $\mathbf{A}$ in a distributed fashion. However, we take random projections in an iterative fashion, until theg row space of the sketch "converges". We estimate the QR factorization from this resulting sketch. Our algorithm is described in Algorithm Algorithm 3. Note that we only aim to compute a constant factor approximation of the QR factors. Hence, computing the $\mathbf{R}$ takes, in the worst case, $O(JdN \log N) = O(\widetilde{N}d \log N)$ time. The QR factorization in each iteration can be updated from its previous estimates efficiently. Computing the final estimate takes about $O(Jd^3)$ time. Finally computing $\hat{\Gamma}(\mathbf{U}_j)$'s takes $O(\widetilde{N}d)$ time, resulting in a total worst case time complexity of $O(\widetilde{N}d \log N)$.

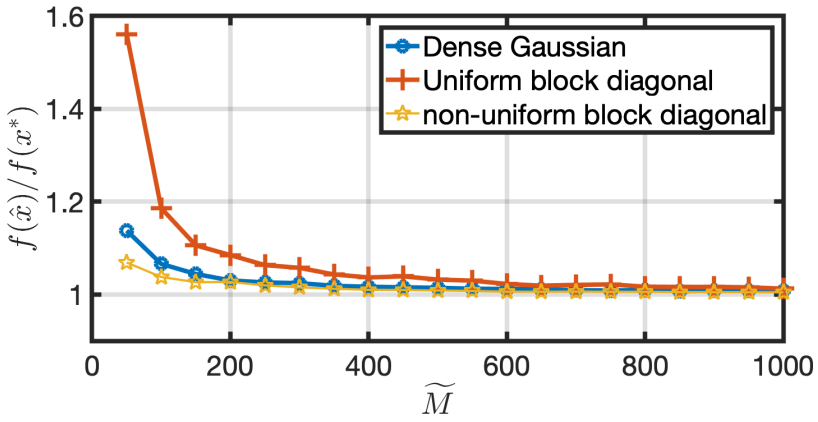


Figure 3.2: $f(\hat{x})/f(x^*)$ *for three sketching matrices: a dense matrix with standard Gaussian entries, a block diagonal matrix with equal sized blocks (uniform diagonal matrix) and a block diagonal matrix with entries designed as in* (Equation 3.8) *(non-uniform diagonal matrix). A ratio close to* 1 *indicates that the sketching matrix is effective in solving* (Equation 3.3)*. When* M_j*'s are chosen appropriately, block diagonal matrices can be as effective as a general matrix.*

3.3 Simulations

We demonstrate the effectiveness of block diagonal sketching matrices by performing experiments on both synthetic and real data. In our first experiment, we demonstrate the importance of choosing the size of the diagonal blocks according to our proposed method given in (Equation 3.8). We use the following parameters: $N = 2000$, $J = 10$, $d = 50$. We design the singular values such that for $\lambda = 0.15$, $\mathbf{sd}_\lambda = 8.5$, but $\text{rank}(\mathbf{A}) = 50$. For each trial, we generate $\mathbf{S}$ with entries drawn from $\mathcal{N}(0, 1/\sqrt{\widetilde{M}})$ and $\mathbf{S}_D$ with the entries of $\mathbf{S}_j$ drawn from $\mathcal{N}(0, 1/\sqrt{M_j})$. In Figure 3.2, we plot $f(\hat{x})/f(x^*)$ averaged over 10 trials for different values of $\widetilde{M}$. In particular, we show that when $M_j = M_0\Gamma(\mathbf{U}_j)$, $\mathbf{S}_D$ has the same rate of decay for $f(\hat{x})/f(x^*)$ as $\mathbf{S}$, and has a worse rate otherwise.

In our next set of experiments, we study performance in terms of prediction accuracy on the YearPredictionMSD dataset. It contains 89 audio features of a set of songs and the task is to predict their release year. The dataset has 463,715 training samples and 51,630 test samples. In this case, we use diagonal blocks of the same size. Across 10 independent realizations of $\mathbf{S}$ and $\mathbf{S}_D$, we compute the empirical probability of $f(\hat{x})/f(x^*) \leqslant (1+\epsilon)$ for

various values of ϵ and $\widetilde{M}$. We show phase transition plots in Figure 3.4 which demonstrate that block diagonal matrices are as effective as dense matrices in terms of accuracy, for the same sample complexity.

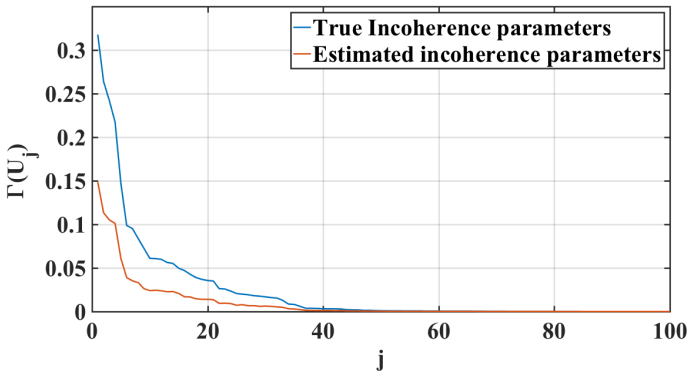


Figure 3.3: *For a test matrix with $J = 100$, $\widetilde{N} = 10000$, the true incoherence values and the estimated values are within a constant factor of each other, shown here in a sorted. Choosing the block sizes M_j proportional to the estimated coherence parameters results in optimal sample complexities.*

We also seek to highlight the computational advantages provided by block matrices. To this end, we compare the sketch compute times for block diagonal matrices with that of SRHT sketching matrices. We consider matrices $\mathbf{A}$ of sizes $2^{18} \times 40$, $2^{20} \times 40$ and $2^{22} \times 40$ and divide them into $J = 2^{10},\ 2^{12},\ 2^{14}$ blocks respectively. In order to ensure fair comparison, we replace the SRHT matrix with randomly subsampled Fast Fourier transform (FFT) matrix, since both have the same theoretical sketch compute time, but the FFT matrix has very efficient software implementations. The sketch compute times are shown in Table 3.1. Our choice of J renders each block small enough for very efficient computations. This results in block diagonal matrices being much faster compared to the FFT matrix.

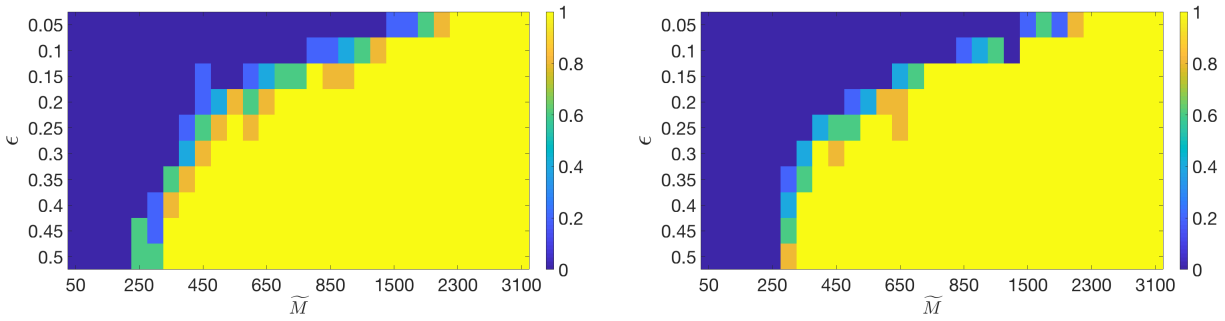


Figure 3.4: *Each plot shows the empirical probability of $f(\hat{x}) \leqslant (1+\epsilon)f(x^*)$ for various values of $\widetilde{M}$, computed using an average over 10 trials. The left pane is for results with dense matrices with sub-Gaussian entries, the right pane for results with block diagonal sketching matrices.*

Table 3.1: *Sketch compute time in sec. for various matrix sizes $\widetilde{N}$ and sketch sizes $\widetilde{M}$. In each cell, the left figure for FFT sketch and the right figure in boldface is for block diagonal matrices.*

Sketch compute time in seconds for large scale matrices				
$\widetilde{N},\ J$	$\widetilde{M} = 600$	$\widetilde{M} = 1400$	$\widetilde{M} = 2200$	$\widetilde{M} = 3000$
$2^{18}, 2^{10}$	$0.26; \mathbf{1.4 \times 10^{-2}}$	$0.26; \mathbf{2 \times 10^{-2}}$	$0.26; \mathbf{3.88 \cdot 10^{-2}}$	$0.26; \mathbf{4.2 \times 10^{-2}}$
$2^{20}, 2^{12}$	$1.16; \mathbf{2.7 \times 10^{-2}}$	$1.16; \mathbf{3.9 \times 10^{-2}}$	$1.16; \mathbf{5.1 \times 10^{-2}}$	$1.16; \mathbf{6.3 \times 10^{-2}}$
$2^{22}, 2^{14}$	$5.87; \mathbf{7.9 \times 10^{-2}}$	$5.87; \mathbf{9.1 \times 10^{-2}}$	$5.87; \mathbf{11 \times 10^{-2}}$	$5.86; \mathbf{11 \times 10^{-2}}$

CHAPTER 4

LOW RANK MATRIX RECOVERY FROM DECENTRALIZED SKETCHES

4.1 Identifying low dimensional subspaces from compressed measurements

A fundamental structural model for data is that the data points lie close to a subspace. In chapter 2, we saw two instances of subspace based models in the domain of array signal processing. For active array imaging, the low-dimensional subspace model helped in designing a more efficient signal acquisition method and for broadband direction-of-arrival estimation, the search over the candidate subspaces could be made sample-efficient due to the low dimensionality.

In both of the above applications, we assumed knowledge of the low-dimensional subspace model that described the observed data. However, identifying the subspace is by itself an interesting problem. When data points lie close to an unknown subspace, the matrix created by concatenating the data vectors has low rank. Identifying the subspace, or equivalently the principal components that best describe the data points is one of the most fundamental problems in signal processing, data science and machine learning, and is usually termed Principal Component Analysis (PCA). In this chapter, we develop a method to perform subspace identification and signal reconstruction from compressed observations when the underlying subspace if unknown.

While PCA is a common tool for dimensionality reduction, denoising, feature extraction and in many other signal processing and machine learning pipelines, we focus here on its usage a tool for linear dimensionality reduction. Consider a set of points in $\mathbb{R}^{d_1}$, drawn from an approximately r-dimensional subspace and let there be d_2 such data points. Then we could first compute the principal components, with a cost of $O(d_1^3)$, and then project data onto the first $r + O(1)$ principal components. While this offers a viable path for both

estimating the principal components and data compression, the $O(d_1^3)$ cost can easily be prohibitive. In this chapter, we will develop a method to compress and recover data points belonging to an unknown low-dimensional subspace from signal-agnostic, random projections of the data.

We cast this problem as a low-rank matrix recovery problem where we wish to recover a set of vectors from a low-dimensional subspace after they have been individually compressed (or "sketched"). More concretely, let $\mathbf{x}_1, \cdots, \mathbf{x}_{d_2}$ be vectors from an unknown r-dimensional subspace in $\mathbb{R}^{d_1}$. We observe the vectors indirectly via linear sketches by corresponding sensing matrices $\mathbf{B}_1, \ldots, \mathbf{B}_{d_2} \in \mathbb{R}^{d_1 \times L}$, where $L \ll d_1$, i.e., the observed measurement vectors are written as

$$\mathbf{y}_i = \mathbf{B}_i^\top \mathbf{x}_i + \mathbf{z}_i, \quad i = 1, \ldots, d_2. \tag{4.1}$$

Our goal is to leverage the low-dimensionality of the ensemble of points and recover them from a number of 'per-vector' measurements $L \ll d_1$. Although individual recovery of each vector is ill-posed, it is still possible to recover $\mathbf{x}_1, \ldots, \mathbf{x}_{d_2}$ jointly by leveraging their mutual structure without knowing the underlying subspace a priori. This indeed results in a low-rank matrix recovery problem with a column-wise observation model.

We are motivated mainly by large-scale inference problems where data is collected in a distributed network or in a streaming setting. In both cases, it is desired to compress the data to lower the communication overhead. In the first scenario, the data is partitioned according to the network structure and each data point must be compressed without accessing the remainders. In the second scenario, memory or computational constraints may limit access to relatively small number of recent data points.

Such compressive and distributive acquisition schemes arise frequently in numerous real-world applications. In next generation high-resolution astronomical imaging systems, an antenna array may be distributed across a wide geographical area to collect data points that have a high dimension but are also heavily correlated (and hence belong to a low-

dimensional subspace). Compression at the node level relieves the overhead to transmit data to a central processing unit [1]. In scientific computing, it is common to generate large scale simulation data that has redundancies that manifest as low-rank structures. For example, simulations in a fluid dynamic system generate large state vectors that have low-rank dynamics [52]. Our observational model describes a kind of on-the-fly compression, where the states are compressed as the system evolves, resulting in efficient communication and storage.

In each of these applications, if the underlying low-dimensional subspace were known a priori, then the projection onto that subspace could have implemented an optimal distortion-free linear compression. Alternatively if the uncompressed data were available, the standard Principal Component Analysis (PCA) might have been used to discover the subspace. Unfortunately, neither is the case. Therefore we approach the recovery as sketching without knowing the latent subspace a priori. It can also be interpreted as a blind compressed sensing problem that recovers the data points and underlying subspace simultaneously from compressed measurements.

The measurement model in (Equation 4.1) is equivalently rewritten as follows: Let $\mathbf{X}_0 \in \mathbb{R}^{d_1 \times d_2}$ be a matrix obtained by concatenating $\mathbf{x}_1, \ldots, \mathbf{x}_{d_1}$. It follows that the rank of $\mathbf{X}_0$ is at most r. The entries of $\mathbf{y}_1, \ldots, \mathbf{y}_{d_2}$ then correspond to noisy linear measurements of $\mathbf{X}_0$, i.e., for $l = 1, \ldots, L$ and $i = 1, \ldots, d_2$, the l^{th} entry of $\mathbf{y}_i$ denoted by $y_{l,i}$ is written as

$$y_{l,i} = \langle \mathbf{A}_{l,i}, \mathbf{X}_0 \rangle + z_{l,i} \quad \text{with} \quad \mathbf{A}_{l,i} = \frac{1}{\sqrt{L}} \mathbf{b}_{l,i} \mathbf{e}_i^\top, \tag{4.2}$$

where $z_{l,i}$, $\mathbf{b}_{l,i}$, and $\mathbf{e}_i$ respectively denote the l^{th} entry of $\mathbf{z}_i$, the l^{th} column of $\mathbf{B}_i$, and the i^{th} column of the identity matrix of size d_2. We propose a convex optimization method to recover $\mathbf{X}_0$ from $\{y_{l,i}\}$ and provide theoretical analysis when $\mathbf{b}_{l,i}$ and $z_{l,i}$ are independent copies of random vectors drawn according to $\mathcal{N}(0, I_{d_1})$ and $\mathcal{N}(0, \sigma^2)$ respectively.

4.2 Tensor products of Banach spaces

4.2.1 Introduction

Recasting the problem of compressing data points drawn from a subspace into a low rank matrix recovery problem opens up many possibilities for designing recovery algorithms and perform theoretical analysis to determine the requires sample complexity L. A popular framework to recover low rank matrices from certain types of linear observations is to use the nuclear norm based recovery algorithms [53, 54]. In this framework, the nuclear norm of a matrix is used as a convex surrogate for the rank of a matrix, thus opening up convex optimization methods as tools for low rank matrix recovery.

While nuclear norm minimization has seen a lot of success, there exist other characterizations of the set of low rank matrices. For example, alternate convex surrogates for low rank matrix recovery have been proposed in the more recent literature [55, 56]. In particular, [57] proposes to use the "max-norm" of a matrix as a convex relaxation to solve the matrix completion problem, whereas [56] used the "mixed-norm" of a matrix to solve for the column-wise observation model described in (Equation 4.2).

All of the above methods can be seen as particular instances of a unifying framework to design recovery algorithms for low rank matrices. We will develop this framework in this section and then focus on the column-wise observation model of (Equation 4.2) in the remainder of the chapter. The unifying framework can be developed by describing matrices as operators between Banach spaces endowed with suitable norms. These norms can then be used to ascribe the matrix itself with certain norms to capture the low-rankness, and further be used to design recovery algorithms for low rank matrices.

Let $\mathbf{X} \in \mathbb{R}^{d_1 \times d_2}$ be a matrix. We can then interpret it as a linear operator from a vector space $\mathbb{R}^{d_2}$ to another vector space $\mathbb{R}^{d_1}$. Then let the domain and range spaces be respectively endowed with the ℓ_p norm and the ℓ_q norm. The vector space of all $d_1 \times d_2$ matrices can then be identified as the tensor product of the two Banach spaces $P =$

$(\mathbb{R}^{d_2}, \|\cdot\|_p)$ and $Q = (\mathbb{R}^{d_1}, \|\cdot\|_q)$ and is denoted as $\ell_{p'} \otimes \ell_q$ (e.g., [58]), where $1/p + 1/p' = 1$. More generally, the set of linear operators from a Banach space S to another space T can be denoted as $S^* \otimes T$, where S^* is the Banach space dual of S.

4.2.2 Tensor product norms

For any given pair of Banach spaces, we can define many norms on their tensor product. [58, 59]. We consider two specific norms: the *injective* and *projective* tensor norms, defined respectively as

$$\|\mathbf{X}\|_\vee = \sup_{\mathbf{u}\in\mathbb{R}^{d_1}, \|\mathbf{u}\|_p=1} \|\mathbf{X}\mathbf{u}\|_q \tag{4.3}$$

and

$$\|\mathbf{X}\|_\wedge = \inf \left\{ \sum_k \|\mathbf{u}_k\|_{p'} \|\mathbf{v}_k\|_q \;\middle|\; \mathbf{X} = \sum_k \mathbf{v}_k \mathbf{u}_k^* \right\}. \tag{4.4}$$

The injective (resp. projective norm) norm is the smallest (resp. largest) norm among the class of all tensor norms. The main insight driving the unified perspective on low rank matrix recovery is that, when we restrict linear operators to those of rank at most r, certain tensor norms become equivalent up to a function of r. This property can then be used to design convex relaxations for the set of rank-r matrices. For example, when $p = q = 2$, it can be easily verified that $\|\mathbf{X}\|_\vee = \|\mathbf{X}\|_2$ and $\|\mathbf{X}\|_\wedge = \|\mathbf{X}\|_*$. It follows from the singular value decomposition that $\|\mathbf{X}\|_2 \leqslant \|\mathbf{X}\|_* \leqslant r\|\mathbf{X}\|_2$. In yet another example, let $p = 1, q = \infty$. Then it can be shown that the injective norm $\|\mathbf{X}\|_\vee = \|\mathbf{X}\|_\infty$, which is the element-wise infinity norm. Further, define the max-norm of a matrix as

$$\|\mathbf{X}\|_{\max} = \inf_{\mathbf{U},\mathbf{V}:\mathbf{U}\mathbf{V}^\top = X} \left\|\mathbf{U}^\top\right\|_{1\to 2} \left\|\mathbf{V}^\top\right\|_{1\to 2}. \tag{4.5}$$

Then, it can be shown that [55]

$$\|\mathbf{X}\|_\infty \leqslant \|\mathbf{X}\|_{\max} \leqslant \sqrt{r}\,\|\mathbf{X}\|_\infty. \tag{4.6}$$

In this case, it has been shown that the max norm is equivalent up to a constant to the projective norm. For $p = 2, q = 2$ ($p = 1, q = \infty$), the sets $\{\mathbf{X} : \|\mathbf{X}\|_\vee \leqslant \alpha, \|\mathbf{X}\|_\wedge \leqslant r\alpha\}$

($\{\mathbf{X} : \|\mathbf{X}\|_\vee \leqslant \alpha, \|\mathbf{X}\|_\wedge \leqslant \sqrt{r}\alpha\}$) are convex relaxations of the set of rank-r matrices. Thus, the convex relaxations proposed in [53, 55, 56] are all particular instances of interpreting matrices as members of the tensor product of of Banach spaces, with each framework adopting particular choices of the Banach spaces.

4.2.3 Tailoring convex regularizations for low rank matrices

With the unifying framework described above in place, a natural question is how to choose the pair of Banach spaces to design a convex relaxation for the set of low rank matrices. This choice is somewhat clear when the rank-r matrix we are trying to recover has a known structured factorization. For example, if $\mathbf{X} = \mathbf{U}\mathbf{V}^\top$ where $\mathbf{U} \in \mathbb{R}^{d_1 \times r}$, $\mathbf{V} \in \mathbb{R}^{r \times d_2}$ and are both sparse, then an intuitive choice is to use the tensor product $\ell_\infty \otimes \ell_1$, since the ℓ_1 norm is known to be sparsity promoting. An extensive numerical study on this subject can be found in [60], where it is demonstrated that knowledge of structured factors can indeed be successfully used in choosing the pair of Banach spaces.

Another potential determining factor in the choice of the Banach spaces is the observation model itself. Consider the matrix completion problem [54, 61], where the goal is to complete a low rank matrix using only noisy observations of a subset of the entries. It is now well understood that matrix completion can be achieved from an optimal number of samples only under certain incoherence conditions that ensure the none of the entries are relatively very large. Constraining the $\|\cdot\|_\infty$ and max-norm is then the natural choice to impose low rankness along with the uniform bound on the entries. It is precisely based on this property that the authors in [55] are able to use the set $\{\mathbf{X} : \|\mathbf{X}\|_\infty \leqslant \alpha,\ \|\mathbf{X}\|_{\max} \leqslant R\}$ to effectively constrain the optimization for low rank matrix completion.

Considering the column-wise compression model defined in (Equation 4.2), we can again use a similar argument to design a novel convex relaxation meant for this particular observation model. Since (Equation 4.2) obtains inner products of individual columns, an equivalent notion of incoherence might be that none of the columns have a relatively

large norm compared to the other columns. This indicates that for this observation model, constraining the maximum column ℓ_2 norm, denoted as $\mathbf{X}_{1\to 2}$ might be the natural choice.

The above discussion forms the basis of the rest of this chapter. We choose $p = 1,\ q = 2$, resulting in $\|\mathbf{X}\|_{\vee} = \|\mathbf{X}\|_{1\to 2}$. Along with this, we define the "mixed-norm" of a matrix, which shares an interlacing property with $\|\cdot\|_{1\to 2}$ characterized by the rank of the matrix. The mixed-norm is equivalent up to a constant to the projective norm of $\ell_\infty \otimes \ell_2$. Further, it is interesting that unlike many tensor norms, the mixed norm can be computed efficiently in a polynomial time, similar to the nuclear norm.This enables efficient implementation of mixed-norm-based low-rank recovery programs. Using the pair $\{\|\mathbf{X}\|_{1\to 2}, \|\mathbf{X}\|_{\text{mixed}}\}$, we design a new convex relaxation method for low rank matrix recovery. We provide upper bounds on the error between the estimate so obtained, and also provide a minimax error bound that shows that the error rate we obtain is tight (up to log factors). Finally, we provide an algorithm to implement our method and also show a strong empirical performance over nuclear norm for the observation model in (Equation 4.2).

4.3 Mixed-norm-based low-rank recovery

Owing to our intuition developed in the previous section, we will now develop the "mixed-norm" based framework for low rank matrix recovery. Our approach is a convex relaxation based approach that is better suited than those found in the literature for the particular observation model described in (Equation 4.2). We propose to use a convex relaxation for the set of low rank matrices that uses two specific matrix norms, the maximum column ℓ_2 norm and the mixed-norm, both of which we define below. For a matrix $\mathbf{X} \in \mathbb{R}^{d_1 \times d_2}$, the maximum ℓ_2 column norm is defined as

$$\|\mathbf{X}\|_{1\to 2} = \max_{j=1\cdots d_2} \|\mathbf{X}\mathbf{e}_j\|_2 , \tag{4.7}$$

where $\mathbf{e}_j$ is the j^{th} standard basis vector. This can be interpreted as the operator norm from the vector space $\ell_1^{d_2}$ to that of $\ell_2^{d_1}$. We define the following matrix norm as the mixed-norm:

$$\|\mathbf{X}\|_{\text{mixed}} = \inf_{\mathbf{U},\mathbf{V}:\mathbf{U}\mathbf{V}^\top = X} \|\mathbf{U}\|_{\text{F}} \|\mathbf{V}^\top\|_{1\to 2}. \tag{4.8}$$

We use the term mixed-norm, since the above definitions is a hybrid of the definitions of the nuclear norm and the max-norm, with nuclear norm using $\|\|_2$ on both the factors, and the max-norm using $\|\|_{1\to 2}$ on both the factors. The two norms described above provide a convex relaxation of the set of rank-r matrices because of the interlacing property given in the following lemma.

Lemma 2 *Let* $\mathbf{X} \in \mathbb{R}^{d_1 \times d_2}$ *satisfy* $\text{rank}(\mathbf{X}) \leqslant r$. *Then*

$$\|\mathbf{X}\|_{1\to 2} \leqslant \|\mathbf{X}\|_{\text{mixed}} \leqslant \sqrt{r}\, \|\mathbf{X}\|_{1\to 2}\,. \tag{4.9}$$

For matrices whose column norms are uniformly bounded by a constant α, their mixed norm is bounded by $\sqrt{r}\alpha$. For $R \geqslant \sqrt{r}\alpha$. By Lemma 2, the set $\kappa(\alpha, R)$ defined by

$$\kappa(\alpha, R) = \{\mathbf{X}\colon\ \|\mathbf{X}\|_{1\to 2} \leqslant \alpha,\ \|\mathbf{X}\|_{\text{mixed}} \leqslant R\} \tag{4.10}$$

contains the set of rank-r matrices with column norms bounded by α. Our proposed method to estimate the ground truth matrix $\mathbf{X}_0$ is given by the following convex optimization program:

$$\hat{\mathbf{X}} \in \operatorname*{argmin}_{\mathbf{X}\in\kappa(\alpha,R)} \sum_{l,i} |y_{l,i} - \langle \mathbf{A}_{l,i}, \mathbf{X}\rangle|^2. \tag{4.11}$$

We have attempted to use the nuclear norm instead of the mixed norm but it was not successful with providing a guarantee at a near optimal sample complexity. Furthermore it also demonstrates worse empirical performance compared to our approach, as we show soon.

Another appealing property of the mixed-norm is that it can be computed in polynomial time using a semidefinite formulation. This renders our proposed estimator readily implementable using general purpose convex solvers. However, to address scalability, we propose an ADMM based framework. We defer further details on efficient computation to section 4.8.

4.4 Related work

The model in (Equation 4.2) has been studied in the context of compressed principal components estimation [62, 63, 64]. These works studied a specific method that computes the underlying subspace though an empirical covariance estimation. While being guaranteed at a near optimal sample complexity, this approach is inherently limited to the linear observation model. On the other hand, our method is more flexible in terms of its potential extension to nonlinear observation models.

Negahban and Wainwright [65] considered the multivariate linear regression problem where a similar model to (Equation 4.2) arises but with a fixed sensing matrix $\mathbf{A}$, i.e., $\mathbf{A}_i = \mathbf{A}$ for all $i = 1, \ldots, d_2$. They showed that a nuclear-norm penalized least squares provides robust recovery at a near optimal sample complexity within a logarithmic factor of the degrees of freedom of rank-r matrices. However, their guarantees applies to an arbitrary fixed ground truth matrix and not to all matrices within the model simultaneously. Our aim is to work with an embedding of the model set $\kappa(\alpha, R)$ and we obtain a *uniform* theoretical guarantee over the entire model set at the cost of using different sensing matrices $\mathbf{A}_i$'s and incoherence of the matrices.

Our solution approach is partly inspired by earlier works on low-rank matrix completion using the max-norm [66, 67, 55]. The pair of max-norm and ℓ_∞ norms is used to relax the set of low-rank matrices to a convex model. We generalize this approach to that of using tensor norms as a proxy for low rank regularization and show that the max-norm and the mixed-norm are particular instances of this general framework. In particular we choose a specific pair of tensor norms in accordance with the structure in the observation model. This leads to a new convex relaxation model of low-rankness, a corresponding optimization formulation, algorithm, and its performance guarantee. Finally, we point out that our method of proofs and the technical tools we use to establish our results are significantly

different from that of [55].

4.5 Recovery guarantees on the mixed-norm based estimator

4.5.1 Upper bound on sample complexity

Our main result, stated in Theorem 9, provides an upper bound on the Frobenius norm of the error between the estimate $\hat{\mathbf{X}}$ and the ground truth matrix $\mathbf{X}_0$. This guarantee holds simultaneously for all matrices $\mathbf{X} \in \kappa(\alpha, R)$ rather than for just a fixed arbitrary matrix $\mathbf{X}_0$. En route to proving our guarantee, we indeed show that $\sum_{l,i}\langle \mathbf{A}_{l,i}, \mathbf{X}\rangle^2$ is well concentrated around its expectation $\|\mathbf{X}\|_{\mathrm{F}}^2$ for all $\mathbf{X} \in \kappa(\alpha, R)$ and hence, the measurements result in an embedding of the set $\kappa(\alpha, R)$ into a low dimension.

Theorem 9 *Let $\kappa(\alpha, R)$ be defined as in* (Equation 4.10). *Suppose that the $\mathbf{b}_{l,i}$ are drawn independently from $\mathcal{N}(\mathbf{0}, \mathbf{I}_{d_1})$, $(z_{i,l})$ are i.i.d. following $\mathcal{N}(0, \sigma^2)$, $d = d_1 + d_2$ and $d_2 \leqslant Ld_2 \leqslant d_1 d_2$. Then, for $R \leqslant \alpha\sqrt{r}$, there exist numerical constants c_1, c_2 such that the estimate $\widehat{\mathbf{X}}$ satisfies*

$$\frac{\|\widehat{\mathbf{X}} - \mathbf{X}_0\|_{\mathrm{F}}^2}{\|\mathbf{X}_0\|_{\mathrm{F}}^2} \leqslant c_1 \cdot \frac{\alpha^2}{\|\mathbf{X}_0\|_{\mathrm{F}}^2 / d_2} \cdot \max\left(1, \frac{\sigma\sqrt{L}}{\alpha}\right) \cdot \sqrt{\frac{r(d_1 + d_2)\log^6 d}{Ld_2}} \tag{4.12}$$

with probability at least $1 - 2\exp(-c_2 R^2 d/\alpha^2)$ for all $\mathbf{X}_0 \in \kappa(\alpha, R)$.

There are a few remarks in order:

- The factor $\alpha^2 d_2 / \|\mathbf{X}_0\|_F^2$ is the ratio between the maximum and the average of the squared column ℓ_2 norm of the ground truth matrix $\mathbf{X}_0$ and represents its degree of *incoherence*. A ratio close to 1 indicates that the columns have similar ℓ_2-norms and results in a lower sample complexity than when the ratio is much larger than 1. This is similar to the dependence on the relative magnitude of each entry in the max-norm-based estimator [55] and the dependence on incoherence in matrix completion problems.

- The second factor is written as $\max(1, \eta)$ where $\eta = \frac{\sigma\sqrt{L}}{\alpha}$ accounts for the noise level in the measurements. Since we take L measurements per column and the measurement operator is isotropic, α^2 is compared against the corresponding noise-variance $\sigma^2 L$.

- If the incoherence term is upper-bounded by a constant and the normalized noise level η satisfies $\eta = \Omega(1)$, then $\widehat{\mathbf{X}}$ obtained from $O(\eta^2 rd \log^6(d)\epsilon^{-2})$ measurements satisfies $\|\widehat{\mathbf{X}} - \mathbf{X}_0\|_{\mathrm{F}}^2 \leqslant \epsilon \|\mathbf{X}\|_{\mathrm{F}}^2$ with high probability.

4.5.2 Proof sketch

We state the key lemmas involved in proving Theorem 9 and point to the tools we use and defer finer details of the proof to the appendix . We begin with the basic optimality condition that relates the estimate $\hat{\mathbf{X}}$ to the ground truth $\mathbf{X}_0$. Let $\mathbf{M} = \hat{\mathbf{X}} - \mathbf{X}_0$. By the triangular inequality, we have $\mathbf{M} \in \kappa(2\alpha, 2R)$. For notational brevity, we assume from now on that $\mathbf{M} \in \kappa(\alpha, R)$. (Neither the main result nor the proofs are affected by this since they involve multiplication with some numerical constants.)

We adapt the first step in the analysis framework of the analogous matrix completion problem [55]. By optimality of the solution and (Equation 4.2), we have

$$\sum_{l,i} \left(y_{l,i} - \langle \mathbf{A}_{l,i}, \hat{\mathbf{X}} \rangle \right)^2 \leqslant \sum_{l,i} \left(y_{l,i} - \langle \mathbf{A}_{l,i}, \mathbf{X}_0 \rangle \right)^2 . \tag{4.13}$$

After substituting $\hat{\mathbf{X}} - \mathbf{X}_0$ by $\mathbf{M}$ and rearranging the terms, we obtain

$$\sum_{l,i} \langle \mathbf{A}_{l,i}, \mathbf{M} \rangle^2 \leqslant 2 \sum_{l,i} \langle \mathbf{A}_{l,i}, \mathbf{M} \rangle z_{l,i}. \tag{4.14}$$

As in [55], we rely on the stochastic nature of the noise. Our strategy is to obtain a lower bound on the quadratic form $\sum_{l,i} \langle \mathbf{A}_{l,i}, \mathbf{M} \rangle^2$ in terms of $\|\mathbf{M}\|_{\mathrm{F}}^2$ and a uniform upper bound on the linear form $\sum_{l,i} \langle \mathbf{A}_{l,i}, \mathbf{M} \rangle z_{l,i}$ over the set $\kappa(\alpha, R)$. We can then bound $\|\mathbf{M}\|_{\mathrm{F}}^2$ uniformly over the set.

Lower bound on the quadratic form

We observe that $\sum_{l,i}\langle \mathbf{A}_{l,i}, \mathbf{M}\rangle^2$ can be reformulated as a quadratic form in standard Gaussian random variables. Let us define

$$\xi = \begin{bmatrix} \mathbf{b}_{1,1} \\ \vdots \\ \mathbf{b}_{L,d_2} \end{bmatrix} \in \mathbb{R}^{Ld_1d_2}. \tag{4.15}$$

Then it follows that $\xi \sim \mathcal{N}(0, I_{Ld_1d_2})$. Therefore, the left-hand side of (Equation A.46) is rewritten as

$$\sum_{l,i}\langle \mathbf{A}_{l,i}, \mathbf{M}\rangle^2 = \left\|\mathbf{Q}_{\mathbf{M}}\xi\right\|^2, \quad \text{where}$$

$$\mathbf{Q}_{\mathbf{M}} = \frac{1}{\sqrt{L}} \begin{bmatrix} \widetilde{\mathbf{M}}_1^\top & 0 & \cdots & 0 \\ 0 & \widetilde{\mathbf{M}}_2^\top & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \widetilde{\mathbf{M}}_{d_2}^\top \end{bmatrix}, \quad \widetilde{\mathbf{M}}_j^\top = \mathbf{I}_L \otimes (\mathbf{M}\mathbf{e}_j)^\top \in \mathbb{R}^{L\times Ld_1}. \tag{4.16}$$

We also have

$$\mathrm{E}\left\|\mathbf{Q}_{\mathbf{M}}\xi\right\|^2 = \|\mathbf{M}\|_{\mathrm{F}}^2$$

We compute a tail estimate on $\sup_{\mathbf{M}\in\kappa(\alpha,R)}\left\|\mathbf{Q}_{\mathbf{M}}\xi\right\|_2^2$ by using the results on suprema of chaos processes [68]. The main result in [68] provides a sharp tail estimate on the supremum of a Gaussian quadratic form maximized over a given set $\mathcal{A}$, which is written as

$$\sup_{\mathbf{A}\in\mathcal{A}} \left|\left\|\mathbf{A}\xi\right\|^2 - \mathrm{E}\left\|\mathbf{A}\xi\right\|^2\right|,$$

by using a chaining argument. By adapting their framework, we obtain the following result.

Lemma 3 *Under the assumptions of Theorem 3, if* $\mathbf{Q}_{\mathbf{M}}$ *and* ξ *are as defined in* (Equation A.61) *and* (Equation A.60)*, then*

$$\sup_{\mathbf{M}\in\kappa(\alpha,R)} \left|\frac{\left\|\mathbf{Q}_{\mathbf{M}}\xi\right\|^2}{d_2} - \frac{\|\mathbf{M}\|_{\mathrm{F}}^2}{d_2}\right| \leqslant cR\sqrt{\frac{d}{Ld_2}}\left(\alpha + \frac{R\sqrt{d}}{\sqrt{Ld_2}}\right)\log^3(d)$$

with probability at least $1 - 2\exp(-cR^2d\alpha^2)$.

From Lemma 3, in the regime where $Ld_2 > R^2 d/\alpha^2$, we can obtain

$$\frac{\sum_{l,i}\langle \mathbf{A}_{l,i}, \mathbf{M}\rangle^2}{d_2} \geqslant \frac{\|\mathbf{M}\|_F^2}{d_2} - cR\alpha\sqrt{\frac{d}{Ld_2}}\log^3 d. \tag{4.17}$$

Upper bound on the right-hand side of (Equation A.46)

We obtain the following uniform upper bound $\sum_{l,i}\langle \mathbf{A}_{l,i}, \mathbf{M}\rangle z_{l,i}$:

Lemma 4 *Under the assumptions of Theorem 3, with probability at least* $1-2\exp(-cR^2d/\alpha^2)$,

$$\sup_{\mathbf{M}\in\kappa(\alpha,R)} \frac{\sum_{l,i}\langle \mathbf{A}_{l,i}, \mathbf{M}\rangle z_{l,i}}{d_2} \leqslant c(\sigma\sqrt{L})R\sqrt{\frac{d}{Ld_2}}\log^3 d. \tag{4.18}$$

To derive Lemma 4, we first express the left-hand side of (Equation 4.18) using a matrix norm. Define

$$|||\mathbf{M}||| := \frac{\|\mathbf{M}\|_{1\to 2}}{\alpha} \vee \frac{\|\mathbf{M}\|_{\text{mixed}}}{R}.$$

Then by the definition of $\kappa(\alpha, R)$ in (Equation 4.10) it follows that the unit ball $B := \{\mathbf{M} : |||\mathbf{M}|||| \leqslant 1\}$ with respect to $|||\cdot|||$ coincides with $\kappa(\alpha, R)$. Therefore, we obtain

$$\sup_{\mathbf{M}\in\kappa(\alpha,R)} \sum_{l,i}\langle \mathbf{A}_{l,i}, \mathbf{M}\rangle z_{l,i} = \sup_{\mathbf{M}\in\kappa(\alpha,R)} \langle \sum_{l,i} z_{l,i}\mathbf{A}_{l,i}, \mathbf{M}\rangle = |||\sum_{l,i} z_{l,i}\mathbf{A}_{l,i}|||_*$$

where $|||\cdot|||_*$ denotes the dual norm. Then, conditioned on $\mathbf{A}_{l,i}$'s, it follows from Theorem 4.7 in [69] that with probability $1-\delta$

$$|||\sum_{l,i} z_{l,i}\mathbf{A}_{l,i}|||_* \leqslant \underbrace{\mathrm{E}_z\,|||\sum_{l,i} z_{l,i}\mathbf{A}_{l,i}|||_*}_{T_1} + \underbrace{\pi\sqrt{\frac{\log(2/\delta)}{2}\sup_{M\in\kappa(\alpha,R)}\sum_{l,i}\langle \mathbf{A}_{l,i}, \mathbf{M}\rangle^2}}_{T_2}. \tag{4.19}$$

The first term T_1 is the *Gaussian complexity* of the sample set $\{A_{l,i}\}$ over the function class $\{\langle M, \cdot\rangle : M \in \kappa(\alpha, R)\}$. This can be (up to a logarithmic factor of the size of the summation) upper-bounded by the corresponding Rademacher complexity ([70], Equation (4.9)) as

$$T_1 \leqslant c\sigma\sqrt{\log(Ld_2+1)}\,\mathrm{E}_{(r_{l,i})}\,|||\sum_{l,i} r_{l,i}\mathbf{A}_{l,i}|||_*, \tag{4.20}$$

where $(r_{l,i})$ is a Rademacher sequence and the expectation is conditioned on $(A_{i,l})$. Then

by the symmetry of the standard Gaussian distribution, we obtain

$$\mathrm{E}_{(r_{l,i})} |||\sum_{l,i} r_{l,i}\mathbf{A}_{l,i}|||_* = \frac{1}{\sqrt{L}} \sup_{\mathbf{M}\in\kappa(\alpha,R)} \left|\sum_{l,i}\langle r_{l,i}\mathbf{b}_{l,i}, \mathbf{M}\mathbf{e}_l\rangle\right| = \frac{1}{\sqrt{L}} \underbrace{\sup_{\mathbf{M}\in\kappa(\alpha,R)} \left|\sum_{l,i}\langle \mathbf{b}_{l,i}, \mathbf{M}\mathbf{e}_l\rangle\right|}_{(\S)}, \tag{4.21}$$

where the second equation holds in the sense of distribution.

Note that (§) is the maximum of linear combinations of Gaussian variables and an upper bound can be obtained using Dudley's inequality [70]. Once we obtain a tail estimate of (§), since (§) no longer depends on the Rademacher sequence $(r_{l,i})$, it can be used to upper-bound T_1 through (Equation 4.20) and (Equation 4.21). An upper bound on T_2 has been already derived in Lemma 3. Combining these upper estimates on T_1 and T_2 results in Lemma 4. From the lower bound on $\sum_{l,i}\langle \mathbf{A}_{l,i}, \mathbf{M}\rangle^2$, we have

$$\frac{\|\mathbf{M}\|_{\mathrm{F}}^2}{d_2} - c\alpha R\sqrt{\frac{d}{Ld_2}}\log^3 d \leqslant \frac{1}{d_2}\sum_{l,i}\langle \mathbf{A}_{l,i}, \mathbf{M}\rangle^2 \leqslant \sup_{\mathbf{M}\in\kappa(\alpha,R)} \frac{\sum_{l,i}\langle \mathbf{A}_{l,i}, \mathbf{M}\rangle z_{l,i}}{d_2}.$$

From Lemma 4, we get the following inequality, which then leads to the final result.

$$\frac{\|\mathbf{M}\|_{\mathrm{F}}^2}{d_2} \leqslant c\log^3 dR\sqrt{\frac{d}{Ld_2}}(\alpha \vee \sigma\sqrt{L})$$

4.6 Role of entropy numbers in providing theoretical guarantees

Both of Lemma 3 and Lemma 4 rely on a key quantity that captures the "complexity" of the set $\kappa(\alpha, R)$. In particular, we need an estimate of the entropy number of the set $\kappa(\alpha, R)$ with respect to the $\|\|_{1\to 2}$ unit ball, which is given by the following Lemma.

Lemma 5 *Let $\kappa(\alpha, R)$ be as in* (Equation 4.10) *and let $B_{1\to 2}$ be the unit ball with respect to $\|\cdot\|_{1\to 2}$. Then there exists a numerical constant c such that*

$$\int_0^\infty \sqrt{\log N(\kappa(\alpha, R), \eta B_{1\to 2})}d\eta \leqslant cR\sqrt{d}\log^{3/2}(d_1 + d_2). \tag{4.22}$$

Here $N(\kappa(\alpha, R), \eta B_{1\to 2})$ denotes the covering number of $\kappa(\alpha, R)$ with respect to the scaled unit ball $\eta B_{1\to 2}$.

Let $B_\wedge$ denote the unit ball with respect to the projective tensor norm in $\ell_\infty^{d_2} \otimes \ell_2^{d_1}$. The injective tensor norm in $\ell_\infty^{d_2} \otimes \ell_2^{d_1}$ reduces to $\|\cdot\|_{1\to 2}$. By its construction, $\kappa(\alpha, R)$ is given as the intersection of two norm balls $\alpha B_{1\to 2}$ and $RB_\wedge$. The proof of Lemma Equation 28 reduces to the computation of the entropy number of the identity map on $\ell_\infty^{d_2} \otimes \ell_2^{d_1}$ from the Banach space with the projective tensor norm to that with the injective tensor norm.

4.7 Information-theoretic lower bound

Theorem 9 provides the rate of convergence for the mixed-norm constrained least squares estimator proposed in (Equation 4.11). A natural question is whether this rate of convergence is optimal, at least over the constraint set $\kappa(\alpha, R)$. We provide such a guarantee by establishing a minimax lower bound for column-wise low rank matrix sensing on the the constraint set $\kappa(\alpha, R)$.

Theorem 10 *Let* $\mathbf{M} \in \mathbb{R}^{d_1 \times d_2}$ *be a rank-r matrix and let* $y_{l,i} = \langle \mathbf{A}_{l,i}, \mathbf{X}_0 \rangle + z_{l,i}$ *with* $\mathbf{A}_{l,i} = \frac{1}{\sqrt{L}} \mathbf{b}_{l,i} \mathbf{e}_i^\top$, *where* $\mathbf{b} \sim \mathcal{N}(0, \mathbf{I}_{d_1}/L)$, $z_{l,i} \sim \mathcal{N}(0, \sigma^2)$. *Further,* $\frac{48\alpha^2}{d_1 \vee d_2} \leqslant R^2$. *Then the minimiax* $\|\cdot\|_F$*-risk is lower bounded as*

$$\inf_{\widehat{\mathbf{M}}} \sup_{\mathbf{M} \in \kappa(\alpha, R)} \frac{1}{d_2} \mathrm{E} \left\| \widehat{\mathbf{M}} - \mathbf{M} \right\|_{\mathrm{F}}^2 \geqslant \frac{\alpha^2}{16} \cdot \min \left(1, \frac{\sigma\sqrt{L}}{\alpha} \sqrt{\frac{r(d_1 + d_2)}{L d_2}} \right). \tag{4.23}$$

where $\kappa(\alpha, R) = \{\mathbf{X} : \|\mathbf{X}\|_{1\to 2} \leqslant \alpha, \|\mathbf{X}\|_{\text{mixed}} \leqslant R\}$.

The rate of convergence in our earlier result in Theorem 9 is:

$$\frac{\|\widehat{\mathbf{M}} - \mathbf{M}\|_{\mathrm{F}}^2}{d_2} \leqslant c_1 \cdot \alpha^2 \cdot \max \left(1, \frac{\sigma\sqrt{L}}{\alpha} \right) \cdot \sqrt{\frac{r(d_1 + d_2) \log^6 d}{L d_2}} \tag{4.24}$$

Comparing the results in (Equation 4.23) and (Equation 4.24), we can conclude that the convergence rate in Theorem 9 is rate-optimal under noise regimes, when $\frac{\sigma\sqrt{L}}{\alpha} > 1$. The proof of Theorem 10 is provided in appendix.

4.8 Algorithms for mixed norm based low rank recovery

Apart from enjoying optimal sample efficiency for the model in (Equation 4.2), the mixed-norm based convex relaxation also be implemented efficiently. To develop the computational algorithm, we first describe how to compute the the mixed-norm of any matrix $\mathbf{X}$.

For any given factorization of $\mathbf{X} = \mathbf{U}\mathbf{V}^\top$, we have $\|\mathbf{U}\|_F^2 = \mathrm{trace}(\mathbf{U}^\top\mathbf{U})$ and $\|\mathbf{V}\|_{1\to 2} = \left\|\mathrm{diag}(\mathbf{V}^\top\mathbf{V})\right\|_\infty$. Further, for any such factorization, we can set $\|\mathbf{U}\|_F = \|\mathbf{V}\|_{1\to 2}$ by rescaling the factors without loss of generality. As a result, the mixed-norm of $\mathbf{X}$ can be computed as

$$\begin{aligned} \|\mathbf{X}\|_{\mathrm{mixed}} \quad = \quad & \min_{\mathbf{W}_{11},\mathbf{W}_{22}} \quad \max(\mathrm{trace}(\mathbf{W}_{11}), \|\mathrm{diag}(\mathbf{W}_{22})\|_\infty) \\ & \text{s.t.} \quad \begin{bmatrix} \mathbf{W}_{11} & \mathbf{X} \\ \mathbf{X}^\top & \mathbf{W}_{22} \end{bmatrix} \succeq \mathbf{0}, \end{aligned} \tag{4.25}$$

where $\mathrm{diag}(\mathbf{W}_{22})$ denotes the vector of the diagonal entries of $\mathbf{W}_{22}$. This semi-definite matrix based formulation comes about by setting $\mathbf{U}\mathbf{U}^\top = \mathbf{W}_{11},\ \mathbf{V}\mathbf{V}^\top = \mathbf{W}_{22},\ \mathbf{U}\mathbf{V}^\top = \mathbf{X}$. Then the optimization routine in (Equation 4.11) can be written as

$$\begin{aligned} & \underset{\mathbf{W}_{11},\mathbf{W}_{22},\mathbf{X}}{\text{minimize}} \quad \textstyle\sum_{l,i} |y_{l,i} - \langle \mathbf{A}_{l,i}, \mathbf{X}\rangle|^2 \\ & \text{subject to} \quad \mathrm{trace}(\mathbf{W}_{11}) \leqslant R, \quad \|\mathrm{diag}(\mathbf{W}_{22})\|_\infty \leqslant R, \\ & \qquad\qquad \|\mathbf{X}\|_{1\to 2} \leqslant \alpha, \quad \mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \mathbf{X} \\ \mathbf{X}^\top & \mathbf{W}_{22} \end{bmatrix} \succeq \mathbf{0}. \end{aligned} \tag{4.26}$$

The program in (Equation 4.26) is now a constrained convex optimization problem over a set of semidefinite (PSD) matrices.

4.8.1 ADMM based fast algorithm

The program in (Equation 4.26) can be implemented using standard convex optimization solvers like SeDuMi. [71]. However, this could result in scaling issues, as run times could be prohibitive in higher dimensions. To address this, we propose to use the ADMM based algorithm [72] which breaks down the optimization problem into smaller problems that can be solved efficiently. Our approach is similar to [73], where the positive semidefinite constraint on $\mathbf{W}$ in (Equation 4.26) is treated separately from the other constraints. We provide an algorithm for the norm-penalized version of (Equation 4.26). By Lagrangian duality, the penalized version and the constrained version are equivalent when the Lagrangian multipliers λ_1 and λ_2 are chosen appropriately.

By introducing an auxiliary variable $\mathbf{T}$, it is straightforward to show that the optimization problem (Equation 4.26) is equivalent to

$$\begin{aligned} \underset{\mathbf{W},\mathbf{T}}{\text{minimize}} \quad & \textstyle\sum_{l,i} |y_{l,i} - \langle \mathbf{A}_{l,i}, \mathbf{W}_{11} \rangle|^2 + \lambda_1 \operatorname{trace}(\mathbf{T}_{11}) + \lambda_2 \left\| \operatorname{diag}(\mathbf{W}_{22}) \right\|_\infty \\ \text{subject to} \quad & \left\| \mathbf{W}_{12} \right\|_{1\to 2} \leqslant \alpha, \;\; \mathbf{T} = \mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{bmatrix}, \;\; \mathbf{T} \succeq 0. \end{aligned} \tag{4.27}$$

In (Equation 4.27), we carry the constraints on $\operatorname{trace}(\mathbf{T}_{11})$ and $\|\operatorname{diag}(\mathbf{W}_{22})\|$ to the objective function by using the Lagrangian formulation. Note that there are other variations possible, with more or fewer constraints carried over to the objective function. The formulation in (Equation 4.27) is amenable to the ADMM algorithm. The augmented Lagrangian of (Equation 4.27) is given by

$$\begin{aligned} L(\mathbf{T}, \mathbf{W}, \mathbf{Z}) = & \sum_{l,i} |y_{l,i} - \langle \mathbf{A}_{l,i}, \mathbf{W}_{11} \rangle|^2 + \lambda_1 \operatorname{trace}(\mathbf{T}_{11}) + \lambda_2 \left\| \operatorname{diag}(\mathbf{W}_{22}) \right\|_\infty \\ & + \langle \mathbf{Z}, \mathbf{T} - \mathbf{W} \rangle + \frac{\rho}{2} \left\| \mathbf{T} - \mathbf{W} \right\|_{\mathrm{F}}^2 + \chi_{\{\mathbf{T} \succeq 0\}} + \chi_{\{\|\mathbf{W}_{12}\|_{1\to 2} \leqslant \alpha\}}, \end{aligned}$$

where $\mathbf{Z}$ is the dual variable and $\chi_{\mathcal{S}}$ is the indicator function of the set $\mathcal{S}$ given as $\chi_{\mathcal{S}}(t) = 0$ if $t \in \mathcal{S}$ and $\chi_{\mathcal{S}}(t) = \infty$ otherwise. The ADMM algorithm then iterates by alternating among $\mathbf{T}$, $\mathbf{W}$ and $\mathbf{Z}$, as shown in Algorithm 4.

Algorithm 4 ADMM algorithm

Initialize: $\mathbf{T}^0$, $\mathbf{W}^0$, $\mathbf{Z}^0$
while not converged **do**
$\mathbf{T}^{k+1} = \underset{\mathbf{T}\succeq 0}{\operatorname{argmin}}\ L(\mathbf{T}, \mathbf{W}^k, \mathbf{Z}^k)$
$\mathbf{W}^{k+1} = \underset{\|\mathbf{W}_{12}\|_{1\to 2}\leqslant\alpha}{\operatorname{argmin}}\ L(\mathbf{T}^{k+1}, \mathbf{W}, \mathbf{Z}^k)$
$\mathbf{Z}^{k+1} = \mathbf{Z}^k + \rho(\mathbf{T}^{k+1} - \mathbf{W}^{k+1})$ =0

Update for T

$$\begin{aligned}\mathbf{T}^{k+1} &= \arg\min_{\mathbf{T}} L(\mathbf{X}, \mathbf{W}^k, \mathbf{Z}^k) \\ &= \arg\min_{\mathbf{T}\succeq\mathbf{0}} \lambda_1\langle\left[\begin{smallmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{smallmatrix}\right], \mathbf{T}\rangle + \langle\mathbf{Z}, \mathbf{T}-\mathbf{W}\rangle + \frac{\rho}{2}\|\mathbf{T}-\mathbf{W}\|_F^2 \\ &= \pi_{\mathcal{S}_+^d}(\mathbf{W}^k - \rho^{-1}\left(\mathbf{Z}^k + \lambda_1\left[\begin{smallmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{smallmatrix}\right]\right))\end{aligned}$$

where $\pi_{\mathcal{S}_+^d}$ denotes the projection operator and $\mathcal{S}_+^d$ s the set of PSD matrices of size d.

Update for W

$$\mathbf{W}^{k+1} = \arg\min_{\mathbf{W}} L(\mathbf{T}^{k+1}, \mathbf{W}, \mathbf{Z}^k)$$

This optimization can be separate into four sub-problems. Let $\mathbf{C} = \mathbf{T}^{k+1} + \rho^{-1}\mathbf{Z}^k$. Let $\widetilde{\mathbf{M}}$ be the matrix obtained by setting the diagonal elements of any matrix $\mathbf{M}$ to 0 and let $q = \operatorname{diag}(\mathbf{C}_{22})$ The four sub-problems are

1. $\mathbf{W}_{12}^{k+1} = \underset{\|\mathbf{W}_{12}\|_{1\to 2}\leqslant\alpha}{\arg\min}\ f(\mathbf{W}_{12}) + \langle\mathbf{Z}_{12}^k, \mathbf{T}_{12}^{k+1} - \mathbf{W}_{12}\rangle + \frac{\rho}{2}\left\|\mathbf{X}_{12}^{k+1} - \mathbf{W}_{12}\right\|_F^2$ where $f(\mathbf{W}_{12}) = \sum_{l,i}|y_{l,i} - \langle A_{l,i}, \mathbf{W}_{12}\rangle|^2$
2. $\mathbf{W}_{11}^{k+1} = \arg\min_{\mathbf{W}_{11}} \|\mathbf{W}_{11} - \mathbf{C}_{11}\|_F^2$
3. $\widetilde{\mathbf{W}}_{22}^{k+1} = \arg\min_{\widetilde{\mathbf{W}}_{22}} \left\|\widetilde{\mathbf{W}}_{22} - \widetilde{\mathbf{C}}_{22}\right\|_F^2$
4. $\operatorname{diag}(\mathbf{W}_{22}^{k+1}) = \underset{u\in\mathbb{R}^{d_2}}{\arg\min}\ \lambda_2\|u\|_\infty + \frac{\rho}{2}\|u - q\|_2^2$

Sub-problem 1 is a least-squares problem which has a closed form solution. Sub-problems 2 and 3 are readily solved by setting $\mathbf{W}_{11}^{k+1} = \mathbf{C}_{11}$ and $\widetilde{\mathbf{W}}_{22}^{k+1} = \widetilde{\mathbf{C}}_{22}$. Sub-problem 4 has a closed form solution as described in [73].

4.9 Experiments

To complement our theoretical results, we observe the empirical performance of the mixed-norm-based method in a set of Monte Carlo simulations. Matrices are set to be of size $1,000 \times 1,000$ and of rank 5. In our experiments we normalize the columns to have the same energy. We observe the estimation error by varying the degree of compression and the signal-to-noise (SNR) ratio. We compare the proposed method to the popular matrix LASSO, which minimizes the least squares loss with a nuclear norm regularizer. We used Algorithm 4 to implement the mixed-norm based method. The nuclear norm minimization approach was implemented using the algorithm provided in [74]. Figure 4.1 shows the obtained simulation results. The estimation error is averaged over 5 trials. The result indicates that the mixed-norm-based estimator outperforms the nuclear-norm-based estimator at both the SNR levels considered.

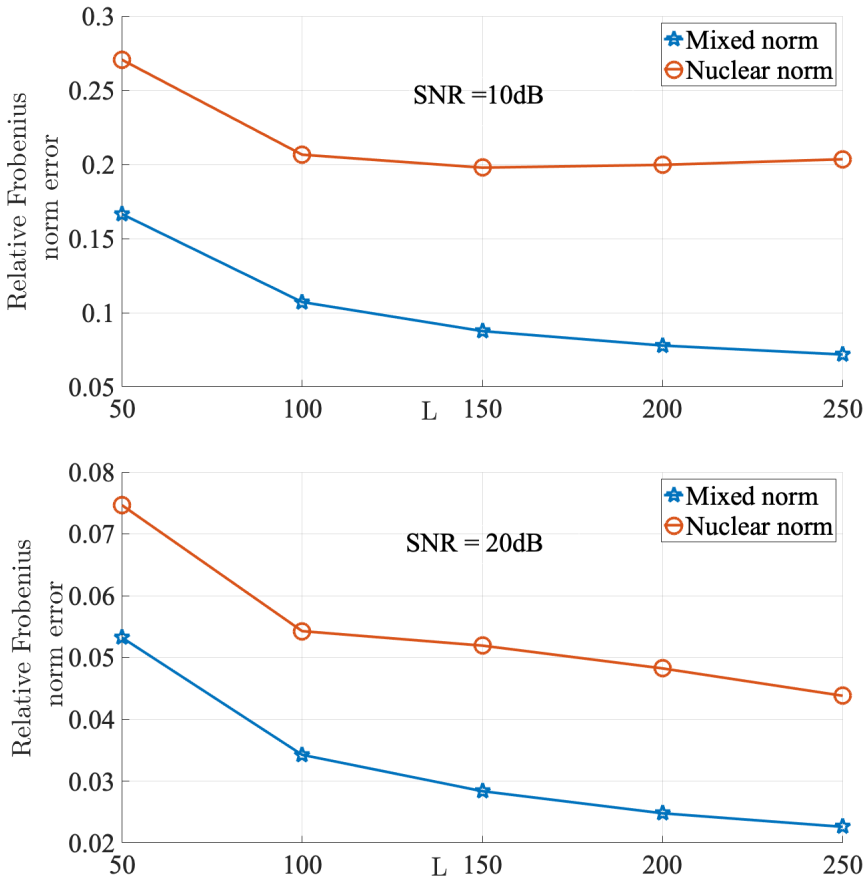


Figure 4.1: *Simulation results comparing the proposed mixed-norm based estimator and the nuclear norm based estimator. The test matrices were of size* $1,000 \times 1,000$ *with rank* 5*. Each data point is computed as an average of 5 trials. Mixed norm estimator is able to achieve much lower errors with fewer measurements compared to the nuclear norm estimator.*

CHAPTER 5

UNIFORM GUARANTEES ON BLIND DECONVOLUTION

5.1 Blind deconvolution as a low rank recovery problem

Blind deconvolution is a fundamental problem that arises in signal processing, communications, imaging. An unknown signal (from a possibly known model family) is convolved with an unknown kernel (also possibly from a known model family) and we only observe this convolution. The task is to then estimate the unknown signal and the kernel. A simple application is image deblurring, where we observe a blurry version of an unknown image and the goal is to reconstruct the underlying sharp image. More complex applications include communication channels, where the receiver must extract the message sent by the transmitter after it has been convolved with the channel response [75]. By itself, the blind deconvolution problem is ill-posed. However, assumptions on the class of signals may lead to a tractable problem.

In this chapter, we will develop a framework to solve the blind deconvolution problem under the assumption that the unknown signal and the convolution kernel lie in known lower-dimensional subspaces. Our problem formulation is similar to that in [75]. However, unlike existing results in the literature, our framework also comes with uniform guarantees over the entire signal class that we consider and also has an optimal noise scaling. This makes our proposed method a viable choice for blind deconvolution under noisy regimes.

Let $\mathbf{w}, \mathbf{x} \in \mathbb{R}^L$ be two signals that are convolved with each other. The standard model of observation for blind deconvolution is then given as

$$\mathbf{y} = \mathbf{w} * \mathbf{x} + \nu, \tag{5.1}$$

where $\nu \in \mathbb{R}^L$ is additive noise, $\mathbf{y} \in \mathbb{R}^L$ are the measurements and $*$ denotes circular

convolution defined as

$$(\mathbf{w} * \mathbf{x})[i] = \sum_{l=1}^{L} \mathbf{w}[l]\mathbf{x}[i-l+1]. \tag{5.2}$$

where $i = 1, \cdots, L$ and the index $i - l + 1$ is assumed to be modulo $\{1, \cdots, L\}$. Given $\mathbf{y}$, we are interested in recovering the signals $\mathbf{h}$ and $\mathbf{m}$.

For observation model above, is it evident that for arbitrary $\mathbf{w}$ and $\mathbf{x}$, the recovery problem is highly ill-posed and there exist many solutions that generate the same set of observations $\mathbf{y}$. To circumvent this, we assume that these two signals $\mathbf{w}$ and $\mathbf{x}$ lie in known subspaces $\mathbf{D}$ and $\mathbf{C}$ respectively, which results in the following form:

$$\mathbf{w} = \mathbf{D}\mathbf{h} \tag{5.3}$$

$$\mathbf{x} = \mathbf{C}\mathbf{m} \tag{5.4}$$

where $\mathbf{D} \in \mathbb{R}^{L\times K}$, $\mathbf{C} \in \mathbb{R}^{L\times N}$, $\mathbf{h} \in \mathbb{R}^{K}$ and $\mathbf{m} \in \mathbb{R}^{N}$. The bases $\mathbf{D}$ and $\mathbf{C}$ are assumed to be known a priori.

After applying the Fourier transform on both sides the observations $\mathbf{y}$ can be written as linear observations of the rank-1 matrix $\mathbf{m}\mathbf{h}^*$ (See [75]). Let $\mathbf{Z}_0 = \mathbf{m}\mathbf{h}^*$ and $\widehat{\mathbf{D}} = \mathbf{F}\mathbf{D}$, where $\mathbf{F} \in \mathbb{C}^{L\times L}$ is the unitary discrete Fourier transform matrix defined as

$$\mathbf{F}(m,n) = \frac{1}{\sqrt{L}} \mathrm{e}^{-j2\pi(m-1)(n-1)/L}. \tag{5.5}$$

Then the blind deconvolution problem can be recast as recovering $\mathbf{Z}_0$ from measurements given as

$$\mathbf{y} = \mathcal{A}(\mathbf{Z}_0\widehat{\mathbf{D}}^*) + \nu \tag{5.6}$$

where $\mathcal{A} : \mathbf{X} \mapsto \{\langle \widehat{\mathbf{c}}_l \mathbf{e}_l^\top, \mathbf{X}\rangle\}_{l=1}^{L}$ and $\widehat{\mathbf{c}}_l^\top$ is the lth row of the matrix $\mathbf{FC}$. Hence, the blind deconvolution problem can be recast as a rank-1 matrix recovery problem.

Assumptions on the bases $\mathbf{D}$ and $\mathbf{C}$: In order to develop a theoretical analysis of the problem, we will assume $\mathbf{C}$ is *generic*, with its entries drawn independently from the standard

Gaussian random distribution. We will also assume that

$$\mathbf{D} = \begin{bmatrix} \mathbf{I}_K \\ \mathbf{0}_{L-K \times K} \end{bmatrix}. \tag{5.7}$$

Note that our choice of $\mathbf{D}$ is time-limited to length K. Our results will also depend on a certain *incoherence* parameter, that describes how uniform the energy spread of the signal $\mathbf{w} = \mathbf{Dh}$ is in the Fourier domain. We denote this parameter as μ and we consider the class of signals $\mathbf{h}$ such that

$$\frac{\|\mathbf{FDh}\|_\infty}{\|\mathbf{FDh}\|_2} \leqslant \frac{\mu}{\sqrt{L}} \tag{5.8}$$

for $1 \leqslant \mu \leqslant \sqrt{L}$. If $\mu = 1$, then we have signals that are perfectly "flat". But as we allow μ to increase, we allow the signal to be more "spiky".

5.2 Tailoring a convex relaxation for blind deconvolution

This reinterpretation of the blind deconvolution problem as a low rank matrix recovery problem opens up a number of low-rank matrix recovery methods as possible tools. These set of tools include the standard nuclear norm based low rank matrix recovery methods that cast the problem as a convex optimization problem. Such a convex relaxation is studied in both [75, 76], where a sample-optimal but instance based theoretical guarantee on the recovery error is given: the theoretical analysis holds for a *fixed* pair of signal and kernel. We are interested in developing a method that can offer a *uniform* guarantee over the set of all candidate signals and kernels.

Our approach to solve the blind deconvolution by using a convex relaxation of the set of rank-1 matrices. However, we will use a different convex relaxation than [75, 76]. The convex relaxation we propose is built on the interpretation of matrices as members of tensor products of Banach spaces that we introduced in section 4.2. In fact, the blind deconvolution measurement model developed in (Equation 5.6) is exactly the same as the column-wise matrix sensing model we considered in Equation 4.2, except for the following

additional structure on the right factor of the underlying low-rank matrix $\mathbf{Z}_0\widehat{\mathbf{D}}^*$:

- The matrix is always of rank 1.
- The right factor always lies in a known K-dimensional subspace
- The right factor obeys the incoherence condition in (Equation 5.8)

While the rank being fixed to 1 can easily be incorporated into the framework developed in chapter 4, the latter pair of differences necessitate a more deliberate convex relaxation. Let us define the $\mathbf{X}_0 = \mathbf{Z}_0\widehat{\mathbf{D}}^*$ and consider factorization of $\mathbf{X}_0$ as

$$\mathbf{X}_0 = \mathbf{U}\mathbf{V}^*. \tag{5.9}$$

Then, the nuclear norm of $\mathbf{X}_0$ can be written as

$$\|\mathbf{X}_0\| = \inf_{\mathbf{U},\mathbf{V}:\mathbf{X}_0=\mathbf{U}\mathbf{V}^*} \|\mathbf{U}\|_F \|\mathbf{V}\|_F . \tag{5.10}$$

Therefore, minimizing the nuclear norm results in minimizing the Frobenius norm of the right factor (equivalently, the ℓ_2 norm, in the rank-1 case). Similarly, from (Equation 4.8), minimizing the mixed norm of $\mathbf{X}_0$ results in minimizing the ℓ_∞ norm of the right factor (in the rank-1 case).

In order to design a suitable convex relaxation for the blind deconvolution problem, we develop over previous approaches by explicitly using the incoherence criteria (Equation 5.8). A bound on the Frobenius norm (or ℓ_2-norm) on the right factor imposed by the nuclear norm, along with (Equation 5.8) implies that the $\|\cdot\|_{1\to 2}$ (or ℓ_∞-norm of rank-1 factors) is also bounded. Intuitively, this results in a bound on **both** of the norms of the right factor. This forms the basis of our new convex relaxation to solve the blind deconvolution problem: We penalize **both** the norms on the right factors, or equivalently, the **maximum** of the Frobenius norm and the $\|\cdot\|_{1\to 2}$ norms.

We first define tensor products with respect to two separate pairs of Banach spaces and then combine them to obtain our characterization of the set of rank-1 matrices. Firstly, The

vector space of $N \times L$ matrices with their row space in the span of $\mathbf{FD}$ can be identified to the tensor product of the following pair of Banach spaces:

$$V_\infty := (\mathrm{span}(\mathbf{FD}), \|\cdot\|_\infty) \tag{5.11}$$

$$U_2 = \ell_2^N \tag{5.12}$$

Then, the vector space of the matrices of interest can be identified to $V_\infty^* \otimes U_2$. Alternatively, if we define

$$V_2 := (\mathrm{span}(\mathbf{FD}), \|\cdot\|_2), \tag{5.13}$$

then the vector space of the matrices of interest can also be identified to $V_2 \otimes U_2$. Further, each such tensor product can itself be endowed with a unique norm, including the injective and projective norms defined in chapter 4.

To design a convex relaxation for the blind deconvolution problem, we consider the following combinations of $V_\infty \otimes U_2$ and $V_2 \otimes U_2$ with the norm scaled by $\frac{\mu}{\sqrt{L}}$:

- We define the "ϵ-interpolation space" as

$$X_\epsilon := V_\infty \otimes_\epsilon U_2 \cap \frac{\mu}{\sqrt{L}} V_2 \otimes_\epsilon U_2, \tag{5.14}$$

  where $\otimes_\epsilon$ indicates the usage of the injective norm on the tensor product (See chapter 4). For a matrix $\mathbf{X}$ interpreted as a member of this Banach space, the corresponding norm is the maximum of the two norms corresponding to each of the tenor product spaces. For the particular tensor products above, the norm of a matrix in the $V_\infty \otimes_\epsilon \ell_2^N$ space is $\|\mathbf{X}\|_{1\to 2}$ and the norm of a matrix in the $\frac{\mu}{\sqrt{L}} V_2 \otimes_\epsilon \ell_2^N$ space is $\frac{\mu}{\sqrt{L}} \|\mathbf{X}\|_2$. The norm in the interpolation space then is given as

$$\|\mathbf{X}\|_\epsilon = \max(\|\mathbf{X}\|_{1\to 2} \frac{\mu}{\sqrt{L}} \|\mathbf{X}\|_{2\to 2}). \tag{5.15}$$

- We define "π-interpolation space" as

$$X_\pi := V_\infty \otimes_\pi U_2 \cap \frac{\mu}{\sqrt{L}} V_2 \otimes_\pi U_2, \tag{5.16}$$

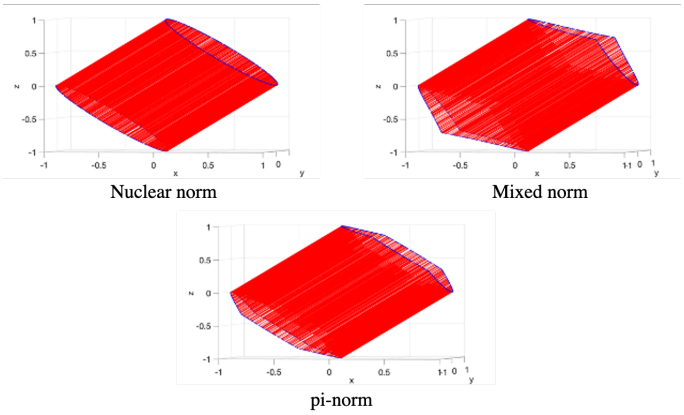


Figure 5.1: *Unit-balls w.r.t the nuclear norm, mixed norm and the π-norm can be visualied by considering 2×2 symmetric matrices. The unit ball w.r.t the π-norm is given by the intersection of the unit balls w.r.t the nuclear and the mixed norms.*

where $\otimes_\pi$ indicates the usage of the projective norm on the tensor product (See chapter 4). The corresponding norm of a matrix in this Banach space is given as

$$\|\mathbf{X}\|_\pi = \max(\|\mathbf{X}\|_{\text{mixed}} \frac{\mu}{\sqrt{L}} \|\mathbf{X}\|_*) \tag{5.17}$$

where the the mixed norm of a matrix is defined as in (Equation 4.8) and $\|\mathbf{X}\|_*$ denotes the nuclear norm.

As before, the pair of $\{\|\cdot\|_\epsilon, \|\|_\pi\}$ help characterize the set of rank-r matrices due to the following property:

Lemma 6 *Let $\mathbf{X} \in \mathbb{R}^{N \times L}$ be such that $\text{rank}(\mathbf{X}) \leqslant r$. Then,*

$$\|\mathbf{X}\|_\epsilon \leqslant \|\mathbf{X}\|_\pi \leqslant r \|\mathbf{X}\|_\epsilon . \tag{5.18}$$

The unit ball of the interpolation-space is given by the intersection of the unit balls w.r.t each of the constituent Banach spaces. For example, the unit ball of the π-interpolation space is given as the intersection of the unit balls w.r.t the mixed norm and the scaled nuclear norm. We illustrate this in the case of 2×2 symmetric matrices in Figure 5.1.

5.3 Blind deconvolution using hybrid norms

From (Equation 5.18), it is clear that the set of rank-r matrices with $\|\mathbf{X}\|_\epsilon \leqslant \alpha$ is contained in the set $\kappa(\alpha, R) = \{\mathbf{X} : \|\mathbf{X}\|_\epsilon \leqslant \alpha, \|\mathbf{X}\|_\pi \leqslant R\}$ when $R \geqslant r\alpha$. Furthermore, if $\mathrm{rank}(\mathbf{X}) = 1$, then $\|\mathbf{X}\|_\pi \leqslant \|\mathbf{X}\|_\epsilon$. The ϵ-norm is then inactive and the π-norm bound sufficiently characterizes the set $\kappa(\alpha, \alpha)$. Our estimator for the ground truth rank-1 matrix $\mathbf{Z}_0$ is given as the following convex optimization problem:

$$\begin{aligned} \widehat{\mathbf{Z}} = \quad & \underset{\mathbf{Z} \in \mathbb{C}^{N \times K}}{\operatorname{argmin}} \qquad \frac{1}{2}\|\hat{\mathbf{y}} - \mathcal{A}(\mathbf{Z}\widehat{\mathbf{D}}^*)\|_2^2 \\ & \text{subject to} \quad \|\mathbf{Z}\widehat{\mathbf{D}}^*\|_\pi \leqslant \alpha, \end{aligned} \tag{5.19}$$

where $\alpha = \left\|\mathbf{Z}_0\widehat{\mathbf{D}}^*\right\|_\epsilon$ and can be tuned as a hyperparameter. We now state our main result that bounds the estimation error as a function of dimension of the signals and signal-to-noise ratio.

Theorem 11 *Let* $\mathbf{D}$ *be the matrix containing the first* K *columns of the identity matrix of size* L *and let* $\mathbf{C}$ *be an* $L \times M$ *matrix with entries drawn from the standard Gaussian random distribution. For any set of coefficients* $\mathbf{h} \in \mathbb{R}^K$ *that obeys* (Equation 5.8) *and* $\mathbf{m} \in \mathbb{R}^m$*, let* $\mathbf{y}$ *be given as in* (Equation 1.32) *and let* $\mathbf{Z}_0 = \mathbf{m}\mathbf{h}^*$*. Then, the solution obtained by solving the program in* (Equation 5.19) *obeys*

$$\frac{\left\|\hat{\mathbf{Z}} - \mathbf{Z}_0\right\|_F^2}{\|\mathbf{Z}_0\|_F^2} \lesssim \sqrt{\frac{K+N}{L}}\left(\mu + \frac{\sigma\sqrt{L}}{\|\mathbf{Z}_0\|_F}\right)\log(1/\delta) \tag{5.20}$$

with probability at least $1 - \delta$ *and* $\lesssim$ *denotes that the inequality holds modulo log factors in the dimensions.*

We can make the following observations on the recovery error.

- **Uniform guarantee:** Unlike existing results on blind deconvolution, the abive theorem provides a ***uniform*** recovery guarantee for all $\mathbf{h}$ that obey (Equation 5.8) and all $\mathbf{m}$.

- **Noise scaling:** Under the regime of large noise ($\frac{\sigma\sqrt{L}}{\|\mathbf{Z}_0\|} \geqslant \mu$), the relative error scales **linearly** with the noise to signal ratio $\frac{\sigma\sqrt{L}}{\|\mathbf{Z}_0\|_F}$. This is dimension free (up to log factors) and matches the state of the art results on the noise performance for blind deconvolution [76] with instance guarantees. However, we note that Theorem 3 does not guarantee perfect recovery in the noiseless case.

- Finally, our proposed method can also be extended to the "blind demixing problem", which is a rank-r version of the blind deconvolution problem. In this problem, r number of individual convolutions are summed to obtain a super-position. Such a model is useful when dealing with multi-channel blind deconvolution in applications such as underwater communication and imaging.

5.3.1 Proof sketch

We state the key lemmas involved in proving Theorem 11 and point to the tools we use and defer finer details of the proof to the appendix. We begin with the basic optimality condition that relates the estimate $\hat{\mathbf{Z}}$ to the ground truth $\mathbf{Z}_0$. Let $\mathbf{M} = \hat{\mathbf{Z}} - \mathbf{Z}_0$. By the triangular inequality, we have $\mathbf{M} \in \kappa(2\alpha, 2\alpha)$. For notational brevity, we assume from now on that $\mathbf{M} \in \kappa(\alpha, \alpha)$. (Neither the main result nor the proofs are affected by this since they involve multiplication with some numerical constants.)

By optimality of the solution and (Equation 5.19), we have

$$\sum_l \left(y_l - \langle \hat{\mathbf{c}}\mathbf{e}_l^\top, \hat{\mathbf{Z}} \rangle\right)^2 \leqslant \sum_l \left(y_l - \langle \hat{\mathbf{c}}\mathbf{e}_l^\top, \mathbf{Z}_0 \rangle\right)^2 \tag{5.21}$$

After substituting $\widehat{\mathbf{Z}} - \mathbf{Z}_0$ by $\mathbf{M}$, since $\widehat{\mathbf{y}} = \mathcal{A}(\mathbf{Z}_0\widehat{\mathbf{D}}) + \nu$, we obtain

$$\sum_{l=1}^{L} \langle \hat{\mathbf{c}}_l \mathbf{e}_l^\mathsf{T}, \mathbf{M}\widehat{\mathbf{D}}^* \rangle^2 \leqslant 2 \sum_{l=1}^{L} \langle \hat{\mathbf{c}}_l \mathbf{e}_l^\mathsf{T}, \mathbf{M}\widehat{\mathbf{D}}^* \rangle \nu_l. \tag{5.22}$$

We then use the following pair of results that provide and upper and lower bound on the right and left hand side of (Equation 5.22).

Lemma 7 *For the setting in* (Equation 5.22)*,*

$$\sum_l \langle \widehat{\mathbf{c}}_l \mathbf{e}_l^{\mathsf{T}}, \mathbf{M}\widehat{\mathbf{D}}^* \rangle^2 \geqslant \left\| M\widehat{\mathbf{D}}^* \right\|_F^2 - \alpha^2 \sqrt{N+K} \frac{\sqrt{L}}{\mu} \log^3(N+K) \sqrt{\log(2\zeta)^{-1}} \tag{5.23}$$

with probability at least $1-\zeta$

Lemma 8 *For the setting in* (Equation 5.22)*,*

$$\sum_l \langle \mathbf{A}_l, \mathbf{M}\widehat{\mathbf{D}}^* \rangle \nu_l \lesssim \sigma\alpha \frac{\sqrt{L}}{\mu} \sqrt{K+N} \sqrt{\log(1/\delta)} + \sigma \sqrt{\alpha^2 \frac{L}{\mu^2} + \alpha^2 \sqrt{N+K} \frac{\sqrt{L}}{\mu}} \sqrt{\log(1/\delta)\log(1/2\zeta)} \tag{5.24}$$

(up to log factors) with probability $1-\zeta-\delta$*.*

From Lemma 8, the upper bound on the linear term can be summarized as

$$\sum_l \langle \mathbf{A}_l, \mathbf{M}\widehat{\mathbf{D}}^* \rangle \nu_l \lesssim \sigma\alpha \frac{\sqrt{L}}{\mu} \sqrt{K+N} + \sigma \sqrt{\alpha^2 \frac{L}{\mu^2} + \alpha^2 \sqrt{N+K} \frac{\sqrt{L}}{\mu}} \tag{5.25}$$

(up to log factors) with probability $1-\zeta-\delta$.

From Lemma 7, the lower bound on the quadratic term from Lemma 7 can be summarized as :

$$\sum_l \langle \mathbf{A}_l, \mathbf{M}\widehat{\mathbf{D}}^* \rangle^2 \geqslant \left\| \mathbf{M}\widehat{\mathbf{D}}^* \right\|_F^2 - \alpha^2 \sqrt{N+K} \frac{\sqrt{L}}{\mu} \log^3(N+K) \sqrt{\log(2\zeta)^{-1}} \tag{5.26}$$

Finally, the result in Theorem 11 can be obtained the results of Lemma 8 and Lemma 7.

5.3.2 Entropy estimates

(Equation 5.22) shows that in order to derive the main result, we need to provide uniform bounds on the random processes $\sum_{l=1}^{L} \langle \widehat{\mathbf{c}}_l \mathbf{e}_l^{\mathsf{T}}, \mathbf{M}\widehat{\mathbf{D}}^* \rangle^2$ and $\sum_{l=1}^{L} \langle \widehat{\mathbf{c}}_l \mathbf{e}_l^{\mathsf{T}}, \mathbf{M}\widehat{\mathbf{D}}^* \rangle \nu_l$ over the set of matrices $\mathbf{M} \in \kappa(\alpha, \alpha)$. Obtaining such a uniform bound ultimately requires a bound on the following *entropy integral*:

$$\int_0^\infty \sqrt{\ln \mathcal{N}(\alpha B_\pi, \|\cdot\|_\epsilon, t)} dt, \tag{5.27}$$

where $\mathcal{N}(S, d, t)$ denotes the covering number of the set S with respect to the metric d and balls of radius t. The entropy integral provides a notion of the complexity of the set αB_π and plays a role in obtaining the final sample complexity bounds. Once a bound on the entropy integral is obtained, the uniform bounds required above can be derived using standard bounds on random processes such as the Dudley's inequality and the result by krahmer et.al, in [77] that provides a generic bound on quadratic Gaussian processes.

Before we provide upper bounds for the entropy integral, we will develop background on some related quantities.

Covering and packing numbers: For symmetric convex bodies D and E, the covering number $N(D, E)$ an the packing number $M(D, E)$ are defined as below:

$$N(D, E) = \min\{\ell | \exists \mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_\ell \in D,\ D \subset \bigcup_{1 \leqslant j \leqslant \ell} (\mathbf{x}_j + E)\} \tag{5.28}$$

$$M(D, E) = \max\{\ell | \exists \mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_\ell \in D,\ \mathbf{x}_i - \mathbf{x}_j \notin E, \forall i \neq j\} \tag{5.29}$$

Dyadic entropy number: Let X, Y be Banach spaces. For an operator $T : X \to Y$, the dyadic entropy number is defined as

$$e_k(T) := \inf\{\eta > 0 \,|\, M(T(B_X), \epsilon B_Y) \leqslant 2^{k-1}\}, \tag{5.30}$$

where B_X, B_Y denote the unit balls in X, Y respectively. A property of the dyadic entropy number that we will use is

$$e_k(RS) \leqslant \|R\|\, e_k(S) \tag{5.31}$$

$$e_k(RS) \leqslant \|S\|\, e_k(R) \tag{5.32}$$

where $\|R\|, \|S\|$ denote the operator norms.

Let $\mathrm{id} : X_\pi \to X_\epsilon$ denote the identity operator from the Banach spaces X_π and X_ϵ defined in (Equation 5.17) and (Equation 5.14). The entropy integral in (Equation 5.27) is equivalent up to a constant to the following sum of *dyadic entropy numbers* [69]:

$$\mathcal{E}_{2,1}(\mathrm{id} : X_\pi \to X_\epsilon) := \sum_{k=0}^{\infty} \frac{e_k(\mathrm{id} : X_\pi \to X_\epsilon)}{\sqrt{k}}, \tag{5.33}$$

where $e_k(\mathrm{id} : X_\pi \to X_\epsilon)$ is the identity operator. Hence, we seek to bound $\mathcal{E}_{2,1}(\mathrm{id} : X_\pi \to X_\epsilon)$ to obtain an upper bound on the entropy integral itself. In order to do so, we make use of the following result that establishes a bound on $\mathcal{E}_{2,1}(\mathcal{T} : X \to Y)$ where $X = \ell_1^n$, $Y = \ell_\infty^m$, where $\mathcal{T}$ is any linear operator.

Lemma 9 (*Maurey's empirical lemma*, [78]) *For any linear operator $T : \ell_1^n \to \ell_\infty^m$,*

$$\mathcal{E}_{2,1}(T) \lesssim \sqrt{1 + \log(m \vee n)}(1 + \log(m \wedge n))^{3/2} \|T\| \tag{5.34}$$

where $\|T\|$ is the operator norm.

In order to use the above lemma, our strategy is to first show that there exist maps $R : X_\pi \to \ell_1^{n'}$ and $S : X_\epsilon \to \ell_\infty^{m'}$ for some n', m'. We can then use the property (Equation 5.31) to obtain

$$e_k(id : X_\pi \to X_\epsilon) \leqslant \|R\| \, \|S\| \, e_k(id : \ell_1^{n'} \to \ell_\infty^{m'}). \tag{5.35}$$

Maurey's empirical lemma can be used to bound $e_k(id : \ell_1^{n'} \to \ell_\infty^{m'})$. If n', m' are not too large compared to dimensions of X_π and X_ϵ and the operator norms $\|R\|, \|S\|$ are not large, then we can obtain a non-trivial upper bound on the entropy integral. This strategy is illustrated as a schematic in Figure 5.2. We provide a detailed construction of the embedding

$$\begin{array}{ccc} \ell_1^{m'} & \xrightarrow{\mathrm{id}} & \ell_\infty^{n'} \\ {\scriptstyle R}\uparrow & & \downarrow{\scriptstyle S} \\ X_\pi & \longrightarrow & X_\epsilon \end{array}$$

Figure 5.2: *Our strategy to use Maurey's empirical lemma to obtain entropy estimates for operators between X_π and X_ϵ.*

maps R, S in the appendix. With this construction, we obtain the following bound on the entropy integral:

Lemma 10 *For the Banach spaces X_π, X_ϵ defined in* (Equation 5.17)*,* (Equation 5.14) *the entropy integral in* (Equation 5.27) *is upper bounded as*

$$\int_0^\infty \sqrt{\log(\alpha B_\pi, \eta B_\epsilon)} d\eta \lesssim \alpha \sqrt{N+K} \log^{3/2} L. \tag{5.36}$$

The bound on the entropy integral is proportional to only $N + K$, which represents the number of degrees of freedom of the underlying signal to be recovered. Such a favorable scaling allows us to derive sample optimal theoretical guarantees on the blind deconvolution recovery algorithm.

CHAPTER 6

CONCLUSIONS

In this thesis, we considered several problems where the underlying data could be modeled using low dimensional subspaces. For each application, we showed that using randomized linear dimensionality reduction the data could be mapped to a dimension roughly proportional to the underlying subspace dimension. This leads to more efficient data acquisition and faster algorithms that can then be scaled to large scale data.

Active array imaging

The first contribution is the development of a signal model for imaging range-limited targets using antenna arrays. While imaging such a target with an array of M elements and using a set of K excitation wavelengths, the received array measurements are drawn from a set K overlapping subspaces. We show that random projections can then be used to collect far fewer than M spatial measurements, owing to the overlap between the subspaces.

Broadband source localization We consider the problem of identifying the spectral support of a superposition of a number of broadband signals. By identifying the a set of N samples of a signal approximately bandlimited to $W < 0.5$ lies approximately in a subspace of dimension $2NW$, we develop algorithms that sequentially identify each source in the signal. Owing to the low-dimensionality, we show that the number of samples required to perform reliable source localization scales as $2N \sum_i W_i$ where N is the number of samples, $W_i < 0.5$ is the bandwidth of each source. When $\sum_i W_i \ll 0.5$, this can lead to sample-efficient broadband source localization.

Sketching using block diagonal matrices We consider the problem of sketched linear regression and sketched matrix multiplication. For each of these problems, we study the utility of block diagonal sketching matrices. For linear regression, we show that the total sample complexity needs to scale only as $O(\text{effective rank})$ of the linear model. For matrix

multiplication, we show that the sample complexity needs to scale only as $O(\text{stable ranks})$ of the multiplicands. This matches existing bounds that use more generic sketching matrices, while allowing for blocked sketching of data.

Low rank matrix recovery for column-wise observations

Low rank matrix recovery is a well-studied problem in the compressed sensing literature. The nuclear norm of a matrix is used as a standard proxy for the rank of a matrix. Our next contribution is to identify that when the underlying matrix, or the observation model have unique structure, it can be used to design tailored estimators of the low rank matrix. We use this framework to design a novel low rank matrix recovery algorithm that enjoys sample optimal theoretical guarantees for the column-wise observation model. We provide tight upper bounds on the estimator error and provide an ADMM based fast algorithm to carry out our proposed optimization program.

Uniform guarantees on blind deconvolution Our final contribution is a novel algorithm to address the problem of blind deconvolution. By following the framework developed earlier to design unique estimators of low rank matrices, we develop a new convex relaxation for the set of low rank matrices. We show that such a convex relaxation can be used to design algorithms that enjoy favorable noise scaling and sample complexity uniformly across the signal class we consider.

Appendices

APPENDIX A

PROOFS

A.1 Proofs of results in chapter 2

A.1.1 Proof of Theorem 4

The main tool we use to prove Theorem 4 and Theorem 5 is polynomial identity testing: a polynomial of a finite total degree is said to be identically zero if the coefficient of every monomial term is zero. Any polynomial which is not identically zero, when evaluated at a random point drawn from a continuous distribution, is non-zero with probability 1, since the set of roots of the polynomial has measure zero with respect to the field of real numbers. In conclusion, a multivariate polynomial evaluated at a multi-dimensional point drawn from a continuous distribution is non-zero with probability 1, if any of the polynomial coefficients are non-zero.

To show that a given matrix is of rank r, we choose a suitable submatrix of size $r \times r$ and show that it has a determinant with at least one non-zero coefficient and therefore not identically zero. For random projections whose entries are drawn from the standard normal distribution, we hence need to show that the sketched matrices have submatrices of desirable sizes that have a non-zero determinant with probability 1.

For ease of notation, we prove the result on the transposed matrices: $A = [A_1\ A_2]$, $\|(I - P_H)A\| = 0$, $H = [A_1P\ A_2P]$, and P is the sketching matrix with entries drawn from the standard normal distribution. For ease of explanation, we assume that A_1 and A_2 are orthonormal matrices: $A_i^T A_i = I$ and $\mathcal{R}(A_1) = \mathcal{R}(A_2)$. We later explain how our proof holds for any two general matrices.

Consider the matrix $Z = A_1^T H = [P\ \ UP]$ where $U = A_1^T A_2$ is an orthogonal matrix. If Z is full column rank, then $\mathrm{rank}(H)$ is full column rank. Any orthogonal matrix can be

decomposed as

$$U = Q^T R Q \tag{A.1}$$

where Q is an orthobasis and R is a block diagonal matrix with 2×2 or 1×1 blocks. The 2×2 blocks are of the form

$$R_i = \begin{bmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{bmatrix}$$

with $\cos\theta_i \pm j\sin\theta_i$ being a pair of complex conjugate eigenvalues of U, $j = \sqrt{-1}$. The 1×1 diagonal blocks are equal to ± 1 and are also eigenvalues of U. (Equation A.1) is referred to as the canonical decomposition of U.

Notice that rank$([P \quad UP])$ = rank$(Q[P \quad UP])$ = rank$([QP \quad RQP])$. Since Q is orthogonal, QP is also a matrix with i.i.d. standard normal variables due to the rotational invariance of the standard normal distribution. Thus we can directly work with the matrix $Z = [P \quad RP]$. The following lemma provides the necessary condition on the multiplicity of any real eigenvalue for $[P \quad RP]$ to be full column rank.

Lemma 11 *Let there be a real eigenvalue of U that has an algebraic multiplicity $M_1 > M/2$. Then for some $M - M_1 < l < M/2$, $Z = [P \quad RP]$ is not full column rank.*

Proof Let λ_o be a real eigenvalue with algebraic multiplicity $M_1 > M/2$ and let $M_2 = M - M_1$. Then

$$Z = [P \quad RP] = \begin{bmatrix} P_1 & \lambda_o P_1 \\ P_2 & R_2 P_2 \end{bmatrix} \tag{A.2}$$

where P_1 is the submatrix formed by the first M_1 rows of P and R_2 is the submatrix formed by the last M_2 rows and columns of R. Let $l = M_2/2 + q$, $q > M_2/2$. Then, the submatrix $[P_2 \quad R_2 P_2]$ has a null space $\mathcal{N}_1$ of dimension $2q > M_2$ and the submatrix $[P_1 \quad \lambda_o P_1]$ has an l dimensional null space $\mathcal{N}_2$ given by the range of $\begin{bmatrix} -\lambda_o I \\ I \end{bmatrix}$. Since $2q + l > 2l$, $\mathcal{N}_1 \cap \mathcal{N}_2 \neq \phi$. The result follows. ■

We now prove some lemmas that provide sufficient conditions on R for which $[P \quad RP]$ is full column rank. We state the lemmas along with their proofs and then use them to prove Theorem 4.

Lemma 12 *Let there be l_o pairs of complex conjugate eigenvalues of R with non-zero imaginary part. Then for any $l \leqslant l_o$, $Z = [P \quad RP]$ has full column rank.*

Proof Let P be expressed as

$$P = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_{\frac{n}{2}} \end{bmatrix}$$

where $P_i \in \mathbb{R}^{2\times l}$. Let $P_{i,j,k}$ denote the (j,k)th element of P_i. Rearranging the columns of Z, we obtain

$$\widehat{Z}_1 = \begin{bmatrix} p_{1,1,1} & \cos\theta_1 p_{1,1,1} - \sin\theta_1 p_{1,2,1} & \cdots \\ p_{1,2,1} & \sin\theta_1 p_{1,1,1} - \cos\theta_1 p_{1,2,1} & \cdots \\ \vdots & \vdots & \\ p_{\frac{n}{2},1,1} & \cos\theta_1 p_{\frac{n}{2},1,1} - \sin\theta_1 p_{\frac{n}{2},2,1} & \cdots \\ p_{\frac{n}{2},2,1} & \sin\theta_1 p_{\frac{n}{2},1,1} - \cos\theta_1 p_{\frac{n}{2},2,1} & \cdots \end{bmatrix} \tag{A.3}$$

Expanding the determinant of $\widehat{Z}$, the coefficient of the term $\prod_{i=1,2,\cdots,l} p_{i,1,i}^2$ is $\prod_{i=1,2,\cdots,l} \sin\theta_i$. For any $l \leqslant M/2$, $\det(\widehat{Z}(2lj-2j+1:2lj, 2lj-2j+1:2lj))$ has a term of the form $\prod_{i=2j-1\ldots 2j+l-1} p_{j,1,j}^2$ with coefficient $\prod_{i=2j-1\ldots 2j+l-1} \sin\theta_i$. If there is a set of l rotations R_i such that $\theta_i \neq 0 \ \forall \ i$, then clearly, the determinant associated with this $2l \times 2l$ diagonal block in $\widehat{Z}$ is not equal to 0 with probability 1 and the result follows. ■

Let the number of complex conjugate pairs of eigenvalues be M_1. We next group as many real eigenvalues as possible into pairs such each pair has distinct eigenvalues. Let the

number of such eigenvalue pairs be M_2. The remaining eigenvalues are all real and equal to each other. Let the number of such eigenvalues be $2M_3$ (this number will be even since M is even).

Lemma 13 *For $l \leqslant M_2$, $Z = [P \quad RP]$ has full column rank.*

Proof Again rearranging the columns of Z as

$$\widehat{Z}_2 = \begin{bmatrix} P_{1,1,1} & \lambda_{1,1}P_{1,1,1} & \cdots & P_{1,1,l} & \lambda_{1,1}P_{1,1,n_2} \\ P_{1,2,1} & \lambda_{1,2}P_{1,2,1} & \cdots & P_{1,2,l} & \lambda_{1,2}P_{1,2,n_2} \\ \vdots & \vdots & & \vdots & \vdots \\ P_{\frac{n}{2},2,1} & \lambda_{n_2,2}P_{\frac{n}{2},2,1} & \cdots & P_{\frac{n}{2},2,n_2} & \lambda_{n_2,2}P_{\frac{n}{2},2,n_2} \end{bmatrix} \tag{A.4}$$

Here $\lambda_{i,j}$ denotes the j^{th} eigenvalue of the i^{th} pair. The coefficient of the $2l$ the degree term $\prod_{i=1,..,l} P_{i,1,i}P_{i,2,i}$ is $\prod_{i=1,..,l} \lambda_{i2} - \lambda_{i1} \neq 0$. The determinant associated with this $2l \times 2l$ diagonal block in $\widehat{Z}$ is not equal to 0 with probability 1. The result follows. ■

Lemma 14 *If $M_3 < M_1$, then for $l < 2M_3$, $Z = [P \quad RP]$ has full column rank with probability 1.*

Proof Consider the submatrix of Z shown below, after rearranging the rows and columns

$$\widehat{Z}_3 = \begin{bmatrix} P_{1,1,1} & c\theta_1 P_{1,1,1} - s\theta_1 P_{1,2,1} & \cdots \\ P_{n_1+n_2+1,1,1} & \lambda_o P_{n_1+n_2+1,1,1} & \cdots \\ P_{1,2,1} & s\theta_1 P_{1,1,1} + c\theta_1 P_{1,2,1} & \cdots \\ P_{n_1+n_2+1,2,1} & \lambda_o P_{n_1+n_2+1,2,1} & \cdots \\ \vdots & \vdots & \\ P_{n_3,2,1} & s\theta_{n_3} P_{n_3,1,1} + c\theta_{n_3} P_{n_3,2,1} & \cdots \\ P_{n_1+n_2+n_3,2,1} & \lambda_o P_{n_1+n_2+n_3,2,1} & \cdots \end{bmatrix} \tag{A.5}$$

The coefficient of the $2l^{\text{th}}$ degree term $\prod_{i=1,..,l/2} \prod_{k=1,2} P_{n_1+n_2+i,k,(i-1)*2+k}P_{i,2,(i-1)*2+k}$ in the determinant of $\widehat{Z}_3$ is equal to $\prod_{i=1,..,l/2} \sin^2\theta_i$.

■

Lemma 15 *If $M_3 > M_1$, then there is a real eigenvalue λ_o with algebraic multiplicity greater than $M/2$.*

Proof of Theorem 4:

With eigenvalues grouped as before, let θ_i parametrize the 2×2 diagonal block of R due to the i^{th} pair of complex conjugate eigenvalues, $\lambda_{j1}, \lambda_{j2}$ denote the 2×2 diagonal block due to the j^{th} pair of real eigenvalues with $\lambda_{j1} \neq \lambda_{j2}$, and let the rest of eigenvalues be equal λ_o (repeated at least M_3 times). Due to Lemma 15, $M_3 < M_1$. Let $M_4 = M_1 - M_3$, and let $l = M/2 = M_3 + M_2 + M_1 = 2M_3 + M_2 + M_4$. Rearranging the columns of Z, we obtain

$$\widehat{Z} = \begin{bmatrix} p_1 & Rp_1 & p_2 & Rp_2 & \cdots & p_l & Rp_l \end{bmatrix} \tag{A.6}$$

Then, using Lemma 12, Lemma 13, Lemma 14, the $2l^{\text{th}}$ degree term in the determinant of $\widehat{Z}$ of the form

$$\prod_{i=1}^{M_4} P^2_{M_3+i,1,M_3+i} \prod_{j=1}^{M_2} P_{M_1+j,1,M_1+j} P_{M_1+j,2,M_1+j} \prod_{k_1=1}^{M_3} \prod_{k_2=1,2} P_{M_1+M_2+k_1,k_2,(k_1-1)*2+k_2} P_{k_1,2,(k_1-1)*2+k_2}$$

$$= \prod_{i=1}^{M_4} \prod_{j=1}^{M_2} \prod_{k_1=1}^{M_3} \prod_{k_2=1,2} P^2_{M_3+i,1,M_3+i} P_{M_1+j,1,M_1+j} P_{M_1+j,2,M_1+j} P_{M_1+M_2+k_1,k_2,(k_1-1)*2+k_2} P_{k_1,2,(k_1-1)*2+k_2}$$

has a coefficient given by

$$\prod_{i=1}^{M_4} \sin\theta_{M_3+i} \prod_{j=1}^{M_2} \lambda_{M_1+j,2} - \lambda_{M_1+j,1} \prod_{k=1}^{M_3} \sin^2\theta_k = \prod_{i=1}^{M_4} \prod_{j=1}^{M_2} \prod_{k=1}^{M_3} \sin\theta_{M_3+i} \sin^2\theta_k (\lambda_{M_1+j,2} - \lambda_{M_1+j,1}) \neq 0.$$

Therefore, there is a $2l \times 2l$ submatrix in Z with non-zero determinant, which renders Z full column rank. Finally, since rank(A) = rank$(A_2) = M$, we have for $l = M/2$, rank(H) = M and $\|(I - P_H)A = 0\|$ for $l \geqslant M/2$. We have hence proved Theorem 4 for the case where A_i's are orthogonal and span the same subspace.

The same result can be extended to any general square matrix $U = A^T B$. The proof for this can be obtained using methods exactly as above, but by operating on the real Schur

decomposition of U. The real Schur decomposition of U can be expressed as

$$U = QRQ^T$$

where R is no longer block diagonal but is a pseudo-upper triangular matrix with either 2×2 or 1×1 blocks along the diagonal that reflect the eigenvalue structure of U.

A.1.2 Proof of Theorem 5

Again for ease of notation, we prove the result on the transposed matrices. We now provide a sufficient condition on an ensemble of matrices $A_1, A_2, \cdots A_k \in \mathbb{R}^{N \times M}$ such that for a random matrix P of size $M \times l$, the matrix $H = \begin{bmatrix} A_1P & A_2P & \cdots & A_kP \end{bmatrix}$ is full column rank.

Towards this end, note that we have

$$H = \begin{bmatrix} A_1P & A_2P & \cdots & A_kP \end{bmatrix} = \begin{bmatrix} A_1VV^TP & A_2VV^TP & \cdots & A_kVV^TP \end{bmatrix}$$

Due to the rotational invariance of the standard normal distribution, we can denote V^TP as P and also denote A_iV as A_i itself without loss of generality. With this in place, we can rearrange the columns of H as

$$\widehat{H} = \begin{bmatrix} A_1p_1 & A_2p_1 & \cdots & A_kp_1 & A_1p_2 & \cdots & A_kp_l \end{bmatrix}$$

Expanding the determinant of the first $kl \times kl$ submatrix of $\widehat{H}$, we can obtain the coefficient of the term $p_{11}^k p_{22}^k \cdots p_{ll}^k$ as

$$\begin{aligned} C &= \det(\begin{bmatrix} A_1^{(1)} & A_2^{(1)} \cdots A_k^{(1)} & A_1^{(2)} & \cdots & A_k^{(l)} \end{bmatrix} \\ &= \det(\begin{bmatrix} A_{1\mathcal{S}} & A_{2\mathcal{S}} & \cdots & A_{k\mathcal{S}}) \end{bmatrix} \\ &= \det(\begin{bmatrix} A_1V_{\mathcal{S}} & A_2V_{\mathcal{S}} & \cdots & A_kV_{\mathcal{S}} \end{bmatrix}) \end{aligned}$$

where we assume $\mathcal{S} = \{1, \cdots, l\}$ without loss of generality and $A_i^{(j)}$ denotes the j^{th} column of A_i. If $\begin{bmatrix} A_1V_{\mathcal{S}} & A_2V_{\mathcal{S}} & \cdots & A_kV_{\mathcal{S}} \end{bmatrix}$ is full column rank, then $C \neq 0$ and the result follows. Finally, for $l = r/K$, we have rank of $H = r$ and hence $(I - P_H A) = 0$

When r is not a multiple of K, the same result can be extended to show that if there exist index sets $\mathcal{S}_1$ and $\mathcal{S}_2$ such that $|\mathcal{S}_2| = \lfloor r/K \rfloor$, $|\mathcal{S}_1| = \lfloor r/K \rfloor + 1$, $\mathcal{S}_2 \subset \mathcal{S}_1$ and an orthobasis V such that $\widehat{M} = \begin{bmatrix} A_1 V_{\mathcal{S}_1} & A_2 V_{\mathcal{S}_1} & \cdots & A_q V_{\mathcal{S}_1} & A_{q+1} V_{\mathcal{S}_2} & A_k V_{\mathcal{S}_2} \end{bmatrix}$ is full column rank, then for $l = \lfloor r/K \rfloor + 1$, $\mathcal{R}(Z) = \mathcal{R}(A)$.

A.2 Proofs of results in chapter 3

A.2.1 Proof of Theorem 7

The fundamental property of a distribution of matrices $\mathcal{D}$ that enables any $\mathbf{S} \sim \mathcal{D}$ to satisfy (8, main paper) is the subspace embedding moment property, defined in [4]:

$$\mathop{\mathrm{E}}_{\mathbf{S}\sim\mathcal{D}} \left\| (\mathbf{SU})^T (\mathbf{SU}) - \mathbf{I} \right\|^l \leqslant \epsilon^l \delta, \tag{A.7}$$

for some $l \geqslant 2$, where ϵ and δ are tolerance parameters that determine the sample complexity and $\mathbf{U}$ is any orthobasis for the span of the columns of $\mathbf{W}$ and $\mathbf{Y}$. Thus, our main goal is to prove the subspace embedding moment property holds for block diagonal sketching matrices.

Our methods differ from the common ϵ-net argument, since using union bound for block diagonal matrices results in a suboptimal sample complexity. The main tools we use are the estimates for the suprema of chaos processes found in [68] and an entropy estimate from the study of restricted isometry properties of block diagonal matrices computed in [45]. We first establish tail bounds on the spectral norm of the matrix

$$\mathbf{\Delta} = (\mathbf{S}_D \mathbf{U})^T (\mathbf{S}_D \mathbf{U}) - \mathbf{I}, \tag{A.8}$$

where $\mathbf{U}$ is an orthobasis for a subspace of dimension d and then bound its moments to establish the subspace embedding moment property.

Suprema of chaos processes

We briefly state here the main result from [68] that provides a uniform bound on the deviation of a Gaussian quadratic form from its expectation. Obtaining a tail bound on the spectral norm of $\mathbf{\Delta}$ is just a particular application of this general framework.

For a given set of matrices $\mathcal{P}$, we define the spectral radius $d_2(\mathcal{P})$, the Frobenius norm radius $d_F(\mathcal{P})$, and the Talagrand functional $\gamma_2(\mathcal{P}, \|\cdot\|_2)$ as

$$\begin{aligned} d_2(\mathcal{P}) &= \sup_{\mathbf{P}\in\mathcal{P}} \|\mathbf{P}\|, \\ d_F(\mathcal{P}) &= \sup_{\mathbf{P}\in\mathcal{P}} \|\mathbf{P}\|_F, \\ \gamma_2(\mathcal{P}, \|\cdot\|_2) &= \int_0^{d_2(\mathcal{P})} \sqrt{\log N(\mathcal{P}, \|\|_2, u)} du, \end{aligned}$$

where $N(\mathcal{P}, \|\|_2, u)$ denotes the covering number of the set $\mathcal{P}$ with respect to balls of radius u in the spectral norm. The main result of [68] then is the following theorem.

Theorem 12 *[Theorem 3.1, [68]] Let $\mathcal{P}$ be a set of matrices and let ϕ be a vector of i.i.d. standard normal entries. Then for $t \geqslant 0$,*

$$\mathbb{P}\left(\sup_{\mathbf{P}\in\mathcal{P}} |\|\mathbf{P}\phi\|^2 - \mathrm{E}\,\|\mathbf{P}\phi\|^2| > c_1 E + t\right) \leqslant 2\mathrm{e}^{-c_2 \min\{\frac{t^2}{V^2}, \frac{t}{U}\}} \tag{A.9}$$

where

$$\begin{aligned} E &= \gamma_2(\mathcal{P})[\gamma_2(\mathcal{P}) + d_F(\mathcal{P})] + d_2(\mathcal{P}) d_F(\mathcal{P}), \\ V &= d_2(\mathcal{P})[\gamma_2(\mathcal{P}) + d_F(\mathcal{P})], \\ U &= d_2^2(\mathcal{P}). \end{aligned}$$

A similar approach of using the results from [68] to analyze block diagonal random matrices was first used in [45] in the context of compressed sensing. However, we target a different set of problems that result in different theoretical considerations and proof techniques.

Tail bound on the spectral norm of the matrix $\mathbf{\Delta}$

We first express $\|\mathbf{\Delta}\|$ as

$$\|\mathbf{\Delta}\| = \sup_{\substack{\mathbf{z}\in\mathbb{R}^d \\ \|\mathbf{z}\|=1}} \left|\mathbf{z}^T(\mathbf{S}_D\mathbf{U})^T(\mathbf{S}_D\mathbf{U})\mathbf{z} - 1\right| \tag{A.10}$$

$$= \sup_{\substack{\mathbf{z}\in\mathbb{R}^d \\ \|\mathbf{z}\|=1}} \left|\|\mathbf{S}_D\mathbf{U}\mathbf{z}\|^2 - \mathrm{E}\,\|\mathbf{S}_D\mathbf{U}\mathbf{z}\|^2\right|. \tag{A.11}$$

For the matrices $\mathbf{S}_j$, let $\mathrm{vec}(\mathbf{S}_j)$ denote their vectorized versions, obtained by stacking the columns one below the other. Let $\mathbf{S}_v = [\mathrm{vec}(\mathbf{S}_1)^T\ \mathrm{vec}(\mathbf{S}_2)^T\ \cdots \mathrm{vec}(\mathbf{S}_J)^T]^T$ be the vector containing all of the $\mathrm{vec}(\mathbf{S}_j)$'s. Note that $\mathbf{S}_v$ is a vector with entries drawn from $\mathcal{N}(0,1)$. We can then express (Equation A.10) as

$$\|\mathbf{\Delta}\| = \sup_{\mathbf{P}_z\in\mathcal{P}} \left|\|\mathbf{P}_z\mathbf{S}_v\|^2 - \mathrm{E}\,\|\mathbf{P}_z\mathbf{S}_v\|^2\right|$$

where $\mathcal{P}$ is defined

$$\mathcal{P} = \left\{ \mathbf{P}_z = \begin{bmatrix} \mathbf{P}_1(z) & 0 & \cdots & 0 \\ 0 & \mathbf{P}_2(z) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{P}_J(z) \end{bmatrix} \right\},$$

$$\mathbf{P}_j(z) = \frac{1}{\sqrt{M_j}} \begin{bmatrix} (U_1 z)^T & 0 & \cdots & 0 \\ 0 & (U_1 z)^T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (U_1 z)^T \end{bmatrix}$$

where $z \in \mathbb{R}^d$ and $\|z\| = 1$. Observe that $\|\mathbf{\Delta}\|$ is then the supremum of the deviation of a Gaussian quadratic form from its expectation, taken over the set $\mathcal{P}$.

We can then compute the corresponding quantities $d_2(\mathcal{P})$, $d_F(\mathcal{P})$ and $\gamma_2(\mathcal{P}, \|\cdot\|_2)$ as

follows. The spectral radius $d_2(\mathcal{P})$ is defined as

$$\begin{aligned}\sup_{\mathbf{P}_z \in \mathcal{P}} \|\mathbf{P}_z\| &= \max_{j, \|z\|_2 = 1} \frac{\|U_j z\|}{\sqrt{M_j}} \\ &\leqslant \min\left(\frac{\sqrt{N}\, \|U_j\|_\infty \, \|z\|_1}{\sqrt{M_j}}, \frac{\|U_j\| \, \|z\|_2}{\sqrt{M_j}} \right) \\ &\leqslant \min\left(\frac{\sqrt{N}\, \|U_j\|_\infty \, \|z\|_1}{\sqrt{M_j}}, \frac{\|U_j\| \, \|z\|_1}{\sqrt{M_j}} \right) \\ &\leqslant \|z\|_1 / \sqrt{M_0} \leqslant \frac{d_1}{\sqrt{M_0}}\end{aligned}$$

where the fourth line follows from the definition of M_j.

The radius in the Frobenius norm $d_F(\mathcal{P})$ is defined as

$$\sup_{\mathbf{P}_z \in \mathcal{P}} \|\mathbf{P}_z\|_F = \sum_j \|U_j z\|^2 = 1.$$

The upper bound for $\gamma_2(\mathcal{P}, \|\cdot\|)$ can be obtained from the Equation (34) in Eftekhari et al., 2015).. In their derivation, they consider a full orthobasis and the set of d-sparse vectors. This bound also holds for a fixed d-dimensional subspace. Hence,

$$\gamma_2(\mathcal{P}, \|\cdot\|) \lesssim \sqrt{\frac{d}{M_0}} \log d \log \widetilde{M} \tag{A.12}$$

Plugging these quantities into Theorem 12, we can obtain Lemma 1.

Lemma 16 *For any orthonormal matrix $\mathbf{U} \in R^{\tilde{N} \times d}$ and a block diagonal matrix $\mathbf{S}_D$ as in Theorem 1, there exists a constant c such that*

$$\mathbb{P}\left(\|\mathbf{\Delta}\| \leqslant c \sqrt{\frac{d \log(2/\delta)}{M_0}} \right) \geqslant 1 - \delta. \tag{A.13}$$

For a desired tolerance ϵ, if $M_0 = \Omega\left(\frac{d \log(2/\delta)}{\epsilon^2}\right)$, $\mathbb{P}\left(\|\mathbf{\Delta}\| \leqslant \epsilon\right) \geqslant 1 - \delta$. This is similar to a subspace embedding guarantee. We now show that this tail bound naturally induces a bound on the moments of $\|\mathbf{\Delta}\|$, from which the main theorems in Section 2 can be proved.

Moment bound on $\|\mathbf{\Delta}\|$

Tail bounds for certain random variables can be translated into bounds on their moments using the following result:

Lemma 17 (Proposition 7.13, [79]) *Suppose that a random variable* q *satisfies, for some* $\gamma > 0$*,*

$$\mathbb{P}\left(|q| \geqslant \mathrm{e}^{1/\gamma}\alpha u\right) \leqslant \beta \mathrm{e}^{-u^{\gamma}/\gamma}$$

for all $u > 0$*. Then, for* $p > 0$*,*

$$\mathrm{E}\,|q|^{p} \leqslant \beta\alpha^{p}(\mathrm{e}\gamma)^{p/\gamma}\Gamma\left(\frac{p}{\gamma}+1\right)$$

where $\Gamma(\cdot)$ *is the Gamma function.*

To adapt this result to bound the moments of the spectral norm of the random matrix $\mathbf{\Delta}$, we can choose $q = \|\mathbf{\Delta}\|$, $\gamma = 2$, $\beta = 1$ and $\mathrm{e}^{-u^2/2} = \delta$. We can then obtain the following result.

Lemma 18 *For any orthonormal matrix* $\mathbf{U} \in R^{\tilde{N}\times d}$ *and a block diagonal matrix* $\mathbf{S}_D$ *as in Theorem 1 and* $M_0 = \Omega\left(\frac{d\log(2/\delta)}{\epsilon^2}\right)$*, then*

$$\mathrm{E}\,\|\mathbf{\Delta}\|^{p} \leqslant \epsilon^{p}\delta \tag{A.14}$$

for $p = (\frac{\log(1/\delta)}{\epsilon^2})$.

Approximate matrix product guarantee

With the moment bound established above, we can now use the framework given by [5] to establish (8, main paper). However, we cannot use their proof directly, since the sample complexity $\widetilde{M}$ in the moment bound in (Equation A.14) is not oblivious to the matrix $\mathbf{U}$. However, once we fix the data matrix, we can adapt the argument used in [5] to show that (8, main paper) holds.

Let $\mathbf{W}$ and $\mathbf{Y}$ be as in (8, main paper). As explained in [5], we can assume that they have orthogonal columns. For a given k as in (8, main paper), let $\mathbf{W}$ and $\mathbf{Y}$ be partitioned into groups of k columns, with $\mathbf{W}_l$ and $\mathbf{Y}_{l'}$ denoting the l^{th} groups. [5] then use the following result in their argument, which follows from (Equation A.14):

$$\mathrm{E}\left\|(\mathbf{S}\mathbf{W}_l)^T(\mathbf{S}\mathbf{Y}_{l'}) - \mathbf{W}_l^T\mathbf{Y}_{l'}\right\|^p \leqslant \epsilon^p \left\|\mathbf{W}_l\right\|^p \left\|\mathbf{Y}_{l'}\right\|^p \delta \tag{A.15}$$

for all pairs (l, l'). This holds since in their setting, the sketching matrices are oblivious to the data matrices.

Although block diagonal matrices are not oblivious, this result holds with for $M_0 = \Omega\left(\frac{2k\log(2/\delta)}{\epsilon^2}\right)$. This is because of the observation that if $\mathbf{U}$ is an orthobasis for the span of $\mathbf{W}$ and $\mathbf{Y}$ and $\mathbf{U}^{l,l'}$ is an orthobasis for the span of $\mathbf{W}_l$ and $\mathbf{Y}_{l'}$, then

$$\Gamma(\mathbf{U}_j^{l,l'}) \leqslant \Gamma(\mathbf{U}_j) \tag{A.16}$$

for all pairs (l, l'). Hence, a given block diagonal sketching matrix $\mathbf{S}_D$ can satisfy (Equation A.15) as well. The rest of the proof remains the same as [5]. This concludes the proof for Theorem 1. Extending this to prove Theorem 2 is straightforward, with $\mathbf{S}_D$ being a particular case of their framework.

A.3 Proofs of results in chapter 4

A.3.1 Proof of Lemma 2

Let $(\mathbf{U}, \mathbf{V})$ be a minimizer to (4). Then by the feasibility it satisfies $\mathbf{X} = \mathbf{U}\mathbf{V}^\top$. Therefore

$$\begin{aligned}
\|\mathbf{X}\|_{1\to 2} &= \left\|\mathbf{U}\mathbf{V}^\top\right\|_{1\to 2} \\
&= \max_i \left\|\mathbf{U}\mathbf{V}^\top\mathbf{e}_i\right\|_2 \leqslant \max_i \|\mathbf{U}\| \left\|\mathbf{V}^\top\mathbf{e}_i\right\|_2 \\
&\leqslant \max_i \|\mathbf{U}\|_{\mathrm{F}} \left\|\mathbf{V}^\top\mathbf{e}_i\right\|_2 \\
&= \|\mathbf{U}\|_{\mathrm{F}} \|\mathbf{V}\|_{2,\infty} = \|\mathbf{X}\|_{\text{mixed}} .
\end{aligned}$$

Suppose that $(\mathbf{U}, \mathbf{V})$ satisfies $\mathbf{M} = \mathbf{U}\mathbf{V}^\top$ and $\mathbf{U}^\top\mathbf{U} = I_r$. Such $(\mathbf{U}, \mathbf{V})$ always exists.

For example, think about the SVD of $\mathbf{M}$. Since $(\mathbf{U}, \mathbf{V})$ is feasible to (3), it follows that

$$\|\mathbf{X}\|_{\text{mixed}} \leqslant \|\mathbf{U}\|_{\text{F}} \left\|\mathbf{V}^\top\right\|_{1\to 2} \leqslant \sqrt{r}\,\|\mathbf{U}\| \left\|\mathbf{V}^\top\right\|_{1\to 2} = \sqrt{r}\left\|\mathbf{V}^\top\right\|_{1\to 2}.$$

On the other hand,

$$\|\mathbf{X}\|_{1\to 2} = \max_i \left\|\mathbf{U}\mathbf{V}^\top e_i\right\|_2 = \max_i \left\|\mathbf{V}^\top e_i\right\|_2 = \left\|\mathbf{V}^\top\right\|_{1\to 2}.$$

We have shown

$$\|\mathbf{X}\|_{\text{mixed}} \leqslant \sqrt{r}\,\|\mathbf{X}\|_{1\to 2}.$$

In summary, we have

$$\|\mathbf{X}\|_{1\to 2} \leqslant \|\mathbf{X}\|_{\text{mixed}} \leqslant \sqrt{r}\,\|\mathbf{X}\|_{1\to 2}.$$

That is, the pair of $\|\cdot\|_{\text{mixed}}$ and $\|\cdot\|_{1\to 2}$ can be also used for a surrogate of the rank of a matrix.

A.3.2 Proof of Lemma 3

We derive a tail estimate on $\sup_{\mathbf{M}} \|\mathbf{Q}_{\mathbf{M}}\xi\|_2^2$ by using the results on suprema of chaos processes [68] summarized in the following theorem.

Theorem 13 (Theorem 3.1 in [68]) *Let $\xi \in \mathbb{R}^n$ be a Gaussian vector with $\mathrm{E}[\xi] = 0$ and $\mathrm{E}[\xi\xi^\top] = I_n$, $\Delta \subset \mathbb{R}^{m\times n}$, and $0 < \zeta < 1$. There exists a numerical constant C such that*

$$\begin{aligned}&\sup_{\mathbf{Q}\in\Delta} |\,\|\mathbf{Q}\xi\|_2^2 - \mathrm{E}[\|\mathbf{Q}\xi\|_2^2]|\\ &\leqslant C\left(E + V\sqrt{\log(2\zeta^{-1})} + U\log(2\zeta^{-1})\right)\end{aligned}$$

holds with probability $1-\zeta$, where

$$\begin{aligned}E &:= \gamma_2(\Delta, \|\cdot\|)\left[\gamma_2(\Delta, \|\cdot\|) + d_{\text{F}}(\Delta)\right],\\ V &:= d_{\text{S}}(\Delta)\left[\gamma_2(\Delta, \|\cdot\|) + d_{\text{F}}(\Delta)\right],\\ U &:= d_{\text{S}}^2(\Delta).\end{aligned}$$

We apply Theorem 14 to the set $\Delta = \{\mathbf{Q}_{\mathbf{M}} : \mathbf{M} \in \kappa(\alpha, R)\}$. The radii of Δ with respect to the Frobenius norm and to the spectral norm are respectively upper-bounded as follows:

$$d_{\mathrm{F}}(\Delta) \leqslant \alpha\sqrt{d_2}$$

and

$$d_{\mathrm{S}}(\Delta) \leqslant \frac{\alpha}{\sqrt{L}}.$$

Let B_{S} denote the unit ball with respect to the spectral norm. Then the γ_2-functional of Δ with respect to the spectral norm is upper-bounded through Dudley's inequality by

$$\begin{aligned}
\gamma_2(\Delta, \|\cdot\|_2) &\leqslant c\int_0^\infty \sqrt{\log N\left(\Delta, \eta B_{\mathrm{S}}\right)}\, d\eta \\
&\leqslant \frac{c}{\sqrt{L}} \int_0^\infty \sqrt{\log N\left(\kappa(\alpha, R), \eta B_{1\to 2}\right)}\, d\eta \\
&\leqslant \frac{c' R\sqrt{d}\log^{3/2} d}{\sqrt{L}},
\end{aligned}$$

where the last inequality follows from Lemma 4.

Then E, U, and V in Theorem 14 are upper-bounded by

$$\begin{aligned}
E &\leqslant \alpha R\sqrt{\frac{(d_1+d_2)d_2}{L}}\log^{3/2} d \\
&\quad + \frac{R^2}{L} d\log^3 d \\
U &\leqslant \frac{\alpha^2}{L} \\
V &\leqslant \frac{\alpha\sqrt{d_2}}{\sqrt{L}}\left(\frac{R\sqrt{d}}{Ld_2}\log^{3/2} d + \alpha\right).
\end{aligned}$$

By plugging in these upper estimates to Theorem 14, we obtain

$$\begin{aligned}
&\sup_{\mathbf{M}\in\kappa(\alpha,R)}\left|\frac{\|\mathbf{Q}_{\mathbf{M}}\xi\|^2}{d_2}-\frac{\|\mathbf{M}\|_{\mathrm{F}}^2}{d_2}\right| \\
&\leqslant c\left(\frac{\alpha R\sqrt{d}\log^{3/2}d}{\sqrt{Ld_2}}+\frac{R^2 d\log^3 d}{Ld_2}\right)+t \\
&\leqslant \frac{cR\sqrt{d}}{\sqrt{Ld_2}}\left(\alpha\log^{3/2}d+\frac{R\sqrt{d}\log^3 d}{\sqrt{Ld_2}}\right)+t
\end{aligned}$$

with probability at least $1-2\exp(-\hat{c}\min(t^2/V^2, t/U))$.

We take $t=\alpha R\sqrt{d}\log^{3/2}d/Ld_2$ not to increase the upper bound in order. This leads to the Lemma 2.

A.3.3 Upper bound on T_1 and T_2

A tail bound for T_1 can be derived by the following lemma [80], which is a direct consequence of the moment version of Dudley's inequality (e.g., p. 263 in [79]) and a version of Markov's inequality (e.g., Proposition 7.11 in [79]).

Lemma 19 *Let $\mu\in\mathbb{C}^n$ be a standard complex Gaussian vector with $\mathrm{E}\,\mu\mu^*=I_n$, and let $\Delta\subset\mathbb{C}^n$, $0<\zeta<e^{1/2}$. Then, there exists constant c such that*

$$\sup_{f\in\Delta}|f^*\mu|\leqslant c\sqrt{\log(\zeta^{-1})}\int_0^\infty\sqrt{\log N(\Delta,\|\cdot\|_2,t)}dt$$

with probability $1-\zeta$.

We apply Lemma 29 to the maximum of linear forms of a Gaussian vector $\mu=[\mathbf{b}_{1,1}^\top\ \cdots\ \mathbf{b}_{L,d_2}^\top]^\top$ over the set $\mathcal{F}=\{f_{\mathbf{M}}:\mathbf{M}\in\kappa(\alpha,R)\}$, where $f_{\mathbf{M}}$ is defined by

$$f_{\mathbf{M}}:=\left[\mathbf{1}_{1,L}\otimes(\mathbf{M}\mathbf{e}_1)^\top\quad\ldots\quad\mathbf{1}_{1,L}\otimes(\mathbf{M}\mathbf{e}_{d_2})^\top\right]^\top.$$

Here $\mathbf{1}_{1,L}$ is the row vector of length L with all entries set to 1. Then we have

$$\begin{aligned}
\|f_{\mathbf{M}}-f_{\mathbf{M}'}\|_2&=\|\mathbf{M}-\mathbf{M}'\|_{\mathrm{F}}\sqrt{L} \\
&\leqslant\|\mathbf{M}-\mathbf{M}'\|_{1\to2}\sqrt{Ld_2}.
\end{aligned}$$

Hence,

$$N(\mathcal{F}, \eta B_2) \leqslant N\left(\kappa(\alpha, R), \frac{\eta}{\sqrt{d_2}} B_\epsilon\right).$$

Combining these quantities and the entropy estimate for $N(\kappa(\alpha, R), \frac{\eta}{\sqrt{d_2}} B_\epsilon)$ with the above lemma, we get

$$\sup_{\mathbf{M} \in \kappa(\alpha, R)} \left| \sum_{l,i} \langle \mathbf{b}_{l,i}, \mathbf{M}\mathbf{e}_i \rangle \right| \leqslant c \log^{1/2} d \sqrt{L} R \sqrt{d} \sqrt{d_2}.$$

Using this, we get

$$T_1 = \mathrm{E}\, ||| \sum_{l,i} \nu_{l,i} \mathbf{A}_{l,i} |||_* \leqslant c\sigma \sqrt{d_2} R \sqrt{d} \log^{3/2} d$$

with probability at least $1 - 2\exp(-cd)$

Using Lemma 2, we have

$$T_2 \leqslant \alpha \sqrt{d_2 \left(\frac{cR}{\alpha} \sqrt{\frac{d_1 + d_2}{L d_2}} + 1 \right)} \log^3 d$$

Note that T_1 dominates T_2 when $L d_2 < d_1 d_2$. In this case, we conclude that

$$\frac{||| \sum_{l,i} \nu_{l,i} \mathbf{A}_{l,i} |||_*}{d_2} \leqslant c\sigma \sqrt{L} R \sqrt{\frac{d}{L d_2}} \log^3 d.$$

A.3.4 Entropy Estimates of Tensor Products

For symmetric convex bodies D and E, the *covering number* $N(D, E)$ and the *packing number* $M(D, E)$ are respectively defined by

$$N(D, E) := \min \left\{ l \,\middle|\, \exists y_1, \ldots, y_l \in D,\ D \subset \bigcup_{1 \leqslant j \leqslant l} (y_j + E) \right\},$$

$$M(D, E) := \max \left\{ l \,\middle|\, \exists y_1, \ldots, y_l \in D,\ y_j - y_k \notin E,\ \forall j \neq k \right\}.$$

Let X, Y be Banach spaces. For $T \in L(X, Y)$, the *dyadic entropy number* [81] is defined by

$$e_k(T) := \inf\{\epsilon > 0 \,|\, M(T(B_X), \epsilon B_Y) \leqslant 2^{k-1}\}.$$

where B_X and B_Y denote unit balls. We will use the following shorthand notation for the

weighted summation of the dyadic entropy numbers:

$$\mathcal{E}_{2,1}(T) := \sum_{k=1}^{\infty} k^{-1/2} e_k(T),$$

which is up to a constant equivalent to the entropy integral $\int_0^\infty \sqrt{\ln N(T(B_X), \epsilon B_Y)} d\epsilon$ [69], which plays a key role in analyzing properties on random linear operators on low-rank matrices.

In this section, we derive the $\mathcal{E}_{2,1}$ of the identity operator from the injective tensor product to the projective tensor product of a set of Banach space pairs. Note that these tensor product spaces are valid Banach spaces too. The main machinery in deriving these estimates is Maurey's empirical method [78], summarized in the following lemma.

Lemma 20 *Let $T \in \ell_\infty^{d_2} \otimes \ell_\infty^{d_1}$. Then*

$$\mathcal{E}_{2,1}(T) \leqslant C\sqrt{1 + \ln(d_1 \vee d_2)}\,(1 + \ln(d_1 \wedge d_2))^{3/2} \|T\|_\vee.$$

In order to apply Lemma 20 to $\ell_\infty^n \otimes \ell_\infty^m$, we use the fact that $\ell_\infty^m \check{\otimes} \ell_\infty^n$ is isometrically isomorphic to ℓ_∞^{mn}. In fact,

$$\|M\|_{\ell_\infty^m \check{\otimes} \ell_\infty^n} = \max_{1 \leqslant j \leqslant n} \|Me_j\|_\infty = \|\mathrm{vec}(M)\|_\infty,$$

where e_j denotes the jth column of the n-by-n identity matrix and $\mathrm{vec}(\cdot) : \mathbb{R}^{m \times n} \to \mathbb{R}^{mn}$ vectorizes $M \in \mathbb{R}^{m \times n}$ by stacking its columns vertically. On the other hand, the trace dual and Banach space dual of $\ell_1^n \hat{\otimes} \ell_1^m$ are $\ell_1^m \check{\otimes} \ell_1^n$ and $\ell_\infty^n \check{\otimes} \ell_\infty^m$, respectively. Therefore, we also have that $\ell_1^m \hat{\otimes} \ell_1^n$ is isotropically isomorphic to ℓ_1^{mn}. With these isometric isomorphisms, Lemma 20 provides the following estimate.

Proposition 1 *There exists a numerical constant C such that*

$$\mathcal{E}_{2,1}(\mathrm{id} : \ell_\infty^m \hat{\otimes} \ell_\infty^n \to \ell_\infty^m \check{\otimes} \ell_\infty^n) \leqslant C\sqrt{m+n}\,(1 + \ln(mn))^{3/2}.$$

To apply Lemma 20 to $\ell_\infty^{d_2} \otimes \ell_p^{d_1}$ with $2 \leqslant p < \infty$, we need the following result that shows embedding of finite dimensional ℓ_p space to ℓ_1 up to a small Banach-Mazur distance.

Lemma 21 ([78, Lemma 5]) *Let $1 < p \leqslant 2$ and $\epsilon > 0$. There is a constant $c(p, \epsilon) > 0$ for which the following property holds: For each d_1, there exists $k \geqslant c(p, \epsilon) d_1$ so that $\ell_1^{d_1}$ contains a subspace $(1 + \epsilon)$-isomorphic to ℓ_p^k, i.e., the Banach-Mazur distance is upper-bounded by $(1 + \epsilon)$.*

Then we obtain the following entropy estimate for $\ell_\infty^{d_2} \otimes \ell_p^{d_1}$ with $2 \leqslant p < \infty$ by combining Lemma 20 and Lemma 21.

Let $2 \leqslant p < \infty$. Then

$$\mathcal{E}_{2,1}(\mathrm{id} : \ell_\infty^{d_2} \mathbin{\widehat{\otimes}} \ell_p^{d_1} \to \ell_\infty^{d_2} \mathbin{\check{\otimes}} \ell_p^{d_1}) \leqslant C\sqrt{d_1 + d_2}\,(1 + \ln(d_1 d_2))^{3/2}.$$

A.3.5 Proof of Theorem 10

We use the following general strategy to establish the lower bound. We first show that there a packing set of $\kappa(\alpha, R)$ of a desirable size and packing density. We then use a multiway hypothesis testing argument and Fano's inequality to establish the minimax lower bound. We use the same argument as used in [55] to show that there is a packing set.

Lemma 22 *Let $r = R^2/\alpha^2$ and $\gamma \leqslant 1$ satisfy $r \leqslant \gamma^2(d_1 \wedge d_2)$ is an integer(?). Then there exists a subset $\mathcal{M} \subset \kappa(\alpha, R)$ with cardinality*

$$|\mathcal{M}| = \left\lceil \exp\left\{ \frac{r(d_1 \vee d_2)}{16\gamma^2} \right\} \right\rceil$$

with the following properties:

1. *Every $\mathbf{M} \in \mathcal{M}$ satisfies that $\mathrm{rank}(\mathbf{M}) \leqslant r/\gamma^2$ and $M_{kl} \in \{\pm\gamma\alpha/\sqrt{d_1}\}$ for all $k \in [d_1]$ and $l \in [d_2]$, thereby*

$$\|\mathbf{M}\|_{1\to 2} = \gamma\alpha, \quad \textit{and} \quad \|\mathbf{M}\|_{\mathrm{F}}^2 / d_2 = \gamma^2\alpha^2.$$

2. *Any two distinct $\mathbf{M}^i, \mathbf{M}^j \in \mathcal{M}$ satisfy*

$$\left\|\mathbf{M}^i - \mathbf{M}^j\right\|_{\mathrm{F}}^2 \geqslant \frac{\gamma^2\alpha^2 d_2}{2}.$$

Proof We adapt the proof of [55, Lemma 3.1] to our setting. The idea is to leverage the empirical method. Without loss of generality, we may assume that $d_2 \geqslant d_1$. (Otherwise we only need to flip d_1 and d_2 in the first assumption.) Let $N = \exp(rd_2/16\gamma^2)$, $B = r/\gamma^2$, and for each $i = 1, \cdots, N$, we draw a random matrix $\mathbf{M}^i$ as follows: The matrix $\mathbf{M}^i$ consists of i.i.d. blocks of dimensions $B \times d_2$, stacked up from top to bottom, with entries of the first block being i.i.d. symmetric random variables, taking values in $\pm\alpha\gamma/\sqrt{d_1}$ such that

$$M^i_{kl} = M^i_{k'l}, \quad \forall l, \forall k, k' : k' \equiv k \,(\mathrm{mod}\, B).$$

It can be verified that all the matrices $\mathbf{M}^1, \ldots, \mathbf{M}^N$ drawn in such a manner satisfy the first property above.

For any $\mathbf{M}^i \neq \mathbf{M}^j$, we have

$$\left\|\mathbf{M}^i - \mathbf{M}^j\right\|_{\mathrm{F}}^2 = \sum_{k,l}(M^i_{kl} - M^j_{kl})^2 \tag{A.17}$$

$$\geqslant \left\lfloor \frac{d_1}{B} \right\rfloor \sum_{k=1}^{B}\sum_{l=1}^{d_2}(M^i_{kl} - M^j_{kl})^2 \tag{A.18}$$

$$= \frac{4\alpha^2\gamma^2}{d_1}\left\lfloor \frac{d_1}{B} \right\rfloor \sum_{k=1}^{B}\sum_{l=1}^{d_2}\delta_{kl}, \tag{A.19}$$

where (δ_{kl}) denotes an array of i.i.d. Bernoulli random variables with mean$1/2$. Hoeffding's inequality implies

$$\mathbb{P}\left(\sum_{k=1}^{B}\sum_{l=1}^{d_2}\delta_{kl} \geqslant \frac{Bd_2}{4}\right) \leqslant e^{-Bd_2/8}.$$

By using the union bound argument over all $\binom{N}{2}$ possible pairs, we obtain that

$$\min_{i\neq j}\left\|\mathbf{M}^i - \mathbf{M}^j\right\|_{\mathrm{F}}^2 > \alpha^2\gamma^2\left\lfloor \frac{d_1}{B} \right\rfloor \frac{Bd_2}{d_1} \geqslant \frac{\alpha^2\gamma^2 d_2}{2}$$

holds with probability at least $1 - \binom{N}{2}\exp\left(-Bd_2/8\right) \geqslant 1/2$. In other words, the second property is satisfied with nonzero probability, thereby, there exists such an instance. This concludes the proof. ∎

Lemma 23 (Equivalence to multiple hypothesis testing) *We have*

$$\inf_{\widehat{\mathbf{M}}} \sup_{\mathbf{M} \in \kappa(\alpha, R)} \mathrm{E} \left\| \widehat{\mathbf{M}} - \mathbf{M} \right\|_{\mathrm{F}}^2 \geqslant \frac{\delta^2}{4} \min_{\widetilde{\mathbf{M}} \in \mathcal{M}} \mathbb{P} \left(\widetilde{\mathbf{M}} \neq \mathbf{M}^* \right), \tag{A.20}$$

where $\mathbf{M}^*$ *is uniformly distributed over the packing set* $\mathcal{M}$ *and* δ *is the packing density.*

Proof Let $\widetilde{\mathbf{M}} = \arg\min_{\mathbf{M} \in \mathcal{M}} \left\| \mathbf{M} - \widehat{\mathbf{M}} \right\|_{\mathrm{F}}$. If there exists $\mathbf{M}^j \in \mathcal{M}$ such that $\mathbf{M}^j \neq \widetilde{\mathbf{M}}$, then there exists $\mathbf{M}^k \in \mathcal{M}$ such that $\mathbf{M}^k \neq \mathbf{M}^j$ and

$$\left\| \widehat{\mathbf{M}} - \mathbf{M}^k \right\|_{\mathrm{F}} \leqslant \left\| \widehat{\mathbf{M}} - \mathbf{M}^j \right\|_{\mathrm{F}}. \tag{A.21}$$

By (Equation A.21) together with the triangle inequality, we obtain

$$\left\| \widehat{\mathbf{M}} - \mathbf{M}^j \right\|_{\mathrm{F}} \geqslant \left\| \mathbf{M}^j - \mathbf{M}^k \right\|_{\mathrm{F}} - \left\| \widehat{\mathbf{M}} - \mathbf{M}_k \right\|_{\mathrm{F}} \geqslant \left\| \mathbf{M}^j - \mathbf{M}^k \right\|_{\mathrm{F}} - \left\| \widehat{\mathbf{M}} - \mathbf{M}^j \right\|_{\mathrm{F}}, \tag{A.22}$$

which implies

$$\left\| \widehat{\mathbf{M}} - \mathbf{M}^j \right\|_{\mathrm{F}} \geqslant \frac{\left\| \mathbf{M}^j - \mathbf{M}^k \right\|_{\mathrm{F}}}{2}.$$

Thus, since every distinct pair in $\mathcal{M}$ is at least δ-separated, we obtain

$$\left\| \widehat{\mathbf{M}} - \mathbf{M}^j \right\|_{\mathrm{F}}^2 \geqslant \frac{\delta^2}{4}.$$

Hence we deduce that

$$\widetilde{\mathbf{M}} \neq \mathbf{M}^j \implies \left\| \widehat{\mathbf{M}} - \mathbf{M}^j \right\|_{\mathrm{F}}^2 \geqslant \frac{\delta^2}{4}. \tag{A.23}$$

Finally, we have

$$\begin{aligned}
\inf_{\widehat{\mathbf{M}}} \sup_{\mathbf{M} \in \kappa(\alpha, R)} \mathrm{E} \left\| \widehat{\mathbf{M}} - \mathbf{M} \right\|_{\mathrm{F}}^2 &\geqslant \inf_{\widehat{\mathbf{M}}} \max_{\mathbf{M}^j \in \mathcal{M}} \mathrm{E} \left\| \widehat{\mathbf{M}} - \mathbf{M}^j \right\|_{\mathrm{F}}^2 \\
&\overset{\text{(a)}}{\geqslant} \inf_{\widehat{\mathbf{M}}} \max_{\mathbf{M}^j \in \mathcal{M}} \frac{\delta^2}{4} \mathrm{P}\left(\right) \left(\left\| \widehat{\mathbf{M}} - \mathbf{M}^j \right\|_{\mathrm{F}}^2 > \frac{\delta^2}{4} \right) \\
&\overset{\text{(b)}}{\geqslant} \frac{\delta^2}{4} \min_{\widetilde{\mathbf{M}} \in \mathcal{M}} \max_{\mathbf{M}^j \in \mathcal{M}} \mathrm{P}\left(\right) \left(\widetilde{\mathbf{M}} \neq \mathbf{M}^j \right) \\
&\overset{\text{(c)}}{\geqslant} \frac{\delta^2}{4} \min_{\widetilde{\mathbf{M}} \in \mathcal{M}} \mathrm{P}\left(\right) \left(\widetilde{\mathbf{M}} \neq \mathbf{M}^* \right),
\end{aligned}$$

where $\mathbf{M}^*$ is drawn uniformly randomly from $\mathcal{M}$. Here (a) holds by Markov's inequality; (b) follows from (Equation A.23); and (c) holds since the worst-case error probability is

larger than the error probability with respect to the uniformly distributed random matrix $\mathbf{M}^*$. This completes the proof. ■

Lemma 24 (Application of Fano's inequality)

$$\mathbb{P}(\widetilde{\mathbf{M}} \neq \mathbf{M}^*) \geqslant 1 - \frac{\binom{|\mathbf{M}|}{2}^{-1} \sum_{k \neq j} \mathbb{E}_{\mathbf{b}_{l,i}} KL(\mathbf{M}^k || \mathbf{M}^j) + \log 2}{\log |\mathcal{M}|} \tag{A.24}$$

$$\tag{A.25}$$

where $KL(\mathbf{M}^k || \mathbf{M}^j)$ *denotes the KL divergence between the distributions of measurements* $y_{l,i}$*'s, conditioned on the measurement model Equation A.26 and* $\mathcal{M}$ *is the packing set derived in Lemma 22.*

KL divergence for our model

To apply Fano's inequality, we need to compute the KL divergence in (Equation A.24). Note that the measurements are generated as per the model in (Equation A.26):

$$y_{l,i} = \left\langle \frac{1}{\sqrt{L}} \mathbf{b}_{l,i} \mathbf{e}_i^\top, \mathbf{M} \right\rangle + z_{l,i} \tag{A.26}$$

Lemma 25 *The KL divergence between the distributions of the measurements generated using two different matrices* $\mathbf{M}^k$ *and* $\mathbf{M}^l$*, conditioned on the sensing vectors* $\mathbf{b}_{l,i}$*'s is given as*

$$\mathrm{KL}(\mathbf{M}^k || \mathbf{M}^j) = \frac{1}{2L\sigma^2} \sum_{l,i} \langle \mathbf{b}_{l,i}, \mathbf{m}_i^k - \mathbf{m}_i^l \rangle \tag{A.27}$$

and

$$\mathbb{E}_{\mathbf{b}_{l,i}} \mathrm{KL}(\mathbf{M}^k || \mathbf{M}^j) = \frac{1}{2\sigma^2} \|\mathbf{M}^k - \mathbf{M}^j\|_{\mathrm{F}}^2 \tag{A.28}$$

Proof We have

$$\mathbb{P}\left(Y | \{\mathbf{b}_{l,i}\}\right) = \prod_{l=1}^{L} \prod_{i=1}^{d_2} \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\left(Y_{l,i} - \langle \mathbf{b}_{l,i}, \mathbf{m}_i \rangle\right)^2 / 2L\sigma^2}.$$

Hence,

$$\mathrm{KL}(\mathbf{M}^k||\mathbf{M}^j) = \int \mathrm{P}\left(\right)(\mathbf{Y}|\mathbf{M}^k) \log\left(\frac{\mathrm{P}\left(\right)(\mathbf{Y}|\mathbf{M}^k)}{\mathrm{P}\left(\right)(\mathbf{Y}|\mathbf{M}^j)}\right) d\mathbf{Y} \tag{A.29}$$

$$= \int \prod_l \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\left(Y_{l,i} - \langle \mathbf{b}_{l,i}, \mathbf{m}_i^k \rangle^2\right)^2 / 2L\sigma^2} \tag{A.30}$$

$$\left[-\sum_l \sum_i \frac{(Y_{l,i} - \langle \mathbf{b}_{l,i}, \mathbf{m}_i^k \rangle^2)}{2L\sigma^2} + \sum_l \sum_i \frac{(Y_{l,i} - \langle \mathbf{b}_{l,i}, \mathbf{m}_i^j \rangle^2)}{2L\sigma^2} \right] d\mathbf{Y} \tag{A.31}$$

$$= \frac{1}{2L\sigma^2} \sum_l \sum_i \langle b_{l,i}, m_i^k - m_i^j \rangle^2 \tag{A.32}$$

We also can conclude that

$$\mathbb{E}_{b_{l,i}} \mathbf{KL}(M^k||M^j) = \frac{1}{2\sigma^2} \left\| M^k - M^j \right\|_{\mathrm{F}}^2 \tag{A.33}$$

■

Further note that $\left\| M^k - M^j \right\|_{\mathrm{F}}^2 \leqslant 4\alpha^2\gamma^2 d_2$ for any distinct k, j. Substituting this, (Equation A.33) and Lemma 22 in (Equation A.24), we get

$$\mathbb{P}(\widetilde{M} \neq M^*) \geqslant 1 - \frac{4\alpha^2\gamma^2 d_2}{2\sigma^2} \frac{16\gamma^2}{r(d_1 \vee d_2)} - \frac{16\gamma^2 \log 2}{r(d_1 \vee d_2)} \tag{A.34}$$

$$\geqslant \frac{1}{2} \tag{A.35}$$

provided that $\gamma^4 \leqslant \frac{\sigma^2 r(d_1 \vee d_2)}{128\alpha^2 d_2}$, and $r(d_1 \vee d_2) \geqslant 48$.

If $\frac{\sigma^2 r(d_1 \vee d_2)}{128\alpha^2 d_2} \geqslant 1$, we can choose $\gamma^2 = 1$. In this case, we obtain

$$\inf_{\widehat{M}} \sup_{M \in \kappa(\alpha, R)} \mathrm{E} \left\| \widehat{M} - M \right\|_{\mathrm{F}}^2 \geqslant \frac{\delta^2}{4} \frac{1}{2} \tag{A.36}$$

$$\geqslant \frac{\alpha^2 d_2}{16}, \tag{A.37}$$

since $\delta = \alpha\gamma\sqrt{d_2/2}$. Hence

$$\inf_{\widehat{M}} \sup_{M \in \kappa(\alpha, R)} \mathrm{E} \frac{1}{d_2} \left\| \widehat{M} - M \right\|_{\mathrm{F}}^2 \geqslant \frac{\delta^2}{4} \frac{1}{2} \tag{A.38}$$

$$\geqslant \frac{\alpha^2}{16}. \tag{A.39}$$

Otherwise, we choose $\gamma^2 = \sqrt{\frac{\sigma^2 r(d_1 \vee d_2)}{128\alpha^2 d_2}}$ and we obtain

$$\inf_{\widehat{M}} \sup_{M \in \kappa(\alpha,R)} \mathrm{E} \left\| \widehat{M} - M \right\|_{\mathrm{F}}^2 \geqslant \frac{\delta^2}{4}\frac{1}{2} \tag{A.40}$$

$$\geqslant \alpha^2 \sigma \sqrt{\frac{r(d_1 \vee d_2)}{128}} \frac{1}{\alpha\sqrt{d_2}} \frac{d_2}{2}\frac{1}{4} \tag{A.41}$$

$$\geqslant \alpha\sigma \sqrt{\frac{r(d_1 \vee d_2)}{128}} \frac{\sqrt{d_2}}{2}. \tag{A.42}$$

since $\delta = \alpha\gamma\sqrt{d_2/2}$. Hence

$$\inf_{\widehat{M}} \sup_{M \in \kappa(\alpha,R)} \mathrm{E} \frac{1}{d_2} \left\| \widehat{M} - M \right\|_{\mathrm{F}}^2 \geqslant \frac{\alpha^2}{16} \frac{\sigma\sqrt{L}}{\alpha} \sqrt{\frac{r(d_1 \vee d_2)}{L d_2}} \tag{A.43}$$

Finally, combining the two results, we have

$$\inf_{\widehat{M}} \sup_{M \in \kappa(\alpha,R)} \frac{1}{d_2} \mathrm{E} \left\| \widehat{M} - M \right\|_{F}^2 \geqslant \min \left(\frac{\alpha^2}{16}, \frac{\alpha^2}{16} \frac{\sigma\sqrt{L}}{\alpha} \sqrt{\frac{r(d_1 + d_2)}{L d_2}} \right) \tag{A.44}$$

A.4 Proofs of results in chapter 5

A.4.1 Proof of Theorem 11

Let $\mathbf{Z}_0 \widehat{\mathbf{D}}^\top$ and $\widehat{\mathbf{Z}}\widehat{\mathbf{D}}^\top$ denote the ground-truth matrix and its estimate, respectively. Due to optimality of $\widehat{\mathbf{Z}}$, we have

$$\left\| \widehat{\mathbf{y}} - \mathcal{A}(\widehat{\mathbf{Z}}\widehat{\mathbf{D}}^\top) \right\|_2^2 \leqslant \left\| \widehat{\mathbf{y}} - \mathcal{A}(\mathbf{Z}_0\widehat{\mathbf{D}}^\top) \right\|_2^2. \tag{A.45}$$

After substituting $\widehat{\mathbf{Z}} - \mathbf{Z}_0$ by $\mathbf{M}$, since $\widehat{\mathbf{y}} = \mathcal{A}(\mathbf{Z}_0\widehat{\mathbf{D}}) + \nu$, we obtain

$$\sum_{l=1}^{L} \langle \widehat{\mathbf{c}}_l \mathbf{e}_l^\mathsf{T}, \mathbf{M}\widehat{\mathbf{D}}^\top \rangle^2 \leqslant 2 \sum_{l=1}^{L} \langle \widehat{\mathbf{c}}_l \mathbf{e}_l^\mathsf{T}, \mathbf{M}\widehat{\mathbf{D}}^\top \rangle \nu_l. \tag{A.46}$$

We then use the following pair of results that provide and upper and lower bound on the right and left hand side of (Equation A.46).

Lemma 26 *For the setting in* (Equation A.46)*,*

$$\sum_{l} \langle \widehat{\mathbf{c}}_l \mathbf{e}_l^\mathsf{T}, \mathbf{M}\widehat{\mathbf{D}}^\top \rangle^2 \geqslant \left\| M\widehat{\mathbf{D}}^\top \right\|_F^2 - \alpha^2 \sqrt{N+K} \frac{\sqrt{L}}{\mu} \log^3(N+K) \sqrt{\log(2\zeta)^{-1}} \tag{A.47}$$

with probability at least $1-\zeta$

Lemma 27 *For the setting in* (Equation A.46)*,*

$$\sum_l \langle \mathbf{A}_l, \mathbf{M}\widehat{\mathbf{D}}^\top \rangle \nu_l \lesssim \sigma\alpha \frac{\sqrt{L}}{\mu}\sqrt{K+N}\sqrt{\log(1/\delta)} + \sigma\sqrt{\alpha^2\frac{L}{\mu^2} + \alpha^2\sqrt{N+K}\frac{\sqrt{L}}{\mu}}\sqrt{\log(1/\delta)\log(1/2\zeta)} \tag{A.48}$$

(up to log factors) with probability $1-\zeta-\delta$*.*

From Lemma 8, the upper bound on the linear term can be summarized as

$$\sum_l \langle \mathbf{A}_l, \mathbf{M}\widehat{\mathbf{D}}^\top \rangle \nu_l \lesssim \sigma\alpha \frac{\sqrt{L}}{\mu}\sqrt{K+N} + \sigma\sqrt{\alpha^2\frac{L}{\mu^2} + \alpha^2\sqrt{N+K}\frac{\sqrt{L}}{\mu}} \tag{A.49}$$

(up to log factors) with probability $1-\zeta-\delta$.

The lower bound on the quadratic term from Lemma 7 can be summarized as :

$$\sum_l \langle \mathbf{A}_l, \mathbf{M}\widehat{\mathbf{D}}^\top \rangle^2 \geqslant \left\|\mathbf{M}\widehat{\mathbf{D}}^\top\right\|_F^2 - \alpha^2\sqrt{N+K}\frac{\sqrt{L}}{\mu}\log^3(N+K)\sqrt{\log(2\zeta)^{-1}} \tag{A.50}$$

Combining the upper and lower bounds, we have (up to log factors)

$$\left\|\mathbf{M}\widehat{\mathbf{D}}^\top\right\|_F^2 \lesssim \alpha^2\sqrt{N+K}\frac{\sqrt{L}}{\mu} + \sigma\alpha\frac{\sqrt{L}}{\mu}\sqrt{K+N} + \sigma\sqrt{\alpha^2\frac{L}{\mu^2} + \alpha^2\sqrt{N+K}\frac{\sqrt{L}}{\mu}} \tag{A.51}$$

$$\lesssim \alpha\frac{\sqrt{L}}{\mu}\left(\alpha\sqrt{N+K} + \sigma\sqrt{K+N} + \sigma\sqrt{1+\frac{\mu\sqrt{K+N}}{\sqrt{L}}}\right) \tag{A.52}$$

$$\lesssim \alpha^2\frac{L}{\mu^2}\left(\frac{\mu\sqrt{K+N}}{\sqrt{L}} + \frac{\mu\sigma\sqrt{K+N}}{\alpha\sqrt{L}} + \frac{\mu\sigma}{\alpha\sqrt{L}}\sqrt{1+\frac{\mu\sqrt{K+N}}{\sqrt{L}}}\right) \tag{A.53}$$

$$\lesssim \alpha^2\frac{L}{\mu^2}\left(\frac{\mu\sqrt{K+N}}{\sqrt{L}} + \frac{\sigma\sqrt{L}}{\alpha\sqrt{L}/\mu}\sqrt{\frac{K+N}{L}} + \frac{\sigma\sqrt{L}}{\alpha\sqrt{L}/\mu}\sqrt{\frac{1}{L}+\frac{\mu\sqrt{K+N}}{L}}\right) \tag{A.54}$$

$$\lesssim \alpha^2 \frac{L}{\mu^2}\left(\frac{\mu\sqrt{K+N}}{\sqrt{L}} + \frac{\sigma\sqrt{L}}{\alpha\sqrt{L}/\mu}\left(\sqrt{\frac{K+N}{L}} + \sqrt{\frac{1}{L} + \frac{\mu\sqrt{K+N}}{L}}\right)\right) \tag{A.55}$$

$$\lesssim \alpha^2 \frac{L}{\mu^2}\left(\frac{\mu\sqrt{K+N}}{\sqrt{L}} + \frac{\sigma\sqrt{L}}{\alpha\sqrt{L}/\mu}\left(\sqrt{\frac{K+N}{L}} + \sqrt{\frac{\mu}{\sqrt{L}}\sqrt{\frac{K+N}{L}}}\right)\right) \tag{A.56}$$

$$\lesssim \alpha^2 \frac{L}{\mu^2}\left(\frac{\mu\sqrt{K+N}}{\sqrt{L}} + \frac{\sigma\sqrt{L}}{\alpha\sqrt{L}/\mu}\sqrt{\frac{K+N}{L}}\right). \tag{A.57}$$

where (Equation A.55) follows from the fact that $\mu\sqrt{K+N} > 1$ and t he last line follows from the assumption $\frac{K+N}{L} > \frac{\mu^2}{L}$. We then obtain the following result:

$$\left\|\mathbf{M}\widehat{\mathbf{D}}^\top\right\|_F^2 \lesssim \alpha^2 \frac{L}{\mu^2}\sqrt{\frac{K+N}{L}}(\mu + \frac{\sigma\sqrt{L}}{\alpha\sqrt{L}/\mu}). \tag{A.58}$$

By setting $\left\|\mathbf{Z}_0\widehat{\mathbf{D}}^\top\right\|_F = \frac{\alpha\sqrt{L}}{\mu}$, we get

$$\frac{\left\|M\widehat{\mathbf{D}}^\top\right\|_F^2}{\left\|\mathbf{Z}_0\widehat{\mathbf{D}}^\top\right\|_F^2} \lesssim \sqrt{\frac{K+N}{L}}\left(\mu + \frac{\sigma\sqrt{L}}{\left\|\mathbf{Z}_0\widehat{\mathbf{D}}^\top\right\|_F}\right) \tag{A.59}$$

holds with high probability.

A.4.2 Proof of Lemma 7

We observe that $\sum_{l=1}^{L}\langle\widehat{\mathbf{c}}_l\mathbf{e}_l^\mathsf{T}, \mathbf{M}\widehat{\mathbf{D}}^\top\rangle^2$ can be reformulated as a quadratic form in standard Gaussian random variables. Let us define

$$\xi = \begin{bmatrix}\widehat{\mathbf{c}}_1 \\ \vdots \\ \widehat{\mathbf{c}}_L\end{bmatrix} \in \mathbb{C}^{LN}. \tag{A.60}$$

Then it follows that $\xi \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{LN})$. Therefore, the left-hand side of (Equation A.46) can be rewritten as

$$\sum_{l=1}^{L}\langle\widehat{\mathbf{c}}_l\mathbf{e}_l^\mathsf{T}, \mathbf{M}\widehat{\mathbf{D}}^\top\rangle^2 = \left\|\mathbf{Q}_\mathbf{M}\xi\right\|^2,$$

where

$$\mathbf{Q}_{\mathbf{M}} = \begin{bmatrix} (\mathbf{M}\widehat{\mathbf{D}}^{\top}\mathbf{e}_1)^{\top} & 0 & \cdots & 0 \\ 0 & (\mathbf{M}\widehat{\mathbf{D}}^{\top}\mathbf{e}_2)^{\top} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (\mathbf{M}\widehat{\mathbf{D}}^{\top}\mathbf{e}_{d_2})^{\top} \end{bmatrix}. \tag{A.61}$$

We also have

$$\mathrm{E}\,\|\mathbf{Q}_{\mathbf{M}}\xi\|^2 = \left\|\mathbf{M}\widehat{\mathbf{D}}^{\top}\right\|_{\mathrm{F}}^2$$

We compute a tail estimate on the supremum of $\|\mathbf{Q}_{\mathbf{M}}\xi\|_2^2$ over the set $\{\mathbf{M} : \mathbf{M}\hat{B}^* \in \alpha B_\pi\}$ where B_π is the π-norm unit ball by using the results on suprema of chaos processes [77]. They derived a sharp tail estimate on the supremum of a Gaussian quadratic form maximized over a given set $\mathcal{S}$, stated below.

Theorem 14 (Theorem 3.1 in [77]) *Let $\xi \in \mathbb{R}^n$ be a Gaussian vector with $\mathrm{E}[\xi] = 0$ and $\mathrm{E}[\xi\xi^{\top}] = I_n$, $\Delta \subset \mathbb{R}^{m\times n}$, and $0 < \zeta < 1$. There exists a numerical constant C such that*

$$\sup_{\mathbf{Q}\in\Delta} |\, \|\mathbf{Q}\xi\|_2^2 - \mathrm{E}[\|\mathbf{Q}\xi\|_2^2]| \leqslant C\left(E + V\sqrt{\log(2\zeta^{-1})} + U\log(2\zeta^{-1})\right)$$

holds with probability $1-\zeta$, where

$$\begin{aligned} E &:= \gamma_2(\Delta, \|\cdot\|)\left[\gamma_2(\Delta, \|\cdot\|) + d_{\mathrm{F}}(\Delta)\right], \\ V &:= d_{\mathrm{S}}(\Delta)\left[\gamma_2(\Delta, \|\cdot\|) + d_{\mathrm{F}}(\Delta)\right], \\ U &:= d_{\mathrm{S}}^2(\Delta). \end{aligned}$$

where d_F is the diameter of Δ in the Frobenius norm, d_{S} is the diameter in the spectral norm and γ_2 is the gamma funcional.

We can use Theorem 14 to bound $\|\mathbf{Q}\xi\|^2$ by defining $\Delta = \{\mathbf{Q}_{\mathbf{M}} : \left\|\mathbf{M}\widehat{\mathbf{D}}^{\top}\right\|_{\pi} \leqslant \alpha\}$. The diameters and the gamma functional for the set Δ are computed below.

Spectral radius of $\mathcal{S}$:

$$d_s(\Delta) = \sup_{\mathbf{Q}_{\mathbf{M}}\in\Delta} \|\mathbf{Q}_{\mathbf{M}}\|_{2\to 2} = \left\|\mathbf{M}\widehat{\mathbf{D}}^{\top}\right\|_{1\to 2} \leqslant \left\|\mathbf{M}\widehat{\mathbf{D}}^{\top}\right\|_{\epsilon} \leqslant \alpha. \tag{A.62}$$

Frobenius radius:

$$d_F(\Delta) \leqslant \left\|\mathbf{M}\widehat{\mathbf{D}}^\top\right\|_F \leqslant \left\|\mathbf{M}\widehat{\mathbf{D}}^\top\right\|_* \leqslant \frac{\alpha\sqrt{L}}{\mu} \tag{A.63}$$

γ_2-functional: To bound the gamma functional, we use the following upper bound [77]:

$$\gamma_2 \lesssim \int_0^\infty \sqrt{\log(\alpha B_\pi, \eta B_\epsilon)} d\eta. \tag{A.64}$$

We provide an upper bound on the right hand side of (Equation A.64) in Lemma Equation 28. We defer the proof of Lemma Equation 28 to subsection A.4.4.

Lemma 28 *The γ_2-functional of the set $\Delta = \{\mathbf{Q}_\mathbf{M} : \|\mathbf{M}\widehat{\mathbf{D}}^\top\|_\pi \leqslant \alpha\}$ with respect to the spectral norm is upper-bounded by*

$$\gamma_2(\Delta) \lesssim \int_0^\infty \sqrt{\log(\alpha B_\pi, \eta B_\epsilon)} d\eta \lesssim \alpha\sqrt{N+K}\log^{3/2} L. \tag{A.65}$$

We can now substitute for these quantities in Theorem 14:

$$E := \gamma_2(\Delta, \|\cdot\|)\left[\gamma_2(\Delta, \|\cdot\|) + d_{\mathrm{F}}(\Delta)\right] \lesssim \alpha^2(N+K)\log^3(N+K) + \alpha^2\sqrt{N+K}\frac{\sqrt{L}}{\mu}$$

$$V := d_{\mathrm{S}}(\Delta)\left[\gamma_2(\Delta, \|\cdot\|) + d_{\mathrm{F}}(\Delta)\right] \lesssim \alpha^2\sqrt{N+K}\log^{3/2}(N+K) + \alpha^2\frac{\sqrt{L}}{\mu}$$

$$U := d_{\mathrm{S}}^2(\Delta) \leqslant \alpha^2.$$

Hence,

$$\begin{aligned}\sup_{\mathbf{Q}_\mathbf{M}\in\Delta} |\, \|\mathbf{Q}_\mathbf{M}\xi\|_2^2 - \mathrm{E}\,\|\mathbf{Q}_\mathbf{M}\xi\|_2^2 \,| &\lesssim E + V\sqrt{\log(2\zeta^{-1})} + U\log(2\zeta^{-1}) \\ &\lesssim \alpha^2(N+K)\log^3(N+K) + \alpha^2\sqrt{N+K}\frac{\sqrt{L}}{\mu}\sqrt{\log(2\zeta)^{-1}})\end{aligned} \tag{A.66}$$

with probability at least $1-\zeta$. We ignore the last term, since it is dominated by the other

two terms. We can then obtain the bound in Lemma 7 as below:

$$\sum_l \langle \widehat{\mathbf{c}}_l \mathbf{e}_l^\mathsf{T}, \mathbf{M}\widehat{\mathbf{D}}^\top \rangle^2 \gtrsim \left\| M\widehat{\mathbf{D}}^\top \right\|_F^2 - \alpha^2 (N+K)\log^3(N+K) - \alpha^2\sqrt{N+K}\frac{\sqrt{L}}{\mu}\sqrt{\log(2\zeta)^{-1}})$$

$$\gtrsim \left\| M\widehat{\mathbf{D}}^\top \right\|_F^2 - \alpha^2\sqrt{N+K}\frac{\sqrt{L}}{\mu}\log^3(N+K)\sqrt{\log(2\zeta)^{-1}}$$

where the last line follows from the fact that $L > N + K$.

A.4.3 Proof of Lemma 8

In order to prove Lemma 8, we will make use of the following inequality from [69]:

Theorem 15 (Theorem 4.7, [69]) *Let g_k be an i.i.d sequence of Gaussian normal random variables on some probability space $(\Omega, \mathcal{Y}, \mathbf{P})$. Let E be a Banach space and let (z_k) be a sequence of elements of E. Let $X = \sum_{k=1}^{\infty} g_k z_k$ and let*

$$\sigma(X) = \sup\left\{ (\sum |\zeta(z_k)|^2)^{1/2} | \zeta \in E^*, \|\zeta\| \leqslant 1 \right\}. \tag{A.67}$$

Then,

$$\forall t > 0, \mathrm{P}\left(|\, \|X\| - \mathrm{E}\,\|X\|\,| > t \right) \leqslant 2\exp(-Kt^2/\sigma(X)^2), \tag{A.68}$$

where K is a numerical constant.

In order to apply this theorem to our problem, we can define the corresponding quantities in the context of our problem. Conditioning on the $\widehat{\mathbf{c}}_l$'s, the entries of the noise vector, ν_l's correspond to g_k's, $\widehat{\mathbf{c}}_l\mathbf{e}_l^\mathsf{T}$'s correspond to the z_k's. Further, let us define the following matrix norm:

$$|||\mathbf{M}||| = ||\mathbf{M}\widehat{\mathbf{D}}^\top||_\pi/\alpha. \tag{A.69}$$

Then, we characterize the linear form in (Equation A.46) as

$$\sup_{\mathbf{M}:\mathbf{M}\widehat{\mathbf{D}}^\top \in \alpha B_\pi} \sum_l \langle \widehat{\mathbf{c}}_l \mathbf{e}_l^\mathsf{T}, \mathbf{M}\widehat{\mathbf{D}}^\top \rangle \nu_l = |||\sum_l \widehat{\mathbf{c}}_l \mathbf{e}_l^\mathsf{T} \nu_l|||_*.$$

Under this setting, from Theorem Equation 15, we have

$$|||\sum_l \widehat{\mathbf{c}}_l \mathbf{e}_l^\mathsf{T} \nu_l|||_* \leqslant \underbrace{\mathrm{E}_\nu \, |||\sum_l \widehat{\mathbf{c}}_l \mathbf{e}_l^\mathsf{T} \nu_l|||_*}_{\text{T1}} + \underbrace{\sigma\pi \sqrt{c\log(2/\delta) \sup_{\mathbf{M}:\mathbf{M}\widehat{\mathbf{D}}^\top \in \alpha B_\pi} \sum_{l=1}^{L} \langle \widehat{\mathbf{c}}_l \mathbf{e}_l^\mathsf{T}, \mathbf{M}\widehat{\mathbf{D}}^\top \rangle^2}}_{\text{T2}} \tag{A.70}$$

holds with probability $1-\delta$.

Bound on T1:

In order to bound **the first term**T1, we use the equivalence (up to a log factor) between the Gaussian and Rademacher complexities of any set. Since the first term in (Equation A.70) is the Gaussian complexity of the set of $\widehat{\mathbf{c}}_l \mathbf{e}_l^\mathsf{T}$'s, we can replace it with the Rademacher complexity:

$$\begin{aligned}
\mathrm{E}_\nu \, |||\sum_l \widehat{\mathbf{c}}_l \mathbf{e}_l^\mathsf{T} \nu_l|||_* &\leqslant c\sigma\sqrt{\log L}\, \mathrm{E}_r \, |||\sum_l \widehat{\mathbf{c}}_l \mathbf{e}_l^\mathsf{T} r_l|||_* \\
&\leqslant c\sigma\sqrt{\log L}\, \mathrm{E}_r \sup_{M:\mathbf{M}\widehat{\mathbf{D}}^\top \in \alpha B_\pi} |\sum_l \langle r_l \widehat{\mathbf{c}}_l \mathbf{e}_l^\mathsf{T}, \mathbf{M}\widehat{\mathbf{D}}^\top \rangle|,
\end{aligned}$$

where r_l's are i.i.d. Rademacher random variables. Due to symmetry of $\widehat{\mathbf{c}}_l$, we have

$$|\sum_l \langle r_l \widehat{\mathbf{c}}_l \mathbf{e}_l^\mathsf{T}, \mathbf{M}\widehat{\mathbf{D}}^\top \rangle| \overset{d}{\sim} |\sum_l \langle \widehat{\mathbf{c}}_l \mathbf{e}_l^\mathsf{T}, \mathbf{M}\widehat{\mathbf{D}}^\top \rangle|. \tag{A.71}$$

Hence, it suffices if we bound $\sup_{M:\mathbf{M}\widehat{\mathbf{D}}^\top \in \alpha B_\pi} |\sum_l \langle \widehat{\mathbf{c}}_l \mathbf{e}_l^\mathsf{T}, \mathbf{M}\widehat{\mathbf{D}}^\top \rangle|$. Note that

$$|\sum_l \langle \widehat{\mathbf{c}}_l \mathbf{e}_l^\mathsf{T}, \mathbf{M}\widehat{\mathbf{D}}^\top \rangle| = |\langle \mathrm{vec}(\mathbf{M}\widehat{\mathbf{D}}^\top), \mathrm{vec}(\{\widehat{\mathbf{c}}_l\}) \rangle| = |\langle f_M, \mathrm{vec}(\{\widehat{\mathbf{c}}_l\}) \rangle| \tag{A.72}$$

where $f_M := \mathrm{vec}(\mathbf{M}\widehat{\mathbf{D}}^\top)$. It can be characterized as a mean-zero Gaussian process over the set $\{\mathbf{M} : \mathbf{M}\widehat{\mathbf{D}}^\top \in \alpha B_\pi\}$. A tail bound can now be derived by the following lemma [80], which is a direct consequence of the moment version of Dudley's inequality (e.g., p. 263 in [79]) and a version of Markov's inequality (e.g., Proposition 7.11 in [79]).

Lemma 29 *Let $\mu \in \mathbb{C}^n$ be a standard complex Gaussian vector with $\mathrm{E}\,\mu\mu^* = I_n$, and let $\Delta \subset \mathbb{C}^n$, $0 < \zeta < e^{1/2}$. Then, there exists constant c such that*

$$\sup_{f\in\Delta} |f^*\mu| \leqslant c\sqrt{\log(\zeta^{-1})} \int_0^\infty \sqrt{\log N(\Delta, \|\cdot\|_2, t)} dt$$

with probability $1-\zeta$.

With f_M defined as in (Equation A.72), we have

$$\|f_M\|_2 = \left\|M\widehat{\mathbf{D}}^\top\right\|_F \leqslant \alpha\frac{\sqrt{L}}{\mu} \tag{A.73}$$

$$\begin{aligned}\int_0^\infty \sqrt{\log N\left(\alpha B_\pi, \frac{\eta}{\sqrt{K}}B_\epsilon\right)}d\eta &= \alpha\frac{\sqrt{L}}{\mu}\int_0^\infty \sqrt{\log N\left(B_\pi, \eta B_\epsilon\right)}d\eta \\ &\lesssim \alpha\frac{\sqrt{L}}{\mu}\sqrt{K+N}\log^{3/2}(K+N).\end{aligned}$$

From Lemma 29, we have

$$\sup_{M:M\hat{B}^*\in\alpha B_\pi} |\sum_l \langle \widehat{\mathbf{c}}_l\mathbf{e}_l^\mathsf{T}, \mathbf{M}\widehat{\mathbf{D}}^\top\rangle| \lesssim \sqrt{\log(1/\delta)}\alpha\frac{\sqrt{L}}{\mu}\sqrt{K+N}\log^{3/2}(K+N) \tag{A.74}$$

holds with probability $1-\delta$. Owing to (Equation A.71), we can conclude that

$$\sup_{M:M\hat{B}^*\in\alpha B_\pi} |\sum_l \langle r_l\widehat{\mathbf{c}}_l\mathbf{e}_l^\mathsf{T}, \mathbf{M}\widehat{\mathbf{D}}^\top\rangle| \lesssim \sqrt{\log(1/\delta)}\alpha\frac{\sqrt{L}}{\mu}\sqrt{K+N}\log^{3/2}(K+N) \tag{A.75}$$

holds with probability $1-\delta$. Therefore,

$$\mathrm{E}_r \sup_{M:M\widehat{\mathbf{D}}^\top\in\alpha B_\pi} |\sum_l \langle r_l\widehat{\mathbf{c}}_l\mathbf{e}_l^\mathsf{T}, \mathbf{M}\widehat{\mathbf{D}}^\top\rangle| \lesssim \sigma\sqrt{\log(1/\delta)}\alpha\frac{\sqrt{L}}{\mu}\sqrt{K+N}\log^{3/2}(K+N) \tag{A.76}$$

holds with probability $1-\delta$.

Hence,

$$\mathrm{E}_\nu |||\sum_l \widehat{\mathbf{c}}_l\mathbf{e}_l^\mathsf{T}\nu_l|||_* \lesssim \sigma\alpha\sqrt{K+N}\frac{\sqrt{L}}{\mu}\sqrt{\log(1/\delta)} \tag{A.77}$$

holds with probability $1-\delta$.

Bound on T2: From (Equation A.66), we also have

$$\sum_{l=1}^L \langle \mathbf{c}_l\mathbf{e}_l^\top, \mathbf{M}\widehat{\mathbf{D}}^\top\rangle^2 \leqslant \left\|\mathbf{M}\widehat{\mathbf{D}}^\top\right\|_F^2 + \alpha^2\sqrt{N+K}\frac{\sqrt{L}}{\mu}\log^3(N+K)\sqrt{\log(2\zeta)^{-1}} \tag{A.78}$$

$$\leqslant \alpha^2\frac{L}{\mu^2} + \alpha^2\sqrt{N+K}\frac{\sqrt{L}}{\mu}\log^3(N+K)\sqrt{\log(2\zeta)^{-1}} \tag{A.79}$$

with probability at least $1-\zeta$. Hence **T2** is upper-bounded as

$$\sigma\pi\sqrt{c\log(2/\delta)\sup_{M:\mathbf{M}\widehat{\mathbf{D}}^\top\in\alpha B_\pi}\sum_{l=1}^{L}\langle\mathbf{A}_l,\mathbf{M}\widehat{\mathbf{D}}^\top\rangle^2}\lesssim\sigma\sqrt{\alpha^2\frac{L}{\mu^2}+\alpha^2\sqrt{N+K}\frac{\sqrt{L}}{\mu}} \tag{A.80}$$

up to a logarithmic factor with probability $1-\zeta-\delta$.

A.4.4 Proof of Lemma 10

Recall that the γ_2-functional in the left-hand side of (Equation A.65) is upper-bounded by Dudley's integral, i.e.

$$\gamma_2(\Delta,\|\cdot\|)\lesssim\int_0^\infty\sqrt{\log N(\Delta,\|\cdot\|,t)}dt. \tag{A.81}$$

Note that, for $\mathbf{Q}_\mathbf{M},\mathbf{Q}_{\mathbf{M}'}\in\Delta$, we have

$$\|\mathbf{Q}_\mathbf{M}-\mathbf{Q}_{\mathbf{M}'}\|_2=\left\|\mathbf{M}\widehat{\mathbf{D}}^\top-\mathbf{M}'\widehat{\mathbf{D}}^\top\right\|_{1\to2}\leqslant\left\|\mathbf{M}\widehat{\mathbf{D}}^\top-\mathbf{M}'\widehat{\mathbf{D}}^\top\right\|_\epsilon,$$

where the inequality follows from the definition of $\|\cdot\|_\epsilon$ in (Equation 5.15). Therefore, by the injectivity of the map $\mathbf{M}\mapsto\mathbf{M}\widehat{\mathbf{D}}^\top$, it follows that

$$N(\Delta,\|\cdot\|,t)\leqslant N(\alpha B_\pi,\|\cdot\|_\epsilon,t).$$

Plugging in this to (Equation A.81) provides

$$\gamma_2(\Delta,\|\cdot\|)\lesssim\int_0^\infty\sqrt{\log N(\alpha B_\pi,\|\cdot\|_\epsilon,t)}dt. \tag{A.82}$$

Recall that the dyadic entropy number of a linear operator $T:X\to Y$ between two Banach spaces X and Y is given by

$$e_k(\mathcal{T}:X\to Y)=\inf\left\{\epsilon>0\,:\,\exists y_1,\ldots,y_{2^{k-1}}\in Y\text{ s.t. }\mathcal{T}(B_X)\in\bigcup_{1\leqslant l\leqslant 2^{k-1}}(y_l+\epsilon B_Y)\right\}. \tag{A.83}$$

Indeed, the right-hand side of (Equation A.82) is equivalent (up to a numerical constant) to the following weighted sum of dyadic entropy numbers:

$$\mathcal{E}_{2,1}(\mathrm{id}:X_\pi\to X_\epsilon):=\sum_{k=0}^{\infty}\frac{e_k(\mathrm{id}:X_\pi\to X_\epsilon)}{\sqrt{k}}. \tag{A.84}$$

Therefore it suffices to derive an upper bound on $\mathcal{E}_{2,1}(\mathrm{id} : X_\pi \to X_\epsilon)$.

Maurey's empirical method

To derive an upper bound on $\mathcal{E}_{2,1}(\mathrm{id} : X_\pi \to X_\epsilon)$, we make use of Maurey's empirical method that upper-bounds the dyadic entropy numbers of a linear operator from ℓ_1^n to ℓ_∞^m, summarized in the following lemma.

Lemma 30 ([78], [82, Lemma 3.4]) *Let $\mathcal{T} : \ell_1^n \to \ell_\infty^m$ be a bounded linear operator. Then we have*

$$\mathcal{E}_{2,1}(\mathcal{T}) \lesssim \sqrt{1 + \log(m \vee n)}\,(1 + \log(m \wedge n))^{3/2}\,\|\mathcal{T}\|_{\mathrm{op}}\,.$$

Our strategy is to embed the domain X_π (resp. co-domain X_ϵ) of $\mathcal{T} = \mathrm{id}$ into ℓ_1^n (resp. ℓ_∞^m) for some suitable $n, m \in \mathbb{N}$. To simplify notation, we introduce two Banach spaces defined by

$$W_\infty := \left(\mathbb{C}^K, \left\|\overline{\mathbf{FD}}\right\|_\infty\right) \tag{A.85}$$

$$W_2 := \left(\mathbb{C}^K, \left\|\overline{\mathbf{FD}}\right\|_2\right). \tag{A.86}$$

Note that $\overline{\mathbf{FD}}$ is an isometry from W_∞ (resp. W_2) to V_∞ (resp. V_2). Furthermore, $\widehat{\mathbf{D}}^\top$ is an isometry from V_∞ (resp. V_2) to W_∞ (resp. W_2). It follows that W_∞ and V_∞ (resp. W_2 and V_2) are equivalent via an isometric bijection. Therefore we have

$$e_k(\mathrm{id} : X_\pi \to X_\epsilon) = e_k\left(\mathrm{id} : W_\infty \otimes_\pi \ell_2^N \cap \frac{\mu}{\sqrt{L}} W_2 \otimes_\pi \ell_2^N \to W_\infty \otimes_\epsilon \ell_2^N \cap \frac{\mu}{\sqrt{L}} W_2 \otimes_\epsilon \ell_2^N\right). \tag{A.87}$$

In the remainder, we derive an upper bound on the right-hand side of (Equation A.87). We achieve this via a series of embedding maps as shown in the commutative diagram in Figure A.1.

We proceed by deriving the operator norm of each of the embedding maps in the commutative diagram.

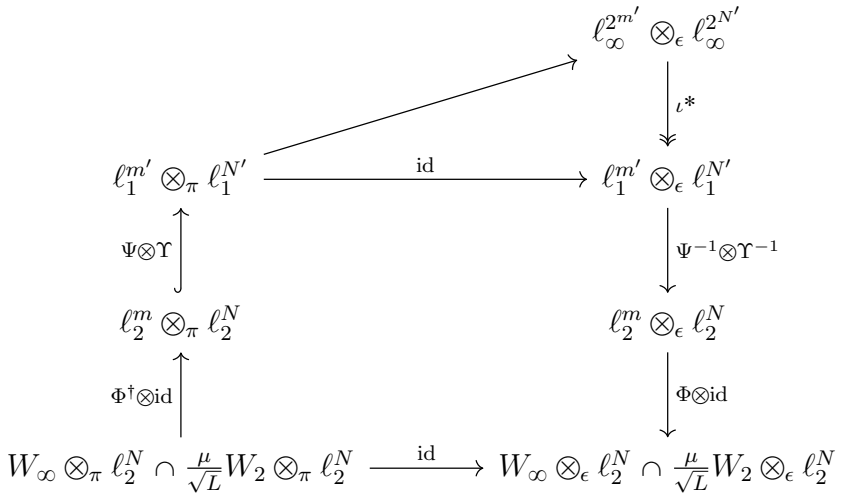


Figure A.1: *Commutative diagram*

Embedding $W_\infty \otimes_\pi \ell_2^N \cap \frac{\mu}{\sqrt{L}} W_2 \otimes_\pi \ell_2^N$ *into* $\ell_2^m \otimes_\pi \ell_2^N$

Our first map embeds $W_\infty \otimes_\pi \ell_2^N \cap \frac{\mu}{\sqrt{L}} W_2 \otimes_\pi \ell_2^N$ into $\ell_2^m \otimes_\pi \ell_2^N$ for a suitable $m > k$. We construct such an embedding map by using a random matrix whose entries are i.i.d. with respect to the standard normal distribution. The following lemma presents the construction of the embedding map and an upper bound on its operator norm.

Lemma 31 *Let*

$$\mathbf{\Phi} = \frac{1}{\sqrt{m}} \widehat{\mathbf{D}}^\top \mathbf{G}^*, \tag{A.88}$$

where $\mathbf{G}$ *is an* $m \times L$ *matrix with i.i.d. entries drawn from* $\mathcal{N}(0,1)$*. Then there exists a numerical constant* C *such that*

$$\left\| \mathbf{\Phi}^\dagger \otimes \mathrm{id} : W_\infty \otimes_\pi \ell_2^N \cap \frac{\mu}{\sqrt{L}} W_2 \otimes_\pi \ell_2^N \to \ell_2^m \otimes_\pi \ell_2^N \right\|_{\mathrm{op}} \leqslant \frac{\sqrt{L}}{\mu} \left(\frac{\sqrt{1+\delta}}{1-\delta} \right)$$

holds with probability $1 - 4\exp(-\delta^2 m)$ *provided*

$$m \geqslant \delta^{-2} K.$$

Proof Note that $\hat{\mathbf{D}}^\top \hat{\mathbf{D}} = \mathbf{I}_K$. Due to the construction of $\mathbf{\Phi}$, it has entries drawn from the complex Gaussian distribution.

Lemma 32 ([83, Theorem 5.65]) *Let* $\mathbf{S} = \frac{1}{\sqrt{m}}\mathbf{R}$*, where* $\mathbf{R} \in \mathbb{R}^{m\times d}$*, with* $\mathbf{R}_{i,j}$ *drawn identically and independently from a sub-Gaussian distribution. If* $m \geqslant d/\delta^2$*, then*

$$\left\|\mathbf{I} - \mathbf{S}^\top\mathbf{S}\right\| \leqslant \delta \tag{A.89}$$

with probability at least $1 - 2\exp(\delta^2 m)$

From Lemma 32, we deduce that

$$\|\mathbf{I} - \mathbf{\Phi}\mathbf{\Phi}^*\| \leqslant \delta \tag{A.90}$$

holds with probability at least $1 - 2\exp(\delta^2 m)$. This also implies

$$\|\mathbf{\Phi}\| \leqslant \sqrt{1+\delta}$$

with probability at least $1 - 2\exp(\delta^2 m)$.

We now derive an upper bound on the operator norm of $\mathbf{\Phi}^\dagger \otimes \mathrm{id}$ in terms of the the operator norms of the following operators:

$$\mathbf{\Phi}^\dagger : W_\infty \cap \frac{\mu}{\sqrt{L}}W_2 \to \ell_2^m \quad \text{and} \quad \mathrm{id} : \ell_2^N \to \ell_2^N.$$

The operator norm is rewritten as

$$\begin{aligned}
&\left\|\mathbf{\Phi}^\dagger \otimes \mathrm{id} : W_\infty \otimes_\pi \ell_2^N \cap \frac{\mu}{\sqrt{L}}W_2 \otimes_\pi \ell_2^N \to \ell_2^m \otimes_\pi \ell_2^N\right\| \\
&= \left\|\mathbf{\Phi}^\dagger \otimes \mathrm{id} : W_\infty \otimes_\pi \ell_2^N \to \ell_2^m \otimes_\pi \ell_2^N\right\| \wedge \left\|\mathbf{\Phi}^\dagger \otimes \mathrm{id} : \frac{\mu}{\sqrt{L}}W_2 \otimes_\pi \ell_2^N \to \ell_2^m \otimes_\pi \ell_2^N\right\| \\
&= \left\|\mathbf{\Phi}^\dagger : W_\infty \to \ell_2^m\right\| \wedge \left\|\mathbf{\Phi}^\dagger : \frac{\mu}{\sqrt{L}}W_2 \to \ell_2^m\right\| \\
&= \left\|\mathbf{\Phi}^\dagger : W_\infty \cap \frac{\mu}{\sqrt{L}}W_2 \to \ell_2^m\right\|.
\end{aligned}$$

where the second and the last line follows from the norm associated with the interpolation space, and the third line is due to the following:

$$\|\mathcal{T} \otimes \mathcal{T}'\|_{\mathrm{op}} = \|\mathcal{T}\|_{\mathrm{op}} \cdot \|\mathcal{T}'\|_{\mathrm{op}}$$

Thus it suffices to derive an upper bound on the operator norm of $\mathbf{\Phi}^\dagger$. Since the norm

dual of W_∞ is W_1 and W_2 is self-dual, we obtain

$$\begin{aligned}
\left\|\mathbf{\Phi}^\dagger\right\|_{\text{op}} &= \left\|\mathbf{\Phi}^*(\mathbf{\Phi}\mathbf{\Phi}^*)^{-1} : W_\infty \cap \frac{\mu}{\sqrt{L}}W_2 \to \ell_2^m\right\|_{\text{op}} \\
&= \left\|(\mathbf{\Phi}\mathbf{\Phi}^*)^{-1}\mathbf{\Phi} : \ell_2^m \to W_1 + \frac{\sqrt{L}}{\mu}W_2\right\|_{\text{op}} \\
&= \left\|(\mathbf{\Phi}\mathbf{\Phi}^*)^{-1}\mathbf{\Phi} : \ell_2^m \to W_1\right\|_{\text{op}} \wedge \left\|(\mathbf{\Phi}\mathbf{\Phi}^*)^{-1}\mathbf{\Phi} : \ell_2^m \to \frac{\sqrt{L}}{\mu}W_2\right\|_{\text{op}} \\
&\leqslant \left\|(\mathbf{\Phi}\mathbf{\Phi}^*)^{-1}\mathbf{\Phi} : \ell_2^m \to \frac{\sqrt{L}}{\mu}W_2\right\|_{\text{op}} \\
&= \frac{\sqrt{L}}{\mu}\left\|\mathbf{\Phi}^*(\mathbf{\Phi}\mathbf{\Phi}^*)^{-1}\right\| \\
&\leqslant \frac{\sqrt{L}}{\mu}\left\|\mathbf{\Phi}\right\|\left\|(\mathbf{\Phi}\mathbf{\Phi}^*)^{-1}\right\|. \qquad \text{(A.91)}
\end{aligned}$$

By plugging in the estimates of $\|\mathbf{\Phi}\|$ and $\|(\mathbf{\Phi}\mathbf{\Phi}^*)^{-1}\|$ into (Equation A.91), we obtain

$$\begin{aligned}
\left\|\mathbf{\Phi}^\dagger\right\|_{\text{op}} &\leqslant \frac{\sqrt{L}}{\mu}\left\|\mathbf{\Phi}^*\right\|\left\|(\mathbf{\Phi}\mathbf{\Phi}^*)^{-1}\right\| \\
&\leqslant \frac{\sqrt{L}}{\mu}\left(\frac{\sqrt{1+\delta}}{1-\delta}\right)
\end{aligned}$$

with probability a least $1 - 4\exp(-\delta^2 m)$. ∎

Embedding $\ell_2^m \otimes_\pi \ell_2^N$ into ℓ_1

Next we embed $\ell_2^m \otimes_\pi \ell_2^N$ a finite dimensional ℓ_1 space so that we can apply Lemma 30. This can be done by embedding each factor of the tensor product $\ell_2^m \otimes_\pi \ell_2^N$ into corresponding ℓ_1 space. Specifically we embed ℓ_2^m (resp. ℓ_2^N) into $\ell_1^{m'}$ (resp. $\ell_1^{N'}$ for suitable m' and N'. Then since $\ell_1^{m'} \otimes_\pi \ell_1^{N'} = \ell_1^{m'N'}$, we obtain the desired embedding. In fact, the embedding between these two particular Banach spaces is well studied in the literature. We use the version by Carl [78] in the following lemma.

Lemma 33 *([78, Lemma 5]) Let $1 < p \leqslant 2$. For each $\epsilon > 0$, there exists a constant $c(p,\epsilon)$ such that for each m', $\ell_1^{m'}$ contains a subspace $(1+\epsilon)$-isomorphic to ℓ_p^m with $m \geqslant c(p,\epsilon)m'$.*

Let $\mathbf{\Psi} : \ell_2^m \to \ell_1^{m'}$ and $\mathbf{\Upsilon} : \ell_2^N \to \ell_1^{N'}$ denote ϵ-accurate embedding maps given by Lemma 33, which satisfy

$$\|\mathbf{\Psi}\|_{\mathrm{op}} \leqslant 1+\epsilon \quad \text{and} \quad \|\mathbf{\Upsilon}\|_{\mathrm{op}} \leqslant 1+\epsilon. \tag{A.92}$$

Then, since $\|\mathbf{\Psi} \otimes \mathbf{\Upsilon}\|_{\mathrm{op}} = \|\mathbf{\Psi}\|_{\mathrm{op}} \|\mathbf{\Upsilon}\|_{\mathrm{op}}$, we have

$$\left\| \mathbf{\Psi} \otimes \mathbf{\Upsilon} : \ell_2^m \otimes \ell_2^N \to \ell_1^{m'} \otimes \ell_1^{N'} \right\| \leqslant (1+\epsilon)^2. \tag{A.93}$$

Let

$$E = \mathbf{\Psi}(\ell_2^m) \subset \ell_1^{m'}$$

and

$$E' = \mathbf{\Upsilon}(\ell_2^N) \subset \ell_1^{N'}.$$

In the next section, we derive an upper bound on the operator norm of an embedding map from $E \otimes_\epsilon E'$ to $W_\infty \otimes_\epsilon \ell_2^N \cap \frac{\mu}{\sqrt{L}} W_2 \otimes_\epsilon \ell_2^N$.

Embedding $E \otimes_\epsilon E'$ into $W_\infty \otimes_\epsilon \ell_2^N \cap \frac{\mu}{\sqrt{L}} W_2 \otimes_\epsilon \ell_2^N$

To complete the commutative diagram in Figure A.1, we derive an upper bound on the the embedding map given by

$$\mathbf{\Phi}\mathbf{\Psi}^{-1} \otimes \mathbf{\Upsilon}^{-1} : E \otimes_\epsilon E' \to W_\infty \otimes_\epsilon \ell_2^N \cap \frac{\mu}{\sqrt{L}} W_2 \otimes_\epsilon \ell_2^N.$$

By the construction of the interpolation space, we have

$$\left\| \mathbf{\Phi}\mathbf{\Psi}^{-1} \otimes \mathbf{\Upsilon}^{-1} : E \otimes_\epsilon E' \to W_\infty \otimes_\epsilon \ell_2^N \cap \frac{\mu}{\sqrt{L}} W_2 \otimes_\epsilon \ell_2^N \right\| \tag{A.94}$$

$$= \left\| \mathbf{\Phi}\mathbf{\Psi}^{-1} \otimes \mathbf{\Upsilon}^{-1} : E \otimes_\epsilon E' \to W_\infty \otimes_\epsilon \ell_2^N \right\| \vee \left\| \mathbf{\Phi}\mathbf{\Psi}^{-1} \otimes \mathbf{\Upsilon}^{-1} : E \otimes_\epsilon E' \to \frac{\mu}{\sqrt{L}} W_2 \otimes_\epsilon \ell_2^N \right\| \tag{A.95}$$

$$= \left\| \mathbf{\Phi}\mathbf{\Psi}^{-1} : E \to W_\infty \right\| \left\| \mathbf{\Psi}^{-1} : E' \to \ell_2^N \right\| \vee \left\| \mathbf{\Phi}\mathbf{\Psi}^{-1} : E \to \frac{\mu}{\sqrt{L}} W_2 \right\| \left\| \mathbf{\Psi}^{-1} : E' \to \ell_2^N \right\|. \tag{A.96}$$

The following lemmas provide upper bounds on the operator norms in the right-hand side of (Equation A.96).

Lemma 34 *For the $\mathbf{\Phi}$ and $\mathbf{\Psi}$ operators defined above, the operator norm of $\mathbf{\Phi}\mathbf{\Psi}^{-1} : E \to W_\infty \cap \frac{\mu}{\sqrt{L}} W_2$ is upper bounded by*

$$\left\| \mathbf{\Phi}\mathbf{\Psi}^{-1} : E \to W_\infty \cap \frac{\mu}{\sqrt{L}} W_2 \right\|_{\text{op}} \lesssim \frac{\max(\mu, \log^{3/2} L)}{\sqrt{L}} \sqrt{\frac{K}{m}}. \tag{A.97}$$

Proof Due to the definition of the interpolation space, we have

$$\left\| \mathbf{\Phi}\mathbf{\Psi}^{-1} : E \to W_\infty \cap \frac{\mu}{\sqrt{L}} W_2 \right\|_{\text{op}} \leqslant \underbrace{\left\| \mathbf{\Phi}\mathbf{\Psi}^{-1} : E \to W_\infty \right\|_{\text{op}}}_{\S} \vee \underbrace{\left\| \mathbf{\Phi}\mathbf{\Psi}^{-1} : E \to \frac{\mu}{\sqrt{L}} W_2 \right\|_{\text{op}}}_{\S\S}$$

Each of the quantities $\S$ and $\S\S$ is upper-bounded as follows.

Claim 1: The operator norm of $\mathbf{\Phi}\mathbf{\Psi}^{-1} : E \to W_\infty$ is upper bounded as

$$\left\| \mathbf{\Phi}\mathbf{\Psi}^{-1} : E \to W_\infty \right\|_{\text{op}} \lesssim \sqrt{\frac{K}{mL}} \log^{3/2} L.$$

Proof [Proof of Claim 1] In order to obtain an upper bound, we make use Chevet's inequality, stated below:

Chevet's inequality: Let $\mathbf{A}$ be an $m \times n$ random matrix with its entries being independent standard Gaussian random variables. Then,

$$\sup_{\substack{\mathbf{x} \in T \\ \mathbf{y} \in S}} \langle \mathbf{A}\mathbf{x}, \mathbf{y} \rangle \leqslant w(T) r(S) + w(S) r(T) + \mu \, r(t) r(S) \tag{A.98}$$

with probability at least $1 - 2\exp(\mu^2)$, where $T,\ S$ are arbitrary bounded sets, $w(\cdot)$ denotes the Gaussian width of a set and $r(\cdot)$ denotes the radius of a set with respect to the Euclidean norm.

We have

$$\underbrace{\left\| \mathbf{\Phi}\mathbf{\Psi}^{-1} : E \to W_\infty \right\|_{\text{op}}}_{\text{A}} = \frac{1}{\sqrt{m}} \sup_{\substack{\mathbf{z} \in \mathbf{\Psi}^{-1} E \\ \mathbf{y} \in \widehat{\mathbf{D}}\widehat{\mathbf{D}}^* B_1^L}} \langle \mathbf{y}, \widehat{\mathbf{G}}\mathbf{z} \rangle$$

Let $T = \{\mathbf{x} = \mathbf{\Psi}^{-1}(E)\}$ and $S = \{\mathbf{x} \in \bar{\widehat{\mathbf{D}}}\widehat{\mathbf{D}}^* B_1^L\}$, where B_1^L is the ℓ_1-unit ball. Then

$$\text{rad}(T) = \sup_{\mathbf{x}\in T} \|\mathbf{x}\|_2 \leqslant 1.$$

Note that this is due to the construction of the map Ψ in Lemma 33, which stipulates that $\|\mathbf{\Psi}\| \, \|\mathbf{\Psi}^{-1}\| \leqslant 1 + \epsilon$. Then, we can design Ψ such that $\|\mathbf{\Psi}\| = 1 + \epsilon$, resulting in $\|\Psi^{-1}\| \leqslant 1$. The conclusion then follows.

$$w(T) = \sup_{\mathbf{x}\in T} \langle \mathbf{x}, \mathbf{g} \rangle \leqslant c\sqrt{\log m'}$$

by standard computation of the Gaussian width of the ℓ_1-unit ball. Similarly, the radius of S is upper-bounded by

$$\text{rad}(S) = \sup_{\mathbf{x}\in S} \|\mathbf{x}\|_2 \leqslant \sup_{\mathbf{y}\in B_1^L} \left\|\widehat{\mathbf{D}}^*\mathbf{y}\right\|_2 = \left\|\widehat{\mathbf{D}}^*\right\|_{1\to 2} \leqslant \sqrt{K/L}.$$

Moreover by Dudley's inequality we obtain an upper bound on the Gaussian width of S given by

$$w(S) = \mathrm{E} \sup_{\mathbf{x}\in\widehat{\mathbf{D}}\widehat{\mathbf{D}}^* B_1^L} \langle \mathbf{x}, \mathbf{g} \rangle \lesssim \sqrt{\frac{K}{L}} \int \sqrt{\log N(S, \|\cdot\|_{1\to 2}, \eta)} d\eta \lesssim \sqrt{\frac{K}{L}} \log^{3/2} L.$$

Finally, we have

$$\left\|\mathbf{\Phi}\mathbf{\Psi}^{-1} : E \to W_\infty\right\| \lesssim \frac{1}{\sqrt{m}} w(T)\text{rad}(S) + w(S)\text{rad}(T) \leqslant \sqrt{\frac{K}{mL}} \log^{3/2} L.$$

■

Claim 1: The operator norm of $\mathbf{\Phi}\mathbf{\Psi}^{-1} : E \to \frac{\mu}{\sqrt{L}} W_2$ is upper-bounded by

$$\left\|\mathbf{\Phi}\mathbf{\Psi}^{-1} : E \to \frac{\mu}{\sqrt{L}} W_2\right\|_{\text{op}} \leqslant \frac{\mu}{\sqrt{Lm}} \left(\sqrt{K} + \sqrt{\log m'}\right).$$

Proof [Proof of Claim 2] The operator norm can be written in a variational form as follows:

$$\left\|\mathbf{\Phi}\mathbf{\Psi}^{-1} : E \to \frac{\mu}{\sqrt{L}} W_2\right\| = \frac{\mu}{\sqrt{Lm}} \sup_{\substack{\mathbf{v}\in\tilde{\widehat{\mathbf{D}}}B_2^K \\ \mathbf{x}\in B_E}} \langle \mathbf{v}, \widehat{\mathbf{D}}^\top \mathbf{G}^* \mathbf{\Psi}^{-1}\mathbf{x}\rangle$$

$$= \frac{\mu}{\sqrt{Lm}} \sup_{\substack{\mathbf{y}\in\tilde{\widehat{\mathbf{D}}}B_2^K =: S \\ \mathbf{z}\in\mathbf{\Psi}^{-1}B_E =: T}} \langle \mathbf{y}, \mathbf{G}^*\mathbf{z}\rangle,$$

where B_E denotes the unit ball in E. We again use Chevet's inequality ((Equation A.98)) to compute an upper bound on the above quantity. The Gaussian width and radius in the ℓ_2-norm of S satisfy

$$w(S) = \mathrm{E}\sup_{\mathbf{x}\in S}\langle \mathbf{x}, \mathbf{g}\rangle \leqslant \sqrt{K};\ \ \mathrm{rad}(S) = \sup_{\mathbf{x}\in B_2^K} \|\mathbf{x}\|_2 = 1.$$

Moreover we have

$$w(T) = \sup_{\mathbf{x}\in T}\langle \mathbf{x}, \mathbf{g}\rangle \leqslant \sqrt{\log m'};\ \ \mathrm{rad}(T) = \sup_{\mathbf{y}\in T} \|\mathbf{y}\|_2 = 1.$$

Then Chevet's inequality implies

$$\left\|\mathbf{\Phi}\mathbf{\Psi}^{-1} : E \to W_2\right\| \lesssim w(T)\mathrm{rad}(S) + w(S)\mathrm{rad}(T) \leqslant \frac{\mu}{\sqrt{Lm}}\left(\sqrt{K} + \sqrt{\log m'}\right).$$

■

We obtain the results in (Equation A.97) by combining the results of Claims 1 and 2. ■

Concluding the proof of Lemma 10

To combine the results in the previous sections with Lemma 30, we utilize the following properties of the dyadic entropy numbers (see e.g. [81]): For any linear operators $\mathcal{R}$ and $\mathcal{S}$ and for any $k \in \mathbb{N}$, we have

$$e_k(\mathcal{R}\mathcal{S}) \leqslant \|\mathcal{R}\|_{\mathrm{op}}\, e_k(\mathcal{S}) \quad \text{and} \quad e_k(\mathcal{R}\mathcal{S}) \leqslant e_k(\mathcal{R})\, \|\mathcal{S}\|_{\mathrm{op}}. \tag{A.99}$$

Then, by applying (Equation A.99) to the commutative diagram in Figure A.1, we obtain

$$\mathcal{E}_{2,1}\left(\mathrm{id} : W_\infty \otimes_\pi \ell_2^N \cap \frac{\mu}{\sqrt{L}} W_2 \otimes_\pi \ell_2^N \to W_\infty \otimes_\epsilon \ell_2^N \cap \frac{\mu}{\sqrt{L}} W_2 \otimes_\epsilon \ell_2^N\right) \tag{A.100}$$

$$\leqslant \mathcal{E}_{1,2}\left(\mathrm{id} : \ell_1^{m'} \otimes_\pi \ell_1^N \to \ell_1^{m'} \otimes_\epsilon \ell_1^N\right) \left\|\mathbf{\Phi}^\dagger \otimes \mathrm{id}\right\| \left\|\mathbf{\Psi} \otimes \mathbf{\Upsilon}\right\| \left\|\mathbf{\Phi}\mathbf{\Psi}^{-1} \otimes \mathbf{\Upsilon}^{-1}\right\| \tag{A.101}$$

where $\mathbf{\Phi}^\dagger \otimes \mathrm{id} : W_\infty \otimes_\pi \ell_2^N \cap \frac{\mu}{\sqrt{L}} W_2 \otimes_\pi \ell_2^N \to \ell_2^m \otimes_\pi \ell_2^N$, $\mathbf{\Psi} \otimes \mathbf{\Upsilon} : \ell_2^m \otimes \ell_2^N \to \ell_1^{m'} \otimes \ell_1^{N'}$ and $\mathbf{\Phi}\mathbf{\Psi}^{-1} \otimes \mathbf{\Upsilon}^{-1} : E \otimes_\epsilon E' \to W_\infty \otimes_\epsilon \ell_2^N \cap \frac{\mu}{\sqrt{L}} W_2 \otimes_\epsilon \ell_2^N$.
Furthermore, there exists an isometric embedding ι^* from $\ell_1^{m'} \otimes_\pi \ell_1^{N'}$ into $\ell_\infty^{2^{m'}} \otimes_\pi \ell_\infty^{2^{N'}}$ [84]. Therefore the map ι^* in Figure A.1 restricted on the image of $\ell_1^{m'} \otimes_\pi \ell_1^{N'}$ has the unit operator norm. Then Lemma 30 and the surjectivity of the dyadic entropy number imply

$$\begin{aligned}\mathcal{E}_{1,2}\left(\mathrm{id} : \ell_1^{m'} \otimes_\pi \ell_1^N \to \ell_1^{m'} \otimes_\epsilon \ell_1^N\right) &\leqslant \mathcal{E}_{2,1}\left(\mathrm{id} : \ell_1^{m'} \otimes \ell_1^{N'} \to \ell_\infty^{2^{m'}} \otimes \ell_\infty^{2^{N'}}\right) \\ &\lesssim \sqrt{m' + N'}\left(1 + \log(m'N')\right)^{3/2}.\end{aligned}$$

By the definition of the injective and projective tensor products, we have

$$\ell_1^{m'} \otimes_\pi \ell_1^{N'} = \ell_1^{m'N'}$$

and

$$\ell_\infty^{2^{m'}} \otimes_\epsilon \ell_\infty^{2^{N'}} = \ell_\infty^{2^{m'+N'}}.$$

Then, by the direct application of Lemma 30, we obtain

$$\begin{aligned}\mathcal{E}_{2,1}\left(\mathrm{id} : \ell_1^{m'} \otimes \ell_1^{N'} \to \ell_\infty^{2^{m'}} \otimes \ell_\infty^{2^{N'}}\right) &\leqslant c\sqrt{1 + \log(2^{m'}2^N)}(1 + \log(m'N))^{3/2} \\ &\lesssim \sqrt{m' + N} \\ &\lesssim \sqrt{K/\delta^2 + N} \\ &\lesssim \sqrt{K + N}\end{aligned} \tag{A.102}$$

where the last line follows by using a suitable constant multiplication factor.

Finally, we substitute the results of the (Equation A.102) (Equation A.92) and (Equation A.97)

in (Equation A.101) to obtain

$$\mathcal{E}_{2,1}\left(\mathrm{id} : W_\infty \otimes_\pi \ell_2^N \cap \frac{\mu}{\sqrt{L}} W_2 \otimes_\pi \ell_2^N \to W_\infty \otimes_\epsilon \ell_2^N \cap \frac{\mu}{\sqrt{L}} W_2 \otimes_\epsilon \ell_2^N\right) \tag{A.103}$$

$$\lesssim \sqrt{K+N}\frac{K}{m}\frac{(1+\epsilon)^2}{1-\delta}\log^{3/2} L$$

$$\lesssim \sqrt{K+N}\log^{3/2} L \tag{A.104}$$

where the last line follows from the fact that $K < m$ and by taking suitable constants to replace ϵ and δ. This proves Lemma 10.

REFERENCES

[1] R. Spencer, "The square kilometre array: The ultimate challenge for processing big data," in *IET Seminar on Data Analytics: Deriving Intelligence and Value from Big Data*, IET, 2013, pp. 1–26.

[2] W. B. Johnson and J. Lindenstrauss, "Extensions of lipschitz mappings into a hilbert space," 1984.

[3] D. Woodruff, "Sketching as a tool for numerical linear algebra," *Found. Trends Theor. Comput. Sci.*, vol. 10, no. 1–2, pp. 1–157, 2014.

[4] H. Avron, K. L. Clarkson, and D. P. Woodruff, "Sharper bounds for regularized data fitting," *arXiv preprint arXiv:1611.03225*, 2016.

[5] M. B. Cohen, J. Nelson, and D. P. Woodruff, "Optimal approximate matrix product in terms of stable rank," *arXiv preprint arXiv:1507.02268*, 2015.

[6] R. Kannan and S. Vempala, "Randomized algorithms in numerical linear algebra," *Acta Numerica*, vol. 26, pp. 95–135, 2017.

[7] D. Ehyaie, "Novel approaches to the design of phased array antennas," PhD thesis, University of Michigan, 2011.

[8] *Project soli*, https://atap.google.com/soli/.

[9] S. Patole and M. Torlak, "Two dimensional array imaging with beam steered data," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5181–5189, 2013.

[10] R. S. Srinivasa, M. A. Davenport, and J. Romberg, "Trading beams for bandwidth: Imaging with randomized beamforming," *SIAM Journal on Imaging Sciences*, vol. 13, no. 1, pp. 317–350, 2020. eprint: https://doi.org/10.1137/19M1242045.

[11] K. Raczkowski, W. De Raedt, B. Nauwelaers, and P. Wambacq, "A wideband beamformer for a phased-array 60ghz receiver in 40nm digital cmos," in *2010 IEEE International Solid-State Circuits Conference - (ISSCC)*, Feb. 2010, pp. 40–41.

[12] E. A. Alwan, S. Balasubramanian, J. G. Atallah, M. LaRue, K. Sertel, W. Khalil, and J. L. Volakis, "Coding-based ultra-wideband digital beamformer with significant hardware reduction," *Analog Integrated Circuits and Signal Processing*, vol. 78, no. 3, pp. 691–703, Mar. 2014.

[13] M. Huang, T. Chi, F. Wang, T. Li, and H. Wang, "A 23-to-30ghz hybrid beamforming mimo receiver array with closed-loop multistage front-end beamformers for full-fov

dynamic and autonomous unknown signal tracking and blocker rejection," in *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, Feb. 2018, pp. 68–70.

[14] T. Chi, M. Huang, S. Li, and H. Wang, "17.7 a packaged 90-to-300ghz transmitter and 115-to-325ghz coherent receiver in cmos for full-band continuous-wave mm-wave hyperspectral imaging," in *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, Feb. 2017, pp. 304–305.

[15] T. Sakamoto, T. Sato, K. Iwasa, and H. Yomo, "Frequency-domain interferometric imaging and velocity vector estimation using networked ultra-wideband 80-ghz array radar systems," in *2017 IEEE International Symposium on Antennas and Propagation USNC/URSI National Radio Science Meeting*, Jul. 2017, pp. 851–852.

[16] J. Johnson, M. Karaman, and B. Khuri-Yakub, "Coherent-array imaging using phased subarrays. Part I: Basic principles," *IEEE Trans. Ultrason., Ferroelect., and Freq. Control*, vol. 52, no. 1, pp. 37–50, 2005.

[17] D. Slepian, "On bandwidth," *Proc. IEEE*, vol. 64, no. 3, pp. 292–300, 1976.

[18] D. Slepian and H. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty — I," *Bell Syst. Tech. J.*, vol. 40, no. 1, pp. 43–63, 1961.

[19] H. Landau and H. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty — II," *Bell Syst. Tech. J.*, vol. 40, no. 1, pp. 65–84, 1961.

[20] D. Slepian, "Prolate spheroidal wave functions, Fourier analysis and uncertainty — IV: Extensions to many dimensions; generalized prolate spheroidal functions," *Bell Syst. Tech. J.*, vol. 43, no. 6, pp. 3009–3057, 1964.

[21] ——, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty — V: The discrete case," *Bell Syst. Tech. J.*, vol. 57, no. 5, pp. 1371–1430, 1978.

[22] H. Landau and H. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty — III: The dimension of the space of essentially time- and band-limited signals," *Bell Syst. Tech. J.*, vol. 41, no. 4, pp. 1295–1336, 1962.

[23] F. Simons, "Slepian functions and their use in signal estimation and spectral analysis," in. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 891–923, ISBN: 978-3-642-01546-5.

[24] F. Simons and D. Wang, "Spatiospectral concentration in the Cartesian plane," *GEM – Int. J. Geomathematics*, vol. 2, no. 1, pp. 1–36, Jun. 2011.

[25] S. Karnik, Z. Zhu, M. Wakin, J. Romberg, and M. Davenport, "The fast slepian transform," *to appear in Appl. Comput. Harmon. Anal.*, 2019.

[26] Z. Zhu, S. Karnik, M. Wakin, M. Davenport, and J. Romberg, "ROAST: Rapid Orthogonal Approximate Slepian Transform," *IEEE Trans. Signal Process.*, vol. 66, no. 22, pp. 5887–5901, 2018.

[27] Z. Zhu, S. Karnik, M. A. Davenport, J. Romberg, and M. Wakin, "The eigenvalue distribution of discrete periodic time-frequency limiting operators," *IEEE Signal Process. Lett.*, vol. 25, no. 1, pp. 95–99, 2018.

[28] N. Halko, P. Martinsson, and J. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM Rev.*, vol. 53, no. 2, pp. 217–288, 2011. eprint: http://dx.doi.org/10.1137/090771806.

[29] B. Li, W. Li, and L. Cui, "New bounds for perturbation of the orthogonal projection," *Calcolo*, vol. 50, no. 1, pp. 69–78, 2013.

[30] P. Boufounos, "Depth sensing using active coherent illumination," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012.

[31] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on pure and applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.

[32] Y. Chi and M. Ferreira Da Costa, "Harnessing sparsity over the continuum: Atomic norm minimization for superresolution," *IEEE Signal Processing Magazine*, vol. 37, no. 2, pp. 39–57, 2020.

[33] M. A. Davenport and M. B. Wakin, "Compressive sensing of analog signals using discrete prolate spheroidal sequences," *Applied and Computational Harmonic Analysis*, vol. 33, no. 3, pp. 438–472, 2012.

[34] S. Karnik, Z. Zhu, M. B. Wakin, J. Romberg, and M. A. Davenport, "The fast slepian transform," *Applied and Computational Harmonic Analysis*, vol. 46, no. 3, pp. 624–652, 2019.

[35] S. Karnik, J. Romberg, and M. A. Davenport, "Fast multitaper spectral estimation," in *2019 13th International conference on Sampling Theory and Applications (SampTA)*, IEEE, 2019, pp. 1–4.

[36] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[37] D. Needell and J. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[38] D. P. Woodruff, "Sketching as a tool for numerical linear algebra," *Found. Trends Theor. Comput. Sci.*, vol. 10, no. 1–2, pp. 1–157, Oct. 2014.

[39] S. Wang, A. Gittens, and M. W. Mahoney, "Sketched ridge regression: Optimization perspective, statistical perspective, and model averaging," in *International Conference on Machine Learning*, 2017, pp. 3608–3616.

[40] J. Yang, X. Meng, and M. Mahoney, "Implementing randomized matrix algorithms in parallel and distributed environments," *Proc. IEEE*, vol. 104, no. 1, pp. 58–92, 2016.

[41] A. Chowdhury, J. Yang, and P. Drineas, "An iterative, sketching-based framework for ridge regression," in *Proceedings of the 35th International Conference on Machine Learning*, J. Dy and A. Krause, Eds., ser. Proceedings of Machine Learning Research, vol. 80, Stockholmsmässan, Stockholm Sweden: PMLR, Oct. 2018, pp. 989–998.

[42] N. Ailon and B. Chazelle, "Approximate nearest neighbors and the fast johnson-lindenstrauss transform," in *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, ACM, 2006, pp. 557–563.

[43] K. L. Clarkson and D. P. Woodruff, "Low-rank approximation and regression in input sparsity time," *Journal of the ACM (JACM)*, vol. 63, no. 6, p. 54, 2017.

[44] R. S. Srinivasa, M. A. Davenport, and J. Romberg, "Sample complexity bounds for localized sketching," S. Chiappa and R. Calandra, Eds., ser. Proceedings of Machine Learning Research, vol. 108, Online: PMLR, 26–28 Aug 2020, pp. 3275–3284.

[45] A. Eftekhari, H. L. Yap, C. J. Rozell, and M. B. Wakin, "The restricted isometry property for random block diagonal matrices," *Applied and Computational Harmonic Analysis*, vol. 38, no. 1, pp. 1–31, 2015.

[46] B. McWilliams, C. Heinze, N. Meinshausen, G. Krummenacher, and H. P. Vanchinathan, "Loco: Distributing ridge regression with random projections," *stat*, vol. 1050, p. 26, 2014.

[47] S. Chen, Y. Liu, M. R. Lyu, I. King, and S. Zhang, "Fast relative-error approximation algorithm for ridge regression," in *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, ser. UAI'15, Amsterdam, Netherlands: AUAI Press, 2015, pp. 201–210, ISBN: 978-0-9966431-0-8.

[48] S. Paul and P. Drineas, "Feature selection for ridge regression with provable guarantees," *Neural Computation*, vol. 28, no. 4, pp. 716–742, 2016, PMID: 26890353. eprint: https://doi.org/10.1162/NECO_a_00816.

[49] H. Avron, K. Clarkson, and D. Woodruff, "Faster kernel ridge regression using sketching and preconditioning," *SIAM Journal on Matrix Analysis and Applications*, vol. 38, no. 4, pp. 1116–1138, 2017. eprint: https://doi.org/10.1137/16M1105396.

[50] P. Drineas, M. Magdon-Ismail, M. W. Mahoney, and D. P. Woodruff, "Fast approximation of matrix coherence and statistical leverage," *Journal of Machine Learning Research*, vol. 13, no. Dec, pp. 3475–3506, 2012.

[51] N. Halko, P.-G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM review*, vol. 53, no. 2, pp. 217–288, 2011.

[52] J. A. Tropp, A. Yurtsever, M. Udell, and V. Cevher, "Streaming low-rank matrix approximation with an application to scientific simulation," *arXiv:1902.08651*, 2019.

[53] S. M. Fazel, "Matrix rank minimization with applications.," 2003.

[54] M. A. Davenport and J. Romberg, "An overview of low-rank matrix recovery from incomplete observations," *IEEE J. Sel. Topics Signal Process*, vol. 10, no. 4, pp. 608–622, Jun. 2016.

[55] T. Cai and W. Zhou, "Matrix completion via max-norm constrained optimization," *Electronic J. Stat.*, vol. 10, no. 1, pp. 1493–1525, 2016.

[56] R. S. Srinivasa, K. Lee, M. Junge, and J. Romberg, "Decentralized sketching of low rank matrices," in *Advances in Neural Information Processing Systems*, 2019, pp. 10 101–10 110.

[57] T. T. Cai, A. Zhang, *et al.*, "Rop: Matrix recovery via rank-one projections," *The Annals of Statistics*, vol. 43, no. 1, pp. 102–138, 2015.

[58] G. Jameson, *Summing and nuclear norms in Banach space theory*. Cambridge University Press, 1987, vol. 8.

[59] R. A. Ryan, *Introduction to tensor products of Banach spaces*. Springer Science & Business Media, 2013.

[60] J. J. Bruer, "Recovering structured low-rank operators using nuclear norms," PhD thesis, California Institute of Technology, 2017.

[61] E. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found, of Comp. Math.*, vol. 9, no. 6, p. 717, 2009.

[62] F. P. A. and S. H., "Memory and computation efficient PCA via very sparse random projections," in *Proceedings of the 31st International Conference on Machine Learning*, vol. 32, Bejing, China, Jun. 2014, pp. 1341–1349.

[63] M. Azizyan, A. Krishnamurthy, and A. Singh, "Extreme compressive sampling for covariance estimation," *arXiv preprint arXiv:1506.00898*, 2015.

[64] H. Qi and S. M. Hughes, "Invariance of principal components under low-dimensional random projection of the data," in *19th IEEE Int. Conf. Image Process.*, Sep. 2012, pp. 937–940.

[65] S. Negahban and M. J. Wainwright, "Estimation of (near) low-rank matrices with noise and high-dimensional scaling," *Ann. Statist.*, vol. 39, no. 2, pp. 1069–1097, Apr. 2011.

[66] N. Srebro, J. Rennie, and T. S. Jaakkola, "Maximum-margin matrix factorization," in *Advances in Neural Information Processing Systems 17*, L. K. Saul, Y. Weiss, and L. Bottou, Eds., MIT Press, 2005, pp. 1329–1336.

[67] N. Linial, S. Mendelson, G. Schechtman, and A. Shraibman, "Complexity measures of sign matrices," *Combinatorica*, vol. 27, no. 4, pp. 439–463, 2007.

[68] F. Krahmer, S. Mendelson, and H. Rauhut, "Suprema of chaos processes and the restricted isometry property," *Communications on Pure and Applied Mathematics*, vol. 67, no. 11, pp. 1877–1904, 2014.

[69] G. Pisier, *The volume of convex bodies and Banach space geometry*. Cambridge University Press, 1999, vol. 94.

[70] M. Ledoux and M. Talagrand, *Probability in Banach Spaces: isoperimetry and processes*. Springer Science & Business Media, 2013.

[71] J. F. Sturm, "Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones," *Optimization methods and software*, vol. 11, no. 1-4, pp. 625–653, 1999.

[72] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[73] E. X. Fang, H. Liu, K. Toh, and W. Zhou, "Max-norm optimization for robust matrix recovery," *Mathematical Programming*, vol. 167, no. 1, pp. 5–35, 2018.

[74] K. Toh and S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems," *Pacific Journal of optimization*, 2010.

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

[76] F. Krahmer and D. Stöger, "On the convex geometry of blind deconvolution and matrix completion," *arXiv preprint arXiv:1902.11156*, 2019.

[77] F. Krahmer, S. Mendelson, and H. Rauhut, "Suprema of chaos processes and the restricted isometry property," *Communications on Pure and Applied Mathematics*, vol. 67, no. 11, pp. 1877–1904, 2014.

[78] B. Carl, "Inequalities of Bernstein-Jackson-type and the degree of compactness of operators in Banach spaces," *Ann. Inst. Fourier (Grenoble)*, vol. 35, no. 3, pp. 79–118, 1985.

[79] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*, 3. Birkhäuser Basel, 2013, vol. 1.

[80] K. Lee, F. Krahmer, and J. Romberg, "Spectral methods for passive imaging: Nonasymptotic performance and robustness," *SIAM Journal on Imaging Sciences*, vol. 11, no. 3, pp. 2110–2164, 2018.

[81] B. Carl and I. Stephani, *Entropy, Compactness and the Approximation of Operators*, ser. Cambridge Tracts in Mathematics. Cambridge University Press, 1990.

[82] M. Junge and K. Lee, "Generalized notions of sparsity and restricted isometry property. part i: A unified framework," *Information and Inference: A Journal of the IMA*, vol. 9, no. 1, pp. 157–193, 2020.

[83] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," *arXiv preprint arXiv:1011.3027*, 2010.

[84] M. Junge, K. Lee, J. Romberg, and R. S. Srinivasa, "Convex relexation of low-rank operators in tensor products of Banach spaces," *in preparation*,