Phase Retrieval from Random One-Bit Measurements

by
Dylan S. Domel-White

A dissertation submitted to the Department of Mathematics,

College of Natural Sciences and Mathematics

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Mathematics

Chair of Committee: Bernhard Bodmann

Committee Member: David Blecher

Committee Member: Simon Foucart

Committee Member: Anna Vershynina

University of Houston
May 2020

ACKNOWLEDGMENTS

There are many people that I would like to acknowledge for playing important roles, big and small, throughout my studies. First, to my fellow graduate students at the University of Houston: thank you for creating a friendly, positive atmosphere for doing mathematics. To name just a few, I would like to acknowledge Ahmed Abouserie, Alan Akil, Tattwamasi Amrutan, Prajakta Bedekar, Kayla Bicol, Jasmine Bhullar, Arianna Cecco, Sarah Chehade, Akshat Das, Matthew Dysart, Iris Emilsdottir, Deepjyoti Ghosh, Mozahid Haque, Gregory Hemenway, Shahzad Kalloo, Nikolaus Karantzas, Ang Li, Wilfredo Molina, Neal Nesbitt, Duong Nguyen, Khanh Nguyen, Basanta Pahari, Nikolaus Panagopoulos, Adrian Radillo, Kazem Safaripoorfatide, Saeed Sarmadi, Worawit Tepsan, Richard Walden, and Kyle Williams. I have enjoyed my time with you, and I hope that I have contributed to making our department a better place. Most importantly, I have to thank my academic older brother Robert Mendez for setting a stellar example of mathematical inquisitiveness and research work ethic. I was very lucky to have Robert around as a mentor and a friend as I first began to engage in mathematical research.

Secondly, I would like to thank the many professors and staff members who have supported me in the UH math department. Specifically, I would like to acknowledge the professors leading our Analysis Research Group, including David Blecher, Bernhard Bodmann, Mehrdad Kalantar, and Anna Vershynina. All of these professors gave me excellent mathematical instruction and helped me achieve at small amount of mathematical maturity. Of course, I must place special emphasis on my Ph.D. advisor Dr. Bernhard Bodmann, who has been an amazing research mentor. During my second year of the Ph.D. program, Dr. Bodmann took a chance and invited me to an upcoming conference on phase retrieval in Hong Kong before and proposed we work together, which was an incredible experience and is an opportunity I am very fortunate to have received. I am thankful for the confidence and generosity he has displayed towards me, in addition to his constantly invaluable advice and guidance.

Lastly, I would like to thank my parents Lori Domel and Cameron White, and my partner Daniella Heysquierdo. Thank you all for your patience and endless support. You have always been incredibly generous with your time and energy, and I am very thankful to you for it. Dad, a large part of the reason I want to pursue a career in academia is thanks to the example you have set. Your kindness and student-focused philosophy is the model I want to achieve with my teaching. Mom, you have always pushed me to be better and to never be complacent, and you have made me a better person through it. I greatly admire your genealogical research abilities and aspire to match them one day. Daniella, thank you for being with me and supporting me during this phase of my life. We have both grown a lot together over the last five years. You have made the journey so much easier for me, and inspired me with your selfless care for those around you. I look forward to the next part of our journey together.

ABSTRACT

Phase retrieval in real or complex Hilbert spaces is the task of recovering a vector, up to an overall unimodular multiplicative constant, from norms of projections onto subspaces. This dissertation deals with phase retrieval of normalized vectors after the norms of projections are quantized by pairwise comparison to retain only one bit of information. In more specific, geometric terms, we choose a sequence of pairs of subspaces in a real or complex Hilbert space and only record which subspace from each pair is closer to the input vector. The recovery algorithm we define uses the qualitative proximity information encoded in the binary measurement of an input vector to assemble an auxiliary matrix, and then chooses a unit vector in the principal eigenspace of this auxiliary matrix as the estimate for the input vector.

For this measurement and recovery procedure, we provide a pointwise bound for fixed input vectors and a uniform bound that controls the worst-case scenario among all inputs. Both bounds hold with high probability with respect to a choice of subspaces from the uniform distribution induced by the action of the orthogonal or unitary group. For real or complex vectors of dimension n, the pointwise bound requires $m \geq C\delta^{-2}n\log(n)$ and the uniform bound $m \geq C\delta^{-2}n^2\log(\delta^{-1}n)$ binary questions in order to achieve a reconstruction accuracy of δ. The accuracy δ is measured by the operator norm of the difference between the rank-one orthogonal projections corresponding to the normalized input vector and its approximate recovery.

After establishing the pointwise and uniform error bounds for noiseless binary measurements, we consider the case of noisy measurements. Noise for a binary-valued measurement takes the form of bit-flips that corrupt the proximity information encoded in the binary measurement. We show that our measurement and recovery scheme is robust in the presence of a percentage of adversarial bit-flips on the order of $\frac{1}{\sqrt{n}}$. We also consider random bit-flips and show in this setting that the mean squared error of reconstruction decays with respect to the number of projections m on the order of $\frac{\log(m)}{m}$.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Background and Preliminaries

This dissertation primarily deals with one-bit phase retrieval, which is a particular type of problem in the larger field of signal acquisition and recovery. This chapter gives a brief introduction to the concepts and mathematical tools needed to understand one-bit phase retrieval. Some knowledge of linear algebra and probability is assumed.

First, Section 1.1 gives a brief outline of signal processing problems that help motivate the one-bit phase retrieval problem. In particular, one-bit compressed sensing is presented as an illuminating example of a signal processing problem that acquires signals through quantized measurements. Section 1.2 provides a more in-depth look at phase retrieval with non-quantized measurements, including the types of problems people try to solve in this field and an overview of major results. In Section 1.3, one-bit phase retrieval is motivated as a natural extension of non-quantized phase retrieval, and general versions of the problems that are addressed in Chapters 2 and 3 are formally stated. Section 1.4 discusses the distributions of random variables, vectors, and matrices that we will make use of, along with concentration inequalities that serve as important probabilistic tools. Lastly, Section 1.5 discusses how one-bit phase retrieval may be viewed as an encoding and decoding problem in rate-distortion theory.

Notation

Let $\mathbb{F}$ denote the field of real numbers or complex numbers, and define $\beta_{\mathbb{F}} = \frac{1}{2}$ if $\mathbb{F} = \mathbb{R}$ and $\beta_{\mathbb{F}} = 1$ if $\mathbb{F} = \mathbb{C}$. Let $\mathbb{T}_{\mathbb{F}} := \{\alpha \in \mathbb{F} : |\alpha| = 1\}$ be the set of unimodular scalars in $\mathbb{F}$ and let $\mathbb{S}_{\mathbb{F}}^{d-1}$ be the set of unit-norm vectors in $\mathbb{F}^d$. The set of rank-k orthogonal projection matrices on $\mathbb{F}^d$ is denoted

$\mathrm{Proj}_{\mathbb{F}}(k,d)$, and the group of unitary matrices on $\mathbb{F}^d$ is denoted $\mathcal{U}_{\mathbb{F}}(d)$. For functions f and g depending on a real-valued parameter x, the notation $f(x) = O(g(x))$ as $x \to a$ means that there exists positive numbers M and ϵ such that $|f(x)| \leq Mg(x)$ for all $|x - a| < \epsilon$. For two random variables X, Y, the notation $X \overset{(\mathrm{d})}{=} Y$ indicates equality in distribution, i.e., $\mathbb{P}\{X \in E\} = \mathbb{P}\{Y \in E\}$ for all measurable sets E.

1.1 Overview of relevant signal processing problems

Signal processing problems deal with acquisition and reconstruction of signals. A signal is represented by a vector in some Hilbert space which is assumed to be finite-dimensional. Acquisition of a signal is the process by which an unknown signal is measured and the measurements are recorded. Signal recovery, or reconstruction, is the inverse problem of determining a signal based on the recorded measurements acquired from it. Methods for signal acquisition and recovery depend on the specific type of signal being considered, but they may be broadly classified as *linear* and *nonlinear* methods. This section gives a brief overview of some of these methods and related fields as context for the one-bit phase retrieval problems that are stated in Section 1.3 and addressed in Chapters 2 and 3.

Linear acquisition and linear recovery with frames

The simplest form of signal acquisition and recovery is based on fundamental principles of linear algebra that are familiar to most mathematicians. To linearly acquire a signal $x \in \mathbb{F}^d$, measurement devices take one-dimensional linear measurements of x of the form $x \mapsto \langle x, f_j \rangle$ for some vectors $f_1, \ldots, f_m \in \mathbb{F}^d$ to yield a measurement vector $\mathcal{M}(x) = (\langle x, f_j \rangle)_{j=1}^m \in \mathbb{F}^m$. Such a measurement map $\mathcal{M}$ is linear and may be expressed via matrix multiplication: if F is the $m \times d$ matrix with $\bar{f}_j$ as its rows, then $\mathcal{M}(x) = Fx$. A linear recovery algorithm would be implemented by a linear map $\mathcal{R} : \mathbb{F}^m \to \mathbb{F}^d$ that is a left inverse for $\mathcal{M}$, i.e., such that $\mathcal{R} \circ \mathcal{M} = I$ where I is the identity map on $\mathbb{F}^d$. By basic linear algebra, a left inverse $\mathcal{R}$ exists if and only if $\mathcal{M}$ is injective, which happens if and only span $\{f_1, \ldots, f_m\} = \mathbb{F}^d$. A collection of vectors with this property is sometimes referred to as a (finite) frame.

Definition 1.1.1. *A sequence of vectors $\mathcal{F} = \{f_1, \ldots, f_m\}$ in $\mathbb{F}^d$ is called a* ***(finite) frame*** *if*

$$\operatorname{span}\{f_1, \ldots, f_m\} = \mathbb{F}^d.$$

The linear measurement map $\mathcal{M}$ mentioned above is injective if and only if the associated collection of vectors $\mathcal{F} = \{f_1, \ldots, f_m\}$ is a frame. If $\mathcal{F}$ is a frame, then there is a canonical linear reconstruction map $\mathcal{R}$ on $\operatorname{Ran}(\mathcal{M})$ by setting $\mathcal{R}(\mathcal{M}(x)) = x$. Injectivity of $\mathcal{M}$ implies this map is well-defined. To see linearity of $\mathcal{R}$, observe that for any $x, y \in \mathbb{F}^d$ and $a \in \mathbb{F}$

$$\mathcal{R}(a\mathcal{M}(x) + \mathcal{M}(y)) = \mathcal{R}(\mathcal{M}(ax + y)) = ax + y = a\mathcal{R}(\mathcal{M}(x)) + \mathcal{R}(\mathcal{M}(y)).$$

Extending this definition of $\mathcal{R}$ to all of $\mathbb{F}^m$ by setting $\mathcal{R}(z) = 0$ for $z \in \operatorname{Ran}(\mathcal{M})^\perp$ yields a linear recovery map with the property that $\mathcal{R} \circ \mathcal{M} = I$. This particular left-inverse for $\mathcal{M}$ is implemented by the matrix $(F^*F)^{-1}F^*$, i.e., $\mathcal{R}(y) = (F^*F)^{-1}F^*y$ for all $y \in \mathbb{F}^m$. Thus, studying linear signal acquisition schemes that admit linear reconstruction algorithms is equivalent to studying frames of vectors for $\mathbb{F}^d$.

The smallest frame possible for $\mathbb{F}^d$ is a basis, i.e., a maximal set of linearly independent vectors. In this case, the associated linear measurement map $\mathcal{M}$ is invertible and there is a unique linear reconstruction algorithm given by $\mathcal{R} = \mathcal{M}^{-1}$, i.e., $\mathcal{R}(x) = F^{-1}x$. On the other hand, the definition of a frame allows for larger collections of vectors, which provide redundancy in the acquired linear measurement $\mathcal{M}(x)$ that can lead to robustness of the linear reconstruction algorithm to various types of error in the measurement process. Different applications encounter different types of errors in measuring or transmitting signals, which lead to different properties that frames must have to handle those errors as well as possible.

The study of frames, their properties, and their behavior for signal processing is called *frame theory* [32, 60]. Generalizations of frames to collections of subspaces rather than vectors, called *fusion frames*, have been studied in the context of recovery from higher rank linear measurements [29, 31, 33]. Although this section has only discussed frames for finite-dimensional Hilbert spaces, historically frame theory began by looking at redundant expansions of signals in infinite-dimensional Hilbert spaces, for example continuous acoustical signals [28, 42]. Much of frame theory in the finite-dimensional case is concerned with finding properties of frames that ensure optimal robustness to certain error models and then constructing frames with those properties. For example, *equal-norm Parseval* frames yield the minimum mean-squared error of reconstruction in the presence of an erasure of one of the linear measurements, i.e., when $\mathcal{M}(x)_j$ is set to 0 for a single index j [30, 54, 62]. Burst erasures are another error model that has been studied, leading to associated frame-theoretic properties for robust recovery [20, 81].

Linear acquisition and nonlinear recovery

In specific applications, nonlinear recovery algorithms using linearly acquired measurements have proven useful. Nonlinear reconstruction methods can provide improvements over linear methods when the signals being considered have some added structure that causes them to lie in a subset of $\mathbb{F}^d$ that is not a subspace. For example, in the field of compressed sensing various convex optimization problems have been shown to allow exact recovery of sparse signals [39, 44, 52].

Definition 1.1.2. *A signal $x \in \mathbb{F}^d$ is called s-**sparse** if the number of non-zero entries of x is at most s.*

The set of s-sparse signals is not a linear subspace, since the sum of two s-sparse vectors can have up to $2s$ nonzero entries. In compressed sensing, it is assumed that measurement devices can take linear measurements of a sparse signal $x \in \mathbb{F}^d$ of the form $\mathcal{M}(x) = (\langle x, f_j \rangle)_{j=1}^m$ for some vectors $\{f_1, \ldots, f_m\} \subset \mathbb{F}^d$. By using the assumption that the input signals considered are s-sparse for some $s \ll d$, a linear measurement $\mathcal{M}$ may be constructed that admits a recovery algorithm $\mathcal{R}$ even if the number of one-dimensional linear measurements is much smaller than the dimension of the signal, i.e., $m \ll d$.

For example, $\mathcal{R}$ can be defined by selecting the minimizer of the convex program

$$\begin{aligned} \underset{y\in\mathbb{F}^d}{\text{minimize}} \quad & \|y\|_1 \\ \text{subject to} \quad & \mathcal{M}(y) = \mathcal{M}(x). \end{aligned} \tag{1}$$

If $m \geq Cs\log(d/s)$ and $\{f_1,\ldots,f_m\}$ are independent standard Gaussian random vectors in $\mathbb{F}^d$, then with high probability all s-sparse vectors x may be recovered from their linear measurement $\mathcal{M}(x) = (\langle x, f_j\rangle)_{j=1}^m$ by finding the minimizer to (1) [39]. The optimization problem given in (1) is called *basis pursuit*, and has been shown to be robust to errors in the linear measurement process [40]. In relevant applications such as Medical Resonance Imaging (MRI), the one-dimensional linear measurements are taken in sequence, and so reducing the number of necessary measurements to fewer than the dimension of the signal can result in a dramatic speed-up of the signal acquisition process [70].

Compressed sensing is closely related to frame theory. Certain classes of signals might not be sparse in the standard orthonormal basis but are sparse with respect to a redundant frame expansion; if $\mathcal{F}$ is a frame with associated matrix F, the linear measurement Fx may be sparse. Signals with this property can still be recovered by basis pursuit from a small number of measurements [23, 82].

Nonlinear acquisition of measurements

Recently, there has been interest in studying problems which require nonlinear signal acquisition methods. For example, in some real-world applications like automatic speech recognition and x-ray crystallography, instead of acquiring one-dimensional linear measurements $\langle x, f_j\rangle$, measurement devices only have access to intensity measurements of the form $\mathcal{M}(x) = (|\langle x, f_j\rangle|^2)_{j=1}^m$ [10, 14, 45, 49, 50, 61, 69, 72, 92]. The problem of reconstructing signals from a collection of intensity measurements is called *phase retrieval.* There are many generalizations of phase retrieval and its related problems, see Section 1.2 for more detail and discussion.

Another type of nonlinear signal acquisition comes from quantized measurements. Quantization means that the measurements take values in some finite alphabet, which is required for sending and representing signals digitally [57]. The most extreme type of quantization that still reveals information about the input signal uses an alphabet of only two elements, which is called *binary quantization* or *one-bit quantization*. Some analysis and empirical evidence has shown that one-bit quantization is the optimal quantization scheme in many practical applications with a low signal-to-noise ratio [68]. Any form of quantization makes exact recovery of all input signals impossible, as there are only a finite number of outputs in the quantization alphabet. Still, it is possible to find recovery maps that provide approximate recovery. Linear reconstruction from quantized frame coefficients has been studied extensively [36, 41, 55, 56, 71, 88]. Both phase retrieval and quantization of measurements are closely related to the field of quantum state tomography, where repeated quantum measurements yield frequencies from a discrete probability distribution that can be used to determine a quantum state [25, 58, 59, 83, 84].

Compressed sensing of sparse signals in $\mathbb{R}^d$ is one particular field where binary quantization of measurements has been studied, yielding recovery guarantees for a computationally feasible optimization problem. In the most basic setup of one-bit compressed sensing, each one-dimensional linear measurement $\langle x, f_j \rangle$ is quantized by applying the signum function, yielding

$$\operatorname{sgn}(\langle x, f_j \rangle) = \begin{cases} 1 & \text{if } \langle x, f_j \rangle \geq 0 \\ -1 & \text{else.} \end{cases}$$

Letting $\Phi : \mathbb{R}^m \to \{-1, 1\}^m$ be the map which applies the signum function to each component, the binary measurement $\Phi(\mathcal{M}(x)) = (\operatorname{sgn}(\langle x, f_j \rangle))_{j=1}^m$ encodes an input signal $x \in \mathbb{R}^d$ into string of 1's and -1's. Assuming unit-norm input signals in $\mathbb{R}^d$, with sufficiently many such one-bit measurements a sparse input signal can still be approximately estimated from its binary measurement via a variety of algorithms.

Boufounos and Baraniuk first investigated the one-bit compressed sensing problem in [21]. They observed problems with treating one-bit quantization as a special case of the noise models used in traditional compressed sensing reconstruction algorithms, and ultimately proposed a new algorithm that seeks a consistent reconstruction, i.e., finds a vector $\hat{x}$ for which $\operatorname{sgn}(\langle \hat{x}, f_j \rangle) = \operatorname{sgn}(\langle x, f_j \rangle)$ for every j. Their algorithm performed better in experiments than traditional compressed sensing algorithms applied to one-bit measurements. In a follow-up paper, theoretical bounds on the reconstruction error achieved by consistent reconstruction were derived, and measurement by random Gaussian vectors was shown to allow near optimal consistent reconstruction up to logarithmic factors [64].

Plan and Vershynin provided the first formal error bounds for a computationally feasible algorithm for one-bit compressed sensing in [78] and related papers [77, 79]. Their algorithm yields a reconstructed signal that is not necessarily consistent with the original one-bit measurements, but is an accurate estimate for the input signal nonetheless. By relaxing the requirement of consistent reconstruction, they find a provable and efficiently implementable algorithm.

The reconstruction algorithm studied in [78] estimates a unit vector x from its binary measurement $\Phi(\mathcal{M}(x))$ via the convex program

$$\begin{aligned} \underset{y}{\text{maximize}} \quad & \sum_{j=1}^{m} \Phi(\mathcal{M}(x))_j \, \langle y, a_j \rangle \\ \text{subject to} \quad & y \in K, \end{aligned} \tag{2}$$

where K is taken to be the convex hull of the set of s-sparse unit vectors. In particular, by letting $m \geq C\delta^{-6} s \log(2d/s)$ and letting $\{a_1, \ldots, a_m\}$ be independent standard Gaussian random vectors, then with high probability

$$\|\hat{x} - x\|^2 \leq \delta \sqrt{\log(e\delta^{-1})}$$

for all s-sparse unit vectors x, where $\hat{x}$ is the solution to the optimization problem (2) [78, Theorem 1.3]. The results derived by Plan and Vershynin are actually much more general than the above statement and include applications to low-rank matrix recovery and other classes of signals. More recent results have introduced new algorithms that improve the dependence of the reconstruction error on the number of bits if the measurement seeks a consistent reconstruction [13, 87, 95].

The main focus of this dissertation is on *one-bit phase retrieval*, which is similar in some ways to one-bit compressed sensing. One-bit phase retrieval is a signal recovery problem where the signal acquisition process has two types of nonlinearity. First, a nonlinear intensity measurement is taken as in regular phase retrieval. Second, binary quantization reduces the information content of the intensity measurement to just a single bit. Section 1.3 describes the one-bit phase retrieval problems that are later solved in Chapters 2 and 3. Section 2.1 presents a signal acquisition method for one-bit phase retrieval, and Section 2.2 gives its associated signal recovery algorithm. This recovery algorithm is nonlinear, and involves computing the principal eigenspace of an auxiliary matrix determined by the one-bit measurements of an input signal.

1.2 Phase retrieval

The *phase retrieval problem* arises in many mathematical applications whenever an unknown signal must be exactly reconstructed or accurately estimated from squared magnitudes of linear measurements. Reconstruction and estimation from linear measurements is a well-understood problem, but taking the squared magnitude of each linear measurement adds a nonlinear twist and requires new algorithms and proof strategies. This situation was first encountered in crystallography and optics, where physical measurement devices can actually record the squared modulus of an evaluation of the Fourier transform of an unknown signal. In these fields, algorithms were developed to reconstruct signals from such magnitude measurements, along with principles for how many magnitude measurements are required for unique reconstruction [14, 45, 49, 50, 61, 69, 72, 92]. Reconstruction based on magnitudes of linear measurements is also encountered in applications such as automatic speech recognition [15, 80], noise reduction for signal processing [5, 46], and astronomical imaging [48]. Earlier theoretical work also studied to what extent the modulus of the Fourier transform could determine signals uniquely [2, 3].

In all of these applications, the unknown signals that are measured are mathematically modeled by vectors in a real or complex Hilbert space $\mathcal{H}$. The Hilbert space $\mathcal{H}$ is often assumed to be finite-dimensional, i.e., $\mathcal{H} = \mathbb{F}^d$ where $d \in \mathbb{N}$. The linear measurements used are typically linear functionals of the form $x \mapsto \langle x, a \rangle$ for some unit vector $a \in \mathbb{F}^d$, and the associated magnitude measurements that are accessible for the phase retrieval problem then have the form $x \mapsto |\langle x, a \rangle|^2$. Higher rank magnitude measurements have also been studied, for example of the form $x \mapsto \|Px\|_2^2$ for some orthogonal projection P on $\mathbb{F}^d$ [8, 34, 43]. Notice that the formulation of magnitude measurements in terms of orthogonal projections generalizes the linear functional case, since $|\langle x, a \rangle|^2 = \|P_a x\|_2^2$ if $a \in \mathbb{F}^d$ is a unit vector and P_a is the orthogonal projection onto span $\{a\}$.

Given a signal $x \in \mathbb{F}^d$, notice that for any orthogonal projection P on $\mathbb{F}^d$ and any unimodular constant $\alpha \in \mathbb{T}_{\mathbb{F}}$ that the magnitude measurement of αx satisfies

$$\|P(\alpha x)\|_2^2 = |\alpha|^2 \|Px\|_2^2 = \|Px\|_2^2 . \tag{3}$$

In other words, the vectors x and αx are indistinguishable under the magnitude measurements used in phase retrieval problems. For this reason, maps of the form $x \mapsto \|Px\|_2^2$ are called *phaseless measurements*: they lose all information about the global phase of the measured signal. Using phaseless measurements means that it does not make sense to talk about exact reconstruction or even approximate estimation in Euclidean norm of a signal x as an individual vector, since $-x$ is a distinct vector that is separated from x by a Euclidean distance of $\|x - (-x)\|_2 = 2$ but which always has the same phaseless measurements as x. For this reason, the goal of signal reconstruction from phaseless measurements must be modified to be reconstruction or estimation of signals in $\mathbb{F}^d$ up to a unimodular multiplicative constant, i.e., up to a global phase factor. Formally, the quotient space $\mathbb{F}^d/\mathbb{T}_{\mathbb{F}}$ denotes the equivalence classes of vectors in $\mathbb{F}^d$ that differ by a unimodular multiplicative constant in $\mathbb{T}_{\mathbb{F}}$: for any $x \in \mathbb{F}^d$, its equivalence class is defined by $[x] = \{\alpha x : \alpha \in \mathbb{T}_{\mathbb{F}}\} \in \mathbb{F}^d/\mathbb{T}_{\mathbb{F}}$.

Remark 1.2.1. There is a natural identification of the quotient space $\mathbb{F}^d/\mathbb{T}_{\mathbb{F}}$ with the space of positive rank-one Hermitian operators on $\mathbb{F}^d$ via the bijection $[x] \mapsto xx^*$. The operator xx^* is defined by $xx^*(y) = \langle y, x \rangle x$. This is a rank-one operator since $\mathrm{Ran}(xx^*) = \mathrm{span}\,\{x\}$. The notation comes from thinking of vectors $x \in \mathbb{F}^d$ as $d \times 1$ matrices, in which case x^* is the conjugate transpose and the matrix product xx^* is then a rank-one $d \times d$ matrix. The map $[x] \mapsto xx^*$ map is well-defined since

$$(\alpha x)(\alpha x)^* = \alpha\bar{\alpha}xx^* = |\alpha|\, xx^* = xx^*$$

for all $\alpha \in \mathbb{T}_{\mathbb{F}}$. It is surjective since all positive rank-one Hermitians X have the form $X = xx^*$ for some $x \in \mathbb{F}^d$ [63]. To see injectivity, observe that if $X = x_1 x_1^*$ and $X = x_2 x_2^*$ for some nonzero $x_1, x_2 \in \mathbb{F}^d$ then

$$\|x_2\|_2^2\, x_2 = x_2 x_2^* x_2 = X x_2 = x_1 x_1^* x_2 = \langle x_2, x_1 \rangle\, x_1 \qquad \Longrightarrow \qquad x_2 = \frac{\langle x_2, x_1 \rangle}{\|x_2\|_2^2} x_1. \tag{4}$$

Letting $\alpha = \frac{\langle x_2, x_1 \rangle}{\|x_2\|_2^2}$, it follows that

$$X = x_2 x_2^* = |\alpha|^2\, x_1 x_1^* = |\alpha|^2\, X.$$

Taking the operator norm of each side shows that $|\alpha|^2 = 1$, hence $\alpha \in \mathbb{T}_{\mathbb{F}}$. Equation 4 says that $x_2 = \alpha x_1$, and thus $[x_1] = [x_2]$ by definition of $\mathbb{F}^d/\mathbb{T}_{\mathbb{F}}$. This shows that the map $[x] \mapsto xx^*$ is in fact a bijection. The phaseless measurements considered in phase retrieval problems may also be expressed in terms of these rank-one Hermitians since $\|Px\|_2^2 = \operatorname{tr}[Pxx^*]$ for any $x \in \mathbb{F}^d$ and orthogonal projection P on $\mathbb{F}^d$. Henceforth, $\mathbb{F}^d/\mathbb{T}_{\mathbb{F}}$ will be implicitly identified with the space of positive rank-one Hermitians, and phase retrieval will be formulated in terms of measuring and reconstructing these operators $X \in \mathbb{F}^d/\mathbb{T}_{\mathbb{F}}$.

Definition 1.2.2. *For a collection of orthogonal projections $\mathcal{P} = \{P_j\}_{j=1}^m$ on $\mathbb{F}^d$, the **phaseless measurement** associated to $\mathcal{P}$ is defined to be the map $\mathcal{M}_{\mathcal{P}} : \mathbb{F}^d/\mathbb{T}_{\mathbb{F}} \to \mathbb{R}^m$ given by*

$$\mathcal{M}_{\mathcal{P}}(X) = (\operatorname{tr}[PX])_{j=1}^m.$$

The goal in phase retrieval problems is to use the phaseless measurement of an input signal $\mathcal{M}_{\mathcal{P}}(X)$ to completely determine the measured operator X, or at least to determine an estimate that is close to X in some metric. This is accomplished by a **reconstruction algorithm**, which is a map $\mathcal{R} : \mathbb{R}^m \to \mathbb{F}^d/\mathbb{T}_\mathbb{F}$ which takes a phaseless measurement and outputs a positive rank-one Hermitian.

1.2.1 Exact phase retrieval

In the absence of noise that could perturb the phaseless measurement $\mathcal{M}_\mathcal{P}(X)$ of a signal $X \in \mathbb{F}^d/\mathbb{T}_\mathbb{F}$ it is desirable to ask for a measurement and reconstruction scheme that recovers the measured signal exactly. The weakest formulation of this problem deals with exact recovery of a single input signal.

Problem 1.2.3 (Exact phase retrieval - fixed input)**.** *Let $X \in \mathbb{F}^d/\mathbb{T}_\mathbb{F}$ be an arbitrary input signal. Choose orthogonal projections $\mathcal{P} = \{P_j\}_{j=1}^m$ on $\mathbb{F}^d$ and a reconstruction algorithm $\mathcal{R}$ as defined above such that*

$$\mathcal{R} \circ \mathcal{M}_\mathcal{P}(X) = X.$$

As a stronger version of Problem 1.2.3, one can search for a single collection of projections and a single reconstruction algorithm that provides exact phase retrieval for all input signals. The slight change in the order of quantifiers makes the desired property much stronger. A measurement and reconstruction scheme that provides simultaneous phase retrieval of all inputs is more useful than one that is only guaranteed to work for a fixed input; once the measurement $\mathcal{M}_\mathcal{P}$ and the reconstruction procedure $\mathcal{R}$ are implemented, they can be used for any encountered unknown signal without modification.

Problem 1.2.4 (Exact phase retrieval - all inputs)**.** *Choose orthogonal projections $\mathcal{P} = \{P_j\}_{j=1}^m$ on $\mathbb{F}^d$ and a reconstruction algorithm $\mathcal{R}$ as defined above such that*

$$\mathcal{R} \circ \mathcal{M}_\mathcal{P}(X) = X$$

for all $X \in \mathbb{F}^d/\mathbb{T}_\mathbb{F}$.

Exact recovery as in Problem 1.2.4 means that composition of the measurement and reconstruction maps $\mathcal{R} \circ \mathcal{M}_{\mathcal{P}}$ is the identity on $\mathbb{F}^d/\mathbb{T}_{\mathbb{F}}$, which implies the phaseless measurement $\mathcal{M}_{\mathcal{P}}$ must be injective. On the other hand, if $\mathcal{M}_{\mathcal{P}}$ is injective, then choosing $\mathcal{R} = \mathcal{M}_{\mathcal{P}}^{-1}$ would result in exact recovery. So Problem 1.2.4 is equivalent to finding orthogonal projections $\mathcal{P} = \{P_j\}_{j=1}^m$ such that the phaseless measurement $\mathcal{M}_{\mathcal{P}}$ is injective. This is more commonly called the *phase retrieval injectivity problem*, and is formally stated below for reference.

Problem 1.2.5 (Phase retrieval injectivity)**.** *Classify collections of projections $\mathcal{P} = \{P_j\}_{j=1}^m$ on $\mathbb{F}^d$ such that $\mathcal{M}_{\mathcal{P}}$ is injective on $\mathbb{F}^d/\mathbb{T}_{\mathbb{F}}$.*

There has been much activity related to the phase retrieval injectivity problem in the last fifteen years, focusing on finding properties of the collection of projections $\mathcal{P}$ that are equivalent to injectivity of $\mathcal{M}_{\mathcal{P}}$ and finding the minimal cardinality out of all collections $\mathcal{P}$ that give injectivity on $\mathbb{F}^d/\mathbb{T}_{\mathbb{F}}$ for a fixed dimension d. The phase retrieval injectivity problem when $\mathcal{P}$ is restricted to consist of only rank-one projections is closely related to frame theory, since $\operatorname{span}\{\operatorname{Ran}(P) : P \in \mathcal{P}\} = \mathbb{F}^d$ is a necessary condition for $\mathcal{M}_{\mathcal{P}}$ to be injective.

In the case when $\mathbb{F} = \mathbb{R}$, and restricting to only rank-one projections for the phaseless measurement, Problem 1.2.5 has been solved completely: any frame of $m = 2d - 1$ rank-one projections on $\mathbb{R}^d$ which satisfies the *complement property* results in a phaseless measurement $\mathcal{M}_{\mathcal{P}}$ which is injective [10, Theorem 2.8], and all collections of $m < 2d - 1$ rank-one projections give a non-injective phaseless measurement [10, Proposition 2.5]. A collection of rank-one orthogonal projections $\mathcal{P}$ on $\mathbb{R}^d$ has the *complement property* if and only if for every subset $\mathcal{S} \subset \mathcal{P}$ either $\bigvee_{P \in \mathcal{S}} P = I$ or $\bigvee_{P \in \mathcal{S}^c} P = I$, where I denotes the identity operator on $\mathbb{F}^d$ and $\bigvee$ denotes the lattice-theoretic join of orthogonal projections. This is a straightforward translation of the usual definition of the complement property for a frame of vectors, found in [10]. In particular, the rank-one projections associated to a generic choice of $2d - 1$ vectors in $\mathbb{F}^d$ will have the complement property [10, Theorem 2.2].

Real phase retrieval injectivity with higher rank projections has also been studied, yielding an elegant characterization of collections $\mathcal{P} = \{P_j\}_{j=1}^m$ of orthogonal projections of arbitrary ranks that yield an injective phaseless measurement: the phaseless measurement $\mathcal{M}_{\mathcal{P}}$ is injective on $\mathbb{R}^d/\mathbb{T}_{\mathbb{R}}$ if and only if $\{P_j x\}_{j=1}^m$ spans the whole space $\mathbb{R}^d$ for all $x \in \mathbb{R}^d$ [43, Theorem 1.1]. Furthermore, a generic choice of $m = 2d - 1$ non-trivial projections of any ranks will yield injectivity, which generalizes the result mentioned previously for a generic choice of rank-one projections [43, Theorem 1.4]. The minimal number of higher rank projections necessary for injectivity has not been settled except in special cases. If $d = 2^k + 1$ for some $k \in \mathbb{N}$, all collections of $m < 2d - 1$ projections fail to give injective phaseless measurements, so the minimal number of projections in this case is in fact $2d - 1$ even when allowing higher-rank projections [43, Theorem 1.6]. Further complicating the picture, there is a collection of $m = 6$ projections providing injectivity on $\mathbb{R}^4$, meaning that the $2d - 1$ is not the minimal number in general for the real case with higher rank projections [94, Theorem 4.2]. Working with slightly more general phaseless measurements, recent work has resolved the minimal measurement number in some more specific cases and provided general non-trivial upper and lower bounds: for example, 6 projections in $\mathbb{R}^4$ is now known to be optimal [93].

For complex signals, even specifying to rank-one projections for the phaseless measurement, the minimal number of projections required for injectivity of $\mathcal{M}_{\mathcal{P}}$ is still unknown. That being said, some bounds on the minimal number have been found and improved over time [10, 35, 93]. It has been shown that $m = 4d - 4$ generic rank-one measurements yield an injective phaseless measurement on $\mathbb{C}^d$ [35], and a constructive method for choosing $4d - 4$ rank-one measurements that give injectivity on $\mathbb{C}^d$ has also been demonstrated [18]. Despite what these results might indicate, it is also known that $m = 4d - 4$ is *not* the minimal number of rank-one projections for complex signals in general, since a collection of 11 rank-one measurements on $\mathbb{C}^4$ has been proven to yield injectivity [91]. Some general bounds on the minimal number are derived in [93], and the exact minimal number is determined in some special cases for slightly more general measurements.

The minimal number of projections for an injective phaseless measurement will not play a big role in one-bit phase retrieval; the above results are stated mainly to serve as a contrast to the typical results used in one-bit signal processing problems. Whereas solutions to the phase retrieval injectivity problem are often constructive and tend to rely on linear algebra or algebraic geometry, the one-bit phase retrieval results derived in Chapter 2 and Chapter 3 are probabilistic in nature.

1.2.2 Stable phase retrieval from noisy measurements

In practice, requiring exact phase retrieval of a signal is unnecessary or impossible due to the presence of noise in the phaseless measurement. In this case, knowing the minimal number of projections to provide injectivity is not particularly useful without guarantees for how well the reconstruction algorithm performs on noisy measurements. Still, it is desirable to accurately estimate the input signal from noisy phaseless measurements under suitable assumptions on the level of noise. To model this scenario, let $\mathcal{E} : \mathbb{R}^m \to \mathbb{R}^m$ be a *noise map*. After taking the phaseless measurement of a signal $\mathcal{M}_{\mathcal{P}}(X)$, the noise map is applied yielding the *noisy phaseless measurement* $\mathcal{E} \circ \mathcal{M}_{\mathcal{P}}$ which is then passed through the reconstruction algorithm. There are many different noise models that can be considered, but for the sake of simplicity consider an additive noise model of the form $\mathcal{E}(y) = y + z$ for some $z \in \mathbb{R}^m$ with bounded ℓ_1-norm $\|z\|_1 < \epsilon$, which is a standard noise model for phase retrieval [26, 67]. To judge the effectiveness of a particular measurement and reconstruction scheme in the presence of noise, the reconstruction error can be measured with either the *operator norm distance* $\|X - Y\|$ or the *Hilbert-Schmidt distance* $\|X - Y\|_{HS} = \sqrt{\operatorname{tr}\left[(X - Y)^*(X - Y)\right]}$.

Problem 1.2.6 (Stable phase retrieval - fixed input)**.** *Consider a fixed input signal $X \in \mathbb{F}^d/\mathbb{T}_{\mathbb{F}}$. For a desired reconstruction accuracy $\delta > 0$ and noise term $z \in \mathbb{R}^m$ with $\|z\|_1 < \epsilon$, choose orthogonal projections $\mathcal{P} = \{P_j\}_{j=1}^m$ and a reconstruction algorithm $\mathcal{R}$ as defined above such that*

$$\left\|\hat{X} - X\right\| < \delta,$$

where $\hat{X} = \mathcal{R}(\mathcal{M}_{\mathcal{P}}(X) + z)$.

Before discussing some of the relevant literature and existing algorithms for stable phase retrieval, it is worth stating a stronger version of Problem 1.2.6 that requires *uniform* reconstruction of all input signals using a fixed measurement and reconstruction procedure. The desire for such a uniform result is motivated by real-world constraints: generating and implementing new set of projections $\mathcal{P}$ for each measured signal is an unnecessary burden if a single $\mathcal{P}$ can work for all signals simultaneously.

Problem 1.2.7 (Uniformly accurate stable phase retrieval)**.** *For a desired reconstruction accuracy $\delta > 0$ and noise term $z \in \mathbb{R}^m$ with $\|z\|_1 < \epsilon$, choose orthogonal projections $\mathcal{P} = \{P_j\}_{j=1}^m$ and a reconstruction algorithm $\mathcal{R}$ as defined above such that*

$$\left\|\hat{X} - X\right\| < \delta,$$

for all $X \in \mathbb{F}^d/\mathbb{T}_{\mathbb{F}}$, where $\hat{X} = \mathcal{R}(\mathcal{M}_{\mathcal{P}}(X) + z)$.

There are many constructions of frames of vectors, or collections of higher-rank projections, that have been shown to allow stable phase retrieval [8, 9, 11, 19]. Finding solutions to Problem 1.2.6 or Problem 1.2.7 as stated is already an interesting problem, but there is one additional constraint that comes with real-world applications: if the solution is to be applied in practice, the reconstruction algorithm $\mathcal{R}$ must be computationally feasible, not abstractly defined and impossible to implement effectively.

The PhaseLift procedure given in [27] is notable for being the first algorithm proven to provide stable phase retrieval of fixed input signals via a computationally feasible reconstruction algorithm, solving Problem 1.2.6 in an efficient way. PhaseLift uses *random* rank-one projections for the phaseless measurement and the reconstruction algorithm is implemented through solving a semidefinite program. The PhaseLift algorithm and its proven guarantees for effectiveness illustrate what solutions to Problems 1.2.6, 1.2.7 and other similar stable recovery problems tend to look like. Some of the details are provided below as an example.

1.2.3 Example: PhaseLift [26, 27]

Let $X \in \mathbb{F}^d/\mathbb{T}_{\mathbb{F}}$ be an arbitrary input signal and let $\mathcal{P} = \{P_j\}_{j=1}^m$ be a collection of rank-one orthogonal projections on $\mathbb{F}^d$. Given the phaseless measurement $b = \mathcal{M}_{\mathcal{P}}(X)$, choose $\mathcal{R}(\mathcal{M}_{\mathcal{P}}(X))$

to be a minimizer of the semidefinite program

$$\begin{aligned} \underset{Y}{\text{minimize}} \quad & \operatorname{tr}[Y] \\ \text{subject to} \quad & \mathcal{M}_{\mathcal{P}}(Y) = b,\ Y \succeq 0. \end{aligned} \tag{5}$$

Minimizing the trace promotes low-rank solutions, and if the number of projections m is large enough then X is the exact solution to the program. In other words, under suitable assumptions the semidefinite program (5) is a computationally feasible reconstruction algorithm for exact phase retrieval. This result is stated formally in the below theorem.

Theorem 1.2.8 (Exact phase retrieval via Phaselift)**.** *Let $X \in \mathbb{F}^d/\mathbb{T}_{\mathbb{F}}$ be an arbitrary input signal and*

$$m \geq c_0 d$$

where c_0 is a dimension-independent constant. If $\mathcal{P} = \{P_j\}_{j=1}^m$ is a collection of rank-one orthogonal projections chosen independently and uniformly at random, then the minimizer of the semidefinite program (5) is X with probability at least $1 - O(\exp(-\gamma m))$ for a constant $\gamma > 0$.

Suppose the phaseless measurement is further passed through an additive noise map as in Problem 1.2.6, i.e., yielding a noisy phaseless measurement $b = \mathcal{M}_{\mathcal{P}}(X) + z$ for some noise term $z \in \mathbb{R}^m$. In this case, the semidefinite program (5) may be modified to deal with the presence of noise by defining $\mathcal{R}(\mathcal{M}_{\mathcal{P}}(X) + z)$ to be the minimizer $\hat{X}$ of the ℓ_1-minimization problem

$$\begin{aligned} \underset{Y}{\text{minimize}} \quad & \|\mathcal{M}_{\mathcal{P}}(Y) - b\|_1 \\ \text{subject to} \quad & Y \succeq 0. \end{aligned} \tag{6}$$

Like the program in (5), this is a computationally tractable minimization problem. With a large enough number of projections m, the algorithm (6) yields stable and accurate reconstruction.

Theorem 1.2.9 (Stable phase retrieval via PhaseLift)**.** *Let $X \in \mathbb{F}^d/\mathbb{T}_{\mathbb{F}}$ be an arbitrary input* *signal, $\delta > 0$ be a desired level of accuracy, and $\epsilon > 0$ be a noise level. If*

$$m \geq c_0 d\epsilon^{-1}\delta^{-1}$$

where c_0 is a dimension-independent constant, $\mathcal{P} = \{P_j\}_{j=1}^m$ is a collection of rank-one orthogonal projections chosen independently and uniformly at random, and $z \in \mathbb{R}^m$ is an additive error term with $\|z\|_1 < \epsilon$, then with probability similar to that in Theorem 1.2.8

$$\left\|\hat{X} - X\right\|_{HS} < \delta,$$

where $\hat{X}$ is the solution to the semidefinite program (6).

Unlike the noiseless case, the minimizer $\hat{X}$ is not guaranteed to be rank-one, but it can be used to find a good rank-one estimate for the input X. Since $\hat{X}$ is a positive Hermitian operator, by the spectral theorem it may be decomposed as $\hat{X} = \sum_{k=1}^d \lambda_k P_k$ for some $\lambda_1 \geq \ldots \geq \lambda_d \geq 0$ and rank-one orthogonal projections P_k. Taking the largest rank-one component $\lambda_1 P_1$ results in a good rank-one approximation to X.

The PhaseLift procedure may be generalized to higher-rank orthogonal projections while retaining stability in the presence of noise. For any rank $1 \leq k \leq d-1$, using $m \geq c_0 d$ independently and uniformly random subspaces of rank k yields accurate reconstruction via (6) with high probability analogously to Theorem 1.2.9 [8, Theorem 5.11]. A modification of the PhaseLift procedure using a different measurement scheme has been shown to provide uniformly accurate stable phase retrieval, solving Problem 1.2.7 [67]. A further development shows that solving a semidefinite program is not even necessary for exact (or stable) phase retrieval of a fixed input: with $m \geq c_o d\log(d)$ random rank-one projections, $\mathcal{M}_{\mathcal{P}}$ is injective (or stable) on $\mathbb{F}^d/\mathbb{T}_{\mathbb{F}}$ with high probability [37]. More in line with results in Section 1.2.1, there is a deterministic construction of $m = 5d - 6$ rank-one projections on $\mathbb{C}^d$ for which the PhaseLift algorithm provides an injective phaseless measurement, and thus exact recovery [66].

1.3 One-bit phase retrieval

Phase retrieval as in Problem 1.2.4 and Problem 1.2.6 can be thought of as signal recovery from linear measurements after losing some of the information contained in each measurement in a nonlinear way; instead of measuring the exact value of each linear functional $\langle x, f_j \rangle$, only the squared magnitude $|\langle x, f_j \rangle|^2$ is retained, and the original input signal x is still to be determined from these magnitude measurements up to a unimodular multiplicative constant. In *one-bit phase retrieval*, even more information is lost from these magnitude measurements until only a single bit of information is retained. Examples of quantization methods include comparing each measured quantity to a fixed threshold or comparing pairs of measured quantities to each other.

One-bit phase retrieval involves two key steps: taking a phaseless binary measurement of the input signal, and then estimating the input signal based on that binary measurement. The first step, taking a phaseless binary measurement, may be split further into taking a non-quantized phaseless measurement as in regular phase retrieval, and then quantizing the measured quantities to one bit of information, a 1 or a 0. Before giving the general statement of the problem, the various maps that are used in these steps must be defined.

Let $K \subset \mathbb{F}^d/\mathbb{T}_{\mathbb{F}}$ be a set of possible signals, where as before $\mathbb{F}^d/\mathbb{T}_{\mathbb{F}}$ represents the equivalence classes of vectors in $\mathbb{F}^d$ that differ by a unimodular multiplicative constant in $\mathbb{F}$. The quotient space $\mathbb{F}^d/\mathbb{T}_{\mathbb{F}}$ is identified with the set of positive rank-one Hermitian operators on $\mathbb{F}^d$ as discussed in Remark 1.2.1. For a collection of orthogonal projections (of any ranks) $\mathcal{P} = \{P_j\}_{j=1}^{m'}$ on $\mathbb{F}^d$, the **(non-quantized) phaseless measurement** given by $\mathcal{P}$ is the map $\mathcal{M}_{\mathcal{P}} : K \to \mathbb{R}^{m'}$ defined by $\mathcal{M}_{\mathcal{P}}(X) = (\mathrm{tr}\,[P_j X])_{j=1}^{m'}$, exactly as defined in Definition 1.2.2. In the one-bit phase retrieval problem, the phaseless measurements are further passed through a **binary quantization map** $\Phi : \mathbb{R}^{m'} \to \{0,1\}^m$. Together, $\mathcal{M}_{\mathcal{P}}$ and Φ give a phaseless binary measurement.

Definition 1.3.1. *For a collection of orthogonal projections $\mathcal{P}$ and a binary quantization map Φ as above, the* ***phaseless binary measurement*** *given by $\mathcal{P}$ and Φ is defined to be the map $\Phi_{\mathcal{P}} := \Phi \circ \mathcal{M}_{\mathcal{P}}$.*

In the last step of one-bit phase retrieval, an element of K is selected based on the binary measurement via an **estimation algorithm** $\mathcal{R} : \{0,1\}^m \to K$. The output of $\mathcal{R}$ based on the phaseless binary measurement of X is denoted $\hat{X} := \mathcal{R}(\Phi_{\mathcal{P}}(X))$ to further simplify notation. The accuracy of the recovery algorithm $\mathcal{R}$ is judged by comparing how close the estimate $\hat{X}$ is to the measured signal X in operator norm distance $\left\|\hat{X} - X\right\|$.

For a given reconstruction algorithm $\mathcal{R} : \{0,1\}^m \to \mathbb{F}^d/\mathbb{T}_{\mathbb{F}}$ to provide uniformly accurate recovery, $\hat{X} = \mathcal{R}(\Phi_{\mathcal{P}}(X))$ must be close to X in operator norm for all input signals X, i.e., $\left\|\hat{X} - X\right\| < \delta$ for some desired accuracy $\delta > 0$. If $\Phi_{\mathcal{P}}(X)$ and $\Phi_{\mathcal{P}}(Y)$ are close together for two distinct input signals $X, Y \in \mathrm{Proj}_{\mathbb{F}}(1,d)$, then ideally X and Y would be close to each other in operator norm, otherwise the reconstruction algorithm would not be robust under a bit-flip error that changes $\Phi_{\mathcal{P}}(X)$ into $\Phi_{\mathcal{P}}(Y)$. The Hamming distance is used to measure the distance between the phaseless binary measurements of different input signals.

Definition 1.3.2. *Let $\Phi_{\mathcal{P}}$ be a phaseless binary measurement. The **measurement Hamming distance** associated with $\Phi_{\mathcal{P}}$ between X and Y is defined to be*

$$d_{\mathcal{P}}(X,Y) := d_H(\Phi_{\mathcal{P}}(X), \Phi_{\mathcal{P}}(Y))$$

where d_H denotes the normalized Hamming distance on $\{0,1\}^m$ defined by

$$d_H(x,y) = \frac{1}{m} \left|\{j : x_j \neq y_j\}\right| .$$

The value $d_{\mathcal{P}}(X,Y)$ gives the fraction of binary questions in the phaseless binary measurement that yield different answers for X and Y as inputs. Note that $d_{\mathcal{P}}$ is a pseudometric, not a metric, since distinct input signals can potentially yield identical phaseless binary measurements.

Remark 1.3.3. Suppose $K \subset \mathbb{F}^d/\mathbb{T}_{\mathbb{F}}$ is an unbounded signal set. Since $\Phi_{\mathcal{P}}$ has a finite range of at most 2^m elements, there exists a binary string $b \in \{0,1\}^m$ such that the pre-image $\Phi_{\mathcal{P}}^{-1}(b)$ is unbounded. The recovery algorithm $\mathcal{R}$ must assign a single element in K to b, but since $\Phi_{\mathcal{P}}^{-1}(b)$

is unbounded, for any choice of $\mathcal{R}(b)$ there exists some $Y \in \Phi_{\mathcal{P}}^{-1}(b)$ with operator norm distance $\|Y - \mathcal{R}(b)\|$ arbitrarily large. In other words, there is no way to accurately estimate all signals in an unbounded signal from a phaseless binary measurement consisting of a finite number of bits.

For this reason, it is assumed that the norm of the input signal is already known, so that without loss of generality only unit-norm input signals need to be considered. This assumption is further justified by the fact that in high dimensions the norm of a random input signal concentrates near a fixed dimension-dependent value [90]. Notice that for any $x \in \mathbb{F}^d$, the operator norm of its associated positive rank-one Hermitian xx^* satisfies $\|xx^*\| = \|x\|_2^2$, so working with unit-norm vectors $x \in \mathbb{S}_{\mathbb{F}}^{d-1}$ may be phrased as working with positive rank-one Hermitians with unit operator norm via the identification in Remark 1.2.1. The positive rank-one Hermitians with operator norm are exactly the rank-one orthogonal projections, so henceforth the space of input signals for one-bit phase retrieval will always be $K = \operatorname{Proj}_{\mathbb{F}}(1, d) \subset \mathbb{F}^d/\mathbb{T}_{\mathbb{F}}$.

Principal angles and operator norm distance

After specializing to unit-norm input signals as stated in Remark 1.3.3, the operator norm distance between two rank-one projections is commonly encountered. For example, a reconstruction algorithm $\mathcal{R}$ outputs a rank-one projection $\hat{X}$, and the operator norm is used to measure how close this recovered signal is to the original input X by computing $\left\|\hat{X} - X\right\|$. The operator norm distance between two rank-one orthogonal projections $X, Y \in \operatorname{Proj}_{\mathbb{F}}(1, d)$ has the property that $\|X - Y\| = \sin(\theta)$ where θ is the principal angle between the one-dimensional subspaces $\operatorname{Ran}(X)$ and $\operatorname{Ran}(Y)$ as defined below.

Definition 1.3.4 ([53])**.** *Given a pair of subspaces $V, W \subset \mathbb{F}^d$ with $\dim(V) = k_1$ and $\dim(W) = k_2$, the* ***principal angles*** *between $\theta_j \in [0, \frac{\pi}{2}]$ are recursively defined for $j = 1, \dots, \min(k_1, k_2)$ by*

$$\cos(\theta_j) = \max_{\substack{v \in V, w \in W \\ \|v\|_2 = \|w\|_2 = 1 \\ \langle v, v_i \rangle = \langle w, w_i \rangle = 0,\ i = 1, \dots, j-1}} |\langle v, w \rangle| = \langle v_j, w_j \rangle .$$

The vectors v_j and w_j are called the ***principal vectors*** *of V and W respectively.*

The fact that $\|X - Y\| = \sin(\theta)$ where θ is the principal angle between the ranges of X and Y follows from a useful fact about the spectral decomposition of $X - Y$ that will also be used in later chapters.

Lemma 1.3.5. *Let $X, Y \in \mathrm{Proj}_{\mathbb{F}}(1, d)$. Then the spectral decomposition of $X - Y$ is*

$$X - Y = \sin(\theta)(A - B)$$

where θ is the principal angle between the subspaces $\mathrm{Ran}(X)$ and $\mathrm{Ran}(Y)$ and $A, B \in \mathrm{Proj}_{\mathbb{F}}(1, d)$ satisfy $\mathrm{Ran}(A) \perp \mathrm{Ran}(B)$. In particular,

$$\|X - Y\| = \sin(\theta).$$

Proof. If $X = Y$, then the principal angle in question is $\theta = 0$, and $\|X - Y\| = 0 = \sin(0)$ trivially. Suppose next that $X \neq Y$. Then $(X - Y)z = 0$ if and only if $Xz = Yz$, which happens if and only if $\langle x, z\rangle = \langle y, z\rangle = 0$. Thus $\ker(X - Y) = \mathrm{span}\left\{\mathrm{Ran}(X), \mathrm{Ran}(Y)\right\}^{\perp}$. Since $X - Y$ is Hermitian, it follows that $\mathrm{Ran}(X - Y)^{\perp} = \ker(X - Y)$, and hence $\mathrm{Ran}(X - Y) = \mathrm{span}\left\{\mathrm{Ran}(X), \mathrm{Ran}(Y)\right\}$.

By the spectral theorem and the fact that $\mathrm{Rank}(X - Y) = 2$ from above, the difference $X - Y$ may be expressed as as

$$X - Y = \lambda_1 A + \lambda_2 B, \tag{7}$$

where $\lambda_1 \geq \lambda_2 \in \mathbb{R}$ and $A, B \in \mathrm{Proj}_{\mathbb{F}}(1, d)$ with $A \perp B$. Clearly $\mathrm{tr}\left[X - Y\right] = 0$, which implies $\lambda_2 = -\lambda_1$ and hence $\lambda_1^2 = \lambda_2^2$. Furthermore, it follows that

$$2\lambda_1^2 = \lambda_1^2 + \lambda_2^2 = \mathrm{tr}\left[(X - Y)^2\right] = \mathrm{tr}\left[X + Y - XY - YX\right] = 2(1 - \mathrm{tr}\left[XY\right]). \tag{8}$$

Letting $x \in \mathrm{Ran}(X)$ and $y \in \mathrm{Ran}(Y)$ be principal vectors for $\mathrm{Ran}(X)$ and $\mathrm{Ran}(Y)$, by definition $\langle x, y\rangle = \cos(\theta)$. Also, since $\|x\|_2 = \|y\|_2 = 1$, the projections X and Y may be expressed as

$X = xx*$ and $Y = yy^*$. Thus

$$\operatorname{tr}[XY] = \operatorname{tr}[xx^*yy^*] = \operatorname{tr}[y^*xxy^*] = \langle x, y\rangle^2 = \cos^2(\theta).$$

From (8) it follows that $\lambda_1^2 = \sin^2(\theta)$, and hence $\lambda_1 = \sin(\theta)$ and $\lambda_2 = -\sin(\theta)$ since $\lambda_1 \geq \lambda_2$ and $\lambda_1 + \lambda_2 = 0$. □

Also, observe that for a unit-norm input $X \in \operatorname{Proj}_{\mathbb{F}}(1, d)$ and orthogonal projection $P \in \operatorname{Proj}_{\mathbb{F}}(k, d)$, the magnitude measurement $\operatorname{tr}[PX] = \cos^2(\theta)$ where θ is the principal angle between $\operatorname{Ran}(P)$ and $\operatorname{Ran}(X)$. If $x \in \operatorname{Ran}(X)$ and $p \in \operatorname{Ran}(P)$ are principal vectors, then $PX = pp^*xx^*$, so

$$\operatorname{tr}[PX] = \operatorname{tr}[pp^*xx^*] = |\langle p, x\rangle|^2 = \cos^2(\theta).$$

Noiseless one-bit phase retrieval

Now that the relevant maps have been defined, the simplest one-bit phase retrieval problem may be stated: accurate reconstruction of a fixed input signal from a phaseless binary measurement and reconstruction algorithm.

Problem 1.3.6 (One-bit phase retrieval - fixed input)**.** *Let $X \in \operatorname{Proj}_{\mathbb{F}}(1, d)$ be an arbitrary input signal. For a desired maximum recovery error $\delta > 0$, choose a phaseless binary measurement $\Phi_{\mathcal{P}}$ and an estimation algorithm $\mathcal{R}$ such that*

$$\left\|\hat{X} - X\right\| < \delta.$$

Solutions to Problem 1.3.6 follow the trend introduced with the PhaseLift algorithm in [27] and [24] of choosing the phaseless measurement by random selection. Empirical results in [74] show that traditional algorithms for phase retrieval do not adapt well to binary quantization of measurements and they instead propose an algorithm based on gradient descent. Results in one-bit compressed sensing can be applied to Problem 1.3.6 as a special case, with some slight modification of the type of measurement used. For example, $m = C\delta^{-4}n$ random one-bit measurements (of the form $X \mapsto \operatorname{sign}(\operatorname{tr}[G_jX])$ for $\{G_j\}_{j=1}^m$ independent matrices with independent standard Gaussian entries) are sufficient to recover $\hat{X}$ with nuclear norm $\operatorname{tr}\left[\left|\hat{X}\right|\right] = 1$ and Hilbert-Schmidt norm $\left\|\hat{X}\right\|_2^2 \leq 1$ such that the Hilbert-Schmidt error satisfies $\left\|\hat{X} - X\right\|_{HS} < \delta$ [78, Section 3.3]. Another result on one-bit phase retrieval also gives comparable asymptotics when using measurements based on comparing pairs of rank-one magnitude measurements [73].

Chapter 2 of this dissertation provides several phaseless binary measurements and an associated reconstruction algorithm that together solve Problem 1.3.6. The reconstruction procedure studied there is called *Principal Eigenspace Programming (PEP)*. With high probability, PEP recovers an estimate $\hat{X}$ of an input signal X that satisfies $\left\|\hat{X} - X\right\| < \delta$ when $m \geq C\delta^{-2}n\log(n)$ random projections of rank equal to half the dimension of the signal are selected for the binary phaseless measurement. Here, C is a constant independent of n and δ. See Theorem 2.4.4 for details. This is an improvement over [78] and [73] in terms of the number of one-bit measurements sufficient to guarantee accurate recovery with high probability.

In the same way that Problem 1.2.4 is a stronger version of Problem 1.2.3 and Problem 1.2.7 is a stronger version of Problem 1.2.6, there is a stronger version of Problem 1.3.6 that is of more practical use: can $\Phi_{\mathcal{P}}$ and $\mathcal{R}$ be selected in such a way that applying the estimation algorithm to the phaseless binary measurement of $X \in \operatorname{Proj}_{\mathbb{F}}(1, d)$ gives a good approximation for X, for all $X \in \operatorname{Proj}_{\mathbb{F}}(1, d)$? In other words, a single phaseless binary measurement and a single reconstruction method should work for all possible input signals. This is called the *uniformly accurate one-bit phase retrieval problem.*

Problem 1.3.7 (Uniformly accurate one-bit phase retrieval)**.** *For a desired maximum recovery error $\delta > 0$, choose a phaseless binary measurement $\Phi_{\mathcal{P}}$ and an estimation algorithm $\mathcal{R}$ such that*

$$\left\|\hat{X} - X\right\| < \delta$$

for all input signals $X \in \operatorname{Proj}_{\mathbb{F}}(1, d)$.

Uniformly accurate signal reconstruction from one-bit measurements is a relatively new problem. The most directly related results come from the field of one-bit compressed sensing, where uniformly accurate reconstruction from one-bit measurements has been studied and recovery guarantees have been proven to hold [78].

In Chapter 2, after providing a solution to Problem 1.3.6, a similar proof strategy as in [78] is used to prove that the binary measurement and reconstruction scheme PEP provides uniformly accurate one-bit phase retrieval of all input signals simultaneously. With high probability, PEP recovers an estimate $\hat{X}$ of an input signal X which satisfies $\left\|\hat{X} - X\right\| < \delta$ for all X for a particular type of phaseless binary measurement of at least $m \geq C\delta^{-2}n^2 \log(\delta^{-1}n)$ bits. See Theorem 2.5.14 for details. To my knowledge, this is the first solution to Problem 1.3.7. Results in [78] on uniformly accurate one-bit compressed sensing may be applied to low-rank matrix recovery as a special case to give a benchmark for one-bit phase retrieval, but it must be noted that the measurements used in compressed sensing do not fit the phaseless measurement model in phase retrieval problems. With that being said, our solution to Problem 1.3.7 improves on [78] in how the number of bits in the binary measurement depends on the desired uniform accuracy.

1.3.1 Noisy one-bit phase retrieval

Like in other signal reconstruction and inverse problems, it is desirable that the reconstruction algorithm for one-bit phase retrieval be robust under measurement error. For one-bit measurements, the most common type of error studied is a bit-flip error, which occurs when one or more of the bits in the binary measurement “flip” from the true value of 1 (or 0) to the incorrect value of 0 (or 1, respectively). Bit-flip errors may be modeled for one-bit phase retrieval by assuming that the phaseless binary measurement $\Phi_{\mathcal{P}}$ passes through a **bit-flip map** $\mathcal{F}_T : \{0,1\}^m \to \{0,1\}^m$ associated to a subset $T \subset \{1, \dots, m\}$, defined by

$$\mathcal{F}_T(x)_j = \begin{cases} 1 - x_j & \text{if } j \in T \\ x_j & \text{else.} \end{cases}$$

The composition $\mathcal{F}_T \circ \Phi_{\mathcal{P}}$ is called the **noisy phaseless binary measurement**, and denoted by $\Phi_{\mathcal{P},T}(X)$ to simplify notation. To further simplify notation, the reconstruction based on the noisy phaseless binary measurement of an input X is denoted $\hat{X}_T := \mathcal{R}(\Phi_{\mathcal{P},T}(X))$.

This dissertation will consider two different noise models for how to select the subset of bit-flips T for the bit-flip map $\mathcal{F}_T$. The first is adversarial noise, which can be though of as "worst case" bit-flips. For adversarial noise, it is assumed that a certain percentage of the bits could flip, and control is sought for the maximum error of reconstruction after all bit-flip maps of that size. The task of performing one-bit phase retrieval of a fixed input vector in the presence of adversarial noise is stated formally in the problem below.

Problem 1.3.8 (Fixed input one-bit phase retrieval, adversarial noise)**.** *Let $X \in \mathrm{Proj}_{\mathbb{F}}(1,d)$ be an arbitrary input signal. For a desired recovery error δ and a maximum bit-flip ratio $\tau > 0$, choose a phaseless binary measurement $\Phi_{\mathcal{P}}$ and an estimation algorithm $\mathcal{R}$ such that*

$$\left\| \hat{X}_T - X \right\| < \delta + r(\tau) \tag{9}$$

for all $T \subset \{1, \dots, m\}$ with $|T| \leq \tau m$.

We include a term $r(\tau)$ in the error bound (9) because adversarial bit-flips may result in increased reconstruction error depending on the ratio of bit-flips performed. Ideally, $r(\tau)$ would depend on the number of bits in the binary measurement m in such a way that $r(\tau) \to 0$ as $m \to \infty$, which would mean for any $\delta_0 > 0$ the number of bits could be chosen to be large enough to ensure $\delta + r(\tau) < \delta_0$ on the right-hand side of (9).

A still more difficult problem is to find measurement and reconstruction algorithms that allow uniformly accurate one-bit phase retrieval as in Problem 1.3.7, but with adversarial bit-flips in the phaseless binary measurement. This problem is formally stated below.

Problem 1.3.9 (Uniformly accurate one-bit phase retrieval, adversarial noise)**.** *For a maximum recovery error δ and a maximum bit-flip ratio $\tau > 0$, choose a phaseless binary measurement $\Phi_{\mathcal{P}}$*

and an estimation algorithm $\mathcal{R}$ as defined above such that

$$\left\|\hat{X}_T - X\right\| < \delta + r(\tau)$$

for all $X \in \mathrm{Proj}_{\mathbb{F}}(1, d)$ and for all $T \subset \{1, \ldots, m\}$ with $|T| \leq \tau m$.

Solutions to Problems 1.3.8 and 1.3.9 are derived in Chapter 3. These results follow as corollaries to the solutions to the noiseless one-bit phase retrieval problems, see Corollary 3.2.1 and Corollary 3.2.2 for details.

1.4 Tools from probability

This section provides definitions of relevant probability distributions, along with a few important probabilistic tools and concepts that are used in proofs. The power of probabilistic methods in signal processing has already been demonstrated in the discussion of phase retrieval in Section 1.2, where several algorithms were mentioned that choose the orthogonal projections $\mathcal{P} = \{P_j\}_{j=1}^m$ at random. For a specific example, Theorems 1.2.8 and 1.2.9 both used rank-one projections chosen independently and uniformly at random from $\mathrm{Proj}_{\mathbb{F}}(1, d)$. These theorems are phrased in such a way that the desirable property (exact/stable reconstruction) holds with high probability with respect to these random projections. This paradigm will be used in Chapter 2 to prove fixed input and uniform reconstruction guarantees for a one-bit phase retrieval algorithm that also randomly selects projections for the phaseless measurement. Chapter 3 shows that these random phaseless measurements are also stable under adversarial and random bit-flips.

Random vectors and related distributions

The concept of random vectors occurs naturally in probability theory; as soon as one investigates two random variables X and Y on $\mathbb{F}$ simultaneously, their joint distribution (X, Y) on $\mathbb{R}^2$ is a random vector that becomes of interest. An $\mathbb{F}$-valued d-dimensional *random vector* is a vector $x = (x_j) \in \mathbb{F}^d$ where each vector component x_j is a random variable on $\mathbb{F}$. An $\mathbb{F}$-valued $d_1 \times d_2$

random matrix is simply a random vector $X = (x_{j,k}) \in \mathbb{F}^{d_1 \times d_2}$ thought of as a $d_1 \times d_2$ matrix. The basic probabilistic tools and concepts for regular $\mathbb{F}$-valued random variables extend to $\mathbb{F}$-valued random vectors and matrices. For example, the expectation of a random vector $x = (x_j) \in \mathbb{F}^d$ is the vector of expected values of its individual entries, i.e., $\mathbb{E}[x] = (\mathbb{E}[x_j])$.

The Gaussian distributions are standard examples of random variables, vectors and matrices. These distributions are defined below, as they are important for constructing other distributions such as random unit vectors and random orthogonal projections.

Definition 1.4.1. *The* ***real-valued Gaussian (or normal) distribution*** *with mean μ and variance σ^2 is the probability distribution on $\mathbb{R}$ given by the probability density function*

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right).$$

The notation $x \sim N(\mu, \sigma^2)$ is used to say that x is a real-valued random variable that has the Gaussian distribution with mean μ and variance σ^2.

Definition 1.4.2. *A d-dimensional* ***real-valued standard Gaussian random vector*** *is a vector $g = (g_j) \in \mathbb{R}^d$ where the entries g_j are independent and each $g_j \sim N(0,1)$. A d-dimensional* ***complex-valued standard Gaussian random vector*** *is a vector $g = (g_j) \in \mathbb{C}^d$ where the entries g_j are complex-valued random variables of the form $g_j = y_j + iz_j$ for independent $y_j \sim N(0, \frac{1}{2})$ and $z_j \sim N(0, \frac{1}{2})$.*

A $d_1 \times d_2$ ***real-valued standard Gaussian random matrix*** *is a real-valued standard Gaussian random vector in $\mathbb{R}^{d_1 \times d_2}$ thought of as a matrix. A $d_1 \times d_2$* ***complex-valued standard Gaussian random matrix*** *is a complex-valued standard Gaussian random vector in $\mathbb{C}^{d_1 \times d_2}$ thought of as a matrix.*

An important property of the Gaussian distribution for random vectors is that it is rotationally invariant, i.e., the distribution is invariant under unitary transformation.

Proposition 1.4.3 (Rotational invariance of standard Gaussian random vector)**.** *Let $g \in \mathbb{F}^d$ be a standard Gaussian random vector and $U \in \mathcal{U}_{\mathbb{F}}(d)$ be a fixed unitary matrix. Then $Ug \overset{(d)}{=} g$.*

In general, the entries of a random vector or random matrix may not be independent of each other. Several particular types of random vectors and matrices with dependent entries will be of interest. One important example is the uniform distribution on unit vectors.

Definition 1.4.4. *Let g be a standard Gaussian random vector in $\mathbb{F}^d$. Then $\frac{g}{\|g\|_2}$, is a random unit vector, called a **uniformly distributed unit vector** in $\mathbb{S}_{\mathbb{F}}^{d-1}$.*

Another important one-dimensional distribution used in later chapters is the beta distribution, which is a probability distribution on the interval $[0,1]$.

Definition 1.4.5. *The **beta distribution** with parameters $a, b > 0$ is the probability distribution on $[0,1]$ given by the probability density function*

$$f(t) = \frac{1}{B(a,b)} t^{a-1}(1-t)^{b-1},$$

where $B(\cdot,\cdot)$ is the beta function defined by

$$B(a,b) := \int_0^1 t^{a-1}(1-t)^{b-1}\ dt.$$

The notation $x \sim \mathrm{Beta}(a,b)$ is used to say that x is a random variable in $[0,1]$ that has the beta distribution with parameters a and b.

The beta distribution arises naturally in the setting of phase retrieval as the distribution of the squared norm of an orthogonal projection of a uniformly distributed unit vector.

Lemma 1.4.6. *Let $x \in \mathbb{S}_{\mathbb{F}}^{d-1}$ be a uniformly distributed unit vector and $P \in \mathrm{Proj}_{\mathbb{F}}(k,d)$ be a fixed rank-k orthogonal projection on $\mathbb{F}^d$. Then $\|Px\|_2^2 \sim \mathrm{Beta}(\beta_{\mathbb{F}}k, \beta_{\mathbb{F}}d)$, where $\beta_{\mathbb{F}} = \frac{1}{2}$ if $\mathbb{F} = \mathbb{R}$ and $\beta_{\mathbb{F}} = 1$ if $\mathbb{F} = \mathbb{C}$.*

Proof. By definition, $x \overset{\text{(d)}}{=} \frac{g}{\|g\|_2}$ where g is a standard Gaussian random vector in $\mathbb{F}^d$. The squared norm of the projection of x can be rewritten using distributional equality to see

$$\|Px\|_2^2 = \langle Px, Px \rangle \overset{\text{(d)}}{=} \frac{1}{\|g\|_2} \langle Pg, Pg \rangle\,.$$

Let U be the unitary matrix such that U^*PU is the projection onto the span of the first k standard basis vectors. Then by the rotational invariance of the distribution of g and the fact that $P^*P = P$ it follows that

$$\frac{1}{\|g\|_2}\langle Pg, Pg\rangle \overset{(d)}{=} \frac{1}{\|g\|_2}\langle PUg, PUg\rangle = \frac{1}{\|g\|_2}\langle U^*PUg, U^*PUg\rangle = \frac{\|U^*PUg\|_2^2}{\|g\|_2^2}.$$

Expanding this ratio of squared norms in terms of the components of g,

$$\frac{\|U^*PUg\|_2^2}{\|g\|_2^2} = \frac{\sum_{j=1}^{k}|g_j|^2}{\sum_{j=1}^{k}|g_j|^2 + \sum_{l=k+1}^{d}|g_l|^2}. \tag{10}$$

If $\mathbb{F} = \mathbb{R}$, then the entries g_j are independent and $g_j \sim N(0,1)$ for every j. Thus $X = sum_{j=1}^{k}|g_j|^2$ is chi-squared distributed with k degrees of freedom, i.e., $X \sim \chi^2(k)$, and similarly $Y = \sum_{l=k+1}^{d}|g_l|^2 \sim \chi^2(d-k)$. The ratio $\frac{X}{X+Y}$ is then beta distributed with parameters $a = \frac{k}{2}$ and $b = \frac{d-k}{2}$. Altogether, this means that

$$\|Px\|_2^2 \overset{(d)}{=} \frac{\sum_{j=1}^{k}|g_j|^2}{\sum_{j=1}^{k}|g_j|^2 + \sum_{l=k+1}^{d}|g_l|^2} = \frac{X}{X+Y} \sim \text{Beta}\left(\frac{k}{2}, \frac{d-k}{2}\right).$$

On the other hand, if $\mathbb{F} = \mathbb{C}$, then multiplying and dividing by 2 gives us

$$\frac{\sum_{j=1}^{k}|g_j|^2}{\sum_{j=1}^{k}|g_j|^2 + \sum_{l=k+1}^{d}|g_l|^2} = \frac{\sum_{j=1}^{k}\left|\sqrt{2}g_j\right|^2}{\sum_{j=1}^{k}\left|\sqrt{2}g_j\right|^2 + \sum_{l=k+1}^{d}\left|\sqrt{2}g_l\right|^2}.$$

By definition, each $g_j = x_j + iy_j$ for independent $x_j, y_j \sim N(0, \frac{1}{2})$, so $\sqrt{2}x_j \sim N(0,1)$ and $\sqrt{2}y_j \sim N(0,1)$. Since $\left|\sqrt{2}g_j\right|^2 = \left|\sqrt{2}x_j\right|^2 + \left|\sqrt{2}y_j\right|^2$, it follows that

$$\frac{\sum_{j=1}^{k}\left|\sqrt{2}g_j\right|^2}{\sum_{j=1}^{k}\left|\sqrt{2}g_j\right|^2 + \sum_{l=k+1}^{d}\left|\sqrt{2}g_l\right|^2} = \frac{\sum_{j=1}^{k}\left(\left|\sqrt{2}x_j\right|^2 + \left|\sqrt{2}y_j\right|^2\right)}{\sum_{j=1}^{k}\left(\left|\sqrt{2}x_j\right|^2 + \left|\sqrt{2}y_j\right|^2\right) + \sum_{l=k+1}^{d}\left(\left|\sqrt{2}x_l\right|^2 + \left|\sqrt{2}y_l\right|^2\right)}.$$

Thus $X = \sum_{j=1}^{k}\left(\left|\sqrt{2}x_j\right|^2 + \left|\sqrt{2}y_j\right|^2\right)$ is chi-squared distributed with $2k$ degrees of freedom, i.e., $X \sim \chi^2(2k)$, and similarly $Y = \sum_{l=k+1}^{d}\left(\left|\sqrt{2}x_l\right|^2 + \left|\sqrt{2}y_l\right|^2\right) \sim \chi^2(2(d-k))$. The ratio $\frac{X}{X+Y}$ is then beta distributed with parameters $a = \frac{2k}{2} = k$ and $b = \frac{2(d-k)}{2} = d - k$. Altogether, this means that

$$\|Px\|_2^2 \overset{(\mathrm{d})}{=} \frac{\sum_{j=1}^{k}\left(\left|\sqrt{2}x_j\right|^2 + \left|\sqrt{2}y_j\right|^2\right)}{\sum_{j=1}^{k}\left(\left|\sqrt{2}x_j\right|^2 + \left|\sqrt{2}y_j\right|^2\right) + \sum_{l=k+1}^{d}\left(\left|\sqrt{2}x_l\right|^2 + \left|\sqrt{2}y_l\right|^2\right)} = \frac{X}{X+Y} \sim \mathrm{Beta}\left(k, d-k\right).$$

□

Random unitaries and orthogonal projections

The group of unitary matrices on $\mathbb{F}^d$, denoted $\mathcal{U}_{\mathbb{F}}(d)$, is a compact group and thus admits a Haar measure γ with $\gamma(\mathcal{U}_{\mathbb{F}}(d)) = 1$ [51, 65]. The Haar measure has the important property of being translation invariant, which means that for any measurable subset $E \subset \mathcal{U}_{\mathbb{F}}(d)$ and any element $U \in \mathcal{U}_{\mathbb{F}}(d)$ the translated sets $UE = \{UV : V \in E\}$ and $EU = \{VU : V \in E\}$ satisfy $\gamma(UE) = \gamma(E) = \gamma(EU)$. The Haar probability measure on $\mathcal{U}_{\mathbb{F}}(d)$ is also referred to the uniform distribution.

Definition 1.4.7. *A* ***uniformly distributed unitary matrix*** *$U \in \mathcal{U}_{\mathbb{F}}(d)$ is a unitary matrix with probability distribution given by the Haar probability measure on $\mathcal{U}_{\mathbb{F}}(d)$. In other words,*

$$\mathbb{P}\left\{U \in E\right\} = \gamma(E)$$

for all measurable subsets $E \subset \mathcal{U}_{\mathbb{F}}(d)$.

Since $\mathcal{U}_{\mathbb{F}}(d)$ acts on the space of rank-k orthogonal projection matrices via conjugation, γ induces a probability measure γ' on $\mathrm{Proj}_{\mathbb{F}}(k, d)$: a subset $E \subset \mathrm{Proj}_{\mathbb{F}}(k, d)$ is measurable with respect to γ' if and only $\{U \in \mathcal{U}_{\mathbb{F}}(d) : U^*VU \in E\}$ is measurable with respect to γ for any fixed $V \in \mathrm{Proj}_{\mathbb{F}}(k, d)$, and $\gamma'(E)$ is defined by $\gamma'(E) := \gamma\left(\{U \in \mathcal{U}_{\mathbb{F}}(d) : U^*VU \in E\}\right)$ for all measurable subsets E [8]. This measure γ' is sometimes referred to as the the **uniform probability measure** on $\mathrm{Proj}_{\mathbb{F}}(k, d)$.

Definition 1.4.8. *A* ***uniformly distributed*** *rank-k projection matrix $P \in \mathrm{Proj}_{\mathbb{F}}(k,d)$ is a projection picked from the uniform probability measure on $\mathrm{Proj}_{\mathbb{F}}(k,d)$. In other words,*

$$\mathbb{P}\{P \in E\} = \gamma'(E) := \gamma\left(\{U \in \mathcal{U}_{\mathbb{F}}(d) : U^*VU \in E\}\right)$$

for all measurable subsets $E \subset \mathrm{Proj}_{\mathbb{F}}(k,d)$.

The translation invariance of the Haar measure on the group of unitaries gives the induced measure on the space of orthogonal projections the property of being rotationally invariant: if P is uniformly distributed in $\mathrm{Proj}_{\mathbb{F}}(k,d)$ and $U \in \mathcal{U}_{\mathbb{F}}(d)$ is fixed, then $U^*PU \overset{(\mathrm{d})}{=} P$. Conjugation by the unitary U "rotates" a projection P, in the sense that U^*PU is the orthogonal projection onto $U^* \mathrm{Ran}(P)$.

There are many equivalent ways to generate a uniformly distributed rank-k projection on $\mathbb{F}^d$. For example, one can take k Gaussian random vectors in $\mathbb{F}^d$ and then form the projection onto their span. A second way is to take a fixed rank-k projection and conjugate it by a Haar distributed random unitary $U \in \mathcal{U}_{\mathbb{F}}(d)$. It can be helpful to think of a "uniformly distributed rank-k projection" as just a "projection onto a uniformly distributed k-dimensional subspace". There are a few more ways to generate uniformly distributed random projections and unit vectors worth mentioning, which are contained in the following lemma.

Lemma 1.4.9. *Let $x \in \mathbb{S}_{\mathbb{F}}^{d-1}$, $P \in \mathrm{Proj}_{\mathbb{F}}(k,d)$, and $U \in \mathcal{U}_{\mathbb{F}}(d)$ each be uniformly distributed as defined above. Let $y \in \mathbb{S}_{\mathbb{F}}^{d-1}$ and $Q \in \mathrm{Proj}_{\mathbb{F}}(k,d)$ be fixed. Then the following facts hold:*

(a) xx^* is uniformly distributed in $\mathrm{Proj}_{\mathbb{F}}(1,d)$

(b) Uy is uniformly distributed in $\mathbb{S}_{\mathbb{F}}^{d-1}$

(c) U^*QU is uniformly distributed in $\mathrm{Proj}_{\mathbb{F}}(k,d)$

(d) $I - P$ is uniformly distributed in $\mathrm{Proj}_{\mathbb{F}}(d-k,d)$.

Proof. All of these facts are proven by showing the given random vector or matrix satisfies the rotational invariance property that characterizes the respective uniform distribution.

For (a), observe first that xx^* defines a random rank-one orthogonal projection since it is Hermitian, has range $\operatorname{Ran}(xx^*) = \operatorname{span}\{x\}$, and is idempotent:

$$(xx^*)(xx^*) = x(x^*x)x^* = \langle x, x\rangle\, xx^* = \|x\|_2^2\, xx^* = xx^*.$$

If $W \in \mathcal{U}_{\mathbb{F}}(d)$ is an arbitrary unitary, then the rotational invariance of the uniform distribution on $\mathbb{S}_{\mathbb{F}}^{d-1}$ yields that

$$W^*xx^*W = (W^*x)(W^*x)^* \overset{(\mathrm{d})}{=} xx^*.$$

Thus the distribution of xx^* is rotationally invariant, hence xx^* is uniformly distributed in $\operatorname{Proj}_{\mathbb{F}}(1, d)$.

For (b), observe first that Uy defines a random unit vector since $\|Uy\|_2 = \|y\|_2 = 1$ because U is an isometry. If $W \in \mathcal{U}_{\mathbb{F}}(d)$ is an arbitrary unitary, then the translation invariance of the uniform distribution on $\mathcal{U}_{\mathbb{F}}(d)$ implies that $WUy \overset{(\mathrm{d})}{=} Uy$. Thus the distribution of Uy is rotationally invariant, hence Uy is uniformly distributed in $\mathbb{S}_{\mathbb{F}}^{d-1}$.

For (c), observe first that U^*QU defines a random rank-k orthogonal projection since it is Hermitian, has range $\operatorname{Ran}(U^*QU) = U^* \operatorname{Ran}(Q)$, and is idempotent:

$$(U^*QU)(U^*QU) = U^*Q(UU^*)QU = U^*Q^2U = U^*QU.$$

If $W \in \mathcal{U}_{\mathbb{F}}(d)$ is an arbitrary unitary, then the translation invariance of the uniform distribution on $\mathcal{U}_{\mathbb{F}}(d)$ yields that

$$W^*U^*QUW = (UW)^*Q(UW) \overset{(\mathrm{d})}{=} U^*QU.$$

Thus the distribution of U^*QU is rotationally invariant, hence U^*QU is uniformly distributed in $\operatorname{Proj}_{\mathbb{F}}(k, d)$.

Lastly, for (d), observe that $I - P$ defines a random rank-$(d-k)$ orthogonal projection since it is Hermitian, has range $\operatorname{Ran}(I - P) = \operatorname{Ran}(P)^{\perp}$, and is idempotent:

$$(I - P)(I - P) = I - 2P + P^2 = I - P.$$

If $W \in \mathcal{U}_{\mathbb{F}}(d)$ is an arbitrary unitary, then by the translation invariance of the uniform distribution on $\mathrm{Proj}_{\mathbb{F}}(k, d)$ it follows that

$$W^*(I - P)W = W^*IW - W^*PW = I - W^*PW \overset{\text{(d)}}{=} I - P.$$

Thus the distribution of $I - P$ is rotationally invariant, hence $I - P$ is uniformly distributed in $\mathrm{Proj}_{\mathbb{F}}(d - k, d)$. □

The rotational invariance of the uniform distribution on $\mathrm{Proj}_{\mathbb{F}}(k, d)$ is useful for determining its expectation. The following lemma determines this expectation, and serves as a useful example of how rotational invariance will be used in proofs.

Lemma 1.4.10. *Let $P \in \mathrm{Proj}_{\mathbb{F}}(k, d)$ be uniformly distributed. Then*

$$\mathbb{E}\left[P\right] = \frac{k}{d} I.$$

Proof. By the rotational invariance of the unitary distribution on $\mathrm{Proj}_{\mathbb{F}}(k, d)$, if $U \in \mathcal{U}_{\mathbb{F}}(d)$ is an arbitrary unitary then $U^*PU \overset{\text{(d)}}{=} P$. Using the linearity of the expected value, it follows that

$$\mathbb{E}\left[P\right] = \mathbb{E}\left[U^*PU\right] = U^*\mathbb{E}\left[P\right]U,$$

or in other words $\mathbb{E}\left[P\right]$ commutes with every unitary on $\mathbb{F}^d$. At this point, one could appeal to Schur's lemma to say that $\mathbb{E}\left[P\right]$ must be a multiple of the identity, $\mathbb{E}\left[P\right] = \lambda I$. Instead, an elementary proof of this fact for this particular case is provided below for the sake of completion.

If $d = 1$ then trivially $\mathbb{E}\left[P\right] = \lambda I$ for some $\lambda \in \mathbb{F}$, so suppose $d > 1$. Since P is a random positive Hermitian operator, $\mathbb{E}\left[P\right]$ is Hermitian and may be decomposed according to the spectral theorem as

$$\mathbb{E}\left[P\right] = \sum_{k=1}^{d} \lambda_k V_k \tag{11}$$

where $\lambda_1 \geq \ldots \geq \lambda_k$ and $V_k \in \mathrm{Proj}_{\mathbb{F}}(1, d)$. Let $\{v_1, \ldots, v_k\}$ be an orthonormal basis for $\mathbb{F}^d$ with

$v_k v_k^* = V_k$, and for each $j_1 \neq j_2$ let U_{j_1,j_2} be the unitary defined by

$$U_{j_1,j_2} v_k = \begin{cases} v_{j_2} & \text{if } k = j_1 \\ v_{j_1} & \text{if } k = j_2 \\ v_k & \text{else.} \end{cases}$$

From this definition, $U^*_{j_1,j_2} \mathbb{E}\left[P\right] U_{j_1,j_2}$ may be expressed as

$$\begin{aligned} U^*_{j_1,j_2} \mathbb{E}\left[P\right] U_{j_1,j_2} &= U^*_{j_1,j_2} \left(\sum_{k=1}^{d} \lambda_k V_k \right) U_{j_1,j_2} \\ &= \sum_{k=1}^{d} \lambda_k U^*_{j_1,j_2} V_k U_{j_1,j_2} \\ &= \lambda_{j_1} V_{j_2} + \lambda_{j_2} V_{j_1} + \sum_{k \neq j_1, j_2} \lambda_k V_k. \end{aligned}$$

Since $U^*_{j_1,j_2} \mathbb{E}\left[P\right] U_{j_1,j_2} = \mathbb{E}\left[P\right]$, it follows that

$$\lambda_{j_1} V_{j_2} + \lambda_{j_2} V_{j_1} = \lambda_{j_1} V_{j_1} + \lambda_{j_2} V_{j_2}. \tag{12}$$

Applying both sides of (12) to the vector v_{j_1} yields $\lambda_{j_1} = \lambda_{j_2}$. Since $j_1 \neq j_2$ were arbitrary indices, it follows that all eigenvalues of $\mathbb{E}\left[P\right]$ are the same, i.e., $\lambda_k = \lambda$ for some $\lambda \in \mathbb{R}$. Thus $\mathbb{E}\left[P\right] = \lambda I$.

To compute λ, the linearity of the trace and the expected value operations show that

$$\lambda d = \operatorname{tr}\left[\lambda I\right] = \operatorname{tr}\left[\mathbb{E}\left[P\right]\right] = \mathbb{E}\left[\operatorname{tr}\left[P\right]\right] = k,$$

and thus $\lambda = \frac{k}{d}$. □

Concentration of measure for random variables

Concentration of measure is a useful phenomenon that occurs for functions of high-dimensional random vectors. Essentially, concentration of measure results say that a function $f : \mathbb{F}^d \to \mathbb{F}$ of a random variable $X \in \mathbb{F}^d$, under suitable regularity assumptions of f and X, will be close to its expected value with high probability. The classic example of concentration of measure is the law of large numbers, which is usually stated as follows.

Theorem 1.4.11 (Weak law of large numbers [47])**.** *Let $(X_j)_{j\in\mathbb{N}}$ be an infinite sequence of i.i.d. Lebesgue integrable $\mathbb{F}$-valued random variables with $\mathbb{E}[X_j] = \mu \in \mathbb{F}$. For each $n \in \mathbb{N}$ define $\hat{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$, the empirical average of the first n random variables. Then for every $\epsilon > 0$*

$$\lim_{n\to\infty} \mathbb{P}\left\{\left|\hat{X}_n - \mu\right| \geq \epsilon\right\} \to 0.$$

Under the assumptions of Theorem 1.4.11, let $Z_n = (X_1, \ldots, X_n) \in \mathbb{F}^n$ and $f_n : \mathbb{F}^d \to \mathbb{F}$ be given by $f(x_1, \ldots, x_n) = \frac{1}{n}\sum_{j=1}^{n} x_j$. Then Z_n is a random vector in $\mathbb{F}^n$ and $\hat{X}_n = f_n(Z_n)$ may be thought of as a function of that random vector. Additionally,

$$\mathbb{E}\left[f_n(Z_n)\right] = \mathbb{E}\left[\frac{1}{n}\sum_{j=1}^{n} X_j\right] = \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\left[X_j\right] = \mu.$$

In this notation, the weak law of large numbers says that for all $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left\{\left|f_n(Z_n) - \mathbb{E}\left[f_n(Z_n)\right]\right| \geq \epsilon\right\} \to 0,$$

i.e., the function f_n of the random vector Z_n is close to its expected value with high probability in high dimensions n.

The weak law of large numbers gives an asymptotic statement about concentration of measure for a specific type of random vector and a specific function (the empirical average) as the dimension grows to infinity. There are other, non-asymptotic measure concentration results that provide numerical bounds on the probability in question and show how quickly the probability decays to zero. A standard example is Chebyshev's inequality, which describes the concentration of measure of random variables with respect to their variance.

Theorem 1.4.12 (Chebyshev's inequality)**.** *Let X be a random variable with $\mathbb{E}[X] = \mu < \infty$ and $0 < \operatorname{var}(X) = \sigma^2 < \infty$. Then for every $t \in \mathbb{R}$*

$$\mathbb{P}\{|X - \mu| \geq t\} \leq \frac{\sigma^2}{t^2}.$$

Chebyshev's inequality can be seen as a quantitative version of the weak law of large numbers with the added assumption of finite variance: using the notation and assumptions of Theorem 1.4.11 and assuming $\operatorname{var}(X_1) = \sigma^2 < \infty$, the empirical average $\hat{X}_n$ is a random variable with variance $\operatorname{var}(\hat{X}_n) = \frac{\sigma^2}{n}$. Thus by Chebyshev's inequality, for any $t > 0$

$$\mathbb{P}\left\{\left|\hat{X}_n - \mu\right| \geq t\right\} \leq \frac{\sigma^2}{nt^2}.$$

Since $\frac{\sigma^2}{nt^2} \to 0$ as $n \to \infty$, the weak law of large numbers follows as a consequence.

More powerful concentration inequalities provide bounds on the probability of deviating substantially from the mean that are sub-exponential in the number of terms in the empirical average. Of particular importance for the proof strategy in Chapter 2 are the Chernoff inequality and the Bernstein inequality. These concentration inequalities are key tools in proving concentration results for one-bit phase retrieval using random projections for the phaseless measurement.

Theorem 1.4.13 (Chernoff inequality for binomial random variables [4, Corollary A.1.7])**.** *Let $X \sim \operatorname{Binom}(m, p)$. Then for any $t > 0$*

$$\mathbb{P}\{|X - mp| \geq t\} \leq 2\exp\left(-\frac{2t^2}{m}\right).$$

Observe that if $X \sim \operatorname{Binom}(m, p)$, then $\mathbb{E}[X] = mp$, so Theorem 1.4.13 gives sub-exponential concentration of binomial random variables around their expected values. If the random variable $\frac{1}{m}X$ is considered instead, which is the empirical average of a sum of i.i.d. Bernoulli random variables, then Theorem 1.4.13 says that

$$\mathbb{P}\left\{\left|\frac{1}{m}X - p\right| \geq t\right\} \leq 2\exp\left(-2mt^2\right).$$

In other words, as $m \to \infty$ the normalized binomial random variable $\frac{1}{m}X$ concentrates sub-exponentially around its expected value p.

Whereas binomial random variables are sums of independent Bernoulli random variables, the Bernstein inequality can be used to give a more general sub-exponential concentration bound for sums of bounded random variables.

Theorem 1.4.14 (Bernstein inequality [89, Theorem 1.6.1])**.** *Let $X_1, \ldots, X_m$ be a finite sequence of independent random variables with $\mathbb{E}[X_j] = 0$ and $|X_j| \leq L$ for all j. Let $Z = \sum_{j=1}^{m} X_j$. Then for any $t > 0$*

$$\mathbb{P}\{|Z| \geq t\} \leq 2\exp\left(-\frac{t^2}{2\operatorname{var}(Z) + \frac{2Lt}{3}}\right).$$

The Bernstein inequality may be extended to cover sums of independent random matrices, rather than just one-dimensional random variables as in Theorem 1.4.14.

Theorem 1.4.15 (Matrix Bernstein inequality [89, Theorem 6.6.1])**.** *Let $X_1, \ldots, X_m$ be a finite sequence of independent Hermitian random matrices in $\mathbb{F}^{d\times d}$ with $\mathbb{E}[X_j] = 0$ and $\|X_j\| \leq L$ for all j. Let $Z = \sum_{j=1}^{m} X_j$, and let $\operatorname{var}(Z)$ be the matrix variance of the sum defined by*

$$\operatorname{var}(Z) := \left\|\mathbb{E}\left[Z^2\right]\right\| = \left\|\sum_{j=1}^{m} \mathbb{E}\left[X_j^2\right]\right\|.$$

Then the expected value of the norm of Z is bounded by

$$\mathbb{E}\left[\|Z\|\right] \leq \sqrt{2v(Z)\log(d)} + \frac{1}{3}L\log(d),$$

and for any $t > 0$

$$\mathbb{P}\{\|Z\| \geq t\} \leq 2d\exp\left(-\frac{t^2}{2\operatorname{var}(Z) + \frac{2Lt}{3}}\right).$$

Theorem 1.4.14, the Bernstein inequality for one-dimensional random variables, can be seen as a special case of the matrix Bernstein inequality: real-valued random variables are just random 1×1 Hermitian matrices, the operator norm of a 1×1 matrix is just the absolute value of its single entry, and the matrix variance of the sum $v(Z)$ is just the regular notion of variance of a random variable.

The matrix Bernstein inequality will mainly be applied to study the concentration of an empirical average of i.i.d. random Hermitian matrices. The following corollary says that such an empirical average experiences concentration of measure that is sub-exponential in the number of i.i.d. copies in the empirical average.

Corollary 1.4.16. *Let $X_1, \ldots, X_m$ be a finite sequence of i.i.d. Hermitian random matrices in $\mathbb{F}^{d\times d}$ with $\mathbb{E}\left[X_j\right] = 0$ and $\|X_j\| \le L$ for all j. Let $\hat{Z} = \frac{1}{m}\sum_{j=1}^m X_j$. Then*

$$\mathbb{E}\left[\left\|\hat{Z}\right\|\right] \le \sqrt{\frac{2\left\|\mathbb{E}\left[X_1^2\right]\right\|\log(d)}{m}} + \frac{L\log(d)}{3m}$$

and for any $t > 0$

$$\mathbb{P}\left\{\left\|\hat{Z}\right\| \ge t\right\} \le 2d\exp\left(-\frac{mt^2}{2\left\|\mathbb{E}\left[X_1^2\right]\right\| + \frac{2Lt}{3}}\right).$$

Proof. Let $Y_j = \frac{1}{m}X_j$. Then $\mathbb{E}\left[Y_j\right] = 0$ and $\|Y_j\| \le \frac{L}{m}$ for all j, and $\hat{Z} = \sum_{j=1}^m X_j$. Also, using the fact the X_j are i.i.d. it follows that

$$v(\hat{Z}) = \left\|\sum_{j=1}^m \mathbb{E}\left[Y_j^2\right]\right\| = m\left\|\mathbb{E}\left[Y_1^2\right]\right\| = \frac{1}{m}\mathbb{E}\left[\|X_1\|^2\right].$$

The results follow by applying Theorem 1.4.15 for $\hat{Z}$. □

The matrix Bernstein inequality may be generalized even further to non-Hermitian and non-square matrices, see [89, Theorem 6.1.1], but the Hermitian case will be sufficient to prove the measure concentration results needed for one-bit phase retrieval.

1.5 Rate-distortion theory and an additional noise model

The one-bit phase retrieval problems of Section 1.3 and other one-bit signal recovery problems such as one-bit compressed sensing can be interpreted in an information theoretic context, making use of concepts from rate-distortion theory and source-channel coding. Rate-distortion theory was first introduced by Shannon to study how well signals drawn from particular distributions could be compressed by measuring the number of bits necessary to achieve a given level of distortion in the recovery procedure [85, 86]. These concepts are briefly discussed in this section, and used to motivate an additional noise model that will be consider in Chapter 3.

In the language of rate-distortion theory, a phaseless binary measurement $\Phi_{\mathcal{P}}$ defined as in Section 1.3 *encodes* an input signal $X \in \mathrm{Proj}_{\mathbb{F}}(1, d)$ and the reconstruction algorithm $\mathcal{R}$ *decodes* the binary string $\Phi_{\mathcal{P}}(X)$ to yield an approximation for the input signal. If $\mathcal{P} = \{P_j\}_{j=1}^m$ is a collection of orthogonal projections on $\mathbb{F}^d$, then the *distortion* of the measurement and recovery scheme $\mathcal{R} \circ \Phi_{\mathcal{P}}$ is the mean squared error

$$D(\Phi_{\mathcal{P}}, \mathcal{R}) := \mathbb{E}_X \left[\left\| \hat{X} - X \right\|^2 \right],$$

where $X \in \mathrm{Proj}_{\mathbb{F}}(1, d)$ is uniformly distributed and $\hat{X} = \mathcal{R} \circ \Phi_{\mathcal{P}}(X)$ is the approximate recovery of X. The mean squared error is a common measure of distortion of an encoding and decoding scheme, and is often used for video or image compression [16, 75].

More generally, the binary string $\Phi_{\mathcal{P}}$ is passed through a noisy channel before the reconstruction algorithm is applied. As discussed in Section 1.3, the noise is implemented via a bit-flip map $\mathcal{F}_T : \{0, 1\}^m \to \{0, 1\}^m$ associated to a subset $T \subset \{1, \ldots, m\}$ that flips the components of a binary string corresponding to the indices in T. The recovered signal based on the phaseless binary measurement that has been flipped on index set T is denoted $\hat{X}_T = \mathcal{R}(\mathcal{F}_T(\Phi_{\mathcal{P}}(X)))$. Unlike the adversarial noise model described in Problems 1.3.8 and 1.2.7, in rate-distortion theory the noisy channel, thought of as the environment the encoded signal is transferred through, is typically assumed to act randomly and not in an adversarial way. This is modeled by selecting the bit-flip index set T randomly in such a way that each index $j \in \{1, \ldots, m\}$ is included in T independently with some fixed probability τ. This distribution on subsets of $\{1, \ldots, m\}$ is referred to as the **binomial distribution** with probability τ. The distortion of a measurement and recovery scheme using a fixed collection of orthogonal projections $\mathcal{P}$ is then computed by also averaging over these binomially distributed bit-flip sets to define

$$D(\Phi_{\mathcal{P}}, \mathcal{R}, \tau) := \mathbb{E}_{X,T}\left[\left\|\hat{X}_T - X\right\|^2\right].$$

In order to gauge the effectiveness of a binary quantization map Φ and reconstruction algorithm $\mathcal{R}$ under the presence of random bit-flips with probability τ, it is useful to study the average distortion as the projections $\mathcal{P} = \{P_j\}_{j=1}^m$ are chosen independently and uniformly at random. This approach avoids questions about the optimality of a particular collection of projections and instead focuses on the performance of the measurement and reconstruction scheme for random projections. The task of devising a measurement and reconstruction scheme for one-bit phase retrieval that achieves an average mean squared error distortion is stated formally below.

Problem 1.5.1 (One-bit phase retrieval average distortion, random i.i.d. bit-flips)**.** *Let $\delta > 0$ be a desired distortion level. Choose m, a binary quantization map Φ, and a recovery algorithm $\mathcal{R}$ such that*

$$\mathbb{E}_{\mathcal{P}}\left[D(\Phi_{\mathcal{P}}, \mathcal{R}, \tau)\right] = \mathbb{E}_{X,T,\mathcal{P}}\left[\left\|\hat{X}_T - X\right\|^2\right] < \delta,$$

where $X \in \mathrm{Proj}_{\mathbb{F}}(1, d)$ is uniformly distributed, $T \subset \{1, \ldots, m\}$ is binomially distributed with probability τ, $\mathcal{P} = \{P_j\}_{j=1}^m$ is a sequence of independent and uniformly distributed orthogonal projections on $\mathbb{F}^d$, and $\hat{X}_T = \mathcal{R}(\mathcal{F}_T(\Phi_{\mathcal{P}}(X)))$.

The measurement and reconstruction scheme for noiseless one-bit phase retrieval in Chapter 2 is shown in Section 3.3 to solve 1.5.1. The number of *bits-per-dimension*, $\frac{m}{d}$, sufficient to achieve a mean squared error of δ is used to judge the effectiveness of this scheme. In the language of rate-distortion theory: the number of bits-per-dimension is the *rate* and the mean squared error is the *distortion*, and reconstruction algorithms are judged by the rate $\frac{m}{d}$ required to achieve a given distortion δ.

Chapter 2

Uniformly Accurate One-Bit Phase Retrieval

This chapter addresses the task of performing approximate phase retrieval from phaseless binary measurements that reveal qualitative information about the measured signal. For this chapter, it is assumed that the phaseless binary measurement is always obtained without any bit-flip errors. See Chapter 3 for the extension to the noisy measurement case with a few different noise models. Additionally, see Section 1.2 for an overview of phase retrieval, Section 1.3 for the basic definitions and problems of one-bit phase retrieval, and Section 1.4 for some definitions and tools from probability theory that will be referenced in proofs throughout this chapter.

Section 2.1 discusses a few types of binary questions that can be used to form a phaseless binary measurement. In particular, *magnitude comparison measurements* are defined, a type of phaseless binary measurement where each bit is determined by comparing two magnitudes $\mathrm{tr}\,[P_1 X]$ and $\mathrm{tr}\,[P_2 X]$ and recording which is larger. Two specific ways to generate random orthogonal projections for a magnitude comparison measurement are identified: independent pairs, and complementary pairs.

Section 2.2 presents an algorithm called *Principal Eigenspace Programming (PEP)* that performs uniformly accurate one-bit phase retrieval, solving Problem 1.3.7. PEP may be implemented by finding the maximizer of a computationally tractable semidefinite program. This algorithm is based on selecting random projections $\mathcal{P}$ for the phaseless binary measurement and assembling an auxiliary matrix $\hat{Q}_{\mathcal{P}}(X)$ based on the quantized magnitudes that in expectation has the input signal X as its principal eigenprojection.

Section 2.3 computes the expectation of the auxiliary matrix $\hat{Q}_{\mathcal{P}}(X)$ used in PEP. The spectral decomposition of this matrix is directly related to the accuracy of PEP, and it varies depending on the phaseless binary measurement used to construct $\hat{Q}_{\mathcal{P}}(X)$. The spectral decomposition is computed for magnitude comparison measurements associated to independent or complementary pairs of projections.

Section 2.4 shows the following pointwise result: for any fixed $X \in \mathrm{Proj}_{\mathbb{F}}(1, 2n)$ and any $\delta > 0$, a random magnitude comparison measurement $\Phi_{\mathcal{P}}$ associated to either independent or complementary pairs of projections gives $\Phi_{\mathcal{P}}(X)$ for which PEP yields a solution $\hat{X}$ that satisfies $\left\|\hat{X} - X\right\| < \delta$ with high probability. See Theorem 2.4.4 for details. This result solves the one-bit phase retrieval problem for fixed input signals, Problem 1.3.6. Measure concentration results described in Section 1.4 are used to show that PEP works with high probability with respect to the choice of random projections.

Much of the effort in Section 2.5 is directed toward getting *uniform* results from the *pointwise* one given by Theorem 2.4.4. This section specifies to complementary magnitude comparison measurements associated to random half-dimensioned projections. The uniform result derived says that for any $\delta > 0$, m may be selected large enough so that a collection of independent uniformly distributed half-dimensioned projections $\mathcal{P} = \{P_j\}_{j=1}^{m}$ will, with high probability, yield complementary magnitude comparison measurements $\Phi_{\mathcal{P}}(X)$ for every $X \in \mathrm{Proj}_{\mathbb{F}}(1, 2n)$ for which the solution $\hat{X}$ to PEP satisfies $\left\|\hat{X} - X\right\| < \delta$. See Theorem 2.5.14 for details. This result solves the uniformly accurate one-bit phase retrieval problem, Problem 1.3.7. According to the uniform result, a large enough random collection of projections has the property that *every* signal is approximately recoverable up to an error of δ from the complementary magnitude comparisons using those projections.

Most work in this chapter may be found in my paper with Bernhard Bodmann [38], which has been submitted for publication. Specifically, [38] dealt with complementary magnitude comparison measurements for half-dimensioned projections as defined in Definition 2.1.6, and derived pointwise and uniform guarantees for this measurement model.

2.1 Phaseless binary measurement models

A solution to either of the noiseless one-bit phase retrieval problems, Problem 1.3.6 for recovery of a fixed input or Problem 1.3.7 for uniformly accurate recovery of all inputs, must first decide upon a phaseless binary measurement $\Phi_{\mathcal{P}}$ as in Definition 1.2.2. This section examines a few possible ways to take a random phaseless binary measurement where each bit of the output has the same information capacity. As mentioned in Remark 1.2.1 and Remark 1.3.3, the equivalence class of a normalized input signal $x \in \mathbb{S}_{\mathbb{F}}^{d-1}$ is identified with the rank-one orthogonal projection onto its span, denoted $xx^* \in \mathrm{Proj}_{\mathbb{F}}(1,d)$. This identification is used to provide geometric intuition for the qualitative information recorded in each individual one-bit measurement.

Types of binary questions

The phaseless binary measurements considered in this dissertation depend upon a random selection of projections $\mathcal{P}$ in such a way that for every $X \in \mathrm{Proj}_{\mathbb{F}}(1,d)$ the entries $(\Phi_{\mathcal{P}}(X)_j)_{j=1}^m$ are independent and identically distributed. In other words, each bit of the binary string $\Phi_{\mathcal{P}}(X)$ has the same information capacity. One way to accomplish this is to assume that the projections $\mathcal{P} = \{P_j\}_{j=1}^m$ are identically distributed, and that each entry of the phaseless binary measurement has the form $\Phi_{\mathcal{P}}(X)_j = \Phi'(\mathcal{M}_{\mathcal{P}_j}(X))$ where $\mathcal{P}_j \subset \mathcal{P}$ with $|\mathcal{P}_j| = s$ for all j, $\mathcal{P}_i \cap \mathcal{P}_j = \emptyset$ for all $i \neq j$, $\mathcal{P}_i$ and $\mathcal{P}_j$ are independent for every $i \neq j$, and $\Phi' : [0,1]^s \to \{0,1\}$. A map like $\Phi' \circ \mathcal{M}_{\mathcal{P}_j}$ is called a *binary question*, because it takes the s magnitude measurements $\mathcal{M}_{\mathcal{P}_j}(X)$ and outputs a binary value. To define a phaseless binary measurement of this form requires defining the binary questions it is comprised of.

One natural binary question compares a pair of traditional phase retrieval magnitude measurements to each other and records which one is larger.

Definition 2.1.1. *Let $P_1 \in \mathrm{Proj}_{\mathbb{F}}(k_1,d)$ and $P_2 \in \mathrm{Proj}_{\mathbb{F}}(k_2,d)$ be orthogonal projections. The* ***magnitude comparison*** *associated to P_1 and P_2 is the map $\phi_{P_1,P_2} : \mathrm{Proj}_{\mathbb{F}}(1,d) \to \{0,1\}$ given* *by*

$$\phi_{P_1,P_2}(X) = \begin{cases} 1 & \textit{if } \operatorname{tr}[P_1X] \geq \operatorname{tr}[P_2X] \\ 0 & \textit{else.} \end{cases}$$

*If $\phi_{P_1,P_2}(X) = 1$, then P_1 is called the projection in $\{P_1, P_2\}$ that is **proximal** to X.*

Recall that if $P \in \mathrm{Proj}_{\mathbb{F}}(k,d)$ and $X \in \mathrm{Proj}_{\mathbb{F}}(1,d)$, then $\operatorname{tr}[PX] = \cos^2(\theta)$ where θ is the principal angle between the subspaces $\mathrm{Ran}(P)$ and $\mathrm{Ran}(X)$. The principal angle is a way to measure the distance between subspaces, see Definition 1.3.4 and the surrounding discussion for details. Thus, a magnitude comparison ϕ_{P_1,P_2} provides a simple way to record qualitative information about the proximity of $\mathrm{Ran}(X)$ to a pair of subspaces $\mathrm{Ran}(P_1)$ and $\mathrm{Ran}(P_2)$. Specifically, $\phi_{P_1,P_2}(X)$ asks "Is $\mathrm{Ran}(X)$ closer to $\mathrm{Ran}(P_1)$ than to $\mathrm{Ran}(P_2)$?", and records the answer "Yes" as a 1 and "No" as a 0. Magnitude comparisons associated to pairs of rank-one projections were used in [73] to prove error bounds for one-bit phase retrieval of fixed input signals.

As a special case, one can consider magnitude comparisons between complementary projections, i.e., projections P_1, P_2 with $P_1 + P_2 = I$.

Definition 2.1.2. *Let $P \in \mathrm{Proj}_{\mathbb{F}}(k,d)$ be an orthogonal projections. The **complementary magnitude comparison** associated to P is the binary question $\phi_P := \phi_{P,I-P}$.*

A complementary magnitude comparison ϕ_P asks the question "Is $\mathrm{Ran}(X)$ closer to $\mathrm{Ran}(P)$ than to its orthogonal complement $\mathrm{Ran}(I-P)$?", and records the answer "Yes" as a 1 and "No" as a 0. Since $\operatorname{tr}[PX]+\operatorname{tr}[(I-P)X] = \operatorname{tr}[X] = 1$ for all $X \in \mathrm{Proj}_{\mathbb{F}}(1,d)$ and projections $P \in \mathrm{Proj}_{\mathbb{F}}(k,d)$ of any rank, $\operatorname{tr}[PX] \geq \operatorname{tr}[(I-P)X]$ if and only if $\operatorname{tr}[PX] \geq \frac{1}{2}$. In other words, the complementary magnitude comparison quantizes the magnitude measurement $\operatorname{tr}[PX]$ by applying a threshold at $\frac{1}{2}$, i.e.,

$$\phi_P(X) := \begin{cases} 1 & \text{if } \operatorname{tr}[PX] \geq \frac{1}{2} \\ 0 & \text{else.} \end{cases} \tag{13}$$

Other threshold values may also be used to quantize magnitude measurements.

Definition 2.1.3. *Let $P \in \mathrm{Proj}_{\mathbb{F}}(k,d)$ be an orthogonal projections. The* ***γ-thresholding question*** *associated to P is the binary question $\varphi_{P,\gamma} : \mathrm{Proj}_{\mathbb{F}}(1,d) \to \{0,1\}$ given by*

$$\varphi_{P,\gamma}(X) := \begin{cases} 1 & \textit{if } \operatorname{tr}[PX] \geq \gamma \\ 0 & \textit{else.} \end{cases}$$

From the above discussion, it follows that $\phi_P = \varphi_{P,\frac{1}{2}}$. Another natural threshold value for γ is the average value of the magnitude measurement $\operatorname{tr}[PX]$ for a uniformly distributed input $X \in \mathrm{Proj}_{\mathbb{F}}(1,d)$.

Definition 2.1.4. *Let $P \in \mathrm{Proj}_{\mathbb{F}}(k,d)$ be an orthogonal projection. The* ***average value thresholding question*** *associated to P is the binary question $\varphi_P := \varphi_{P,\frac{k}{d}}$.*

Observe that if $X \in \mathrm{Proj}_{\mathbb{F}}(1,d)$ is a uniformly distributed rank-one orthogonal projection and $P \in \mathrm{Proj}_{\mathbb{F}}(k,d)$ is fixed, then the average value of $\operatorname{tr}[PX]$ is precisely $\frac{k}{d}$. To see this, first observe that by linearity of the expected value $\mathbb{E}[\operatorname{tr}[PX]] = \operatorname{tr}[P\mathbb{E}[X]]$. By Lemma 1.4.10 it follows that $\mathbb{E}[X] = \frac{1}{d}I$, thus

$$\operatorname{tr}[P\mathbb{E}[X]] = \frac{1}{d}\operatorname{tr}[P] = \frac{k}{d},$$

as claimed. Thus φ_P quantizes the magnitude measurement $\operatorname{tr}[PX]$ by comparing it to the threshold $\mathbb{E}[\operatorname{tr}[PX]]$, as its name "average value thresholding question" suggests. Put another way, φ_P asks the question "Is $\operatorname{tr}[PX]$ larger than its average if X were uniformly distributed?", and the answer "Yes" is encoded as a 1 and "No" is encoded as a 0.

If $X \in \mathrm{Proj}_{\mathbb{F}}(1,d)$ and $P \in \mathrm{Proj}_{\mathbb{F}}(k,d)$ are both uniformly distributed, and $Y \in \mathrm{Proj}_{\mathbb{F}}(1,d)$ and $Q \in \mathrm{Proj}_{\mathbb{F}}(k,d)$ are fixed, then $\operatorname{tr}[PY] \overset{(\mathrm{d})}{=} \operatorname{tr}[QX]$. This holds by using the rotational invariance of the uniform distributions for orthogonal projections and the cyclic property of the trace: if $U \in \mathcal{U}_{\mathbb{F}}(d)$ is a Haar distributed random unitary, then $P \overset{(\mathrm{d})}{=} U^*QU$ and $X \overset{(\mathrm{d})}{=} UYU^*$ by Lemma 1.4.9, and so

$$\operatorname{tr}[PY] \overset{(\mathrm{d})}{=} \operatorname{tr}[U^*QUY] = \operatorname{tr}[QUYU^*] \overset{(\mathrm{d})}{=} \operatorname{tr}[QX].$$

Hence, φ_P may also be interpreted as asking the question "Is $\mathrm{tr}\left[PX\right]$ larger than its average if P were uniformly distributed?" and recording the answer in a binary value.

This dissertation will not investigate thresholding quantization in full generality, only in the specific case of $\gamma = \frac{1}{2}$ when it coincides with complementary magnitude comparisons as mentioned above. In particular, if $P \in \mathrm{Proj}_{\mathbb{F}}(n, 2n)$ then the complementary magnitude comparison ϕ_P, the $\frac{1}{2}$-thresholding question $\varphi_{P,\frac{1}{2}}$, and the average value thresholding question φ_P all coincide.

Phaseless binary measurements via magnitude comparison

The goal of this chapter is to show that accurate phase retrieval may be achieved with the qualitative proximity information gained from a sufficiently large set of binary questions. The results derived are for phaseless binary measurements for that are formed by taking magnitude comparisons as defined in Definition 2.1.1 for many different pairs of projections.

Definition 2.1.5. *If $\mathcal{P} = \{P_j\}_{j=1}^{2m}$ is a collection of orthogonal projections on $\mathbb{F}^d$, then the* ***magnitude comparison measurement*** *associated to $\mathcal{P}$ is the phaseless binary measurement $\Phi_{\mathcal{P}} : \mathrm{Proj}_{\mathbb{F}}(1, d) \to \{0,1\}^m$ defined by*

$$\Phi_{\mathcal{P}}(X)_i = \phi_{P_i, P_{m+i}}(X)$$

for each $i = 1, \ldots, m$. In other words,

$$\Phi_{\mathcal{P}}(X)_i = \begin{cases} 1 & \textit{if } \mathrm{tr}\left[P_i X\right] \geq \mathrm{tr}\left[P_{m+i} X\right] \\ 0 & \textit{else.} \end{cases}$$

Magnitude comparison measurements are a particular choice of $\Phi_{\mathcal{P}}$ that fit the general definition of a phaseless binary measurement as given in Definition 1.2.2. As a special case, one can consider magnitude comparison measurements associated to complementary pairs of projections.

Definition 2.1.6. *If $\mathcal{P} = \{P_j\}_{j=1}^m$ is a collection of orthogonal projections on $\mathbb{F}^d$, the* ***complementary magnitude comparison measurement*** *associated to $\mathcal{P}$ is the phaseless binary measurement $\Phi_{\mathcal{P}} : \mathrm{Proj}_{\mathbb{F}}(1, d) \to \{0, 1\}^m$ defined by*

$$\Phi_{\mathcal{P}}(X)_i = \phi_{P_i, I-P_i}(X)$$

for each $i = 1, \ldots m$. In other words,

$$\Phi_{\mathcal{P}}(X)_i = \begin{cases} 1 & \textit{if } \operatorname{tr}[P_i X] \geq \operatorname{tr}[(I-P)X] \\ 0 & \textit{else} \end{cases} = \begin{cases} 1 & \textit{if } \operatorname{tr}[P_i X] \geq \frac{1}{2} \\ 0 & \textit{else.} \end{cases}$$

This corresponds to the magnitude comparison measurement associated to the collection of projections

$$\mathcal{P} \cup (I - \mathcal{P}) = \{P_1, \ldots, P_m, I - P_1, \ldots, I - P_m\}.$$

Measurement by random projections

Due to the absence of an intuitive way to construct "optimal" collections of orthogonal projections for magnitude comparison measurements, the projections will be chosen uniformly at random and results will be stated with respect to this probability distribution. The uniform probability measure on $\mathrm{Proj}_{\mathbb{F}}(k, d)$ is induced by the Haar measure of the unitary group $\mathcal{U}_{\mathbb{F}}(d)$, and is characterized by the property of being rotationally invariant. In other words, if P is uniformly distributed in $\mathrm{Proj}_{\mathbb{F}}(k, d)$ then for any $U \in \mathcal{U}_{\mathbb{F}}(d)$ the random projection UPU^* satisfies $UPU^* \overset{\text{(d)}}{=} P$. See Section 1.4 for more details, such as how to generate a uniformly distributed orthogonal projection matrix.

Two methods of selecting uniformly distributed orthogonal projections for a magnitude comparison measurement will be considered for the pointwise results derived in Section 2.4. For the first, all projections are assumed to have the same rank and be independent, i.e., $\mathcal{P} = \{P_j\}_{j=1}^{2m}$ is an independent sequence of uniformly distributed projections in $\mathrm{Proj}_{\mathbb{F}}(k, d)$. The second model chooses m independent projections of a fixed rank for a complementary magnitude comparison measurement. By (13), this second model is equivalent to thresholding the magnitudes at $\frac{1}{2}$. Both of these methods result in a phaseless binary measurement $\Phi_{\mathcal{P}}$ where the entries of $\Phi_{\mathcal{P}}(X)$ are i.i.d. for each $X \in \mathrm{Proj}_{\mathbb{F}}(1, d)$, i.e., each entry carries the same information capacity. In Section 2.5, only complementary magnitude comparison measurements associated to uniformly distributed half-dimensioned projections will be considered.

2.2 Estimation algorithm - Principal Eigenspace Programming (PEP)

The overarching goal of this chapter is to show that the outcome of a phaseless binary measurement can be used to accurately and uniformly estimate all input signals via a computationally tractable reconstruction algorithm. This section defines the particular reconstruction algorithm $\mathcal{R}$ that will be used for estimation of signals from a magnitude comparison measurement as defined in Definition 2.1.5.

Suppose an unknown input signal $X \in \mathrm{Proj}_{\mathbb{F}}(1, d)$ is measured with a magnitude comparison measurement $\Phi_{\mathcal{P}}$ associated with a collection of random orthogonal projections $\mathcal{P}$ to obtain the binary vector $\Phi_{\mathcal{P}}(X)$. The information gained from these measurements will not in general completely determine the rank-one projection X corresponding to the input signal: in fact, the sets $\Phi_{\mathcal{P}}^{-1}(\{y\})$ for $y \in \{0, 1\}^m$ partition $\mathrm{Proj}_{\mathbb{F}}(1, d)$, and many of these preimages can consist of infinitely many signals that all yield the same phaseless binary measurement. Still, with enough binary questions the measurement $\Phi_{\mathcal{P}}(X)$ may be used to construct a projection $\hat{X}$ which approximates X. A consistent reconstruction would seek an element $\hat{X}$ in the feasible set, that is, the set of all Y consistent with the binary measurement in the sense that $\Phi_{\mathcal{P}}(Y) = \Phi_{\mathcal{P}}(X)$ [22]. A natural error bound for such a reconstruction strategy would then result from the diameter of the feasible set, which intuitively will be small if $\mathcal{P}$ is suitably large and uniformly distributed.

The reconstruction method considered in this dissertation relaxes the perfect consistency condition, but still achieves accurate recovery with a computationally feasible semidefinite programming algorithm. The approximate reconstruction of X is conveniently described in terms of auxiliary projections obtained from the binary measurement $\Phi_{\mathcal{P}}(X)$.

Definition 2.2.1. *Given an input signal $X \in \mathrm{Proj}_{\mathbb{F}}(1,d)$ and a magnitude comparison measurement $\Phi_{\mathcal{P}}$ associated to orthogonal projections $\mathcal{P} = \{P_j\}_{j=1}^{2m}$, the* ***proximal projections*** *are defined to be $\hat{P}_j(X) := \Phi_{\mathcal{P}}(X)_j P_j + (1 - \Phi_{\mathcal{P}}(X)_j) P_{m+j}$. In other words,*

$$\hat{P}_j(X) := \begin{cases} P_j & \textit{if } \operatorname{tr}[P_j X] \geq \operatorname{tr}[P_{m+j} X] \\ P_{m+j} & \textit{else.} \end{cases}$$

The ***empirical average of the proximal projections*** *is defined to be*

$$\hat{Q}_{\mathcal{P}}(X) := \frac{1}{m} \sum_{j=1}^{m} \hat{P}_j(X).$$

In other words, $\hat{P}_j(X)$ picks out of the two options P_j and P_{m+j} the projection that is proximal to X in the sense of Definition 2.1.1. If $\Phi_{\mathcal{P}}$ is a complementary magnitude comparison measurement associated to orthogonal projections $\mathcal{P} = \{P_j\}_{j=1}^{m}$, then by definition $P_{m+j} = I - P_j$, and so

$$\hat{P}_j(X) = \begin{cases} P_j & \text{if } \operatorname{tr}[P_j X] \geq \frac{1}{2} \\ I - P_j & \text{else.} \end{cases}$$

Given a sequence $\mathcal{P} = \{P_j\}_{j=1}^{m}$, the binary vector $\Phi_{\mathcal{P}}(X)$ is encoded into the sequence of proximal projections $\{\hat{P}_j(X)\}_{j=1}^{m}$, and $\hat{Q}_{\mathcal{P}}(X)$ is the empirical average of this auxiliary sequence.

The recovery algorithm takes the binary measurement $\Phi_{\mathcal{P}}(X)$ and outputs $\hat{X} := \mathcal{R}(\Phi_{\mathcal{P}}(X))$ which is the rank-one orthogonal projection onto the eigenspace of $\hat{Q}_{\mathcal{P}}(X)$ corresponding to the largest eigenvalue.

Definition 2.2.2. *Given an input signal $X \in \mathrm{Proj}_{\mathbb{F}}(1,d)$ and a magnitude comparison measurement $\Phi_{\mathcal{P}}$ associated to orthogonal projections $\mathcal{P} = \{P_j\}_{j=1}^{2m}$,* ***Principal Eigenspace Programming (PEP)*** *is the reconstruction algorithm $\mathcal{R}$ defined by $\mathcal{R}(\Phi_{\mathcal{P}}(X)) = \hat{X}$ where $\hat{X}$ is a rank-one orthogonal projection onto the principal eigenspace of $\hat{Q}_{\mathcal{P}}(X)$.*

This choice of $\hat{X}$ is the solution to the semidefinite program

$$\begin{aligned} &\underset{Y}{\text{maximize}} && \operatorname{tr}\left[\hat{Q}_{\mathcal{P}}(X)Y\right] \\ &\text{subject to} && Y \succeq 0, \operatorname{tr}[Y] \le 1. \end{aligned} \tag{14}$$

Finding the maximizer to (14) is referred to as Principal Eigenspace Programming (PEP) because it amounts to maximizing the Rayleigh quotient [63, Section 4.2] for $\hat{Q}_{\mathcal{P}}(X)$, which finds its principal eigenspace. This special class of semidefinite programs can be implemented efficiently [76, Chapter 4]. To be more specific, $\hat{X}$ is given by finding a unit eigenvector corresponding to the largest eigenvalue $\hat{x} = \arg\max_{u \in \mathbb{S}_{\mathbb{F}}^{d-1}} \left\langle \hat{Q}_{\mathcal{P}}(X)u, u \right\rangle$ and then setting $\hat{X} = \hat{x}\hat{x}^*$.

Since $\hat{Q}_{\mathcal{P}}(X)$ is a positive self-adjoint operator, it may be decomposed according to the spectral theorem as a linear combination of mutually orthogonal rank-one projections $\hat{Q}_{\mathcal{P}}(X) = \sum_{i=1}^{2n} \lambda_i E_i$, where $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_{2n} \ge 0$. Thus, any positive self-adjoint trace normalized operator with range contained in the principal eigenspace of $\hat{Q}_{\mathcal{P}}(X)$ is a solution to (PEP). If in addition λ_1 is strictly larger than λ_2 (which happens with probability 1 for our random measurement models), then its principal eigenspace is one-dimensional, and so $\hat{X} = E_1$ is the unique solution to (PEP). Corollary 2.3.7 will show that $\mathbb{E}\left[\hat{Q}_{\mathcal{P}}(X)\right] = \mu_1 X + \mu_2(I - X)$ with $\mu_1 > \mu_2$, i.e., the input signal X is "on average" the principal eigenspace of $\hat{Q}_{\mathcal{P}}(X)$. The matrix Bernstein inequality, Theorem 1.4.15, is used to show for large m that $\hat{Q}_{\mathcal{P}}(X)$ concentrates around its expectation, and this concentration passes to the principal eigenspace to show $\hat{X} \approx X$. The spectral gap $\mu_1 - \mu_2$ of $\mathbb{E}\left[\hat{Q}_{\mathcal{P}}(X)\right]$ measures how well on average PEP can distinguish X from other rank-one projections, and will play an important role in proving reconstruction guarantees for PEP and establishing its robustness to noise.

2.3 Spectral decompositions for PEP

In this section, the expectation of the empirical average of proximal projections, $\mathbb{E}\left[\hat{Q}_{\mathcal{P}}(X)\right]$, is computed for magnitude comparison measurements comprised of independent or complementary pairs of projections. Recovery guarantees for PEP will rely in large part on the spectral decomposition of $\mathbb{E}\left[\hat{Q}_{\mathcal{P}}(X)\right]$.

First, a simple fact is shown about the distribution of the non-quantized phaseless measurement $\mathrm{tr}\left[PX\right]$ for a fixed input signal X and random rank-k orthogonal projection. Since $\mathrm{tr}\left[PX\right] = \cos^2(\theta)$ where θ is the principal angle between $\mathrm{Ran}(X)$ and $\mathrm{Ran}(P)$, this lemma can also be interpreted as giving the distribution of the cosine squared of the principal angle between a random k-dimensional subspace and a fixed one-dimensional subspace in $\mathbb{F}^d$.

Lemma 2.3.1. *Let $X \in \mathrm{Proj}_{\mathbb{F}}(1, d)$ be fixed and $P \in \mathrm{Proj}_{\mathbb{F}}(k, d)$ be uniformly distributed. Then $\mathrm{tr}\left[PX\right] \sim \mathrm{Beta}(\beta_{\mathbb{F}} k, \beta_{\mathbb{F}}(d-k))$.*

Proof. Let $x \in \mathrm{Ran}(X)$ with $\|x\|_2 = 1$, which implies $xx^* = X$. Then we may rewrite $\mathrm{tr}\left[PX\right]$ in terms of x as $\mathrm{tr}\left[PX\right] = \|Px\|_2^2$. Recall from Lemma 1.4.9 that if $U \in \mathcal{U}_{\mathbb{F}}(d)$ is uniformly distributed and $P' \in \mathrm{Proj}_{\mathbb{F}}(1, d)$ is fixed, then $U^* P' U \overset{(\mathrm{d})}{=} P$. In particular, this means that

$$\|Px\|_2^2 \overset{(\mathrm{d})}{=} \left\|U^* P' U x\right\|_2^2 = \left\|P' U x\right\|_2^2,$$

where the last equality follows by the fact U^* is a unitary matrix and hence an isometry. Altogether, these facts tell us that

$$\mathrm{tr}\left[PX\right] \overset{(\mathrm{d})}{=} \left\|P' U x\right\|_2^2.$$

By Lemma 1.4.9 we know that Ux is a uniformly distributed unit vector, hence $\|P'Ux\|_2^2$ is the squared-norm of a fixed projection of a uniformly distributed unit vector in $\mathbb{S}_{\mathbb{F}}^{d-1}$. Thus Lemma 1.4.6 implies

$$\left\|P' U x\right\|_2^2 \sim \mathrm{Beta}(\beta_{\mathbb{F}} k, \beta_{\mathbb{F}}(d-k)).$$

Combining all these steps then shows

$$\mathrm{tr}\left[PX\right] \overset{(\mathrm{d})}{=} \left\|P' U x\right\|_2^2 \sim \mathrm{Beta}(\beta_{\mathbb{F}} k, \beta_{\mathbb{F}}(d-k)).$$

□

The expectation of the empirical average of proximal projections associated to a magnitude comparison measurement can now be derived from the distribution of $\operatorname{tr}[PX]$ given in Lemma 2.3.1. First, consider the case where each magnitude comparison comes from an i.i.d. pair of uniformly distributed projections.

Proposition 2.3.2. *Let $d \geq 2$ and $X \in \operatorname{Proj}_{\mathbb{F}}(1,d)$ be fixed. If $\Phi_{\mathcal{P}}$ is the magnitude comparison measurement associated to an independent sequence of uniformly distributed orthogonal projections $\mathcal{P} = \{P_j\}_{j=1}^{2m} \subset \operatorname{Proj}_{\mathbb{F}}(k,d)$, then*

$$\mathbb{E}\left[\hat{Q}_{\mathcal{P}}(X)\right] = \mu_1 X + \mu_2 (I - X)$$

where

$$\mu_1 = \frac{k}{d} + \frac{2B(2\beta_{\mathbb{F}}k, 2\beta_{\mathbb{F}}(d-k))}{\beta_{\mathbb{F}}d\ B(\beta_{\mathbb{F}}k, \beta_{\mathbb{F}}(d-k))^2}, \qquad \mu_2 = \frac{k}{d} - \frac{2B(2\beta_{\mathbb{F}}k, 2\beta_{\mathbb{F}}(d-k))}{\beta_{\mathbb{F}}d(d-1)\ B(\beta_{\mathbb{F}}k, \beta_{\mathbb{F}}(d-k))^2}.$$

Proof. By definition of $\hat{Q}_{\mathcal{P}}(X)$ from Definition 2.2.1 and the fact that the $P_j's$ are identically distributed, we have

$$\mathbb{E}\left[\hat{Q}_{\mathcal{P}}(X)\right] = \mathbb{E}\left[\hat{P}_1(X)\right].$$

If $U \in \mathcal{U}_{\mathbb{F}}(d)$ is an arbitrary unitary that fixes X, i.e., such that $U^*XU = X$, then the distribution of $\hat{P}_1(X)$ is invariant under conjugation by U. Indeed, by definition of $\hat{P}_1(X)$, the cyclic property of the trace, and the rotational invariance of the uniform distribution on orthogonal projections

$$\begin{aligned} U^*\hat{P}_1(X)U &= \begin{cases} U^*P_1U & \text{if } \operatorname{tr}[P_1X] \geq \operatorname{tr}[P_{m+1}X] \\ U^*P_{m+1}U & \text{else} \end{cases} \\ &= \begin{cases} U^*P_1U & \text{if } \operatorname{tr}[U^*P_1UX] \geq \operatorname{tr}[U^*P_{m+1}UX] \\ U^*P_{m+1}U & \text{else} \end{cases} \\ &\overset{(\mathrm{d})}{=} \hat{P}_1(X). \end{aligned}$$

This implies that the expectation of $\hat{P}_1(X)$ is invariant under conjugation by any unitary that fixes X, i.e.,

$$U^*\mathbb{E}\left[\hat{P}_1(X)\right]U = \mathbb{E}\left[U^*\hat{P}_1(X)U\right] = \mathbb{E}\left[\hat{P}_1(X)\right].$$

Since $\hat{P}_1(X)$ is a random Hermitian matrix, its expectation $\mathbb{E}\left[\hat{P}_1(X)\right]$ is Hermitian and thus has a unique spectral decomposition $\mathbb{E}\left[\hat{P}_1(X)\right] = \sum_{k=1}^{l} \lambda_k V_k$ for some $\lambda_1 > \ldots > \lambda_l$ and orthogonal projections V_k with $\mathrm{Ran}(V_{k_1}) \perp \mathrm{Ran}(V_{k_2})$ for all $k_1 \neq k_2$. If U is a unitary with $U^*XU = X$, then

$$\sum_{k=1}^{l} \lambda_k V_k = \mathbb{E}\left[\hat{P}_1(X)\right] = U^*\mathbb{E}\left[\hat{P}_1(X)\right]U = \sum_{k=1}^{l} \lambda_k U^* V_k U,$$

and by uniqueness of the spectral decomposition this implies $U^*V_kU = V_k$ for all $k = 1, \ldots, l$. Thus each eigenspace of $\mathbb{E}\left[\hat{P}_1(X)\right]$ is invariant under conjugation by all unitaries that fix X. The only subspaces with this property are $\mathrm{Ran}(X)$ and $\mathrm{Ran}(I - X)$, and thus $\mathbb{E}\left[\hat{P}_1(X)\right]$ has spectral decomposition of the form

$$\mathbb{E}\left[\hat{P}_1(X)\right] = \mu_1 X + \mu_2(I - X)$$

for some μ_1, μ_2.

To determine the exact value of μ_1, we use the law of total expectation to see

$$\begin{aligned}\mu_1 = \mathbb{E}\left[\mathrm{tr}\left[\hat{P}_1(X)X\right]\right] &= \mathbb{E}\left[\mathrm{tr}\left[\hat{P}_1(X)X\right] \;\middle|\; \mathrm{tr}\left[P_1X\right] \geq \mathrm{tr}\left[P_{m+1}X\right]\right]\mathbb{P}\left\{\mathrm{tr}\left[P_1X\right] \geq \mathrm{tr}\left[P_{m+1}X\right]\right\} \\ &+ \mathbb{E}\left[\mathrm{tr}\left[\hat{P}_1(X)X\right] \;\middle|\; \mathrm{tr}\left[P_1X\right] < \mathrm{tr}\left[P_{m+1}X\right]\right]\mathbb{P}\left\{\mathrm{tr}\left[P_1X\right] < \mathrm{tr}\left[P_{m+1}X\right]\right\}.\end{aligned}$$

By the definition of $\hat{P}_1(X)$ and the fact that $P_1 \overset{(\mathrm{d})}{=} P_{m+1}$, we have

$$\begin{aligned}\mathrm{tr}\left[\mathbb{E}\left[\hat{P}_1(X)X\right] \;\middle|\; \mathrm{tr}\left[P_1X\right] \geq \mathrm{tr}\left[P_{m+1}X\right]\right] &= \mathbb{E}\left[\mathrm{tr}\left[P_1X\right] \;\middle|\; \mathrm{tr}\left[P_1X\right] \geq \mathrm{tr}\left[P_{m+1}X\right]\right] \\ &= \mathbb{E}\left[\mathrm{tr}\left[\hat{P}_{m+1}(X)X\right] \;\middle|\; \mathrm{tr}\left[P_{m+1}X\right] \geq \mathrm{tr}\left[P_1X\right]\right] \\ &= \mathbb{E}\left[\mathrm{tr}\left[\hat{P}_1(X)X\right] \;\middle|\; \mathrm{tr}\left[P_1X\right] < \mathrm{tr}\left[P_{m+1}X\right]\right].\end{aligned}$$

Since $\mathbb{P}\left\{\operatorname{tr}\left[P_1X\right] \geq \operatorname{tr}\left[P_{m+1}X\right]\right\} = \mathbb{P}\left\{\operatorname{tr}\left[P_1X\right] < \operatorname{tr}\left[P_{m+1}X\right]\right\} = \frac{1}{2}$, these steps show that

$$\mu_1 = \mathbb{E}\left[\operatorname{tr}\left[\hat{P}_1(X)X\right]\right] = \mathbb{E}\left[\operatorname{tr}\left[P_1X\right] \;\middle|\; \operatorname{tr}\left[P_1X\right] \geq \operatorname{tr}\left[P_{m+1}X\right]\right]. \tag{15}$$

Since P_1 and P_{m+1} are independent uniformly distributed projections in $\operatorname{Proj}_{\mathbb{F}}(k,d)$, by Lemma 2.3.1 the joint distribution $\left(\operatorname{tr}\left[P_1X\right], \operatorname{tr}\left[P_{m+1}X\right]\right)$ has probability density function defined on $[0,1]^2$ by

$$p(x,y) = \frac{1}{B(\beta_{\mathbb{F}}k, \beta_{\mathbb{F}}(d-k))^2}(xy)^{\beta_{\mathbb{F}}k-1}((1-x)(1-y))^{\beta_{\mathbb{F}}(d-k)-1}.$$

We can express the conditional expectation in (15) with this density function to get

$$\begin{aligned}
\mu_1 &= \frac{2}{B(\beta_{\mathbb{F}}k, \beta_{\mathbb{F}}(d-k))^2}\int_0^1\int_0^x x\cdot(xy)^{\beta_{\mathbb{F}}k-1}((1-x)(1-y))^{\beta_{\mathbb{F}}(d-k)-1}\,dydx \\
&= \frac{2}{B(\beta_{\mathbb{F}}k, \beta_{\mathbb{F}}(d-k))}\int_0^1 \mathbb{P}\left\{b \leq x\right\} x^{\beta_{\mathbb{F}}k}(1-x)^{\beta_{\mathbb{F}}(d-k)-1}\,dx,
\end{aligned} \tag{16}$$

where $b \sim \operatorname{Beta}(\beta_{\mathbb{F}}k, \beta_{\mathbb{F}}(d-k))$. Letting $c \sim \operatorname{Beta}(\beta_{\mathbb{F}}k+1, \beta_{\mathbb{F}}(d-k))$, a property of the cumulative distribution function of beta random variables from [1] says that

$$\mathbb{P}\left\{b \leq x\right\} = \mathbb{P}\left\{c \leq x\right\} + \frac{x^{\beta_{\mathbb{F}}k}(1-x)^{\beta_{\mathbb{F}}(d-k)}}{\beta_{\mathbb{F}}k\; B(\beta_{\mathbb{F}}k, \beta_{\mathbb{F}}(d-k))}.$$

Substituting this identity into (16), and using the fact that the expected value of a random variable evaluated by its own cumulative distribution function is $\frac{1}{2}$, we have

$$\mu_1 = \frac{B(\beta_{\mathbb{F}}k+1, \beta_{\mathbb{F}}(d-k))}{B(\beta_{\mathbb{F}}k, \beta_{\mathbb{F}}(d-k))} + \frac{2B(2\beta_{\mathbb{F}}k+1, 2\beta_{\mathbb{F}}(d-k))}{\beta_{\mathbb{F}}k\; B(\beta_{\mathbb{F}}k, \beta_{\mathbb{F}}(d-k))^2}.$$

Another property of the beta function from [1] shows that

$$\frac{B(\beta_{\mathbb{F}}k+1, \beta_{\mathbb{F}}(d-k))}{B(\beta_{\mathbb{F}}k, \beta_{\mathbb{F}}(d-k))} = \frac{k}{d}$$

and similarly

$$\frac{2B(2\beta_{\mathbb{F}}k+1, 2\beta_{\mathbb{F}}(d-k))}{\beta_{\mathbb{F}}k\ B(\beta_{\mathbb{F}}k, \beta_{\mathbb{F}}(d-k))^2} = \frac{2B(2\beta_{\mathbb{F}}k, 2\beta_{\mathbb{F}}(d-k))}{\beta_{\mathbb{F}}d\ B(\beta_{\mathbb{F}}k, \beta_{\mathbb{F}}(d-k))^2},$$

yielding the desired expression for μ_1. The value of μ_2 follows from the fact that $\operatorname{tr}\left[\mathbb{E}\left[\hat{Q}_{\mathcal{P}}(X)\right]\right] = k$, and so $\mu_1 + (d-1)\mu_2 = k$. □

As mentioned in Section 2.2, the spectral gap of the empirical average of the proximal projections is an important quantity for determining how well an input signal may be identified based on its phaseless measurement. The spectral gap for a magnitude comparison measurement associated to pairs of independent projections is bounded in the next corollary, and it is shown to be maximized for half-dimensioned projections, i.e., of rank $\frac{1}{2}d$.

Corollary 2.3.3. *Let $X \in \mathrm{Proj}_{\mathbb{F}}(1,d)$ be fixed. If $\Phi_{\mathcal{P}}$ is the magnitude comparison measurement associated to an independent sequence of uniformly distributed orthogonal projections $\mathcal{P} = \{P_j\}_{j=1}^{2m} \subset \mathrm{Proj}_{\mathbb{F}}(k,d)$, then the spectral gap $\mu_1(k,d) - \mu_2(k,d)$ of $\mathbb{E}\left[\hat{Q}_{\mathcal{P}}(X)\right]$ is*

$$\mu_1 - \mu_2 = \frac{2B(2\beta_{\mathbb{F}}k, 2\beta_{\mathbb{F}}(d-k))}{\beta_{\mathbb{F}}(d-1)\ B(\beta_{\mathbb{F}}k, \beta_{\mathbb{F}}(d-k))^2},$$

and there are constants c_0, c_1 independent of k and d such that

$$c_0 \frac{\sqrt{k(d-k)}}{d^{3/2}} \leq \mu_1 - \mu_2 \leq c_1 \frac{\sqrt{k(d-k)}}{d^{3/2}}.$$

In particular, for a fixed dimension d the spectral gap is maximized by using rank-$\left\lceil \frac{1}{2}d \right\rceil$ projections.

Proof. Using the expressions for μ_1 and μ_2 as given in Proposition 2.3.2, the spectral gap is

$$\mu_1 - \mu_2 = \frac{2B(2\beta_{\mathbb{F}}k, 2\beta_{\mathbb{F}}(d-k))}{\beta_{\mathbb{F}}(d-1)\ B(\beta_{\mathbb{F}}k, \beta_{\mathbb{F}}(d-k))^2}. \tag{17}$$

Using the fact that $B(\alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}$ from [1], where $\Gamma(x)$ is the gamma function defined by

$$\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}\ dx,$$

we may express the spectral gap in terms of gamma functions as

$$\mu_1 - \mu_2 = \frac{2}{\beta_{\mathbb{F}}(d-1)} \cdot \frac{\Gamma(2\beta_{\mathbb{F}}k)\Gamma(2\beta_{\mathbb{F}}(d-k))\Gamma(\beta_{\mathbb{F}}d)^2}{\Gamma(2\beta_{\mathbb{F}}d)\Gamma(\beta_{\mathbb{F}}k)^2\Gamma(\beta_{\mathbb{F}}(d-k))^2}. \tag{18}$$

Treating k as a real-valued parameter in the interval $[1, d-1]$, we can use calculus to find the maximum value of the spectral gap as a function of k. To do this, we employ the digamma function ψ_0 which is defined by $\Gamma'(x) = \Gamma(x)\psi_0(x)$ for $x > 0$. Letting $a = \frac{2\Gamma(\beta_{\mathbb{F}}d)^2}{\beta_{\mathbb{F}}(d-1)\Gamma(2\beta_{\mathbb{F}}d)}$, $g(k) = \Gamma(2\beta_{\mathbb{F}}k)\Gamma(2\beta_{\mathbb{F}}(d-k))$ and $h(k) = \Gamma(\beta_{\mathbb{F}}k)^2\Gamma(\beta_{\mathbb{F}}(d-k))^2$ we have

$$f(k) := \mu_1 - \mu_2 = a \cdot \frac{g(k)}{h(k)}.$$

Taking the derivative of f with the quotient rule yields

$$f'(k) = f(k) \cdot 2\beta_{\mathbb{F}} \left[\psi_0(2\beta_{\mathbb{F}}k) - \psi_0(2\beta_{\mathbb{F}}(d-k)) - \psi_0(\beta_{\mathbb{F}}k) + \psi_0(\beta_{\mathbb{F}}(d-k))\right].$$

If $j(k) = 2\beta_{\mathbb{F}} \left[\psi_0(2\beta_{\mathbb{F}}k) - \psi_0(2\beta_{\mathbb{F}}(d-k)) - \psi_0(\beta_{\mathbb{F}}k) + \psi_0(\beta_{\mathbb{F}}(d-k))\right]$, then since $f(k) > 0$ we have $f'(k) = 0$ if and only if $j(k) = 0$, and $\operatorname{sgn}(f'(k)) = \operatorname{sgn}(j(k))$. By the multiplication theorem for the gamma function [1], $\psi_0(2x) = \log(2) + \frac{1}{2}\psi_0(x) + \frac{1}{2}\psi_0(x + \frac{1}{2})$, and thus

$$j(k) = \beta_{\mathbb{F}} \left[\psi_0\left(\beta_{\mathbb{F}}k + \frac{1}{2}\right) - \psi_0(\beta_{\mathbb{F}}k) + \psi_0(\beta_{\mathbb{F}}(d-k)) - \psi_0\left(\beta_{\mathbb{F}}(d-k) + \frac{1}{2}\right)\right].$$

Clearly, $j(\frac{1}{2}d) = 0$, and $j(\frac{1}{2}d + r) = -j(\frac{1}{2}d - r)$. For $k \in [1, \frac{1}{2}d)$, since ψ_0 is strictly increasing and ψ_0' is strictly decreasing, we have

$$\psi_0(\beta_{\mathbb{F}}(d-k)) - \psi_0(\beta_{\mathbb{F}}k) \geq \psi_0\left(\beta_{\mathbb{F}}(d-k) + \frac{1}{2}\right) - \psi_0\left(\beta_{\mathbb{F}}k + \frac{1}{2}\right),$$

and hence $j(k) > 0$. By symmetry, this implies $j(k) < 0$ for $k \in (\frac{1}{2}d, d-1]$. Thus the maximum value of f occurs at $\frac{1}{2}d$.

For explicit bounds on the spectral gap, we use Stirling's approximation for the gamma function to approximate the beta functions in (18). In particular, we use the form of Stirling's approximation given in [7] which says that for all $x > 0$ there exists some $0 < r < 1$ such that

$$\Gamma(x) = \sqrt{2\pi} x^{x-\frac{1}{2}} \exp\left(-x + \frac{r}{12x}\right). \tag{19}$$

To simplify notation, let $a = \beta_{\mathbb{F}} k$, $b = \beta_{\mathbb{F}}(d-k)$, and $c = \beta_{\mathbb{F}} d$. Then Stirling's formula says

$$B(2\beta_{\mathbb{F}} k, 2\beta_{\mathbb{F}}(d-k)) = \frac{\sqrt{\pi} a^{2a-\frac{1}{2}} b^{2b-\frac{1}{2}}}{c^{2c-\frac{1}{2}}} \cdot \exp\left(\frac{r_1}{16a} + \frac{r_2}{16b} - \frac{r_3}{16c}\right),$$

and

$$B(\beta_{\mathbb{F}} k, \beta_{\mathbb{F}}(d-k))^2 = \frac{2\pi a^{2a-1} b^{2b-1}}{c^{2c-1}} \cdot \exp\left(\frac{r_4}{4a} + \frac{r_5}{4b} - \frac{r_6}{4c}\right),$$

where $0 < r_j < 1$ for $j = 1, \ldots, 6$. Making these substitutions in (17), we have

$$\mu_1 - \mu_2 = \frac{\sqrt{ab}}{\beta_{\mathbb{F}} \sqrt{\pi c}(d-1)} \cdot \exp\left(\frac{r_1}{16a} + \frac{r2}{16b} - \frac{r_3}{16c} - \frac{r_4}{4a} - \frac{r_5}{4b} + \frac{r_6}{4c}\right).$$

Substituting our expressions for a, b, and c, we observe that

$$\frac{\sqrt{ab}}{\beta_{\mathbb{F}} \sqrt{\pi c}(d-1)} = \frac{\sqrt{k(d-k)}}{\sqrt{\beta_{\mathbb{F}} \pi d}(d-1)}.$$

The exponential multiplicative factor may be bounded above and below by a constant. □

From this analysis, it follows that for fixed k, as $d \to \infty$ the spectral gap satisfies $\mu_1 - \mu_2 = O(\frac{1}{d})$. On the other hand, if $k = \alpha d$ for some fixed α and $d \to \infty$, then $\mu_1 - \mu_2 = O(\frac{1}{\sqrt{d}})$. In particular, $k = \lceil \frac{1}{2} d \rceil$ gives the maximal spectral gap.

The next proposition gives the expectation of the empirical average of proximal projections for a complementary magnitude comparison measurement.

Proposition 2.3.4. *Let $X \in \mathrm{Proj}_{\mathbb{F}}(1,d)$ be fixed. If $\Phi_{\mathcal{P}}$ is the complementary magnitude comparison measurement associated to an independent sequence of uniformly distributed orthogonal* *projections* $\mathcal{P} = \{P_j\}_{j=1}^m \subset \mathrm{Proj}_{\mathbb{F}}(k,d)$*, then*

$$\mathbb{E}\left[\hat{Q}_{\mathcal{P}}(X)\right] = \mu_1(k,d)X + \mu_2(k,d)(I - X)$$

for

$$\begin{aligned}\mu_1 &= \frac{k}{d}p_k + \frac{d-k}{d}(1-p_k) + \frac{2}{\beta_{\mathbb{F}} d 2^{\beta_{\mathbb{F}} d} B(\beta_{\mathbb{F}} k, \beta_{\mathbb{F}}(d-k))}\\ \mu_2 &= \frac{k}{d}p_k + \frac{d-k}{d}(1-p_k) - \frac{2}{\beta_{\mathbb{F}} d(d-1) 2^{\beta_{\mathbb{F}} d} B(\beta_{\mathbb{F}} k, \beta_{\mathbb{F}}(d-k))},\end{aligned}$$

where $p_k = \mathbb{P}\left\{\mathrm{tr}\left[PX\right] \geq \frac{1}{2}\right\}$ *for a uniformly distributed* $P \in \mathrm{Proj}_{\mathbb{F}}(k,d)$*.*

Proof. By definition of $\hat{Q}_{\mathcal{P}}(X)$ from Definition 2.2.1 and the fact that the $P_j's$ are identically distributed, we have

$$\mathbb{E}\left[\hat{Q}_{\mathcal{P}}(X)\right] = \mathbb{E}\left[\hat{P}_1(X)\right].$$

If $U \in \mathcal{U}_{\mathbb{F}}(d)$ is an arbitrary unitary that fixes X, i.e., such that $U^*XU = X$, then the distribution of $\hat{P}_1(X)$ is invariant under conjugation by U. Indeed, by definition of $\hat{P}_1(X)$ for a complementary magnitude comparison, the cyclic property of the trace, and the rotational invariance of the uniform distribution on orthogonal projections, we have

$$\begin{aligned}U^*\hat{P}_1(X)U &= \begin{cases} U^*P_1U & \text{if } \mathrm{tr}\left[P_1X\right] \geq \frac{1}{2} \\ U^*(I-P_1)U & \text{else}\end{cases}\\ &= \begin{cases} U^*P_1U & \text{if } \mathrm{tr}\left[U^*P_1UX\right] \geq \frac{1}{2} \\ U^*(I-P_1)U & \text{else}\end{cases}\\ &\overset{(\mathrm{d})}{=} \hat{P}_1(X).\end{aligned}$$

This implies that the expectation of $\hat{P}_1(X)$ is invariant under conjugation by any unitary that fixes

X, i.e.,

$$U^*\mathbb{E}\left[\hat{P}_1(X)\right]U = \mathbb{E}\left[U^*\hat{P}_1(X)U\right] = \mathbb{E}\left[\hat{P}_1(X)\right].$$

Since $\hat{P}_1(X)$ is a random Hermitian matrix, its expectation $\mathbb{E}\left[\hat{P}_1(X)\right]$ is Hermitian. The invariance of $\mathbb{E}\left[\hat{P}_1(X)\right]$ under conjugation by all unitaries that fix X implies that it has a spectral decomposition of the form

$$\mathbb{E}\left[\hat{P}_1(X)\right] = \mu_1 X + \mu_2(I - X)$$

for some $\mu_1, \mu_2 \geq 0$ as in the proof of Lemma 1.4.10.

To determine the exact value of μ_1, we compute

$$\begin{aligned}\mu_1 &= \operatorname{tr}\left[\mathbb{E}\left[\hat{P}_1(X)\right]X\right]\\ &= \mathbb{E}\left[\operatorname{tr}\left[P_1X\right] \mid \operatorname{tr}\left[P_1X\right] \geq \frac{1}{2}\right]\mathbb{P}\left\{\operatorname{tr}\left[P_1X\right] \geq \frac{1}{2}\right\}\\ &+ \mathbb{E}\left[\operatorname{tr}\left[(I-P_1)X\right] \mid \operatorname{tr}\left[(I-P_1)X\right] > \frac{1}{2}\right]\mathbb{P}\left\{\operatorname{tr}\left[(I-P_1)X\right] > \frac{1}{2}\right\}.\end{aligned}$$

If $v_k = \mathbb{E}\left[\operatorname{tr}\left[PX\right] \mid \operatorname{tr}\left[PX\right] \geq \frac{1}{2}\right]$ and $p_k = \mathbb{P}\left\{\operatorname{tr}\left[PX\right] \geq \frac{1}{2}\right\}$ for a uniformly distributed $P \in \operatorname{Proj}_{\mathbb{F}}(k,d)$, then using the fact that $\operatorname{tr}\left[P_1X\right] + \operatorname{tr}\left[(I-P_1)X\right] = 1$ we have

$$\mu_1 = v_k p_k + v_{d-k} p_{d-k}. \tag{20}$$

To compute v_k, observe that $\operatorname{tr}\left[PX\right] \sim \operatorname{Beta}(\beta_{\mathbb{F}}k, \beta_{\mathbb{F}}(d-k))$ for a uniformly distributed $P \in \operatorname{Proj}_{\mathbb{F}}(k,d)$ by Lemma 2.3.1. Using the probability density function from Definition 1.4.5 we compute

$$v_k p_k = \frac{1}{B(\beta_{\mathbb{F}}k, \beta_{\mathbb{F}}(d-k))}\int_{\frac{1}{2}}^{1} x \cdot x^{\beta_{\mathbb{F}}k-1}(1-x)^{\beta_{\mathbb{F}}(d-k)-1}\,dx.$$

Let $u = x^{\beta_{\mathbb{F}}k}$ and $dv = (1-x)^{\beta_{\mathbb{F}}(d-k)} - 1$, so $du = \beta_{\mathbb{F}}kx^{\beta_{\mathbb{F}}k-1}dx$ and $v = -\frac{1}{\beta_{\mathbb{F}}(d-k)}(1-x)^{\beta_{\mathbb{F}}(d-k)}$, and integrate by parts to get

$$v_k p_k = \frac{1}{\beta_{\mathbb{F}}(d-k)2^{\beta_{\mathbb{F}}d}B(\beta_{\mathbb{F}}k, \beta_{\mathbb{F}}(d-k))} + \frac{k}{d-k}\left(p_k - v_k p_k\right).$$

Rearranging this equation and solving for $v_k p_k$ yields

$$v_k p_k = \frac{1}{\beta_{\mathbb{F}} d 2^{\beta_{\mathbb{F}} d} B(\beta_{\mathbb{F}} k, \beta_{\mathbb{F}}(d-k))} + \frac{k}{d} p_k.$$

Thus by (20) we have

$$\mu_1 = \frac{2}{\beta_{\mathbb{F}} d 2^{\beta_{\mathbb{F}} d} B(\beta_{\mathbb{F}} k, \beta_{\mathbb{F}}(d-k))} + \frac{k}{d} p_k + \frac{d-k}{d}(1-p_k).$$

Observe that $\mu_1 + (d-1)\mu_2 = \operatorname{tr}\left[\mathbb{E}\left[\hat{Q}_{\mathcal{P}}(X)\right]\right] = k p_k + (d-k)(1-p_k)$, which yields the desired expression for μ_2. □

The spectral gap for a complementary magnitude comparison measurement is bounded in the next corollary, and it is also shown to be maximized for half-dimensioned projections, i.e., of rank $\frac{1}{2}d$.

Corollary 2.3.5. *Let $X \in \mathrm{Proj}_{\mathbb{F}}(1,d)$ be fixed. If $\Phi_{\mathcal{P}}$ is the complementary magnitude comparison measurement associated to an independent sequence of uniformly distributed orthogonal projections $\mathcal{P} = \{P_j\}_{j=1}^m \subset \mathrm{Proj}_{\mathbb{F}}(k,d)$, then the spectral gap $\mu_1(k,d) - \mu_2(k,d)$ of $\mathbb{E}\left[\hat{Q}_{\mathcal{P}}(X)\right]$ is*

$$\mu_1 - \mu_2 = \frac{1}{\beta_{\mathbb{F}}(d-1) 2^{\beta_{\mathbb{F}} d - 1} B(\beta_{\mathbb{F}} k, \beta_{\mathbb{F}}(d-k))},$$

and there are constants c_0, c_1 independent of k and d such that

$$c_0 r(k,d) \leq \mu_1 - \mu_2 \leq c_1 r(k,d).$$

where $r(k,d) = \frac{\sqrt{k(d-k)}}{d^{3/2}} \left(\frac{d}{2k}\right)^{\beta_{\mathbb{F}} k} \left(\frac{d}{2(d-k)}\right)^{\beta_{\mathbb{F}}(d-k)}$. In particular, for a fixed dimension d, the spectral gap is maximized by using rank-$\left\lceil \frac{1}{2} d \right\rceil$ projections.

Proof. To bound the spectral gap, we have from Proposition 2.3.4 that

$$\mu_1 - \mu_2 = \frac{1}{\beta_{\mathbb{F}}(d-1) 2^{\beta_{\mathbb{F}} d - 1} B(\beta_{\mathbb{F}} k, \beta_{\mathbb{F}}(d-k))}.$$

Rewriting the beta function in terms of gamma functions, we have

$$\mu_1 - \mu_2 = \frac{\Gamma(\beta_{\mathbb{F}} d)}{\beta_{\mathbb{F}}(d-1)2^{\beta_{\mathbb{F}} d-1}\Gamma(\beta_{\mathbb{F}} k)\Gamma(\beta_{\mathbb{F}}(d-k))}. \tag{21}$$

Treating k as a real-valued parameter in the interval $[1, d-1]$, we can use calculus to find the maximum value of the spectral gap as a function of k. Letting $a = \frac{\Gamma(\beta_{\mathbb{F}} d)}{\beta_{\mathbb{F}}(d-1)2^{\beta_{\mathbb{F}} d-1}}$ and $g(k) = \Gamma(\beta_{\mathbb{F}} k)\Gamma(\beta_{\mathbb{F}}(d-k))$, we see that

$$f(k) := \mu_1 - \mu_2 = a \cdot \frac{1}{g(k)}.$$

Taking the derivative of f yields

$$f'(k) = f(k) \cdot \left[\psi_0(\beta_{\mathbb{F}}(d-k)) - \psi_0(\beta_{\mathbb{F}} k)\right].$$

Since the digamma function is strictly increasing, we see that $f'(\frac{1}{2}d) = 0$, $f'(k) > 0$ for $k \in [1, \frac{1}{2}d)$, and $f'(k) < 0$ for $k \in (\frac{1}{2}d, d-1]$. Thus the spectral gap is maximized by using rank-$\lceil \frac{1}{2}d \rceil$ projections.

Using Stirling's approximation as in our proof of Proposition 2.3.2, we have

$$B(\beta_{\mathbb{F}} k, \beta_{\mathbb{F}}(d-k))^{-1} = \frac{\sqrt{\beta_{\mathbb{F}}}\, d^{\beta_{\mathbb{F}} d - \frac{1}{2}}}{\sqrt{2\pi}\, k^{\beta_{\mathbb{F}} k - \frac{1}{2}} (d-k)^{\beta_{\mathbb{F}}(d-k) - \frac{1}{2}}} \cdot \exp\left(-\frac{r_1}{8\beta_{\mathbb{F}} k} - \frac{r_2}{8\beta_{\mathbb{F}}(d-k)} + \frac{r_3}{8\beta_{\mathbb{F}} d}\right)$$

for $0 < r_1, r_2, r_3 < 1$. Observe that

$$\frac{d^{\beta_{\mathbb{F}} d - \frac{1}{2}}}{k^{\beta_{\mathbb{F}} k - \frac{1}{2}} (d-k)^{\beta_{\mathbb{F}}(d-k) - \frac{1}{2}}} = \sqrt{\frac{k(d-k)}{d}} \left(\frac{d}{k}\right)^{\beta_{\mathbb{F}} k} \left(\frac{d}{d-k}\right)^{\beta_{\mathbb{F}}(d-k)}.$$

from which the desired bounds follow by substitution into (21). □

Complementary magnitude comparison measurements associated to half-dimensioned projections, i.e., rank-n projections on $\mathbb{F}^{2n}$, are of particular interest. In this case, the beta distribution of $\operatorname{tr}[PX]$ has nice properties that allow the expression for $\mathbb{E}\left[\hat{Q}_{\mathcal{P}}(X)\right]$ given in Proposition 2.3.4

to be simplified. These elementary properties are stated in the following lemma.

Lemma 2.3.6. *Let $X \in \mathrm{Proj}_{\mathbb{F}}(1, 2n)$ be fixed and $P \in \mathrm{Proj}_{\mathbb{F}}(n, 2n)$ be uniformly distributed. Then* $\mathrm{tr}\,[PX] \sim \mathrm{Beta}(\beta_{\mathbb{F}} n, \beta_{\mathbb{F}} n)$ *has the following properties:*

(a) The distribution of $\mathrm{tr}\,[PX]$ *is symmetric about* $\frac{1}{2}$

(b) $\mathbb{E}\,[\mathrm{tr}\,[PX]] = \frac{1}{2}$

(c) $\mathbb{P}\left\{\mathrm{tr}\,[PX] \geq \frac{1}{2}\right\} = \mathbb{P}\left\{\mathrm{tr}\,[PX] \leq \frac{1}{2}\right\} = \frac{1}{2}$

(d) $\mathrm{var}(\mathrm{tr}\,[PX]) = \frac{1}{4(2\beta_{\mathbb{F}} n - 1)}$.

Proof. The fact that $\mathrm{tr}\,[PX] \sim \mathrm{Beta}(\beta_{\mathbb{F}} n, \beta_{\mathbb{F}} n)$ follows directly from Lemma 2.3.1. By Definition 1.4.5, this means that $\mathrm{tr}\,[PX]$ has the probability density function

$$f(t) = \frac{1}{B(\beta_{\mathbb{F}} n, \beta_{\mathbb{F}} n)} t^{\beta_{\mathbb{F}} n - 1}(1-t)^{\beta_{\mathbb{F}} n - 1}.$$

For (a), observe that for any $s \in [0, \frac{1}{2}]$ we have

$$f\left(\frac{1}{2} + s\right) = \frac{1}{B(\beta_{\mathbb{F}} n, \beta_{\mathbb{F}} n)} \left(\frac{1}{2} + s\right)^{\beta_{\mathbb{F}} n - 1} \left(\frac{1}{2} - s\right)^{\beta_{\mathbb{F}} n - 1} = f\left(\frac{1}{2} - s\right).$$

Thus f is symmetric about $\frac{1}{2}$.

To see (b), we could use the fact that symmetry of the probability density function about $\frac{1}{2}$ further implies that $\mathbb{E}\,[\mathrm{tr}\,[PX]] = \frac{1}{2}$ since the expected value of a symmetric distribution is exactly its point of symmetry. For the sake of completeness, we observe that this holds for this Beta distribution: $\mathbb{E}\left[\mathrm{tr}\,[PX] - \frac{1}{2}\right]$ may be expressed as

$$\mathbb{E}\left[\mathrm{tr}\,[PX] - \frac{1}{2}\right] = \frac{1}{B(\beta_{\mathbb{F}} n, \beta_{\mathbb{F}} n)} \int_0^1 \left(t - \frac{1}{2}\right) t^{\beta_{\mathbb{F}} n - 1}(1-t)^{\beta_{\mathbb{F}} n - 1}\, dt.$$

Making the change of variables $t = \frac{1}{2} + s$, we have

$$\mathbb{E}\left[\operatorname{tr}[PX] - \frac{1}{2}\right] = \frac{1}{B(\beta_{\mathbb{F}} n, \beta_{\mathbb{F}} n)} \int_{-\frac{1}{2}}^{\frac{1}{2}} s \left(\frac{1}{2} + s\right)^{\beta_{\mathbb{F}} n - 1} \left(\frac{1}{2} - s\right)^{\beta_{\mathbb{F}} n - 1} \, ds.$$

Since $g(s) = s$ is an odd function and $h(s) = \left(\frac{1}{2} + s\right)^{\beta_{\mathbb{F}} n - 1} \left(\frac{1}{2} - s\right)^{\beta_{\mathbb{F}} n - 1}$ is an even function, we conclude $g(s)h(s)$ is odd and thus $\int_{-\frac{1}{2}}^{\frac{1}{2}} g(s)h(s) \, ds = 0$. Thus

$$\mathbb{E}\left[\operatorname{tr}[PX] - \frac{1}{2}\right] = \frac{1}{B(\beta_{\mathbb{F}} n, \beta_{\mathbb{F}} n)} \int_{-\frac{1}{2}}^{\frac{1}{2}} g(s)h(s) \, ds = 0,$$

so $\mathbb{E}[\operatorname{tr}[PX]] = \frac{1}{2}$.

For (c), since the interval $[0, \frac{1}{2}]$ is the reflection of the interval $[\frac{1}{2}, 1]$ about the point $\frac{1}{2}$, symmetry of the distribution of $\operatorname{tr}[PX]$ about $\frac{1}{2}$ implies that the probability of $\operatorname{tr}[PX]$ being in $[0, \frac{1}{2}]$ is equal to the probability that it is in $[\frac{1}{2}, 1]$. Thus

$$2\mathbb{P}\left\{\operatorname{tr}[PX] \geq \frac{1}{2}\right\} = \mathbb{P}\left\{\operatorname{tr}[PX] \geq \frac{1}{2}\right\} + \mathbb{P}\left\{\operatorname{tr}[PX] \leq \frac{1}{2}\right\} = 1,$$

which implies

$$\mathbb{P}\left\{\operatorname{tr}[PX] \geq \frac{1}{2}\right\} = \mathbb{P}\left\{\operatorname{tr}[PX] \leq \frac{1}{2}\right\} = \frac{1}{2}.$$

Lastly, for (d), we compute the variance by first evaluating $\mathbb{E}\left[\operatorname{tr}[PX]^2\right]$ via integrating by parts. Using the probability density function for $\operatorname{tr}[PX]$, we see

$$\mathbb{E}\left[\operatorname{tr}[PX]^2\right] = \frac{1}{B(\beta_{\mathbb{F}} n, \beta_{\mathbb{F}} n)} \int_0^1 t^{\beta_{\mathbb{F}} n + 1} (1 - t)^{\beta_{\mathbb{F}} n - 1} \, dt.$$

Letting $u = t^{\beta_{\mathbb{F}} n + 1}$ and $dv = (1-t)^{\beta_{\mathbb{F}} n - 1} \, dt$, we have $du = (\beta_{\mathbb{F}} n + 1) t^{\beta_{\mathbb{F}} n} \, dt$ and $v = -\frac{1}{\beta_{\mathbb{F}} n}(1-t)^{\beta_{\mathbb{F}} n}$.

Integration by parts says

$$\begin{aligned}\int t^{\beta_{\mathbb{F}}n+1}(1-t)^{\beta_{\mathbb{F}}n-1}\ dt &= \int u\ dv \\ &= uv - \int v\ du \\ &= -\frac{1}{\beta_{\mathbb{F}}n}t^{\beta_{\mathbb{F}}n+1}(1-t)^{\beta_{\mathbb{F}}n} + \frac{\beta_{\mathbb{F}}n+1}{\beta_{\mathbb{F}}n}\int t^{\beta_{\mathbb{F}}n}(1-t)^{\beta_{\mathbb{F}}n}\ dt,\end{aligned}$$

so we can evaluate the definite integral

$$\begin{aligned}\int_0^1 t^{\beta_{\mathbb{F}}n+1}(1-t)^{\beta_{\mathbb{F}}n-1}\ dt &= \left[-\frac{1}{\beta_{\mathbb{F}}n}t^{\beta_{\mathbb{F}}n+1}(1-t)^{\beta_{\mathbb{F}}n}\right]_0^1 + \frac{\beta_{\mathbb{F}}n+1}{\beta_{\mathbb{F}}n}\int_0^1 t^{\beta_{\mathbb{F}}n}(1-t)^{\beta_{\mathbb{F}}n}\ dt \\ &= \frac{\beta_{\mathbb{F}}n+1}{\beta_{\mathbb{F}}n}B(\beta_{\mathbb{F}}n+1,\beta_{\mathbb{F}}n+1).\end{aligned}$$

Thus the second moment of $\mathrm{tr}\,[PX]$ is

$$\begin{aligned}\mathbb{E}\left[\mathrm{tr}\,[PX]^2\right] &= \frac{1}{B(\beta_{\mathbb{F}}n,\beta_{\mathbb{F}}n)}\int_0^1 t^{\beta_{\mathbb{F}}n+1}(1-t)^{\beta_{\mathbb{F}}n-1}\ dt \\ &= \frac{(\beta_{\mathbb{F}}n+1)\ B(\beta_{\mathbb{F}}n+1,\beta_{\mathbb{F}}n+1)}{\beta_{\mathbb{F}}n\ B(\beta_{\mathbb{F}}n,\beta_{\mathbb{F}}n)}. \qquad (22)\end{aligned}$$

By using properties of the Beta function, we have

$$B(\beta_{\mathbb{F}}n+1,\beta_{\mathbb{F}}n+1) = B(\beta_{\mathbb{F}}n,\beta_{\mathbb{F}}n+1)\frac{\beta_{\mathbb{F}}n}{2\beta_{\mathbb{F}}n+1} = B(\beta_{\mathbb{F}}n,\beta_{\mathbb{F}}n)\frac{\beta_{\mathbb{F}}n}{2(2\beta_{\mathbb{F}}n+1)}. \qquad (23)$$

Putting together (22) and (23) yields

$$\mathbb{E}\left[\mathrm{tr}\,[PX]^2\right] = \frac{\beta_{\mathbb{F}}n+1}{2(2\beta_{\mathbb{F}}n+1)},$$

and so the variance of $\mathrm{tr}\,[PX]$ is

$$\begin{aligned}\mathrm{var}(\mathrm{tr}\,[PX]) &= \mathbb{E}\left[\mathrm{tr}\,[PX]^2\right] - \mathbb{E}\,[\mathrm{tr}\,[PX]]^2 \\ &= \frac{\beta_{\mathbb{F}} n + 1}{2(2\beta_{\mathbb{F}} n + 1)} - \frac{1}{4} \\ &= \frac{2(\beta_{\mathbb{F}} n + 1) - (2\beta_{\mathbb{F}} n + 1)}{4(2\beta_{\mathbb{F}} n + 1)} \\ &= \frac{1}{4(2\beta_{\mathbb{F}} n + 1)}.\end{aligned}$$

□

Using some of the above facts about the distribution of the phaseless measurement $\mathrm{tr}\,[PX]$ for a random half-dimensioned projection P, the expressions for the spectral decomposition given in Proposition 2.3.4 may be simplified for the case of half-dimensioned projections. We state these simplified expressions in the following corollary.

Corollary 2.3.7. *Let $X \in \mathrm{Proj}_{\mathbb{F}}(1, 2n)$ be fixed. If $\Phi_{\mathcal{P}}$ is the complementary magnitude comparison measurement associated to an independent sequence of uniformly distributed orthogonal projections $\mathcal{P} = \{P_j\}_{j=1}^m \subset \mathrm{Proj}_{\mathbb{F}}(n, 2n)$, then*

$$Q(X) = \mu_1 X + \mu_2 (I - X)$$

for

$$\mu_1 = \frac{1}{2} + \frac{1}{\beta_{\mathbb{F}} n 4^{\beta_{\mathbb{F}} n} B(\beta_{\mathbb{F}} n, \beta_{\mathbb{F}} n)}, \qquad \mu_2 = \frac{1}{2} - \frac{1}{\beta_{\mathbb{F}} n (2n-1) 4^{\beta_{\mathbb{F}} n} B(\beta_{\mathbb{F}} n, \beta_{\mathbb{F}} n)}.$$

In addition to the bounds for the spectral gap given by Corollary 2.3.5, later results will require asymptotically correct bounds for $\mu_1 - \mu_2$ with explicit constants for the case of half-dimensioned projections. First, asymptotically correct bounds are given for the beta function in the special case where both arguments are the same, i.e., $B(x, x)$ for some $x > 0$. Afterwards, the beta function bounds will be used to bound $\mu_1 - \mu_2$ for half-dimensioned complementary magnitude comparisons.

Proposition 2.3.8. *For all $x > 0$,*

$$\frac{2\sqrt{\pi}}{4^x\sqrt{x}} \cdot \exp\left(-\frac{1}{24x}\right) \leq B(x,x) \leq \frac{2\sqrt{\pi}}{4^x\sqrt{x}} \cdot \exp\left(\frac{1}{6x}\right).$$

Proof. From (19), we get an expression for the beta function of the form

$$\begin{aligned} B(x,x) &= \frac{\Gamma(x)^2}{\Gamma(2x)} \\ &= \frac{2\pi x^{2x-1}\exp\left(-2x+\frac{r_1}{6x}\right)}{\sqrt{2\pi}(2x)^{2x-\frac{1}{2}}\exp\left(-2x+\frac{r_2}{24x}\right)} \\ &= \frac{2\sqrt{\pi}}{4^x\sqrt{x}} \cdot \exp\left(\frac{4r_1-r_2}{24x}\right). \end{aligned}$$

The inequalities follow from the fact that $0 < r_1, r_2 < 1$. □

Finally, the bounds in Proposition 2.3.8 give asymptotically correct bounds for the spectral gap $\mu_1 - \mu_2$. While not strictly necessary to prove our pointwise and uniform results, these bounds will be used to compute explicit constants arising in the sufficient number of projections for accurate one-bit phase retrieval of a fixed input signal.

Lemma 2.3.9. *Let μ_1 and μ_2 be as in Corollary 2.3.7. Then we have*

$$\frac{\sqrt{n}}{(2n-1)\sqrt{\beta_{\mathbb{F}}\pi}} \cdot \exp\left(-\frac{1}{6\beta_{\mathbb{F}}n}\right) \leq \mu_1 - \mu_2 \leq \frac{\sqrt{n}}{(2n-1)\sqrt{\beta_{\mathbb{F}}\pi}} \cdot \exp\left(\frac{1}{24\beta_{\mathbb{F}}n}\right).$$

Proof. From the expressions derived in Corollary 2.3.7 we have

$$\begin{aligned} \mu_1 - \mu_2 &= \frac{1}{\beta_{\mathbb{F}}n4^{\beta_{\mathbb{F}}n}B\left(\beta_{\mathbb{F}}n,\beta_{\mathbb{F}}n\right)} + \frac{1}{\beta_{\mathbb{F}}n(2n-1)4^{\beta_{\mathbb{F}}n}B\left(\beta_{\mathbb{F}}n,\beta_{\mathbb{F}}n\right)} \\ &= \frac{2}{\beta_{\mathbb{F}}(2n-1)4^{\beta_{\mathbb{F}}n}B(\beta_{\mathbb{F}}n,\beta_{\mathbb{F}}n)}. \end{aligned} \tag{24}$$

By the bounds on the beta function given in Proposition 2.3.8, we have that

$$\frac{2\sqrt{\pi}}{4^{\beta_{\mathbb{F}}n}\sqrt{\beta_{\mathbb{F}}n}}\exp\left(-\frac{1}{24\beta_{\mathbb{F}}n}\right) \leq B(\beta_{\mathbb{F}}n,\beta_{\mathbb{F}}n) \leq \frac{2\sqrt{\pi}}{4^{\beta_{\mathbb{F}}n}\sqrt{\beta_{\mathbb{F}}n}} \cdot \exp\left(\frac{1}{6\beta_{\mathbb{F}}n}\right).$$

Applying these inequalities to (24) gives the desired bounds for $\mu_1 - \mu_2$. □

2.4 Accurate recovery for a fixed input

This section derives results on the statistics of signal recovery using PEP by considering a fixed input signal while a random collection of projections is used for the phaseless binary measurement. The phaseless binary measurements $\Phi_{\mathcal{P}}$ considered are as in Proposition 2.3.2 and Proposition 2.3.4, i.e., magnitude comparison measurements associated to uniformly distributed independent or complementary pairs of projections. Reconstruction by PEP is defined in Definition 2.2.2. To summarize the entire procedure: for $X \in \mathrm{Proj}_{\mathbb{F}}(1, 2n)$, each bit of the binary string $\Phi_{\mathcal{P}}(X)$ identifies whether $\mathrm{Ran}(X)$ is closer to $\mathrm{Ran}(P_1)$ or $\mathrm{Ran}(P_2)$ for some projections P_1, P_2. PEP then recovers a rank-one projection by taking the principal eigenprojection of the empirical average of the proximal projections, $\hat{Q}_{\mathcal{P}}(X)$. The main goal of this section is to prove that PEP provides accurate recovery of an input signal when sufficiently many random projections are used for the phaseless binary measurement, i.e., when m, the number of bits in the binary measurement, is large enough. This result is contained in Theorem 2.4.4, and gives a solution to the one-bit phase retrieval problem for a fixed input signal, see Problem 1.3.6. The derivation of the result proceeds in three steps:

(1) If the orthogonal projections $\mathcal{P}$ for the magnitude comparison measurement $\Phi_{\mathcal{P}}$ of X are chosen uniformly and independently, then the empirical average of the proximal projections has the expectation $Q(X) := \mathbb{E}\left[\hat{Q}_{\mathcal{P}}(X)\right] = \mu_1 X + \mu_2 (I - X)$ where $0 < \mu_2 < \mu_1$ are constants. In particular, X is the projection onto the eigenspace corresponding to the largest eigenvalue of $Q(X)$.

(2) The empirical average of the proximal projections $\hat{Q}_{\mathcal{P}}(X)$ concentrates near its expectation $Q(X)$.

(3) The orthogonal projection onto the eigenspace of $\hat{Q}_{\mathcal{P}}(X)$ corresponding to its largest eigenvalue concentrates near X.

Step (1) was completed in Section 2.3, so the proof proceeds with step (2).

Concentration of $\hat{Q}_{\mathcal{P}}(X)$ near $Q(X)$

Since the empirical average of the proximal projections $\hat{Q}_{\mathcal{P}}(X)$ is, after all, an empirical average, by the law of large numbers it should concentrate tightly around its expectation $Q(X)$ as the number of measurements m goes to infinity. While the traditional law of large numbers only applies to sequences of one-dimensional random variables, there are generalizations to sequences of random matrices. The matrix Bernstein inequality is used to show concentration of measure for $\hat{Q}_{\mathcal{P}}(X)$, see Theorem 1.4.15 and Corollary 1.4.16 [89, Theorem 1.6.2].

Lemma 2.4.1. *Let $X \in \mathrm{Proj}_{\mathbb{F}}(1,d)$ be fixed. Let $\Phi_{\mathcal{P}}$ be a magnitude comparison measurement as in Proposition 2.3.2 or a complementary magnitude comparison measurement as in Proposition 2.3.4. Then*

$$\mathbb{E}\left[\left\|\hat{Q}_{\mathcal{P}}(X) - Q(X)\right\|\right] \leq \sqrt{\frac{2\log(2d)\max(\mu_1 - \mu_1^2, \mu_2 - \mu_2^2)}{m}} + \frac{\log(2d)\max(\mu_1, 1-\mu_2)}{3m},$$

and for any $t > 0$,

$$\mathbb{P}\left\{\left\|\hat{Q}_{\mathcal{P}}(X) - Q(X)\right\| \geq t\right\} \leq 2d\exp\left(-\frac{t^2 m}{2\max(\mu_1 - \mu_1^2, \mu_2 - \mu_2^2) + \frac{2}{3}\max(\mu_1, 1-\mu_2)t}\right).$$

Proof. Let $S_j = \frac{1}{m}(\hat{P}_j(X) - Q(X))$. Then $\mathbb{E}[S_j] = 0$ and $\|S_j\| \leq \frac{1}{m}\max(\mu_1, 1-\mu_2)$ for all $j = 1, \ldots, m$. Note that $Z := \sum_{j=1}^m S_j = \hat{Q}_{\mathcal{P}}(X) - Q(X)$. Additionally, since $\hat{P}_j(X)$ is a projection and $\mathbb{E}\left[\hat{P}_j(X)\right] = Q(X)$ for all j, we may bound the matrix variance

$$\begin{aligned}
v(Z) &= \left\|\sum_{j=1}^m \mathbb{E}\left[S_j^2\right]\right\| \\
&= \frac{1}{m}\left\|\mathbb{E}\left[(\hat{P}_j(X) - Q(X))^2\right]\right\| \\
&= \frac{1}{m}\left\|Q(X) - Q(X)^2\right\| \\
&= \max(\mu_1 - \mu_1^2, \mu_2 - \mu_2^2).
\end{aligned}$$

The expectation bound and probability bound now follow from applying the matrix Bernstein Inequality as in Theorem 1.4.15. □

As a corollary, the probability in Lemma 2.4.1 can be applied to the case of a complementary magnitude comparison measurement associated to a collection of half-dimensioned projections.

Corollary 2.4.2. *Let $X \in \mathrm{Proj}_{\mathbb{F}}(1, 2n)$ be fixed. Let $\Phi_{\mathcal{P}}$ a complementary magnitude comparison measurement associated to an independent sequence of projections $\mathcal{P} = \{P_j\}_{j=1}^{m} \subset \mathrm{Proj}_{\mathbb{F}}(n, 2n)$. Then for any $0 < t < 1$,*

$$\mathbb{P}\left\{\left\|\hat{Q}_{\mathcal{P}}(X) - Q(X)\right\| \geq t\right\} \leq 4n \exp\left(-\frac{6t^2 m}{7}\right).$$

In particular, if $m \geq \frac{7}{6} t^{-2} \log(4n\rho^{-1})$ then

$$\mathbb{P}\left\{\left\|\hat{Q}_{\mathcal{P}}(X) - Q(X)\right\|\right\} \leq \rho.$$

Proof. For this particular phaseless measurement, we may bound $\max(\mu_1 - \mu_1^2, \mu_2 - \mu_2^2) \leq \frac{1}{4}$ and $\max(\mu_1, 1 - \mu_2)t \leq 1$ and apply Lemma 2.4.1 to get

$$\mathbb{P}\left\{\left\|\hat{Q}_{\mathcal{P}}(X) - Q(X)\right\| \geq t\right\} \leq 4n \exp\left(-\frac{t^2 m}{\frac{1}{2} + \frac{2}{3}}\right) = 4n \exp\left(-\frac{6t^2 m}{7}\right).$$

Additionally, if $m \geq \frac{7}{6} t^{-2} \log(4n\rho^{-1})$ then this probability bound satisfies

$$4n \exp\left(-\frac{6t^2 m}{7}\right) = \exp\left(\log(4n) - \frac{6t^2 m}{7}\right) \leq \rho.$$

□

Concentration of $\hat{X}$ near X (Solution to Problem 1.3.6)

Lemma 2.4.1 says that, with enough random projections for the phaseless binary measurement, with high probability $\hat{Q}_{\mathcal{P}}(X)$ is close to $Q(X)$ in operator norm. When it is sufficiently close, then the eigenspace of $\hat{Q}_{\mathcal{P}}(X)$ corresponding to its maximum eigenvalue will also be close to X. To see this, Lemma 1.3.5 is invoked to extract a factor of $\left\|\hat{X}-X\right\|$ from the operator difference $\hat{X}-X$. This will let us prove another key lemma that will let us control the perturbation of the principal eigenspace of $Q(X)$ under our approximation of it by the empirical average $\hat{Q}_{\mathcal{P}}(X)$.

Lemma 2.4.3. *Let $X, Y \in \mathrm{Proj}_{\mathbb{F}}(1,d)$ with $X \neq Y$ and $Q = \mu_1 X + \mu_2(I-X)$ for $\mu_1, \mu_2 \in \mathbb{F}$. Then*

$$\|X-Y\| = (\mu_1-\mu_2)^{-1}\operatorname{tr}\left[Q(A-B)\right]$$

where $A, B \in \mathrm{Proj}_{\mathbb{F}}(1,2n)$ are the mutually orthogonal projections in the spectral decomposition $X - Y = \|X-Y\|(A-B)$ given in Lemma 1.3.5.

Proof. Let θ be the principal angle between the subspaces associated to X and Y. Then we can pick $x, y, z \in \mathbb{S}_{\mathbb{F}}^{2n-1}$ with $x \perp z$ such that $X = xx^*, Y = yy^*$ and $y = \cos(\theta)x + \sin(\theta)z$. Then

$$Y = yy^* = \cos^2(\theta)xx^* + \sin^2(\theta)zz^* + \sin(\theta)\cos(\theta)(xz^* + zx^*).$$

Since $Q(X) = \mu_1 X + \mu_2(I-X)$, both x and z are eigenvectors of $Q(X)$ with eigenvalues μ_1 and μ_2 respectively, thus

$$\begin{aligned}\operatorname{tr}\left[Q(X)(xz^*+zx^*)\right] &= \operatorname{tr}\left[Q(X)xz^*\right] + \operatorname{tr}\left[Q(X)zx^*\right]\\ &= \operatorname{tr}\left[\mu_1 xz^*\right] + \operatorname{tr}\left[\mu_2 zx^*\right]\\ &= \mu_1\langle x, z\rangle + \mu_2\langle z, x\rangle = 0.\end{aligned}$$

Thus we can evaluate $\operatorname{tr}\left[Q(X)(X-Y)\right]$ to get

$$\begin{aligned}\operatorname{tr}\left[Q(X)(X-Y)\right] &= \operatorname{tr}\left[Q(X)\left(xx^* - \cos^2(\theta)xx^* - \sin^2(\theta)zz^* - \sin(\theta)\cos(\theta)(xz^*+zx^*)\right)\right]\\ &= \sin^2(\theta)\left(\operatorname{tr}\left[Q(X)xx^*\right] - \operatorname{tr}\left[Q(X)zz^*\right]\right)\\ &= (\mu_1-\mu_2)\sin^2(\theta).\end{aligned}$$

Since $\sin(\theta) = \|X - Y\|$, the above chain of equalities says that

$$\operatorname{tr}\left[Q(X)(X - Y)\right] = (\mu_1 - \mu_2)\left\|X - Y\right\|^2 .$$

By Lemma 1.3.5, there exist mutually orthogonal rank-one projections $A, B \in \mathrm{Proj}_{\mathbb{F}}(1, 2n)$ such that $X - Y = \|X - Y\|\,(A - B)$. Thus

$$\|X - Y\| \operatorname{tr}\left[Q(X)(A - B)\right] = \operatorname{tr}\left[Q(X)(X - Y)\right] = (\mu_1 - \mu_2)\left\|X - Y\right\|^2 ,$$

and by the hypothesis that $X \neq Y$ we may cancel a factor of $\|X - Y\|$ on each side and divide by the spectral gap to yield the desired equality. □

Note that $Q(X) = \mathbb{E}\left[\hat{Q}_{\mathcal{P}}(X)\right]$ has the necessary spectral decomposition for Lemma 2.4.3 when the phaseless binary measurement $\Phi_{\mathcal{P}}$ is chosen randomly as in Proposition 2.3.2 or Proposition 2.3.4. Applying Lemma 2.4.3 and the concentration inequality from Lemma 2.4.1 leads to the pointwise error bound for approximate recovery of a fixed input signal using PEP.

Theorem 2.4.4. *Let $X \in \mathrm{Proj}_{\mathbb{F}}(1, d)$ be a fixed input signal, and let $\Phi_{\mathcal{P}}$ be a magnitude comparison measurement as in Proposition 2.3.2 or a complementary magnitude comparison measurement as in Proposition 2.3.4. Then*

$$\mathbb{E}\left[\left\|\hat{X} - X\right\| > \delta\right] \leq \sqrt{\frac{8\log(2d)\max(\mu_1 - \mu_1^2, \mu_2 - \mu_2^2)}{m(\mu_1 - \mu_2)^2}} + \frac{2\log(2d)\max(\mu_1, 1 - \mu_2)}{3m(\mu_1 - \mu_2)}$$

and for any $0 < \delta < 1$

$$\mathbb{P}\left\{\left\|\hat{X} - X\right\| > \delta\right\} \leq 2d\exp\left(-\frac{(\mu_1 - \mu_2)^2\delta^2 m}{8\max(\mu_1 - \mu_1^2, \mu_2 - \mu_2^2) + \frac{8}{3}\max(\mu_1, 1 - \mu_2)(\mu_1 - \mu_2)\delta}\right) \qquad (25)$$

where $\hat{X}$ is the projection onto the principal eigenspace of $\hat{Q}_{\mathcal{P}}(X)$.

Proof. From Lemma 2.4.3, we may express the error in approximating X by $\hat{X}$ as

$$\left\|\hat{X} - X\right\| = (\mu_1 - \mu_2)^{-1} \operatorname{tr}\left[Q(X)(A - B)\right], \tag{26}$$

where $A, B \in \mathrm{Proj}_{\mathbb{F}}(1, d)$ are the orthogonal projections from the spectral decomposition of the difference $X - \hat{X} = \left\|X - \hat{X}\right\| (A - B)$.

Since $\hat{X}$ is the projection onto the principal eigenspace of $\hat{Q}_{\mathcal{P}}(X)$, we know $\operatorname{tr}\left[\hat{Q}_{\mathcal{P}}(X)\hat{X}\right] \geq \operatorname{tr}\left[\hat{Q}_{\mathcal{P}}(X)\hat{X}\right]$, and thus we have inequalities

$$\begin{aligned}
\operatorname{tr}\left[\hat{Q}_{\mathcal{P}}(X)(\hat{X} - X)\right] \geq 0 &\implies \operatorname{tr}\left[\hat{Q}_{\mathcal{P}}(X)(B - A)\right] \geq 0 \\
&\implies (\mu_1 - \mu_2)^{-1} \operatorname{tr}\left[\hat{Q}_{\mathcal{P}}(X)(B - A)\right] \geq 0.
\end{aligned}$$

This shows that adding $(\mu_1 - \mu_2)^{-1} \operatorname{tr}\left[\hat{Q}_{\mathcal{P}}(X)(B - A)\right]$ to the right-hand side of (26) gives an upper bound

$$\left\|\hat{X} - X\right\| \leq (\mu_1 - \mu_2)^{-1} \operatorname{tr}\left[(Q(X) - \hat{Q}(X))(A - B)\right].$$

Since $Q(X) - \hat{Q}(X)$ is Hermitian, we know that for any rank-one projection $Z \in \mathrm{Proj}_{\mathbb{F}}(1, d)$ that

$$\lambda_d(Q(X) - \hat{Q}(X)) \leq \operatorname{tr}\left[(Q(X) - \hat{Q}(X))Z\right] \leq \lambda_1(Q(X) - \hat{Q}(X)),$$

where $\lambda_1(\cdot)$ and $\lambda_d(\cdot)$ denote the largest and smallest eigenvalues of the operator, and that both bounds are sharp by choosing Z to be a rank-one projection onto the eigenspace of $Q(X) - \hat{Q}_{\mathcal{P}}(X)$ corresponding to its smallest (respectively, largest) eigenvalue. Thus

$$\operatorname{tr}\left[(Q(X) - \hat{Q}(X))(A - B)\right] \leq \lambda_1(Q(X) - \hat{Q}(X)) - \lambda_d(Q(X) - \hat{Q}(X)).$$

Since the operator norm of a Hermitian matrix is the largest of the magnitudes of its eigenvalues, we know $\left\|\hat{Q}_{\mathcal{P}}(X)-Q(X)\right\|=\max\left[\left|\lambda_1(Q(X)-\hat{Q}(X))\right|,\left|\lambda_d(Q(X)-\hat{Q}(X))\right|\right]$. It follows that

$$\begin{aligned}\lambda_1(Q(X)-\hat{Q}(X))-\lambda_d(Q(X)-\hat{Q}(X)) &= \left|\lambda_1(Q(X)-\hat{Q}(X))-\lambda_d(Q(X)-\hat{Q}(X))\right|\\ &\leq \left|\lambda_1(Q(X)-\hat{Q}(X))\right|+\left|\lambda_d(Q(X)-\hat{Q}(X))\right|\\ &\leq 2\left\|\hat{Q}_{\mathcal{P}}(X)-Q(X)\right\|.\end{aligned}$$

Altogether these steps show that

$$\left\|\hat{X}-X\right\|\leq(\mu_1-\mu_2)^{-1}\operatorname{tr}\left[(Q(X)-\hat{Q}(X))(A-B)\right]\leq 2(\mu_1-\mu_2)^{-1}\left\|Q(X)-\hat{Q}(X)\right\|. \qquad (27)$$

The result then follows by bounding $\left\|\hat{Q}_{\mathcal{P}}(X)-Q(X)\right\|$ with high probability using Lemma 2.4.1 with $t=\frac{1}{2}(\mu_1-\mu_2)\delta$.

□

Theorem 2.4.4 provides a variety of solutions to the one-bit phase retrieval problem for fixed input signals, Problem 1.3.6. Specifically, it gives recovery guarantees when $\Phi_{\mathcal{P}}$ is the magnitude comparison measurement associated to an independent sequence $\mathcal{P}=\{P_j\}_{j=1}^{2m}\subset\operatorname{Proj}_{\mathbb{F}}(k,d)$, and when $\Phi_{\mathcal{P}}$ is the complementary magnitude comparison measurement associated to an independent sequence $\mathcal{P}=\{P_j\}_{j=1}^{m}\subset\operatorname{Proj}_{\mathbb{F}}(k,d)$. The probability bound in (25) is of main interest, as it ensures that if m is chosen large enough then the random choice of projections will accurately recover X with high probability. Theorem 2.4.4 may be applied to a few natural choices of the rank k of the projections used for the measurement. First, consider independent pairs of rank-one projections.

Corollary 2.4.5. *Let $X\in\operatorname{Proj}_{\mathbb{F}}(1,d)$ be a fixed input signal, $0<\delta<1$ be a desired level of accuracy, and $0<\rho<1$ be an acceptable failure probability. There is a constant C such that if*

$$m\geq C\delta^{-2}d\log(2d\rho^{-1})$$

and $\Phi_{\mathcal{P}}$ is the magnitude comparison measurement associated to an independent sequence of projections $\mathcal{P} = \{P_j\}_{j=1}^{2m} \subset \mathrm{Proj}_{\mathbb{F}}(1, d)$, then with probability at least $1 - \rho$ we have

$$\left\| \hat{X} - X \right\| < \delta$$

where $\hat{X}$ is the projection onto the principal eigenspace of $\hat{Q}_{\mathcal{P}}(X)$.

Proof. By Proposition 2.3.2, we know $Q(X) = \mu_1 X + \mu_2(I - X)$ where

$$\mu_1 = \frac{1}{d} + \frac{2B(2\beta_{\mathbb{F}}, 2\beta_{\mathbb{F}}(d-1))}{\beta_{\mathbb{F}} d\ B(\beta_{\mathbb{F}}, \beta_{\mathbb{F}}(d-1))^2}, \qquad \mu_2 = \frac{1}{d} - \frac{2B(2\beta_{\mathbb{F}}, 2\beta_{\mathbb{F}}(d-1))}{\beta_{\mathbb{F}} d(d-1)\ B(\beta_{\mathbb{F}}, \beta_{\mathbb{F}}(d-1))^2}.$$

In particular, $\mu_1, \mu_2 = O(\frac{1}{d})$ and similarly $\mu_1 - \mu_2 = O(\frac{1}{d})$ by Corollary 2.3.3. Thus, by Theorem 2.4.4 there exists a constant $c > 0$ such that

$$\begin{aligned}\mathbb{P}\left\{ \left\| \hat{X} - X \right\| > \delta \right\} &\leq 2d \exp\left(-\frac{(\mu_1 - \mu_2)^2 \delta^2 m}{8 \max(\mu_1 - \mu_1^2, \mu_2 - \mu_2^2) + \frac{8}{3} \max(\mu_1, 1 - \mu_2)(\mu_1 - \mu_2)\delta} \right) \\ &\leq 2d \exp\left(-\frac{c\delta^2 m}{d} \right).\end{aligned}$$

If $m \geq c^{-1}\delta^{-2} d \log(2d\rho^{-1})$ then $\mathbb{P}\left\{ \left\| \hat{X} - X \right\| > \delta \right\} \leq \rho$. □

Corollary 2.4.5 can be seen as an improvement of the pointwise result in [73]. In particular, in a fixed dimension d using $m = O(\delta^{-2})$ rank-one magnitude comparison is sufficient to guarantee phase retrieval with high probability, as opposed to the $O(\delta^{-4})$ shown in [73].

On the opposite end of the spectrum, the magnitude comparison measurement associated to independent pairs of half-dimensioned projection can be considered in Theorem 2.4.4.

Corollary 2.4.6. *Let $X \in \mathrm{Proj}_{\mathbb{F}}(1, 2n)$ be a fixed input signal, $0 < \delta < 1$ be a desired level of accuracy, and $0 < \rho < 1$ be an acceptable failure probability. There is a constant C such that if*

$$m \geq C\delta^{-2} n \log(4n\rho^{-1})$$

and $\Phi_{\mathcal{P}}$ is the magnitude comparison measurement associated to an independent sequence of projections $\mathcal{P} = \{P_j\}_{j=1}^{2m} \subset \mathrm{Proj}_{\mathbb{F}}(n, 2n)$, then with probability at least $1-\rho$ we have

$$\left\| \hat{X} - X \right\| < \delta$$

where $\hat{X}$ is the projection onto the principal eigenspace of $\hat{Q}_{\mathcal{P}}(X)$.

Proof. By Proposition 2.3.2, we know $Q(X) = \mu_1 X + \mu_2(I - X)$ where

$$\mu_1 = \frac{1}{2} + \frac{B(2\beta_{\mathbb{F}} n, 2\beta_{\mathbb{F}} n)}{\beta_{\mathbb{F}} n \; B(\beta_{\mathbb{F}} n, \beta_{\mathbb{F}} n)^2}, \qquad \mu_2 = \frac{1}{2} - \frac{B(2\beta_{\mathbb{F}} n, 2\beta_{\mathbb{F}} n)}{\beta_{\mathbb{F}} n(2n-1) \; B(\beta_{\mathbb{F}} n, \beta_{\mathbb{F}} n)^2}.$$

In particular, $\mu_1 - \mu_2 = O(\frac{1}{\sqrt{n}})$ by Corollary 2.3.3. Thus, by Theorem 2.4.4 there exists a constant $c > 0$ such that

$$\begin{aligned} \mathbb{P}\left\{ \left\| \hat{X} - X \right\| > \delta \right\} &\le 4n \exp\left(-\frac{(\mu_1 - \mu_2)^2 \delta^2 m}{8 \max(\mu_1 - \mu_1^2, \mu_2 - \mu_2^2) + \frac{8}{3} \max(\mu_1, 1 - \mu_2)(\mu_1 - \mu_2)\delta} \right) \\ &\le 4n \exp\left(-\frac{c\delta^2 m}{n} \right). \end{aligned}$$

If $m \ge c^{-1}\delta^{-2} d \log(2d\rho^{-1})$ then $\mathbb{P}\left\{ \left\| \hat{X} - X \right\| > \delta \right\} \le \rho$. □

The minimal constants C in Corollary 2.4.5 and Corollary 2.4.6 may be different, but otherwise the number of projections used is the same. In other words, half-dimensioned projections provide accurate one-bit phase retrieval via PEP with the same number of bits per measurement as rank-one projections, up to an absolute constant. A similar proof shows that the same guarantee holds for complementary magnitude comparison measurements using half-dimensioned projections, with a potentially different constant C. In this case, an explicit value for C is computed.

Corollary 2.4.7. *Let $X \in \mathrm{Proj}_{\mathbb{F}}(1, 2n)$ be a fixed input signal, $0 < \delta < 1$ be a desired level of accuracy, and $0 < \rho < 1$ be an acceptable failure probability. There is a constant C such that if*

$$m \ge C\delta^{-2} n \log(4n\rho^{-1}) \tag{28}$$

and $\Phi_{\mathcal{P}}$ is the complementary magnitude comparison measurement associated to an independent sequence of projections $\mathcal{P} = \{P_j\}_{j=1}^{m} \subset \mathrm{Proj}_{\mathbb{F}}(n, 2n)$, then with probability at least $1 - \rho$ we have

$$\left\| \hat{X} - X \right\| < \delta$$

where $\hat{X}$ is the projection onto the principal eigenspace of $\hat{Q}_{\mathcal{P}}(X)$. In particular, we can take $C = \frac{28}{3}\beta_{\mathbb{F}}\pi$.

Proof. In the proof of Theorem 2.4.4, we instead apply Lemma 2.4.2 to bound the probability that $\left\| \hat{Q}_{\mathcal{P}}(X) - Q(X) \right\| \geq \frac{1}{2}(\mu_1 - \mu_2)\delta$, yielding that

$$\mathbb{P}\left\{ \left\| \hat{X} - X \right\| > \delta \right\} \leq 4n \exp\left(-\frac{6}{7}(\mu_1 - \mu_2)^2 \delta^2 m \right).$$

Explicit bounds for $\mu_1 - \mu_2$ for this phaseless binary measurement are given in Lemma 2.3.9, and imply that $(\mu_1 - \mu_2)^2 \geq \frac{1}{8\beta_{\mathbb{F}}\pi n}$, and so

$$4n \exp\left(-\frac{6}{7}(\mu_1 - \mu_2)^2 \delta^2 m \right) \leq \exp\left(\log(4n) - \frac{3\delta^2 m}{28\beta_{\mathbb{F}}\pi n} \right).$$

If $m \geq \frac{28\beta_{\mathbb{F}}\pi}{3}\delta^{-2} n \log(4n\rho^{-1})$, then this probability is bounded by ρ. □

Thus, all three of these choices of phaseless binary measurement yield roughly equivalent reconstruction guarantees, up to a dimension-independent constant. Theorem 2.4.4 also gives complete freedom to choose the probability of failure ρ, whereas [73] ensured recovery with a predefined probability of failure $\rho = O(d^{-2})$. Any of the phaseless binary measurements considered in Theorem 2.4.4 — in particular those specified by Corollaries 2.4.5, 2.4.6, and 2.4.7 — can ensure success with high probability by taking $\rho = d^{-\alpha}$ for some $\alpha > 0$, or with overwhelming probability by taking $\rho = \exp(-d)$. For overwhelming probability of success, the number of sufficient bits for the binary measurement increases by a factor of the dimension, as the next corollary says.

Corollary 2.4.8. *Let $X \in \mathrm{Proj}_{\mathbb{F}}(1, 2n)$ and $0 < \delta < 1$ be a desired level of accuracy. There is a* *constant C such that if*

$$m \geq C\delta^{-2}n^2$$

and and $\Phi_{\mathcal{P}}$ is a phaseless binary measurement associated to an independent sequence of projections $\mathcal{P}$ as in Corollary 2.4.5, 2.4.6, or 2.4.7, then with probability at least $1 - \exp(-2n)$

$$\left\|\hat{X} - X\right\| < \delta,$$

where $\hat{X}$ is the solution to PEP with input $\Phi_{\mathcal{P}}(X)$.

The solution to Problem 1.3.6 given by Corollary 2.4.7, i.e., using a complementary magnitude comparison measurement associated to half-dimensioned projections, is of particular interest. It is this phaseless measurement model that is exclusively studied to derive uniform recovery results in Section 2.5. In the notation of Problem 1.3.6, this solution is as follows: for signal reconstruction of $X \in \mathrm{Proj}_{\mathbb{F}}(1, 2n)$, let m be as in (28) and generate m independent uniformly distributed projections of rank-n. Let $\Phi_{\mathcal{P}}$ be the complementary magnitude comparison measurement as in Corollary 2.4.7 and $\mathcal{R}$ be the recovery algorithm given by PEP as in Definition 2.2.2. Then with probability at least $1 - \rho$, this choice of $\Phi_{\mathcal{P}}$ and $\mathcal{R}$ gives one-bit phase retrieval of X, i.e., $\|\mathcal{R}(\Phi_{\mathcal{P}}(X)) - X\| < \delta$.

See Figure 1 for a plot showing how our theoretical bound from Corollary 2.4.7 on the sufficient number of measurements to achieve an accuracy of δ relates to experimental results. The single line separate from the cluster represents the upper bound on δ given by Corollary 2.4.7. The MATLAB code used to generate this plot is included in Appendix A.1.

2.5 Uniformly accurate recovery

As discussed in Section 1.3, uniformly accurate recovery of all input signals is a desirable property for a measurement and reconstruction scheme for one-bit phase retrieval. The task of devising a measurement and reconstruction scheme with this property was formally stated in Problem 1.3.7, called the uniform one-bit phase retrieval problem. In this section the result from Theorem 2.4.4, and more specifically Corollary 2.4.7, is extended to show that the recovery error using PEP is

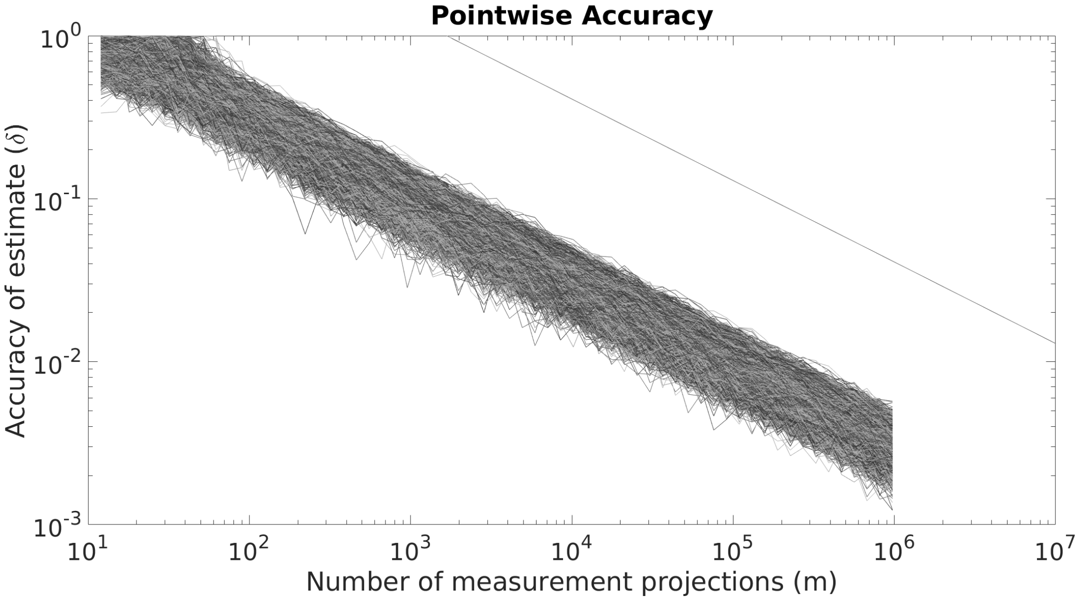

Figure 1: Plot showing the accuracy of recovery using PEP compared to the theoretical upper bound in Corollary 2.4.7 for 7200 independent complementary magnitude comparison measurements, each using a collection of 10^6 half-dimensioned projections on $\mathbb{R}^{16}$.

small uniformly across all input vectors $X \in \mathrm{Proj}_{\mathbb{F}}(1, 2n)$ for a single complementary magnitude comparison measurement $\Phi_{\mathcal{P}}$ associated to a random collection of half-dimensioned projections. In other words, the same measurement and reconstruction scheme used for one-bit phase retrieval of fixed input signals also provides a solution to Problem 1.3.7, although a greater number m of projections will be required to achieve a desired level of uniform accuracy. See Section 2.1 for details about the specific phaseless binary measurement we are considering, and Section 2.2 for the definition of the reconstruction algorithm PEP.

The strategy for extending from the pointwise result in Corollary 2.4.7 to the uniform result given in Theorem 2.5.14 consists of the following steps:

(1) Using sufficiently many random projections, $\hat{Q}_{\mathcal{P}}(X)$ concentrates near $Q(X)$ for all X in an ϵ-net of $\mathrm{Proj}_{\mathbb{F}}(1, 2n)$.

(2) With high probability the measurement Hamming distance between a pair $X, Y \in \mathrm{Proj}_{\mathbb{F}}(1, 2n)$

concentrates near its expected value $\mathbb{E}\left[d_{\mathcal{P}}(X,Y)\right]$, uniformly for all such pairs. See Definition 1.3.2 for the definition of the measurement Hamming distance.

(3) The expected value of the measurement Hamming distance $\mathbb{E}\left[d_{\mathcal{P}}(X,Y)\right]$ is bounded above by the operator norm distance $\|X-Y\|$.

(4) The eigenspace of $\hat{Q}_{\mathcal{P}}(X)$ corresponding to its largest eigenvalue concentrates near X uniformly for all $X \in \mathrm{Proj}_{\mathbb{F}}(1,2n)$.

As previously mentioned, in this section the phaseless binary measurement is always assumed to be a complementary magnitude comparison associated to a collection of random half-dimensioned projections $\mathcal{P} \subset \mathrm{Proj}_{\mathbb{F}}(n,2n)$.

2.5.1 Concentration of $\hat{Q}_{\mathcal{P}}(X)$ near $Q(X)$ uniformly on a net

Recall that a subset $\mathcal{N}_\epsilon \subset \mathrm{Proj}_{\mathbb{F}}(1,2n)$ is called an *ϵ-net* if for every $X \in \mathrm{Proj}_F(1,2n)$ there exists some $Y \in \mathcal{N}_\epsilon$ such that $\|X-Y\| < \epsilon$. This section shows that the empirical average of the proximal projections concentrates uniformly for all input signals in an ϵ-net of $\mathrm{Proj}_{\mathbb{F}}(1,2n)$. This fact relies on the existence of ϵ-nets of $\mathrm{Proj}_{\mathbb{F}}(1,2n)$ with explicit cardinality bounds. First, a basic inequality is derived that related the Euclidean distance between unit vectors to the operator norm distance between their associated rank-one projections.

Lemma 2.5.1. *Let $d \in \mathbb{N}$. Then for all $x,y \in \mathbb{S}_{\mathbb{F}}^{d-1}$*

$$\|xx^* - yy^*\| \leq \|x-y\|_2 \,.$$

Proof. Let θ be the principal angle between the subspaces associated to xx^* and yy^*, and recall $\|xx^*-yy^*\| = \sin(\theta)$. Since $\theta \in [0,\frac{\pi}{2}]$ we know $0 \leq \cos(\theta) \leq 1$, and thus

$$\|xx^*-yy^*\|^2 = \sin^2(\theta) = (1+\cos(\theta))(1-\cos(\theta)) \leq 2(1-\cos) = 2-2\cos(\theta).$$

Since $\cos(\theta)$ is the maximum value of $|\langle x', y'\rangle|$ over unit vectors $x' \in \mathrm{span}\,\{x\}$ and $y' \in \mathrm{span}\,\{y\}$, it follows that

$$2 - 2\cos(\theta) \leq 2 - 2\,|\langle x, y\rangle| \leq 2 - 2\Re\langle x, y\rangle = \langle x - y, x - y\rangle = \|x - y\|_2^2 .$$

Combining these inequalities yields the desired bound. □

Lemma 2.5.1 can now be used to prove the existence of ϵ-nets of $\mathrm{Proj}_{\mathbb{F}}(1, 2n)$ with explicit cardinality bounds. This follows from the analogous results for ϵ-nets of $\mathbb{S}_{\mathbb{F}}^{2n-1}$.

Lemma 2.5.2. *For any* $0 < \epsilon \leq 1$*, there exists an* ϵ*-net* $\mathcal{N}_\epsilon$ *for* $\mathrm{Proj}_{\mathbb{F}}(1, 2n)$ *with respect to the operator norm with cardinality satisfying*

$$\log|\mathcal{N}_\epsilon| \leq 4\beta_{\mathbb{F}} n \log(3\epsilon^{-1}).$$

Proof. By the standard volume bound for the covering number of the sphere in real Euclidean space [12], and the fact that $\mathbb{S}_{\mathbb{C}}^{2n-1}$ is naturally isometric to $\mathbb{S}_{\mathbb{R}}^{4n-1}$ in the Euclidean distance, for every $\epsilon > 0$ there exists an ϵ-net $\mathcal{N}'_\epsilon$ for $\mathbb{S}_{\mathbb{F}}^{2n-1}$ (with respect to the Euclidean distance) with cardinality satisfying

$$|\mathcal{N}'_\epsilon| \leq \left(\frac{3}{\epsilon}\right)^{4\beta_{\mathbb{F}} n} .$$

By Lemma 2.5.1, $\mathcal{N}_\epsilon := \{xx^* : x \in \mathcal{N}'_\epsilon\}$ is an ϵ-net for $\mathrm{Proj}_{\mathbb{F}}(1, 2n)$ with the desired cardinality bound. □

With control of the cardinality of epsilon-nets for $\mathrm{Proj}_{\mathbb{F}}(1, 2n)$, using Lemma 2.4.1 and a union bound shows that with sufficiently many measurements $\hat{Q}_{\mathcal{P}}(X)$ concentrates near $Q(X)$ uniformly for all X in an epsilon-net of $\mathrm{Proj}_{\mathbb{F}}(1, 2n)$.

Lemma 2.5.3. *Let* $\epsilon > 0$ *and* $\mathcal{N}_\epsilon$ *be an* ϵ*-net of* $\mathrm{Proj}_{\mathbb{F}}(1, 2n)$ *such that* $\log|\mathcal{N}_\epsilon| \leq 4\beta_{\mathbb{F}} n \log(3\epsilon^{-1})$*,* $0 < \delta < 1$ *be a desired level of concentration, and* $0 < \rho < 1$ *be an acceptable failure probability. If*

$$m \geq \frac{7}{6}\delta^{-2}\left[\log(4n) + 4\beta n \log(3\epsilon^{-1}) + \log(\rho^{-1})\right]$$

and $\mathcal{P} = \{P_j\}_{j=1}^m$ *is an independent sequence of uniformly distributed projections in* $\mathrm{Proj}_{\mathbb{F}}(n, 2n)$*, then with probability at least* $1-\rho$

$$\left\| \hat{Q}_{\mathcal{P}}(X) - Q(X) \right\| < \delta$$

for all $X \in \mathcal{N}_\epsilon$*.*

Proof. By Corollary 2.4.2 and our assumption on m, for each $X \in \mathcal{N}_\epsilon$ we know

$$\mathbb{P}\left\{ \left\| \hat{Q}_{\mathcal{P}}(X) - Q(X) \right\| \geq \delta \right\} \leq \exp\left(-4\beta n \log(3\epsilon^{-1}) + \log(\rho^{-1})\right),$$

so by looking at the complement of these events we have for $X \in \mathcal{N}_\epsilon$ that

$$\mathbb{P}\left\{ \left\| \hat{Q}_{\mathcal{P}}(X) - Q(X) \right\| < \delta \right\} \geq 1 - \exp\left(-4\beta n \log(3\epsilon^{-1}) + \log(\rho^{-1})\right).$$

By taking a union bound over all $X \in \mathcal{N}_\epsilon$ it follows that

$$\begin{aligned} \mathbb{P}\left\{ \left\| \hat{Q}_{\mathcal{P}}(X) - Q(X) \right\| < \delta \text{ for all } X \in \mathcal{N}_\epsilon \right\} &\geq 1 - |\mathcal{N}_\epsilon| \exp\left(-4\beta n \log(3\epsilon^{-1}) + \log(\rho^{-1})\right) \\ &\geq 1 - \exp\left(\log(|\mathcal{N}_\epsilon|) - 4\beta n \log(3\epsilon^{-1}) + \log(\rho^{-1})\right) \\ &= 1 - \rho. \end{aligned}$$

□

2.5.2 Uniform concentration of measurement Hamming distance

The main goal of this section is to prove Theorem 2.5.9, which says that with sufficiently many measurements, with high probability the measurement Hamming distance $d_{\mathcal{P}}(X, Y)$ concentrates uniformly near its expected value for all pairs $X, Y \in \mathrm{Proj}_{\mathbb{F}}(1, 2n)$ simultaneously. It is relatively simple to show that this happens for fixed X and Y, but showing that it holds uniformly for all such pairs requires more complicated techniques. To this end, the *t-soft Hamming distance* will be defined similarly as in Plan and Vershynin's *Dimension reduction by random hyperplane tessellations* [79]. The t-soft Hamming distance admits a type of continuity property which will be used to show uniform concentration of the measurement Hamming distance near its expected value over all of $\mathrm{Proj}_{\mathbb{F}}(1,2n)$, Theorem 2.5.9.

In order to motivate the definition of the t-soft Hamming distance, define for any pair of input signals $X, Y \in \mathrm{Proj}_{\mathbb{F}}(1,2n)$ the set

$$\mathcal{S}_{X,Y} := \{P \in \mathrm{Proj}_{\mathbb{F}}(n,2n) : \phi_P(X) \neq \phi_P(Y)\},$$

i.e., the set of projections P that yield different one-bit measurements of X and Y under the complementary magnitude comparison ϕ_P. If $P \in \mathcal{S}_{X,Y}$, then P is said to *separate* X and Y. By the definition of the binary question ϕ_P, Definition 2.1.2, there is an equivalent expression for $\mathcal{S}_{X,Y}$ given by

$$\begin{aligned}\mathcal{S}_{X,Y} =&\{P \in \mathrm{Proj}_{\mathbb{F}}(n,2n) : \mathrm{tr}\,[PX] < \frac{1}{2} \leq \mathrm{tr}\,[PY]\}\\ &\cup \{P \in \mathrm{Proj}_{\mathbb{F}}(n,2n) : \mathrm{tr}\,[PY] < \frac{1}{2} \leq \mathrm{tr}\,[PX]\}.\end{aligned}$$

For a sequence $\mathcal{P} = \{P_j\}_{j=1}^m \subset \mathrm{Proj}_{\mathbb{F}}(n,2n)$, recall that the measurement Hamming distance between X and Y is the fraction of projections in $\mathcal{P}$ that separate X and Y, i.e.,

$$d_{\mathcal{P}}(X,Y) = \frac{1}{m}\left|\{j : P_j \in \mathcal{S}_{X,Y}\}\right|.$$

With this expression for the measurement Hamming distance in mind, define subsets

$$\begin{aligned}\mathcal{S}_{X,Y}^t := &\{P \in \mathrm{Proj}_{\mathbb{F}}(n,2n) : \mathrm{tr}\,[PX] + t < \frac{1}{2} \leq \mathrm{tr}\,[PY] - t\}\\ &\cup \{P \in \mathrm{Proj}_{\mathbb{F}}(n,2n) : \mathrm{tr}\,[PY] + t < \frac{1}{2} \leq \mathrm{tr}\,[PX] - t\}\end{aligned}$$

for all $t \in [-\frac{1}{2}, \frac{1}{2}]$. If $P \in \mathcal{S}_{X,Y}^t$ then P is said to *t-separate* X and Y. In a similar fashion to the measurement Hamming distance, the t-soft Hamming distance is defined to be the fraction of projections in $\mathcal{P}$ that t-separate X and Y.

Definition 2.5.4. *Given a sequence of orthogonal projections $\mathcal{P} = \{P_j\}_{j=1}^m$ in $\mathrm{Proj}_{\mathbb{F}}(n, 2n)$ and $t \in [-\frac{1}{2}, \frac{1}{2}]$, define the t-**soft Hamming distance** between input projections $X, Y \in \mathrm{Proj}_{\mathbb{F}}(1, 2n)$ to be*

$$d_{\mathcal{P}}^t(X, Y) := \frac{1}{m} \left| \{j : P_j \in \mathcal{S}_{X,Y}^t\} \right|.$$

The goal is to prove uniform concentration results for the measurement Hamming distance, but its discontinuity causes problems with standard ϵ-net arguments. The t-soft Hamming distance helps work around this discontinuity by adjusting the parameter t which determines how strict the criteria should be for determining whether each binary question φ_{P_j} distinguishes X and Y. This is reflected in the fact that for $t_1 \leq 0 \leq t_2$ we have $\mathcal{S}_{X,Y}^{t_2} \subset \mathcal{S}_{X,Y} \subset \mathcal{S}_{X,Y}^{t_1}$.

The addition of this extra parameter gives rise to a type of continuity for $d_{\mathcal{P}}^t(X, Y)$ where both t and the projections X and Y are allowed to vary. If the projections X, Y are perturbed by a small amount in operator norm, then the Hamming distance between the measurements of the perturbed X and Y can be controlled by slightly increasing/decreasing the parameter t. This fact is contained in the following proposition.

Proposition 2.5.5. *Let $\mathcal{P} = \{P_j\}_{j=1}^m$ be a sequence of projections in $\mathrm{Proj}_{\mathbb{F}}(n, 2n)$, $t \in [-\frac{1}{2}, \frac{1}{2}]$, $0 < \epsilon < 1$, and $X_0, Y_0, X, Y \in \mathrm{Proj}_{\mathbb{F}}(1, 2n)$ such that $\|X - X_0\| < \epsilon$ and $\|Y - Y_0\| < \epsilon$. Then*

$$d_{\mathcal{P}}^{t+\epsilon}(X, Y) \leq d_{\mathcal{P}}^t(X_0, Y_0) \leq d_{\mathcal{P}}^{t-\epsilon}(X, Y)$$

Proof. To see this chain of inequalities, we show the subset inclusions

$$\mathcal{S}_{X,Y}^{t+\epsilon} \subset \mathcal{S}_{X_0,Y_0}^t \subset \mathcal{S}_{X,Y}^{t-\epsilon}, \tag{29}$$

which imply

$$\left|\left\{j : P_j \in \mathcal{S}_{X,Y}^{t+\epsilon}\right\}\right| \leq \left|\{j : P_j \in \mathcal{S}_{X_0,Y_0}^t\}\right| \leq \left|\left\{j : P_j \in \mathcal{S}_{X,Y}^{t-\epsilon}\right\}\right|$$

and hence the desired inequality for the soft Hamming distances in (2.5.5).

To this end, suppose $P \in \mathcal{S}^{t+\epsilon}_{X,Y}$. Then, without loss of generality, we may assume that

$$\operatorname{tr}[PY] + t + \epsilon < \frac{1}{2} < \operatorname{tr}[PX] - t - \epsilon.$$

Since P is a projection we have $|\operatorname{tr}[P(Y_0 - Y)]| \leq \|Y - Y_0\| < \epsilon$, so

$$\operatorname{tr}[PY_0] + t = \operatorname{tr}[PY] - \operatorname{tr}[P(Y - Y_0)] + t \leq \operatorname{tr}[PY] + t + \epsilon < \frac{1}{2}$$

and also

$$\operatorname{tr}[PX_0] - t = \operatorname{tr}[PX] - \operatorname{tr}[P(X - X_0)] - t \geq \operatorname{tr}[PX] - t - \epsilon > \frac{1}{2}.$$

Thus $P \in \mathcal{S}^{t}_{X_0,Y_0}$, which shows $\mathcal{S}^{t+\epsilon}_{X,Y} \subset \mathcal{S}^{t}_{X_0,Y_0}$.

The second set inclusion in (29) follows from above by swapping the roles of X, Y with X_0, Y_0 and replacing t with $t - \epsilon$. □

The next lemma gives a concentration result for for the t-soft Hamming distance between two fixed vectors. It follows from the fact that

$$\begin{aligned} d^t_{\mathcal{P}}(X,Y) &= \frac{1}{m}\left|\left\{j : P_j \in \mathcal{S}^t_{X,Y}\right\}\right| \\ &= \frac{1}{m}\sum_{j=1}^{m} \mathbb{1}_{\mathcal{S}^t_{X,Y}}(P_j), \end{aligned} \tag{30}$$

where $\mathbb{1}_{\mathcal{S}^t_{X,Y}}(\cdot)$ is the indicator function of the set $\mathcal{S}^t_{X,Y}$.

Lemma 2.5.6. *Let $\mathcal{P} = \{P_j\}_{j=1}^m$ be an independent sequence of uniformly distributed projections in $\operatorname{Proj}_{\mathbb{F}}(n, 2n)$, $t \in [-\frac{1}{2}, \frac{1}{2}]$ be a parameter for the soft Hamming distance, $\delta > 0$ be a desired level of concentration, and $X, Y \in \operatorname{Proj}_{\mathbb{F}}(1, 2n)$ be fixed. Then*

$$\mathbb{P}\left\{\left|d^t_{\mathcal{P}}(X,Y) - \mathbb{E}\left[d^t_{\mathcal{P}}(X,Y)\right]\right| > \delta\right\} \leq 2\exp\left(-2\delta^2 m\right).$$

Proof. As mentioned above,

$$d_{\mathcal{P}}^t(X,Y) = \frac{1}{m}\sum_{j=1}^{m} \mathbb{1}_{\mathcal{S}_{X,Y}^t}(P_j).$$

Since the projections $\mathcal{P}$ are independent and identically distributed, the outputs of the indicator function are i.i.d. Bernoulli random variables: for each j, $\mathbb{1}_{\mathcal{S}_{X,Y}^t}(P_j) = 1$ with probability $p = \mathbb{P}\left\{\mathcal{S}_{X,Y}^t\right\}$ and $\mathbb{1}_{\mathcal{S}_{X,Y}^t}(P_j) = 0$ with probability $1-p$. Thus $m \cdot d_{\mathcal{P}}^t(X,Y) \sim \operatorname{Binom}(m,p)$. Observe from (30), $p = \mathbb{E}\left[d_{\mathcal{P}}^t(X,Y)\right]$. The result then follows from a standard Chernoff inequality for binomial random variables, see Theorem 1.4.13 and the ensuing discussion. □

Next, Lemma 2.5.6 and the bound on the size of ϵ-nets of $\operatorname{Proj}_{\mathbb{F}}(1,2n)$ given in Lemma 2.5.2 are used to take a union bound. The result is a bound for the probability that the t-soft Hamming distance is close to its expectation for all pairs of projections in an ϵ-net simultaneously.

Proposition 2.5.7. *Let $\epsilon > 0$ and $\mathcal{N}_\epsilon$ be an ϵ-net of $\operatorname{Proj}_{\mathbb{F}}(1,2n)$ such that $\log|\mathcal{N}_\epsilon| \leq 4\beta_{\mathbb{F}} n \log(3\epsilon^{-1})$. Also, let $t \in [-\frac{1}{2},\frac{1}{2}]$ be a parameter for the soft Hamming distance, $\delta > 0$ be a desired level of accuracy, and $0 < \rho < 1$ be an acceptable failure probability. If*

$$m \geq \frac{1}{2}\delta^{-2}\left(8\beta_{\mathbb{F}} n \log(3\epsilon^{-1}) + \log(\rho^{-1})\right) \tag{31}$$

and $\mathcal{P} = \{P_j\}_{j=1}^m$ is an independent sequence of uniformly distributed projections in $\operatorname{Proj}_{\mathbb{F}}(n,2n)$, then with probability at least $1-\rho$

$$\left|d_{\mathcal{P}}^t(X,Y) - \mathbb{E}\left[d_{\mathcal{P}}^t(X,Y)\right]\right| \leq \delta$$

for all $X, Y \in \mathcal{N}_\epsilon$.

Proof. By Proposition 2.5.6 and taking a union bound over all $\binom{|\mathcal{N}_\epsilon|}{2} \leq \frac{1}{2}|\mathcal{N}_\epsilon|^2$ pairs in $\mathcal{N}_\epsilon \times \mathcal{N}_\epsilon$, we have that

$$\mathbb{P}\left\{\left|d_{\mathcal{P}}^t(X,Y) - \mathbb{E}\left[d_{\mathcal{P}}^t(X,Y)\right]\right| \leq \delta, \text{ for all } (X,Y) \in \mathcal{N}_\epsilon \times \mathcal{N}_\epsilon\right\} \geq 1 - |\mathcal{N}_\epsilon|^2 \exp\left(-2\delta^2 m\right).$$

Using our bound on the cardinality of $|\mathcal{N}_\epsilon|$ and our assumption about m we have

$$\begin{aligned}
|\mathcal{N}_\epsilon|^2 \exp\left(-2\delta^2 m\right) &= \exp\left(2\log(|\mathcal{N}_\epsilon|) - 2\delta^2 m\right) \\
&\leq \exp\left(8\beta_{\mathbb{F}} n \log(3\epsilon^{-1}) - 2\delta^2 m\right) \\
&\leq \rho.
\end{aligned}$$

□

Proposition 2.5.5 showed that both entries in the t-soft Hamming distance may be perturbed slightly in exchange for perturbing the value of t. The following proposition addresses how varying t affects the expected difference of the t-soft Hamming distance from the measurement Hamming distance.

Proposition 2.5.8. *Let $\mathcal{P} = \{P_j\}_{j=1}^m$ be an independent sequence of uniformly distributed projections in $\mathrm{Proj}_{\mathbb{F}}(n, 2n)$, $t \in [-\frac{1}{2}, \frac{1}{2}]$ be a parameter for the soft Hamming distance, and $X, Y \in \mathrm{Proj}(1, 2n)$ be fixed. Then*

$$\left|\mathbb{E}\left[d_{\mathcal{P}}^t(X,Y) - d_{\mathcal{P}}(X,Y)\right]\right| \leq \frac{8\,|t|\,\sqrt{\beta_{\mathbb{F}} n}}{\sqrt{\pi}}.$$

Proof. Because the t-soft and regular Hamming distances are linear combinations of indicator functions, and the fact that the P_j are i.i.d., we have

$$\begin{aligned}
\left|\mathbb{E}\left[d_{\mathcal{P}}^t(X,Y) - d_{\mathcal{P}}(X,Y)\right]\right| &= \left|\mathbb{E}\left[\frac{1}{m}\sum_{j=1}^m \mathbb{1}_{\mathcal{S}_{X,Y}^t}(P_j)\right] - \mathbb{E}\left[\frac{1}{m}\sum_{j=1}^m \mathbb{1}_{\mathcal{S}_{X,Y}}(P_j)\right]\right| \\
&= \left|\mathbb{E}\left[\mathbb{1}_{\mathcal{S}_{X,Y}^t}(P_1) - \mathbb{1}_{\mathcal{S}_{X,Y}}(P_1)\right]\right|.
\end{aligned}$$

By Jensen's inequality we may bring the absolute value inside the expected value to get an upper bound

$$\left|\mathbb{E}\left[\mathbb{1}_{\mathcal{S}_{X,Y}^t}(P_1) - \mathbb{1}_{\mathcal{S}_{X,Y}}(P_1)\right]\right| \leq \mathbb{E}\left[\left|\mathbb{1}_{\mathcal{S}_{X,Y}^t}(P_1) - \mathbb{1}_{\mathcal{S}_{X,Y}}(P_1)\right|\right]. \tag{32}$$

Since $|\mathbb{1}_A - \mathbb{1}_B| = \mathbb{1}_{A\triangle B}$, where $A\triangle B$ is the symmetric difference

$$A\triangle B = (A \setminus B) \cup (B \setminus A),$$

it follows from (32) that

$$\begin{aligned}\mathbb{E}\left[\left|\mathbb{1}_{\mathcal{S}^t_{X,Y}}(P_1) - \mathbb{1}_{\mathcal{S}_{X,Y}}(P_1)\right|\right] &= \mathbb{E}\left[\mathbb{1}_{\mathcal{S}^t_{X,Y}\triangle\mathcal{S}_{X,Y}}(P_1)\right] \\ &= \mathbb{P}\left\{P_1 \in \mathcal{S}^t_{X,Y}\triangle\mathcal{S}_{X,Y}\right\},\end{aligned}$$

and hence

$$\left|\mathbb{E}\left[d^t_{\mathcal{P}}(X,Y) - d_{\mathcal{P}}(X,Y)\right]\right| \leq \mathbb{P}\left\{P_1 \in \mathcal{S}^t_{X,Y}\triangle\mathcal{S}_{X,Y}\right\}. \tag{33}$$

We break up this symmetric difference into two disjoint pieces

$$\mathbb{P}\left\{P_1 \in \mathcal{S}^t_{X,Y}\triangle\mathcal{S}_{X,Y}\right\} = \mathbb{P}\left\{P_1 \in \mathcal{S}^t_{X,Y} \setminus \mathcal{S}_{X,Y}\right\} + \mathbb{P}\left\{P_1 \in \mathcal{S}_{X,Y} \setminus \mathcal{S}^t_{X,Y}\right\}$$

and look at two cases. First, if $t > 0$ then $\mathcal{S}^t_{X,Y} \setminus \mathcal{S}_{X,Y}$ is empty, and

$$\mathcal{S}_{X,Y} \setminus \mathcal{S}^t_{X,Y} \subset \left\{\left|\mathrm{tr}\left[P_1 X\right] - \frac{1}{2}\right| < t\right\} \bigcup \left\{\left|\mathrm{tr}\left[P_1 Y\right] - \frac{1}{2}\right| < t\right\}.$$

Similarly, if $t < 0$ then $\mathcal{S}_{X,Y} \setminus \mathcal{S}^t_{X,Y}$ is empty and again

$$\mathcal{S}^t_{X,Y} \setminus \mathcal{S}_{X,Y} \subset \left\{\left|\mathrm{tr}\left[P_1 X\right] - \frac{1}{2}\right| < -t\right\} \bigcup \left\{\left|\mathrm{tr}\left[P_1 Y\right] - \frac{1}{2}\right| < -t\right\},$$

Since $\mathrm{tr}\left[P_1 X\right] \overset{(d)}{=} \mathrm{tr}\left[P_1 Y\right]$, in both cases (and trivially if $t = 0$) we have

$$\mathbb{P}\left\{P_1 \in \mathcal{S}^t_{X,Y}\triangle\mathcal{S}_{X,Y}\right\} \leq 2\mathbb{P}\left\{\left|\mathrm{tr}\left[P_1 X\right] - \frac{1}{2}\right| < |t|\right\}. \tag{34}$$

By Lemma 2.3.1 we know $\mathrm{tr}\left[P_1 X\right] \sim \mathrm{Beta}(\beta_{\mathbb{F}} n, \beta_{\mathbb{F}} n)$, and so we can bound this probability using the the probability density function of the beta distribution given in Definition 1.4.5. We see that

$$\mathbb{P}\left\{\left|\operatorname{tr}\left[P_1 X\right]-\frac{1}{2}\right|<|t|\right\}=\frac{1}{B(\beta_{\mathbb{F}} n, \beta_{\mathbb{F}} n)} \int_{\frac{1}{2}-t}^{\frac{1}{2}+t}[x(1-x)]^{\beta_{\mathbb{F}} n-1}\ dx, \tag{35}$$

and since $x(1-x) \leq \frac{1}{4}$ for all $x \in [0,1]$ we may bound this integral by

$$\int_{\frac{1}{2}-t}^{\frac{1}{2}+t}[x(1-x)]^{\beta_{\mathbb{F}} n-1}\ dx \leq 2|t|\left(\frac{1}{4}\right)^{\beta_{\mathbb{F}} n-1}.$$

Thus

$$\mathbb{P}\left\{\left|\operatorname{tr}\left[P_1 X\right]-\frac{1}{2}\right|<|t|\right\} \leq \frac{8|t|}{4^{\beta_{\mathbb{F}} n} B(\beta_{\mathbb{F}} n, \beta_{\mathbb{F}} n)}.$$

Using the lower bound for the beta function given in Lemma 2.3.8 then yields

$$\mathbb{P}\left\{\left|\operatorname{tr}\left[P_1 X\right]-\frac{1}{2}\right|<|t|\right\} \leq \frac{4|t| \sqrt{\beta_{\mathbb{F}} n}}{\sqrt{\pi}}. \tag{36}$$

The result follows from combining equation (33) with inequalities (34) and (36). □

The above lemmas and propositions give all the necessary tools to prove that the measurement Hamming distance concentrates uniformly near its expected value for all pairs in $\mathrm{Proj}_{\mathbb{F}}(1,2n)$.

Theorem 2.5.9. *Let $0<\delta<1$ be a desired level of concentration and $0<\rho<1$ be an acceptable failure probability. If*

$$m \geq 2\delta^{-2}\left(8\beta_{\mathbb{F}} n \log\left(96\sqrt{\frac{\beta_{\mathbb{F}} n}{\pi}}\delta^{-1}\right)+\log(2\rho^{-1})\right)$$

and $\mathcal{P}=\{P_j\}_{j=1}^m$ is a collection of independent uniformly distributed projections in $\mathrm{Proj}_{\mathbb{F}}(n,2n)$, then with probability at least $1-\rho$

$$\left|d_{\mathcal{P}}(X,Y)-\mathbb{E}\left[d_{\mathcal{P}}(X,Y)\right]\right|<\delta \tag{37}$$

for all $X, Y \in \mathrm{Proj}_{\mathbb{F}}(1, 2n)$.

Proof. Let $\epsilon = \frac{\sqrt{\pi}}{32\sqrt{\beta_{\mathbb{F}} n}}\delta$ and let $\mathcal{N}_\epsilon$ be an ϵ-net of $\mathrm{Proj}_{\mathbb{F}}(1, 2n)$ with $\log |\mathcal{N}_\epsilon| \leq 4\beta_{\mathbb{F}} n \log(3\epsilon^{-1})$ as in Lemma 2.5.2. By our assumption on m, Proposition 2.5.7 says that

$$\mathbb{P}\left\{ |d_{\mathcal{P}}^{\epsilon}(X, Y) - \mathbb{E}\left[d_{\mathcal{P}}^{\epsilon}(X, Y)\right]| > \frac{\delta}{2} \text{ for some } X, Y \in \mathcal{N}_\epsilon \right\} \leq \frac{\rho}{2}$$

and also

$$\mathbb{P}\left\{ \left|d_{\mathcal{P}}^{-\epsilon}(X, Y) - \mathbb{E}\left[d_{\mathcal{P}}^{-\epsilon}(X, Y)\right]\right| > \frac{\delta}{2} \text{ for some } X, Y \in \mathcal{N}_\epsilon \right\} \leq \frac{\rho}{2},$$

and so with probability at least $1 - \rho$ we have $\left|d_{\mathcal{P}}^{\pm\epsilon}(X, Y) - \mathbb{E}\left[d_{\mathcal{P}}^{\pm\epsilon}(X, Y)\right]\right| \leq \frac{\delta}{2}$ for all $X, Y \in \mathcal{N}_\epsilon$ (call this event $\mathcal{A}$).

Suppose that $\mathcal{A}$ occurs, and consider an arbitrary pair $X, Y \in \mathrm{Proj}(1, 2n)$. Let $X_0, Y_0 \in \mathcal{N}_\epsilon$ such that $\|X - X_0\| < \epsilon$ and $\|Y - Y_0\| < \epsilon$. By Proposition 2.5.5 we know that

$$d_{\mathcal{P}}(X, Y) \leq d_{\mathcal{P}}^{\epsilon}(X_0, Y_0) \leq d_{\mathcal{P}}^{2\epsilon}(X, Y).$$

These inequalities together with $\mathcal{A}$ holding imply

$$\begin{aligned} d_{\mathcal{P}}(X, Y) &\leq d_{\mathcal{P}}^{\epsilon}(X_0, Y_0) \\ &\leq \mathbb{E}\left[d_{\mathcal{P}}^{\epsilon}(X_0, Y_0)\right] + \frac{\delta}{2} \\ &\leq \mathbb{E}\left[d_{\mathcal{P}}^{2\epsilon}(X, Y)\right] + \frac{\delta}{2}. \end{aligned} \tag{38}$$

By Proposition 2.5.8 we may bound the difference in expected values of the 2ϵ-soft Hamming distance and the measurement Hamming distance by $\left|\mathbb{E}\left[d_{\mathcal{P}}^{2\epsilon}(X, Y)\right] - \mathbb{E}\left[d_{\mathcal{P}}(X, Y)\right]\right| \leq \frac{16\epsilon\sqrt{\beta_{\mathbb{F}} n}}{\sqrt{\pi}}$. From our choice of ϵ, this implies

$$\mathbb{E}\left[d_{\mathcal{P}}^{2\epsilon}(X, Y)\right] \leq \mathbb{E}\left[d_{\mathcal{P}}(X, Y)\right] + \frac{\delta}{2}. \tag{39}$$

Together, (38) and (39) show that $d_{\mathcal{P}}(X,Y) \leq \mathbb{E}\left[d_{\mathcal{P}}(X,Y)\right] + \delta$.

Similarly, using Proposition 2.5.5 again shows that $d_{\mathcal{P}}(X,Y) \geq d_{\mathcal{P}}^{-\epsilon}(X_0,Y_0) \geq d_{\mathcal{P}}^{-2\epsilon}(X,Y)$, and since $\mathcal{A}$ holds we have

$$d_{\mathcal{P}}(X,Y) \geq d_{\mathcal{P}}^{-\epsilon}(X_0,Y_0) \geq \mathbb{E}\left[d_{\mathcal{P}}^{-\epsilon}(X_0,Y_0)\right] - \frac{\delta}{2} \geq \mathbb{E}\left[d_{\mathcal{P}}^{-2\epsilon}(X,Y)\right] - \frac{\delta}{2}.$$

Using Proposition 2.5.8 as above but for $t = -\epsilon$ yields $d_{\mathcal{P}}(X,Y) \geq \mathbb{E}\left[d_{\mathcal{P}}(X,Y)\right] - \delta$. □

Bounding $\mathbb{E}\left[d_{\mathcal{P}}(X,Y)\right]$ above by $\|X-Y\|$

Theorem 2.5.9 says that when the measurement projections are chosen uniformly and independently, then $d_{\mathcal{P}}(X,Y)$ concentrates near its expected value $\mathbb{E}\left[d_{\mathcal{P}}(X,Y)\right]$. As observed in the proof of Lemma 2.5.6, $\mathbb{E}\left[d_{\mathcal{P}}(X,Y)\right] = \mathbb{P}\left\{P \in \mathcal{S}_{X,Y}\right\}$, where P is a single uniformly distributed projection in $\mathrm{Proj}_{\mathbb{F}}(n,2n)$.

When $n = 1$, then X, Y, and P are all rank-one projections on $\mathbb{F}^2$ and if P is uniformly distributed then it is easy to see that

$$\mathbb{P}\left\{P \in \mathcal{S}_{X,Y}\right\} \leq \|X-Y\|.$$

This section shows that this upper bound holds for arbitrary n and for $\mathbb{F} = \mathbb{C}$, see Proposition 2.5.12. Deriving this bound requires understanding the joint distribution of $(\mathrm{tr}\left[PX\right], \mathrm{tr}\left[PY\right])$.

By rotational invariance of the distribution of P we may assume that $\mathrm{Ran}(X)$ and $\mathrm{Ran}(Y)$ are in the two-dimensional subspace spanned by e_1 and e_2, the first two standard basis vectors. Viewed as matrices, this means that all entries of X and Y are zero outside of the top-left 2×2 submatrix. Furthermore, if $\tilde{P}, \tilde{X}$, and $\tilde{Y}$ are the top-left 2×2 submatrices of their respective matrices then $(\mathrm{tr}\left[PX\right], \mathrm{tr}\left[PY\right]) = (\mathrm{tr}\left[\tilde{P}\tilde{X}\right], \mathrm{tr}\left[\tilde{P}\tilde{Y}\right])$. We study the joint distribution of $(\mathrm{tr}\left[PX\right], \mathrm{tr}\left[PY\right])$ through the submatrix $\tilde{P}$ acting on $\mathbb{F}^2$.

Since P is Hermitian, so is $\tilde{P}$. Thus we may write $\tilde{P} = \lambda_1 E_1 + \lambda_2 E_2$ where $\lambda_1 \geq \lambda_2$ are the eigenvalues of $\tilde{P}$ and $E_1 \perp E_2$ are the projections onto their corresponding eigenspaces. We write

$\lambda(\tilde{P}) := (\lambda_1, \lambda_2) \in [0,1]^2$, and $E(\tilde{P}) := (E_1, E_2) \in \mathrm{Proj}_{\mathbb{F}}(1,2) \times \mathrm{Proj}_{\mathbb{F}}(1,2)$. By the rotational invariance of P, E_1 and E_2 are both uniformly distributed in $\mathrm{Proj}_2(1,2)$, but they are not independent: $E_2 = I - E_1$ since Hermitian matrices have mutually orthogonal eigenspaces. Note also that $\lambda(\tilde{P})$ and $E(\tilde{P})$ are independent of each other by the rotational invariance of $\tilde{P}$. The distribution of $\lambda(\tilde{P})$ is given in the following lemma.

Lemma 2.5.10. *Let $n \geq 2$ and $P \in \mathrm{Proj}_{\mathbb{F}}(n, 2n)$ be uniformly distributed. Then $\lambda(\tilde{P})$ has probability density function p_n on $\mathcal{D} := \{(x,y) \in [0,1]^2 : y \leq x\}$ defined by*

$$p_n(x,y) := M_n^{-1}(x-y)^{2\beta_{\mathbb{F}}}\left[x(1-x)y(1-y)\right]^{\beta_{\mathbb{F}}(n-1)-1},$$

with the normalization constant

$$M_n := \iint_{\mathcal{D}} (x-y)^{2\beta_{\mathbb{F}}}\left[x(1-x)y(1-y)\right]^{\beta_{\mathbb{F}}(n-1)-1}\, dxdy = \begin{cases} \frac{2}{n-1}B(n-1,n-1) & \text{if } \mathbb{F} = \mathbb{R} \\ \\ \frac{1}{8n-4}B(n-1,n-1)^2 & \text{if } \mathbb{F} = \mathbb{C}. \end{cases}$$

Proof. The probability density functions are given by [6, Proposition 4.1.4] with $p = 2$, $q = 2n-2$, $r = n-2$ and $s = n-2$. It only remains to compute the normalization constants M_n.

Suppose $\mathbb{F} = \mathbb{R}$. Then $p_n(x,y) = M_n^{-1}(x-y)\left[x(1-x)y(1-y)\right]^{\frac{n-3}{2}}$. Define the functions

$$\begin{aligned} f_n(x,y) &= -\frac{1}{n-1}\left[x(1-x)\right]^{\frac{n-3}{2}}\left[y(1-y)\right]^{\frac{n-1}{2}} \\ g_n(x,y) &= -\frac{1}{n-1}\left[x(1-x)\right]^{\frac{n-1}{2}}\left[y(1-y)\right]^{\frac{n-3}{2}}. \end{aligned}$$

With these definitions, we have $p_n = M_n^{-1}(\frac{\partial g_n}{\partial x} - \frac{\partial f_n}{\partial y})$ on $\mathcal{D}$. So by Green's theorem,

$$1 = \iint_{\mathcal{D}} p_n(x,y)\, dxdy = M_n^{-1}\oint_{\partial\mathcal{D}} f_n dx + g_n dy,$$

where $\partial\mathcal{D}$ is the boundary of $\mathcal{D}$. Note that f_n and g_n both vanish on the boundary of $\mathcal{D}$ except for the diagonal $\Delta := \{(x,y) \in \mathcal{D} : x = y\}$, so we only need to compute the line integral over Δ. Parameterizing Δ by $x(t) = y(t) = 1 - t$ for $t \in [0,1]$, we see

$$\begin{aligned} M_n = \oint_{\partial \mathcal{D}} f_n dx + g_n dy &= -\int_0^1 \left[f_n(x(t),y(t)) + g_n(x(t),y(t))\right]\ dt \\ &= \frac{2}{n-1}\int_0^1 t^{n-2}(1-t)^{n-2}\ dt \\ &= \frac{2}{n-1}B(n-1,n-1). \end{aligned}$$

Next, we consider the case when $\mathbb{F} = \mathbb{C}$. Then $p_n(x,y) = M_n^{-1}(x-y)^2\left[x(1-x)y(1-y)\right]^{n-2}$. By symmetry of the distribution under the map $(x,y) \mapsto (y,x)$ we see

$$\begin{aligned} 1 &= \iint_{\mathcal{D}} p_n(x,y)\ dxdy \\ &= \frac{1}{2}\iint_{[0,1]^2} p_n(x,y)\ dxdy \\ &= \frac{M_n^{-1}}{2}\iint_{[0,1]^2} (x-y)^2\left[x(1-x)y(1-y)\right]^{n-2}\ dxdy. \end{aligned}$$

Expanding $(x-y)^2 = x^2 - 2xy + y^2$ and using linearity of the integral and the definition of the beta distribution, we have

$$\begin{aligned} M_n &= \left(\mathbb{E}\left[b^2\right] - \mathbb{E}\left[b\right]^2\right)B(n-1,n-1)^2 \\ &= \operatorname{var}(b)\cdot B(n-1,n-1)^2 \end{aligned}$$

where $b \sim \operatorname{Beta}(n-1,n-1)$. By Lemma 2.3.6, this beta-distributed random variable has variance $\operatorname{var}(b) = \frac{1}{4(2n-1)}$, which determines M_n. □

Let $\mathcal{D}_{\text{Sep}} := \{(x,y) \in \mathcal{D} : y < \frac{1}{2} < x\}$. Then $\lambda(\tilde{P}) \in \mathcal{D}_{\text{Sep}}$ if and only if there exist projections $A, B \in \operatorname{Proj}_{\mathbb{F}}(1,2)$ such that $\tilde{P} \in \mathcal{S}_{A,B}$. This is true because $\lambda_1 = \max_{A' \in \operatorname{Proj}(1,2)} \operatorname{tr}\left[PA'\right]$ and $\lambda_2 = \max_{B' \in \operatorname{Proj}(1,2n)} \operatorname{tr}\left[PB'\right]$. In particular, $P \in \mathcal{S}_{X,Y}$ requires $\lambda(\tilde{P}) \in \mathcal{D}_{\text{Sep}}$. The probability that $\lambda(\tilde{P}) \in \mathcal{D}_{\text{Sep}}$ may be computed as follows.

Lemma 2.5.11. *If $n \geq 2$, and $P \in \mathrm{Proj}_{\mathbb{F}}(n, 2n)$ is uniformly distributed, then*

$$\mathbb{P}\left\{\lambda(\tilde{P}) \in \mathcal{D}_{Sep}\right\} = \begin{cases} \frac{B\left(\frac{n-1}{2}, \frac{n-1}{2}\right)}{2^n B(n-1,n-1)} \to \frac{1}{\sqrt{2}} & \text{if } \mathbb{F} = \mathbb{R} \\ \\ \frac{1}{2} + \frac{4n-2}{(n-1)^2 2^{4n-4} B(n-1,n-1)^2} \to \frac{1}{2} + \frac{1}{\pi} & \text{if } \mathbb{F} = \mathbb{C}. \end{cases}$$

Proof. First, suppose $\mathbb{F} = \mathbb{R}$, so $p_n(x,y) = M_n^{-1}(x-y)\left[x(1-x)y(1-y)\right]^{\frac{n-3}{2}}$. Then,

$$\mathbb{P}\left\{\lambda(\tilde{P}) \in \mathcal{D}_{\mathrm{Sep}}\right\} = M_n^{-1} \int_0^{\frac{1}{2}} \int_{\frac{1}{2}}^1 (x-y)\left[x(1-x)y(1-y)\right]^{\frac{n-3}{2}} dxdy.$$

By linearity and Fubini's theorem, we get

$$\int_0^{\frac{1}{2}} \int_{\frac{1}{2}}^1 (x-y)\left[x(1-x)y(1-y)\right]^{\frac{n-3}{2}} dxdy = \frac{1}{4}\left(\mathbb{E}\left[b \mid b \geq \frac{1}{2}\right] - \mathbb{E}\left[b \mid b \leq \frac{1}{2}\right]\right) B\left(\frac{n-1}{2}, \frac{n-1}{2}\right)^2,$$

where $b \sim \mathrm{Beta}\left(\frac{n-1}{2}, \frac{n-1}{2}\right)$. Calculating these conditional expectations as in Corollary 2.3.7 we get

$$\mathbb{E}\left[b \mid b \geq \frac{1}{2}\right] - \mathbb{E}\left[b \mid b \leq \frac{1}{2}\right] = \frac{1}{(n-1)2^{n-3} B\left(\frac{n-1}{2}, \frac{n-1}{2}\right)},$$

and combining this with Lemma 2.5.10 yields

$$\mathbb{P}\left\{\lambda(\tilde{P}) \in \mathcal{D}_{\mathrm{Sep}}\right\} = \frac{B\left(\frac{n-1}{2}, \frac{n-1}{2}\right)}{2^n B(n-1, n-1)}. \tag{40}$$

Next, suppose $\mathbb{F} = \mathbb{C}$, so $p_n(x,y) = M_n^{-1}(x-y)^2\left[x(1-x)y(1-y)\right]^{n-2}$. Then,

$$\mathbb{P}\left\{\lambda(\tilde{P}) \in \mathcal{D}_{\mathrm{Sep}}\right\} = \int_0^{\frac{1}{2}} \int_{\frac{1}{2}}^1 p_n(x,y)\ dxdy.$$

By using the expression for M_n given in Proposition 2.5.10, expanding $(x-y)^2 = x^2 - 2xy + y^2$, and rewriting integrals in terms of expectations of beta-distributed random variables, we see

$$\begin{aligned}\int_0^{\frac{1}{2}} \int_{\frac{1}{2}}^1 p_n(x,y)\ dxdy &= \frac{8n-4}{B(n-1,n-1)^2} \int_0^{\frac{1}{2}} \int_{\frac{1}{2}}^1 (x-y)^2 \left[x(1-x)y(1-y)\right]^{n-2}\ dxdy \\ &= (4n-2)\left(\mathbb{E}\left[b^2\right] - \mathbb{E}\left[b \mid b \geq \frac{1}{2}\right] \cdot \mathbb{E}\left[b \mid b \leq \frac{1}{2}\right]\right)\end{aligned}$$

where $b \sim \mathrm{Beta}(n-1,n-1)$. We know that $\mathbb{E}\left[b^2\right] = \mathbb{E}\left[b\right]^2 + \mathrm{var}(b) = \frac{1}{4} + \frac{1}{8n-4}$ by Lemma 2.3.6. The conditional expectations may be evaluated as in the proof of Corollary 2.3.7 to yield

$$\mathbb{E}\left[b \mid b \geq \frac{1}{2}\right] = \frac{1}{2} + \frac{1}{(n-1)2^{2n-2}B(n-1,n-1)} = 1 - \mathbb{E}\left[b \mid b \leq \frac{1}{2}\right].$$

Thus the product of conditional expectations is

$$\mathbb{E}\left[b \mid b \geq \frac{1}{2}\right] \cdot \mathbb{E}\left[b \mid b \leq \frac{1}{2}\right] = \frac{1}{4} - \frac{1}{(n-1)^2 2^{4n-4} B(n-1,n-1)^2}.$$

Putting this all together yields

$$\begin{aligned}\mathbb{P}\left\{\lambda(\tilde{P}) \in \mathcal{D}_{\mathrm{Sep}}\right\} &= (4n-2)\left(\frac{1}{8n-4} - \frac{1}{(n-1)^2 2^{4n-4} B(n-1,n-1)^2}\right) \\ &= \frac{1}{2} + \frac{4n-2}{(n-1)^2 2^{4n-4} B(n-1,n-1)^2}. \end{aligned} \tag{41}$$

The asymptotic limit of $\mathbb{P}\left\{\lambda(\tilde{P}) \in \mathcal{D}_{\mathrm{Sep}}\right\}$ as $n \to \infty$ follows from the bounds on the beta function given in Proposition 2.3.8. When $\mathbb{F} = \mathbb{R}$, by (40) we have for $n \geq 2$ that

$$\frac{1}{\sqrt{2}} \exp\left(-\frac{1}{4(n-1)}\right) \leq \mathbb{P}\left\{\lambda(\tilde{P}) \in \mathcal{S}_{\mathrm{Sep}}\right\} \leq \frac{1}{\sqrt{2}} \cdot \exp\left(\frac{3}{8(n-1)}\right),$$

so clearly $\mathbb{P}\left\{\lambda(\tilde{P}) \in \mathcal{S}_{\mathrm{Sep}}\right\} \to \frac{1}{\sqrt{2}}$.

When $\mathbb{F} = \mathbb{C}$, by (41) we have for $n \geq 2$ that

$$\frac{1}{2} + \frac{2n-1}{2\pi(n-1)} \cdot \exp\left(-\frac{1}{3(n-1)}\right) \leq \mathbb{P}\left\{\lambda(\tilde{P}) \in \mathcal{S}_{\mathrm{Sep}}\right\} \leq \frac{1}{2} + \frac{2n-1}{2\pi(n-1)} \exp\left(\frac{1}{12(n-1)}\right),$$

so it follows that $\mathbb{P}\left\{\lambda(\tilde{P}) \in \mathcal{S}_{\text{Sep}}\right\} \to \frac{1}{2} + \frac{1}{\pi}$. □

The bound for $\mathbb{P}\{P \in \mathcal{S}_{X,Y}\}$ in terms of the operator norm distance $\|X - Y\|$ follows by analyzing the distribution of $\lambda(\tilde{P})$.

Proposition 2.5.12. *Let $P \in \mathrm{Proj}_{\mathbb{F}}(n, 2n)$ be uniformly distributed, then for any $X, Y \in \mathrm{Proj}_{\mathbb{F}}(1, 2n)$*

$$\mathbb{P}\{P \in \mathcal{S}_{X,Y}\} \leq \|X - Y\|.$$

Proof. The case when $n = 1$ is simple and was mentioned previously, so we consider here $n \geq 2$. Further, without loss of generality, assume $\mathrm{Ran}(X), \mathrm{Ran}(Y) \subset \mathrm{Ran}(E)$ where E is the orthogonal projection onto $\mathrm{span}\{e_1, e_2\}$. By conditioning, $\mathbb{P}\{P \in \mathcal{S}_{X,Y}\} = \mathbb{E}\left[\mathbb{P}\left\{P \in \mathcal{S}_{X,Y} \mid \lambda(\tilde{P})\right\}\right]$. By the definition of $\mathcal{D}_{\text{Sep}}$ we see that $\mathbb{P}\left\{P \in \mathcal{S}_{X,Y} \mid \lambda(\tilde{P})\right\} = 0$ if $\lambda(\tilde{P}) \in \mathcal{D}_{\text{Sep}}^c$. Hence

$$\mathbb{E}\left[\mathbb{P}\left\{P \in \mathcal{S}_{X,Y} \mid \lambda(\tilde{P})\right\}\right] = \mathbb{E}\left[\mathbb{P}\left\{P \in \mathcal{S}_{X,Y} \mid \lambda(\tilde{P})\right\} \mathbb{1}_{\mathcal{D}_{\text{Sep}}}(\lambda(\tilde{P}))\right].$$

Suppose now that $\lambda(\tilde{P}) \in \mathcal{D}_{\text{Sep}}$, and first consider the case when $\mathbb{F} = \mathbb{R}$. Then $\mathrm{Proj}_{\mathbb{R}}(1, 2)$ can be viewed as $\mathbb{S}^1_{\mathbb{R}}$ with its opposite points identified, and $E(\tilde{P})$ is a (uniformly distributed) random pair of antipodal points in this quotient space. Letting $E_1 = e_1 e_1^*$ and $E_2 = e_2 e_2^*$, we may parameterize $\mathrm{Proj}_{\mathbb{R}}(1, 2)$ by $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ via $\phi \mapsto E_\phi := (\cos(\phi)e_1 + \sin(\phi)e_2)(\cos(\phi)e_1 + \sin(\phi)e_2)^* = \cos^2(\phi)E_1 + \sin^2(\phi)E_2 + \sin(\phi)(\cos(\phi)(e_1 e_2^* + e_2 e_1^*)$. We see that $\mathrm{tr}\left[\tilde{P}E_\phi\right] = \lambda_1 \cos^2(\phi) + \lambda_2 \sin^2(\phi) = \lambda_1 - (\lambda_1 - \lambda_2)\sin^2(\phi)$. Since $\mathrm{tr}\left[\tilde{P}E_0\right] = \lambda_1 > \frac{1}{2}$ and $\mathrm{tr}\left[\tilde{P}E_{\frac{\pi}{2}}\right] = \lambda_2 < \frac{1}{2}$, there exists some $\phi_h \in (0, \frac{\pi}{2})$ such that $\mathrm{tr}\left[\tilde{P}E_{\phi_h}\right] = \mathrm{tr}\left[\tilde{P}E_{-\phi_h}\right] = \frac{1}{2}$. In fact, $\phi_h = \arcsin\left(\sqrt{\frac{\lambda_1 - \frac{1}{2}}{\lambda_1 - \lambda_2}}\right)$. We see that $\mathrm{tr}\left[\tilde{P}E_\phi\right] > \frac{1}{2}$ for $\phi \in (-\phi_h, \phi_h)$, and $\mathrm{tr}\left[\tilde{P}E_\phi\right] < \frac{1}{2}$ for $\phi \in [-\frac{\pi}{2}, -\phi_h) \cup (\phi_h, \frac{\pi}{2}]$. In our quotient space picture, $E_{-\phi_h}$ is the reflection of the point E_{ϕ_h} across the vertical line between E_1 and E_2.

All of this goes to show that $\lambda(\tilde{P})$ determines ϕ_h, which along with the orientation of E_1 determines which rank-one projections in $\mathrm{Ran}(E)$ that P separates. In the quotient space picture, the open arc between E_{ϕ_h} and $E_{-\phi_h}$ containing E_1 represents the rank-one projections with measurements greater than $\frac{1}{2}$, and the other arc represents those with measurements less than $\frac{1}{2}$. Let

$w = \min\{2\phi_h, \pi - 2\phi_h\}$, which is the length of the smallest of these two arcs. If $w \leq \theta$, then $\mathbb{P}\left\{P \in \mathcal{S}_{X,Y} \mid \lambda(\tilde{P})\right\} = \frac{2w}{\pi} \leq \frac{2}{\pi}\theta$. If $w > \theta$, then $\mathbb{P}\left\{P \in \mathcal{S}_{X,Y} \mid \lambda(\tilde{P})\right\} = \frac{2\theta}{\pi}$. So

$$\begin{aligned}\mathbb{E}\left[\mathbb{P}\left\{P \in \mathcal{S}_{X,Y} \mid \lambda(\tilde{P})\right\} \mathbb{1}_{\mathcal{D}_{\text{Sep}}}(\lambda(\tilde{P}))\right] &\leq \mathbb{E}\left[\frac{2\theta}{\pi} \mathbb{1}_{\mathcal{D}_{\text{Sep}}}(\lambda(\tilde{P}))\right] \\ &= \frac{2\theta}{\pi}\mathbb{P}\left\{\lambda(\tilde{P}) \in \mathcal{D}_{\text{Sep}}\right\} \\ &\leq \|X - Y\|.\end{aligned}$$

Next, we consider the case when $\mathbb{F} = \mathbb{C}$, in which case $\text{Proj}_{\mathbb{C}}(1,2)$ can be identified with the Bloch sphere [17]. By rotational invariance, $E(\tilde{P})$ is a pair of (uniformly distributed) antipodal points on the sphere, and $\lambda(\tilde{P})$ determines which pairs of projections are separated by P. If e_1 and e_2 satsify $e_1e_1^* = E_1$ and $e_2e_2^* = E_2$, and $e_{\phi,\psi} := \cos(\frac{\phi}{2})e_1 + e^{i\psi}\sin(\frac{\phi}{2})e_2$ for $\phi \in [0,\pi], \psi \in [0, 2\pi]$, then $E_{\phi,\psi}$ lies on the circle of points in the Bloch sphere at an angle of ϕ from E_1. Moreover, this parameterization shows that $\operatorname{tr}\left[\tilde{P}E_{\phi,\psi_1}\right] = \operatorname{tr}\left[\tilde{P}E_{\phi,\psi_2}\right]$ for all ϕ, ψ_1, and ψ_2. By continuity, there must exist some $\phi_h \in [0,\pi]$ such that $\operatorname{tr}\left[\tilde{P}E_{\phi_h,\psi}\right] = \frac{1}{2}$ for all $\psi \in [0,2\pi]$. In fact, we can calculate $\phi_h = 2\arcsin(\sqrt{\frac{\lambda_1 - \frac{1}{2}}{\lambda_1 - \lambda_2}})$. The open spherical cap centered at E_1 of angle ϕ_h consists exactly of those projections $A \in \text{Proj}_{\mathbb{C}}(1,2)$ such that $\operatorname{tr}\left[\tilde{P}A\right] > \frac{1}{2}$, and the complimentary cap consists of those for which $\operatorname{tr}\left[\tilde{P}A\right] < \frac{1}{2}$. See Figure 2.5.2 for an illustration.

Conditioning on $\lambda(\tilde{P})$ determines the opening angles of these two spherical caps, which are oriented along a random diameter determined by $E(\tilde{P})$. The projections X, Y are two fixed points on the Bloch sphere at an angle of 2θ, and are separated by P if and only if they are not in the same cap. Let $w = \min\{\phi_h, \pi - \phi_h\}$, which is the smallest opening angle of these two caps. If $w \leq \theta$, then any cap of angle w containing X cannot contain Y (and vice versa), so $\mathbb{P}\left\{P \in \mathcal{S}_{X,Y} \mid \lambda(\tilde{P})\right\}$ is just twice the normalized area of a cap of angle w (which is just its normalized height), i.e.,

$$\mathbb{P}\left\{P \in \mathcal{S}_{X,Y} \mid \lambda(\tilde{P})\right\} = 1 - \cos(w) \leq 1 - \cos(\theta) \leq \sin(\theta) = \|X - Y\|.$$

If $w > \theta$, then it is possible for both X and Y to be in a cap of opening angle w. In this case,

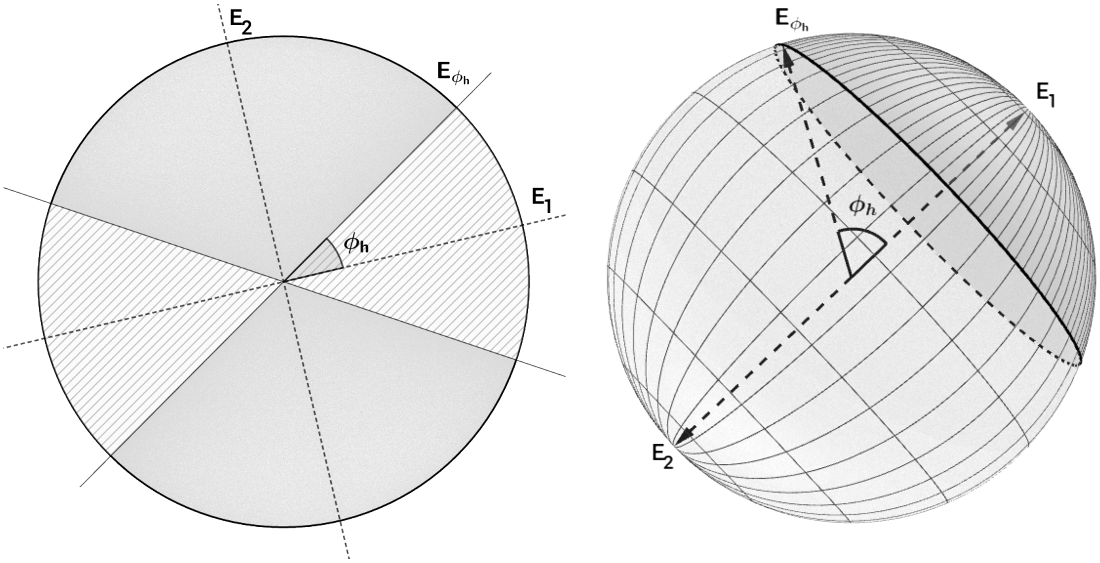

Figure 2: The submatrix $\tilde{P}$ divides $\mathrm{Proj}_{\mathbb{F}}(1,2)$ into two disjoint sets based on whether the Hilbert-Schmidt inner product with $\tilde{P}$ is greater or less than $\frac{1}{2}$ (Left: $\mathbb{F} = \mathbb{R}$; Right: $\mathbb{F} = \mathbb{C}$).

$\mathbb{P}\left\{P \in \mathcal{S}_{X,Y} \mid \lambda(\tilde{P})\right\}$ is just the normalized area of the symmetric difference of spherical caps of angle w centered at X and Y. The intersection of these two caps contains a spherical cap of angle $w - \theta$ centered at the geodesic midpoint of X and Y, so for this case

$$\mathbb{P}\left\{P \in \mathcal{S}_{X,Y} \mid \lambda(\tilde{P})\right\} \leq \cos(w - \theta) - \cos(w) \leq \sin(\theta) = \|X - Y\|.$$

where the last inequality follows since $w \leq \frac{\pi}{2}$. Thus we have

$$\begin{aligned}\mathbb{E}\left[\mathbb{P}\left\{P \in \mathcal{S}_{X,Y} \mid \lambda(\tilde{P})\right\} \mathbb{1}_{\mathcal{D}_{\mathrm{Sep}}}(\lambda(\tilde{P}))\right] &\leq \|X - Y\| \, \mathbb{P}\left\{\lambda(\tilde{P}) \in \mathcal{D}_{\mathrm{Sep}}\right\} \\ &\leq \|X - Y\|.\end{aligned}$$

□

A uniform bound for the measurement Hamming distance in terms of the operator norm distance now follows directly by combining Theorem 2.5.9 with Proposition 2.5.12.

Corollary 2.5.13. *Let $\delta > 0$ be a desired level of uniform concentration and $0 < \rho < 1$ be an* *acceptable failure probability. If*

$$m \geq \frac{1}{2}\delta^{-2}\left(8\beta_{\mathbb{F}}n\log\left(96\sqrt{\frac{\beta_{\mathbb{F}}n}{\pi}}\delta^{-1}\right)+\log(2\rho^{-1})\right),$$

and $\mathcal{P} = \{P_j\}_{j=1}^m$ *is an independent sequence of uniformly distributed projections in* $\mathrm{Proj}_{\mathbb{F}}(n, 2n)$*, then with probability at least* $1-\rho$

$$d_{\mathcal{P}}(X, Y) \leq \|X - Y\| + \delta$$

for all $X, Y \in \mathrm{Proj}_{\mathbb{F}}(1, 2n)$*.*

Uniform guarantee for accurate recovery

Uniformly accurate one-bit phase retrieval follows from the uniform concentration of the measurement Hamming distance given by Theorem 2.5.9 and the uniform bound in terms of operator norm from Corollary 2.5.13.

Theorem 2.5.14. *Let* $0 < \delta < 1$ *be a desired level of uniform accuracy,* $0 < \rho < 1$ *be an acceptable failure probability, and set* $\epsilon = \frac{(\mu_1 - \mu_2)}{8}\delta$*. If*

$$m \geq 2\epsilon^{-2}\left(8\beta_{\mathbb{F}}n\log\left(96\sqrt{\frac{\beta_{\mathbb{F}}n}{\pi}}\epsilon^{-1}\right)+\log(4\rho^{-1})\right) \tag{42}$$

and $\mathcal{P} = \{P_j\}_{j=1}^m$ *is an independent sequence of uniformly distributed projections in* $\mathrm{Proj}_{\mathbb{F}}(n, 2n)$*, then with probability at least* $1-\rho$

$$\left\|\hat{X} - X\right\| < \delta$$

for all $X \in \mathrm{Proj}(1, 2n)$*, where* $\hat{X}$ *is the solution to PEP with input* $\Phi_{\mathcal{P}}(X)$*.*

Proof. Let $\mathcal{N}_\epsilon$ be an ϵ-net for $\mathrm{Proj}_{\mathbb{F}}(1, 2n)$ such that $\log|\mathcal{N}_\epsilon| \leq 4\beta_{\mathbb{F}}n\log(3\epsilon^{-1})$ as in Lemma 2.5.2. By our choice of m, Lemma 2.5.3 says that with probability greater than $1 - \frac{\rho}{2}$ we have for all all $X \in \mathcal{N}_\epsilon$ that $\left\|\hat{Q}_{\mathcal{P}}(X) - Q(X)\right\| \leq \epsilon$ (call this event $\mathcal{A}$). Also by our choice of m, Corollary 2.5.13

says that with probability at least $1-\frac{\rho}{2}$ we have $d_{\mathcal{P}}(X,Y) \le \|X-Y\| + \epsilon$ for all $X, Y \in \mathrm{Proj}_{\mathbb{F}}(1,2n)$ (call this event $\mathcal{B}$).

Suppose that $\mathcal{A}$ and $\mathcal{B}$ both occur, which happens with probability at least $1-\rho$ via a union bound, and consider an arbitrary $X \in \mathrm{Proj}_{\mathbb{F}}(1,2n)$. We know from (27) in the proof of Theorem 2.4.4 that

$$\left\|\hat{X} - X\right\| \le 2(\mu_1 - \mu_2)^{-1} \left\|\hat{Q}_{\mathcal{P}}(X) - Q(X)\right\|. \tag{43}$$

To bound the right-hand side of this last inequality we pass to the ϵ-net $\mathcal{N}_\epsilon$ by picking $X_0 \in \mathcal{N}_\epsilon$ with $\|X - X_0\| < \epsilon$. Then by the triangle inequality

$$\left\|\hat{Q}_{\mathcal{P}}(X) - Q(X)\right\| \le \left\|\hat{Q}_{\mathcal{P}}(X) - \hat{Q}_{\mathcal{P}}(X_0)\right\| + \left\|\hat{Q}_{\mathcal{P}}(X_0) - Q(X_0)\right\| + \|Q(X_0) - Q(X)\|. \tag{44}$$

Next, we examine each of the three terms on the right side of (44). To bound the first term, note that $\left|\left\{j : \hat{P}_j(X) \neq \hat{P}_j(X_0)\right\}\right| = m \cdot d_{\mathcal{P}}(X, X_0)$. By this identity and the triangle inequality, we have

$$\begin{aligned}
\left\|\hat{Q}_{\mathcal{P}}(X) - \hat{Q}_{\mathcal{P}}(X_0)\right\| &= \left\|\frac{1}{m} \sum_{j:\hat{P}_j(X)\neq\hat{P}_j(X_0)} \hat{P}_j(X) - \hat{P}_j(X_0)\right\| \\
&\le \frac{1}{m} \sum_{j:\hat{P}_j(X)\neq\hat{P}_j(X_0)} \left\|\hat{P}_j(X) - \hat{P}_j(X_0)\right\| \\
&= d_{\mathcal{P}}(X, X_0),
\end{aligned}$$

and since $\mathcal{A}$ holds we have

$$d_{\mathcal{P}}(X, X_0) \le \|X - X_0\| + \epsilon < 2\epsilon.$$

Since $\mathcal{B}$ holds and $X_0 \in \mathcal{N}_\epsilon$, we can bound the second term of (44) by $\left\|\hat{Q}_{\mathcal{P}}(X_0) - Q(X_0)\right\| \le \epsilon$. Lastly, Corollary 2.3.7 gives $Q(X) - Q(X_0) = (\mu_1 - \mu_2)(X - X_0)$, and so we can bound the third term by

$$\|Q(X_0) - Q(X)\| = (\mu_1 - \mu_2)\|X - X_0\| \le (\mu_1 - \mu_2)\epsilon.$$

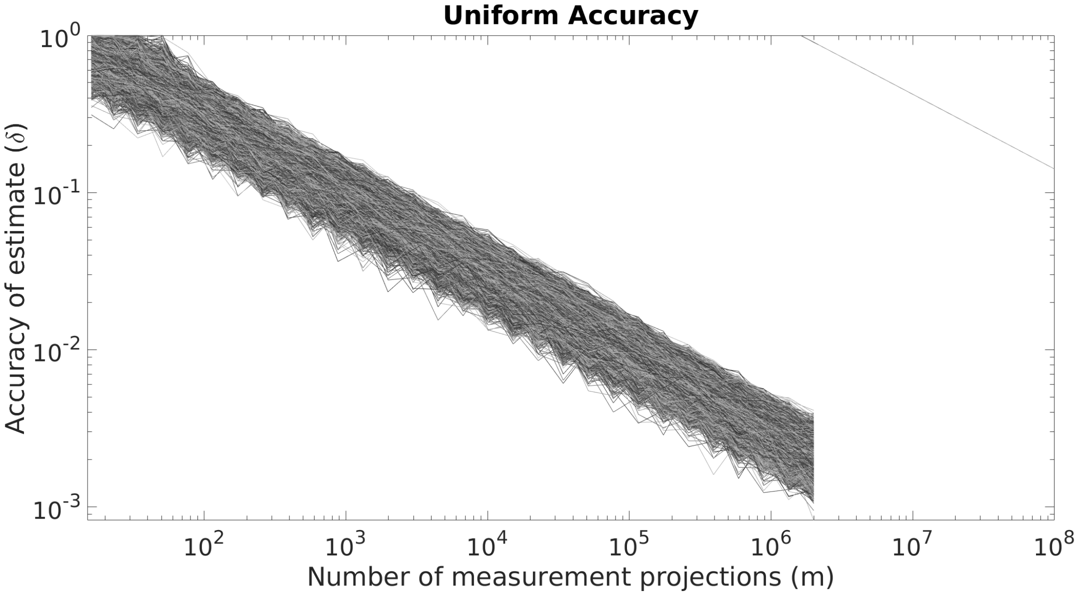

Figure 3: Plot showing the empirical accuracy of recovery versus the theoretical error bound from Theorem 2.5.14 for 15000 random vectors in $\mathbb{R}^{16}$ using PEP with a fixed collection of $2 \cdot 10^6$ measurement projections.

Using these three bounds together in (44) along with our choice of ϵ gives

$$\left\|\hat{Q}_{\mathcal{P}}(X) - Q(X)\right\| \leq 3\epsilon + (\mu_1 - \mu_2)\epsilon \leq \frac{1}{2}(\mu_1 - \mu_2)\delta,$$

which substituted into (43) yields $\left\|\hat{X} - X\right\| < \delta$. □

See Figure 3 for a plot showing how our bound on the sufficient number of measurements to achieve a uniform accuracy of δ relates to experimental results. The single line separate from the cluster represents the upper bound on δ given by Theorem 2.5.14. The MATLAB code used to generate this plot is included in Appendix A.2.

As in the pointwise case, Theorem 2.5.14 allows control over the probability of failure by adjusting the value of ρ in (42). In particular, letting $\rho = \exp(-2n)$ ensures that the generated projections allow uniformly accurate one-bit phase retrieval with overwhelming probability with respect to the dimension of the input signals, i.e., the failure rate decays exponentially with respect to $2n$. In the pointwise case, this resulted in gaining an additional factor of n in the number of measurement projections, see Corollary 2.4.8. In the uniform case, however, the asymptotics remain the same. This is stated as a corollary, and follows by the fact that $(\mu_1 - \mu_2)^{-1} = O(\sqrt{n})$ from Lemma 2.3.9.

Corollary 2.5.15. *Let $0 < \delta < 1$ be a desired level of uniform accuracy. There exists a constant C such that if*

$$m \geq C\delta^{-2}n^2 \log(\delta^{-1}n)$$

and $\mathcal{P} = \{P_j\}_{j=1}^m$ is an independent sequence of uniformly distributed projections in $\mathrm{Proj}_{\mathbb{F}}(n, 2n)$, then with probability at least $1 - \exp(-n)$

$$\left\|\hat{X} - X\right\| < \delta$$

for all $X \in \mathrm{Proj}(1, 2n)$, where $\hat{X}$ is the solution to PEP with input $\Phi_{\mathcal{P}}(X)$.

Chapter 3

Noisy One-Bit Phase Retrieval

Chapter 2 provided solutions to Problem 1.3.6 and Problem 1.3.7, the fixed input and uniform one-bit phase retrieval problems with noiseless measurements, in Theorem 2.4.4 and Theorem 2.5.14 respectively. In Section 1.3 we also stated versions of these problems in the presence of bit-flip errors that result in a noisy phaseless binary measurement. Problem 1.3.8 states the fixed input one-bit phase retrieval problem in the presence of adversarial noise, and Problem 1.2.7 states the analogous problem for uniformly accurate recovery with adversarial noise. This chapter provides solutions to both of these problems as corollaries to Theorem 2.4.4 and Theorem 2.5.14 respectively, see Section 3.2 below. These error bounds for adversarial noise were included in [38] in the case of a complementary magnitude comparison measurement associated to half-dimensioned projections.

Section 1.5 provided motivation from rate-distortion theory for looking at a second noise model: random bit-flips where each bit in the phaseless binary measurement is flipped independently with some fixed probability τ. Using the mean squared error as a measure of distortion by the measurement and reconstruction scheme, Section 3.3 provides a solution to Problem 1.5.1 below.

3.1 Reconstruction from noisy measurements

Both the adversarial noise model discussed in Section 1.3 and the random bit-flip noise model discussed in Section 1.5 are formulated in terms of a bit-flip map $\mathcal{F}_T$ which flips all entries of the phaseless binary measurement $\Phi_{\mathcal{P}}$ which have indices in $T \subset \{1, \ldots, m\}$. Once an input signal X is measured to obtain the noisy phaseless binary measurement $\mathcal{F}_T(\Phi_{\mathcal{P}}(X))$, PEP can be used as defined in Section 2.2 to estimate X. The recovery constructs an auxiliary matrix analogous to the empirical average $\hat{Q}_{\mathcal{P}}(X)$ and then finds its principal eigenspace.

Definition 3.1.1. *Given an input signal $X \in \mathrm{Proj}_{\mathbb{F}}(1, d)$, a magnitude comparison measurement $\Phi_{\mathcal{P}}$ associated to orthogonal projections $\mathcal{P} = \{P_j\}_{j=1}^{2m}$, and a bit-flip subset $T \subset \{1, \dots, m\}$, define the* ***bit-flipped empirical average of the proximal projections*** *to be*

$$\hat{Q}_{\mathcal{P},T}(X) := \frac{1}{m}\left[\sum_{j \in T}[\Phi_{\mathcal{P}}(X)_j P_{m+j} + (1 - \Phi_{\mathcal{P}}(X)_j)P_j] + \sum_{j \in T^c} \hat{P}_j(X)\right].$$

Observe that $\hat{Q}_{\mathcal{P},T}(X)$ is just $\hat{Q}_{\mathcal{P}}(X)$ with the summands with indices in $j \in T$ flipped from P_j to P_{m+j} and vice versa. If $\Phi_{\mathcal{P}}$ is the complementary magnitude comparison measurement associated to orthogonal projections $\mathcal{P} = \{P_j\}_{j=1}^{m}$, then

$$\hat{Q}_{\mathcal{P},T}(X) = \frac{1}{m}\left[\sum_{j \in T}(I - \hat{P}_j(X)) + \sum_{j \in T^c} \hat{P}_j(X)\right]$$

With the bit-flipped empirical average $\hat{Q}_{\mathcal{P},T}(X)$, the noisy reconstruction algorithm $\mathcal{R}$ is defined as in Definition 2.2.2 by setting $\hat{X}_T := \mathcal{R}(\mathcal{F}_T(\Phi_{\mathcal{P}}(X)))$ to be the maximizer of

$$\begin{aligned} &\underset{Y}{\text{maximize}} && \mathrm{tr}\left[\hat{Q}_{\mathcal{P},T}(X)Y\right] \\ &\text{subject to} && Y \succeq 0, \mathrm{tr}\,[Y] \leq 1. \end{aligned} \tag{45}$$

This is the same optimization problem as in the noiseless case, so it is still referred to as Principal Eigenspace Programming, or PEP. The maximizer $\hat{X}_T$ of (45) is the rank-one orthogonal projection onto the principal eigenspace of the bit-flipped empirical average $\hat{Q}_{\mathcal{P},T}(X)$.

3.2 Adversarial noise

With high probability, the phaseless binary measurements $\Phi_{\mathcal{P}}$ studied in Theorem 2.4.4 and the recovery algorithm PEP defined in Section 3.1 are robust to adversarial bit-flip errors in the binary measurement of a fixed input signal. This can be seen via a small addition to the proof of

Theorem 2.4.4.

Corollary 3.2.1. *Let $X \in \operatorname{Proj}_{\mathbb{F}}(1,d)$ be fixed and $\Phi_{\mathcal{P}}$ be a magnitude comparison measurement as in Proposition 2.3.2 or a complementary magnitude comparison measurement as in Proposition 2.3.4. Then with probability at least*

$$1 - 2d \exp\left(- \frac{(\mu_1 - \mu_2)^2 \delta^2 m}{8 \max(\mu_1 - \mu_1^2, \mu_2 - \mu_2^2) + \frac{8}{3} \max(\mu_1, 1 - \mu_2)(\mu_1 - \mu_2)\delta} \right)$$

we have

$$\left\| \hat{X}_T - X \right\| \leq \delta + 2(\mu_1 - \mu_2)^{-1} \frac{|T|}{m}$$

for all $T \subset \{1, \ldots, m\}$, where $\hat{X}_T$ is the projection onto the principal eigenspace of $\hat{Q}_{\mathcal{P},T}(X)$ and μ_1, μ_2 are the eigenvalues of $Q(X)$.

Proof. By Lemma 2.4.3, we have

$$\left\| \hat{X}_T - X \right\| = (\mu_1 - \mu_2)^{-1} \operatorname{tr}\left[Q(X)(A - B) \right],$$

where $X - \hat{X}_T = \left\| X - \hat{X}_T \right\| (A - B)$ is the spectral decomposition given by Lemma 1.3.5. Since $\hat{X}_T$ is the projection onto the principal eigenspace of $\hat{Q}_{\mathcal{P},T}(X)$, we have

$$\operatorname{tr}\left[\hat{Q}_{\mathcal{P},T}(X)(\hat{X}_T - X) \right] \geq 0 \implies (\mu_1 - \mu_1)^{-1} \operatorname{tr}\left[\hat{Q}_{\mathcal{P},T}(X)(B - A) \right] \geq 0.$$

Thus we have an upper bound for $\left\| \hat{X}_T - X \right\|$ given by

$$\begin{aligned} \left\| \hat{X}_T - X \right\| &\leq (\mu_1 - \mu_2)^{-1} \operatorname{tr}\left[(Q(X) - \hat{Q}_{\mathcal{P},T}(X))(A - B) \right] && (46) \\ &\leq 2(\mu_1 - \mu_2)^{-1} \left\| \hat{Q}_{\mathcal{P},T}(X) - Q(X) \right\|. && (47) \end{aligned}$$

By the triangle inequality, it follows that

$$\left\| \hat{Q}_{\mathcal{P},T}(X) - Q(X) \right\| \leq \left\| \hat{Q}_{\mathcal{P},T}(X) - \hat{Q}_{\mathcal{P}}(X) \right\| + \left\| \hat{Q}_{\mathcal{P}}(X) - Q(X) \right\|.$$

The normalized Hamming distance between $\Phi_{\mathcal{P}}(X)$ and $\mathcal{F}_T(\Phi_{\mathcal{P}}(X))$ is equal to $\frac{|T|}{m}$, which implies

$$\begin{aligned} \left\| \hat{Q}_{\mathcal{P},T}(X) - \hat{Q}_{\mathcal{P}}(X) \right\| &= \left\| \frac{1}{m} \sum_{j \in T} (2\Phi_{\mathcal{P}}(X)_j - 1)(P_{m+j} - P_j) \right\| \\ &\leq \frac{1}{m} \sum_{j \in T} \|(2\Phi_{\mathcal{P}}(X)_j - 1)(P_{m+j} - P_j)\| \\ &\leq \frac{|T|}{m}. \end{aligned}$$

The result follows by using Lemma 2.4.1 to bound the probability that $\left\| \hat{Q}_{\mathcal{P}}(X) - Q(X) \right\| \geq \frac{1}{2}(\mu_1 - \mu_2)\delta$. □

Corollary 3.2.1 shows that recovery via PEP from these types of phaseless binary measurements provide one-bit phase retrieval of a fixed input signal in the presence of adversarial bit-flips, solving Problem 1.3.8. Corollaries 2.4.5, 2.4.6, and 2.4.7 may be used to show how the number of bits required for the noisy measurement depends on the dimension, desired accuracy, and probability of failure, for three specific phaseless binary measurements. In the notation of Problem 1.3.8, by restricting to bit-flip sets of size at most τm then Corollary 3.2.1 gives an adversarial bit-flip error term that depends on τ of the form $r(\tau) = 2(\mu_1 - \mu_2)^{-1}\tau$. The spectral gap $\mu_1 - \mu_2$ depends on the type of phaseless binary measurement used and the ranks of the random projections. As shown in Corollary 2.3.3 and Corollary 2.3.5, the maximal spectral gap for the measurements we considered occurs when either independent or complementary pairs of half-dimensioned projections are used for a magnitude comparison measurement. In these cases, the spectral gap is $\mu_1 - \mu_2 = O(\frac{1}{\sqrt{d}})$, which implies that a ratio of $\tau = O(\frac{1}{\sqrt{d}})$ adversarial bit-flips can occur and PEP will still recover X accurately. For a magnitude comparison measurement with independent pairs of rank-one projections, the spectral gap is $\mu_1 - \mu_2 = O(\frac{1}{d})$, meaning only a ratio $\tau = O(\frac{1}{d})$ bit-flips can be corrected to ensure accurate recovery.

As in the pointwise case, the proof of Theorem 2.5.14 can be modified to show that uniform recovery using PEP is also robust to adversarial bit-flip errors occurring in a noisy measurement.

Recall that Theorem 2.5.14 worked with a specific phaseless binary measurement: the complementary magnitude comparison measurement associated to a collection of half-dimensioned projections.

Corollary 3.2.2. *Let δ, ρ, m, $\{P_j\}_{j=1}^m$, and $\Phi_{\mathcal{P}}$ be as in Theorem 2.5.14. Then with probability at least $1-\rho$, for all $X \in \mathrm{Proj}_{\mathbb{F}}(1,2n)$*

$$\left\|\hat{X}_T - X\right\| \leq \delta + 2(\mu_1-\mu_2)^{-1}\frac{|T|}{m},$$

for all $T \subset \{1,\ldots,m\}$, where $\hat{X}_T$ is the output of PEP applied to the bit-flipped empirical average $\hat{Q}_{\mathcal{P},T}(X)$ and μ_1,μ_2 are the eigenvalues of $Q(X)$.

Proof. Following the same steps as in Corollary 3.2.1, we arrive at (46) which says

$$\left\|\hat{X}_T - X\right\| \leq 2(\mu_1-\mu_2)^{-1}\left\|\hat{Q}_{\mathcal{P},T}(X) - Q(X)\right\|.$$

Using the triangle inequality, we expand

$$\begin{aligned}\left\|\hat{Q}_{\mathcal{P},T}(X) - Q(X)\right\| &\leq \left\|\hat{Q}_{\mathcal{P},T}(X) - \hat{Q}_{\mathcal{P}}(X)\right\| + \left\|\hat{Q}_{\mathcal{P}}(X) - Q(X)\right\| \\ &\leq \tau + \left\|\hat{Q}_{\mathcal{P}}(X) - Q(X)\right\|.\end{aligned}$$

Bounding the probability that $\left\|\hat{Q}_{\mathcal{P}}(X) - Q(X)\right\| < \frac{1}{2}(\mu_1-\mu_2)\delta$ for all X with high probability proceeds exactly as in Theorem 2.5.14. □

Corollary 3.2.2 shows that our measurement and reconstruction scheme provides uniform one-bit phase retrieval in the presence of adversarial bit-flips, solving Problem 1.3.9. If a maximum of τm bit-flips are allowed, then Corollary 3.2.2 gives an adversarial bit-flip error term that depends on τ of the form $r(\tau) = 2(\mu_1-\mu_2)^{-1}\tau$ as in the pointwise case. This implies that $(\mu_1-\mu_2) = O(\frac{1}{\sqrt{n}})$ adversarial bit-flips may be allowed in the phaseless binary measurement of any signal and PEP will still provide accurate reconstruction.

3.3 Random independent bit-flips

This section addresses the noise model outlined in Section 1.5, where each bit in the phaseless binary measurement is flipped independently with a fixed probability τ. The mean squared error is computed for the measurement and reconstruction scheme that uses the complementary magnitude comparison measurement associated to a random collection of half-dimensioned projections as defined in Definition 2.1.6 and the noisy reconstruction algorithm given by PEP in Section 3.1.

First, the expectation of the bit-flipped empirical average $\hat{Q}_{\mathcal{P},T}(X)$ is computed. Its spectral decomposition in a similar form as that of $Q(X) = \mathbb{E}\left[\hat{Q}_{\mathcal{P}}(X)\right]$ given in Corollary 2.3.7, but the eigenvalues and spectral gap depend on the bit-flip probability τ in addition to the dimension of the input signals.

Proposition 3.3.1. *Let $X \in \mathrm{Proj}_{\mathbb{F}}(1, 2n)$ be a fixed input signal and $0 \leq \tau < \frac{1}{2}$ be the probability of a bit-flip. If $\mathcal{P} = \{P_j\}_{j=1}^m$ is an independent sequence of uniformly distributed projections in $\mathrm{Proj}_{\mathbb{F}}(n, 2n)$ and $T \subset \{1, \ldots, m\}$ is binomially distributed with probability τ and independent of $\mathcal{P}$, then*

$$Q_T(X) := \mathbb{E}\left[\hat{Q}_{\mathcal{P},T}(X)\right] = \gamma_1 X + \gamma_2 (I - X),$$

where

$$\gamma_1 = (1 - 2\tau)\mu_1 + \tau, \qquad \gamma_2 = (1 - 2\tau)\mu_2 + \tau.$$

Here, μ_1, μ_2 are the eigenvalues of $Q(X)$ given in Corollary 2.3.7. In particular, the spectral gap of $Q_T(X)$ is

$$\gamma_1 - \gamma_2 = (1 - 2\tau)(\mu_1 - \mu_2).$$

Proof. By definition of the bit-flipped empirical average, independence of $\mathcal{P}$ and T, and linearity of expectation, we have

$$\begin{aligned}\mathbb{E}\left[\hat{Q}_{\mathcal{P},T}(X)\right] &= \mathbb{E}_T\left[\mathbb{E}_{\mathcal{P}}\left[\frac{1}{m}\left[\sum_{j\in T}(I-\hat{P}_j(X))+\sum_{j\in T^c}\hat{P}_j(X)\right]\,\middle|\, T\right]\right]\\ &= \mathbb{E}_T\left[\frac{1}{m}\left[\sum_{j\in T}\left(I-\mathbb{E}_{\mathcal{P}}\left[\hat{P}_j(X)\right]\right)+\sum_{j\in T^c}\mathbb{E}_{\mathcal{P}}\left[\hat{P}_j(X)\right]\right]\right].\end{aligned}$$

Since the projections P_j are i.i.d., and $\mathbb{E}\left[\hat{P}_j\right] = Q(X)$ by definition, it follows that

$$\frac{1}{m}\left[\sum_{j\in T}\left(I-\mathbb{E}\left[\hat{P}_j(X)\right]\right)+\sum_{j\in T^c}\mathbb{E}\left[\hat{P}_j(X)\right]\right] = \frac{|T|}{m}(I-Q(X))+\frac{m-|T|}{m}Q(X).$$

Since $\mathbb{E}\left[|T|\right] = \tau m$ by definition of the binomial distribution with probability τ, we have

$$\begin{aligned}\mathbb{E}\left[\hat{Q}_{\mathcal{P},T}(X)\right] &= \mathbb{E}_T\left[\frac{|T|}{m}(I-Q(X))+\frac{m-|T|}{m}Q(X)\right]\\ &= \tau(I-Q(X))+(1-\tau)Q(X)\\ &= (1-2\tau)Q(X)+\tau I.\end{aligned}$$

The spectral decomposition of $Q_T(X)$ then follows by using the spectral decomposition of $Q(X)$ given in Corollary 2.3.7. □

The next lemma computes the concentration of the bit-flipped empirical average around its expectation.

Lemma 3.3.2. *Let $X \in \mathrm{Proj}_{\mathbb{F}}(1,2n)$ be a fixed input signal and $0 < \tau < \frac{1}{2}$ be the probability of a bit-flip. If $\mathcal{P} = \{P_j\}_{j=1}^m$ is an independent sequence of uniformly distributed projections in $\mathrm{Proj}_{\mathbb{F}}(n,2n)$ and $T \subset \{1,\ldots,m\}$ is binomially distributed with probability τ and independent of $\mathcal{P}$, then*

$$\mathbb{E}\left[\left\|\hat{Q}_{\mathcal{P},T}(X)-Q_T(X)\right\|\right] \le \sqrt{\frac{2\log(4n)}{m}}+\frac{\log(4n)}{3m}$$

and

$$\mathbb{P}\left\{\left\|\hat{Q}_{\mathcal{P},T}(X) - Q_T(X)\right\| \geq t\right\} \leq 4n \exp\left(-\frac{mt^2}{3}\right)$$

Proof. Define random matrices S_j depending on the random projections $\mathcal{P}$ and the random subset T by

$$S_j = \begin{cases} \frac{1}{m}((I - \hat{P}_j(X)) - Q_T(X)) & \text{if } j \in T \\ \frac{1}{m}(\hat{P}_j(X) - Q_T(X)) & \text{else.} \end{cases}$$

Then the S_j's are independent by the independence of $\mathcal{P}$ and T, and $Z := \sum_{j=1}^m S_j = \hat{Q}_{\mathcal{P},T}(X) - Q_T(X)$. Also, the S_j's are mean-zero since

$$\begin{aligned}\mathbb{E}\left[S_j\right] &= \frac{1}{m}\mathbb{E}_{\mathcal{P}}\left[\tau((I - \hat{P}_j(X)) - Q_T(X)) + (1-\tau)(\hat{P}_j(X) - Q_T(X))\right] \\ &= \frac{1}{m}\left[(1-2\tau)Q(X) + \tau I - Q_T(X)\right] \\ &= 0,\end{aligned}$$

where the last equality follows from the spectral decomposition for $Q_T(X)$ given in Proposition 3.3.1. Using the triangle inequality, we may obtain the bound $\mathbb{E}\left[\|S_j\|\right] \leq \frac{1}{m}$, and we can bound the matrix variance of the sum by observing

$$\left\|S_j^2\right\| = \|S_j\|^2 \leq \frac{1}{m^2},$$

and hence by the fact that the S_j's are i.i.d. and Jensen's inequality we have

$$\operatorname{var}(Z) = \left\|\sum_{j=1}^m \mathbb{E}\left[S_j^2\right]\right\| = m\left\|\mathbb{E}\left[S_j^2\right]\right\| \leq m\mathbb{E}\left[\left\|S_j^2\right\|\right] \leq \frac{1}{m}.$$

The expectation and probability bounds then follow from applying the matrix Bernstein inequality, Theorem 1.4.15. □

Now the mean squared error of the noisy measurement and recovery scheme may be computed.

Theorem 3.3.3. *Let $0 \leq \tau < \frac{1}{2}$ be a bit-flip probability and assume*

$$m \geq 24(1-2\tau)^{-2}(\mu_1 - \mu_2)^{-2}\log(4n).$$

If $X \in \mathrm{Proj}_{\mathbb{F}}(1, 2n)$ be uniformly distributed, $\mathcal{P} = \{P_j\}_{j=1}^m$ is a sequence of uniformly distributed orthogonal projections in $\mathrm{Proj}_{\mathbb{F}}(n, 2n)$, and $T \subset \{1, \ldots, m\}$ is binomially distributed with probability τ, then

$$\mathbb{E}_{X,T,\mathcal{P}}\left[\left\|\hat{X}_T - X\right\|^2\right] \leq \frac{12\log\left(\frac{e}{3}mn(1-2\tau)^2(\mu_1-\mu_2)^2\right)}{m(1-2\tau)^2(\mu_1-\mu_2)^2}$$

where $\hat{X}_T$ is the solution to PEP applied to the bit-flipped empirical average $\hat{Q}_{\mathcal{P},T}(X)$.

Proof. By the same argument as in Lemma 2.4.3, we have

$$\left\|\hat{X}_T - X\right\| = (\gamma_1 - \gamma_2)^{-1}\operatorname{tr}\left[Q_T(X)(A-B)\right],$$

where $X - \hat{X}_T = \left\|\hat{X}_T - X\right\|(A - B)$ is the spectral decomposition given by Lemma 1.3.5. Since $\hat{X}_T$ is the projection onto the principal eigenspace of $\hat{Q}_{\mathcal{P},T}(X)$, for any $X, \mathcal{P}$, and T we have

$$\begin{aligned}\operatorname{tr}\left[\hat{Q}_{\mathcal{P},T}(X)(\hat{X}_T - X)\right] \geq 0 &\implies \operatorname{tr}\left[\hat{Q}_{\mathcal{P},T}(X)(A-B)\right] \geq 0 \\ &\implies (\gamma_1-\gamma_2)^{-1}\operatorname{tr}\left[\hat{Q}_{\mathcal{P},T}(X)(A-B)\right] \geq 0.\end{aligned}$$

Thus we have an upper bound that holds for all $X, \mathcal{P}$, and T given by

$$\left\|\hat{X}_T - X\right\| \leq 2(\gamma_1-\gamma_2)^{-1}\left\|\hat{Q}_{\mathcal{P},T}(X) - Q_T(X)\right\|. \tag{48}$$

Now we begin to compute the mean squared error. First, observe by conditioning on X that

$$\mathbb{E}_{X,T,\mathcal{P}}\left[\left\|\hat{X}_T - X\right\|^2\right] = \mathbb{E}_X\left[\mathbb{E}_{T,\mathcal{P}}\left[\left\|\hat{X}_T - X\right\|^2 \mid X\right]\right].$$

Given X, let $\mathcal{A}_t$ be the event that $\left\|\hat{X}_T - X\right\| < t$ for some parameter $t \in [0,1]$ to be optimized over later. By conditioning on $\mathcal{A}_t$, we have

$$\mathbb{E}_{T,\mathcal{P}}\left[\left\|\hat{X}_T - X\right\|^2\right] \leq t^2 + \mathbb{P}\left\{\mathcal{A}_t^c\right\}.$$

From Lemma 3.3.2 and the upper bound in (48) we know that

$$\begin{aligned}\mathbb{P}\left\{\mathcal{A}_t^c\right\} &= \mathbb{P}\left\{\left\|\hat{X}_T - X\right\| \geq t\right\}\\ &\leq \mathbb{P}\left\{2(\gamma_1-\gamma_2)^{-1}\left\|\hat{Q}_{\mathcal{P},T}(X) - Q_T(X)\right\| \geq t\right\}\\ &\leq \mathbb{P}\left\{\left\|\hat{Q}_{\mathcal{P},T}(X) - Q_T(X)\right\| \geq \frac{1}{2}(\gamma_1-\gamma_2)t\right\}\\ &\leq 4n\exp\left(-\frac{m(\gamma_1-\gamma_2)^2t^2}{12}\right).\end{aligned}$$

This yields for all $t \in [0,1]$ that

$$\mathbb{E}_{T,\mathcal{P}}\left[\left\|\hat{X}_T - X\right\|^2\right] \leq t^2 + 4n\exp\left(-\frac{m(\gamma_1-\gamma_2)^2t^2}{12}\right). \tag{49}$$

Define $f(t) = t^2 + 4n\exp\left(-\frac{m(\gamma_1-\gamma_2)^2t^2}{12}\right)$. We want to minimize f on $[0,1]$ to yield an optimal upper bound. We accomplish this through basic calculus. First, observe that the derivative of f is

$$f'(t) = 2t - 8\alpha tn\exp\left(-\alpha t^2\right),$$

where $\alpha = \frac{1}{12}m(\gamma_1-\gamma_2)^2$. We see that $f'(0) = 0$ and $f'(t) = 0$ for $t \neq 0$ if and only if

$$\begin{aligned}2t - 8\alpha tn\exp\left(-\alpha t^2\right) = 0 &\iff 2t = 8\alpha tn\exp\left(-\alpha t^2\right)\\ &\iff \frac{1}{4\alpha n} = \exp\left(-\alpha t^2\right)\\ &\iff \log(4\alpha n) = \alpha t^2.\end{aligned}$$

By assumption $4\alpha n \geq 1$, so this equation has one positive solution given by

$$t_0 = \sqrt{\frac{\log(4\alpha n)}{\alpha}},$$

and $t_0 \leq 1$ since $\alpha \geq 2\log(4n)$. The same algebraic manipulations shows us that f is decreasing on the interval $\left(0, \sqrt{\frac{\log(4\alpha n)}{\alpha}}\right)$ and increasing on the interval $\left(\sqrt{\frac{\log(4\alpha n)}{\alpha}}, \infty\right)$, so the minimum value of f on $[0,1]$ occurs at $t_0 = \min\left(\sqrt{\frac{\log(4\alpha n)}{\alpha}}, 1\right)$. Plugging this optimal parameter into (49) yields

$$\begin{aligned}\mathbb{E}_{T,\mathcal{P}}\left[\left\|\hat{X}_T - X\right\|^2\right] &\leq t_0^2 + 4n\exp\left(-\frac{m(\gamma_1-\gamma_2)^2 t_0^2}{12}\right)\\ &\leq \frac{\log(4e\alpha n)}{\alpha}.\end{aligned}$$

Since this bound holds for all $X \in \mathrm{Proj}_{\mathbb{F}}(1, 2n)$, it follows that

$$\begin{aligned}\mathbb{E}_X\left[\mathbb{E}_{T,\mathcal{P}}\left[\left\|\hat{X}_T - X\right\|^2 \mid X\right]\right] &\leq \mathbb{E}_X\left[\frac{\log(4e\alpha n)}{\alpha}\right]\\ &= \frac{\log(4e\alpha n)}{\alpha}.\end{aligned}$$

Substituting $\alpha = \frac{1}{12}m(\gamma_1-\gamma_2)^2$ and using the fact that $\gamma_1 - \gamma_2 = (1-2\tau)(\mu_1-\mu_2)$ by Proposition 3.3.1 gives the desired bound on the mean squared error. □

See Figure 4 for a plot showing how this bound for mean squared error relates to experimental results. The dashed line represents the theoretical mean squared error bound given by Theorem 3.3.3. The MATLAB code used to generate this plot is included in Appendix A.3.

For a fixed bit-flip probability τ and a fixed dimension $2n$, Theorem 3.3.3 says that the mean squared error of this measurement and reconstruction scheme decays on the order of $\frac{\log(m)}{m}$. The following corollary gives a more precise bound on how large m must be in order to achieve a mean squared error of δ.

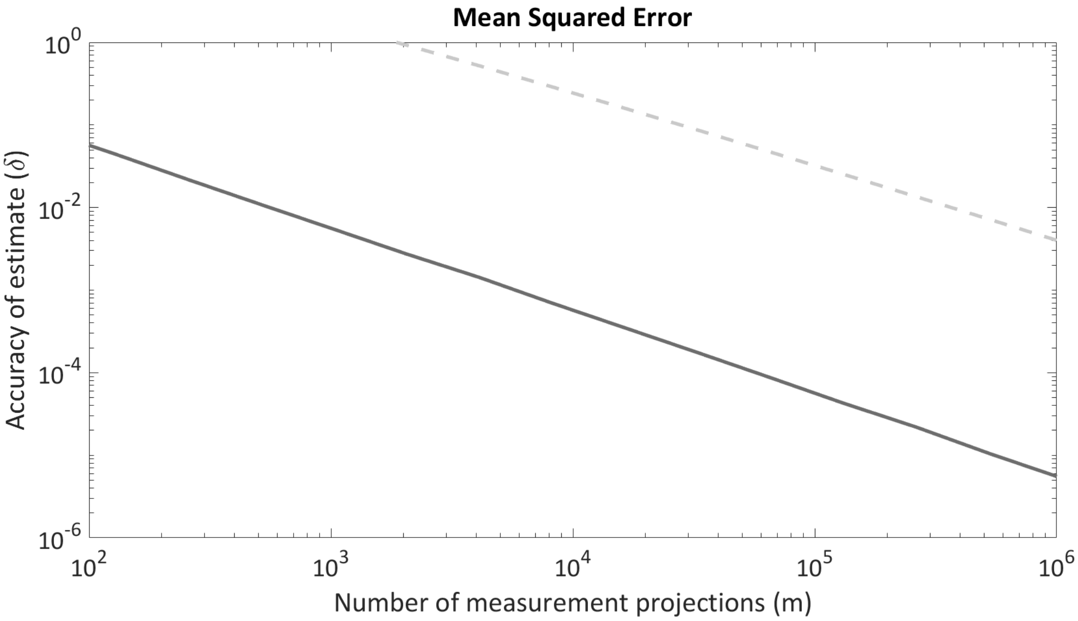

Figure 4: Plot showing the empirical mean squared error of recovery for 1000 random collections of 2^j rank-4 projections on $\mathbb{R}^8$ for $j = 7, \ldots, 20$ versus the theoretical error bound given in Theorem 3.3.3.

Corollary 3.3.4. *Let $\delta > 0$ be a desired mean squared error distortion, $0 \leq \tau < \frac{1}{2}$ be a bit-flip probability, and assume*

$$m \geq 24\delta^{-1}(1 - 2\tau)^{-2}(\mu_2 - \mu_2)^{-2} \log(4en).$$

If $X \in \mathrm{Proj}_{\mathbb{F}}(1, 2n)$ is uniformly distributed, $\mathcal{P} = \{P_j\}_{j=1}^m$ is a sequence of uniformly distributed orthogonal projections in $\mathrm{Proj}_{\mathbb{F}}(n, 2n)$, and $T \subset \{1, \ldots, m\}$ is binomially distributed with probability τ, then

$$\mathbb{E}_{X,T,\mathcal{P}}\left[\left\|\hat{X}_T - X\right\|^2\right] \leq \delta + \delta\frac{\log(2\delta^{-1})}{2\log(4en)},$$

where $\hat{X}_T$ solution to PEP applied to the bit-flipped empirical average $\hat{Q}_{\mathcal{P},T}(X)$.

Proof. By our choice of m, we may apply Theorem 3.3.3 to see

$$\begin{aligned}\frac{12\log\left(\frac{e}{3}mn(1-2\tau)^2(\mu_1-\mu_2)^2\right)}{m(1-2\tau)^2(\mu_1-\mu_2)^2} &\leq \frac{\log(8en\delta^{-1}\log(4en))}{2\delta^{-1}\log(4en)} \\ &= \frac{\log(4en)}{2\delta^{-1}\log(4en)} + \frac{\log(2\delta^{-1}\log(4en))}{2\delta^{-1}\log(4en)} \\ &= \frac{\delta}{2}\left[1 + \frac{\log(\log(4en))}{\log(4en)} + \frac{\log(2\delta^{-1})}{\log(4en)}\right] \\ &\leq \delta + \delta\frac{\log(2\delta^{-1})}{2\log(4en)}.\end{aligned}$$

□

In particular, for $\delta > \frac{1}{n}$ choosing $m \geq C\delta^{-1}(1-2\tau)^{-2}n\log(4en)$ random projections is sufficient to achieve a mean squared error on the order of δ. In other words, a complementary magnitude comparison measurement associated to uniformly distributed half-dimensioned projections requires $\frac{m}{2n} \geq C\delta^{-1}(1-2\tau)^{-2}\log(4en)$ bits-per-dimension for accurate encoding and decoding via PEP of input signals. This shows that the number of bits-per-dimension grows only logarithmically as the dimension grows.

Appendix A

MATLAB Code

A.1 Empirical results - Pointwise error

The following code was used to generate the plot in Figure 1, which shows how the theoretical pointwise error bound from Theorem 2.4.4 relates to empirical results for complementary magnitude comparison measurements associated to half-dimensioned projections.

```
% [n] = half the dimension of the input signal.
n = 8;

% [X] = the input signal to be measured. Without loss of generality, X is
% assumed to be the rank-one projection onto the first standard basis vector.
I = eye(2*n, 2*n);
x = I(:,1);
X = x*x';

% [steps] = total number of steps. Maximum number of projections is
% [step_factor]^[steps].
%
% [step_factor] = number of projections to use for the measurement is
% multiplied by this factor at each step.
%
% [samples] = number of independent collections of projections that are
% generated and used for reconstruction at each step.
steps = 63;
step_factor = 1.2;
samples = 7200;
m_values = ceil(10*1.2.^1:steps);

% [errors] = array that stores the error of reconstruction for each sample
% of each random choice of projections.
errors = zeros(steps, samples);


for it = 1:samples
    % [m] = number of projections to use for measurement and reconstruction
    % in this step.

    % [Q_hat] = the empirical average of proximal projections, for this
    % sample of projections.
    Q_hat = zeros(2*n, 2*n);
    m_total = 0;

    % Compute the empirical average of proximal projections and use it for
    % reconstruction, keeping track of the reconstruction error along the
    % way.
    for j=1:steps
        batch_end = m_values(j);
        if j==1
            batch_start = 1;
        else
            batch_start = m_values(j-1)+1;
        end
        batch_size = batch_end - batch_start;

        Q_hat = m_total/(m_total + batch_size) * Q_hat;

        for k = 1:batch_size
            % [P] = a uniformly distributed rank-[n] projections
            A = randn(2*n, n);
            P = A*(A'*A)^(-1)*A';

            % Compute the phaseless measurement, quantize it, and assemble
            % the empirical average of proximal projections [Q_hat].
            if (trace(P*X) > .5)
                Q_hat = Q_hat + 1/(m_total + batch_size)*P;
            else
                Q_hat = Q_hat + 1/(m_total + batch_size)*(I - P);
            end
        end

        % [X_hat] = the projection onto the principal eigenspace of [Q_hat],
        % which is the estimate for [X] using PEP.
        [x_hat,e] = eigs(Q_hat,1);
        X_hat = x_hat*x_hat';

        % Store the reconstruction error for this step of this sample.
        errors(j, it) = norm(X_hat - X);

        m_total = m_total + batch_size;
    end
end

loglog(m_values'.*ones(steps, 1), errors)
```

A.2 Empirical results - Uniform error

The following code was used to generate the plot in Figure 3, which shows how the theoretical uniform error bound from Theorem 2.5.14 relates to empirical results for complementary magnitude comparison measurements associated to half-dimensioned projections.

```
% [n] = half the dimension of the input signal.
n = 8;
I = eye(2*n, 2*n);

% [steps] = total number of steps. Total number of projections is
% 10*[step_factor]^[steps].
%
% [step_factor] = number of projections to use for the measurement is
% multiplied by this factor at each step.
%
% [input_samples] = number of random input signals used to test the
% reconstruction error of a fixed collection of projections
steps = 30;
step_factor = 1.5021;
input_samples = 15000;
m_values = ceil(10*step_factor.^1:steps);

% [m_max] = total number of projections.
%
% [projs] = random collection of [m_max] independent, uniformly distributed
% rank-[n] projections.
m_max = ceil(10*step_factor^steps);
projs = zeros(2*n, 2*n, m_max);
for j=1:m_max
    A = randn(2*n, n);
    projs(:,:,j) = A*(A'*A)^(-1)*A';
end

% [errors] = array that stores the error of reconstruction for each random
% signal at each step.
errors = zeros(steps, input_samples);

for it = 1:input_samples
    % [X] = uniformly distributed rank-one projections, i.e., a random input
    % signal.
    x = randn(2*n,1);
    x= x/norm(x);
    X = x*x';

    % [Q_hat] = empirical average of proximal projections for the input
    % signal [X].
    Q_hat = zeros(2*n, 2*n);
    m_total = 0;

    % Compute the empirical average of proximal projections in batches and
    % use it for reconstruction, keeping track of the reconstruction error
    % along the way.
    for j=1:steps
        batch_end = m_values(j);
        if j==1
            batch_start = 1;
        else
            batch_start = m_values(j-1)+1;
        end
        batch_size = batch_end - batch_start;
        batch_projs = projs(:,:,batch_start:batch_end);

        Q_hat = m_total/(m_total + batch_size) * Q_hat;

        for k = 1:batch_size
            P = batch_projs(:,:,k);
            % Compute the phaseless measurement, quantize it, and assemble
            % the empirical average of proximal projections [Q_hat].
            if (trace(P*X) > .5)
                Q_hat = Q_hat + 1/(m_total + batch_size)*P;
            else
                Q_hat = Q_hat + 1/(m_total + batch_size)*(I - P);
            end
        end

        % [X_hat] = the projection onto the principal eigenspace of [Q_hat],
        % which is the estimate for [X] using PEP.
        [x_hat,e] = eigs(Q_hat,1);
        X_hat = x_hat*x_hat';

        % Store the reconstruction error for this step of this input signal.
        errors(j, it) = norm(X_hat - X);

        m_total = m_total + batch_size;
    end
end

loglog(m_values'.*ones(steps, 1), errors)
```

A.3 Empirical results - Mean squared error

The following code was used to generate the plot in Figure 4, which shows how the theoretical mean squared error bound from Theorem 3.3.3 relates to empirical results for complementary magnitude comparison measurements associated to half-dimensioned projections.

```
% [n] = half the dimension of the input signal.
%
% [tau] = bit-flip probability.
n = 4;
tau = 1/16;

% [X] = the input signal to be measured. Without loss of generality, X is
% assumed to be the rank-one projection onto the first standard basis
% vector.
I = eye(2*n, 2*n);
x = I(:,1);
X = x*x';

```

```
% [steps] = total number of steps. Maximum number of projections is
% [step_factor]^[steps].
%
% [step_factor] = number of projections to use for the measurement is
% multiplied by this factor at each step.
%
% [samples] = number of independent collections of projections that are
% generated and used for reconstruction at each step.
steps = 20;
step_factor = 2;
samples = 1000;

% [errors] = array that stores the error of reconstruction for each sample of
% each random choice of projections.
errors = zeros(samples, steps);

% [m_values] = array that stores the number of projections used for the
% phaseless measurement at each step.
m_values = zeros(1, steps);



for it=1:steps
    % [m] = number of projections to use for measurement and reconstruction
    % in this step.
    m = step_factor^it;
    m_values(it) = m;

    % Generate [samples] independent collections of [m] projections and use
    % them for noisy measurement and recoery
    for s = 1:samples

        % [Q_T] = the bit-flipped empirical average of proximal
        % projections, which is computed during this for-loop.
        Q_T = zeros(2*n ,2*n);

        % Generate [m] random projections, measure [X], and assemble the
        % bit-flipped empirical average of the proximal projections.
        for j=1:m

            % [P] = a uniformly distributed rank-[n] projection.
            A = randn(2*n, n);
            P = A*(A'*A)^(-1)*A';

            % Compute the phaseless measurement, quantize it, and apply a
            % bit-flip with probability [tau]. The outcome of this
            % bit-flipped binary question is used to assemble the empirical
            % average [Q_T].
            if trace(P*X) > .5
                if rand() > tau
                    Q_T = Q_T + P/m;
                else
                    Q_T = Q_T + (I - P)/m;
                end
            else
                if rand() > tau
```

```
                    Q_T = Q_T + (I - P)/m;
                else
                    Q_T = Q_T + P/m;
                end
            end
        end

        % [X_T] = the projection onto the principal eigenspace of [Q_T],
        % which is the estimate for [X] using PEP.
        [x_T, e] = eigs(Q_T, 1);
        X_T = x_T*x_T';

        % Store the squared error of reconstruction for this sample of
        % projections.
        errors(s, it) = norm(X - X_T)^2;
    end
end

% [mean_squared_errors] = vector of empirical averages of the squared error
% of reconstruction, for each of the [steps] values of [m] considered.
mean_squared_errors = 1/samples*sum(errors);

loglog(m_values, mean_squared_errors)
```

Bibliography

[1] Abramowitz, M., and Stegun, I. A. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, vol. 55. US Government printing office, 1948.

[2] Akutowicz, E. J. On the determination of the phase of a Fourier integral. I. *Trans. Amer. Math. Soc. 83* (1956), 179–192.

[3] Akutowicz, E. J. On the determination of the phase of a Fourier integral. II. *Proc. Amer. Math. Soc. 8* (1957), 234–238.

[4] Alon, N., and Spencer, J. H. *The probabilistic method*, 4th ed. Wiley Publishing, 2016.

[5] Anderson, B. D., and Moore, J. B. *Optimal filtering.* Courier Corporation, 2012.

[6] Anderson, G. W., Guionnet, A., and Zeitouni, O. *An introduction to random matrices*, vol. 118 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, 2010.

[7] Artin, E. *The gamma function.* Courier Dover Publications, 2015.

[8] Bachoc, C., and Ehler, M. Signal reconstruction from the magnitude of subspace components. *IEEE Trans. Inform. Theory 61*, 7 (2015), 4015–4027.

[9] Balan, R., Bodmann, B. G., Casazza, P. G., and Edidin, D. Painless reconstruction from magnitudes of frame coefficients. *Journal of Fourier Analysis and Applications 15*, 4 (2009), 488–501.

[10] Balan, R., Casazza, P., and Edidin, D. On signal reconstruction without phase. *Applied and Computational Harmonic Analysis 20*, 3 (2006), 345 – 356.

[11] Bandeira, A. S., Cahill, J., Mixon, D. G., and Nelson, A. A. Saving phase: injectivity and stability for phase retrieval. *Appl. Comput. Harmon. Anal. 37*, 1 (2014), 106–125.

[12] Baraniuk, R., Davenport, M., DeVore, R., and Wakin, M. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation 28*, 3 (Dec 2008), 253–263.

[13] Baraniuk, R. G., Foucart, S., Needell, D., Plan, Y., and Wootters, M. Exponential decay of reconstruction error from binary measurements of sparse signals. *IEEE Transactions on Information Theory 63*, 6 (2017), 3368–3385.

[14] Bates, R., and Mnyama, D. The status of practical fourier phase retrieval. *Advances in Electronics and Electron physics 67* (1986), 1–64.

[15] Becchetti, C., and Ricotti, K. P. *Speech recognition: Theory and C++ implementation (with CD).* John Wiley & Sons, 2008.

[16] Berger, T. Rate-distortion theory. *Wiley Encyclopedia of Telecommunications* (2003).

[17] Bloch, F. Nuclear induction. *Phys. Rev. 70* (Oct 1946), 460–474.

[18] Bodmann, B. G., and Hammen, N. Stable phase retrieval with low-redundancy frames. *Advances in Computational Mathematics 41*, 2 (May 2014), 317–331.

[19] Bodmann, B. G., and Hammen, N. Algorithms and error bounds for noisy phase retrieval with low-redundancy frames. *Applied and Computational Harmonic Analysis 43*, 3 (2017), 482–503.

[20] Bodmann, B. G., and Singh, P. K. Burst erasures and the mean-square error for cyclic Parseval frames. *IEEE Transactions on Information Theory 57*, 7 (2011), 4622–4635.

[21] Boufounos, P. T., and Baraniuk, R. G. 1-bit compressive sensing. In *2008 42nd Annual Conference on Information Sciences and Systems* (March 2008), pp. 16–21.

[22] Boyd, S., and Vandenberghe, L. *Convex optimization.* Cambridge University Press, 2004.

[23] Candes, E. J., Eldar, Y. C., Needell, D., and Randall, P. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis 31*, 1 (2011), 59–73.

[24] Candès, E. J., Eldar, Y. C., Strohmer, T., and Voroninski, V. Phase retrieval via matrix completion. *SIAM J. Imaging Sci. 6*, 1 (2013), 199–225.

[25] Candes, E. J., Eldar, Y. C., Strohmer, T., and Voroninski, V. Phase retrieval via matrix completion. *SIAM Review 57*, 2 (2015), 225–251.

[26] Candès, E. J., and Li, X. Solving quadratic equations via phaselift when there are about as many equations as unknowns. *Foundations of Computational Mathematics 14*, 5 (2014), 1017–1026.

[27] Candès, E. J., Strohmer, T., and Voroninski, V. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics 66*, 8 (Nov 2012), 1241–1274.

[28] Casazza, P. G., et al. The art of frame theory. *Taiwanese Journal of Mathematics 4*, 2 (2000), 129–201.

[29] Casazza, P. G., Fickus, M., Mixon, D. G., Wang, Y., and Zhou, Z. Constructing tight fusion frames. *Applied and Computational Harmonic Analysis 30*, 2 (2011), 175–187.

[30] Casazza, P. G., and Kovačević, J. Equal-norm tight frames with erasures. *Advances in Computational Mathematics 18*, 2-4 (2003), 387–430.

[31] Casazza, P. G., and Kutyniok, G. Frames of subspaces. *Contemporary Mathematics 345* (2004), 87–114.

[32] Casazza, P. G., and Kutyniok, G. *Finite frames: Theory and applications.* Springer, 2012.

[33] Casazza, P. G., Kutyniok, G., and Li, S. Fusion frames and distributed processing. *Applied and Computational Harmonic Analysis 25*, 1 (2008), 114–132.

[34] Casazza, P. G., and Woodland, L. M. Phase retrieval by vectors and projections. In *Operator methods in wavelets, tilings, and frames*, vol. 626 of *Contemp. Math.* Amer. Math. Soc., Providence, RI, 2014, pp. 1–17.

[35] Conca, A., Edidin, D., Hering, M., and Vinzant, C. An algebraic characterization of injectivity in phase retrieval. *Appl. Comput. Harmon. Anal. 38*, 2 (2015), 346–356.

[36] Cvetkovic, Z. Source coding with quantized redundant expansions: Accuracy and reconstruction. In *Proceedings DCC'99 Data Compression Conference (Cat. No. PR00096)* (1999), IEEE, pp. 344–353.

[37] Demanet, L., and Hand, P. Stable optimizationless recovery from phaseless linear measurements. *Journal of Fourier Analysis and Applications 20*, 1 (2014), 199–221.

[38] Domel-White, D., and Bodmann, B. G. Phase retrieval by random binary questions: Which complementary subspace is closer?, 2019.

[39] Donoho, D. L. Compressed sensing. *IEEE Transactions on Information Theory 52*, 4 (2006), 1289–1306.

[40] Donoho, D. L., and Elad, M. On the stability of the basis pursuit in the presence of noise. *Signal Processing 86*, 3 (2006), 511–532.

[41] Dragotti, P. L., Kovačević, J., and Goyal, V. K. Quantized oversampled filter banks with erasures. In *Proceedings DCC 2001. Data Compression Conference* (2001), IEEE, pp. 173–182.

[42] Duffin, R. J., and Schaeffer, A. C. A class of nonharmonic fourier series. *Transactions of the American Mathematical Society 72*, 2 (1952), 341–366.

[43] Edidin, D. Projections and phase retrieval. *Applied and Computational Harmonic Analysis 42*, 2 (Mar 2017), 350–359.

[44] Eldar, Y. C., and Kutyniok, G. *Compressed sensing: theory and applications.* Cambridge University Press, 2012.

[45] Elser, V. Phase retrieval by iterated projections. *JOSA A 20*, 1 (2003), 40–55.

[46] Ephraim, Y., and Malah, D. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing 32*, 6 (1984), 1109–1121.

[47] Feller, W. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons, Inc., New York-London-Sydney, 1968.

[48] Fienup, C., and Dainty, J. Phase retrieval and image reconstruction for astronomy. *Image Recovery: Theory and Application 231* (1987), 275.

[49] Fienup, J. R. Reconstruction of an object from the modulus of its fourier transform. *Opt. Lett. 3*, 1 (Jul 1978), 27–29.

[50] Fienup, J. R. Phase retrieval algorithms: a comparison. *Applied Optics 21*, 15 (1982), 2758–2769.

[51] Folland, G. B. *A course in abstract harmonic analysis*, vol. 29. CRC Press, 2016.

[52] Foucart, S., and Rauhut, H. A mathematical introduction to compressive sensing. *Bull. Am. Math 54* (2017), 151–165.

[53] Galántai, A., and Hegedűs, C. J. Jordan's principal angles in complex vector spaces. *Numerical Linear Algebra with Applications 13*, 7 (2006), 589–598.

[54] Goyal, V. K., Kovačević, J., and Kelner, J. A. Quantized frame expansions with erasures. *Applied and Computational Harmonic Analysis 10*, 3 (2001), 203–233.

[55] Goyal, V. K., Kovačević, J., and Vetterli, M. Quantized frame expansions as source-channel codes for erasure channels. In *Proceedings DCC'99 Data Compression Conference (Cat. No. PR00096)* (1999), IEEE, pp. 326–335.

[56] Goyal, V. K., Vetterli, M., and Thao, N. T. Quantized overcomplete expansions in $\mathbb{R}^n$: Analysis, synthesis, and algorithms. *IEEE Transactions on Information Theory 44*, 1 (1998), 16–31.

[57] Gray, R. M., and Neuhoff, D. L. Quantization. *IEEE Transactions on Information Theory 44*, 6 (1998), 2325–2383.

[58] Gross, D., Liu, Y.-K., Flammia, S. T., Becker, S., and Eisert, J. Quantum state tomography via compressed sensing. *Physical Review Letters 105*, 15 (2010), 150401.

[59] Guta, M., Kahn, J., Kueng, R. J., and Tropp, J. A. Fast state tomography with optimal error bounds. *Journal of Physics A: Mathematical and Theoretical* (2020).

[60] Han, D., Kornelson, K., Weber, E., and Larson, D. *Frames for undergraduates*, vol. 40. American Mathematical Soc., 2007.

[61] Hayes, M. The reconstruction of a multidimensional sequence from the phase or magnitude of its fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing 30*, 2 (1982), 140–154.

[62] Holmes, R. B., and Paulsen, V. I. Optimal frames for erasures. *Linear Algebra and its Applications 377* (2004), 31–51.

[63] Horn, R. A., and Johnson, C. R. *Matrix Analysis*, 2nd ed. Cambridge University Press, New York, 2013.

[64] Jacques, L., Laska, J. N., Boufounos, P. T., and Baraniuk, R. G. Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors. *IEEE Transactions on Information Theory 59*, 4 (2013), 2082–2102.

[65] Johansson, K. On random matrices from the compact classical groups. *Annals of Mathematics* (1997), 519–545.

[66] Kech, M. Explicit frames for deterministic phase retrieval via phaselift. *Applied and Computational Harmonic Analysis 45*, 2 (2018), 282–298.

[67] Krahmer, F., and Liu, Y.-K. Phase retrieval without small-ball probability assumptions: Recovery guarantees for phaselift. In *2015 International Conference on Sampling Theory and Applications (SampTA)* (2015), IEEE, pp. 622–626.

[68] Laska, J. N., and Baraniuk, R. G. Regime change: Bit-depth versus measurement-rate in compressive sensing. *IEEE Transactions on Signal Processing 60*, 7 (2012), 3496–3505.

[69] Liu, G. Fourier phase retrieval algorithm with noise constraints. *Signal Processing 21*, 4 (1990), 339–347.

[70] Lustig, M., Donoho, D., and Pauly, J. M. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine 58*, 6 (2007), 1182–1195.

[71] Mehrotra, S., and Chou, P. A. On optimal frame expansions for multiple description quantization. In *2000 IEEE International Symposium on Information Theory (Cat. No. 00CH37060)* (2000), IEEE, p. 176.

[72] Millane, R. P. Phase retrieval in crystallography and optics. *J. Opt. Soc. Am. A 7*, 3 (Mar 1990), 394–411.

[73] Mroueh, Y., and Rosasco, L. On efficiency and low sample complexity in phase retrieval. In *2014 IEEE International Symposium on Information Theory* (June 2014), pp. 931–935.

[74] Mukherjee, S., and Seelamantula, C. S. Phase retrieval from binary measurements. *IEEE Signal Processing Letters 25*, 3 (2018), 348–352.

[75] Ortega, A., and Ramchandran, K. Rate-distortion methods for image and video compression. *IEEE Signal Processing Magazine 15*, 6 (1998), 23–50.

[76] Parlett, B. N. *The symmetric eigenvalue problem*, vol. 20 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1998. Corrected reprint of the 1980 original.

[77] Plan, Y., and Vershynin, R. One-bit compressed sensing by linear programming. *Comm. Pure Appl. Math. 66*, 8 (2013), 1275–1297.

[78] Plan, Y., and Vershynin, R. Robust 1-bit compressed sensing and sparse logistic regression: a convex programming approach. *IEEE Trans. Inform. Theory 59*, 1 (2013), 482–494.

[79] Plan, Y., and Vershynin, R. Dimension reduction by random hyperplane tessellations. *Discrete Comput. Geom. 51*, 2 (2014), 438–461.

[80] Rabiner, L., and Juang, B.-H. *Fundamentals of speech recognition.* Prentice-Hall, Inc., 1993.

[81] Rath, G., and Guillemot, C. Performance analysis and recursive syndrome decoding of dft codes for bursty erasure recovery. *IEEE Transactions on Signal Processing 51*, 5 (2003), 1335–1350.

[82] Rauhut, H., Schnass, K., and Vandergheynst, P. Compressed sensing and redundant dictionaries. *IEEE Transactions on Information Theory 54*, 5 (2008), 2210–2219.

[83] Renes, J. M., Blume-Kohout, R., Scott, A. J., and Caves, C. M. Symmetric informationally complete quantum measurements. *Journal of Mathematical Physics 45*, 6 (2004), 2171–2180.

[84] Scott, A. Tight informationally complete quantum measurements. *Journal of Physics A: Mathematical and General 39* (10 2006).

[85] Shannon, C. E. A mathematical theory of communication. *Bell System Technical Journal 27*, 3 (1948), 379–423.

[86] Shannon, C. E. Coding theorems for a discrete source with a fidelity criterion. *IRE National Convention Record 4* (1959), 142–163.

[87] Shi, H.-J. M., Case, M., Gu, X., Tu, S., and Needell, D. Methods for quantized compressed sensing. In *2016 Information Theory and Applications Workshop (ITA)* (2016), IEEE, pp. 1–9.

[88] Thao, N. T., and Vetterli, M. Lower bound on the mean-squared error in oversampled quantization of periodic signals using vector quantization analysis. *IEEE Transactions on Information Theory 42*, 2 (1996), 469–479.

[89] Tropp, J. A. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning 8*, 1-2 (2015), 1–230.

[90] Vershynin, R. *High-dimensional probability: An introduction with applications in data science*, vol. 47. Cambridge University Press, 2018.

[91] Vinzant, C. A small frame and a certificate of its injectivity. In *2015 International Conference on Sampling Theory and Applications (SampTA)* (May 2015), pp. 197–200.

[92] Walther, A. The question of phase retrieval in optics. *Optica Acta 10* (1963), 41–49.

[93] Wang, Y., and Xu, Z. Generalized phase retrieval: Measurement number, matrix recovery and beyond. *Applied and Computational Harmonic Analysis 47*, 2 (2019), 423 – 446.

[94] Xu, Z. The minimal measurement number for low-rank matrix recovery. *Applied and Computational Harmonic Analysis 44*, 2 (2018), 497 – 508.

[95] Zhang, L., Yi, J., and Jin, R. Efficient algorithms for robust one-bit compressive sensing. In *International Conference on Machine Learning* (2014), pp. 820–828.