Quantum Binary State Discrimination with Overcompleteness

by

Catherine Aiko Medlock

B.S. Electrical Science and Engineering, Massachusetts Institute of Technology (2016)
B.S. Physics, Massachusetts Institute of Technology (2016)
M.Eng. Electrical Engineering and Computer Science, Massachusetts Institute of Technology (2017)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2021

Author .....
Department of Electrical Engineering and Computer Science
August 27, 2021

Certified by .....
Alan V. Oppenheim
Ford Professor of Engineering
Thesis Supervisor

Accepted by .....
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Quantum Binary State Discrimination with Overcompleteness

by

Catherine Aiko Medlock

Submitted to the Department of Electrical Engineering and Computer Science
on August 27, 2021, in partial fulfillment of the
requirements for the degree of
Doctor of Science in Electrical Engineering and Computer Science

Abstract

The central topics of this thesis are operating characteristics for binary hypothesis testing in classical and quantum settings and overcomplete quantum measurements for quantum binary state discrimination. With this we explore decision and measurement operating characteristics defined as the tradeoff between probability of detection and probability of false alarm as parameters are varied. The thesis specifically addresses the Neyman-Pearson optimality of receiver operating characteristics when they are generated using threshold tests on the score variable rather than threshold tests on the likelihood ratio. The analysis applies to any scalar score variable. In the quantum setting, informationally overcomplete POVMs are explored to provide more robust quantum binary state discrimination schemes. We focus on equal trace rank one or Etro POVMs, which can be specified by arrangements of points on a sphere that we refer to as an Etro sphere.

Thesis Supervisor: Alan V. Oppenheim
Title: Ford Professor of Engineering

Acknowledgments

Al, in addition to being my mentor and research advisor you have become a dear friend and family member to me. Any description I could provide of the many ways you have positively impacted me and my life would be inadequate. Instead, I've chosen to share a few memories and stories from the past five years that are emblematic of our interactions.

In the fall of 2015 you and Professor Randy Davis agreed to take me on as a UROP the following semester. In one of our meetings you offered to give me a desk in the DSPG suite. For the rest of the semester, anytime I thought about that desk I would become almost giddy and any other stress I might be feeling would just melt away. It was the first of many confidence boosts you would give me by showing me that you had confidence in me.

In the summer of 2016 I put together a set of slides for my final internship presentation at Digital Cognition Technologies, Inc., which came out of that UROP project. When we sat down in the terminal room I was expecting to cruise right past the title slide onto the "more important" ones with the technical content. Instead we spent upwards of 35 minutes talking about authorship. I learned how important it is to consider the intricacies of which names are included and how they're ordered, and how having a frank discussion about the topic with everyone involved is essential. Even more importantly, that discussion highlighted for me how you would give me huge amounts of your time purely for the sake of mentoring me, even when there are plenty of other things you could be doing.

In the spring of 2019 we went to ICASSP in Brighton. After the panel you were on about education in signal processing, we walked to dinner at a restaurant near the conference center. We talked about the conference, but also about things completely unrelated – like why you decided not to bike at all while in England. It was very relaxing and we were in no hurry throughout the evening. I had fish pie and a delicious sticky toffee pudding for dessert. For some reason that particular dinner sticks out in my mind as a time when I really felt like we were just friends, even if on paper you

were my research advisor.

In the fall of 2020 we had a research meeting on campus. It was probably our hundredth meeting that took place in your office, with no time constraints on either side of the table and the whiteboard in play. After the first hour and a half (we were just getting started), we decided to pause for lunch. I mentioned that I would be gone for a bit to buy something to eat, but instead you got a paper plate and gave me half of your lunch. In return I made coffee for us with the coffeemaker I keep in my office. After we ate I went back to trying to convince you I was right about the technical concept we were discussing (I wasn't).

Just a few months ago in the spring of 2021 you sent an email to Mike Casey, an editor and our contact for the Foundations and Trends article we were finishing up. You asked him if it would be appropriate for me to list the article as "invited" on my resume. This was just one of many instances when you asked questions or took action purely for my benefit. There was no real benefit to you and no one would have noticed if you had done nothing at all. But I wouldn't have known the right questions to ask, so you stepped in and did it for me.

At one point early on, when I was still afraid that I might say something "stupid" that would make you regret taking me on as a student, you stated explicitly to me that I had your unconditional acceptance and approval. I don't have the words to tell you how impactful that was. Your confidence and belief in me over the years has morphed into confidence in myself, which has made me a much happier person with a fuller life. If I could choose one word to describe your advising and mentoring style, without a doubt it would be: empowering. In summary, my six words to describe our interaction and relationship are: Above all else, an empowering friendship.

Jim, you have been one of my biggest role models since you were my 6.341 recitation instructor in the fall of 2015. I greatly admire how you're able to effortlessly command a room, and yet at the same time your quiet personality makes you approachable and warm. I hope that I can one day achieve these same traits. You have consistently shown that you are willing to go out of your way to provide me with opportunities, including organizing a trip to Lincoln for me where I was able to

give a presentation and tour several different groups. That visit gave me invaluable experience as well as exposure to a place I would not have otherwise been able to see. And of course, I left with an excellent impression and am very happy to be able to go back! October 4th will always be a special date for us both. I am also very grateful for your willingness to attend countless dry runs of conference and internship presentations, giving me feedback to improve both my presentation style and pointing out mistakes or adjustments in the slides themselves.

Ike, the first time I met with you in person I was extremely intimidated. I remember feeling shocked at how quiet and humble you were. I was in awe of how quick you were to acknowledge and celebrate others' ideas, and since then I have made a conscious effort to emulate that in my own actions. You have given me advice at critical points during my time at MIT and it has stuck with me. You helped me understand that the process of interviewing a job involves some creativity on my part in that it's my responsibility to imagine how I would fit in best at the company and how I could make the best uses of the resources available. And of course, you told me that a thesis should be written at least in part as a document to my future self. In a time of very high stress as I try to finish everything up, that comment removed a lot of tension and helped me remember why I found all of this so interesting and beautiful in the first place.

Petros, I first met you in the fall of 2016 during the weekly DSPG group meetings. You contributed to every discussion with a seemingly endless amount of thoughtful comments, intriguing ideas, and just pure technical skill. I was in such awe that I almost didn't dare talk to you. I'm so glad that I was able to take 6.347 – far and away one of my favorite classes throughout all my time at MIT – because it allowed me to see that while you were just as knowledgeable as I had thought, you weren't really so scary. In fact, you were very friendly and humble and never made any of the students feel self-conscious about asking questions. That impression solidified over the next few years as I slowly got to know you better. I learned that you would never hesitate to go out of your way to give me help or advice – you spent many hours on Zoom giving me advice about job hunting, providing feedback on my job interview talk,

and helping me understand the various health benefit options for different potential employers. Another example that has stuck in my mind is when I recently asked you which option I should select on the arXiv list of licenses. I was expecting a one sentence response telling me which one to click. Instead I got a multi-paragraph message describing the different types of licenses and why I might want one over the other. You got no benefit or applause for that, but you did it because it helped someone else.

Markus, I can actually remember the exact moment when I called my parents in my freshman year to tell them that a professor had responded to my email and was willing to take me on as a UROP. You are my longest standing mentor at MIT and I look up to you in so many ways. One thing you said that I have always remembered is that your philosophy regarding UROPs was to give everyone a chance, with the understanding that if it wasn't what they expected or if they found they didn't enjoy it, they might leave, and that was fine with you. I really respect that philosophy in part because it shows that even though you're extremely accomplished and successful, you don't take yourself too seriously and you would never consider yourself above taking even the least experienced students under your wing. The two summers I spent at CERN were some of the happiest months of my life. I can still walk the entire route from the apart-hotel to the office in my head (42 minutes) and I still remember the post-lunch ice cream breaks we all took together (the first time I tried Magnum ice cream bars). I also remember taking questions at the end of a presentation about electron efficiencies while you and Aram mouthed the answers to me from the crowd. A perfect representation of our interactions was in the spring of 2018 when I was graduating with my M.Eng. degree. I hadn't worked for you for a few years by that point but to my great delight, we still kept in touch. You told me that you were planning on attending the commencement ceremony. I responded, confused, "Why? I thought you didn't have have any students graduating?" And you said, "Well you're graduating, right?"

George, you are one of the best professors I know at MIT and one of the most student-oriented. What I remember most from taking 6.02 and 6.011 with you, aside

from your crystal clear lectures and endless patience in office hours, is that it was somehow made clear without ever being explicitly said that you truly cared about the students as people, completely independently of their relationship to the class. Very few professors have ever given me personally that same feeling as strongly as you did. I still am not sure exactly what you did to convey that message, but it came through in your kindheartedness and the way you spoke and acted to students. I greatly respect that and hope to model myself after you in that way. One moment that stands out in my mind as a display of your style as a professor and a person happened in the spring of 2015. Two freshmen, Matthew Nehring and Christina Tourant, died tragically within days of each other. You came to know that many students in the class knew one or both of them personally. You addressed the issue during lecture and also sent out an email later in the day. In it you described times of struggle in your own career and how you dealt with them. You said in part, "One realization that slowly grew on me during these museum visits was that I had to find and nurture and express my own talents. I was always going to be surrounded by people who did most (if not all!) things better than me (this is even more true now that I am at MIT). I needed to be where I could support and encourage and celebrate their excellence while not belittling or discounting or suppressing what I was good at." I had never heard a professor be so candid before and I was amazed. I respect you so much for telling students that you yourself, someone we all saw as essentially incapable of doing anything wrong, actually knew what it felt like to feel out of place. I have gone back to that email multiple times over the years. True to form, even years after I was a student in your classes you have always shown genuine interest in my well-being and my research.

Andy, I don't think I can overstate how much my life has improved since you joined the group. You're such a smart and motivated student, which I really admire, but even more importantly you're an excellent friend. When we first started working together in the spring and summer of 2019, I was amazed by how quickly you got up to speed on the material and how consistently you produced results. I secretly hoped that you would come to grad school somewhere in Boston so that we could continue

to be friends and talk about signal processing. Needless to say I was overjoyed when you chose MIT. Since you've arrived here some of my favorite times at the office have been the casual coffees and lunches we've gotten together. It's a perfect break in the day and a chance to have a conversation with someone who knows exactly what I'm going through. You've also been a great technical colleague and have gently pointed out many mistakes that I've made. I admire how you're able to be quietly confident without being overconfident, and how driven you are to learn about the physical implementation of superconducting qubits. Anything I know about superconducting qubits is thanks to you. I respect your willingness to try new things even with people you don't know, like playing soccer on Wednesdays with me and Thomas and Thomas' coworkers. And we still need to go to an auto-cross event with Thomas' brother Robert! All in all, I can sincerely say that I hope we will remain close friends and share many more experiences together in the years to come.

There are many other intellectual colleagues and family members who had a positive impact on me during my time at MIT. I would like to thank the following people for answering technical questions, providing moral support, and being there as mentors and friends to me: Tom Baran, Amanda Beck, John Buck, Randall Davis, Sefa Demirtas, Dan Dudgeon, Yonina Eldar, Meir Feder, Megan Fuller, Jane Lai, Bruce Musicus, Pablo Martinez-Nuevo, Lucas Nissenbaum, Will Oliver, Arthur Redfern, Maya Said, William Souillard-Mandar, Daniel Schemmel, Guolong Su, Jessica Weaver, Lizhong Zheng.

To my family, in no particular order: Mom, I owe you my life and we both already know it, so I won't waste ink here. You would need to edit it anyways so I'm saving us both the effort. Papa, technically I owe you my life as much as I owe it to Mama. Mariko, you said you would prefer cash to a written acknowledgement so thanks for nothing and check your bank account. Edilson, thank you for feeding me all the time. Miles, thanks for giving me so many hugs. Lila, I'm going to keep trying to make you not scared of me. Kimiko, I assume you're with Mariko. Liyuan, I hope I can be as stylish and badass as you are someday. Keiko, I assume you're with Kimiko and Mariko. Ale, please teach me to be better at soccer. Mom/Patti, thanks for being

my Maine mom and always bringing me and Thomas your blueberry pie. Chris/Dad, thanks for always understanding what it's like to struggle through graduate school. Robert, thanks for getting me that coffee the first time I met you at Thanksgiving. Marian, thanks for greeting me with a big hug that same Thanksgiving. I'll never forget either of those moments. Jo and Geir, thanks for always giving us a place to stay and for your hilarious senses of humor. Alexander, I never have as much fun at 7-11 as when I'm there with you. Bixby, thanks for always asking for snuggles on the couch.

Thomas, for better or worse you've completely changed who I am. Thanks for taking care of me and I still think we have enough room for another pet.

Contents

List of Figures

List of Tables

Nomenclature

Symbols

H	Random quantity representing the true hypothesis, $H \in \{H_0, H_1\}$ 42
q_i	For $i = 0, 1$, the prior probabilities 43
S	Score variable 43
s	A specific realization or sample value of the score variable 43
$f_i(\cdot)$	For $i \in \{0, 1\}$, the conditional probability distributions of the score variable 43
$\mathcal{D}$	Decision region of a binary decision rule 43
$\hat{H}$	Random quantity representing final decision, $\hat{H} \in \{H_0, H_1\}$ 43
P_f	Probability of false alarm 44
P_d	Probability of detection 44
P_e	Probability of error 44
$g_f(\cdot)$	Mapping from LRT threshold to corresponding probability of false alarm for a scalar score variable 53
$g_d(\cdot)$	Mapping from LRT threshold to corresponding probability of detection for a scalar score variable 53

F_{can} Synthesis map of the canonical dual frame 79

$\mathcal{H}$ Hilbert space of dimension d 86

d Dimension of $\mathcal{H}$ 86

$|V\rangle\rangle$ Modified bra-ket notation used to indicate that a Hermitian operator V acting on $\mathcal{H}$ is being considered as an element of the operator space $\mathcal{V}$ 86

$\mathbf{A}_0$ Analysis map of an operator-valued frame 86

$\mathbf{F}_0$ Synthesis map of an operator-valued frame 86

$\mathcal{U}$ Subspace of the operator space $\mathcal{V}$ spanned by the identity operator . . 87

$\mathcal{U}^\perp$ Subspace of the operator space VV containing all trace zero operators; orthogonal complement of $\mathcal{U}$ 87

I Identity operator on $\mathcal{H}$ 87

$\{|F_k\rangle\rangle, 1 \leq k \leq M\}$ Frame for $\mathcal{V}$ when $\mathcal{V}$ is an operator-valued vector space 89

$\{|W_k\rangle\rangle, 1 \leq k \leq M\}$ Orthonormal basis for $\mathcal{W}$ when $\mathcal{W}$ is an operator-valued vector space 89

σ_j For $1 \leq j \leq 3$, the Pauli operators 89

$\{\hat{a}_k, 1 \leq k \leq M\}$ Observed and imprecise counterparts of the $\{a_k\}$ 91

ρ Density operator of a quantum system 99

$\{A_k, 1 \leq k \leq M\}$ Measurement operators associated with a quantum measurement 100

$\{p(k), 1 \leq k \leq M\}$ Probabilities of the different measurement outcomes of a quantum measurement 100

$\{E_k, 1 \leq k \leq M\}$ The elements of a POVM 102

L Number of quantum mechanical systems produced by physical environment or preparation procedure 104

ρ_i For $i \in \{0, 1\}$, one of the two density operators corresponding to the two possible hypotheses 104

$\{Q_k, 1 \leq k \leq M\}$ Operators derived from the elements of a POVM and used to define a tight IC POVM 131

$\{\mathbf{c}_k, 1 \leq k \leq M\}$ Etro vectors of a qubit Etro POVM 132

$\mathbf{S}$ Vector-valued score variable of relative frequencies 134

$\mathbf{s}$ A specific realization or sample of $\mathbf{S}$ 134

$\mathbf{r}_i$ For $i \in \{0, 1\}$, the Bloch vector of ρ_i 134

Acronyms

ROC Receiver operating characteristic 41

MPE Minimum probability of error 45

LRT Likelihood ratio test 47

MOC Measurement operating characteristic 51

DOC Decision operating characteristic 51

SVT Score variable threshold test 54

POVM Positive operator-valued measure 102

IC Informationally complete 102

QMS Quantum mechanical system 104

QDOC Quantum decision operating characteristic 107

QMOC Quantum measurement operating characteristics 109

Chapter 1

Introduction

Binary decisions guide our everyday lives in situations both critical and trivial. The choices made by politicians and physicians may have impact on a global or individual scale. Perhaps less consequential is whether or not we choose to carry an umbrella on a cloudy day. Any choice made inherently involves a conscious, subconscious, or formal tradeoff between benefits and detriments. The defense of a country may come at the cost of soldiers' lives, the prolongation of life at the cost of an individual patient's quality of life, the ability to keep dry in a downpour at the cost of the wasted effort of carrying an umbrella on a sunny day. In some cases our analysis of the compounding factors may be informal and the worst case outcome fairly inconsequential. But when the worst case outcome could have severe consequences as, for example, in a clinical setting or when deciding whether or not to fire a missile, it is much more desirable to have a structured analysis and process for arriving at a final decision. This may be a complicated task for many reasons, including the fact that the assignment of relative costs to the outcomes of the two possible decisions is often a judgement call itself. We may also lack a historical dataset that is large enough to allow for accurate estimation of important quantities such as the a priori probabilities.

The overarching theme of this thesis is the study of such binary hypothesis testing problems with a particular focus on the problem of quantum binary state discrimination. In Sections 1.1 and 1.2 below we provide some context regarding these problems and describe several important quantities that will reappear many times in the the-

sis, including the probabilities of false alarm and detection. Among the factors that differentiate typical viewpoints on binary hypothesis testing in the classical versus quantum settings is the meaning of the word "measurement". Section 1.3 contains a few remarks about measurement and gives a sense of our perspective. In Section 1.4 we provide a very brief review of an area in quantum mechanics where overcompleteness and redundancy have been exploited. This is in preparation for discussions later in the thesis about how overcompleteness can potentially be utilized in the context of quantum binary state discrimination. Section 1.5 contains an outline of the remaining chapters.

1.1 Binary Hypothesis Testing

Mathematically, the objective of a binary hypothesis testing problem is to make a decision as to which of two possible hypotheses, denoted as H_0 or H_1, is true in some optimal way based on the outcomes of measurements or observations made on an input. The measurements or observations are assumed to result in a set of numerical values that may be concatenated to form a vector or alternatively may be combined algorithmically into a single value. In either case the result is frequently modeled as a realization of a random variable referred to as the score variable. A particular realization of the score variable can be viewed as being drawn from one of two probability distributions conditioned on the true hypothesis. Based in part on these probability distributions, classical decision theory is utilized to make a final decision in an optimal way with respect to a specific criterion. Uncertainty in the final decision is caused by overlap in the two conditional probability distributions resulting from, for example, noise in a physical environment or variability in a population of subjects.

Historically a quantity considered to be of significance for this problem is the probability of error, denoted as P_e and defined as the probability of identifying H_0 to be true when H_1 is in fact true or vice versa. Other probabilities that may be of interest are (i) the probability of detection, denoted by P_d and defined as the

probability of deciding that H_1 is true given that it is indeed true, (ii) the probability of a miss, denoted by P_m and defined as the probability of deciding that H_0 is true given that in fact H_1 is true, and (iii) the probability of false alarm, denoted by P_f and defined as the probability of deciding that H_1 is true given that H_0 is in fact true. Also of importance are the a priori or prior probabilities, denoted by q_0 and q_1, associated with whether H_0 or H_1 is true apart from any measurement, observation, or decision. Various of these probabilities are of course connected mathematically through the rules of probability.

Since in many scenarios the prior probabilities are difficult or impossible to assess, it has become common in many contexts to formulate the decision making process without explicitly requiring knowledge of these probabilities. One approach that has become widespread for accomplishing this is to focus on the tradeoff between P_f and P_d, often displayed using what is commonly referred to as a receiver operating characteristic or ROC. ROCs originated in the radar signal detection community, where they were used to characterize systems that detected the presence or absence of military targets during World War II [75]. The use of ROCs has become increasingly prevalent in a very broad set of application areas including biostatistics and machine learning [5, 27, 50, 66, 71, 73, 84]. More generally, the term operating characteristic can be used to refer to any representation of the tradeoff between P_f and P_d as one or more parameters of a hypothesis testing system is varied. Many results of this thesis are directly related to operating characteristics or are illustrated using different types of operating characteristics.

1.2 Quantum Binary State Discrimination

Linear algebra is a core part of the mathematical language of signal processing including as it is applied to binary hypothesis testing problems. A canonical example is the problem of discriminating between two discrete time finite length signals in the presence of independent and identically distributed (IID) Gaussian noise. As described further in Chapter 2, implementation of the minimum probability of error decision

strategy can be accomplished using what is referred to as a matched filter that is then followed by a sampler. The problem essentially boils down to discriminating between two vectors belonging to a finite dimensional Hilbert space under the assumption that IID Gaussian noise has been applied to their expansion coefficients with respect to a given orthonormal basis. Geometric intuition suggests and a mathematical derivation confirms that the minimum probability of error solution corresponds essentially to projecting the observed vector onto the direction defined by the difference of the two vectors we wish to distinguish. And indeed it is well-understood that the matched filter can be interpreted in this way.

The problem of discriminating between two finite dimensional vectors in the presence of additive noise is in some ways straightforward, but in fact it becomes increasingly rich and interesting in the presence of modified assumptions and constraints. The modifications considered in this thesis can be roughly grouped into two categories: (i) those stemming from frame theory and leading to an assumption that the expansion coefficients correspond not a linearly dependent set of vectors rather than a basis, and (ii) those stemming from the postulates of quantum mechanics. Specifically, the problem of quantum binary state discrimination is merely a special case of the problem of discriminating between two vectors in a finite dimensional Hilbert space with the caveat that as a result of the postulates of quantum mechanics, the noise values on the expansion coefficients are correlated both with each other and with the true underlying vector. Modifying the quantum measurement used for discrimination can be done in such a way that the expansion coefficients correspond to linearly dependent vectors instead of linearly independent vectors. To make these connections more precise we now state the problem of quantum binary state discrimination.

In the binary hypothesis testing problem considered in this thesis, H_0 and H_1 each correspond to a specific physical environment or preparation procedure that has resulted in a collection of L quantum mechanical systems (QMSs), each of which can be described by the same quantum state. The pre-decision operator is comprised of a specific quantum measurement performed separately on each of the individual QMSs. We assume for simplicity that the chosen quantum measurement has a finite number

M of possible outcomes. The score variable is the M-element vector containing the relative frequencies of each possible measurement outcome. This problem along with its many variations is often referred to as quantum binary state discrimination and plays an important role in quantum communication systems [38, 19, 76, 3, 80, 79, 40]. In that context each possible state represents a different transmitted message. More broadly, quantum hypothesis testing can be seen as a way to read out the information that has been computed by other quantum technologies and is contained in the state of a quantum system [10, 72]. The measurement strategy that achieves minimum probability of error was derived by Helstrom [38].

As described in more detail in Chapter 4, a quantum state can always be represented by a Hermitian operator ρ referred to as a density operator acting on a Hilbert space. We will always assume that the Hilbert space is finite dimensional for simplicity. An M-outcome quantum measurement always has an associated set of M Hermitian operators $\{E_k, 1 \leq k \leq M\}$ acting on the same Hilbert space. The $\{E_k\}$ always satisfy the definition of what is referred to as a positive operator-valued measure (POVM). ρ and the $\{E_k\}$ can be interpreted as elements of a common vector space $\mathcal{V}$ that can be referred to as an operator space, and the probabilities $\{p(k)\}$ mentioned above can be interpreted as basis expansion coefficients of ρ whenever the $\{E_k\}$ span $\mathcal{V}$. The $\{E_k\}$ can also be chosen to be linearly dependent while still spanning $\mathcal{V}$, in which case the $\{p(k)\}$ contain some amount of redundancy.

If the expansion coefficients (i.e., probabilities) corresponding to each density operator could be determined precisely then the two hypotheses could be perfectly distinguished. What prevents this from being possible is that in the problem formulation given above, there are only L QMSs available for measurement. Thus only relative frequencies can be obtained. A vector of relative frequencies can be expressed as a vector of true probabilities added to an error vector. But the distribution of the error vector depends directly on the true probabilities, and additionally the error values must add to zero and so they are also correlated with each other. This is why as stated above, the error values on the expansion coefficients are correlated with each other and with the true hypothesis.

1.3 Classical Versus Quantum Measurement

Quantum binary state discrimination naturally involves a tradeoff between P_f and P_d and therefore it also involves the notion of an operating characteristic. Interestingly, operating characteristics of any kind are significantly less prevalent in the quantum binary state discrimination literature. Perhaps one of the principal reasons for this is that although there are many similarities between the classical and quantum scenarios, there are also some fundamental differences that stem from the underlying differences between the postulates of classical versus quantum physics. Of particular importance and as described in Chapter 4 are the stipulations made by the postulates of quantum mechanics about the state of a quantum system and about the concept of quantum measurement.

To give a sense of the differences in how measurement is thought of in classical and quantum settings, we provide here a brief discussion that takes its inspiration in part from the delicate and comprehensive treatment of the topic in Chapter 1 of [58]. In all settings it is necessary to make a distinction between the word "measurement" as it refers to a specified experimental setup in a hypothetical laboratory and as it refers to the laws of classical or quantum physics that model our knowledge of the behavior of the laboratory equipment and its physical interaction with the object or system we wish to measure. Our main concern in this thesis is the latter connotation of the word – all of our discussion can be aptly described as applications of the laws of quantum physics to predict the behavior of hypothetical laboratory equipment.

In the classical world it is common to assume that the role of measurement is to expose the pre-existing value of some property of the object or system of interest. To measure the weight of a block of wood is to acquire knowledge of an intrinsic property that it has. Measurements as performed in a laboratory are interactive processes between the measuring apparatus and the object or system being measured. Placing the wood on a scale causes the beam to tip to one side. The effect of the interaction on the apparatus and/or the measured system can be thought of as the outcome of the measurement. We interpret the tipping of the beam to a certain

position as meaning that the weight of the wood is higher or lower than a specific value. While measurements performed on quantum systems also necessarily involve interaction between the apparatus and the measured system, and while the effects of the interaction can again be thought of as the outcome of the measurement, it is not the case that the alterations to the apparatus can in general be interpreted as a result of intrinsic, pre-existing properties of the measured system. Related to this statement is the fact that quantum measurements are in general not reproducible. Two identically prepared quantum systems may interact with two identical measurement apparatus and it would be entirely consistent with the laws of the quantum physics for the effects on both the apparatus and the system to be different. And this in comparison with an analogous classical scenario – if two identical blocks of wood were placed on identical scales, we would of course expect the two beams to tip to the same position. It is important to note, however, that the assumed reproducibility of classical measurements do not take into account the presence of random error. In a classical communication setting the receiver might measure two copies of the same incoming bit sequence. But if the copies are each subject to different errors produced by the communication channel, the measured bits will in general be different. As is made more exact in Chapter 4, in this thesis we will be focused on quantum measurements in which the possible effects of the measurement interaction on any hypothetical laboratory equipment and on any measured system – that is, the possible measurement outcomes – can only happen in finite number M of ways. The possibilities will be indexed by an integer k with $1 \leq k \leq M$. The laws of quantum physics allow us to predict the probabilities $\{p(k)\}$ of the various outcomes and it is these probabilities that are our main interest.

1.4 Overcompleteness in Quantum Mechanics

Quantum measurements that employ redundant, or overcomplete, representations of the state of the system being measured have been studied in the context of quantum state estimation. Roughly, the goal of that problem is to reconstruct an unknown

quantum state from estimates of its probabilities corresponding to a particular quantum measurement. This is mathematically analogous to reconstructing an unknown vector from imprecise versions of its expansion coefficients with respect to a given set of linearly independent or linearly dependent vectors that span the space in question.

In the case of quantum binary state discrimination the two vectors we wish to distinguish are elements of an operator space $\mathcal{V}$. It was mentioned in Section 1.2 that the POVM $\{E_k\}$ associated with a given quantum measurement can be chosen to span $\mathcal{V}$. When this is the case the POVM satisfies the definition of what is referred to as an informationally complete (IC) POVM, which is one that maps every quantum state to a unique probability distribution over the possible measurement outcomes [2, 8, 23, 24, 28, 29, 30, 62, 64, 67, 81, 82, 83]. When the $\{E_k\}$ span $\mathcal{V}$ and are linearly dependent, it is referred to as being informationally overcomplete (IOC). Due to a fundamental result that states that the elements of an IC or IOC POVM always form a frame for $\mathcal{V}$, there is strong overlap in the mathematical analysis of quantum state estimation with the field of frame theory, including, for example, a consideration of the optimal dual frame for reconstruction [82]. Less attention has been given to how overcompleteness might benefit quantum state discrimination.

1.5 Outline

In Chapter 2 we establish our notation and terminology surrounding general binary hypothesis testing problems. Multiple widely used criteria for determining the optimal binary decision rule for a given score variable are reviewed, all of which lead to threshold tests performed on the likelihood ratio. In scenarios where the information needed to implement the optimal likelihood ratio test is not fully known, a common strategy is to threshold the score variable as opposed to thresholding the likelihood ratio. A central topic of Chapter 2 is the development of a condition that, when satisfied, guarantees that the ROC generated using threshold tests on the score variable is equivalent to the ROC that would have been generated using threshold tests on the likelihood ratio.

Chapter 3 summarizes our viewpoint on several core concepts of linear algebra and

frame theory. The objective is to introduce the mathematical machinery and notation necessary to apply the concepts to quantum measurement in Chapters 4 and 5. In Chapter 4 we review the quantum state and measurement postulates and use them to precisely state the quantum binary state discrimination problem considered in this thesis. Helstrom's minimum probability of error solution is reviewed and examples of two types of operating characteristics are presented.

In Chapter 5 we phrase the quantum binary state discrimination problem using the language of linear algebra applied to operator spaces. The operator space $\mathcal{V}$ in question contains all density operators and POVM elements. A direct sum decomposition of $\mathcal{V}$ into two orthogonal subspaces leads us to define a counterpart to the well-known Bloch ball and Bloch sphere for a special class of POVMs that we refer to as equal trace rank one (Etro) POVMs. An Etro POVM corresponding to a qubit measurement can be fully specified by M points on what we refer to as an Etro sphere of radius $\sqrt{2}/M$.

In Chapter 6 we present an exploratory investigation into how overcompleteness in Etro POVMs can be exploited in the context of qubit binary state discrimination. Specifically, Etro POVMs constructed from arrangments of points corresponding to the vertices of a Platonic solid have been of particular interest in the quantum state estimation community due in large part to the fact that they all provide straightforward state reconstruction formulas. State discrimination does not require state reconstruction, allowing for the construction of Etro POVMs using various point arrangements on the sphere. In the qubit binary state discrimination problem considered Chapter 6, we assume that the overall alignment of the two states relative to the coordinates of the Bloch sphere is unknown. Helstrom's minimum probability of error solution is irrelevant in this case since its implementation requires complete knowledge of the two states. We compare a selection of Etro POVMs based on the minimum and maximum probabilities of error they can achieve over all possible rotational alignments of the two states relative to the Bloch sphere.

Chapter 7 provides a selection of concluding remarks and suggestions for topics of future study.

Chapter 2

Operating Characteristics for Binary Hypothesis Testing

After first defining a simple framework that encompasses general binary hypothesis testing problems, in Section 2.2 we review optimal binary decision strategies with respect to the minimum probability of error, minimum risk or Bayes' cost, and Neyman-Pearson criteria. All of these criteria lead to the family of likelihood ratio test (LRT) decision rules, which can sometimes but not always be recast as what we refer to as score variable threshold test (SVT) decision rules. A principal contribution of the thesis is described in Section 2.5. It addresses the relation between SVT and LRT receiver operating characteristics or ROCs for a scalar score variable, and in particular when the LRT ROC, which is optimal with respect to all of the criteria mentioned above, can be recovered from a sub-optimal SVT ROC. ROCs can also be referred to as decision operating characteristics in our terminology.

2.1 Framework

A given binary hypothesis testing problem might belong to one of a number of more specific problem types. For example, an established binary hypothesis testing problem in the signal processing community is that of determining whether an incoming waveform consists only of noise or of noise in addition to a pre-determined signal.

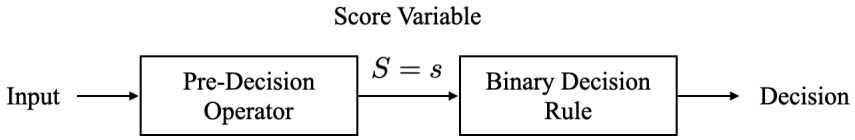


Figure 2-1: Binary hypothesis testing framework.

Another common objective is to determine which out of a pre-determined pair of signals an incoming waveform represents. These two problems are typically referred to as signal detection and signal discrimination or signal classification, respectively. In a machine learning setting, the problem of determining to which of two categories a given data point belongs is often referred to as binary classification. Additional examples are given at the end of Section 2.1 and are described using the terminology introduced next. Note however that the discussion of Chapter 3 pertains to general binary hypothesis testing problems and not to one specific subcategory.

The framework that we consider in this thesis is shown in Figure 2-1. It starts with an input that may take many different forms depending on context. For instance, the input may be a patient who is healthy or ill or it may be a collection of quantum mechanical systems that was produced by one of two preparation procedures. The true state of the input is assumed to be dependent on a random quantity H representing the true hypothesis, where $H = H_0$ or $H = H_1$. The objective is to make a binary decision about whether $H = H_0$ or $H = H_1$ in an optimal way with respect to a specific optimality criterion. The a priori or prior probabilities that each hypothesis is true will be denoted as

$$P(H = H_i) = q_i\,, \quad i \in \{0,\, 1\}. \tag{2.1}$$

Note that the values of the $\{q_i\}$ may not be readily available and may instead need to be estimated using available data and application-specific modeling. As described in more detail in Section 2.4, the absence of the use of priors in the formulation of decision, or receiver, operating characteristic curves is an important advantage to that approach. The first step in making a binary decision about the true hypothesis

is to process the input using what we refer to as a pre-decision operator. This results in a sample of a random variable sometimes referred to as the score variable. We will denote the score variable itself by upper case S, the sample value of the score variable by lower case s, and will assume for simplicity that S is real-valued. The conditional probability distributions of the score variable corresponding to each possible hypothesis, also sometimes referred to as likelihood functions, will be denoted as

$$P(S = s \,|\, H = H_i) = f_i(s)\,, \quad i \in \{0, 1\}. \tag{2.2}$$

The decision about the true hypothesis can be viewed as a decision about whether a specific value of s was drawn from $f_0(\cdot)$ or $f_1(\cdot)$. When $f_0(s) > 0$ and $f_1(s) > 0$, it is impossible to identify the true hypothesis with certainty. This problem has been studied extensively in the field of classical decision theory. In that context each binary decision rule is typically described using a decision region $\mathcal{D}$ that is a subset of the sample space of S. The final decision, which will be represented by a second random quantity $\hat{H}$, where $\hat{H} = H_0$ or $\hat{H} = H_1$. An error is made when $\hat{H} \neq H$. If s lies in $\mathcal{D}$ then we set $\hat{H} = H_1$. Otherwise we set $\hat{H} = H_0$.[1] This can be summarized as

$$\hat{H} = \begin{cases} H_1 \text{ if } s \in \mathcal{D} \\ H_0 \text{ if } s \notin \mathcal{D}. \end{cases} \tag{2.3}$$

Two of the most common families of decision regions are those corresponding to likelihood ratio tests and those corresponding to what we refer to in [47] as score variable threshold tests.

The probabilities of false alarm (P_f) and detection (P_d) can both be written as functions of the conditional distributions $\{f_i(\cdot)\}$ and the decision region $\mathcal{D}$. Recall that the probability of false alarm is defined as the conditional probability that we decide $\hat{H} = H_1$ given that $H = H_0$. The probability of detection is the conditional

[1] There also exist randomized decision strategies in which each value of the score variable is associated with a certain probability of deciding that $\hat{H} = H_1$ or $\hat{H} = H_0$, but we will not be considering those in this thesis.

probability that $\hat{H} = H_1$ given that $H = H_0$. We have

$$P_f = P(\hat{H} = H_1 | H = H_0) = \int_{s \in \mathcal{D}} ds\, f_0(s) \tag{2.4a}$$

$$P_d = P(\hat{H} = H_1 | H = H_1) = \int_{s \in \mathcal{D}} ds\, f_1(s). \tag{2.4b}$$

In Equations (2.4) and throughout Chapter 2, we have arbitrarily assumed that S is continuous and have thus used an integral instead of a sum to calculate probability values. The results can be appropriately modified for the discrete case, and indeed in Chapter 4 we will assume that S is discrete. The probability of error, denoted by P_e can be expressed as $P_e = q_0\, P_f + q_1\, (1 - P_d)$. Obviously, the values of P_f and P_d inherently depend on $f_0(\cdot)$, $f_1(\cdot)$, and $\mathcal{D}$. Operating characteristics can be generated by varying parameters of the pre-decision operator, which effectively changes $f_0(\cdot)$ and $f_1(\cdot)$, as well as by varying the binary decision rule which is done by varying $\mathcal{D}$. These two possibilities correspond to the two classes of the operating characteristics described in Sections 2.3 and 2.4.

Example 2.1. *In a typical radar signal detection problem the input to the pre-decision operator is the waveform reflected by a target and received by the radar system following the emission of an electromagnetic pulse by the radar transmitter. The pre-decision operator is a linear filter followed by a sampler. The score variable is the sampled value at a specified time at the output of the filter. The value of this score variable is used to make a binary decision about whether or not a target is present.*

Example 2.2. *In a medical decision-making scenario, the input is a series of clinical measurements made on a patient and the pre-decision operator might be a machine learning algorithm that combines the measurements into a single number. The score variable is this composite number and its value is used to make a decision about whether the patient is healthy or ill.*

Example 2.3. *Consider a scenario in which the input is a realization x of a real-valued Gaussian random variable X with zero mean and variance of either σ_0^2 or*

σ_1^2, i.e., $H = H_i$, where $i \in \{0, 1\}$. To distinguish between the two hypotheses with minimum probability of error, we set the score variable to $s = x^2$ and the decision region $\mathcal{D}$ to $\mathcal{D} = \{s : s \geq \gamma\}$, where $\gamma \geq 0$ is a fixed threshold value that depends on the prior probabilities. This choice of decision region corresponds to what we will refer to as a score variable threshold test or SVT. Decision strategies that minimize probability of error are the topic of Section 2.2.1. This specific example is elaborated on further in Chapter 2.

In practice and when possible, it can be useful to relate the decision region $\mathcal{D}$ connected to the score variable to a corresponding decision region connected to the input. For a given $\gamma \geq 0$, we have $s \geq \gamma$ exactly when $x \geq \sqrt{\gamma}$ or $x \leq -\sqrt{\gamma}$. A one-sided threshold decision region on s is equivalent to a symmetric two-sided threshold decision region on x.

2.2 Optimal Decision Rules with respect to Common Criteria

In Section 2.2 we briefly review decision rules that are optimal with respect to the minimum probability of error, minimum Bayes' cost, and Neyman-Pearson criteria. Identifying the optimal decision rule for a given criterion amounts to finding the associated optimal decision region $\mathcal{D}$. All of the optimal decision regions have the form of an LRT with some threshold $\eta \geq 0$.

2.2.1 Minimum Probability of Error

One of the most common optimality criteria used in binary decision making is the minimum probability of error or MPE criterion. To find the optimal decision region we start by writing the total probability of error, denoted as P_e, as an expectation over all possible values of the score variable S,

$$P_e = \int ds\, f_S(s)\, P_{e|s}. \tag{2.5}$$

Here $f_S(s)$ is the overall probability distribution function of S and $P_{e|s}$ is the conditional probability of error given that $S = s$. $f_S(s)$ is non-negative for all values of s, so to minimize P_e it is sufficient to minimize $P_{e|s}$ for each value of s individually. To see how this can be achieved, recall that if s lies in the decision region $\mathcal{D}$ then we decide $\hat{H} = H_1$. An error is made when s lies in $\mathcal{D}$ but in fact $H = H_0$. The reverse is true for when s does not lie in $\mathcal{D}$. $P_{e|s}$ can be written as

$$P_{e|s} = \begin{cases} P(H = H_0|S = s) \text{ if } s \in \mathcal{D} \\ P(H = H_1|S = s) \text{ if } s \notin \mathcal{D}. \end{cases} \tag{2.6}$$

The conditional probability $P(H = H_i|S = s)$ for $i \in \{0, 1\}$ is referred to as the a posteriori, or posterior, probability that $H = H_i$ given that $S = s$. To minimize $P_{e|s}$ we should choose the hypothesis that has the maximum a posteriori probability conditioned on the observation $S = s$. Thus the optimal decision rule with respect to the MPE criterion is

$$\hat{H} = \begin{cases} H_1 \text{ if } P(H = H_1|S = s) \geq P(H = H_0|S = s) \\ H_0 \text{ if } P(H = H_1|S = s) < P(H = H_0|S = s). \end{cases} \tag{2.7}$$

Note that the values of $S = s$ for which the a posteriori probabilities are equal can be associated with either final decision without affecting the total probability of error. Applying Bayes' rule to the $P(H = H_i|S = s)$ yields

$$P(H = H_i|S = s) = \frac{P(S = s|H = H_i)\, P(H = H_i)}{f_S(s)} = \frac{f_i(s)\, q_i}{f_S(s)}, \quad i \in \{0, 1\}. \tag{2.8}$$

Cancelling the factors of $f_S(s)$ on both sides of the inequalities and rearranging leads to the following equivalent decision strategy,

$$\hat{H} = \begin{cases} H_1 \text{ if } \dfrac{f_1(s)}{f_0(s)} \geq \dfrac{q_0}{q_1} \\ H_0 \text{ if } \dfrac{f_1(s)}{f_0(s)} < \dfrac{q_0}{q_1}. \end{cases} \tag{2.9}$$

The quantity $f_1(s)/f_0(s)$ is referred to as the likelihood ratio associated with the value $S = s$, and a decision rule that applies a threshold to the likelihood ratio is termed a likelihood ratio test or LRT. In the terminology of decision regions introduced above, the optimal MPE decision region $\mathcal{D}_{\text{MPE}}$ is an LRT with threshold value $\eta = q_0/q_1$,

$$\mathcal{D}_{\text{MPE}} = \left\{ s : \frac{f_1(s)}{f_0(s)} \geq \frac{q_0}{q_1} \right\}. \tag{2.10}$$

It may also be desirable in some cases to assign different relative cost values to the different possible decision scenarios – a detection, a false alarm, etc. The expected cost incurred over all values of S is sometimes referred to as the Bayes' cost and the corresponding optimal decision rule is the one that minimizes Bayes' cost. If c_{ij} is the cost of deciding $\hat{H} = \hat{H}_i$ when in truth $H = H_j$, then the probability of error corresponds to the special case where $c_{01} = c_{10} = 1$ and $c_{00} = c_{11} = 0$. A parallel analysis to the one described above can be used to show that the optimal decision rule with respect to the minimum Bayes's cost criterion is

$$\hat{H} = \begin{cases} H_1 \text{ if } \dfrac{f_1(s)}{f_0(s)} \geq \dfrac{q_0(c_{10} - c_{00})}{q_1(c_{01} - c_{11})} \\ H_0 \text{ if } \dfrac{f_1(s)}{f_0(s)} < \dfrac{q_0(c_{10} - c_{00})}{q_1(c_{01} - c_{11})}. \end{cases} \tag{2.11}$$

Thus, the minimum Bayes' cost decision rule is an LRT with threshold $\eta = [q_0(c_{10} - c_{00})]/[q_1(c_{01} - c_{11})$. Its decision region $\mathcal{D}_{\text{BC}}$ is

$$\mathcal{D}_{\text{BC}} = \left\{ s : \frac{f_1(s)}{f_0(s)} \geq \frac{q_0(c_{10} - c_{00})}{q_1(c_{01} - c_{11})} \right\}. \tag{2.12}$$

As we summarize below, the optimal decision region for the Neyman-Pearson criterion has a very similar form.

2.2.2 Neyman-Pearson Criterion

While the minimum probability of error and minimum Bayes' cost criteria are intuitively desirable in that they minimize the notion of average cost over many decisions,

implementation of the resulting optimal decision rules may be impractical if the priors are unknown and difficult to estimate. The minimum Bayes' cost criterion also requires us to assign relative costs to the different possible decisions, which may be a highly subjective task with no obvious or clear answer. Another common optimality criterion used in classical binary hypothesis testing scenarios involves placing bounds on either the probability of false alarm or the probability of a missed detection. As an example, in the radar community P_f is often constrained to be below 10^{-6} since false detection of a target can trigger costly actions and a waste of expensive resources. In this and other similar situations, a reasonable objective is to maximize P_d subject to a given tolerable upper bound on P_f. This is referred to as the Neyman-Pearson criterion. The optimal Neyman-Pearson decision rule is an LRT with a threshold value η that is chosen to ensure that P_f is exactly equal to its upper bound [39, 51], i.e., the optimal Neyman-Pearson decision region $\mathcal{D}_{\text{NP}}$ is

$$\mathcal{D}_{\text{NP}} = \left\{ s : \frac{f_1(s)}{f_0(s)} \geq \eta_0 \right\} \text{ where } \eta_0 \text{ is chosen s.t. } P_f = \alpha. \tag{2.13}$$

Recall that the threshold value η_0 affects the value of P_f through the integral given in Equation (2.4a). An informal argument [51] that provides intuition as to why Equation (2.13) is optimal with respect to the Neyman-Pearson criterion is as follows. Assume that the decision region $\mathcal{D}$ has been chosen to be Neyman-Pearson optimal. Then by definition it is impossible to modify it in such a way that P_d is increased while P_f stays the same. Mathematically we can think of modification of the decision region as taking two small portions of the real axis, one that lies in $\mathcal{D}$ and is denoted as the interval $[s, s + ds]$ and one that lies outside of $\mathcal{D}$ and is denoted as the interval $[s', s' + ds']$, and interchanging their decision region assignments. In other words, we remove the interval $[s, s + ds]$ from $\mathcal{D}$ and add the interval $[s', s' + ds']$. According to Equations (2.4), the resulting changes in P_f and P_d are

$$\Delta P_f = f_0(s')\, ds' - f_0(s)\, ds \tag{2.14a}$$

$$\Delta P_d = f_1(s')\, ds' - f_1(s)\, ds. \tag{2.14b}$$

If we assume that the value of P_f stays the same ($\Delta P_f = 0$), then since the original decision region was Neyman-Pearson optimal we know by definition that the value of P_d must have stayed the same or decreased ($\Delta P_d \leq 0$). Applying these conditions to Equations (2.14) and combining them together leads to the requirement that

$$\frac{f_1(s')\,ds'}{f_0(s')\,ds'} \geq \frac{f_1(s)\,ds}{f_0(s)\,ds}. \tag{2.15}$$

After cancelling the factors of ds and ds', the right-hand side of the inequality is equal to the likelihood ratio at the point $S = s$, which lay in the original, Neyman-Pearson optimal decision region $\mathcal{D}$. Similarly, the left-hand side is the likelihood ratio as the point $S = s'$, which lay outside of this region. Since the intervals $[s, s + ds]$ and $[s', s' + ds']$ were arbitrary so long as they lay inside or outside of $\mathcal{D}$, respectively, Equation (2.15) says that for the Neyman-Pearson optimal decision region $\mathcal{D}$, the likelihood ratio for values of the score variable lying inside $\mathcal{D}$ is always greater than or equal to the likelihood ratio for values lying outside $\mathcal{D}$. In other words, the Neyman-Pearson optimal decision regions represent a threshold test on the likelihood ratio.

Example 2.4. *It was stated in Example 2.1 that in a typical radar signal detection problem the pre-decision operator is a linear filter followed by a sampler. We summarize here a well-known example [39, 51] in which the filter and sampler are designed to compute the likelihood ratio of the incoming samples. Consider a scenario in which the samples $x[n]$ of an incoming waveform consist only of noise or of noise in addition to a pre-determined, finite length signal $y[n]$,*

$$x[n] = \begin{cases} w[n] \text{ if } H = H_0 \\ \\ w[n] + y[n] \text{ if } H = H_1 \end{cases}, \quad 1 \leq n \leq N. \tag{2.16}$$

In Equation (2.16), $w[n]$ is assumed to be an independent and identically-distributed zero-mean Gaussian random process with variance σ^2. The conditional PDFs of the N samples, $f_i(x[1], \ldots, x[N])$ for $i \in \{0, 1\}$, are also Gaussian and their ratio can be

expressed as

$$\frac{f_1(x[1],\ldots,x[N])}{f_0(x[1],\ldots,x[N])} = \exp\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{N} y[n]^2 + \frac{1}{\sigma^2}\sum_{n=1}^{N} x[n]\,y[n]\right]. \tag{2.17}$$

Straightforward algebra leads to the conclusion that for an LRT threshold value η_0, the likelihood ratio is greater than or equal to η_0 whenever

$$\sum_{n=1}^{N} x[n]\,y[n] \geq \sigma^2 \ln(\eta_0) + \frac{1}{2}\sum_{n=1}^{N} y[n]^2. \tag{2.18}$$

The sum on the left-hand side of the inequality can be computed by inputting the incoming samples $x[n]$ into a linear filter whose impulse response is $h[n] = y[-n]$ and sampling the output of the filter at the appropriate time. $h[n]$ is commonly referred to as a matched filter since it is "matched" to $y[n]$. The value of η_0 could be chosen to be optimal with respect to minimum probability of error, minimum Bayes' cost, or the Neyman-Pearson criterion.

2.3 Measurement Operating Characteristics

It is common in many practical scenarios for the optimal pre-decision operator to be only partially known. A lack of information about the two possible hypotheses for instance, may make it impossible to fully parameterize the optimal pre-decision operator for a given optimality criterion. In such a case it may be desirable to fix the binary decision rule while varying some parameter or parameters of the pre-decision operator as a way of determining or at least estimating its optimal form. An operating characteristic can be generated by plotting or tabulating the values of P_f and P_d obtained for each individual pre-decision operator and we refer to operating characteristics generated in this way as measurement operating characteristics or MOCs. If the input to the system is a series of sample values $x[n]$, for instance, and the optimal pre-decision operator is known to be a filter of a given bandwidth, then the center frequency of the filter might be varied while the binary decision rule is kept

fixed. A comparison of empirical values of the probability of error achieved with each center frequency could be used to obtain a rough estimate of its optimal (with respect to minimum probability of error) value. MOCs generated in the quantum setting are discussed in Chapter 4.

2.4 Receiver Operating Characteristics

When considering a specific optimality criterion under a fixed set of conditions – fixed priors, for example – the primary goal is to find the single optimal decision rule with respect to that criterion and those conditions. But it is often very useful to consider entire families of decision rules that are optimal with respect to potentially different criteria and for possibly different sets of conditions. Receiver operating characteristics or ROCs are a useful tool that allows us to accomplish exactly this. In reference to Figure 2-1, ROCs are generated by fixing the pre-decision operator, varying the decision region of the binary decision rule, and plotting the resulting probabilities of false alarm and detection. For the sake of consistency with the literature we will continue to refer to them as ROCs. But to be more consistent with the analogous curve introduced in Chapter 4 for the quantum binary state discrimination problem, they could equally well be referred to as decision operating characteristics or DOCs. Our purpose for defining this slightly different terminology is to emphasize that these types of operating characteristics need not be considered specific to traditional applications such as radar signal detection or clinical decision making.

Among the ways of utilizing ROCs, it has become common in many communities to compare two decision-making strategies based on global properties of their corresponding ROC. Such a comparison is inherently difficult because of the fundamental difference between metrics used to compare individual decision rules and metrics used to compare entire ROCs, which represent collections of decision rules. It is less clear how to interpret the latter in terms of realizable differences in performance since ultimately only a single rule can be used. Nevertheless, the area under an ROC or AUC is one such property that is widely used in the literature and in practice. There

is significant debate over whether the AUC is a reasonable metric despite its popularity and many alternatives have been proposed although not widely accepted. For additional details we refer the reader to [34, 45].

Throughout the remainder of Chapter 2, we assume for simplicity that $f_0(\cdot)$ and $f_1(\cdot)$ are continuous and strictly positive functions. This implies that the likelihood ratio function $f_1(\cdot)/f_0(\cdot)$ is continuous. We assume in addition that the likelihood ratio function is not constant over any finite interval. In Sections 2.4.1 and 2.4.2 below we review known properties of LRT and SVT ROCs and introduce notation for their parametric formulas with respect to the LRT or SVT threshold value. Our main motivation in doing so is to set the stage for the results presented in Section 2.5 that relate to when an SVT ROC is optimal and how, if it is not, the LRT ROC can be recovered. These results can be extended to more general scalar score variables. However, the analysis is more complicated and does not lead to additional insight, so we do not address this more general case.

2.4.1 LRT ROCs

The LRT ROC associated with a given score variable may be obtained by plotting the values of P_f and P_d resulting from all possible LRT thresholds. Each possible operating point on the LRT ROC is optimal with respect to the MPE criterion for some combination of priors, the minimum Bayes' cost criterion for some combination of prior probabilities and relative costs, and the Neyman-Pearson criterion for some upper bound on the value of P_f. By looking at the entire operating characteristic, we can see the optimal operating points with respect to each of these criteria for all possible sets of prior probabilities, all possible relative costs, and all possible upper bounds on P_f.

It will be advantageous to introduce notation for the parametric formulas of the LRT ROC of a given score variable, where the parameter being varied along the curve

is the LRT threshold value. We define the functions $g_f(\cdot)$ and $g_d(\cdot)$ as

$$P_f^{\mathrm{LRT}} = g_f(\eta) = \int_{\mathcal{D}_{\mathrm{LRT}}(\eta)} ds\, f_0(s) \tag{2.19a}$$

$$P_d^{\mathrm{LRT}} = g_d(\eta) = \int_{\mathcal{D}_{\mathrm{LRT}}(\eta)} ds\, f_1(s). \tag{2.19b}$$

When $\eta = +\infty$, $\mathcal{D}_{\mathrm{LRT}}(\eta)$ is empty and $g_f(\eta) = g_d(\eta) = 0$. At the other extreme when $\eta = 0$, $\mathcal{D}_{\mathrm{LRT}}(\eta)$ contains the entire real line and $g_f(\eta) = g_d(\eta) = 1$. $g_f(\cdot)$ and $g_d(\cdot)$ are always non-increasing in η, since for two threshold values $\eta_0 \leq \eta_1$, $\mathcal{D}_{\mathrm{LRT}}(\eta_1)$ always lies within $\mathcal{D}_{\mathrm{LRT}}(\eta_0)$. Under the current assumptions $g_f(\cdot)$ and $g_d(\cdot)$ are continuous and strictly decreasing, so they are invertible.

Next we briefly review well-known properties of ROCs generated using LRTs and SVTs applied to a given score variable. Ultimately these properties will be helpful in connecting the SVT ROC and LRT ROC of a given score variable, including when they are identical and when, if they are not identical, one can be obtained from the other. The properties are as follows.

(i) The slope of an LRT ROC curve at the point $(P_f^{\mathrm{LRT}}, P_d^{\mathrm{LRT}}) = (g_f(\eta_0), g_d(\eta_0))$ associated with a fixed threshold value η_0 is equal to η_0. That is,

$$\left.\frac{dP_d^{\mathrm{LRT}}}{dP_f^{\mathrm{LRT}}}\right|_{P_f^{\mathrm{LRT}} = g_f(\eta_0)} = \frac{g_d'(\eta_0)}{g_f'(\eta_0)} = \eta_0, \tag{2.20}$$

where $g_f'(\cdot)$ and $g_d'(\cdot)$ denote the derivatives of $g_f(\cdot)$ and $g_d(\cdot)$, respectively.

(ii) LRT ROC curves are concave.

A derivation of Equation (2.20) can be found in many classical decision theory textbooks (see, for example, [39]) and relies mainly on a change of variables in the integrals in Equations (2.19) from an integration over all possible score variable values to an integration over all possible likelihood ratio values. The mathematical details are not relevant to the focus of this article, so we omit them. Property (ii) follows directly from Equation (2.20) in combination with the monotonicity of $g_f(\cdot)$ and $g_d(\cdot)$.

As the LRT threshold value η decreases from $+\infty$ to 0, we move from left to right along the curve and the slope decreases. This is evident in the LRT ROCs shown in Examples 2.5 and 2.6 below. Another way of stating this is that concavity is a *necessary* condition for the Neyman-Pearson optimality of an ROC curve. Under the current assumptions, LRT ROCs are necessarily strictly concave.

2.4.2 SVT ROCs

Another commonly used family of decision regions stems from the somewhat simpler strategy of thresholding the score variable itself, rather than thresholding the likelihood ratio.[2] We will refer to such a strategy as a score variable threshold test or SVT [47]. Each member of the SVT family of decision regions has the form

$$\mathcal{D}_{\mathrm{SVT}}(\gamma) = \{s : s \geq \gamma\} \quad \text{for some real number } \gamma. \tag{2.21}$$

Again, the SVT ROC associated with a given score variable may be obtained by varying the threshold γ over all possible values and plotting the corresponding values of P_f and P_d.

SVTs are especially common in scenarios where ROCs are generated using empirical datasets. In these contexts the score variable is typically a finely-tuned combination of many measurements, possibly computed by applying a machine learning algorithm to a vector of feature values. Thus, it is often less amenable to mathematical analysis and in particular to accurate modeling of the distributions $f_0(\cdot)$ and $f_1(\cdot)$. In principle this does not preclude the use of LRTs, since $f_0(\cdot)$ and $f_1(\cdot)$ can be estimated from histograms derived from training data. However, reliable estimation of probability densities from empirical data is well-known to be a difficult problem [56, 63]. Estimation of the likelihood ratio from empirical data is even more difficult because small errors in the estimate of the denominator of the ratio can lead to large

[2] Of course, we may always redefine the score variable to be the likelihood ratio random variable, i.e., the random variable $S' = f_1(S)/f_0(S)$ where S is the original score variable. An LRT performed with respect to the original score variable may then be reinterpreted as an SVT performed with respect to the new score variable. But this may not be a feasible strategy if the conditional distributions $f_0(\cdot)$ and $f_1(\cdot)$ are inaccessible.

errors in the estimate of the ratio itself. It is in part for this reason that other decision strategies besides LRTs, including SVTs as a particularly common choice, are used in many practical binary hypothesis testing situations.

For an SVT ROC we define the functions $h_f(\cdot)$ and $h_d(\cdot)$ as

$$P_f^{\mathrm{SVT}} = h_f(\gamma) = \int_{\mathcal{D}_{\mathrm{SVT}}(\gamma)} ds\, f_0(s) \tag{2.22a}$$

$$P_d^{\mathrm{SVT}} = h_d(\gamma) = \int_{\mathcal{D}_{\mathrm{SVT}}(\gamma)} ds\, f_1(s). \tag{2.22b}$$

Equation (2.22) can be simplified by defining $F_i(\cdot)$ to be the cumulative distribution function (CDF) of $f_i(\cdot)$,

$$F_i(u) = \int_{-\infty}^{u} ds\, f_i(s), \quad i \in \{0,\, 1\}, \tag{2.23}$$

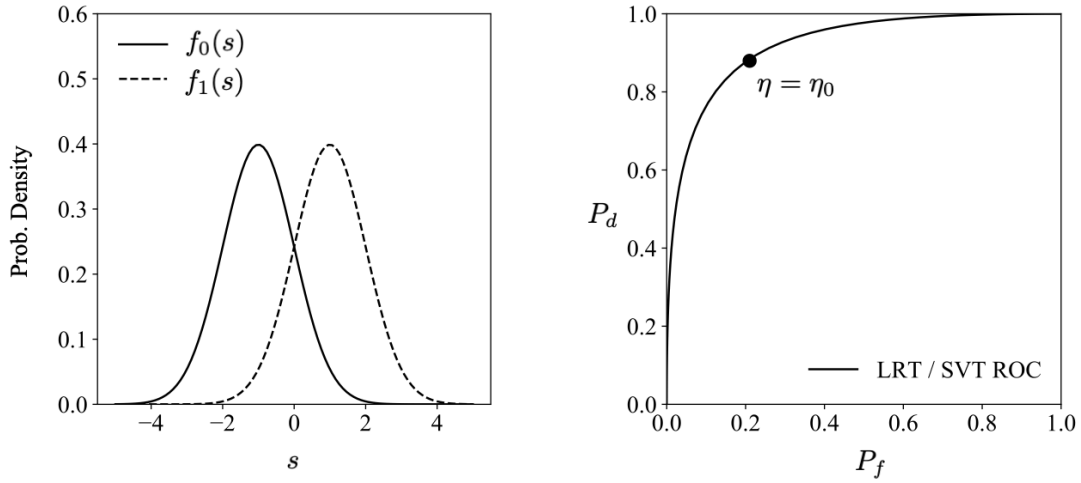
for any real number u. Equation (2.22) can then be rewritten as

$$P_f^{\mathrm{SVT}} = h_f(\gamma) = 1 - F_0(\gamma) \tag{2.24a}$$

$$P_d^{\mathrm{SVT}} = h_d(\gamma) = 1 - F_1(\gamma). \tag{2.24b}$$

When $\gamma = +\infty$, $\mathcal{D}_{\mathrm{SVT}}(\gamma)$ is empty and $h_f(\gamma) = h_d(\gamma) = 0$. When $\gamma = -\infty$, $\mathcal{D}_{\mathrm{SVT}}(\gamma)$ is the whole real line and $h_f(\gamma) = h_d(\gamma) = 1$. Since $F_0(\cdot)$ and $F_1(\cdot)$ are non-decreasing in γ, $h_f(\cdot)$ and $h_d(\cdot)$ are non-increasing in γ. Alternatively, $h_f(\cdot)$ and $h_d(\cdot)$ are non-increasing in γ since for any two thresholds $\gamma_0 \leq \gamma_1$, $\mathcal{D}_{\mathrm{SVT}}(\gamma_1)$ is always contained within $\mathcal{D}_{\mathrm{SVT}}(\gamma_0)$. Under the current assumptions $h_f(\cdot)$ and $h_d(\cdot)$ are strictly decreasing and therefore invertible.

Example 2.5. *Let the conditional distributions $f_0(\cdot)$ and $f_1(\cdot)$ be Gaussian with a common variance σ^2 but different means denoted by μ_0 and μ_1, respectively. An example with $\sigma^2 = 1$, $\mu_0 = -1$, and $\mu_1 = 1$ is shown in Figure 2-2a. For these values of σ, μ_0, and μ_1, the likelihood ratio function $\ell(\cdot) = f_1(\cdot)/f_0(\cdot)$ is strictly monotonic and therefore invertible. As a result, we have $\mathcal{D}_{LRT}(\eta) = \mathcal{D}_{SVT}(\ell^{-1}(\eta))$ and*



(a) (b)

Figure 2-2: (a) Gaussian conditional distributions with variance $\sigma^2 = 1$ and mean $\mu_0 = -1$ or $\mu_1 = 1$. (b) LRT and SVT ROCs, which are identical for these conditional distributions.

$\mathcal{D}_{SVT}(\gamma) = \mathcal{D}_{LRT}(\ell(\gamma))$ *for all* η *and all* γ *and the LRT and SVT ROCs are identical. The LRT ROC is shown in Figure 2-2b. Each point on the curve corresponds to a specific LRT threshold. The parametric formula for the LRT ROC as a function of the LRT threshold are*

$$P_f^{LRT} = g_f(\eta) = 1 - \Phi\left(\frac{\ell^{-1}(\eta) - \mu_0}{\sigma}\right) \tag{2.25a}$$

$$P_d^{LRT} = g_d(\eta) = 1 - \Phi\left(\frac{\ell^{-1}(\eta) - \mu_1}{\sigma}\right), \tag{2.25b}$$

where $\Phi(\cdot)$ *is the CDF of the standard normal distribution. It can be verified through straightforward algebra that for any* $\eta_0 \geq 0$*, we have* $g_d'(\eta_0)/g_f'(\eta_0) = \eta_0$*. The parametric formulas for the SVT ROC as a function of the SVT threshold are*

$$P_f^{SVT} = h_f(\gamma) = 1 - \Phi\left(\frac{\ell(\gamma) - \mu_0}{\sigma}\right) \tag{2.26a}$$

$$P_d^{SVT} = h_d(\gamma) = 1 - \Phi\left(\frac{\ell(\gamma) - \mu_1}{\sigma}\right). \tag{2.26b}$$

Example 2.6. *Let the conditional distributions $f_0(\cdot)$ and $f_1(\cdot)$ be Gaussian with zero mean but different variances denoted by σ_0^2 and σ_1^2, respectively. An example with $\sigma_0^2 = 0.45$ and $\sigma_1^2 = 1.25$ is shown in Figure 2-3a. For these values of σ_0 and σ_1, the likelihood ratio function $\ell(\cdot) = f_1(\cdot)/f_0(\cdot)$ is an even function of s that is strictly decreasing for $s < 0$ and strictly increasing for $s \geq 0$. Since $\ell(\cdot)$ is not invertible, the LRT and SVT ROCs are different as shown in Figure 2-3b. The parametric formulas for the LRT ROC as a function of the LRT threshold are*

$$P_f^{LRT} = g_f(\eta) = 2 - 2\,\Phi\left(\frac{u}{\sigma_0}\right) \tag{2.27a}$$

$$P_d^{LRT} = g_d(\eta) = 2 - 2\,\Phi\left(\frac{u}{\sigma_1}\right) \tag{2.27b}$$

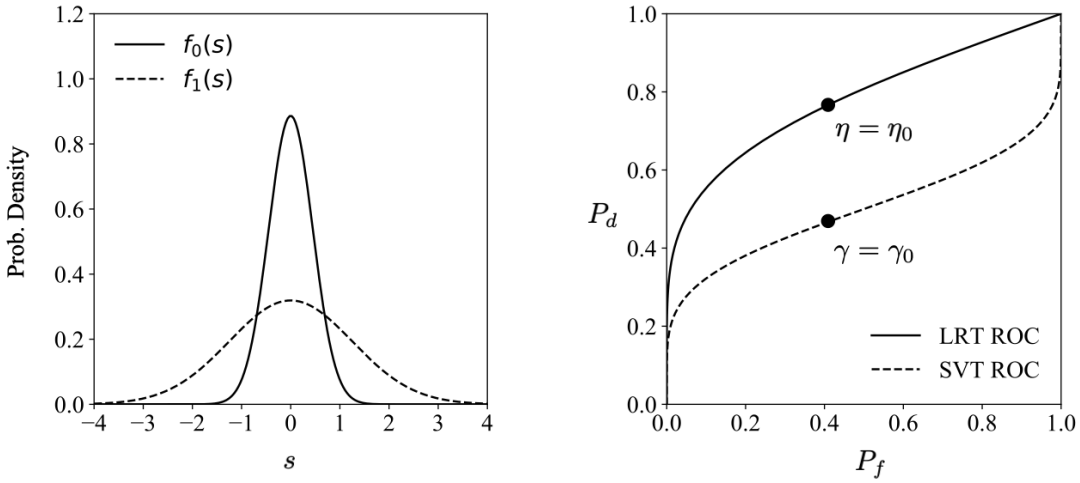
where $u \geq 0$ is the unique non-negative value that satisfies $\ell(u) = \eta$ and $\Phi(\cdot)$ is the CDF of the standard normal distribution. Again it can be verified through straightforward algebra that for any $\eta_0 \geq 0$, we have $g_d'(\eta_0)/g_f'(\eta_0) = \eta_0$. The parametric formulas for the SVT ROC as a function of the SVT threshold are

$$P_f^{SVT} = h_f(\gamma) = 1 - \Phi\left(\frac{\ell(\gamma)}{\sigma_0}\right) \tag{2.28a}$$

$$P_d^{SVT} = h_d(\gamma) = 1 - \Phi\left(\frac{\ell(\gamma)}{\sigma_1}\right). \tag{2.28b}$$

2.5 Relation between LRT and SVT ROCs

For a given score variable, we might be interested in whether or not its SVT ROC is identical to its LRT ROC. When this is the case, we can perform optimal MPE and Neyman-Pearson decision rules by performing SVTs instead of LRTs. This means in particular that we do not need to estimate the conditional distributions of the score variable, nor do we need to estimate the likelihood ratio function. The question is equivalent to asking when any LRT decision region $\mathcal{D}_{\text{LRT}}(\eta)$ can be written as an equivalent SVT decision region $\mathcal{D}_{\text{SVT}}(\gamma)$ and vice versa. It is straightforward to



(a) (b)

Figure 2-3: (a) Gaussian conditional distributions with mean $\mu = 0$ and variance $\sigma_0^2 = 0.45$ or $\sigma_1^2 = 1.25$. (b) LRT and SVT ROCs.

see that this is the case only when the likelihood ratio $f_1(s)/f_0(s)$ is an invertible function of the score variable, because then if we define $\ell(s) = f_1(s)/f_0(s)$ we have $\mathcal{D}_{\text{LRT}}(\eta) = \mathcal{D}_{\text{SVT}}(\ell^{-1}(\eta))$ and $\mathcal{D}_{\text{SVT}}(\gamma) = \mathcal{D}_{\text{LRT}}(\ell(\gamma))$. Of course, this is not true in general.

Since the SVT and LRT ROCs of a given score variable are not necessarily the same, unlike an LRT ROC, there is no reason a priori to assume that an SVT ROC need be concave. An interesting question is whether or not, if the SVT ROC of a given score variable *is* concave, it must be identical to the LRT ROC of that score variable. The answer turns out to be surprisingly simple, relying only on a calculation of the slope of an SVT ROC as a function of the SVT threshold and is addressed in Section 2.5.1.

In Section 2.5.1 we state a condition under which SVT ROCs are guaranteed to be Neyman-Pearson optimal. We describe a procedure that can be used to recover the optimal ROC from a non-optimal SVT ROC, i.e., an SVT ROC for which this condition is not met, in Section 2.5.2.

2.5.1 Optimality of a Concave SVT ROC

A principal result presented in [47] is that if an ROC that was generated using SVTs on a given score variable is concave, then it is guaranteed to be the LRT ROC for that score variable. In other words, concavity is a *sufficient* condition for the Neyman-Pearson optimality of the SVT ROC of a given score variable. To show that this is true, let $h'_f(\cdot)$ and $h'_d(\cdot)$ denote the derivatives of $h_f(\cdot)$ and $h_d(\cdot)$, respectively, as defined in Equation (2.24). A key observation is that due to the properties of the CDFs $\{F_i(\cdot)\}$, we have $h'_f(\gamma) = f_0(\gamma)$ and $h'_d(\gamma) = f_1(\gamma)$. And thus at the point on the curve corresponding to the SVT threshold γ_0, the derivative of the curve is

$$\left.\frac{dP_d^{\mathrm{SVT}}}{dP_f^{\mathrm{SVT}}}\right|_{P_f^{\mathrm{SVT}}=h_f(\gamma_0)} = \frac{h'_d(\gamma_0)}{h'_f(\gamma_0)} = \frac{f_1(\gamma_0)}{f_0(\gamma_0)}. \tag{2.29}$$

Equation (2.29) guarantees that if an SVT ROC is concave, then it must be identical to the LRT ROC of the underlying score variable. This is because if the SVT ROC is concave, then the current assumptions guarantee that it will be strictly concave. Its slope will therefore be an invertible (strictly decreasing) function of P_f^{SVT}. Since P_f^{SVT} is itself an invertible (strictly decreasing) function of γ, the slope of the curve will also be an invertible (strictly increasing) function of γ. But according to Equation (2.29), the slope of the curve as a function of γ is simply equal the likelihood ratio function. In summary, if the SVT ROC is concave then the likelihood ratio function must an invertible function of the SVT threshold, or equivalently an invertible function of the score variable. This implies that the SVT and LRT ROC of the score variable must be identical, proving the result. Note that this result implicitly yields a method for checking whether or not the likelihood ratio function is a monotonic function of the score variable without explicitly computing it for all values of s. Specifically, we may simply generate the SVT ROC and if it is concave, then the likelihood ratio function is necessarily monotonic in the score variable.

The above result is different from the statement in [77], which says that given any concave curve with endpoints at $(0, 0)$ and $(1, 1)$, one can always construct a pair of

conditional distributions for which that curve is the LRT ROC. In that context, the curve and the distributions are strictly abstract and the curve need not have been generated in any particular way relating to the distributions (in fact, it need not have been generated in any particular way at all, i.e. it is essentially just an arbitrary continuous map from the interval $[0, 1]$ to itself). On the other hand, the result stemming from Equation (2.29) says that if the given curve (i) is an ROC that was generated using SVTs on a *specific* pair of distributions associated with a given score variable and (ii) is strictly concave, then the curve is optimal *for those distributions*.

The fact that the SVT ROC of a given score variable is Neyman-Pearson optimal if it is concave leaves open the question of what can be said about a given score variable if its SVT ROC is not concave. In this case the SVT ROC is not identical to the LRT ROC of the score variable and thus it is not Neyman-Pearson optimal. However, as we show in Section 2.5.2, it is still possible to recover the LRT ROC of the score variable from its SVT ROC. Moreover, the recovery does not depend on any knowledge of the conditional distributions of the score variable.

2.5.2 Constructing the Optimal ROC from a Non-Concave SVT ROC

In Section 2.5.2 we develop a procedure for constructing the LRT ROC of a score variable directly from its SVT ROC. It is assumed of course that the SVT ROC is not concave, since otherwise it would already be optimal according to Section 2.5.1. Consider for a moment the scenario where the functions $P_f^{\text{SVT}} = h_f(\gamma)$ and $P_d^{\text{SVT}} = h_d(\gamma)$ are known for all SVT thresholds γ. Equivalently the SVT threshold associated with each point on the SVT ROC is known. A simple way of constructing the LRT ROC would be to differentiate $h_f(\cdot)$ and $h_d(\cdot)$ with respect to γ to recover $f_0(\cdot)$ and $f_1(\cdot)$, respectively. Then LRTs could be directly performed for all LRT thresholds to compute the functions $P_f^{\text{LRT}} = g_f(\eta)$ and $P_d^{\text{LRT}} = g_d(\eta)$. If, on the other hand, P_d^{SVT} is known as a function of P_f^{SVT} but neither one is known as a function of the SVT threshold, i.e., the functions $h_f(\cdot)$ and $h_d(\cdot)$ are unknown, then it is less clear

how to recover the LRT ROC. This scenario is the focus of the current discussion.

An example is shown in Figure 2-4 for concreteness and ease of visualization. The conditional PDFs $f_0(\cdot)$ and $f_1(\cdot)$ shown in the left-hand panel were designed specifically to generate distinctly different SVT and LRT ROCs. For any $\eta_0 \geq 0$, the following procedure allows us to recover $P_f^{\text{LRT}} = g_f(\eta_0)$ and $P_d^{\text{LRT}} = g_d(\eta_0)$. A detailed explanation of the underlying logic follows.

1. Identify the segments of the SVT ROC for which the slope $dP_d^{\text{SVT}}/dP_f^{\text{SVT}}$ is greater than or equal to η_0. These segments are highlighted in green for $\eta_0 = 1$ in Figure 2-5.

2. Add the segments together end-to-end to compute the location of the desired point on the LRT ROC. The point for $\eta_0 = 1$ is marked by a black circle in Figure 2-5. Mathematically, this can be done by recording the changes in P_f^{SVT} and P_d^{SVT} over each segment. Let these changes be denoted by $\Delta P_f^{(j)}$ and $\Delta P_d^{(j)}$ where j is an index over segments. $P_f^{\text{LRT}} = g_f(\eta_0)$ and $P_d^{\text{LRT}} = g_d(\eta_0)$ can be computed by summing the $\Delta P_f^{(j)}$ and $\Delta P_d^{(j)}$ values,

$$P_f^{\text{LRT}} = g_f(\eta_0) = \sum_j \Delta P_f^{(j)} \tag{2.30a}$$

$$P_d^{\text{LRT}} = g_d(\eta_0) = \sum_j \Delta P_d^{(j)}. \tag{2.30b}$$

An explanation of the procedure is shown in Figure 2-5. The graphs in Figure 2-5a show P_f^{SVT}, P_d^{SVT}, and the derivative $dP_d^{\text{SVT}}/dP_f^{\text{SVT}}$ as functions of the score variable. According to Equation (2.29) the derivative is equal to the likelihood ratio function. Note that these graphs are caricatures used only for visualization, since the procedure does not require explicit knowledge of any of the aforementioned quantities as functions of the score variable. A fixed LRT threshold value $\eta_0 \geq 0$ identifies multiple disjoint regions of s for which $f_1(s)/f_0(s) \geq \eta_0$, highlighted in green for $\eta_0 = 1$ in the figure. Together these regions comprise $\mathcal{D}_{\text{LRT}}(\eta_0)$. Each individual region j covers an interval $[a_j, b_j]$ with $a_j < b_j$ and corresponds to the segment in the

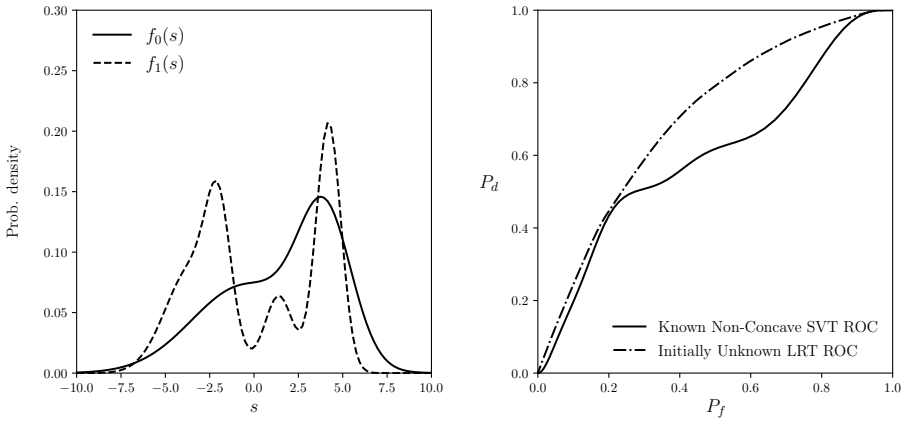


Figure 2-4: Sample conditional PDFs $f_0(\cdot)$ and $f_1(\cdot)$ along with the corresponding SVT and LRT ROCs. Assuming that the SVT ROC is known, the objective is to construct the LRT ROC.

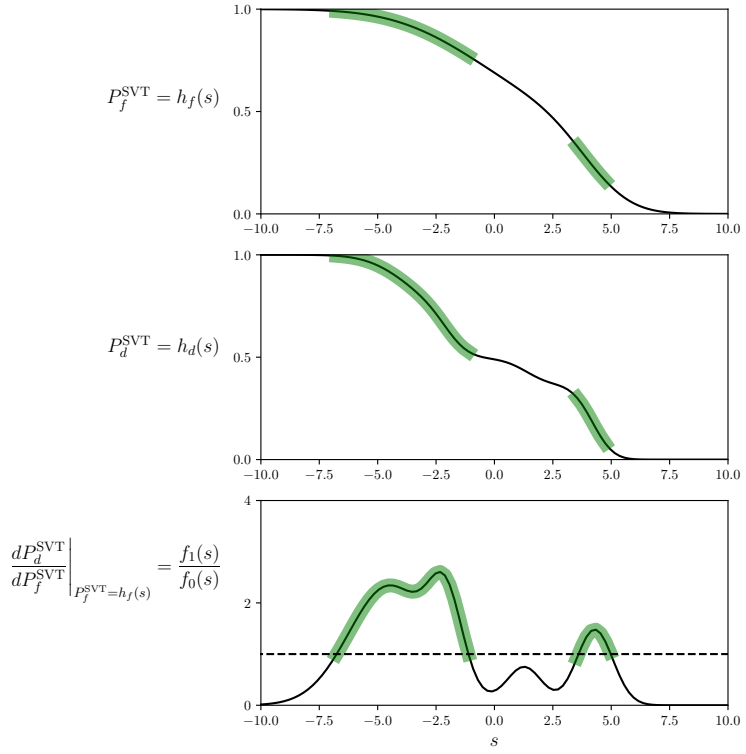
SVT ROC curve with endpoints $(h_f(b_j), h_d(b_j))$ and $(h_f(a_j), h_d(a_j))$. The integrals of $f_0(\cdot)$ and $f_1(\cdot)$ over the region, shown in Figure 2-5b, can be expressed as

$$\int_{a_j}^{b_j} ds\, f_0(s) = (1 - F_0(a_j)) - (1 - F_0(b_j)) = h_f(a_j) - h_f(b_j) \tag{2.31a}$$

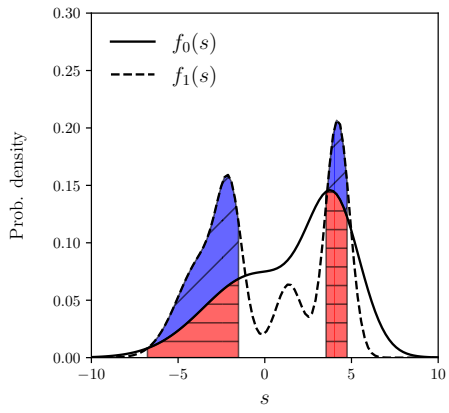
$$\int_{a_j}^{b_j} ds\, f_1(s) = (1 - F_1(a_j)) - (1 - F_1(b_j)) = h_d(a_j) - h_d(b_j) \tag{2.31b}$$

which are simply the changes in P_f^{SVT} and P_d^{SVT} between the endpoints of the segment. Summing these changes over all regions corresponds to summing the integrals of $f_0(\cdot)$ and $f_1(\cdot)$ over each disjoint portion of $\mathcal{D}_{\text{LRT}}(\eta)$. The resulting LRT ROC made by varying η_0 over its entire range is illustrated in Figure 2-5c.

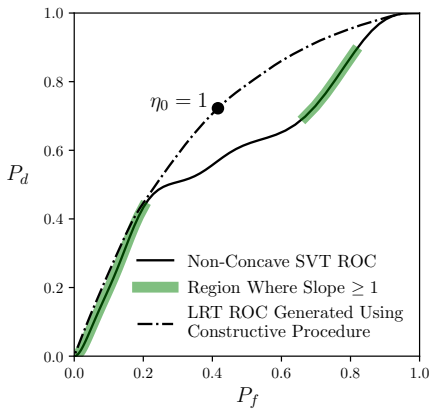
This procedure is different than the use of randomization to replace a convex region on an ROC curve by the straight line connecting its endpoints [78, 60]. In that case, a biased coin is flipped and the result dictates whether the decision region of the first endpoint or that of the second endpoint is used. It is straightforward to show that the effective probabilities of false alarm and detection then lie on the straight line between the endpoints. However, the resulting curve is not Neyman-Pearson optimal. One way of seeing this is to observe that the LRT ROC curve of a continuous score variable, even in the absence of the assumptions made in this paper, can never have any linear regions – it must either be continuous and strictly concave or discontinuous



(a)



(b)



(c)

Figure 2-5: (a) Probability of false alarm, probability of detection, and derivative of SVT ROC as functions of the score variable. The highlighted regions represent regions where the derivative of the curve is greater than or equal to $\eta_0 = 1$. (b) Integrals of the conditional PDFs over the LRT decision region $\mathcal{D}_{\mathrm{LRT}}(\eta_0)$ for $\eta_0 = 1$. (c) Non-concave SVT ROC and LRT ROC generated using the procedure given in the text.

and strictly concave over each of its disjoint regions.

Suppose that for a certain value of $\eta_0 \geq 0$, we wish to not only compute $g_f(\eta_0)$ and $g_d(\eta_0)$ but also to identify the decision region $\mathcal{D}_{\text{LRT}}(\eta_0)$. If the functions $h_f(\cdot)$ and $h_d(\cdot)$ are known then as previously stated, we can simply differentiate them to recover $f_0(\cdot)$ and $f_1(\cdot)$, respectively, and then compute the decision region analytically. But the constructive procedure above also implicitly provides a method for identifying $\mathcal{D}_{\text{LRT}}(\eta_0)$ without requiring explicit computation of the conditional PDFs or their ratio. Specifically, we may plot the derivative of the SVT ROC curve as a function of the SVT threshold as shown in Figure 2-5a and then read the decision region $\mathcal{D}_{\text{LRT}}(\eta_0)$ directly off the graph by checking where the derivative is greater than or equal to η_0.

Chapter 3

A Perspective on Linear Algebra and Frame Representations

In Chapter 2 we discussed a framework for general binary hypothesis testing systems and then elaborated on two ways of designing the decision region of the binary decision rule – using an SVT or an LRT. By contrast, Chapters 4 through 6 will be directed more towards the design of the pre-decision operator, and specifically for the problem of quantum binary state discrimination. A thorough and precise statement of the problem requires use of the postulates of quantum mechanics, which themselves are often phrased using language and concepts stemming from linear algebra. In particular, according to the postulates of quantum mechanics, quantum states and measurements can be mathematically represented through the vectors and operators associated with a given Hilbert space. The Hilbert spaces considered in this thesis are assumed to be finite-dimensional for simplicity, and as such many of the results rely heavily on linear algebra as it pertains to finite-dimensional Hilbert spaces. The purpose of Chapter 3 aside from establishing our notation and terminology is to summarize our viewpoint on this topic by reviewing a selection of key results that are used in Chapters 4 through 6. Of particular importance are the concepts and results related to frame representations of a given finite-dimensional Hilbert space.

We review basic concepts related to finite-dimensional Hilbert spaces, inner products, and the concept of the Hermitian adjoint of a linear transformation in Section

3.1. A crucial concept is the relationship of the range and nullspace of a given linear transformation with those of its adjoint. In Section 3.2 we give our perspective on a selection of core concepts from the field of frame theory. The technical results reviewed in Section 3.2 have been long established, but our viewpoint on them, which involves lifting a vector in one Hilbert space to a different vector in a larger space using a given frame representation, is less traditional. Our main motivation in discussing frame representations is to apply the concepts to operator-valued Hilbert spaces as described in Section 3.4. The robustness of frame representations to certain types of error affecting the reconstruction of an unknown vector is described in Section 3.5. While the purpose of Chapter 3 is to set the stage for the discussions of binary hypothesis testing for quantum systems in Chapters 4 through 6, but none of the concepts discussed are specific to quantum mechanics.

3.1 Hilbert Spaces

By definition a Hilbert space is a vector space that is equipped with an inner product and that is complete.[1] We will denote by $\mathcal{V}$ and $\mathcal{W}$ two Hilbert spaces with dimensions $\dim \mathcal{V} = N$ and $\dim \mathcal{W} = M$. Several key results of the thesis concern the dependence of certain performance metrics on the value of M for fixed N. For simplicity we will assume that both $\mathcal{V}$ and $\mathcal{W}$ are equipped with the same inner product. It will always be assumed that the corresponding field F is either the real numbers $\mathbb{R}$ or the complex numbers $\mathbb{C}$. We have arbitrarily chosen to phrase much of the following discussion only in terms of $\mathcal{V}$, but we emphasize that the concepts apply to any finite-dimensional Hilbert space.

To be consistent with the relevant quantum mechanics literature we will use Dirac's bra-ket notation, in which a vector in $\mathcal{V}$ is represented by a ket (for example, $|v\rangle$) and its conjugate transpose is represented by a bra (for example, $\langle v|$).

[1] Completeness means that if the sum of the lengths of an infinite sequence of vectors in $\mathcal{V}$ converges to some finite number, then the sum of the vectors themselves converge to a vector in $\mathcal{V}$, $\sum_{i=1}^{\infty} ||v_i|| < \infty$ implies $\sum_{i=1}^{\infty} |v_i\rangle \in \mathcal{V}$. This is sometimes informally described as $\mathcal{V}$ having "no holes". For brevity we do not expound on the concept of completeness further.

For a specified basis $\{|u_n\rangle, 1 \leq n \leq N\}$ for $\mathcal{V}$, we will occasionally use the notation $|v\rangle = [c_1, \ldots, c_N]^T$ as shorthand to indicate that $|v\rangle = \sum_n c_n |u_n\rangle$. Unless otherwise noted the inner product between two vectors $|v_1\rangle, |v_2\rangle \in \mathcal{V}$ will be denoted by $\langle v_1|v_2\rangle$ and the squared norm of a vector $|v\rangle \in \mathcal{V}$ will be defined as the inner product of $|v\rangle$ with itself, denoted by $||v||^2 = \langle v|v\rangle$. The angle θ between two vectors $|v_1\rangle, |v_2\rangle \in \mathcal{V}$ is defined via the relation $||v_1||\,||v_2||\cos\theta = \langle v_1|v_2\rangle$. Inner products are discussed in more detail in Section 3.1.2.

3.1.1 Linear Transformations

A linear transformation $T : \mathcal{V} \to \mathcal{W}$ is a mapping from vectors in $\mathcal{V}$ to vectors in $\mathcal{W}$ with the property that

$$|T(a\,v_1 + b\,v_2)\rangle = a\,|T(v_1)\rangle + b\,|T(v_2)\rangle \tag{3.1}$$

for all $|v_1\rangle, |v_2\rangle \in \mathcal{V}$ and all $a, b \in F$. Note that when a vector $|v\rangle$ is used as the input to a linear transformation T, the output will denoted interchangeably by $T\,|v\rangle$, $|Tv\rangle$, or $|T(v)\rangle$. When $\mathcal{W} = \mathcal{V}$, T is typically referred to as a linear operator on $\mathcal{V}$. The nullspace and range of T, denoted by $N(T)$ and $R(T)$, respectively, are defined as

$$N(T) = \{\, |v\rangle \in \mathcal{V} \,|\, T(v) = 0 \,\} \tag{3.2a}$$

$$R(T) = \{\, T(v) \text{ for all } |v\rangle \in \mathcal{V} \,\}. \tag{3.2b}$$

It can be shown that $N(T)$ has the properties required to be a subspace of $\mathcal{V}$ and similarly $R(T)$ is a subspace of $\mathcal{W}$. If dim $R(T) = d$ for some $0 \leq d \leq M$, then T is said to have rank d. A fundamental result in linear algebra states that dim $N(T)$ + dim $R(T) = M$.

It is often of great value to analyze vectors in $\mathcal{V}$ in terms of their unique components in multiple subspaces. This notion can be formalized using the concept of a direct sum decomposition. Given two subspaces $\mathcal{U}_1$ and $\mathcal{U}_2$ of $\mathcal{V}$, $\mathcal{V}$ can be written as the direct sum $\mathcal{V} = \mathcal{U}_1 \oplus \mathcal{U}_2$ if every $|v\rangle \in \mathcal{V}$ can be written uniquely as the sum of

two components with one in each subspace,

$$|v\rangle = |u_1\rangle + |u_2\rangle\,, \quad |u_1\rangle \in \mathcal{U}_1, \quad |u_2\rangle \in \mathcal{U}_2. \tag{3.3}$$

Consider the pair of linear operators $\mathcal{P}_1$ and $\mathcal{P}_2$ on $\mathcal{V}$ defined to satisfy

$$\mathcal{P}_1\,|v\rangle = |u_1\rangle\,, \quad \mathcal{P}_2\,|v\rangle = |u_2\rangle\,, \tag{3.4}$$

for all $|v\rangle \in \mathcal{V}$, where $|u_1\rangle$ and $|u_2\rangle$ are defined as in Equation (3.3). For $i \in \{1, 2\}$, $\mathcal{P}_i$ extracts the component of $|v\rangle$ lying in $\mathcal{U}_i$ with respect to the decomposition $\mathcal{V} = \mathcal{U}_1 \oplus \mathcal{U}_2$ and is referred to as a projection operator – or projector for short – onto $\mathcal{U}_i$. More generally, a projection operator is defined as any idempotent linear operator $\mathcal{P}$, i.e., a linear operator satisfying $\mathcal{P}^2 = \mathcal{P}$, on $\mathcal{V}$. An operator with these properties is a projector onto its range $R(\mathcal{P})$ and we always have $\mathcal{V} = R(\mathcal{P}) \oplus N(\mathcal{P})$.

It is important to note that for a given subspace $\mathcal{U}$ of $\mathcal{V}$, there are many different projection operators onto $\mathcal{U}$ with each one corresponding to a different decomposition of $\mathcal{V}$ as $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}'$ for some other subspace $\mathcal{U}'$. When $\mathcal{U}'$ is the orthogonal complement of $\mathcal{U}$, denoted as $\mathcal{U}' = \mathcal{U}^\perp$, each one of the corresponding pair of projectors satisfy the definition of a so-called orthogonal projector and together they are said to form a complete set of orthogonal projectors. The precise definition of orthogonal projectors is stated in Section 3.1.3 as it relies on the definition of the Hermitian adjoint of a linear transformation.

The concept of a direct sum decomposition along with its corresponding set of projection operators generalizes in a straightforward manner to sets of up to N subspaces. For example, assume that $\{|u_n\rangle\,, 1 \leq n \leq N\}$ is a basis for $\mathcal{V}$ and let $\{\mathcal{U}_n\}$ be the set of one-dimensional subspaces spanned by the individual basis vectors. A given vector $|v\rangle \in \mathcal{V}$ can always be expressed as a sum of a unique set of N components, with one component lying in each of the $\{\mathcal{U}_n\}$. Thus $\mathcal{V} = \mathcal{U}_1 \oplus \cdots \oplus \mathcal{U}_N$.

3.1.2 Inner Products

A given vector space may be equipped with multiple inner products, and the one used in technical analysis is important because it defines the geometry of the space in some sense. The notion of the angle between two vectors and the definition of orthogonality, for instance, both stem directly from the definition of the inner product. By definition an inner product is a mapping $G : \mathcal{V} \times \mathcal{V} \to F$ that satisfies the following properties for all $|v_1\rangle, |v_2\rangle, |v_3\rangle \in \mathcal{V}$ and for all $a, b \in F$,

$$G(v_1, v_2) = G(v_2, v_1)^*, \tag{3.5a}$$

$$G(a\,v_1 + b\,v_2, v_3) = a\,G(v_1, v_3) + b\,G(v_2, v_3), \tag{3.5b}$$

$$G(v_1, a\,v_2 + b\,v_3) = a^*\,G(v_1, v_2) + b^*\,G(v_1, v_3), \tag{3.5c}$$

$$G(v_1, v_1) > 0 \text{ whenever } |v_1\rangle \neq 0. \tag{3.5d}$$

where the superscript $*$ signifies complex conjugation. If F is equal to $\mathbb{R}$, then complex conjugation can of course be disregarded. As with linear transformations, note that for two vectors $|v_1\rangle, |v_2\rangle \in \mathcal{V}$ we will typically write $G(|v_1\rangle, |v_2\rangle)$ as simply $G(v_1, v_2)$ for the sake of notational clarity. For a given inner product $G(\cdot, \cdot)$, the squared norm of a vector $|v\rangle$ is defined to be the inner product of $|v\rangle$ with itself, $G(v, v)$. The angle θ between two vectors $|v_1\rangle, |v_2\rangle \in \mathcal{V}$ is defined via the relation $\sqrt{G(v_1, v_1)\,G(v_2, v_2)}\cos\theta = G(v_1, v_2)$. $|v_1\rangle$ and $|v_2\rangle$ are said to be orthogonal if $G(v_1, v_2) = 0$. A given basis for $\mathcal{V}$, that is, any set of linearly independent vectors that span $\mathcal{V}$, is referred to as orthonormal if each of the basis vectors has unit norm and if they are collectively pairwise orthogonal.

In Section 3.2 we will make use of the following convenient fact involving the definition of multiple inner products on the same Hilbert space. In the absence of any other constraints, we can always assume that an arbitrary basis $\{|u_n\rangle, 1 \leq n \leq N\}$ for $\mathcal{V}$ is orthonormal with respect to the $\langle\cdot|\cdot\rangle$ inner product. If it were not, we could always construct an inner product $G(\cdot, \cdot)$ under which the $\{|u_n\rangle\}$ were orthonormal along with an invertible function relating $G(\cdot, \cdot)$ to $\langle\cdot|\cdot\rangle$. To see why this is true,

let $\{|u_n\rangle\}$ and $\{|e_n\rangle\}$ be two different bases for $\mathcal{V}$ and assume that the $\{|e_n\rangle\}$ are orthonormal with respect to the $\langle\cdot|\cdot\rangle$ inner product but that the $\{|u_n\rangle\}$ are not. Thus we have $\langle e_m|e_n\rangle = \delta_{mn}$ for all $1 \leq m, n \leq N$, where δ_{mn} takes the value 1 if $m = n$ and 0 otherwise. Consider a mapping $G : \mathcal{V} \times \mathcal{V} \to F$ that is defined to satisfy

$$G(u_m, u_n) = \delta_{mn} \tag{3.6}$$

for all $1 \leq m, n \leq N$ and also to satisfy Equations (3.5a) to (3.5c). These constraints together imply that $G(\cdot,\cdot)$ also satisfies Equation (3.5d). Therefore, $G(\cdot,\cdot)$ is a valid inner product on $\mathcal{V}$ and the $\{|u_n\rangle\}$ are orthonormal with respect to this inner product. Let $|v\rangle$ be an arbitrary nonzero vector in $\mathcal{V}$, then $|v\rangle$ can always be written as $|v\rangle = \sum_n c_n |u_n\rangle$ for some combination of scalars $\{c_n\}$, at least one of which is nonzero. We have

$$G(v,v) = G\left(\sum_{m=1}^{N} c_m |u_m\rangle, \sum_{n=1}^{N} c_n |u_n\rangle\right) = \sum_{m,n=1}^{N} c_m^* c_n G(u_m, u_n) = \sum_{m,n=1}^{N} |c_n|^2 > 0. \tag{3.7}$$

To relate $G(\cdot,\cdot)$ back to the $\langle\cdot|\cdot\rangle$ inner product, we now show that there is an invertible linear operator L on $\mathcal{V}$ with the property that for all $|v_1\rangle, |v_2\rangle \in \mathcal{V}$,

$$G(L|v_1\rangle, L|v_2\rangle) = \langle v_1|v_2\rangle . \tag{3.8}$$

It is straightforward to show that to satisfy Equation (3.8) it is sufficient to have $G(L|u_m\rangle, L|u_n\rangle) = \langle u_m|u_n\rangle$ for all $1 \leq i, j \leq N$. The derivation relies on the fact that any $|v_1\rangle, |v_2\rangle \in \mathcal{V}$ can be written as linear combinations of the $\{|u_n\rangle\}$ and on the properties of $G(\cdot,\cdot)$ and $\langle\cdot|\cdot\rangle$. To find a suitable operator L we first express $|u_i\rangle$ and $|u_j\rangle$ for any $1 \leq i, j \leq N$ in terms of the $\{|e_n\rangle\}$, $|u_i\rangle = \sum_m c_m |e_m\rangle$ and $|u_j\rangle = \sum_n d_n |e_n\rangle$. $G(L|u_i\rangle, L|u_j\rangle)$ can be written as

$$G(L|u_i\rangle, L|u_j\rangle) = G\left(\sum_{m=1}^{N} c_m L|e_m\rangle, \sum_{n=1}^{N} d_n L|e_n\rangle\right) = \sum_{m,n=1}^{N} c_m^* d_n G(L|e_m\rangle, L|e_n\rangle) \tag{3.9}$$

and $\langle u_i|u_j\rangle$ can be written as

$$\langle u_i|u_j\rangle = \left\langle \sum_{m=1}^{N} c_m \left|e_m\right\rangle \middle| \sum_{n=1}^{N} d_n \left|e_n\right\rangle \right\rangle = \sum_{k=1}^{N} c_n^* \, d_n. \tag{3.10}$$

Note that in Equation (3.10) we have used the orthonormality of the $\{|e_n\rangle\}$ with respect to the $\langle\cdot|\cdot\rangle$ inner product. For the two to be equal, it is sufficient to have $G(L\left|e_m\right\rangle, L\left|e_n\right\rangle) = \delta_{mn}$ for all $1 \leq m, n \leq N$. A suitable linear operator can be defined by the relation

$$L\left|e_n\right\rangle = \left|u_n\right\rangle \tag{3.11}$$

for all $1 \leq n \leq N$. Because the $\{|e_n\rangle\}$ and $\{|u_n\rangle\}$ are both bases for $\mathcal{V}$, L is clearly invertible.

3.1.3 Hermitian Adjoints

The concept of the Hermitian adjoint of a linear transformation is fundamental to many applications of linear algebra [4, 74]. In what follows we review the definition of the adjoint as well as several of its key properties. A crucial takeaway is a set of relations that connect the range and nullspace of a linear transformation to those of its adjoint. Let $T : \mathcal{V} \to \mathcal{W}$ be a linear transformation. The Hermitian adjoint – or adjoint for short – of T is defined as a linear transformation $T^\dagger : \mathcal{W} \to \mathcal{V}$ satisfying

$$\langle w|Tv\rangle = \langle T^\dagger w|v\rangle \tag{3.12}$$

for all $|v\rangle \in \mathcal{V}$ and all $|w\rangle \in \mathcal{W}$. It is always unique for a given inner product under the assumptions made in this thesis. When T is a linear operator and $T = T^\dagger$, T is referred to as a Hermitian or self-adjoint operator. The adjoint of the adjoint is always the original transformation, i.e., $(T^\dagger)^\dagger = T$. This follows from taking the complex conjugate of both sides of Equation (3.12) and applying Equation (3.5a), leading to $\langle Tv|w\rangle = \langle v|T^\dagger w\rangle$. Given a linear transformation $T = \sum_m |y_m\rangle\langle x_m|$ where the $\{|x_m\rangle\}$ are elements of $\mathcal{V}$ and the $\{|y_m\rangle\}$ are elements of $\mathcal{W}$, it is straightforward to

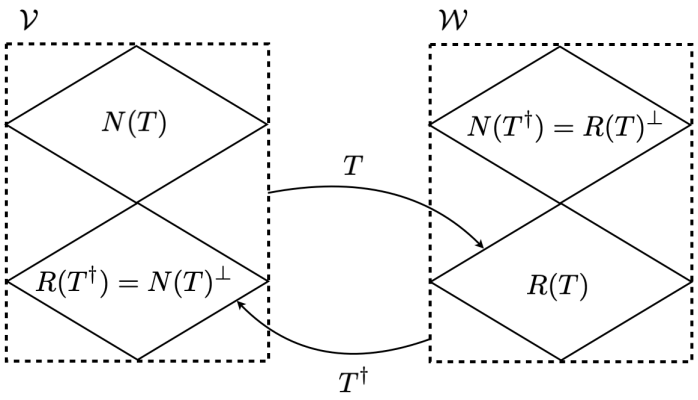


Figure 3-1: Relationship between the ranges and nullspaces of a linear transformation $T : \mathcal{V} \to \mathcal{W}$ and its adjoint $T^\dagger : \mathcal{W} \to \mathcal{V}$.

show that the adjoint $T^\dagger$ can always be written as $T^\dagger = \sum_m |x_m\rangle \langle y_m|$.

A fundamental result of linear algebra states that $N(T^\dagger) = R(T)^\perp$. Since $(T^\dagger)^\dagger = T$ this implies by symmetry that $N(T) = R(T^\dagger)^\perp$. These relations are illustrated in Figure 3-1 and can be derived as follows. We first show that $N(T^\dagger) \subset R(T)^\perp$. Assuming that $|w_1\rangle$ is an arbitrary element of $N(T^\dagger)$, we wish to show that $|w_1\rangle$ is orthogonal to all elements of $R(T)$. Let $|w_2\rangle$ be an arbitrary element of $R(T)$, then by definition there is some $|v\rangle \in \mathcal{V}$ such that $|w_2\rangle = T|v\rangle$. $|w_1\rangle$ must be orthogonal to $|w_2\rangle$ because

$$\langle w_1|w_2\rangle = \langle Tv|w_2\rangle = \langle v|T^\dagger w_2\rangle = 0. \tag{3.13}$$

We now show that $R(T)^\perp \subset N(T^\dagger)$, and this proves that $N(T^\dagger) = R(T)^\perp$. Assuming that $|w\rangle$ is an arbitrary element of $R(T)^\perp$, we wish to show that it must also be an element of $N(T^\dagger)$. By definition $|w\rangle$ must be orthogonal to all elements of $R(T)$, i.e., $\langle Tv|w\rangle = 0$ for all $|v\rangle \in \mathcal{V}$. This expression can be rewritten as

$$\langle Tv|w\rangle = \langle v|T^\dagger w\rangle = 0 \tag{3.14}$$

for all $|v\rangle \in \mathcal{V}$, and this implies that $T^\dagger|w\rangle = 0$, i.e., $|w\rangle \in N(T^\dagger)$.

As mentioned in Section 3.1.1, a projection operator on $\mathcal{V}$ is defined as any linear, idempotent operator on $\mathcal{V}$. An operator $\mathcal{P}$ with these properties is a projector onto its range $R(\mathcal{P})$ and we always have $\mathcal{V} = R(\mathcal{P}) \oplus N(\mathcal{P})$. By definition an orthogonal

projection operator is a projection operator that is also Hermitian, $\mathcal{P}^\dagger = \mathcal{P}$. And in this case it is straightforward to show that we always have $N(\mathcal{P}) = R(\mathcal{P})^\perp$, so the corresponding direct sum decomposition of $\mathcal{V}$ utilizes two orthogonal subspaces.

3.2 Frame Representations

As is well-known, a given vector in a finite-dimensional Hilbert space can always be represented in terms of its coefficients with respect to a fixed basis for the space and a basis expansion corresponds to a complete representation of each vector. Frames are a generalization of bases allowing for an overcomplete representation of a vector in the space as a linear combination of linearly dependent vectors. In effect, the coefficients in an overcomplete frame expansion can be viewed as corresponding to multiple linear combinations of the coefficients in a basis expansion. Among the advantages of an overcomplete representation is the redundancy of the information in the coefficients representing the vector. Consequently frames and frame representations often provide an important mechanism for describing, analyzing and implementing robust vector representations that are less sensitive to errors in the coefficients representing the vectors. Constructing an overcomplete representation can be as simple as replicating each basis vector multiple times, but there are of course a variety of more strategic ways of introducing and exploiting redundancy. Extensive research has been devoted to this topic and its many extensions in the field of frame theory [13, 16, 17, 41, 42]. We emphasize that the many core concepts of frame theory that are not essential to the discussions of subsequent chapters are not included here. For a comprehensive review of the foundations of frame theory, we refer the reader to, for example, [13, 16, 17].

Throughout Section 3.2 we assume that $\mathcal{V}$ is a subspace of $\mathcal{W}$, implying that $\dim \mathcal{V} \leq \dim \mathcal{W}$, i.e., $N \leq M$. We will denote the orthogonal projection operator from $\mathcal{W}$ onto $\mathcal{V}$ by $\mathcal{P}_\mathcal{V}$. The notation $\{|f_k\rangle, 1 \leq k \leq M\}$ will always be used to denote a frame for $\mathcal{V}$ as defined below and the notation $\{|w_k\rangle, 1 \leq k \leq M\}$ will always be used to refer to an orthonormal basis for $\mathcal{W}$. As explained in Section 3.1.2, in the absence of any other assumptions no generality is lost by assuming that the

$\{|w_k\rangle\}$ are orthonormal. Note that while the $\{|w_k\rangle\}$ are a basis for $\mathcal{W}$ they are in general neither a basis nor a frame for $\mathcal{V}$, since not every linear combination of them necessarily lies in $\mathcal{V}$.

3.2.1 Definition of a Frame

Consider any set of M vectors $\{|f_k\rangle, 1 \leq k \leq M\}$ that lie in and span $\mathcal{V}$ but that are not necessarily linearly independent. Since $\mathcal{V}$ is finite-dimensional, any set of vectors with these properties form what is referred to as a frame for $\mathcal{V}$. More generally, an M-element frame for $\mathcal{V}$ is defined as any set of M vectors $\{|f_k\rangle\}$ in $\mathcal{V}$ that satisfy

$$C\,||v||^2 \leq \sum_{k=1}^{M} |\,\langle f_k|v\rangle\,|^2 \leq D\,||v||^2, \quad \forall\,|v\rangle \in \mathcal{V} \tag{3.15}$$

for some $0 < C \leq D < \infty$ [16]. When C and D are set to form the tightest possible bounds, they are typically referred to as upper and lower frame bounds of $\{|f_k\rangle\}$, respectively. The requirement that $C > 0$ ensures that the frame vectors span $\mathcal{V}$. Equation (3.15) can additionally be extended to include continuous frames and frames with a countably infinite number of elements but for simplicity, in this thesis we will only address the case where there are a finite number M of frame vectors. Unlike in finite dimensions, in infinite dimensions Equation (3.15) is not necessarily satisfied by any set of vectors that lie in and span $\mathcal{V}$.

3.2.2 Analysis and Synthesis Operators and Maps

Associated with any frame for $\mathcal{V}$ are two linear transformations referred to as the analysis and synthesis operators of the frame [13]. The analysis operator A takes as its input any $|v\rangle \in \mathcal{V}$ and generates a set of frame coefficients defined by $\{a_k = \langle f_k|v\rangle, 1 \leq k \leq M\}$. The synthesis operator F takes as its input any set $\{c_k, 1 \leq k \leq M\}$ of coefficients and produces as its output the vector $\sum_k c_k\,|f_k\rangle \in \mathcal{V}$. The $\{c_k\}$ used as input to the synthesis operator do not necessarily need to have been obtained by applying the analysis operator to some $|v\rangle \in \mathcal{V}$. Indeed, there may not be any $|v\rangle \in \mathcal{V}$

such that $c_k = \langle f_k|v\rangle$ for $1 \leq k \leq M$.

In this thesis we take a particular perspective that requires the definition of the following two linear operators on $\mathcal{W}$, derived from A and F and denoted by A_0 and F_0. Since A_0 and F_0 are closely related to A and F but are not strictly identical, we refer to them as the analysis and synthesis maps of the frame to avoid ambiguity. The analysis map A_0 maps any vector $|v\rangle \in \mathcal{V}$ to a specific vector $|w\rangle \in \mathcal{W}$ according to the relation

$$|v\rangle \in \mathcal{V} \longrightarrow |w\rangle = A_0 |v\rangle = \sum_{k=1}^{M} \langle f_k|v\rangle |w_k\rangle = \sum_{k=1}^{M} a_k |w_k\rangle \in \mathcal{W}. \tag{3.16}$$

Because the $\{|f_k\rangle\}$ span $\mathcal{V}$, A_0 has rank N, $R(A_0)$ is an N-dimensional subspace of $\mathcal{W}$ and $N(A_0) = \mathcal{V}^{\perp}$ is an $(M-N)$-dimensional subspace of $\mathcal{W}$. An important implication is that A_0 is left-invertible, meaning that there is a linear operator (not necessarily unique) that recovers $|v\rangle$ from $A_0 |v\rangle$ for every $|v\rangle \in \mathcal{V}$. As explained in Section 3.2.3, each left-inverse of A_0 is connected to one member of a collection of frames referred to as the dual frames of $\{|f_k\rangle\}$. Since the $\{|w_k\rangle\}$ are orthonormal we have $||A_0 |v\rangle ||^2 = \sum_k |a_k|^2$, so the definition of a frame given in Equation 3.15 can be rewritten as

$$C ||v||^2 \leq ||A_0 |v\rangle ||^2 \leq D ||v||^2, \forall |v\rangle \in \mathcal{V}. \tag{3.17}$$

For a given frame $\{|f_k\rangle\}$ for $\mathcal{V}$, the analysis map A_0 can always be expressed as

$$A_0 = \sum_{k=1}^{M} |w_k\rangle \langle f_k| . \tag{3.18}$$

The action of A_0 is summarized schematically in Figure 3-2. Both sides of the diagram represent decompositions of $\mathcal{W}$ into a direct sum of two orthogonal subspaces, $\mathcal{W} = \mathcal{V} \oplus \mathcal{V}^{\perp}$ and $\mathcal{W} = R(A_0) \oplus R(A_0)^{\perp}$.

The synthesis map F_0 maps any vector $|w\rangle \in \mathcal{W}$ to a specific vector $|v\rangle \in \mathcal{V}$ in a way that relies on the basis coefficients of $|w\rangle$ with respect to the $\{|w_k\rangle\}$ basis.

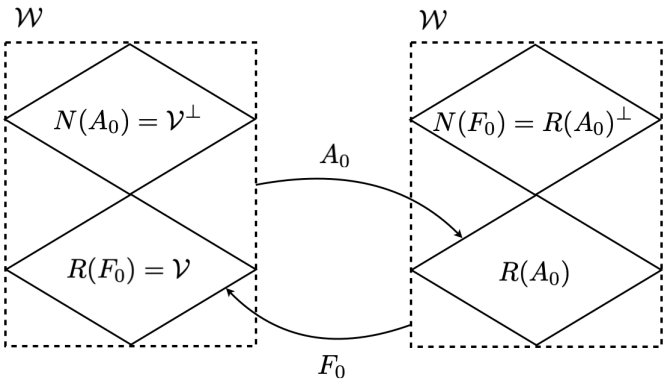


Figure 3-2: The analysis map A_0 takes vectors in $\mathcal{V}$ to a (possibly) different subspace of $\mathcal{W}$ with the same dimension as $\mathcal{V}$. It takes vectors in $\mathcal{V}^\perp$ to the zero vector. The synthesis map F_0 takes vectors in $R(A_0)$ to the subspace $\mathcal{V}$. It takes vectors in $R(A_0)^\perp$ to the zero vector.

Specifically, F_0 is defined by the relation

$$|w\rangle = \sum_{k=1}^{M} c_k |w_k\rangle \in \mathcal{W} \longrightarrow |v\rangle = F_0 |w\rangle = \sum_{k=1}^{M} c_k |f_k\rangle \in \mathcal{V}. \tag{3.19}$$

Since the $\{|w_k\rangle\}$ are orthonormal, the basis coefficients $\{c_k\}$ can be expressed as $c_k = \langle w_k|w\rangle$ for $1 \leq k \leq M$ and F_0 can always be expressed as

$$F_0 = \sum_{k=1}^{M} |f_k\rangle \langle w_k| . \tag{3.20}$$

The action of F_0 is also shown in Figure 3-2. Since the $\{|f_k\rangle\}$ span $\mathcal{V}$, the range of F_0 is $R(F_0) = \mathcal{V}$. If the $\{|f_k\rangle\}$ are linearly dependent, then F_0 has a non-trivial nullspace $N(F_0)$ that is an $(M-N)$-dimensional subspace of $\mathcal{W}$. It is easily verified that $F_0 = A_0^\dagger$ and thus $A_0 = F_0^\dagger$.[2] As noted in Section 3.1.3 this implies that $N(A_0) = R(F_0) \perp$ and $N(F_0) = R(A_0)^\perp$.

For some frames, the synthesis map can also be written as $F_0 = \sum_k |f_k\rangle \langle g_k|$ for a set of basis vectors $\{|g_k\rangle\}$ for $\mathcal{W}$ that is different from $\{|w_k\rangle\}$. This implies that we could have started with $\{|g_k\rangle\}$ as a basis for $\mathcal{W}$ instead of with $\{|w_k\rangle\}$, and we would

[2]The analysis and synthesis operators are also adjoints of each other, $F^\dagger = A$ and $A^\dagger = F$.

have arrived at the same synthesis map. To see why this is true, note that instead of defining F_0 using Equation (3.19) we could equivalently define it according to the relation

$$F_0 \left|w_k\right\rangle = \left|f_k\right\rangle, \quad 1 \leq k \leq M. \tag{3.21}$$

Equation (3.19) then follows by linearity. For Equation (3.21) to be true the $\{|g_j\rangle\}$ must satisfy the relation

$$\sum_{j=1}^{M} \left|f_j\right\rangle \left\langle g_j|w_k\right\rangle = \left|f_k\right\rangle, \quad 1 \leq k \leq M. \tag{3.22}$$

We can then expand the $\{|f_j\rangle\}$ and $\{|g_j\rangle\}$ as linear combinations of the $\{|w_j\rangle\}$ to arrive at a system of linear equations in which the unknowns are the basis coefficients of the $\{|g_j\rangle\}$. Depending on the frame, the equations may or may not have multiple solutions. Another way to look at it is to note that if the $\{|f_k\rangle\}$ are linearly dependent, then there is a linear combination of them that is equal to zero. Let $\{c_k\}$ be a set of coefficients such that $\sum_k c_k \left|f_k\right\rangle = 0$. Then to satisfy Equation (3.22) it is sufficient to have

$$\left\langle g_j|w_k\right\rangle = \delta_{jk}\, c_k, \quad 1 \leq j, k \leq M. \tag{3.23}$$

This is again a system of linear equations that may or may not have more than one solution depending on the frame.

We emphasize that the operators A_0 and F_0 have been introduced primarily for the purpose of providing us with a convenient interpretation of the analysis and synthesis operations of the frame $\{|f_k\rangle\}$ as operations acting on the larger space $\mathcal{W}$. By definition A_0 and F_0 implicitly depend on our choice of basis vectors $\{|w_k\rangle\}$ and we are free to choose the $\{|w_k\rangle\}$ in such a way that the interpretation of A_0 and F_0 is simplified as much as possible. Note that an arbitrary linear operator T acting on $\mathcal{W}$ that has rank N and satisfies $N(T) = \mathcal{V}^{\perp}$ can always be written in the form of Equation (3.18) and can thus be interpreted as the analysis map of a given frame for $\mathcal{V}$. Similarly, an arbitrary linear operator T acting on $\mathcal{W}$ that has rank N and satisfies $R(T) = \mathcal{V}$ can always be written in the form of Equation (3.20) and can thus

be interpreted as the synthesis map of a given frame for $\mathcal{V}$.

3.2.3 Dual Frames

The concept of a dual frame arises naturally when considering how an arbitrary vector $|v\rangle \in \mathcal{V}$ can be written as a linear combination of a given set of frame vectors $\{|f_k\rangle\}$. Given an arbitrary vector $|v\rangle \in \mathcal{V}$, consider the problem of obtaining a set of coefficients $\{\tilde{a}_k\}$ such that

$$|v\rangle = \sum_{k=1}^{M} \tilde{a}_k \, |f_k\rangle \, . \tag{3.24}$$

Since the $\{|f_k\rangle\}$ may be linearly dependent the solution is in general not unique. A very useful and established approach to finding a suitable set of $\{\tilde{a}_k\}$ is by using a so-called dual frame of $\{|f_k\rangle\}$. A frame $\{|\tilde{f}_k\rangle\}$ for $\mathcal{V}$ is referred to as a dual frame of $\{|f_k\rangle\}$ if

$$|v\rangle = \sum_{k=1}^{M} \langle \tilde{f}_k | v\rangle \, |f_k\rangle \, , \quad \forall \, |v\rangle \in \mathcal{V}. \tag{3.25}$$

A dual frame is always guaranteed to exist [16], and as we will show below if $\{|\tilde{f}_k\rangle\}$ is dual to $\{|f_k\rangle\}$ then the reverse is also true. Clearly, Equation (3.24) is satisfied by setting $\tilde{a}_k = \langle \tilde{f}_k | v\rangle$ for $1 \leq k \leq M$, where $\{|\tilde{f}_k\rangle\}$ is any dual frame of $\{|f_k\rangle\}$. When a vector $|v\rangle \in \mathcal{V}$ is written in the form of Equation (3.25), $\{|\tilde{f}_k\rangle\}$ is typically referred to as the analysis frame while $\{|f_k\rangle\}$ is referred to as the synthesis frame. Correspondingly, if the analysis map of $\{|\tilde{f}_k\rangle\}$ is denoted as $\tilde{A}_0$, then Equation (3.25) has the equivalent forms

$$|v\rangle = \sum_{k=1}^{M} |f_k\rangle \, \langle \tilde{f}_k | v\rangle = F_0 \, \tilde{A}_0 \, |v\rangle \, , \quad \forall \, |v\rangle \in \mathcal{V} \tag{3.26a}$$

$$F_0 \, \tilde{A}_0 = \mathcal{P}_{\mathcal{V}}. \tag{3.26b}$$

Given a frame $\{|f_k\rangle\}$ for $\mathcal{V}$, the dual frame of $\{|f_k\rangle\}$ is only unique when the frame vectors are linearly independent in which case they form a basis for $\mathcal{V}$. When the frame vectors are linearly dependent, one way of characterizing the set of all dual

frames is to consider the coefficient vector $|\tilde{w}\rangle$ corresponding to a particular dual frame and a particular $|v\rangle \in \mathcal{V}$,

$$|\tilde{w}\rangle = \tilde{A}_0 \, |v\rangle = \sum_{k=1}^{M} \tilde{a}_k \, |w_k\rangle \in \mathcal{W}, \tag{3.27}$$

with squared norm $||\tilde{w}||^2 = \sum_k \tilde{a}_k^2$. In general, distinct dual frames lead to distinct coefficient vectors. The dual frame that results in the minimum squared norm $||\tilde{w}||^2$ is

$$|\tilde{f}_k\rangle = (F_0 \, A_0)^{-1} \, |f_k\rangle \, , \quad 1 \leq k \leq M. \tag{3.28}$$

A derivation of this fact is included in Section 3.2.4. The dual frame defined by Equation (3.28) is referred to as the canonical dual frame of $\{|f_k\rangle\}$ [16]. Its synthesis map, which we will denote by F_{can}, is equal to $F_{\text{can}} = (F_0 \, A_0)^{-1} F_0$. Its analysis map is $A_{\text{can}} = F_{\text{can}}^{\dagger} = A_0 (F_0 \, A_0)^{-1}$, where we have used the fact that for an invertible linear operator T, we have $(T^{-1})^{\dagger} = (T^{\dagger})^{-1}$. It is significant that the nullspace of F_{can} and the range of A_0 are related via $N(F_{\text{can}}) = R(A_0)^{\perp}$. This property is utilized in Section 3.5. Note that if $\{|\tilde{f}_k\rangle\}$ is the canonical dual of $\{|f_k\rangle\}$, then the reverse is also true. This can be verified by interchanging the roles of the $\{|f_k\rangle\}$ and $\{|\tilde{f}_k\rangle\}$ in Equation (3.28) and substituting in the expressions for F_{can} and A_{can}.

The concept of a dual frame also arises when considering how an arbitrary vector $|v\rangle \in \mathcal{V}$ can be linearly reconstructed from its frame coefficients $\{a_k = \langle f_k | v \rangle\}$. Here $\{|f_k\rangle\}$ is assumed to be a fixed frame for $\mathcal{V}$. Given the $\{a_k\}$, consider the problem of finding a synthesis frame $\{|\tilde{f}_k\rangle\}$ such that

$$|v\rangle = \sum_{k=1}^{M} a_k \, |\tilde{f}_k\rangle \, , \quad \forall \, |v\rangle \in \mathcal{V}. \tag{3.29}$$

Equation (3.29) states that $\{|f_k\rangle\}$ is a dual frame of $\{|\tilde{f}_k\rangle\}$. It has the equivalent

forms

$$|v\rangle = \tilde{F}_0 \, A_0 \, |v\rangle \, , \quad \forall \, |v\rangle \in \mathcal{V} \tag{3.30a}$$

$$\tilde{F}_0 \, A_0 = \mathcal{P}_{\mathcal{V}}. \tag{3.30b}$$

Taking the adjoint of both sides of Equations (3.26b) and (3.30b) shows that they too are equivalent. In summary, Equations (3.25), (3.26), (3.29), and (3.30) are all equivalent and a dual frame could be defined according to any of them. The canonical dual frame is of significance in relation to Equation (3.29) because, as explained in Section 3.5, setting $\{|\tilde{f}_k\rangle\}$ to be the canonical dual frame of $\{|f_k\rangle\}$ minimizes the expected reconstruction error when an unknown vector $|v\rangle \in \mathcal{V}$ is estimated from imprecise versions of the $\{a_k\}$.

3.2.4 The Canonical Dual Frame

Two significant properties of the canonical dual frame were stated above in Section 3.2.3. The first is that for a specific $|v\rangle \in \mathcal{V}$ and a specific frame $\{|f_k\rangle\}$ for $\mathcal{V}$, the canonical dual frame leads to the coefficient vector $|\tilde{w}\rangle = \sum_k \langle f_k|v\rangle \, |w_k\rangle$ with minimum norm. The second is that it minimizes the expected reconstruction error when imprecise frame coefficients are used to estimate an unknown vector. In Section 3.2.4 we derive the first of these properties. A derivation of the second is given in Section 3.5.

Let $|v\rangle$ be an arbitrary vector in $\mathcal{V}$ and let $\{|f_k\rangle\}$ be a frame for $\mathcal{V}$. We wish to find the dual frame $\{|\tilde{f}_k\rangle\}$ of $\{|f_k\rangle\}$ that minimizes the squared norm of the coefficient vector $\tilde{A}_0 \, |v\rangle = \sum_k \langle \tilde{f}_k|v\rangle \, |w_k\rangle$. It is sufficient to solve for the analysis map $\tilde{A}_0$ of the optimal dual frame. Denoting the synthesis map of $\{|f_k\rangle\}$ by F_0, the problem can be formulated as

$$\underset{\tilde{A}_0 \, : \, \mathcal{V} \to \mathcal{W}}{\text{minimize}} \quad ||\tilde{A}_0 \, |v\rangle \, ||^2 \tag{3.31a}$$

$$\text{subject to} \quad F_0 \tilde{A}_0 \, |v\rangle = |v\rangle \tag{3.31b}$$

The optimal coefficient vector must satisfy $\tilde{A}_0 |v\rangle \in R(A_0)$. To see why this is true, note that $\tilde{A}_0 |v\rangle$ can always be written as the sum of a component in $R(A_0)$ and a component in $R(A_0)^\perp = N(F_0)$,

$$\tilde{A}_0 |v\rangle = |w_1\rangle + |w_2\rangle \tag{3.32}$$

where $|w_1\rangle \in R(A_0)$ and $|w_2\rangle \in N(F_0)$. We have $||\tilde{A}_0 |v\rangle||^2 = ||w_1||^2 + ||w_2||^2$ and $F_0\tilde{A}_0 |v\rangle = F_0 |w_1\rangle$. Assume that Equation (3.32) holds for a given dual frame. If $|w_2\rangle$ were nonzero, then we could always find a different dual frame with analysis map $\hat{A}_0$ satisfying $\hat{A}_0 |v\rangle = |w_1\rangle$. Equation (3.31b) would still be satisfied ($F_0\hat{A}_0 |v\rangle = F_0 |w_1\rangle = |v\rangle$) and the new coefficient vector would have smaller squared norm ($||\hat{A}_0 |v\rangle||^2 \leq ||\tilde{A}_0 |v\rangle||^2$). Next note that since $|v\rangle$ was assumed to be arbitrary, Equation (3.31b) implies that $\dim R(\tilde{A}_0) \geq N$. Since $\dim R(A_0) = N$ according to Section 3.2.2, the requirements that $\tilde{A}_0 |v\rangle \in R(A_0)$ for arbitrary $|v\rangle \in \mathcal{V}$ and $\dim R(\tilde{A}_0) \geq N$ together imply that the optimal analysis map satisfies $R(\tilde{A}_0) = R(A_0)$. Therefore, by definition of $R(A_0)$ we must have $\tilde{A}_0 |v\rangle = A_0 |x\rangle$ for some $|x\rangle \in \mathcal{V}$. Substituting into Equation (3.31b), we find that $F_0\tilde{A}_0 |v\rangle = F_0A_0 |x\rangle$. It is straightforward to show that the operator (F_0A_0), often referred to as the frame operator of $\{|f_k\rangle\}$, is always invertible. Thus, $|x\rangle = (F_0A_0)^{-1} |v\rangle$ and so $\tilde{A}_0 |v\rangle = A_0 |x\rangle = A_0(F_0A_0)^{-1} |v\rangle$. Again using the fact that $|v\rangle$ was assumed to be arbitrary, this implies that $\tilde{A}_0 = A_0(F_0A_0)^{-1}$, which is equal to the analysis map of the canonical dual frame.

3.3 Parseval Frames and Naimark's Theorem

Reconstructing an unknown vector from its frame coefficients $\{a_k\}$ using the canonical dual frame requires the inversion of the the operator $(F_0\, A_0)$, a task that can lead to issues of computational complexity or instability. Tight frames are an important class of frames that circumvent these issues due to the fact that they are self-dual up to a constant factor. Parseval frames are tight frames for which the constant is equal to one. As detailed below, the synthesis and analysis maps of a Parseval frame

can be made to have a particularly simple form by choosing the $\{|w_k\rangle\}$ according to Naimark's Theorem.

3.3.1 Parseval Frames

A tight frame $\{|f_k\rangle\}$ for $\mathcal{V}$ is one that satisfies Parseval's identity [16] up to a constant factor,

$$\sum_{k=1}^{M} |\langle f_k|v\rangle|^2 = C\,||v||^2 \;\forall\; |v\rangle \in \mathcal{V} \tag{3.33}$$

for some $C > 0$. In terms of the analysis map A_0 of $\{|f_k\rangle\}$, Equation (3.33) can be written as $||A_0|v\rangle||^2 = C||v||^2$ for all $|v\rangle \in \mathcal{V}$. When $C = 1$ the frame is referred to as a Parseval frame. Orthonormal bases are a special case of Parseval frames with $M = N$. In reference to Equation (3.15), Equation (3.33) is equivalent to the statement the the frame bounds of $\{|f_k\rangle\}$ are equal, $C = D$.

Parseval frames are always self-dual. This follows from the fact that the sum in Equation (3.33) can be alternately expressed as

$$\sum_{k=1}^{M} |\langle f_k|v\rangle|^2 = \sum_{k=1}^{M} \langle v|f_k\rangle\langle f_k|v\rangle = \langle v|\left(\sum_{k=1}^{M} |f_k\rangle\langle f_k|v\rangle\right). \tag{3.34}$$

For the above expression to be equal to $||v||^2 = \langle v|v\rangle$ for all $|v\rangle \in \mathcal{V}$, we must have

$$\sum_{k=1}^{M} |f_k\rangle\langle f_k|v\rangle = |v\rangle \;\forall\; |v\rangle \in \mathcal{V}. \tag{3.35}$$

And Equation (3.35) states by definition that $\{|f_k\rangle\}$ is a dual frame of itself. Consequently when $\{|f_k\rangle\}$ is a Parseval frame, the task of reconstructing a vector $|v\rangle$ from the collection of coefficients $\{a_k = \langle f_k|v\rangle\}$ using the canonical dual frame is especially straightforward. A given Parseval frame is in fact its own canonical dual frame [16]. This is because Equation (3.35) implies that $(F_0\,A_0) = I$ where I is the identity operator on $\mathcal{V}$. Substituting into Equation (3.28) leads to

$$|\tilde{f}_k\rangle = (F_0\,A_0)^{-1}\,|f_k\rangle = I^{-1}\,|f_k\rangle = |f_k\rangle\,, \quad 1 \le k \le M. \tag{3.36}$$

A similar line of logic to the one given above can be used to show that a frame that is self-dual is itself always a Parseval frame.

3.3.2 Naimark's Theorem

Naimark's Theorem is well-known in both the frame theory and quantum physics communities as well as in operator theory more generally. The version stated below will perhaps be most familiar to readers with a background in frame theory [14, 16]. The version more typically used in the quantum physics community is stated in terms of positive operator-valued measures, which are defined in Chapter 4, and often arises in the context of the physical realizability of non-standard measurements [55, 57]. As a result we review the latter version of the theorem in Chapter 4. In the statement of Naimark's Theorem given below we continue to assume that $\mathcal{V}$ and $\mathcal{W}$ are finite dimensional, but we emphasize that this is not its most general form.

Naimark's Theorem. *As typically stated in the terminology of frame theory: A frame* $\{|f_k\rangle, 1 \leq k \leq M\}$ *for* $\mathcal{V}$ *is a Parseval frame if and only if there exists an orthonormal basis* $\{|w_k\rangle, 1 \leq k \leq M\}$ *for* $\mathcal{W}$ *such that*

$$\mathcal{P}_{\mathcal{V}} |w_k\rangle = |f_k\rangle, \quad 1 \leq k \leq M. \tag{3.37}$$

A derivation of Naimark's Theorem is included below. We will refer to Equation (3.37) as Naimark's identity for convenience. Note that for a given frame $\{|f_k\rangle\}$ for $\mathcal{V}$, it is trivial to construct a set of basis vectors $\{|w_k\rangle\}$ for $\mathcal{W}$ satisfying Naimark's identity *if* they are not required to be orthonormal. For example, if $\{|u_k\rangle, N+1 \leq k \leq M\}$ is an orthonormal basis for $\mathcal{V}^{\perp}$ then setting $|w_k\rangle = |f_k\rangle$ for $1 \leq k \leq N$ and $|w_k\rangle = |f_k\rangle + |u_k\rangle$ for $N+1 \leq k \leq M$ is sufficient. Naimark's Theorem guarantees that when $\{|f_k\rangle\}$ is a Parseval frame, we can always construct the $\{|w_k\rangle\}$ in such a way that they satisfy Naimark's identity *and* are orthonormal.

Let $\{|f_k\rangle\}$ be an arbitrary frame for $\mathcal{V}$ and assume that there exists an orthonormal basis $\{|w_k\rangle\}$ for $\mathcal{W}$ satisfying Naimark's identity. To show that $\{|f_k\rangle\}$ must be a

Parseval frame, note that an arbitrary vector $|v\rangle \in \mathcal{V}$ can always be written as

$$|v\rangle = \sum_{k=1}^{M} \langle w_k|v\rangle \, |w_k\rangle = \sum_{k=1}^{M} b_k \, |w_k\rangle \,, \tag{3.38}$$

where we have defined $b_k = \langle w_k|v\rangle$ for $1 \leq k \leq M$. Since the $\{|w_k\rangle\}$ are orthonormal, the squared norm of $|v\rangle$ is equal to the sum of the squared magnitudes of the $\{b_k\}$, $||v||^2 = \langle v|v\rangle = \sum_k |b_k|^2$. On the other hand, since the $\{|w_k\rangle\}$ satisfy Naimark's identity, we also have $b_k = \langle w_k|v\rangle = \langle f_k|v\rangle$ for all $1 \leq k \leq M$, i.e., the basis coefficients and the frame coefficients of $|v\rangle$ are always identical. This is because the component of $|w_k\rangle$ in $\mathcal{V}^\perp$ has no impact on the value of its inner product with $|v\rangle$. Thus,

$$\sum_{k=1}^{M} |\, \langle f_k|v\rangle \,|^2 = \sum_{k=1}^{M} |b_k|^2 = ||v||^2 \text{ for all } |v\rangle \in \mathcal{V}, \tag{3.39}$$

which implies by definition that $\{|f_k\rangle\}$ is a Parseval frame.

Now assume that $\{|f_k\rangle\}$ is a Parseval frame. To show that there always exists an orthonormal basis $\{|w_k\rangle\}$ for $\mathcal{W}$ satisfying Naimark's identity, let $\{|e_k\rangle\}$ be an arbitrary orthonormal basis for $\mathcal{W}$ and consider expanding each of the $\{|f_k\rangle\}$ as a linear combination of the $\{|e_k\rangle\}$,

$$|f_k\rangle = \sum_{j=1}^{M} \langle e_j|f_k\rangle \, |e_j\rangle = \sum_{j=1}^{M} c_{jk} \, |e_j\rangle \,, \quad 1 \leq k \leq M, \tag{3.40}$$

where we have defined $c_{jk} = \langle e_j|f_k\rangle$ for $1 \leq j, k \leq M$. Then the vectors

$$|w_k\rangle = \sum_{\ell=1}^{M} c_{k\ell} \, |e_\ell\rangle \,, \quad 1 \leq k \leq M \tag{3.41}$$

form an orthonormal basis for $\mathcal{W}$. Since $\mathcal{W}$ has dimension M, to verify that the M vectors $\{|w_k\rangle\}$ span $\mathcal{W}$ it is sufficient to verify that they are orthonormal. To verify that the $\{|w_k\rangle\}$ are orthonormal assume that $|v\rangle = \sum_m b_m \, |e_m\rangle$ is an arbitrary vector in $\mathcal{V}$ where $b_m = \langle e_m|v\rangle$ for $1 \leq m \leq M$. Substituting the expansions of $|v\rangle$ and of

the $\{|f_k\rangle\}$ into the left-hand side of Equation (3.35) and rearranging leads to

$$\sum_{k=1}^{M} |f_k\rangle \langle f_k|v\rangle = \sum_{k=1}^{M} \left(\sum_{j=1}^{M} c_{jk} |e_j\rangle \right) \left(\sum_{\ell=1}^{M} c_{\ell k}^* \langle e_\ell| \right) \left(\sum_{m=1}^{M} b_m |e_m\rangle \right) \tag{3.42a}$$

$$= \sum_{j=1}^{M} \left[\sum_{\ell=1}^{M} b_\ell \left(\sum_{k=1}^{M} c_{jk} c_{\ell k}^* \right) \right] |e_j\rangle \,. \tag{3.42b}$$

For the above expression to be equal to the right-hand side of Equation (3.35), $|v\rangle = \sum_j b_j |e_j\rangle$ for an arbitrary $|v\rangle \in \mathcal{V}$, we must have $\sum_k = c_{jk} c_{\ell k}^* = \delta_{j\ell}$ which implies that $\langle w_\ell | w_j \rangle = \delta_{j\ell}$, i.e., the $\{|w_k\rangle\}$ are orthonormal.

3.3.3 Synthesis and Analysis Maps of a Parseval Frame

If $\{|f_k\rangle\}$ is a Parseval frame and the $\{|w_k\rangle\}$ are chosen to satisfy Naimark's identity, then the analysis and synthesis maps of $\{|f_k\rangle\}$ are $A_0 = F_0 = \mathcal{P}_\mathcal{V}$. To show that this is true, first note that as mentioned above, for an arbitrary vector $|v\rangle \in \mathcal{V}$, the basis coefficients $\{b_k = \langle w_k|v\rangle\}$ and the frame coefficients $\{a_k = \langle f_k|v\rangle\}$ of $\mathcal{V}$ are identical. This leads to

$$A_0 |v\rangle = \sum_{k=1}^{M} |w_k\rangle \langle f_k|v\rangle = \sum_{k=1}^{M} b_k |w_k\rangle = |v\rangle \,. \tag{3.43}$$

For this to be true for all $|v\rangle \in \mathcal{V}$ we must have $A_0 = \mathcal{P}_\mathcal{V}$ and thus $F_0 = A_0^\dagger = \mathcal{P}_\mathcal{V}$.

An alternate derivation relies on the fact that since Parseval frames are self-dual we have $(F_0 A_0) = \mathcal{P}_\mathcal{V}$ according to Equation (3.26b). A_0 can be expressed as

$$A_0 = \sum_{k=1}^{M} |w_k\rangle \langle f_k| = \mathcal{P}_\mathcal{V} \left(\sum_{k=1}^{M} |f_k\rangle \langle f_k| \right) = \mathcal{P}_\mathcal{V} F_0 A_0 = \mathcal{P}_\mathcal{V}, \tag{3.44}$$

where we have used the fact that $\mathcal{P}_\mathcal{V}^2 = \mathcal{P}_\mathcal{V}$. Since orthogonal projection operators are Hermitian, taking the adjoint of both sides leads to the conclusion that $A_0^\dagger = F_0 = \mathcal{P}_\mathcal{V}$.

3.4 Frame Representations of Operator Spaces

The goal of Section 3.4 is to extend the discussion of frame representations to vector spaces $\mathcal{V}$ whose elements are Hermitian operators acting on a given Hilbert space $\mathcal{H}$ of dimension d. Such a vector space is sometimes referred to as an operator space. Unlike in Chapter 4, in Section 3.4 we do not assume that $\mathcal{H}$ necessarily represents the state space of a quantum system. Rather, the concepts addressed apply to any finite-dimensional Hilbert space. A key perspective that we take is the geometric characterization of positive semidefinite operators using a ball and sphere in operator space when $\mathcal{H}$ has dimension 2. The concepts are applied to operator spaces in quantum mechanics in Chapters 4 and 6.

For the remainder of the thesis, $\mathcal{V}$ and $\mathcal{W}$ will be used to denote operator spaces defined in relation to a given Hilbert space $\mathcal{H}$. And in certain contexts we may wish to consider an element V of $\mathcal{V}$ alternately as an operator acting on an element of $\mathcal{H}$ or as a "vector" in $\mathcal{V}$ (that is, an element of the operator-valued vector space $\mathcal{V}$). Following a combination of the conventions in [24, 67], when we wish to emphasize that a Hermitian operator V on $\mathcal{H}$ is being used as an element of $\mathcal{V}$ we will denote it using modified bra-ket notation as $|V\rangle\rangle$. The inner product between any two operators $V_1, V_2 \in \mathcal{V}$ will be denoted as $\langle\langle V_1|V_2\rangle\rangle$. A specific expression for $\langle\langle V_1|V_2\rangle\rangle$ is given in Equation (3.49) below. The same notation carries over to elements of $\mathcal{W}$. Linear operators acting on $\mathcal{W}$ ("superoperators" [67]) will be denoted using bold font. For example, $\mathbf{A}_0$ and $\mathbf{F}_0$ will denote the analysis and synthesis maps, respectively, of a given frame for $\mathcal{V}$.

3.4.1 Definitions of $\mathcal{V}$ and $\mathcal{W}$

Assume that $\mathcal{H}$ is a Hilbert space of dimension d. The set of all Hermitian operators on $\mathcal{H}$ forms an operator space $\mathcal{V}$ over the real numbers with dimension $N = d^2$. $\mathcal{V}$

can always be decomposed into the two orthogonal subspaces $\mathcal{U}$ and $\mathcal{U}^\perp$, defined as

$$\mathcal{U}^\perp = \text{span}\{I\}, \tag{3.45a}$$

$$\mathcal{U} = \text{span}\{V \in \mathcal{V} : \langle\langle I|V\rangle\rangle = \text{Tr}(V) = 0\}, \tag{3.45b}$$

where I denotes the identity operator on $\mathcal{H}$ and $\text{Tr}(\cdot)$ is the trace operator. $\mathcal{U}$ is the span of all trace 0 operators in $\mathcal{V}$. It has dimension $(N-1) = (d^2-1)$ and is always isomorphic to $\mathbb{R}^{d^2-1}$ [67]. $\mathcal{U}^\perp$ is the span of the identity and has dimension 1. Given an an arbitrary operator $V \in \mathcal{V}$, the orthogonal projection of V onto $\mathcal{U}^\perp$ is always equal to $\mathcal{P}_{\mathcal{U}^\perp}(V) = \langle\langle I|V\rangle\rangle\,|I\rangle\rangle/d = \text{Tr}(V)\,|I\rangle\rangle/d$, where the factor of $1/d$ accounts for the fact that $|I\rangle\rangle/\sqrt{d}$ has unit norm. Therefore, V can always be written as

$$|V\rangle\rangle = \mathcal{P}_{\mathcal{U}^\perp}|V\rangle\rangle + \mathcal{P}_{\mathcal{U}}|V\rangle\rangle = \frac{\text{Tr}(V)}{\sqrt{d}}\frac{|I\rangle\rangle}{\sqrt{d}} + \mathcal{P}_{\mathcal{U}}|V\rangle\rangle. \tag{3.46}$$

For an arbitrary real number τ, V has trace τ if and only if $\mathcal{P}_{\mathcal{U}^\perp}(V) = \tau\,|I\rangle\rangle/d$. The set of all elements in $\mathcal{V}$ with trace τ then forms a hyperplane in $\mathcal{V}$ that is orthogonal to the identity.

There are many ways of constructing a larger operator space $\mathcal{W}$ that contains $\mathcal{V}$. As an example, consider extending $\mathcal{H}$ to a larger space $\mathcal{H}'$ of dimension $d' > d$. $\mathcal{H}'$ can be expressed as the direct sum of $\mathcal{H}$ and its orthogonal complement $\mathcal{H}^\perp$, where $\mathcal{H}^\perp$ has dimension $(d'-d)$. Informally, we may define $\mathcal{W}$ to be the real span of all Hermitian operators on $\mathcal{H}'$ that are "block-diagonal" with respect to the direct sum decomposition $\mathcal{H}' = \mathcal{H} \oplus \mathcal{H}^\perp$. Mathematically this can be phrased as follows. Given a Hermitian operator V on $\mathcal{H}$, V can always be expressed as

$$V = \sum_{i=1}^{d} a_i\,|x_i\rangle\,\langle x_i|\,, \tag{3.47}$$

where the eigenvalues $\{a_i\}$ are real and the eigenvectors $\{|x_i\rangle\}$ form an orthonormal basis for $\mathcal{H}$. Since the $\{|x_i\rangle\}$ are also elements of $\mathcal{H}'$, V can also be viewed as a Hermitian operator acting on $\mathcal{H}'$. It maps all vectors in $\mathcal{H}^\perp$ to the zero vector.

Similarly, given a Hermitian operator U on $\mathcal{H}^\perp$, U can always be written as

$$U = \sum_{i=1}^{d'-d} b_i \left|y_i\right\rangle \left\langle y_i\right|, \tag{3.48}$$

where the eigenvalues $\{b_i\}$ are real and the eigenvectors $\{|y_i\rangle\}$ form an orthonormal basis for $\mathcal{H}^\perp$. Since the $\{|y_i\rangle\}$ are also elements of $\mathcal{H}'$, U can also be viewed as a Hermitian operator on $\mathcal{H}'$ that maps all vectors in $\mathcal{H}$ to the zero vector. We define $\mathcal{W}$ as the real span of all operators on $\mathcal{H}'$ that can be written in the form of either Equation (3.47) or (3.48) – that is, the set of all linear combinations of such operators with real coefficients. When constructed in this way, $\mathcal{W}$ has dimension $d^2 + (d'-d)^2 = N + (d'-d)^2$. If we desire $\mathcal{W}$ to have a specific dimension $M > d^2$, we can always choose d' to be large enough so that $N + (d'-d)^2 > M$ and then redefine $\mathcal{W}$ to be a subspace of itself with dimension M. We will assume going forward that a suitable operator space $\mathcal{W}$, constructed for example according to the procedure just described, has been specified. The inner product between any two elements $|W_1\rangle\rangle, |W_2\rangle\rangle \in \mathcal{W}$ will be denoted by $\langle\langle W_1|W_2\rangle\rangle$ and defined by the relation

$$\langle\langle W_1|W_2\rangle\rangle = \sum_{i=1}^{d'} \gamma_i \left\langle e_i\right| W_2 \left|e_i\right\rangle \text{ where } W_1 = \sum_{i=1}^{d'} \gamma_i \left|e_i\right\rangle \left\langle e_i\right|. \tag{3.49}$$

In Equation (3.49), the $\{\gamma_i\}$ are the eigenvalues of W_1 and the $\{|e_i\rangle\}$, which lie in $\mathcal{H}'$, are its eigenvectors. Because all elements of $\mathcal{W}$ can be written as a linear combination of operators of the form of Equations (3.47) and (3.48), each of the $\{|e_i\rangle\}$ lie either in $\mathcal{H}$ or in $\mathcal{H}^\perp$. It is straightforward to verify that the function defined in Equation (3.49) satisfies all the properties of a valid inner product function on $\mathcal{W}$. It is in fact a special case of the well-known Hilbert-Schmidt or trace inner product [49].

3.4.2 Operator-Valued Frames

For clarity we repeat the definition of a frame using operator space notation. Any set of operators $\{F_k, 1 \leq k \leq M\}$ that lie in and span an operator space $\mathcal{V}$ form a frame for $\mathcal{V}$. More generally an M-element frame for $\mathcal{V}$ is defined as any set of operators

$\{F_k\}$ that lie in $\mathcal{V}$ and satisfy

$$C\,||V||^2 \leq \sum_{k=1}^{M} |\langle\langle F_k|V\rangle\rangle|^2 \leq D\,||V||^2 \tag{3.50}$$

for some $0 < C \leq D < \infty$ and for all $|V\rangle\rangle \in \mathcal{V}$ [67]. We will always assume that the values of C and D are set to form the tightest possible bounds, in which case they are referred to as the frame bounds of $\{F_k\}$. A tight frame for $\mathcal{V}$ is one whose frame bounds are equal. In keeping with the current notation, from this point forward we will use $\{W_k,\, 1 \leq k \leq M\}$ to denote an orthonormal basis for $\mathcal{W}$, and $\{F_k,\, 1 \leq k \leq M\}$ to denote a frame for $\mathcal{V}$.

Regardless of whether the number of frame vectors is finite or infinite, the definition of an operator frame given in Equation (3.50) may also be extended to include to the notion of a generalized operator frame with respect to a given measure. In the terminology of [67], a set of operators satisfying Equation (3.50) is referred to as a generalized operator frame with respect to the counting measure.

3.4.3 Operator Space for $\mathcal{H} = \mathbb{C}^2$

We explicitly describe the operator space $\mathcal{V}$ when $\mathcal{H} = \mathbb{C}^2$, i.e., $d = 2$. Our intent aside from providing a concrete example in low dimensions is to also present some geometric intuition regarding where operators with constant trace and positive semidefinite operators lie in $\mathcal{V}$. The concepts presented in Section 3.4.3 are relevant to the simulations presented in Chapter 6 involving qubit density operators. For generalizations to values of $d > 2$, we refer the reader to, for example, [29, 67].

When $\mathcal{H} = \mathbb{C}^2$, $\mathcal{V}$ has dimension $d^2 = 4$. The set of operators $\{I/\sqrt{2}, \sigma_1/\sqrt{2}, \sigma_2/\sqrt{2}, \sigma_3/\sqrt{2}\}$ where $\{\sigma_1, \sigma_2, \sigma_3\}$ are the Pauli operators [49] is a commonly used basis for $\mathcal{V}$. It is an orthonormal basis with respect to the inner product defined by Equation (3.49). We have $\mathcal{U}^\perp = \text{span}\{I/\sqrt{2}\}$ and $\mathcal{U} = \text{span}\{\sigma_1/\sqrt{2}, \sigma_2/\sqrt{2}, \sigma_3/\sqrt{2}\}$. The Pauli operators will prove to be a convenient choice of orthonormal basis for $\mathcal{U}$ in the context of quantum mechanics as they are directly related to the representation of an arbitrary qubit density operator in terms of its Bloch vector.

Given an arbitrary operator $V \in \mathcal{V}$, V can always be written as a linear combination of the operators $\{I/\sqrt{2}, \sigma_1/\sqrt{2}, \sigma_2/\sqrt{2}, \sigma_3/\sqrt{2}\}$,

$$V = c_0 \frac{|I\rangle\rangle}{\sqrt{2}} + c_1 \frac{|\sigma_1\rangle\rangle}{\sqrt{2}} + c_2 \frac{|\sigma_2\rangle\rangle}{\sqrt{2}} + c_3 \frac{|\sigma_3\rangle\rangle}{\sqrt{2}}. \tag{3.51}$$

In Equation (3.51) the basis expansion coefficients are $c_0 = \langle\langle I|V\rangle\rangle/\sqrt{2} = \mathrm{Tr}(V)/\sqrt{2}$ and $c_i = \langle\langle \sigma_i|V\rangle\rangle/\sqrt{2} = \mathrm{Tr}(\sigma_i V)/\sqrt{2}$ for $1 \leq i \leq 3$. A crucial relation that forms the foundation of much of the discussion in Chapters 4 and 6 is that if V is positive semidefinite then we always have

$$\sqrt{c_1^2 + c_2^2 + c_3^2} \leq \mathrm{Tr}(V)/\sqrt{2}, \tag{3.52}$$

with equality if and only if V has rank one. Equation (3.52) can be derived by solving for the eigenvalues of V in terms of the $\{c_i\}$ and requiring them to be non-negative. One way of interpreting Equation (3.52) is as follows. Given a positive semidefinite operator V with basis expansion coefficients $\{c_i\}$, there is always an associated closed ball in $\mathbb{R}^3$ of radius $\mathrm{Tr}(V)/\sqrt{2}$. The column vector $\mathbf{c} = [c_1, c_2, c_3]^T$ corresponds to coefficients of the orthogonal projection of V onto $\mathcal{U}$ and always lies within the ball. $\mathbf{c}$ lies on the surface of the ball, that is, on the sphere of radius $\mathrm{Tr}(V)/\sqrt{2}$, when V has rank one.

Example 3.1. To help in providing an intuitive geometric picture, we temporarily define $\mathcal{V} = \mathbb{R}^3$ with dimension $N = 3$ and orthonormal basis $\{|b_0\rangle, |b_1\rangle, |b_2\rangle\}$. An arbitrary vector $|x\rangle \in \mathbb{R}^3$ can always be expressed as

$$|x\rangle = c_0 \, |b_0\rangle + c_1 \, |b_1\rangle + c_2 \, |b_1\rangle , \tag{3.53}$$

where $c_i = \langle b_i|x\rangle$ for $0 \leq i \leq 2$. As shown in Figure 3-3, the set of vectors in $\mathbb{R}^3$ that satisfy $c_0 = 2^{-1/2}$ lie on a hyperplane while the set of vectors that satisfy $c_0^2 \geq c_1^2 + c_2^2$ lie on or within a cone. The set of vectors that satisfy both of the constraints lies at the intersection of the hyperplane and the cone, which takes the form of an $(N-1) = 2$ dimensional ball, i.e., a circle.

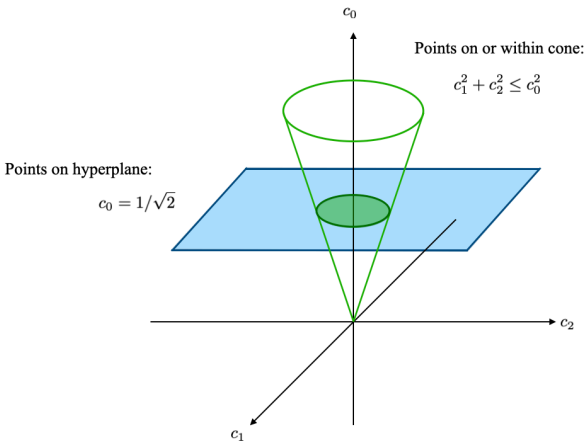


Figure 3-3: Illustration of the constraints described in Example 3.1.

3.5 Robustness of Frame Representations

Given a vector $|v\rangle \in \mathcal{V}$ and an analysis frame $\{|f_k\rangle\}$ for $\mathcal{V}$, $|v\rangle$ can always be represented in terms of its frame coefficients $\{a_k = \langle f_k|v\rangle\}$, for example by using Equation (3.29). An important problem in classical signal processing is that of obtaining an accurate representation of $|v\rangle$ using only imprecise versions of the $\{a_k\}$ after they have been subjected to some source of error, such as quantization error. In Sections 3.5.1 and 3.5.2 we describe a version of this problem that incorporates a specific model for the error source in more detail. We emphasize that there are many highly sophisticated ways of exploiting the redundancy in the $\{a_k\}$ to obtain increasingly accurate representations of $|v\rangle$ (see, for example, [12, 20] and references therein). However, their details are beyond the discussion in Sections 3.5.1 and 3.5.2, and beyond the scope of this thesis. Note also that as this topic is not directly relevant to binary hypothesis testing, some readers may wish to proceed directly to Chapter 4.

We assume that $|v\rangle$ is approximated by a vector $|\hat{v}\rangle$ that is computed by simply replacing the true coefficients $\{a_k\}$ by their imprecise counterparts $\{\hat{a}_k\}$ in Equation (3.29),

$$|\hat{v}\rangle = \sum_{k=1}^{M} \hat{a}_k \, |\tilde{f}_k\rangle \,, \tag{3.54}$$

where the synthesis frame $\{|\tilde{f}_k\rangle\}$ is a dual frame of $\{|f_k\rangle\}$. The final error vector is

defined as $|v_e\rangle = |\hat{v}\rangle - |v\rangle$ and the objective is to find the frame $\{|\tilde{f}_k\rangle\}$ that minimizes $\mathcal{E} = \mathbb{E}[||v_e||^2]$, where the expectation is taken over all possible values of the $\{\hat{a}_k\}$. The central conclusions of Sections 3.5.1 and 3.5.2 are that the canonical dual frame is optimal with respect to minimizing $\mathcal{E}$ and that when $\{|f_k\rangle\}$ is a so-called equal norm tight frame, the minimum value of $\mathcal{E}$ as obtained by reconstruction with the canonical dual is proportional to Δ^2/M, where Δ^2 is the variance of the individual error values. Throughout Section 3.5 we will use notation corresponding to vector-valued vector spaces, but we emphasize that all of the analysis applies equally well to operator spaces. Indeed, in Section 5.5 we will apply the same analysis to elements of an operator space in the context of quantum state estimation.

3.5.1 Optimality of the Canonical Dual

Assume that the observed coefficients are $\{\hat{a}_k = a_k + e_k\}$ and that the individual error values $\{e_k\}$ have zero mean, variance σ^2, and are pairwise uncorrelated,

$$\mathbb{E}[e_k] = 0, \qquad 1 \leq k \leq M, \tag{3.55a}$$

$$\mathbb{E}[e_j e_k] = \delta_{jk}, \quad 1 \leq j,\, k \leq M. \tag{3.55b}$$

Equations (3.55) have been shown to be a useful model mathematically in certain scenarios, despite not always being literally true in practice (see, for example, Chapter 4 of [51]). Substituting $\{\hat{a}_k = a_k + e_k\}$ into Equation (3.54) yields $|\hat{v}\rangle = |v\rangle + |v_e\rangle$, where $|v_e\rangle = \sum_k e_k\,|\tilde{f}_k\rangle$ is the final error vector. Let $|w_e\rangle = \sum_k e_k\,|w_k\rangle$ be the element of $\mathcal{W}$ whose coefficients in the $\{|w_k\rangle\}$ basis are the $\{e_k\}$. Denoting the synthesis map of $\{|\tilde{f}_k\rangle\}$ by $\tilde{F}_0$, the final error vector can also be expressed as $|v_e\rangle = \tilde{F}_0\,|w_e\rangle$. We wish to find the synthesis frame that minimizes $\mathcal{E} = \mathbb{E}[\,||\tilde{F}_0\,|w_e\rangle\,||^2\,]$ where the expectation is taken over all possible values of the $\{\hat{a}_k\}$ or equivalently over all possible values of the $\{e_k\}$.

We now show that the optimal synthesis frame is the canonical dual of the analysis frame [31]. This is true even when the $\{e_k\}$ have possibly different variances, as long as they remain pairwise uncorrelated. The problem of minimizing $\mathcal{E}$ can be formulated

as

$$\underset{\tilde{F}_0 : \mathcal{W} \to \mathcal{V}}{\text{minimize}} \quad \mathbb{E}\left[||\tilde{F}_0 \, |w_e\rangle \, ||^2 \right] \tag{3.56a}$$

$$\text{subject to} \quad \tilde{F}_0 \, A_0 \, |v\rangle = |v\rangle \text{ for all } |v\rangle \in \mathcal{V}, \tag{3.56b}$$

where the minimization is performed over all linear operators $\tilde{F}_0$ from $\mathcal{W}$ to $\mathcal{V}$ that satisfy Equation (3.56b). Equation (3.56b) states that $\tilde{F}_0$ must be a left-inverse of A_0, which effectively specifies that $\tilde{F}_0$ must be the synthesis operator of a frame that is dual to the analysis frame. A left-inverse is guaranteed to exist because as stated in Section 3.2.2, A_0 has rank N. Assume that $\tilde{F}_0$ is an arbitrary left-inverse of A_0. To fully specify $\tilde{F}_0$, it is both necessary and sufficient to specify separately its actions on $R(A_0)$ and $R(A_0)^\perp$. Its action on $R(A_0)$ is fully constrained by Equation (3.56b), whereas its action on $R(A_0)^\perp$ has no effect on Equation (3.56b) and can be chosen to minimize $\mathcal{E}$.

Assume that $\{|u_k\rangle \, , 1 \leq k \leq N\}$ is an orthonormal basis for $R(A_0)$ and that $\{|u_k\rangle \, , N+1 \leq k \leq M\}$ is an orthonormal basis for $R(A_0)^\perp$. The vector $|w_e\rangle$ can always be expressed as

$$|w_e\rangle = |w_{e1}\rangle + |w_{e2}\rangle = \sum_{k=1}^{N} c_k \, |u_k\rangle + \sum_{k=N+1}^{M} c_k \, |u_k\rangle \, . \tag{3.57}$$

Clearly, $|w_{e1}\rangle \in R(A_0)$ and $|w_{e2}\rangle \in R(A_0)^\perp$. Note that since the $\{c_k\}$ are related to the $\{e_k\}$ by an orthogonal transformation in $\mathcal{W}$, they also have zero mean, variance σ^2, and are pairwise uncorrelated. Next note that since each of the $\{|u_k\rangle \, , 1 \leq k \leq N\}$ is an element of $R(A_0)$, by definition there is a unique vector $|v_k\rangle \in \mathcal{V}$ satisfying $A_0 \, |v_k\rangle = |u_k\rangle$ for all $1 \leq k \leq N$. Equation (3.56b) implies that $\{\tilde{F}_0 \, |u_k\rangle = |v_k\rangle \, , 1 \leq k \leq N\}$. Now it only remains to specify the vectors $\{\tilde{F}_0 \, |u_k\rangle \, , N+1 \leq k \leq M\}$. We have $\mathcal{E} = \mathbb{E}[||\tilde{F}_0 \, |w_{e1}\rangle + \tilde{F}_0 \, |w_{e2}\rangle \, ||^2$ and since the $\{c_k\}$ are uncorrelated all of the cross

terms are equal to zero. Thus,

$$\mathcal{E} = \mathbb{E}\left[||\tilde{F}_0 |w_e\rangle ||^2\right] = \mathbb{E}\left[\sum_{k=1}^{N} c_k^2 ||v_k||^2 + \sum_{k=N+1}^{M} c_k^2 ||\tilde{F}_0 |w_k\rangle ||^2\right] \tag{3.58a}$$

$$= \sum_{k=1}^{N} \mathbb{E}[c_k^2] ||v_k||^2 + \sum_{k=N+1}^{M} \mathbb{E}[c_k^2] ||\tilde{F}_0 |w_k\rangle ||^2 \tag{3.58b}$$

$$= \sigma^2 \sum_{k=1}^{N} ||v_k||^2 + \sigma^2 \sum_{k=N+1}^{M} ||\tilde{F}_0 |w_e\rangle ||^2. \tag{3.58c}$$

Since the value of the first sum is fixed and since all terms in both sums must be non-negative, the minimal value is obtained when the second sum is equal to zero, which happens when $\{\tilde{F}_0 |w_k\rangle = 0, N+1 \leq k \leq M\}$. Therefore, the optimal left-inverse $\tilde{F}_0$ inverts A_0 over $R(A_0)$ and acts as the zero operator on $R(A_0)^\perp$. The unique left-inverse with these properties is the Moore-Penrose pseudoinverse, denoted by A_0^+, of A_0 [16]. Explicitly, A_0^+ can be expressed as

$$A_0^+ = (A_0^\dagger A_0)^{-1} A_0^\dagger = (F_0 A_0)^{-1} F_0. \tag{3.59}$$

We have $A_0^+ = F_{\text{can}}$ and therefore the optimal left-inverse is $\tilde{F}_0 = A_0^+ = F_{\text{can}}$. The optimal synthesis frame is the canonical dual frame of $\{|f_k\rangle\}$. The minimum value of $\mathcal{E}$ is

$$\mathcal{E}_{\min} = \mathbb{E}\left[||F_{\text{can}} |w_e\rangle||^2\right] = \mathbb{E}\left[||F_{\text{can}} |w_{e1}\rangle||^2\right], \tag{3.60}$$

where $|w_{e1}\rangle \in R(A_0)$ is defined as in Equation (3.57).

3.5.2 Application to Equal-Norm Tight Frames

We simplify the expression for $\mathcal{E}_{\min}$ for the case where $\{|f_k\rangle\}$ is a tight frame for $\mathcal{V}$ with the additional property that all of the frame vectors have the same norm. Such a frame is typically referred to as an equal norm tight frame (ENTF) [15, 16].

Mathematically, we assume that

$$\sum_{k=1}^{M} |\langle f_k|v\rangle|^2 = C\,||v||^2 \;\forall\; |v\rangle \in \mathcal{V}, \tag{3.61a}$$

$$||f_k|| = B, \quad 1 \leq k \leq M, \tag{3.61b}$$

for some constants $B, C > 0$. ENTFs are utilized, for example, in the context of oversampling in classical signal processing to reduce the effect of quantization noise on a bandlimited signal (see Appendix A.7 in [48]). They are also of interest in the quantum physics community in the form of tight IC POVMs as used for quantum state estimation [67].

It was shown in Section 3.5.1 that $\mathcal{E}_{\min} = \mathbb{E}[\,||F_{\text{can}}\,|w_{e1}\rangle||^2\,]$. To further simplify this expression, first note that since $|w_{e1}\rangle$ is an element of $R(A_0)$, there must be some $|v\rangle \in \mathcal{V}$ such that $A_0\,|v\rangle = |w_{e1}\rangle$. We have $F_{\text{can}}A_0\,|v\rangle = |v\rangle$ and thus $||F_{\text{can}}\,|w_{e1}\rangle\,||^2 = ||v||^2$. By definition of a tight frame we have $||w_{e1}||^2 = ||v||^2/C$, and combining these two expressions leads to $||F_{\text{can}}\,|w_{e1}\rangle\,||^2 = ||w_{e1}||^2/C$. Recall that the $\{c_k\}$ have zero-mean, variance σ^2, and are pairwise uncorrelated. This implies that $\mathbb{E}[||w_{e1}||^2] = N\Delta^2$ and so $\mathcal{E}_{\min} = \mathbb{E}[\,|||w_{e1}\rangle||^2\,]/C = N\Delta^2/C$. To emphasize the dependence of $\mathcal{E}_{\min}$ on the number M of frame vectors, we utilize the relationship $CN = MB^2$ which is true for any ENTF satisfying Equations (3.61) [15]. This leads to

$$\mathcal{E}_{\min} = \frac{N^2\,\Delta^2}{M\,B^2}. \tag{3.62}$$

Note that $\mathcal{E}_{\min}$ is proportional to Δ^2/M, so a reduction in the variance of the error values and an increase in the number of frame vectors both individually lead to higher quality reconstruction. When the variances of the $\{e_k\}$ are not assumed to be identical for all values of k, $\{\mathbb{E}[e_k^2] = \Delta_k^2\}$, we have $\mathbb{E}[||\,|w_{e1}\rangle\,||^2] = (N/M)\sum_k \Delta_k^2$ and therefore

$$\mathcal{E}_{\min} = \frac{N}{MC}\sum_{k=1}^{M}\Delta_k^2 = \frac{N^2}{M^2B^2}\sum_{k=1}^{M}\Delta_k^2. \tag{3.63}$$

As expected, Equation (3.63) reduces to Equation (3.62) when $\Delta_k^2 = \Delta^2$ for all

$1 \leq k \leq M$.

Chapter 4

Operating Characteristics for Quantum Binary State Discrimination

In Chapter 4 we state the quantum binary state discrimination problem considered in this thesis using the terminology and notation developed in Chapter 2. In Chapter 5 the same problem is phrased and interpreted using the mathematical concepts related to linear algebra and operator spaces that were introduced in Chapter 3. The terminology that we use surrounding quantum measurement is discussed briefly in Section 4.1. The postulates of quantum mechanics that are relevant to this thesis are stated in Section 4.2. This allows us to formally state the quantum binary state discrimination problem in Section 4.3. In Section 4.4 we review Helstrom's well-known result regarding minimum probability of error decision strategies for quantum binary state discriminiation. Helstrom's result is the counterpart to the classical MPE decision rules reviewed in Section 2.2.1. Examples of decision and measurement operating characteristics in the quantum setting are given in Sections 4.5 and 4.6.

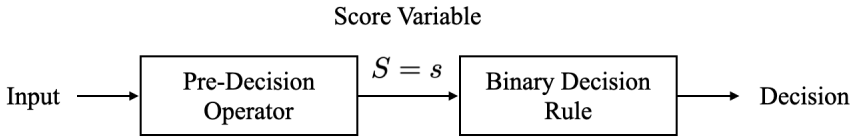


Figure 4-1: Binary hypothesis testing framework.

4.1 Preliminaries

There are many ways in which classical and quantum systems differ and correspondingly so do many of the issues related to hypothesis testing. Much of the terminology related to quantum mechanics is phrased somewhat differently depending on whether it is presented or described more from a physical and experimental perspective or from a mathematical perspective. Quantum phenomena inherently occur in the physical world. However, the fundamental underpinnings of the mathematical analysis of quantum phenomena rely on a representation of quantum states as vectors or operators in a Hilbert space. While the mathematics provides the tools to make predictions about the outcome of experiments, the experiments themselves occur in the physical world. As succinctly phrased by Asher Peres in his book [58],

> *Quantum phenomena do not occur in a Hilbert space. They occur in a laboratory.*

Our focus in the remainder of the monograph is on the mathematics and the representation of quantum states and operations on those states abstractly in Hilbert space. And the terminology that we use will correspond to that representation. Consequently in this preliminary section we define several significant terms as we will be using them in the subsequent discussion. For our purposes, the *state* of a quantum system will refer to the density operator associated with the physical procedure used to prepare the system in a laboratory. The density operator is a mathematical representation capturing all that is known about the system prior to measurements on it. And the ways in which information can be obtained about the state through measurement are constrained by the postulates of quantum mechanics. A key aspect

of the postulates is the meaning of and constraints on the concept of measurement. In all scenarios it is necessary to make a distinction between the word *measurement* as it refers to a specified experimental setup in a real or hypothetical laboratory and as it refers to the laws of classical or quantum physics that model our knowledge of the interaction of the laboratory equipment with the object or system we wish to measure. In this thesis we borrow from the terminology in [49] in which every quantum measurement is "described by a collection of measurement operators $\{A_k\}$...operating on the state space of the system being measured". We use the term *measurement* to refer to the collection of operators $\{A_k\}$. We will assume in addition that the index k satisfies $1 \leq k \leq M$. When the measurement is made the state of system being measured changes in a probabilistic manner to a new state whose value depends on both the original state and on one of the $\{A_k\}$. Thus there are M possible measurement outcomes that can occur, each associated with a value of the index k in the set of operators. We will only be concerned with the value of the index k representing the operator used to compute the post-measurement state, and not with the value of the post-measurement state itself, and consequently we will use the term *measurement outcome* to refer to that index.

4.2 The Postulates of Quantum Mechanics

While there are in total four postulates of quantum mechanics, in Section 4.2 and throughout this thesis we focus on the two postulates that relate specifically to this monograph. Both are loosely paraphrased from Chapter 2 of [49].

Quantum State Postulate. *The state of an isolated physical system can be represented by a density operator ρ that acts on a complex Hilbert space $\mathcal{H}$. $\mathcal{H}$ is often referred to as the state space of the system. We assume for convenience that $\mathcal{H}$ is finite dimensional with dimension d. A given density operator can always be written in the form*

$$\rho = \sum_{j=1}^{D} a_j \left|\psi_j\right\rangle \left\langle\psi_j\right|, \tag{4.1}$$

where $D > 0$ is an integer, the $\{a_j\}$ are probabilities, and the $\{|\psi_j\rangle\}$ are unit vectors in $\mathcal{H}$ that are often referred to as state vectors. For a given density operator ρ, it is well-known that the value of D, the $\{a_j\}$, and the $\{|\psi_j\rangle\}$ are in general not unique. We remark on this fact further below.

Quantum Measurement Postulate. *Quantum measurements are described by a collection $\{A_k\}$ of measurement operators that act on the state space $\mathcal{H}$ of the system being measured. Each value of the index k corresponds to a different possible measurement outcome. In this thesis we will assume for convenience that there are a finite number, M, of elements, and that $1 \leq k \leq M$. Measurement elements satisfy a completeness relation on $\mathcal{H}$,*

$$\sum_{k=1}^{M} A_k^\dagger A_k = I, \tag{4.2}$$

where I is the identity operator on $\mathcal{H}$. If the state of the system is described by the density operator $\rho = \sum_j a_j |\psi_j\rangle \langle\psi_j|$ immediately before a measurement described by the operators $\{A_k\}$, then with probability

$$p(k) = \sum_{j=1}^{D} a_j \langle\psi_j| A_k^\dagger A_k |\psi_j\rangle \tag{4.3}$$

the kth measurement outcome occurs. It is often convenient to write $p(k)$ using the trace operator $\mathrm{Tr}(\cdot)$,

$$p(k) = \mathrm{Tr}\left(A_k^\dagger A_k \rho\right). \tag{4.4}$$

A derivation of Equation (4.4) from Equation (4.3) is given below. Once observed, the kth measurement outcome indicates that the state of the system has collapsed to the kth post-measurement state, denoted by ρ_k. The value of ρ_k is determined by the pre-measurement state ρ and the measurement operator A_k and is not relevant to this monograph. An exact expression can be found in [49].

A density operator in the form of Equation (4.1) represents a quantum system that has been prepared in the state $\rho_j = |\psi_j\rangle \langle\psi_j|$ with probability a_j [49]. If only one of the $\{a_j\}$ is non-zero, i.e., $a_m = 1$ for some $1 \leq m \leq D$ and $a_j = 0$ for $j \neq m$,

then $\rho = |\psi_m\rangle \langle\psi_m|$ is said to represent a pure state. The state vector $|\psi_m\rangle$ is itself also often referred to as a pure state. If more than one of the $\{a_j\}$ is non-zero, then ρ is referred to as a mixed state – that is, a probabilistic mixture of the pure states $\{|\psi_j\rangle\}$. Regardless of the state vectors and probabilities that are used, Equation (4.1) implies that ρ is always a positive semidefinite Hermitian operator with trace 1. This is because

$$\mathrm{Tr}(\rho) = \sum_{j=1}^{D} a_j \ \mathrm{Tr}\left(|\psi_j\rangle \langle\psi_j|\right) = \sum_{j=1}^{D} a_j = 1, \tag{4.5}$$

where we have used the linearity of the trace and the fact that for any vector $|x\rangle \in \mathcal{H}$ we have $\mathrm{Tr}\left(|x\rangle \langle x|\right) = ||x||^2$. Since ρ is Hermitian it can always be written in terms of its eigenvalues and eigenvectors as

$$\rho = \sum_{j=1}^{d} \lambda_j |x_j\rangle \langle x_j|, \tag{4.6}$$

where the $\{\lambda_j\}$ are real and the $\{|x_j\rangle\}$ are orthogonal. Equation (4.6) is a special case of Equation (4.1) with $D = d$ and $\{a_j = \lambda_j\}$. The $\{\lambda_j\}$ are always valid probabilities as a result of the fact that ρ is positive semidefinite and has trace 1. Therefore, a given density operator ρ can always be interpreted as the state of a system that has been prepared in the state $\rho_j = |x_j\rangle \langle x_j|$ with probability λ_j. For convenience, we will continue to specify density operators using their eigendecompositions as in Equation (4.6).

Regarding the quantum measurement postulate, Equation (4.4) can be derived from Equation (4.3) via

$$p(k) = \sum_{j=1}^{D} a_j \ \mathrm{Tr}\left(\langle\psi_j| A_k^\dagger A_k |\psi_j\rangle\right) = \mathrm{Tr}\left(\sum_{j=1}^{D} a_j \ \langle\psi_j| A_k^\dagger A_k |\psi_j\rangle\right) = \mathrm{Tr}\left(A_k^\dagger A_k \, \rho\right). \tag{4.7}$$

In Equation (4.7) we have used both the fact that the trace of a scalar is itself and the cyclic property of the trace, $\mathrm{Tr}(A\,B) = \mathrm{Tr}(B\,A)$ for any two suitable linear operators A and B. Aside from notational convenience, as noted in Chapter 3 the trace operator is also useful because it can be interpreted as an inner product function.

Throughout this thesis we will only be concerned with the probability distribution of measurement outcomes, $\{p(k), 1 \leq k \leq M\}$, and not with the corresponding post-measurement states. To that end, note that the $\{p(k)\}$ depend on the measurement operators $\{A_k\}$ only through the operators $\{E_k = A_k^\dagger A_k\}$. By construction the $\{E_k\}$ have the properties

$$E_k = E_k^\dagger \qquad \text{(Hermiticity)} \tag{4.8a}$$

$$\langle x| E_k |x\rangle \geq 0, \text{ for all } |x\rangle \in \mathcal{H} \qquad \text{(positive semidefiniteness)} \tag{4.8b}$$

$$\sum_{k=1}^{M} E_k = I \qquad \text{(completeness).} \tag{4.8c}$$

In functional analysis, a collection of operators satisfying these three properties is referred to as a positive operator-valued measure or POVM [7]. Distinct quantum measurements can have the same corresponding POVM because replacing each A_k by UA_k, where U is a unitary operator on $\mathcal{H}$, preserves the relation $E_k = A_k^\dagger A_k$. It is worth explicitly writing Equations (4.3) and (4.4) in terms of the operator E_k and the eigenvectors and eigenvalues of a given density operator ρ,

$$p(k) = \sum_{j=1}^{d} \lambda_j \langle x_j| E_k |x_j\rangle = \text{Tr}(E_k \rho). \tag{4.9}$$

Equation (4.9) is a crucial relation that is the basis for much of the discussion in Chapter 5. A main focus of Chapter 5 is a particular class of POVMs referred to as informationally complete or IC POVMs. An IC POVM is one that maps each possible density operator to a unique sequence of probabilities [8]. Explicitly, given two density operators ρ_1 and ρ_2 as well as an IC POVM $\{E_k\}$, we have

$$p_1(k) = p_2(k), \quad 1 \leq k \leq M, \tag{4.10}$$

where $p_i(k) = \text{Tr}(E_k \rho_i)$ for $i = 1, 2$, if and only if $\rho_1 = \rho_2$. An important result regarding IC POVMs that is reviewed in Chapter 5 connects each IC POVM to an

overcomplete representation of an operator-valued vector space containing all valid density operators.

A different class of POVMs correspond to the class of quantum measurements referred to as standard measurements, also sometimes referred to as projective or von Neumman measurements. A standard quantum measurement is one for which the measurement operators $\{A_k\}$ form a complete set of orthogonal projectors on $\mathcal{H}$. The POVM elements $\{E_k = A_k^\dagger A_k\}$ of a standard measurement also form a complete set of orthogonal projectors on $\mathcal{H}$. This follows from the fact that orthogonal projection operators are Hermitian and idempotent, so $A_k^\dagger A_k = A_k^2 = A_k$ for all $1 \leq k \leq M$ for a standard measurement. The reverse is also true – if the elements a given POVM $\{E_k\}$ form a complete set of orthogonal projectors on $\mathcal{H}$, then all associated quantum measurements must be standard measurements.

Example 4.1. Consider a density operator $\rho = |\psi\rangle \langle\psi|$ that represents a pure state along with a standard measurement whose elements have the form $\{A_k = |v_k\rangle \langle v_k|\}$ for some orthonormal basis $\{|v_k\rangle\}$ of $\mathcal{H}$. The corresponding standard POVM is $\{E_k = |v_k\rangle \langle v_k|\}$. It is straightforward to verify that the $\{E_k\}$ satisfy the three conditions specified in Equation (4.8). When the measurement is made, the kth measurement outcome occurs with probability

$$p(k) = \mathrm{Tr}\left(E_k\, \rho\right) = |\, \langle v_k|\psi\rangle\, |^2. \tag{4.11}$$

Equation (4.11) states that the kth measurement outcome occurs with a probability equal to the squared magnitude of the component of $|x\rangle$ in the direction of $|v_k\rangle$. If $|x\rangle$ is orthogonal to $|v_k\rangle$ then the kth measurement outcome has zero probability of occurring.

4.3 Quantum Binary State Discrimination

For the remainder of the thesis we consider the problem where the two possible hypotheses H_0 and H_1 correspond to two possible physical environments or preparation

procedures that have resulted in L quantum mechanical systems that can all be described by the same density operator, either ρ_0 or ρ_1. As in Chapter 2, the prior probabilities will continue to be denoted by $P(H = H_0) = q_0$ and $P(H = H_1) = q_1$. The eigendecompositions of ρ_0 and ρ_1 will be denoted as

$$\rho_0 = \sum_{j=1}^{d} a_j \left|x_j\right\rangle \left\langle x_j\right|, \tag{4.12a}$$

$$\rho_1 = \sum_{j=1}^{d} b_j \left|y_j\right\rangle \left\langle y_j\right|, \tag{4.12b}$$

where the $\{|x_j\rangle\}$ and $\{|y_j\rangle\}$ each form orthonormal bases of $\mathcal{H}$ and the $\{a_j\}$ and $\{b_j\}$ are probabilities. To avoid ambiguity between the discrimination or hypothesis testing system and the quantum mechanical system, from this point forward we will abbreviate the latter as the QMS. To discriminate between the two hypotheses, each of the L QMSs is measured individually using a quantum measurement whose associated POVM is $\{E_k, 1 \leq k \leq M\}$. The score variable is equal to the vector of relative frequencies corresponding to the frequency of occurrence of each of the M possible outcomes. The decision region $\mathcal{D}$ of the binary decision rule is some subset of the set of all possible relative frequency vectors.

Throughout Chapter 4 we will assume that $L = 1$ for simplicity. In this case the score variable can equivalently be thought of as being equal to one of the index values $1 \leq k \leq M$ and the decision region $\mathcal{D}$ of the binary decision rule is then some subset of $\{1, 2, \ldots, M\}$. For a given decision region $\mathcal{D}$, the conditional distributions of the score variable are

$$f_0(k) = \sum_{j=1}^{d} a_j \left\langle x_j\right| E_k \left|x_j\right\rangle = \mathrm{Tr}\left(E_k \rho_0\right), \quad 1 \leq k \leq d, \tag{4.13a}$$

$$f_1(k) = \sum_{j=1}^{d} b_j \left\langle y_j\right| E_k \left|y_j\right\rangle = \mathrm{Tr}\left(E_k \rho_1\right), \quad 1 \leq k \leq d. \tag{4.13b}$$

Then in analogy with Equation (2.4), the probabilities of false alarm and detection

are

$$P_f = \sum_{k\in\mathcal{D}} f_0(k) = \sum_{k\in\mathcal{D}} \mathrm{Tr}\left(E_k\,\rho_0\right) = \mathrm{Tr}\left[\left(\sum_{k\in\mathcal{D}} E_k\right)\rho_0\right] \tag{4.14a}$$

$$P_d = \sum_{k\in\mathcal{D}} f_1(k) = \sum_{k\in\mathcal{D}} \mathrm{Tr}\left(E_k\,\rho_1\right) = \mathrm{Tr}\left[\left(\sum_{k\in\mathcal{D}} E_k\right)\rho_1\right]. \tag{4.14b}$$

It is not uncommon to assume that the quantum measurement that constitutes the pre-decision operator only has 2 possible outcomes, i.e., $M = 2$. This implies that the score variable can only take on two possible values, which is significant because it implies in turn that the decision region of the binary decision rule can only take on four possible values: $\mathcal{D} = \{\}$ (the empty set), $\mathcal{D} = \{1\}$, $\mathcal{D} = \{2\}$, or $\mathcal{D} = \{1, 2\}$. Recall that classical ROCs are generated by varying $\mathcal{D}$ in order to achieve different operating points in the P_f-P_d plane, with distinct operating points corresponding to distinct decision regions. When $M = 2$ in a quantum binary hypothesis testing system, there are only four possible operating points on an operating characteristic analogous to a classical ROC. Moreover, two of those operating points are $(P_f, P_d) = (0, 0)$ and $(P_f, P_d) = (1, 1)$, which correspond to ignoring the outcome of the measurement and consistently deciding either $\hat{H} = H_0$ or $\hat{H} = H_1$, respectively. This lack of flexibility is different from classical ROCs, which are typically used in scenarios where there is a large range – possibly even a continuous range – of potential operating points that are "weighed" against each other using various optimality criteria. It is of course important to remember that there are many alternative formulations of quantum binary hypothesis testing in which this is not the case.

When we expand to the case were $L > 1$ in Chapter 5, we will denote the score variable by $\mathbf{S}$ as opposed to S to emphasize that it is a vector-valued random variable as opposed to a scalar random variable. A particular realization $\mathbf{s}$ will be denoted by the column vector $\mathbf{s} = [n_1/L, \ldots, n_M/L]^T$ where n_k is the number of occurrences of the kth measurement outcome. Clearly, $\sum_k n_k = L$. The conditional distributions $f_0(\cdot)$ and $f_1(\cdot)$ of the score variable are multinomial distributions.

4.4 Minimum Probability of Error Decision Rules

In analogy with the classical MPE decision rules described in Section 2.2.1, we summarize Helstrom's well-known result [38] regarding discrimination between two fixed density operators with minimum probability of error. We refer the reader to [38] for the complete derivation and a generalization of the result to the minimum risk error criterion. From this point forward, the word "optimal" will be used specifically to describe systems that achieve minimum probability of error unless otherwise specified. Assume that the pre-decision operator is a quantum measurement with POVM $\{E_1, E_2\}$ and that $\mathcal{D} = \{2\}$. That is, if the measurement outcome is $s = 1$ then the final decision is $\hat{H}_0$ and if the measurement outcome is $s = 2$ then the final decision is $\hat{H}_1$. The probability of error can be expressed as

$$P_e = q_1 - q_1 \text{ Tr}\left[E_1 \left(\rho_1 - \frac{q_0}{q_1}\rho_0\right)\right]. \tag{4.15}$$

Helstrom's result utilizes the orthonormal eigenvectors $\{|z_j\rangle, 1 \leq j \leq d\}$ and real eigenvalues $\{\eta_j, 1 \leq j \leq d\}$ of the operator $(\rho_1 - (q_0/q_1)\rho_0)$. This operator or a scaled version of it is sometimes referred to as the Lagrange operator [ref]. Helstrom showed that the probability of error is minimized when E_1 is the orthogonal projector onto the subspace $\mathcal{U}_1 = \text{span}\{|z_j\rangle : \eta_j \geq 0\}$. Since $E_1 + E_2 = I$ this implies that E_2 must be the orthgonal projector onto the subspace $\mathcal{U}_1^{\perp} = \text{span}\{|z_j\rangle : \eta_j < 0\}$, where the superscript $\perp$ indicates an orthogonal complement. Note that the $|z_j\rangle$ for which $\eta_j = 0$ may be included in either subspace without changing the probability of error. The optimal POVM elements can be written as

$$E_1 = \sum_{j:\eta_j \geq 0} |z_j\rangle \langle z_j|, \quad E_2 = \sum_{j:\eta_j < 0} |z_j\rangle \langle z_j|. \tag{4.16}$$

Helstrom noted that an equivalent way of achieving minimum probability of error is to use the d-outcome POVM with elements $E_k = |z_k\rangle \langle z_k|, \quad 1 \leq k \leq d$. If the measurement outcome is $s = k$ where $\eta_k \geq 0$, then the final decision is $\hat{H}_1$, otherwise the final decision is $\hat{H}_0$. Equivalently, $\mathcal{D} = \{k : \eta_k \geq 0\}$. Both of these POVMs have

the property that the elements form complete sets of orthogonal projectors on $\mathcal{H}$, so they each correspond to standard quantum measurements.

4.5 Decision Operating Characteristics for Quantum Systems

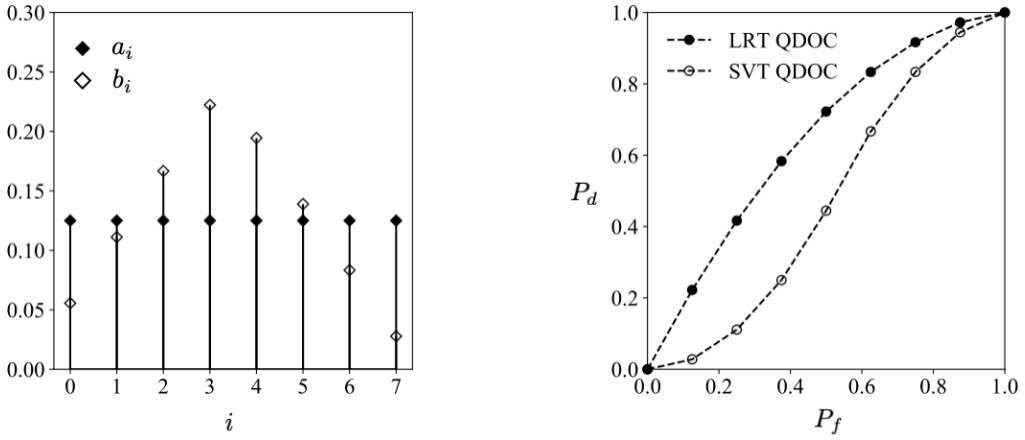
A performance characteristic analogous to classical ROCs can be made for the quantum case by fixing the quantum measurement that constitutes the pre-decision operator and varying the decision region of the binary decision rule. We refer to such an operating characteristic as a quantum decision operating characteristic or QDOC. In the example below we show how the result presented in Section 2.5.1 can be applied to QDOCs.

Example 4.2. For this example we set the dimension of $\mathcal{H}$ to $d = 8$ and we set $|x_j\rangle = |y_j\rangle = |e_j\rangle$, $1 \leq j \leq 8$, where $\{|e_j\rangle\}$ is any orthonormal basis for $\mathcal{H}$. Note that $\mathcal{H}$ is isomorphic to $\mathbb{C}^8$. The probabilities $\{a_j\}$ and $\{b_j\}$ are arbitrarily chosen to be the uniform distribution and an asymmetric triangular distribution, respectively, as shown in Figure 4-2a. We have $a_j = 1/8$ for $1 \leq j \leq 8$ and $b_1 = 2/32, b_2 = 4/32, b_3 = 6/32, b_4 = 8/32, b_5 = 7/32, b_6 = 5/32, b_7 = 3/32, b_8 = 1/32$. We assume that the pre-decision operator is an 8-outcome standard quantum measurement with associated POVM elements $E_k = |e_k\rangle \langle e_k|$, $1 \leq k \leq 8$. According to Equation (4.13) the conditional distributions of the score variable are

$$f_0(k) = \sum_{j=1}^{8} a_j \langle e_j|e_k\rangle \langle e_k|e_j\rangle = a_k, \quad 1 \leq k \leq 8, \tag{4.17a}$$

$$f_1(k) = \sum_{j=1}^{8} b_j \langle e_j|e_k\rangle \langle e_k|e_j\rangle = b_k, \quad 1 \leq k \leq 8, \tag{4.17b}$$

where we have used the fact that the $\{|e_j\rangle\}$ are orthonormal. The LRT QDOC for this POVM is indicated by the solid black circles shown in Figure 4-2b. Unlike an LRT decision region, an SVT decision region and therefore an SVT QDOC inherently



(a) (b)

Figure 4-2: (a) Conditional distributions of the score variable as given in Equation (4.12). (b) QDOCs generated using LRT or SVT decision regions.

depends on the choice of ordering of the POVM elements. Since the index values $k \in \{1, \ldots, 8\}$ represent convenient labels corresponding to the possible measurement outcomes as opposed to actual numerical values, the ordering is arbitrary. Distinct orderings correspond to distinct shapes of the conditional PMFs $f_0(\cdot)$ and $f_1(\cdot)$. In Figure 4-2a we have assumed the natural ordering from $k = 1$ to $k = 8$, and this results in the SVT QDOC represented by the hollow black circles in Figure 4-2b. Linear interpolation was used between the points to aid in visualization of the shapes of the curves. Of course, any operating point on any of the line segments could be achieved using randomization between two LRT or SVT decision regions [39]. The constructive procedure described in Section 2.5.2 could be used to reconstruct the LRT QDOC from the SVT QDOC without any explicit knowledge of ρ_0, ρ_1, or any of the $\{E_k\}$. The same would be true for any two density operators ρ_0 and ρ_1 (whose eigenvectors may or may not be the same) along with any POVM $\{E_k\}$.

4.6 Measurement Operating Characteristics for Quantum Systems

An analogous operating characteristic to the CMOCs discussed in Section 2.3 for the quantum case can be generated by keeping the decision regions of the binary decision rule fixed while varying the parameters of the quantum measurement that constitutes the pre-decision operator. We refer to this type of operating characteristic as a quantum measurement operating characteristic or QMOC. The operating characteristics defined by Bodor and Koniorczyk in [9] are QMOCs in our terminology. Examples 4.3 and 4.4 below below were directly motivated by the analysis and examples given in [9].

In Example 4.3 we set the dimension of $\mathcal{H}$ to $d = 2$ and demonstrate the effects of various parameters of ρ_0 and ρ_1 on the shape of the QMOC generated using all possible standard measurements. As noted in [9], the optimal operating points for all possible prior probabilities q_0 and q_1 lie on an ellipse. It is also pointed out in [9] that this is not true in general for $d > 2$. For arbitrary mixed states with $d > 2$, the collection of optimal operating points for all possible priors do not lie on an ellipse, but rather on a series of disjoint segments in the P_f-P_d plane. Using as motivation the simulations in [9], we demonstrate this in Example 4.4. We additionally demonstrate in Example 4.4, as is also shown in [9], that the operating points corresponding to a large number of randomly chosen standard POVMs (some of which are not optimal for any set of prior probabilities) form clusters in the P_f-P_d plane. Each cluster corresponds to a different pair of values for the ranks of the POVM elements.

Example 4.3. For this example we set $d = 2$, so $\mathcal{H}$ is isomorphic to $\mathbb{C}^2$. As in Equations (4.12) we denote the eigenvectors and eigenvalues of ρ_0 by $\{|x_i\rangle\}$ and $\{a_i\}$, respectively. We arbitrarily set $a_1 = 1/15$ and $a_2 = 14/15$. We may always express the $\{|x_i\rangle\}$ and $\{|y_i\rangle\}$ as column vectors whose elements are the coefficients in a basis

expansion in the $\{|x_i\rangle\}$ basis (see Section 3.1),

$$|x_1\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad |x_2\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{4.18a}$$

$$|y_1\rangle = \begin{bmatrix} \cos(\beta/2) \\ \sin(\beta/2) \end{bmatrix}, \quad |y_2\rangle = \begin{bmatrix} -\sin(\beta/2) \\ \cos(\beta/2) \end{bmatrix}. \tag{4.18b}$$

The angle β satisfies $\cos(\beta/2) = \langle x_1|y_1\rangle = \langle x_2|y_2\rangle$ and is a measure of the degree of separation between the ρ_0 and ρ_1. The pre-decision operator is assumed to be a standard measurement with associated POVM $\{E_1 = |v_1\rangle\langle v_1|, E_2 = |v_2\rangle\langle v_2|\}$, where

$$|v_1\rangle = \begin{bmatrix} -\sin(\theta/2) \\ \cos(\theta/2) \end{bmatrix}, \quad |v_2\rangle = \begin{bmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{bmatrix} \tag{4.19}$$

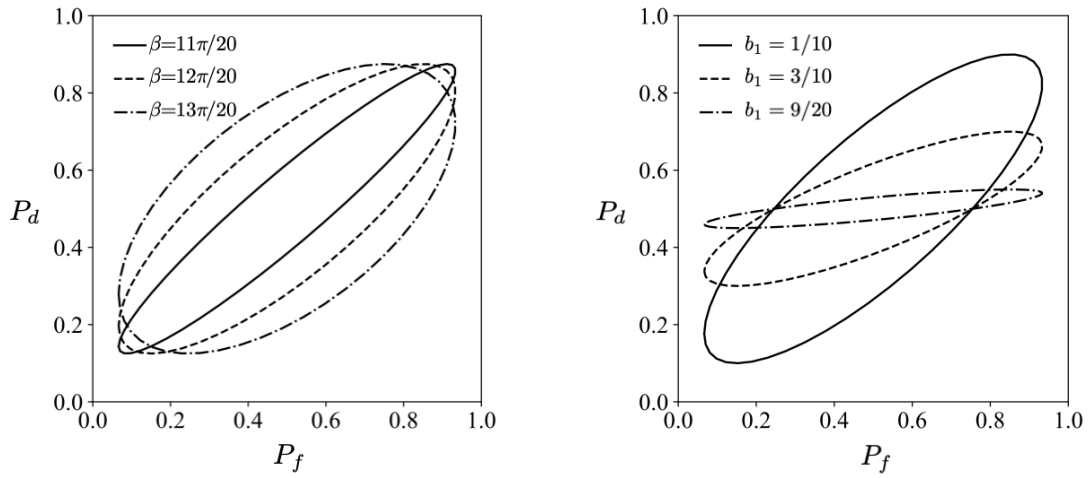
for some angle θ. The decision region of the binary decision rule is $\mathcal{D} = \{2\}$. A QMOC can be generated by fixing the values of β, b_1, and b_2 and varying the angle θ. It is straightforward to show that

$$P_f = \mathrm{Tr}(E_2\,\rho_0) = a_1\,\cos^2\left(\frac{\theta}{2}\right) + a_2\,\sin^2\left(\frac{\theta}{2}\right) \tag{4.20a}$$

$$P_d = \mathrm{Tr}(E_2\,\rho_1) = b_1\,\cos^2\left(\frac{\theta-\beta}{2}\right) + b_2\,\sin^2\left(\frac{\theta-\beta}{2}\right). \tag{4.20b}$$

In Section 4.6.1 we show that Equations (4.20) correspond to the parametric formula for an ellipse. This was stated but not explicitly proven in [9]. Explicit formulas for the parameters of the ellipse in terms of the $\{a_i\}$, the $\{b_i\}$, and α are also given in Section 4.6.1.

Figure 4-3 shows a collection of QMOCs each generated by fixing the values of β, b_1, and b_2 and varying the angle θ. In Figure 4-3a, b_1 and b_2 are arbitrarily fixed to $b_1 = 3/4$ and $b_2 = 1/4$ and each QMOC corresponds to a different value of β. As β approaches 0, the eccentricity of the ellipse increases. In Figure 4-3b, β is arbitrarily fixed to $\beta = \pi/5$ while b_1 and b_2 are varied. As b_1 and b_2 approach $1/2$, the ellipse becomes more concentrated around the line $P_d = 1/2$. Indeed, it is straightforward



(a) (b)

Figure 4-3: (a) QMOCs generated with $d = 2$ for a fixed ρ_0 by varying the parameters of ρ_1 and the standard measurement that constitutes the pre-decison operator.

to show that the QMOC is inscribed in the rectangle with sides $P_f = \min\{a_1, a_2\}$, $P_f = \max\{a_1, a_2\}$, $P_d = \min\{b_1, b_2\}$, $P_d = \max\{b_1, b_2\}$.

Example 4.4. We now set $d = 8$, so $\mathcal{H}$ is isomorphic to $\mathbb{C}^8$, and describe the collection of operating points that results from performing Helstrom's MPE decision strategy for a range of prior probabilities q_0 and q_1. When the operating points of a large number of randomly chosen standard measurements are plotted, the result is a series of clusters in the P_f-P_d plane.

The eigenvectors of ρ_0 are set to $|x_j\rangle = |e_j\rangle$, $1 \leq j \leq 8$, while the eigenvectors of ρ_1 are the vectors in an arbitrarily chosen orthonormal basis of $\mathbb{C}^8$. The $\{a_i\}$ and $\{b_i\}$ are arbitrarily set to $a_1 = 1/141, a_2 = 5/141, a_3 = 10/141, a_4 = 15/141, a_5 = 20/141, a_6 = 25/141, a_7 = 30/141, a_8 = 35/141$ and $b_1 = 35/141, b_2 = 30/141, b_3 = 25/141, b_4 = 20/141, b_5 = 15/141, b_6 = 10/141, b_7 = 5/141, b_8 = 1/141$. The prior probabilities q_0 and q_1 are varied over their entire ranges from 0 to 1. For each pair of priors, the Lagrange operator $(\rho_1 - (q_0/q_1)\rho_0)$ is formed and its eigendecomposition is computed in order to identify Helstrom's POVM elements E_1 and E_2 as defined in

Equations (4.16). The MPE operating point then has coordinates

$$P_f = \mathrm{Tr}\left(E_2\,\rho_0\right) = \sum_{j=1}^{8} a_j \left\langle x_j\right| E_2 \left|x_j\right\rangle , \tag{4.21a}$$

$$P_d = \mathrm{Tr}\left(E_2\,\rho_1\right) = \sum_{j=1}^{8} b_j \left\langle y_j\right| E_2 \left|y_j\right\rangle . \tag{4.21b}$$

The result is the collection of upper operating points shown in Figure 4-4. They form $(d-1) = 7$ disjoint segments, in addition to the points $(0,0)$ (optimal for $q_1 = 0$) and $(1,1)$ (optimal for $q_1 = 1$). This is characteristic of the type of plot that results from other arbitrary density operators ρ_0 and ρ_1 and for other values of $d > 2$. As noted in [9], each pair of prior probabilities q_0 and q_1 corresponds to a different decomposition of $\mathcal{H}$ in terms of Helstrom's orthogonal subspaces $\mathcal{U}_1$ and $\mathcal{U}_2$. The discontinuities between the segments in Figure 4-4 correspond to changes in the dimension of $\mathcal{U}_2$ (equivalently, the number of non-negative eigenvalues of $(\rho_1 - (q_1/q_0)\rho_0)$). The exception to this pattern is the case where ρ_0 and ρ_1 represent two pure states with $d > 2$, since in that case the problem essentially reduces to the case where $d = 2$, with the effective state space being the two-dimensional subspace spanned by the two pure states. In that case as stated in Example 4.3, the optimal operating points for all sets of priors lie on an ellipse.

There are of course many different ways to decompose $\mathcal{H}$ into a combination of two orthogonal subspaces. Each decomposition corresponds to a different (potentially suboptimal) two-outcome standard measurement that can be used to distinguish between ρ_0 and ρ_1. When randomly chosen two-outcome standard measurements are used in this way, the corresponding operating points form a series of $(d-1)$ clusters in the P_f-P_d plane that are apparently centered along the line $P_f = P_d$. This is shown by the lower operating points in Figure 4-4. Each cluster corresponds to a different pair of dimensions for the orthogonal subspaces [9]. The fact that the clusters contain points that are not on any of the disjoint segments of optimal operating points is a reflection of the fact that not every decomposition of $\mathcal{H}$ into two orthogonal subspaces is optimal for some set of priors.

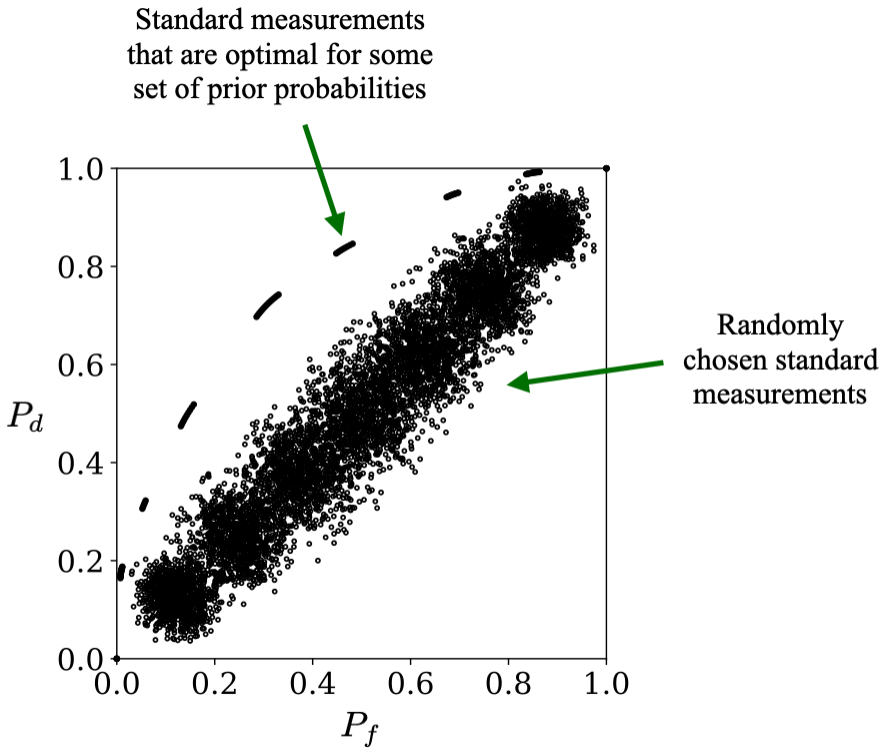


Figure 4-4: Operating points obtained by 2-outcome standard measurements performed on arbitrarily chosen density operators ρ_0 and ρ_1 with $d = 8$. *Upper segments of operating points:* Minimum probability of error operating points for a range of prior probabilities, $0 \leq q_1 \leq 1$. *Lower clusters of operating points:* Operating points obtained by randomly chosen two-outcome standard measurements. Many of these measurements are not optimal for any pair of prior probabilities.

4.6.1 QMOCs Generated using Standard Qubit Measurements are Ellipses

We show that any QMOC generated according to the method described in Example 4.3 of Section 4.6, in which two-outcome quantum measurements with associated standard POVMs are used to distinguish between arbitrary qubit density matrices ρ_0 and ρ_1 with $d = 2$, is an ellipse. More specifically, it is a rotated ellipse in the P_f-P_d plane centered at the point $(1/2, 1/2)$. The derivation also applies to the case where ρ_0 and ρ_1 represent two pure states with $d > 2$, as long as the standard POVMs used to generate the QMOC have the following properties: The first two elements of the POVM, E_1 and E_2, should be analogous to those defined by Equation (4.19), but with the additional requirement that $|v_1\rangle$ and $|v_2\rangle$ should lie in the plane defined by the two pure states. The other measurement elements must therefore project onto subspaces of the orthogonal complement of that plane. Again the final decision is $\hat{H}_1$ if the measurement outcome associated with E_2 occurs and $\hat{H}_0$ if the measurement outcome associated with E_2 occurs. The other possible outcomes have zero probability of occurring and can be associated with either final decision. Essentially, this reduces the problem to that of distinguishing between two pure states with $d = 2$.

The coordinates of the QMOC in terms of the angle θ are

$$P_f = \mathrm{Tr}(E_1\rho_0) = a_0 \cos^2\left(\frac{\theta}{2}\right) + a_1 \sin^2\left(\frac{\theta}{2}\right) \tag{4.22a}$$

$$P_d = \mathrm{Tr}(E_1\rho_1) = b_0 \cos^2\left(\frac{\theta-\beta}{2}\right) + b_1 \sin^2\left(\frac{\theta-\beta}{2}\right). \tag{4.22b}$$

Assuming for the moment that this is the parametric equation of a rotated ellipse centered at $(1/2, 1/2)$, we can center the ellipse at the origin and use trigonometric identities to derive equations for the centered coordinates,

$$P_f - \frac{1}{2} = \frac{a_0 - a_1}{2}\cos\theta \tag{4.23a}$$

$$P_f - \frac{1}{2} = \frac{b_0 - b_1}{2}\cos(\theta-\beta). \tag{4.23b}$$

For ease of notation we now make the substitutions

$$x = P_f - \frac{1}{2}, \quad y = P_d - \frac{1}{2}, \quad a = \frac{a_0 - a_1}{2}, \quad b = \frac{b_0 - b_1}{2}, \tag{4.24}$$

and introduce the functions $f_x(\cdot)$ and $f_y(\cdot)$, so that the centered coordinates become

$$x = f_x(\theta) = a\cos\theta \tag{4.25a}$$

$$y = f_y(\theta) = b\cos(\theta - \beta). \tag{4.25b}$$

(Note that the x and y above should not be confused with the $\{|x_i\rangle\}$ and $\{|y_i\rangle\}$ in Equations (4.12).) The objective now is to show that $x = f_x(\theta)$ and $y = f_y(\theta)$ represent a rotated ellipse centered at the origin. That is, the objective is to show that they can be written in the form

$$x = g_x(t) = q\cos\phi\cos t - r\sin\phi\sin t \tag{4.26a}$$

$$y = g_y(t) = q\sin\phi\cos t + r\cos\phi\sin t \tag{4.26b}$$

for some angle of rotation ϕ from the horizontal, semi-major axis q, semi-minor axis r, and parameter t (which will prove inconsequential for our purposes). The functions $g_x(\cdot)$ and $g_y(\cdot)$ have been introduced for convenience. We can solve for the parameters q, r, ϕ in terms of the known values of a, b, β by using Equations (4.25) and (4.26) to find the points on each ellipse with maximum x- and y-values and then setting their coordinates equal to one another. Taking the derivative of $f_x(\theta)$ and setting it to zero, we find that the point with maximum x-value occurs at $\theta_x = 0$ and has coordinates $(f_x(0), f_y(0)) = (a, b)$. The point with maximum y-value occurs at $\theta_y = \beta$ and has coordinates $(f_x(\beta), f_y(\beta)) = (a\cos\beta, b)$. Similarly, the point on the ellipse described by Equations (4.26) with maximum x-value occurs at $t_x = \tan^{-1}(-(r/q)\tan\phi)$ and

has coordinates

$$g_x(t_x) = \sqrt{q^2 \cos^2\phi + r^2 \sin^2\phi} \tag{4.27a}$$

$$g_y(t_x) = \frac{q^2 - r^2}{\sqrt{q^2/\sin^2\phi + r^2/\cos^2\phi}}. \tag{4.27b}$$

The point with maximum y-value occurs at $t_y = \tan^{-1}(r/(q\tan\phi))$ and has coordinates

$$g_x(t_y) = \frac{q^2 - r^2}{\sqrt{q^2/\cos^2\phi + r^2/\sin^2\phi}} \tag{4.28a}$$

$$g_y(t_y) = \sqrt{q^2 \sin^2\phi + r^2 \cos^2\phi}. \tag{4.28b}$$

Setting $f_x(0) = g_x(t_x)$, $f_y(0) = g_y(t_x)$, $f_x(\beta) = g_x(t_y)$, and $f_y(\beta) = g_y(t_y)$ and solving for q, r, and ϕ in terms of a, b, and β yields

$$\phi = \frac{1}{2}\tan^{-1}\left(\frac{2ab\cos\beta}{a^2 - b^2}\right) \tag{4.29a}$$

$$q = \left[\frac{1}{2}\left(a^2 + b^2 + \frac{a^2 - b^2}{\cos(2\phi)}\right)\right]^{1/2} \tag{4.29b}$$

$$r = \left[\frac{1}{2}\left(a^2 + b^2 - \frac{a^2 - b^2}{\cos(2\phi)}\right)\right]^{1/2}. \tag{4.29c}$$

It can be verified through straightforward algebra that when ϕ, q, and r are given by Equations (4.29), the coordinates $x = f_x(\theta)$, $y = f_y(\theta)$ in Equation (4.25) satisfy the equation that defines an ellipse: $Ax^2 + Bxy + Cy^2 + D = 0$ with $B^2 - 4AC < 0$,

where

$$A = q^2 \sin^2 \phi + r^2 \cos^2 \phi \tag{4.30a}$$

$$B = 2(q^2 - r^2) \sin \phi \cos \phi \tag{4.30b}$$

$$C = q^2 \cos^2 \phi + r^2 \sin^2 \phi \tag{4.30c}$$

$$D = -q^2 r^2. \tag{4.30d}$$

This verifies our initial assumption that $x = f_x(\theta)$ and $y = f_y(\theta)$ are the coordinates of an ellipse that is centered at the origin, rotated by an angle ϕ from the horizontal, and has semi-major axis q and semi-minor axis r. The original QMOC is the same ellipse centered at the point $(1/2, 1/2)$.

Chapter 5

An Operator Space View of Quantum Binary State Discrimination

The main objective of Chapter 5 is to utilize the mathematical methodology developed in Chapter 3 to interpret the process of quantum measurement and the problem of quantum binary state discrimination as described in Chapter 4. In Section 5.1 we apply the concepts of operator spaces outlined in Section 3.4 to a quantum mechanical setting. This includes a brief discussion of Naimark's Theorem as it is typically stated in the quantum mechanics literature and a comparison to the version of the theorem given in Section 3.2. We emphasize that there are many alternative formulations of the connection between quantum measurements and classical frame theory [8, 25, 26, 6, 11, 62]. Informationally complete and overcomplete POVMs are formally defined in Section 5.2. POVMs corresponding to qubit measurements, which we refer to as qubit POVMs for brevity, are of particular interest throughout Chapters 5 and 6. In Section 5.3 we define a class of qubit POVMs referred to as equal trace rank one or Etro POVMs. The representation of a qubit Etro POVM through M points on what we refer to as an Etro sphere is exactly analogous to the representation of a pure state qubit density operator through a point on the Bloch sphere. Simulations regarding the use of Etro POVMs constructed using Platonic solids for qubit binary state discrimination are presented in Section 5.4. In Chapter 6 we generalize this to POVMs specified by other sets of points on an Etro sphere.

5.1 Operator Spaces in Quantum Mechanics

Throughout Chapter 5, $\mathcal{H}$ will always represent the d-dimensional state space of a QMS and $\mathcal{V}$ will denote the d^2-dimensional operator space of all Hermitian operators acting on $\mathcal{H}$. ρ will denote an arbitrary density operator on $\mathcal{H}$ and $\{E_k, 1 \leq k \leq M\}$ will denote an arbitrary POVM on $\mathcal{H}$. ρ and the $\{E_k\}$ are all elements of $\mathcal{V}$ by definition and they are also all positive semidefinite. We have $\mathrm{Tr}(\rho) = 1$ and $\mathrm{Tr}(E_k) \geq 0$ for all $1 \leq k \leq M$. In terms of the operator-valued inner product defined in Equation (3.49), the measurement outcome probabilities $\{p(k)\}$ can be expressed as

$$p(k) = \mathrm{Tr}(E_k\,\rho) = \langle\langle E_k|\rho\rangle\rangle, \quad 1 \leq k \leq M. \tag{5.1}$$

When the $\{E_k\}$ form a frame for $\mathcal{V}$, the $\{p(k)\}$ are equal to the frame coefficients of ρ with respect to the $\{E_k\}$. In Section 5.2 we review an important result that states the $\{E_k\}$ form a frame for $\mathcal{V}$ if and only if $\{E_k\}$ is an IC POVM.

When $\mathcal{H}$ represents the state space of a qubit, decomposing ρ into the sum of its orthogonal projections onto $\mathcal{U}$ and $\mathcal{U}^\perp$ naturally leads to the definition of the commonly used Bloch ball. Decomposing each of the $\{E_k\}$ in the same way will lead to the definition of what we refer to as an Etro sphere whose radius depends on M. Since these decompositions do not inherently rely on $\mathcal{H}$ having dimension $d = 2$, we state them as generally as possible before specifying that $d = 2$ in Section 5.3. According to Equation (3.46), we have

$$|\rho\rangle\rangle = \frac{1}{\sqrt{d}}\,\frac{|I\rangle\rangle}{\sqrt{d}} + \mathcal{P}_{\mathcal{U}}\,|\rho\rangle\rangle \tag{5.2a}$$

$$|E_k\rangle\rangle = \frac{\mathrm{Tr}(E_k)}{\sqrt{d}}\,\frac{|I\rangle\rangle}{\sqrt{d}} + \mathcal{P}_{\mathcal{U}}\,|E_k\rangle\rangle, \quad 1 \leq k \leq M. \tag{5.2b}$$

The requirement that $\sum_k E_k = I$ can be interpreted in terms of Equation (5.2b). Specifically, summing both sides of Equation (5.2b) over all values of k yields

$$|I\rangle\rangle = \frac{1}{d}\left(\sum_{k=1}^{M}\mathrm{Tr}(E_k)\right)|I\rangle\rangle + \left(\sum_{k=1}^{M}\mathcal{P}_{\mathcal{U}}\,|E_k\rangle\rangle\right). \tag{5.3}$$

Equation (5.3) implies that $\sum_k \mathrm{Tr}(E_k) = d$ and, since $|I\rangle\rangle$ is orthogonal to all elements of $\mathcal{U}$, that the sum of the $\{\mathcal{P}_{\mathcal{U}} |E_k\rangle\rangle\}$ must be equal to zero. In Section 5.3 we apply these concepts as well as those described in Section 3.4.3 specifically to POVMs of a qubit, leading to the definition of an Etro sphere.

5.1.1 Naimark's Theorem

We briefly review a version of Naimark's Theorem as it frequently appears in the quantum mechanics literature. In contrast to the variation given in Chapter 3 which was stated in terms of vectors in a Hilbert space, the variation discussed here is stated in terms of Hermitian operators on a given Hilbert space.

Recall that a standard quantum measurement is one whose associated measurement operators $\{A_k\}$ form a complete set of orthogonal projectors on the state space $\mathcal{H}$ of the measured system. The POVM associated with a standard measurement is $\{E_k = A_k^\dagger A_k = A_k\}$. Its elements also form a complete set of orthogonal projectors on $\mathcal{H}$. It is straightforward to show that if the elements of a POVM have this property, then any corresponding quantum measurement must be a standard measurement. A non-standard measurement is of course one whose associated POVM does not have this property. Roughly, Naimark's Theorem as it is typically invoked in quantum mechanics refers to the fact that a non-standard measurement can always be implemented by coupling the QMS that we wish to measure with an ancilla system, possibly performing a unitary operation on the joint system, and then performing a standard measurement on the joint system. Mathematically this boils down to expanding the Hilbert system of the original QMS using a tensor product with the Hilbert system of the ancilla, then choosing an orthonormal basis for the larger space that reproduces the desired outcome probabilities. One reason this is important is that typically only standard measurements are performed in a laboratory, so imposing a mathematical constraint of only using standard measurements has practical consequences concerning implementation. For more details we refer the interested reader to the fascinating references [...] [52, 53, 59, 68] What follows in the remainder of Section 5.1.1 is by no means a full summary of the issues surrounding this topic.

According to the postulates of quantum mechanics, when two QMSs are considered as a joint system the state space of the joint system is the tensor product of the individual state spaces (see Chapter 2 of [49]). Furthermore, the state of one component of a larger joint system can be found by taking the partial trace of state of the joint system. For simplicity in Section 5.1.1 we will only consider pure states and we will only consider quantum measurements whose POVM elements all have rank-one. In addition it will be simplest to use the state vector formalism for quantum states rather than the density operator formalism that we have found preferable in the rest of thesis. The results can be generalized using linearity and the fact that an arbitrary POVM element can be decomposed in many ways as a a sum of rank-one Hermitian elements. Consider a QMS with state vector $|\psi\rangle$ in state space $\mathcal{H}_A$ of dimension d, The objective is to perform a quantum measurement with measurement operators $\{A_k, 1 \leq k \leq M\}$ and corresponding POVM $\{E_k = A_k^\dagger A_k\}$. We assume that $M > d$ implying that the measurement is non-standard. Since the $\{E_k\}$ are rank-one by assumption we have $\{E_k = A_k = |f_k\rangle \langle f_k|\}$ for some vectors $\{|f_k\rangle\}$ in $\mathcal{H}_A$. The $\{|f_k\rangle\}$ form a Parseval frame for $\mathcal{H}_A$ because of the requirement that $\sum_k E_k = I_A$, where I_A is the identity operator on $\mathcal{H}_A$. The measurement outcome probabilities are $\{p(k) = |\langle f_k|\psi\rangle|^2 = \mathrm{Tr}(E_k |\psi\rangle \langle\psi|)\}$ and the post-measurement states are $\{|\psi_k\rangle = A_k |\psi\rangle\}$ up to normalization.

When the original QMS is coupled with an ancilla system with state space $\mathcal{H}_B$, the state space of the joint system is $\mathcal{H}_J = \mathcal{H}_A \otimes \mathcal{H}_B$ where the subscript J stands for joint. We assume that the ancilla system starts in an arbitrary pure state $|\phi\rangle$, implying that the joint system starts in the state $|\psi\rangle \otimes |\phi\rangle$. We may wish to perform a unitary operation on the joint system, resulting in the state $U(|\psi\rangle \otimes |\phi\rangle)$. Consider performing a standard measurement with rank-one measurement operators $\{C_k\}$ and corresponding POVM elements $\{D_k = C_k = |u_k\rangle \langle u_k|\}$ on the joint system in state $U(|\psi\rangle \otimes |\phi\rangle)$. Note that the $\{C_k\}$ and $\{D_k\}$ are operators on $\mathcal{H}_J$ and the $\{|u_k\rangle\}$ form an orthonormal basis for $\mathcal{H}_J$. The measurement outcome probabilities are $\{p_J(k) = |\langle u_k|U(|\psi\rangle \otimes |\phi\rangle)\rangle|^2\}$ and the post-measurement states of the joint system are $\{|\alpha_k\rangle = C_k U(|\psi\rangle \otimes |\phi\rangle)\}$. The post-measurement states $\{|\beta_k\rangle\}$ of the original system can be

obtained by taking the partial trace over $\mathcal{H}_B$ of the $\{|\alpha_k\rangle\}$, $\{|\beta_k\rangle = \mathrm{Tr}_B(|\alpha_k\rangle\langle\alpha_k|)\}$. We are interested in one direction of Naimark's theorem as it is typically invoked in the quantum setting, which states that the $\{C_k\}$ can always be chosen so that $\{p(k) = p_J(k)\}$ and $\{|\psi_k\rangle = |\beta_k\rangle\}$. (A similar statement in the opposite direction can also be made, but that is not included here.) The following examples illustrate in a somewhat preliminary way the concepts that make this statement true. Note that the matching of the post-measurement states are not always of interest, sometimes it is satisfactory to only have $\{p(k) = p_J(k)\}$.

The version of Naimark's Theorem stated in Section 3.3.2 states that there exists an orthonormal basis $\{|w_k\rangle\}$ of some M-dimensional space $\mathcal{H}_\mathcal{W}$ such that $\mathcal{P}_{\mathcal{H}_A}|w_k\rangle = |f_k\rangle$ for all $1 \leq k \leq M$, where $\mathcal{P}_{\mathcal{H}_A}$ is the orthogonal projection operator from $\mathcal{H}_\mathcal{W}$ onto $\mathcal{H}_A$. If the $\{|f_k\rangle\}$ are written in column-vector form as $|f_k\rangle = [f_{k1}\ldots f_{kd}]^T$ with respect to an arbitrary orthonormal basis for $\mathcal{H}_A$, the proof in matrix form essentially boils down to showing that the rows of the matrix

$$G = \begin{bmatrix} f_{11} & \cdots & \cdots & f_{M1} \\ \vdots & \vdots & \vdots & \vdots \\ f_{1d} & \cdots & \cdots & f_{Md} \end{bmatrix} \tag{5.4}$$

are orthogonal. Then Gram-Schmidt can be used to find $(M-d)$ more orthogonal rows, and the columns of the augmented square matrix G', where

$$G' = \begin{bmatrix} f_{11} & \cdots & \cdots & f_{M1} \\ \vdots & \vdots & \vdots & \vdots \\ f_{1d} & \cdots & \cdots & f_{Md} \\ b_{1,d+1} & \cdots & \cdots & b_{M,d+1} \\ \vdots & \vdots & \vdots & \vdots \\ b_{1M} & \cdots & \cdots & b_{MM} \end{bmatrix} \tag{5.5}$$

represent the coefficients of the $\{|w_k\rangle\}$ in some arbitrary orthonormal basis for $\mathcal{H}_\mathcal{W}$. Because of Naimark's identity, the $\{|w_k\rangle\}$ satisfy $|\langle w_k|\psi\rangle|^2 = |\langle f_k|\psi\rangle|^2 = p(k)$ for all $1 \leq k \leq M$. But this is not enough to specify a suitable set of measurement operators $\{C_k\}$ on the larger space, because unless M is a multiple of d, $\mathcal{H}_\mathcal{W}$ can-

not necessarily be expressed as a tensor product of $\mathcal{H}_A$ with another Hilbert space. And this is the crucial difference between the classical and quantum settings. The larger space acted on by G' can be thought of as a direct sum of $\mathcal{H}_A$ with a second, $(M-d)$-dimensional Hilbert space. But in the quantum setting we need to embed the coefficients $\{b_{k,d+1},\ldots,b_{k,M}\}$ for $1 \leq k \leq M$ in a tensor product space whose dimension will by definition be a multiple of d. Examples 5.1 and 5.2 below illustrate two ways of doing so. We assume $d = 2$, so $\mathcal{H}_A \simeq \mathbb{R}^2$, and $M = 3$ for the sake of concreteness and our goal is to specify a suitable set of $\{|u_k\rangle\}$ using the elements of the matrix G'.

Example 5.1. *This technique was described in [Naimark's Theorem and Quantum Inseparability]. Starting from the vectors $|f_k\rangle = [f_{k1}, f_{k2}]^T$ for $1 \leq k \leq 3$, we let $\mathcal{H}_B \simeq \mathbb{R}^2$ and*

$$|u_k\rangle = \begin{bmatrix} f_{k1} \\ f_{k2} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} + b_{3k} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} f_{k1} \\ f_{k2} \\ b_{k3} \\ 0 \end{bmatrix}, \quad 1 \leq k \leq 3 \tag{5.6a}$$

$$|u_4\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}. \tag{5.6b}$$

The tensor product space $\mathcal{H}_J = \mathcal{H}_A \otimes \mathcal{H}_B$ has dimension $d \times (M - d + 1) = 2 \times 2 = 4$. Essentially the idea is to embed each extra coefficient $\{b_{k,d+1},\ldots,b_{kM}\}$ into its own d-dimensional subspace. Coefficients corresponding to the same value of k are "placed" in the same subspace. Because of Naimark's identity the outcome probabilities satisfy $\{p_J(k) = p(k)\}$. The post-measurement states, however, do not satisfy $\{|\psi_k\rangle = |\beta_k\rangle\}$.

Example 5.2. *This technique was described in [49]. Again starting from the vectors $|f_k\rangle = [f_{k1}, f_{k2}]^T$ for $1 \leq k \leq 3$, we let $\mathcal{H}_B \simeq \mathbb{R}^3$ and set the initial state of the ancilla to $|\phi\rangle = [1, 0, 0]^T$. If $|\psi\rangle = [\psi_1, \psi_2]^T$ is the initial state of the original system then the*

initial state of the joint system is

$$|\psi\rangle \otimes |\phi\rangle = \begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \psi_1 \\ \psi_2 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{5.7}$$

Let the desired post-measurement states be $\{|\psi_k\rangle = A_k |\psi\rangle = [\gamma_{k1}, \gamma_{k2}]^T\}$ *up to normalization. We next perform a unitary operation* U *on the joint system in such a way that the state of the joint system is transformed to*

$$U\left(|\psi\rangle \otimes |\phi\rangle\right) = \begin{bmatrix} \gamma_{11} \\ \gamma_{12} \\ \gamma_{21} \\ \gamma_{22} \\ \gamma_{31} \\ \gamma_{32} \end{bmatrix}. \tag{5.8}$$

Such a transformation always exists due to the properties of the $\{A_k\}$*. We now define the* $\{|u_k\rangle\}$ *by summing pairs of orthonormal basis vectors for* $\mathcal{H}_J$ *rather than*

as orthonormal basis vectors directly,

$$|u_1\rangle = \begin{bmatrix}1\\0\end{bmatrix} \otimes \begin{bmatrix}1\\0\\0\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix} \otimes \begin{bmatrix}1\\0\\0\end{bmatrix} = \begin{bmatrix}1\\1\\0\\0\\0\\0\end{bmatrix} \tag{5.9a}$$

$$|u_2\rangle = \begin{bmatrix}1\\0\end{bmatrix} \otimes \begin{bmatrix}0\\1\\0\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix} \otimes \begin{bmatrix}0\\1\\0\end{bmatrix} = \begin{bmatrix}0\\0\\1\\1\\0\\0\end{bmatrix} \tag{5.9b}$$

$$|u_3\rangle = \begin{bmatrix}1\\0\end{bmatrix} \otimes \begin{bmatrix}0\\0\\1\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix} \otimes \begin{bmatrix}0\\0\\1\end{bmatrix} = \begin{bmatrix}0\\0\\0\\0\\1\\1\end{bmatrix} \tag{5.9c}$$

It can be verified that this scheme achieves $\{p_J(k) = |\langle u_k|U(|\psi\rangle \otimes |\phi\rangle)\rangle|^2 = \gamma_{k1}^2 + \gamma_{k2}^2 = |\langle f_k|\psi\rangle|^2\}$ *as desired. This follows from the fact that for* $1 \leq k \leq M$ *we have* $A_k = |f_k\rangle\langle f_k| / ||f_k||$*, so* $|\psi_k\rangle = A_k|\psi\rangle = \langle f_k|\psi\rangle |f_k\rangle / ||f_k||$ *and thus* $\gamma_{k1}^2 + \gamma_{k2}^2 = |\langle f_k|\psi\rangle|^2$*. The desired post-measurement states are also achieved. To see why this is true, assume that the measurement outcome corresponding to* $k = 2$ *is observed. Then the post-measurement state of the joint system is*

$$|\alpha_2\rangle = \begin{bmatrix}0\\0\\\gamma_{21}\\\gamma_{22}\\0\\0\end{bmatrix}. \tag{5.10}$$

The post-measurement state of the original system can be found by taking the partial trace over $\mathcal{H}_B$, which corresponds to taking all of the $d \times 1$ blocks from $|\alpha_2\rangle$ and summing them together, i.e., summing together the 1st, 3rd, and 5th elements of $|\alpha_2\rangle$ and summing together the 2nd, 4th, and 6th elements. This leads to $|\beta_2\rangle = [0 + \gamma_{21} + 0, 0 + \gamma_{22} + 0]^T = [\gamma_{21}, \gamma_{22}]^T$, as desired. And of course the same holds true for the other possible measurement outcomes.

Our goal has been to present two methods of extending the original Hilbert space into a larger, tensor product space in such a way that an orthonormal basis of the larger space reproduces the correct measurement outcome probabilities. There are many others and plenty of interesting research in this area. In the context of this thesis it is interesting to note that the partial trace can also be viewed according to its action in operator space. Specifically, it is known that the partial trace operator, over $\mathcal{H}_B$ for instance, is the adjoint of the linear mapping that takes any density operator ρ acting on $\mathcal{H}_A$ to the joint density operator $\rho \otimes I_B$ acting on $\mathcal{H}_J = \mathcal{H}_A \otimes \mathcal{H}_B$, where I_B is the identity operator on $\mathcal{H}_B$. This suggests that Naimark's Theorem could also be phrased using the linear algebra of operator spaces.

5.2 Informationally Complete and Overcomplete POVMs

The definition of an informationally complete or IC POVM as a POVM that maps each possible density operator to a unique sequence of probabilities does not employ any notation or terminology associated with frame representations. This is why we chose to introduce IC POVMs in Section 4.2 following the statement of the quantum measurement postulate. However, a particularly useful way of thinking about and analyzing IC POVMs relies on the following fundamental result: Given an arbitrary set of operators $\{U_k, 1 \leq k \leq M\}$ in $\mathcal{V}$, $\{U_k\}$ is an IC POVM if and only if $\{U_k\}$ is both a POVM and a frame for $\mathcal{V}$ [8, 23, 67]. It will be convenient moving forward to summarize this statement in two parts using the labels (i), (ii), and (iii) to refer to

the relevant properties of the $\{U_k\}$,

$$\begin{array}{ll} \text{(i) } \{U_k\} \text{ is a POVM} & \text{(i) } \{U_k\} \text{ is a POVM} \\ \text{(ii) } \{U_k\} \text{ maps every density operator } \rho \in \mathcal{V} \text{ to a} & \implies \text{(iii) } \{U_k\} \text{ span } \mathcal{V} \\ \quad \text{unique sequence of coefficients } \{\langle\langle U_k|\rho\rangle\rangle\} & \end{array} \quad (5.11a)$$

$$\begin{array}{ll} \text{(i) } \{U_k\} \text{ is a POVM} & \text{(i) } \{U_k\} \text{ is a POVM} \\ \text{(ii) } \{U_k\} \text{ maps every density operator } \rho \in \mathcal{V} \text{ to a} & \impliedby \text{(iii) } \{U_k\} \text{ span } \mathcal{V}. \\ \quad \text{unique sequence of coefficients } \{\langle\langle U_k|\rho\rangle\rangle\} & \end{array} \quad (5.11b)$$

Note that since $\mathcal{V}$ is finite-dimensional, the $\{U_k\}$ span $\mathcal{V}$ if and only if they form a frame for $\mathcal{V}$. The terms "minimal IC POVM" and "informationally overcomplete (IOC) POVM" are sometimes used to differentiate between those IC POVMs whose elements are linearly indpendent and thus form a basis for $\mathcal{V}$ and those whose elements are linearly dependent, respectively [2, 18, 29, 67, 81, 82]. Equations (5.11) can be generalized to include the case where $\mathcal{V}$ is infinite-dimensional and to include generalized operator-valued frames [67], but for simplicity we do not consider those scenarios in this thesis.

To derive Equation (5.11b), assume that a set of operators $\{U_k\}$ in $\mathcal{V}$ satisfies the definition of a POVM and spans $\mathcal{V}$. Then as stated above $\{U_k\}$ must be a frame for $\mathcal{V}$. Its analysis map $\mathbf{A}_0$ can always be written as $\mathbf{A}_0 = \sum_{k=1}^{M} |W_k\rangle\rangle\langle\langle U_k|$. To show that $\{U_k\}$ is IC, it is sufficient to show that if two density operators have the same probability sequences with respect to this POVM, then they must be identical. This follows from the fact that since the $\{U_k\}$ span $\mathcal{V}$, no $V \in \mathcal{V}$ is orthogonal to all of them. Therefore, if $\mathbf{A}_0 |V\rangle\rangle = 0$ for some $V \in \mathcal{V}$ then we must have $V = 0$. Consider

the action of $\mathbf{A}_0$ on two arbitrary density operators ρ_1, $\rho_2 \in \mathcal{V}$. We have

$$\mathbf{A}_0 \left|\rho_1\right\rangle\rangle = \sum_{k=1}^{M} \left|W_k\right\rangle\rangle \langle\langle U_k|\rho_1\rangle\rangle = \sum_{k=1}^{M} p_1(k) \left|W_k\right\rangle\rangle, \tag{5.12a}$$

$$\mathbf{A}_0 \left|\rho_2\right\rangle\rangle = \sum_{k=1}^{M} \left|W_k\right\rangle\rangle \langle\langle U_k|\rho_2\rangle\rangle = \sum_{k\in\mathcal{K}} p_2(k) \left|W_k\right\rangle\rangle, \tag{5.12b}$$

where $p_i(k) = \langle\langle U_k|\rho_i\rangle\rangle$ for $i = 1, 2$ is defined as in Equation (5.1). If $p_1(k) = p_2(k)$ for $1 \leq k \leq M$, then $\mathbf{A}_0 \left|\rho_1 - \rho_2\right\rangle\rangle = 0$ implying that $\left|\rho_1 - \rho_2\right\rangle\rangle = 0$, i.e., $\rho_1 = \rho_2$.

Equation (5.11a) is more subtle as demonstrated through its comparison with the following related statement. Given an arbitrary set of operators $\{U_k\}$ in $\mathcal{V}$, it is straightforward to show by contradiction that if the $\{U_k\}$ map every $V \in \mathcal{V}$ to a unique sequence of coefficients $\{\langle\langle U_k|V\rangle\rangle\}$, then the $\{U_k\}$ must span $\mathcal{V}$. If instead the $\{U_k\}$ are only required to map every *density operator* in $\mathcal{V}$ to a unique sequence of coefficients, i.e., only condition (ii) is satisfied in Equation (5.11a), then they must span $\mathcal{U}$ but they do not necessarily span all of $\mathcal{V}$. This follows from the fact that an all density operators have constant trace and thus a constant orthogonal projection onto $\mathcal{U}^\perp$, so they are only distinguished by their orthogonal projections onto $\mathcal{U}$. Assume now that the $\{U_k\}$ satisfy both conditions (i) and (ii) on the left-hand side of Equation (5.11a). Then the $\{U_k\}$ must span $\mathcal{U}$ and they must also satisfy $\sum_k U_k = I$. Since I spans $\mathcal{U}^\perp$ by definition, this implies that the $\{U_k\}$ also span $\mathcal{U}^\perp$ and therefore they span all of $\mathcal{V}$. This line of reasoning also leads to the conclusion that if the $\{U_k\}$ form a POVM, then for the $\{U_k\}$ to span $\mathcal{V}$ it is sufficient for their orthogonal projections onto $\mathcal{U}$ to span $\mathcal{U}$. This statement is applied to a class a POVMs that we refer to as Etro POVMs in Section 5.3.

IC POVMs are commonly studied in the context of quantum state estimation [1, 18, 62, 67, 61, 81, 82], in which the objective is to reconstruct an unknown density operator from its probability values stemming from a given POVM. Obviously, the ability to recover an arbitrary density operator using only the probability values requires the POVM to be IC. But even if an IC POVM is employed, exact recovery of the probability values can only be achieved if we are able to measure an infinitely

large ensemble of systems, all prepared in the unknown state we wish to estimate. This is in general not possible in practice, and one motivation for using IOC POVMs is to mitigate the error caused by finite sample size estimations of the probabilities. This topic is also a motivation for the simulations presented in Section 5.4.

Another important issue in the use of IC POVMs to estimate unknown quantum states is that the reconstruction procedure implicitly requires computation of the dual frame of the POVM elements. This is in general a difficult task because it requires the inversion of a linear operator on $\mathcal{V}$, which is itself a "superoperator" [67]. Thus, IC POVMs whose duals are more easily computed are of great interest to the quantum physics community. Tight IC POVMs were introduced by Scott in [67] and are some of the most extensively studied.

5.2.1 Tight Informationally Complete POVMs

A tight IC POVM could be naturally defined as an IC POVM whose elements form a tight frame for $\mathcal{V}$. However, the definition is in fact slightly more nuanced as it takes into account the fact that all density operators lie within a hyperplane of $\mathcal{V}$. Briefly, the underlying logic is that when the hyperplane containing all density operators is shifted to the origin, it is identical to the subspace $\mathcal{U}$ of $\mathcal{V}$. The elements of a POVM may always be scaled and shifted to lie in $\mathcal{U}$, and when the scaled and shifted versions of the POVM elements form what is referred to as a tight frame for $\mathcal{U}$ with respect to the trace measure, the POVM is referred to as a tight IC POVM.

Given a density operator ρ and a POVM $\{E_k\}$, the operators $(\rho - I/d)$ and $\{S_k = E_k/\operatorname{Tr}(E_k) - I/d\}$ are all elements of $\mathcal{U}$. In [67] a tight IC POVM was defined as a POVM for which the $\{S_k\}$ satisfy

$$\sum_{k=1}^{M} \operatorname{Tr}(E_k)\, |\langle\langle S_k|V\rangle\rangle|^2 = C\, ||V||^2 \text{ for all } V \in \mathcal{U}, \tag{5.13}$$

for some constant $C > 0$. Comparing Equation (5.13) to the definition of an operator-valued frame in Equation (3.50) for the case where $C = D$, it is clear that the only difference (aside from the substitution of $\mathcal{U}$ for $\mathcal{V}$) is the extra factor of $\operatorname{Tr}(E_k)$ in

each term of the sum. This factor is the reason that in the terminology of [67], any set of operators $\{S_k\}$ satisfying Equation (5.13) are said to form a tight frame for $\mathcal{U}$ with respect to the trace measure. All POVM elements must have non-negative trace, and thus Equation (5.13) may be re-written using the operators $\{Q_k = \sqrt{\mathrm{Tr}(E_k)}\, S_k\}$, resulting in the equivalent form

$$\sum_{k=1}^{M} |\langle\langle Q_k|V\rangle\rangle|^2 = C\,||V||^2 \text{ for all } V \in \mathcal{U}. \tag{5.14}$$

Therefore, in our terminology a tight IC POVM is a POVM for which the operators $\{Q_k = E_k/\sqrt{\mathrm{Tr}(E_k)} - \sqrt{\mathrm{Tr}(E_k)}\, I/d\}$ form a tight frame for $\mathcal{U}$.

5.3 The Bloch Sphere and The Etro Spheres

We now apply the discussion given in Section 3.4.3 to the case where $\mathcal{H}$ represents the state of a qubit. We have $d = 2$ implying that $\mathcal{V}$ has dimension $d^2 = 4$. As stated in Section 3.4.3, the operators $\{\sigma_1/\sqrt{2}, \sigma_2/\sqrt{2}, \sigma_3/\sqrt{2}\}$ form an orthonormal basis for $\mathcal{U}$. Therefore, Equations (5.2) can be rewritten as

$$\rho = \frac{1}{\sqrt{2}}\frac{|I\rangle\rangle}{\sqrt{2}} + r_1\frac{|\sigma_1\rangle\rangle}{\sqrt{2}} + r_2\frac{|\sigma_2\rangle\rangle}{\sqrt{2}} + r_3\frac{|\sigma_3\rangle\rangle}{\sqrt{2}}, \tag{5.15a}$$

$$E_k = \frac{\mathrm{Tr}(E_k)}{\sqrt{2}}\frac{|I\rangle\rangle}{\sqrt{2}} + c_{k1}\frac{|\sigma_1\rangle\rangle}{\sqrt{2}} + c_{k2}\frac{|\sigma_2\rangle\rangle}{\sqrt{2}} + c_{k3}\frac{|\sigma_3\rangle\rangle}{\sqrt{2}}, \quad 1 \le k \le M, \tag{5.15b}$$

where $r_i = \langle\langle\sigma_i|\rho\rangle\rangle/\sqrt{2}$ for $1 \le i \le 3$ and $c_{ki} = \langle\langle\sigma_i|E_k\rangle\rangle/\sqrt{2}$ for $1 \le k \le M$ and $1 \le i \le 3$. Since ρ is positive semidefinite, it has an associated closed ball in $\mathbb{R}^3$ with radius $1/\sqrt{2}$. The column vector $\mathbf{r} = [r_1, r_2, r_3]^T$ always lies within the ball or on the sphere corresponding to the surface of the ball. It lies on the sphere when ρ has rank 1 and thus represents a pure state. The ball and sphere correspond within a constant factor to the very commonly used Bloch ball and Bloch sphere, which are typically assumed to have unit radius [49]. The column vector $\mathbf{r}$ is proportional to the Bloch vector of ρ. All of the $\{E_k\}$ are also positive semidefinite and therefore they also each have an associated closed ball in $\mathbb{R}^3$ whose surface is of course a sphere

in $\mathbb{R}^3$. The ball associated with E_k has radius $\mathrm{Tr}(E_k)/\sqrt{2}$ and the column vector $\mathbf{c}_k = [c_{k1}, c_{k2}, c_{k3}]^T$ always lies within that ball or on the sphere corresponding to its surface. It lies on the sphere when E_k has rank one. We will refer to $\mathbf{c}_k$ as an Etro vector. Note that Equation (5.3) implies that the $\{\mathbf{c_k}\}$ must always sum to zero. The probabilities in Equation (5.1) can also be written in terms of $\mathbf{r}$ and the $\{\mathbf{c_k}\}$. Substituting Equations (5.15) into Equation (5.1) yields

$$p(k) = \frac{\mathrm{Tr}(E_k)}{2} + \mathbf{c_k} \cdot \mathbf{r} = \frac{1}{M} + \mathbf{c_k} \cdot \mathbf{r}, \tag{5.16}$$

where $\cdot$ denotes the standard dot product in $\mathbb{R}^3$.

Of particular interest in this thesis are qubit POVMs that we will refer to as equal trace rank one or Etro POVMs. Unsurprisingly, an Etro POVM is one for which $\mathrm{Tr}(E_k) = 2/M$ for $1 \leq k \leq M$ and for which each of the $\{E_k\}$ has rank one, implying that $\sqrt{c_{k1}^2 + c_{k2}^2 + c_{k3}^2} = \mathrm{Tr}(E_k)/\sqrt{2} = \sqrt{2}/M$ for $1 \leq k \leq M$. When this is the case, all of the $\{E_k\}$ have the same associated ball in $\mathbb{R}^3$ with radius $\sqrt{2}/M$. The $\{\mathbf{c}_k\}$ all lie on a sphere of radius $\sqrt{2}/M$ that we refer to as an Etro sphere. Explicitly, an Etro sphere is one of a class of spheres in $\mathbb{R}^3$, each with radius $\sqrt{2}/M$ for some integer M. An M-element Etro POVM can be fully specified by M vectors $\{\mathbf{c}_k\}$ lying on the Etro sphere of radius $\sqrt{2}/M$. It can equivalently be specified by M points on the Etro sphere of radius $\sqrt{2}/M$ with each point corresponding to the endpoint of one of the $\{\mathbf{c}_k\}$. It follows from Section 5.2 that a qubit Etro POVM is IC if and only if the $\{\mathbf{c_k}\}$, which represent the orthogonal projections of the $\{E_k\}$ onto $\mathcal{U}$, span $\mathbb{R}^3$ [29]. While the definition of an Etro POVM could clearly be applied to higher dimensions, in this thesis we use the term Etro POVM to refer specifically to those corresponding to qubit measurements.

An Etro POVM satisfies the definition of a a tight IC POVM when its Etro vectors $\{\mathbf{c_k}\}$ form a tight frame for $\mathbb{R}^3$. Specifically, substituting the definition of the $\{Q_k\}$ from Section 5.2.1 into Equation (5.15b) and utilizing the fact that $\mathrm{Tr}(E_k) = 2/M$

for $1 \leq k \leq M$ yields

$$Q_k = \frac{1}{\sqrt{\mathrm{Tr}(E_k)}} \left(c_{k1} \frac{|\sigma_1\rangle\rangle}{\sqrt{2}} + c_{k2} \frac{|\sigma_2\rangle\rangle}{\sqrt{2}} + c_{k3} \frac{|\sigma_3\rangle\rangle}{\sqrt{2}} \right) \tag{5.17a}$$

$$= \sqrt{\frac{M}{2}} \left(c_{k1} \frac{|\sigma_1\rangle\rangle}{\sqrt{2}} + c_{k2} \frac{|\sigma_2\rangle\rangle}{\sqrt{2}} + c_{k3} \frac{|\sigma_3\rangle\rangle}{\sqrt{2}} \right), \quad 1 \leq k \leq M. \tag{5.17b}$$

According to Equation (5.17), the $\{Q_k\}$ form a tight frame for $\mathcal{U}$ when the vectors $\{\sqrt{M/2}\mathbf{c_k}\}$, or equivalently the vectors $\{\mathbf{c_k}\}$, form a tight frame for $\mathbb{R}^3$.

Example 5.3. POVMs constructed using Platonic solids are used often in the literature [81, 82, 21, 70]. In the terminology of this thesis, a POVM constructed using the Platonic solid with M vertices for $M \in \{4, 6, 8, 12, 20\}$ is an M-element Etro POVM whose Etro vectors $\{\mathbf{c}_k\}$ correspond to the vertices of that Platonic solid inscribed in the corresponding Etro sphere. When the Platonic solid is an octahedron ($M = 6$), the POVM is typically described as having been constructed from three mutually unbiased bases (MUBs) for the state space of the qubit. All POVMs constructed from the Platonic solids are tight IC POVMs.

Example 5.4. Consider a qubit POVM $\{E_1, E_2\}$ whose elements form a complete set of orthogonal projectors onto $\mathcal{H}$. As stated in Section 4.2, this type of POVM always corresponds to a standard quantum measurement. It is straightforward to verify that $\{E_1, E_2\}$ is an Etro POVM whose corresponding Etro sphere has radius $1/\sqrt{2}$ and is thus identical to the Bloch sphere. The Etro vectors $\{\mathbf{c}_1, \mathbf{c}_2\}$ must satisfy $\mathbf{c}_1 + \mathbf{c}_2 = 0$, implying that they point in opposite directions on the Etro sphere. Helstrom's optimal POVM for distinguishing between two fixed qubit density operators is one example of this type of POVM.

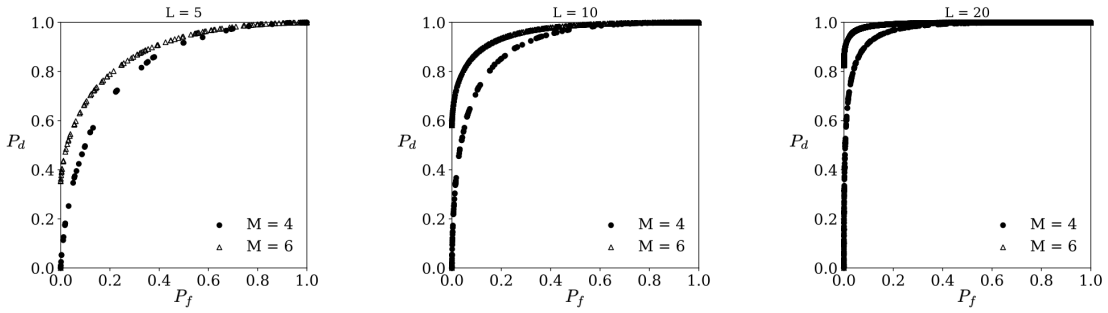
5.4 Qubit State Discrimination using Platonic Solids

Starting in Section 5.4 and continuing on through Chapter 6, we return to the problem stated in Chapter 4 in which an M-element POVM is used to distinguish between the possibilities that L QMSs all have density operator ρ_0 or all have density operator ρ_1.

In Chapter 4 we assumed that $L = 1$, but we now begin to explore what happens as L is increased. As stated in Section 4.3, we will denote the score variable by $\mathbf{S}$ as opposed to S to emphasize that it is a vector-valued random variable as opposed to a scalar random variable. A particular realization $\mathbf{s}$ will be denoted by the column vector $\mathbf{s} = [n_1/L, \ldots, n_M/L]^T$ where n_k is the number of occurrences of the kth measurement outcome and $\sum_k n_k = L$. The conditional distributions $f_i(\mathbf{S})$ for $i \in \{0, 1\}$ are multinomial distributions. We also assumed in Chapter 4 that the decision region of the binary decision rule could be any subset of the possible relative frequency vectors. Throughout the rest of the thesis we assume that only decision regions corresponding to an LRT with threshold $\eta = q_0/q_1$ for some prior probabilities are used.

In Example 5.5 below, we utilize POVMs constructed using Platonic solids, which were defined in Example 5.3. These very preliminary simulations set the stage for Chapter 6 where we consider POVMs constructed from other arrangements of points on an Etro sphere. POVMs constructed using Platonic solids are of significant interest in the context of qubit state estimation [1, 18, 62, 67, 61, 81, 82, 83] but have been utilized less often for binary hypothesis testing. We present evidence through simulation that there is a tradeoff in discrimination performance between the number L of identically-prepared QMSs and the number M of POVM elements. Note that as stated in Example 5.3, all POVMs constructed using Platonic solids are Etro POVMs.

Example 5.5. In this example we arbitrarily set the Bloch vectors of ρ_0 and ρ_1 to $\mathbf{r}_0 = (1/\sqrt{2})[0, 0, 1]^T$ and $\mathbf{r}_1 = (1/\sqrt{2})[\cos\phi \sin\theta, \sin\phi \sin\theta, \cos\theta]^T$, where $\theta = 2\pi/3$ and $\phi = \pi/3$. Note that as mentioned in Section 5.3, there is an extra factor of $(1/\sqrt{2})$ in comparison to Bloch vectors as they are typically defined in the literature. The LRT QDOCs corresponding to POVMs constructed using a tetrahedron ($M = 4$) and an octahedron ($M = 6$) inscribed in the Etro sphere and to $L = 5$, 10, 20 are shown in Figure 5-1. The plots reflect a tradeoff in discrimination performance between M and L. For a fixed value of L, increasing the value of M leads to better detection as reflected by the superior QDOC. On the other hand, for a fixed value of M increasing the value of L also leads to better detection.



(a) (b) (c)

Figure 5-1: LRT QDOCs for $L \in \{5, 10, 20\}$ and IC POVMs constructed from Platonic solids with $M = 4$ (tetrahedron) and $M = 6$ (octahedron) vertices. The density operators ρ_0 and ρ_1 that were used for this example are specified in Example 5.5.

5.5 Robustness of Informationally Overcomplete POVMs

As stated in Section 5.2, IOC POVMs have been shown to be of great utility in the context of quantum state estimation. The problem addressed in Section 5.5 was designed specifically to demonstrate how the discussion in Section 3.5 can be applied to quantum mechanics. Readers interested only in quantum binary state discrimination and not quantum state estimation may wish to proceed to Chapter 6.

An important difference between Sections 3.5 and 5.5 is that in Section 3.5 we assumed that the observed frame coefficients were affected by error values that were pairwise uncorrelated. By contrast, in Section 5.5 we assume that the error values stem from estimating probability values using relative frequencies. This implies that they are correlated with each other since the set of all relative frequencies must of course add to 1. A whitening transformation can be used to compensate for the correlations. But it is important to note that, as explained in more detail below, the identification of such a transformation requires knowledge of the true density matrix. The optimal solution is thus impractical in the context of quantum state estimation, where the true density matrix is of course assumed to be unknown. The optimal solution is perhaps more relevant to the problem of linearly reconstructing a known density operator using quantized versions of its frame coefficients, where the

frame coefficients are the probability values derived from the density operator and a given IC POVM and their quantized counterparts are a corresponding set of relative frequencies. This is reminiscent of the linear reconstruction of a known bandlimited signal using quantized versions of its sample values [51].

5.5.1 Quantum State Estimation

We emphasize that the following scenario was designed to be parallel to the one described in Section 3.5. Let ρ be an unknown density operator and let $\{E_k\}$ be an arbitrary tight IC POVM. In the following variation of the problem stated in Section 3.5, the unknown vector $|v\rangle$ lying in $\mathcal{V}$ is replaced by the shifted operator $|\rho - I/d\rangle\rangle$ lying in the subspace $\mathcal{U}$ of $\mathcal{V}$ and the analysis frame $\{|f_k\rangle\}$ is replaced by the operators $\{Q_k = E_k/\sqrt{\mathrm{Tr}(E_k)} - \sqrt{\mathrm{Tr}(E_k)}I/d\}$. By definition of a tight IC POVM, the $\{Q_k\}$ form a tight frame for $\mathcal{U}$. We will denote the frame bound of the $\{Q_k\}$ by C and will additionally assume that they all have norm B. Thus, in analogy with Equation (3.61) the $\{Q_k\}$ are an ENTF for $\mathcal{U}$ satisfying

$$\sum_{k=1}^{M} |\langle\langle Q_k|V\rangle\rangle|^2 = C\,||V||^2 \text{ for all } V \in \mathcal{U}, \tag{5.18a}$$

$$||Q_k|| = B, \quad 1 \leq k \leq M. \tag{5.18b}$$

Note that in addition to the relation $CN = MB^2$ which is satisfied by any ENTF including those outside of the context of quantum mechanics, the constants B and C in Equation (5.18) must also satisfy various other constraints deriving from the required properties of the POVM elements $\{E_k\}$. Specifically, the $\{E_k\}$ must by definition be positive semidefinite and add to the identity. In the absence of any source of error, the operator $|\rho - I/d\rangle\rangle$ can always be reconstructed from its frame coefficients $\{a_k = \langle\langle Q_k|\rho - I/d\rangle\rangle\}$. Our objective is to estimate $|\rho - I/d\rangle\rangle$ given only a set of observed coefficients $\{\hat{a}_k\}$, not necessarily equal to the $\{a_k\}$, in such a way that

$$\mathcal{E} = \mathbb{E}\left[||\rho_e||^2\right] = \mathbb{E}\left[||\hat{\rho} - \rho||^2\right] = \mathbb{E}\left[\mathrm{Tr}\left((\hat{\rho} - \rho)^2\right)\right] \tag{5.19}$$

is minimized. In Equation (5.19), the expectation is taken over all possible values of the $\{\hat{a}_k\}$. As in Section 3.5, we assume that the estimated operator $|\hat{\rho} - I/d\rangle\rangle$ has the form

$$\hat{\rho} - \frac{I}{d} = \sum_{k=1}^{M} \hat{a}_k \tilde{Q}_k \tag{5.20}$$

where $\{\tilde{Q}_k\}$ is any dual frame of $\{Q_k\}$.

Clearly, the optimal dual frame depends on the distribution of the $\{\hat{a}_k\}$. In Section 3.5 it was assumed that $\{\hat{a}_k = a_k + e_k\}$ where the $\{e_k\}$ had zero mean, variance σ^2, and were pairwise uncorrelated. And under these assumptions, the optimal dual frame was the canonical dual frame of $\{Q_k\}$. In the current discussion, however, we assume that the $\{\hat{a}_k\}$ are derived from a relative frequency vector generated using the procedure described in Section 5.4. Specifically, we assume that there are L identically-prepared QMSs whose states are all described by the density operator ρ. Each QMS is measured using a quantum measurement with corresponding POVM $\{E_k, 1 \leq k \leq M\}$, and therefore the probabilities of the M measurement outcomes are $\{p(k) = \langle\langle E_k|\rho\rangle\rangle\}$. The L measurement outcomes result in a set of relative frequencies $\{\hat{p}(k) = n_k/L\}$ where n_k is the number of occurrences of the kth measurement outcome. Note that the process of obtaining the $\{\hat{p}(k)\}$ from the $\{p(k)\}$ can be viewed as probabilistic uniform quantization. The $\{p(k)\}$ and $\{\hat{p}(k)\}$ can be related to the $\{a_k\}$ and the $\{\hat{a}_k\}$, respectively, as follows. Direct substitution using the definition of the $\{Q_k\}$ leads to

$$a_k = \langle\langle Q_k|\rho - I/d\rangle\rangle = \frac{p(k)}{\sqrt{\mathrm{Tr}(E_k)}} - \frac{\sqrt{\mathrm{Tr}(E_k)}}{d}, \quad 1 \leq k \leq M. \tag{5.21}$$

We then define

$$\hat{a}_k = \frac{\hat{p}(k)}{\sqrt{\mathrm{Tr}(E_k)}} - \frac{\sqrt{\mathrm{Tr}(E_k)}}{d}, \quad 1 \leq k \leq M. \tag{5.22}$$

We have $\{\hat{a}_k = a_k + e_k\}$ where $e_k = (\hat{p}_k - p(k))/\sqrt{\mathrm{Tr}(E_k)}$ for $1 \leq k \leq M$. As stated in Section 5.4, the $\{e_k\}$ are distributed according to a multinomial distribution with

parameters L and $\{p(k)\}$. It can be shown using straightforward algebra that

$$\mathbb{E}[e_k] = 0, \quad 1 \leq k \leq M \tag{5.23a}$$

$$\mathbb{E}[e_j e_k] = \begin{cases} \dfrac{p(k)\,(1-p(k))}{L\ \mathrm{Tr}(E_k)} \text{ if } j = k \\ \dfrac{-p(j)\,p(k)}{L\,\sqrt{\mathrm{Tr}(E_j)\,\mathrm{Tr}(E_k)}} \text{ if } j \neq k \end{cases}, \quad 1 \leq j,\, k \leq M. \tag{5.23b}$$

The optimal synthesis frame $\{\tilde{Q}_k\}$ can be found by whitening the $\{e_k\}$ and then computing the canonical dual of the effective analysis frame. It is important to note, however, that whitening the $\{e_k\}$ requires knowledge of the $\{p(k)\}$ and thus of ρ itself.

Chapter 6

Qubit State Discrimination on the Etro Spheres

As mentioned briefly in Chapter 5, POVMs constructed from Platonic solids are of interest in the quantum state estimation literature because they are tight IC POVMs, implying that they are self-dual up to a constant. This makes reconstruction of an unknown state from estimates of its frame coefficients in the form of the relative frequencies particularly straightforward. In the context of quantum state discrimination, however, it is not necessary to reconstruct the state from the relative frequency vector. This suggests that it might be interesting to explore constructing POVMs from other arrangements of points on an Etro sphere.

Just as a pure state qubit density operator can be specified by a single point on the Bloch sphere, as shown in Chapter 5 each element of an Etro POVM associated with a qubit measurement can be specified by a single point on the Etro sphere. Considerable previous work has focused on POVMs constructed from one of the five Platonic solids [21, 46, 70, 81, 82]. In our terminology these are Etro POVMs constructed from sets of points corresponding to the vertices of one of the Platonic solids. A main motivation of Chapter 6 is to present an exploratory and preliminary investigation into the utility for quantum binary state discrimination of Etro POVMs constructed from other distributions of M points on an Etro sphere. Our findings are also summarized in [22].

As in Section 5.4 we assume that an M-element Etro POVM $\{E_k\}$ with corresponding Etro vectors $\{\mathbf{c_k}\}$ is used to discriminate between the possibilities that L identically-prepared qubits all have density operator ρ_0 or all have density operator ρ_1. We continue to denote the prior probabilities by $P(H_i) = q_i$ for $i \in \{0, 1\}$. The $\{\mathbf{c_k}\}$ lie on the Etro sphere with radius $\sqrt{2}/M$ and always satisfy $\sum_k \mathbf{c_k} = 0$. The Bloch vectors $\mathbf{r}_0$ and $\mathbf{r}_1$ of the two pure states are separated by a relative angle α and are known only up to an overall rotation on the Bloch sphere. We may equivalently assume that $\mathbf{r}_0$ and $\mathbf{r}_1$ are known exactly but that the overall alignment of the $\{\mathbf{c_k}\}$ relative to the Etro sphere is unknown, i.e., the relative rotational orientation of the Bloch sphere and the Etro sphere is unknown. The performance of each POVM is measured according to its minimum and maximum probabilities of error, denoted as $\min P_e$ and $\max P_e$, over all possible relative orientations as well as their difference. A smaller value of $(\max P_e - \min P_e)$ suggests that the corresponding POVM is less sensitive to changes in the relative orientations of the Bloch and Etro spheres. The exploratory simulations presented in Chapter 6 leave open the question of what the optimal POVM is with respect to its sensitivity to changes in the relative orientation of the Bloch and Etro spheres.

6.1 Optimal Distributions of M Points on a Sphere

An Etro POVM $\{E_k\}$ can always be fully specified by its M Etro vectors $\{\mathbf{c_k}\}$, or equivalently by the M endpoints of those vectors which all lie on the Etro sphere of radius $\sqrt{2}/M$. Intuition suggests that maximally spreading the endpoints on the sphere will tend to reduce the variation in performance over all possible orientations.

Various approaches to and criteria for evenly distributing M points on a sphere have been reported in the literature [35, 43, 65]. We first consider distributions of points that correspond to the vertices of a Platonic solid or an Archimedean solid, in addition to to distributions of points that minimize Riesz s-energy for a given value of M, subject to the constraint that the $\{\mathbf{c}_k\}$ must sum to zero. In three dimensions

the Riesz s-energy of a set of M vectors $\{\mathbf{c}_k\}$ of equal length is defined as

$$E(s) = \begin{cases} \displaystyle\sum_{1\leq j<k\leq M} \log \|\mathbf{c_j} - \mathbf{c_k}\|^{-1} & \text{if } s = 0 \qquad (6.1a) \\ \displaystyle\sum_{1\leq j<k\leq M} \|\mathbf{c_j} - \mathbf{c_k}\|^{-s} & \text{if } s \geq 0. \qquad (6.1b) \end{cases}$$

In Equation (6.1), $\|\mathbf{c}_j - \mathbf{c}_k\|$ often denotes the Euclidean distance between $\mathbf{c_j}$ and $\mathbf{c_k}$ but it could also be defined as the great circle distance between $\mathbf{c_j}$ and $\mathbf{c_k}$. Minimizing $E(0)$ is equivalent to maximizing the product of distances between points. Minimizing $E(1)$ is equivalent to minimizing the electric potential energy of a system of point charges located at the endpoints of the vectors. As $s \to \infty$, only the two closest points contribute to the sum and minimizing $E(s)$ corresponds to maximizing nearest neighbor distance.

In the simulations presented in Section 6.2, we also consider distributions of points that were computed numerically by Sloane et al. [69] to be optimal with respect to the maximum convex hull volume, maximum nearest neighbor distance, and minimum covering radius criteria. The latter criteria are defined as

$$\max_{\mathbf{c_1},\ldots,\mathbf{c_M}} \min_{1\leq j,\,k\leq M} ||\mathbf{c_j} - \mathbf{c_k}|| \qquad \text{(max. nearest neighbor distance)} \qquad (6.2a)$$

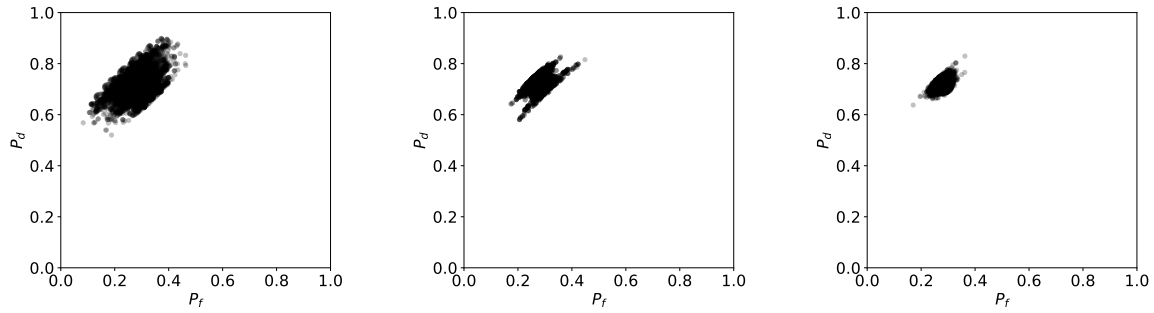
$$\min_{\mathbf{c_1},\ldots,\mathbf{c_M}} \max_{\mathbf{x}:||\mathbf{x}||=1} \min_{1\leq k\leq M} ||\mathbf{c_j} - \mathbf{c_k}|| \qquad \text{(min. covering radius)} \qquad (6.2b)$$

It is important to note, however, that these solutions were computed *without* the constraint that the$\{\mathbf{c}_k\}$ sum to zero. Many of the optimal solutions sum to a vector whose norm is very close to zero. Consequently for our exploratory purposes we chose to compensate by appending an extra vector $\epsilon = -\sum_k \mathbf{c_k}$ to the $\{\mathbf{c}_k\}$ with corresponding POVM element E_0. Intuitively this would not be expected to affect any broad trends observed in the results, since for all simulations presented we required $||\epsilon|| \leq 10^{-8}$.

6.2 Results and Simulations

A sampling of our results for various combinations of the parameter values M, $0 \leq \alpha \leq \pi$, and $0 \leq q_1 \leq 1$ is shown in Tables 6.1 to 6.3 for $L = 5$. Extended results for $L = 1$ and $L = 5$ are given in Appendix A. The rotational orientation of a specific set of M points on an Etro sphere, was varied by first choosing an arbitrary point as the north pole and then incrementing the azimuth and elevation angles of that point over their full ranges using a step size of $\pi/50$. The trends we describe next were not consistent over all sets of parameter values. For fixed α and q_1, we observed that larger values of M typically correspond to lower $\max P_e$ but higher $\min P_e$ over all possible orientations (see Table 6.1 and Figure 6-1). Apparently, the optimal sets of $M = 6$ points with respect to all of the criteria considered here look very much like the vertices of an octahedron. This is reflected in Table 6.1 by the identical performance of all criteria for $M = 6$. Note that the QMOCs in Figure 6-1 become more concentrated around a central region for larger values of M, indicating less sensitivity to changes in the rotational orientation of the Etro sphere. Furthermore for a fixed value of $M \in \{4, 6, 8, 12\}$, the Platonic solid with M vertices is not necessarily the best arrangement of points in terms of its sensitivity to rotation. This can be seen, for example, in Table 6.1 for the case where $M = 6$. For fixed q_1, the decrease in sensitivity with M is more pronounced for smaller values of α (see Table 6.2), which makes intuitive sense since smaller values of α correspond to Bloch vectors states that are more collinear and thus more sensitive to small changes in the relative orientation of the Bloch and Etro spheres. For fixed M, larger values of α and values of q_1 that are further from $1/2$ generally lead to lower $\min P_e$ and $\max P_e$ (see Table 6.3 and Figure 6-2). In Figure 6-2 each row corresponds to a specific value of M and each column corresponds to a specific value of α. The MOCs become more concentrated around a central region as M increases down each column. The increase in concentration is more pronounced for columns corresponding to smaller values of α.

Our exploratory investigation suggests that various arrangements of points that



(a) (b) (c)

Figure 6-1: QMOCs generated using $\alpha = \pi/4$, $q_1 = 1/2$, and Etro POVMs constructed from M points on an Etro sphere with minimum covering radius. Each operating point represents the P_f and P_d values obtained by a specific rotational orientation of the M points and an LRT with threshold $\eta = q_0/q_1$. (a) $M = 4$ (b) $M = 6$ (c) $M = 8$.

are well-spread with respect to the chosen metrics perform well with respect to their sensitivity to changes in the relative rotational orientation of the Bloch and Etro spheres. We focused on metrics that promote evenly spread distributions of points since we assumed that all relative orientations of the Bloch and Etro spheres were equally likely. If this were not the case, it would be intuitively expected that distributions of points with higher concentrations in certain regions of an Etro sphere would be more desirable. This might be the case if, for example, in a particular application the two hypotheses corresponded to the L qubits being prepared in a density operator ρ drawn from one of two distributions over all possible density operators, each localized around a particular region of the Bloch sphere.

Table 6.1: Minimum and maximum probabilities of error for different distributions of M points on a sphere. The values below were generated using $\alpha = \pi/2$ and $q_1 = 1/2$.

M	**Description**	**Min** P_e	**Max** P_e	**Difference**
4	Tetrahedron	0.076	0.214	0.138
	Max Convex Hull Vol	0.076	0.213	0.137
	Max N.N. Dist	0.076	0.214	0.138
	Min Covering Radius	0.077	0.211	0.135
	Min Riesz 0-energy	0.076	0.213	0.137
5	(No Platonic Solid)	N/A	N/A	N/A
	Max Convex Hull Vol	0.097	0.224	0.126
	Max N.N. Dist	0.101	0.221	0.120
	Min Covering Radius	0.092	0.223	0.131
	Min Riesz 0-energy	0.091	0.198	0.107
6	Octahedron	0.111	0.174	0.063
	Max Convex Hull Vol	0.111	0.174	0.063
	Max N.N. Dist	0.111	0.174	0.063
	Min Covering Radius	0.111	0.174	0.063
	Min Riesz 0-energy	0.111	0.174	0.063
7	(No Platonic Solid)	N/A	N/A	N/A
	Max Convex Hull Vol	0.121	0.199	0.079
	Max N.N. Dist	N/A	N/A	N/A
	Min Covering Radius	0.120	0.196	0.076
	Min Riesz 0-energy	0.112	0.199	0.087
8	Cube	0.122	0.170	0.048
	Max Convex Hull Vol	0.118	0.168	0.050
	Max N.N. Dist	0.113	0.173	0.060
	Min Covering Radius	0.116	0.160	0.044
	Min Riesz 0-energy	0.124	0.160	0.036

Table 6.2: Minimum and maximum probabilities of error for the M points on a sphere that minimize Riesz 0-energy with the Euclidean distance metric. The values below were generated using $q_1 = 1/2$.

α	M	**Min** P_e	**Max** P_e	**Difference**
$\pi/4$	4	0.223	0.343	0.119
	5	0.235	0.332	0.097
	6	0.255	0.316	0.061
	7	0.252	0.332	0.080
	8	0.261	0.305	0.043
	9	0.271	0.300	0.029
	10	0.268	0.296	0.028
$\pi/2$	4	0.076	0.213	0.137
	5	0.091	0.198	0.107
	6	0.111	0.174	0.063
	7	0.112	0.199	0.087
	8	0.124	0.160	0.036
	9	0.129	0.156	0.027
	10	0.132	0.151	0.019
$3\pi/4$	4	0.029	0.121	0.092
	5	0.044	0.116	0.072
	6	0.063	0.092	0.029
	7	0.059	0.118	0.060
	8	0.066	0.091	0.025
	9	0.068	0.085	0.016
	10	0.070	0.086	0.015

Table 6.3: Minimum and maximum probabilities of error for the $M = 6$ points on the sphere that minimize covering radius.

q_1	α	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.095	0.123	0.028
	$\pi/2$	0.046	0.086	0.040
	$3\pi/4$	0.023	0.052	0.029
1/4	$\pi/4$	0.177	0.224	0.047
	$\pi/2$	0.081	0.133	0.051
	$3\pi/4$	0.042	0.075	0.033
3/8	$\pi/4$	0.234	0.286	0.053
	$\pi/2$	0.103	0.155	0.052
	$3\pi/4$	0.055	0.086	0.031
1/2	$\pi/4$	0.255	0.317	0.061
	$\pi/2$	0.111	0.174	0.063
	$3\pi/4$	0.063	0.092	0.029

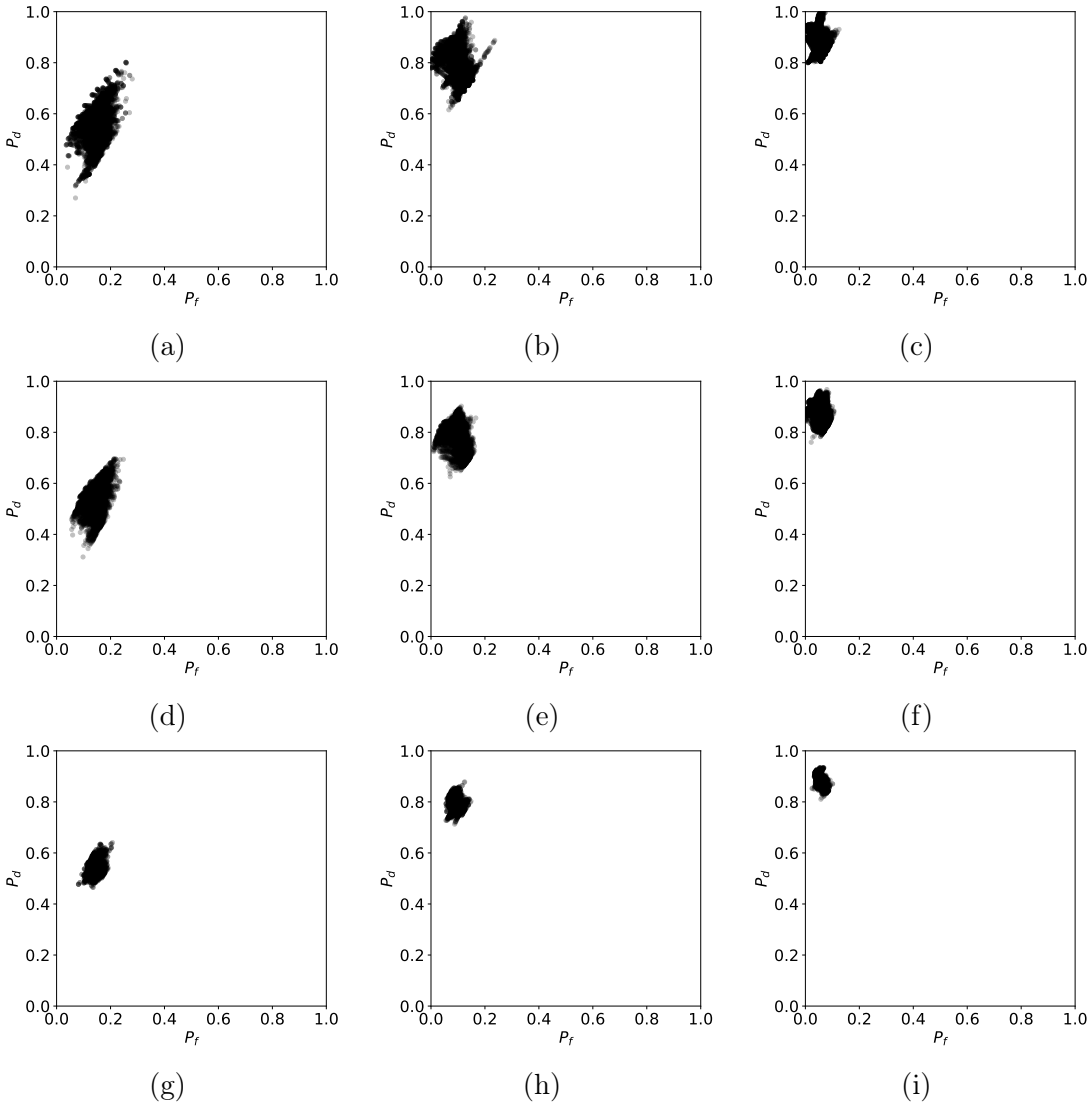


Figure 6-2: QMOCs generated using $q_1 = 3/8$, and Etro POVMs constructed from M points on an Etro sphere with maximum nearest neighbor distance. Each operating point represents the P_f and P_d values obtained by a specific rotational orientation of the M points and an LRT with threshold $\eta = q_0/q_1$. (a) $M = 4, \alpha = \pi/4$ (b) $M = 4, \alpha = \pi/2$ (c) $M = 4, \alpha = 3\pi/4$ (d) $M = 5, \alpha = \pi/4$ (e) $M = 5, \alpha = \pi/2$ (f) $M = 5, \alpha = 3\pi/4$ (g) $M = 8, \alpha = \pi/4$ (h) $M = 8, \alpha = \pi/2$ (i) $M = 8, \alpha = 3\pi/4$.

Chapter 7

Concluding Remarks and Future Work

In this thesis we explored a variety of issues surrounding binary hypothesis testing problems. The first major issue related to an optimality condition for ROCs generated using SVTs in addition to a constructive procedure for obtaining the optimal ROC when the condition is not met. The second set of issues related to interpreting the problem of quantum binary state discrimination using the language of linear algebra applied to operator spaces. Inspired in large part by the use of redundancy in classical frame theory for designing robust signal processing systems, we explored how overcompleteness could be exploited in the context of quantum binary state discrimination. We used decision and measurement operating characteristics as a way of evaluating and comparing different binary hypothesis testing systems. Along the way we defined a counterpart to the Bloch sphere for the class of Etro POVMs and demonstrated how it might be useful to design qubit Etro POVMs by drawing inspiration from other fields where distributing points on a sphere is an important problem. Sections 7.1 to 7.3 below summarize the main points and describe a variety of possibilities for future work.

7.1 Optimal ROCs from Sub-Optimal ROCs

A principal result of Chapter 2 stated that an ROC that was generated using SVTs applied to a scalar score variable is guaranteed to be the Neyman-Pearson optimal ROC for that score variable if it is concave. A procedure was given for constructing the optimal ROC from a non-concave SVT ROC. One of the most interesting aspects of these results is the fact that they make use of the information inherently contained in an ROC, *independent* of the score variable used to generate it. For the purposes of identifying and obtaining the optimal ROC for a given score variable, they allow an SVT ROC to be detached from any and all analytical models and real-world data. We chose to focus on ROCs generated using SVTs, but an opportunity for future work would be to ask similar questions of ROCs generated using other types of decision regions. We also chose to focus on identifying and obtaining the optimal ROC (or DOC in our terminology), but we did not ask the question of what the optimal MOC is nor did we investigate possible answers to that question that already exist in the literature. It would be interesting to find out if there is a parallel situation for MOCs, i.e., the optimal MOC is known and accepted but often not explicitly generated. Finally, all of our results pertain to scalar score variables. A possible area of future study might extend these ideas to vector-valued score variables.

7.2 Frame Theory and Quantization for Quantum Binary State Discrimination

In Chapters 4 through 6 the same basic structure of a pre-decision operator followed by a binary decision rule is applied to exploring operating characteristics for quantum binary hypothesis testing. However the nature of the pre-decision operator and corresponding score vector are fundamentally different than in the classical case. This is a direct consequence of the fact that the physics of quantum mechanics is fundamentally different than the physics of classical mechanics and that measurement outcomes on quantum systems are inherently probabilistic. While there are a number

of ways in which quantum systems and quantum states can be described mathematically, we have chosen the representation in terms of the density operator and the representation of the pre-decision operator in terms of a POVM which consists of an indexed set of M Hermitian, positive semidefinite operators that sum to the identity. In Chapter 4 we summarize the key postulates from quantum mechanics that govern the formulation and development of the quantum binary hypothesis testing problem that is considered in this thesis. One specific way of viewing the underlying problem as we formulate it is to imagine two possible physical environments corresponding to H_0 and H_1 that we would like to distinguish between. And that any quantum system prepared by or associated with each would have associated with it one of two known density operators. The hypothesis testing system to be designed and evaluated through its operating characteristics is based on knowledge of each of the two density operators. The pre-decision operator is then a specified or previously designed POVM. We assume that L independent quantum mechanical systems or QMSs prepared by only one of the environments are available for measurement and that we are able to determine the index of the outcome after each measurement. The pre-decision operator generates an M-element vector of relative frequencies corresponding to the number of occurrences of each of the M possible measurement outcomes. This vector is used as our score variable. The classic paper by Helstrom specifies the two-element POVM and the decision boundary for minimum probability of error or equivalently the decision boundary on the two-element probability vector based on the POVM. Helstrom also noted that a d-element POVM with each element corresponding to the outer product of the eigenvectors of the density operator associated with H_0 can equivalently be used. Here d is the dimension of the Hilbert space that the density operator acts on.

As in the earlier discussion of classical hypothesis testing, we separate the discussion of operating characteristics into QDOCs and QMOCs. The first assumes the pre-decision operator, i.e., the POVM, has been specified and the operating characteristics correspond to the LRT decision rule applied to the score variable vector resulting from the POVM. QMOCs correspond to keeping fixed the decision region

for the decision operator and varying the pre-decision operator or POVM. Our presentation of QMOCs is primarily intended to introduce the concept and to illustrate it with a simple example.

Chapters 4 through 6 were directed at design of the pre-decision operator, i.e. the POVM used to generate the score variable vector of relative frequencies where M can in general be larger than the dimensionality of the two density operators to be distinguished. This in effect corresponds to utilizing an overcomplete or redundant characterization and measurement process. It is well-understood that overcomplete representations of elements in a vector space through the use of a larger set than necessary of linearly dependent vectors often has the advantage of providing robustness to errors in the coefficients. A powerful and often used vector space methodology for overcomplete representation of vector spaces is that of frames and that is the methodology that we exploit in designing overcomplete POVMs. Consequently before utilizing frame representations of vector spaces, we summarize in Chapter 3 our perspective and the notation and key properties that we exploit in Chapters 5 and 6. This includes the basics of frame theory. In effect, frames correspond to sets of linearly dependent vectors that span the space, i.e. every vector in the space can be constructed as a linear combination of the frame vectors. A basis set of vectors is of course a valid frame but more generally, with linearly dependent frame vectors, the set of coefficients representing any vector is not unique, and the representation is overcomplete which offers redundancy and an opportunity for robustness. The specific viewpoint that we take is that the space $\mathcal{V}$ being represented is a subspace of some larger space $\mathcal{W}$. And that the overcomplete frame representation of a vector in $\mathcal{V}$ can also be associated with a unique vector in the larger space $\mathcal{W}$. We introduced the notion of analysis and synthesis maps which are closely connected to the more traditional analysis and synthesis operators. Section 3.4 then extends this discussion of frame representations to vector spaces in which the vectors are operators. The notation and perspective associated with operator spaces as developed in Chapter 3 forms the basis for the discussion of the quantum binary state discrimination problem stated in Chapter 4 and expanded upon in Chapters 5 and 6.

Viewing quantum binary state discrimination through the lens of linear algebra applied to operator spaces was a central component of this thesis. We found this perspective to be intriguing in part because it naturally directed us towards the identification of a variety of signal processing methodologies that could be applied to this problem. As just one example, the signal processing literature on issues stemming from quantization error is rich and highly sophisticated (see [32] and references therein). Techniques such as sigma-delta quantization and noise shaping more generally have been studied extensively (see, for example, [12]) and might have potential for use in quantum binary state discrimination. The reason is that as mentioned in Chapter 2, a vector of relative frequencies can be viewed as the output of a probabilistic uniform quantizer applied to the corresponding vector of true probabilities. We briefly explored how first order sigma-delta could be applied to quantum state estimation as a way of more accurately representing an unknown density operator in terms of a set of "quantization levels" corresponding to possible relative frequency values. The hope was that we could formulate a problem and solution analogous to the problem of accurately reconstructing a bandlimited continuous time signal using quantized versions of its sample values. The issue we ran into is that in the latter scenario, the error values stemming from the effect of the quantizer are assumed to be known. In both an estimation setting and a discrimination setting the effect of the "quantizer" is not known, so to adapt the technique the error itself would also need to be estimated. This would be a very interesting possibility for future study.

Another example is the notion of generalized sampling [54], which roughly corresponds to gathering information about an unknown vector by taking its inner products with other known vectors in the same space. In this sense measuring a QMS can be considered as "sampling" its density operator since the probabilities of the measurement outcomes are the inner products of the density operator with the POVM elements of the measurement. It would be interesting to investigate how other sampling paradigms that have been studied for signal processing applications could be applied in this setting. This is already an active area of research in some cases. Compressive sensing has been studied in the context of classical signal detection [36, 37]

and also quantum state estimation and detection [33, 44]. The question of what it means to obtain nonuniform samples of a quantum state may also be interesting to consider.

7.3 Geometric Design of Qubit POVMs

Chapter 5 applies the methodology in Chapter 3 to density operators, POVM elements and in particular utilizing frame theory to characterize informationally overcomplete (IOC) POVMs. We specifically address the characterization of the density operators and POVM elements in operator space associated with qubits. In this case, density operators in this operator space are characterized by the well established Bloch ball and Bloch sphere. As we develop in Chapter 5, when we restrict the POVM elements to all be equal trace rank one operators in addition to being Hermitian and positive semi-definite, they each can also be characterized by a point on the surface of a ball, i.e., on a sphere. We refer to these as the Etro ball and Etro sphere which, for an M-element POVM have radius $\sqrt{2}/M$. In other words, with the restrictions above, an M-element POVM can be specified by M points on an Etro sphere.

Our discussion of POVM design in Chapter 6 is specifically based on selecting the M points on the Etro sphere from which the POVM is constructed. While a frame in operator space by definition spans the space and provides a valid complete or overcomplete representation of any operator in the space, it will not necessarily correspond to a valid POVM since for completeness POVMs have the additional constraint that all the elements must sum to the identity operator. Consequently in designing an IOC POVM it is necessary to ensure that it is both a frame for the operator space and a valid POVM. A commonly referenced class of IOC POVMs are those constructed using Platonic solids with M vertices where $M \in \{4, 6, 8, 12, 20\}$. In this case, the points on the Etro sphere are the M vertices of the corresponding Platonic solid. Example 5.5 in Section 5.4 utilized Platonic solids with $M = 4$ and $M = 6$ vertices and with $L = 5, 10$ and 20 to illustrate, in a very preliminary way, the effect of increasing either L or M for the QDOC for two somewhat arbitrarily chosen

density operators. As is clear in the example increasing either L or M improves the discrimination, a totally anticipated result. A strong candidate for future work is a much more detailed exploration of how L and M individually affect the discrimination with a broader set of examples. Also it would clearly be of interest to understand the tradeoff between L and M. Increasing L of course involves increasing the number of identically prepared QMSs, whereas increasing M involves increasing the number of POVM elements at the measurement stage. Essentially $1/L$ can be interpreted as a quantization step size for the relative frequencies, which suggests that an analysis involving quantization as it is typically viewed in signal processing might be of use.

The inspiration for Chapter 6 comes from the role of Platonic solids in quantum measurement and their potential use as illustrated in Example 5.5 for quantum state discrimination. As stated in Section 5.3 the vertices of each Platonic solid can be inscribed in an Etro sphere and define valid POVMs. In Chapter 6 we consider, in a somewhat preliminary and exploratory way, other possible distributions of points on the Etro sphere as the basis for POVMs to be used for qubit state discrimination. The problem considered is again discriminating between two qubits with a known orientation on the Bloch sphere with respect to each other but unknown orientation with respect to the Bloch sphere itself. Phrased differently, this corresponds to the rotational orientation of the Etro sphere with respect to the Bloch sphere being unknown. The simulations in Chapter 6 consider the minimum and maximum probabilities of error in discrimination over all possible rotations of the two spheres relative to each other and for a variety of choices for the POVMs resulting from distributing M points on the Etro sphere. While there are no strong conclusions to be drawn from these very preliminary simulations, Chapter 6 offers an initial approach to the design of IOC POVMs. The underlying ideas could be extended both analytically and through further simulations in many different ways. As an example, the problem could be formally considered as a nonlinear optimization problem in which the goal is to minimize the quantity $(\max P_e - \min P_e)$, as defined in Chapter 6, over all possible POVMs. The cost function could also be based on Bayes' cost rather than probability of error, or on a different of discrimination performance.

Furthermore, in Chapter 6 we arbitrarily assumed that all relative orientations of the Bloch and Etro spheres were equally likely. It would be interesting to know what distributions of orientations are actually achievable using a specific physical implementation of a qubit. For instance, assume that H_0 and H_1 correspond to two specified laboratory procedures that prepare L qubits in state ρ_0 or ρ_1. But assume as well that the procedure can only prepare the qubits in that state to within some error, corresponding for example to a small region on the surface of the Bloch sphere concentrated around the desired state. That distribution of states could then be factored into the design of the ideal POVM for discrimination. Other geometric techniques besides finding distributions of points that are maximally spread would likely come into play, and would need to be combined with the mathematical constraints stemming from the postulates of quantum mechanics. Another possible area of study concerns the fact that we only considered local, non-adaptive measurement schemes. In other words, we assumed that each QMS was measured individually and that the measurements performed were not dependent on any previous outcomes. We did not consider how collective and adaptive measurement schemes could be interpreted or exploited in the framework of operator spaces. Finally, we focused on Etro POVMs for simplicity and did not consider the question of how to choose the traces of the POVM elements to optimize discrimination performance. We also did not explicitly consider the generalization of Etro POVMs to dimensions larger than 2.

Appendix A

Chapter 6 Extended Results

Tables A.1 to A.34 below contain the minimum and maximum probabilities of error ($\min P_e$ and $\max P_e$) corresponding to a wide range of possible rotational orientations of various sets of M points on the corresponding Etro sphere. More specifically, for a given set of M points on an Etro sphere, the orientation was varied by choosing an arbitrary point as the north pole and then incrementing the azimuth and elevation angles of that point over their full ranges using a step size of $\pi/50$.

Table A.1: $\min P_e$ and $\max P_e$ for a tetrahedron ($M = 4$).

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.115	0.125	0.010	0.083	0.125	0.042
	$\pi/2$	0.092	0.125	0.033	0.027	0.105	0.078
	$3\pi/4$	0.071	0.125	0.054	0.008	0.067	0.060
1/4	$\pi/4$	0.225	0.250	0.025	0.155	0.239	0.084
	$\pi/2$	0.177	0.250	0.2073	0.047	0.166	0.119
	$3\pi/4$	0.139	0.250	0.111	0.015	0.098	0.083
3/8	$\pi/4$	0.326	0.375	0.049	0.202	0.318	0.115
	$\pi/2$	0.255	0.375	0.120	0.063	0.198	0.135
	$3\pi/4$	0.205	0.332	0.127	0.023	0.109	0.087
1/2	$\pi/4$	0.390	0.422	0.032	0.224	0.344	0.121
	$\pi/2$	0.296	0.356	0.060	0.076	0.214	0.138
	$3\pi/4$	0.233	0.311	0.078	0.029	0.123	0.094

Table A.2: $\min P_e$ and $\max P_e$ for an octahedron ($M = 6$).

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.118	0.125	0.007	0.095	0.123	0.028
	$\pi/2$	0.103	0.125	0.022	0.046	0.087	0.041
	$3\pi/4$	0.089	0.125	0.036	0.023	0.052	0.029
1/4	$\pi/4$	0.233	0.250	0.017	0.177	0.224	0.048
	$\pi/2$	0.202	0.250	0.048	0.081	0.135	0.053
	$3\pi/4$	0.176	0.229	0.053	0.042	0.075	0.033
3/8	$\pi/4$	0.342	0.371	0.029	0.233	0.287	0.053
	$\pi/2$	0.287	0.309	0.023	0.103	0.156	0.053
	$3\pi/4$	0.232	0.272	0.040	0.055	0.086	0.031
1/2	$\pi/4$	0.390	0.436	0.047	0.255	0.317	0.061
	$\pi/2$	0.296	0.382	0.086	0.111	0.174	0.063
	$3\pi/4$	0.233	0.346	0.113	0.063	0.094	0.031

Table A.3: $\min P_e$ and $\max P_e$ for a cube ($M = 8$).

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.120	0.125	0.005	0.101	0.123	0.022
	$\pi/2$	0.108	0.125	0.017	0.055	0.086	0.031
	$3\pi/4$	0.098	0.125	0.027	0.031	0.052	0.021
1/4	$\pi/4$	0.238	0.250	0.012	0.185	0.223	0.038
	$\pi/2$	0.214	0.250	0.036	0.091	0.133	0.042
	$3\pi/4$	0.182	0.228	0.046	0.051	0.073	0.022
3/8	$\pi/4$	0.346	0.371	0.025	0.240	0.288	0.048
	$\pi/2$	0.289	0.315	0.026	0.116	0.154	0.039
	$3\pi/4$	0.232	0.272	0.040	0.063	0.084	0.021
1/2	$\pi/4$	0.390	0.422	0.032	0.258	0.312	0.053
	$\pi/2$	0.296	0.356	0.060	0.122	0.170	0.048
	$3\pi/4$	0.233	0.311	0.078	0.068	0.093	0.025

Table A.4: $\min P_e$ and $\max P_e$ for an icosahedron ($M = 12$).

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.122	0.125	0.003	0.107	0.120	0.012
	$\pi/2$	0.114	0.125	0.011	0.065	0.076	0.011
	$3\pi/4$	0.107	0.120	0.014	0.038	0.046	0.008
1/4	$\pi/4$	0.242	0.250	0.008	0.196	0.213	0.017
	$\pi/2$	0.218	0.228	0.010	0.105	0.116	0.011
	$3\pi/4$	0.188	0.214	0.026	0.059	0.066	0.008
3/8	$\pi/4$	0.349	0.359	0.010	0.253	0.270	0.017
	$\pi/2$	0.291	0.305	0.014	0.128	0.138	0.009
	$3\pi/4$	0.245	0.256	0.011	0.070	0.077	0.007
1/2	$\pi/4$	0.397	0.412	0.015	0.274	0.290	0.016
	$\pi/2$	0.309	0.338	0.028	0.137	0.145	0.008
	$3\pi/4$	0.251	0.288	0.037	0.074	0.081	0.007

Table A.5: $\min P_e$ and $\max P_e$ for a set of $M = 4$ points on a sphere with minimum Riesz 0-energy.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.115	0.125	0.010	0.082	0.125	0.043
	$\pi/2$	0.092	0.125	0.033	0.027	0.105	0.077
	$3\pi/4$	0.071	0.125	0.054	0.008	0.068	0.060
1/4	$\pi/4$	0.225	0.250	0.025	0.155	0.239	0.084
	$\pi/2$	0.177	0.250	0.073	0.047	0.167	0.119
	$3\pi/4$	0.139	0.250	0.111	0.015	0.098	0.082
3/8	$\pi/4$	0.326	0.375	0.049	0.203	0.318	0.115
	$\pi/2$	0.255	0.375	0.120	0.063	0.198	0.135
	$3\pi/4$	0.205	0.332	0.127	0.022	0.110	0.088
1/2	$\pi/4$	0.390	0.422	0.032	0.223	0.343	0.119
	$\pi/2$	0.296	0.355	0.059	0.076	0.213	0.137
	$3\pi/4$	0.233	0.311	0.077	0.029	0.121	0.092

Table A.6: $\min P_e$ and $\max P_e$ for a set of $M = 5$ points on a sphere with minimum Riesz 0-energy.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.117	0.125	0.008	0.089	0.125	0.036
	$\pi/2$	0.098	0.125	0.027	0.038	0.101	0.064
	$3\pi/4$	0.082	0.125	0.043	0.015	0.066	0.051
1/4	$\pi/4$	0.230	0.250	0.020	0.165	0.236	0.071
	$\pi/2$	0.192	0.250	0.058	0.066	0.157	0.091
	$3\pi/4$	0.161	0.250	0.089	0.028	0.091	0.063
3/8	$\pi/4$	0.336	0.375	0.039	0.216	0.307	0.091
	$\pi/2$	0.269	0.339	0.070	0.084	0.183	0.010
	$3\pi/4$	0.212	0.300	0.088	0.037	0.105	0.068
1/2	$\pi/4$	0.392	0.433	0.042	0.235	0.332	0.097
	$\pi/2$	0.300	0.377	0.077	0.091	0.198	0.107
	$3\pi/4$	0.239	0.339	0.100	0.044	0.116	0.072

Table A.7: $\min P_e$ and $\max P_e$ for a set of $M = 6$ points on a sphere with minimum Riesz 0-energy.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.118	0.125	0.007	0.095	0.123	0.028
	$\pi/2$	0.103	0.125	0.022	0.045	0.086	0.041
	$3\pi/4$	0.089	0.125	0.036	0.023	0.052	0.029
1/4	$\pi/4$	0.233	0.250	0.017	0.177	0.224	0.048
	$\pi/2$	0.202	0.250	0.048	0.082	0.132	0.050
	$3\pi/4$	0.176	0.229	0.053	0.042	0.074	0.032
3/8	$\pi/4$	0.342	0.371	0.029	0.234	0.287	0.053
	$\pi/2$	0.287	0.310	0.023	0.103	0.155	0.052
	$3\pi/4$	0.232	0.273	0.041	0.055	0.085	0.030
1/2	$\pi/4$	0.390	0.436	0.047	0.255	0.316	0.061
	$\pi/2$	0.296	0.392	0.086	0.111	0.174	0.063
	$3\pi/4$	0.239	0.339	0.100	0.063	0.092	0.029

Table A.8: $\min P_e$ and $\max P_e$ for a set of $M = 7$ points on a sphere with minimum Riesz 0-energy.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.119	0.125	0.006	0.097	0.123	0.026
	$\pi/2$	0.106	0.125	0.019	0.050	0.092	0.042
	$3\pi/4$	0.094	0.125	0.031	0.027	0.055	0.028
1/4	$\pi/4$	0.236	0.250	0.014	0.178	0.228	0.050
	$\pi/2$	0.209	0.250	0.041	0.083	0.143	0.060
	$3\pi/4$	0.175	0.219	0.045	0.045	0.080	0.036
3/8	$\pi/4$	0.343	0.364	0.021	0.232	0.297	0.065
	$\pi/2$	0.278	0.316	0.038	0.103	0.169	0.066
	$3\pi/4$	0.230	0.285	0.056	0.054	0.094	0.040
1/2	$\pi/4$	0.396	0.445	0.049	0.252	0.332	0.080
	$\pi/2$	0.308	0.398	0.090	0.112	0.199	0.087
	$3\pi/4$	0.249	0.367	0.118	0.059	0.118	0.060

Table A.9: $\min P_e$ and $\max P_e$ for a set of $M = 8$ points on a sphere with minimum Riesz 0-energy.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.120	0.125	0.005	0.101	0.123	0.022
	$\pi/2$	0.108	0.125	0.017	0.056	0.087	0.031
	$3\pi/4$	0.098	0.125	0.027	0.031	0.055	0.024
1/4	$\pi/4$	0.238	0.250	0.012	0.186	0.224	0.038
	$\pi/2$	0.214	0.250	0.036	0.092	0.131	0.038
	$3\pi/4$	0.183	0.234	0.050	0.051	0.076	0.025
3/8	$\pi/4$	0.347	0.374	0.027	0.242	0.285	0.043
	$\pi/2$	0.281	0.317	0.037	0.115	0.155	0.040
	$3\pi/4$	0.228	0.291	0.062	0.062	0.089	0.027
1/2	$\pi/4$	0.392	0.414	0.022	0.261	0.305	0.043
	$\pi/2$	0.300	0.341	0.041	0.124	0.160	0.036
	$3\pi/4$	0.239	0.293	0.054	0.066	0.091	0.025

Table A.10: $\min P_e$ and $\max P_e$ for a set of $M = 9$ points on a sphere with minimum Riesz 0-energy.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.120	0.125	0.005	0.103	0.122	0.019
	$\pi/2$	0.110	0.125	0.015	0.059	0.083	0.025
	$3\pi/4$	0.101	0.125	0.024	0.034	0.050	0.016
1/4	$\pi/4$	0.239	0.250	0.011	0.189	0.221	0.032
	$\pi/2$	0.216	0.238	0.022	0.098	0.127	0.028
	$3\pi/4$	0.184	0.216	0.032	0.053	0.071	0.018
3/8	$\pi/4$	0.348	0.361	0.013	0.247	0.280	0.033
	$\pi/2$	0.279	0.322	0.043	0.121	0.149	0.028
	$3\pi/4$	0.230	0.283	0.054	0.065	0.082	0.017
1/2	$\pi/4$	0.397	0.413	0.016	0.271	0.300	0.029
	$\pi/2$	0.310	0.339	0.029	0.129	0.156	0.027
	$3\pi/4$	0.251	0.289	0.038	0.068	0.085	0.016

Table A.11: $\min P_e$ and $\max P_e$ for a set of $M = 10$ points on a sphere with minimum Riesz 0-energy.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.121	0.125	0.004	0.105	0.121	0.017
	$\pi/2$	0.112	0.125	0.013	0.061	0.080	0.019
	$3\pi/4$	0.103	0.125	0.022	0.035	0.049	0.014
1/4	$\pi/4$	0.240	0.250	0.010	0.192	0.217	0.025
	$\pi/2$	0.217	0.232	0.015	0.100	0.121	0.021
	$3\pi/4$	0.188	0.212	0.024	0.055	0.070	0.015
3/8	$\pi/4$	0.350	0.359	0.009	0.249	0.276	0.028
	$\pi/2$	0.285	0.316	0.031	0.123	0.144	0.020
	$3\pi/4$	0.233	0.278	0.046	0.066	0.081	0.015
1/2	$\pi/4$	0.397	0.415	0.018	0.268	0.296	0.028
	$\pi/2$	0.310	0.344	0.033	0.132	0.151	0.019
	$3\pi/4$	0.252	0.296	0.043	0.070	0.086	0.015

Table A.12: $\min P_e$ and $\max P_e$ for a set of $M = 11$ points on a sphere with minimum Riesz 0-energy.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.121	0.125	0.004	0.106	0.121	0.016
	$\pi/2$	0.113	0.125	0.012	0.063	0.082	0.019
	$3\pi/4$	0.105	0.125	0.020	0.035	0.050	0.015
1/4	$\pi/4$	0.241	0.250	0.009	0.192	0.218	0.026
	$\pi/2$	0.215	0.236	0.020	0.101	0.124	0.022
	$3\pi/4$	0.185	0.217	0.033	0.054	0.071	0.017
3/8	$\pi/4$	0.348	0.362	0.014	0.248	0.278	0.029
	$\pi/2$	0.285	0.315	0.030	0.124	0.145	0.021
	$3\pi/4$	0.236	0.275	0.039	0.065	0.082	0.017
1/2	$\pi/4$	0.397	0.419	0.022	0.268	0.297	0.029
	$\pi/2$	0.310	0.350	0.040	0.131	0.151	0.020
	$3\pi/4$	0.251	0.304	0.053	0.070	0.086	0.016

Table A.13: $\min P_e$ and $\max P_e$ for a set of $M = 12$ points on a sphere with minimum Riesz 0-energy.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.122	0.125	0.003	0.107	0.120	0.012
	$\pi/2$	0.114	0.125	0.011	0.065	0.076	0.011
	$3\pi/4$	0.107	0.120	0.014	0.038	0.046	0.008
1/4	$\pi/4$	0.242	0.250	0.008	0.196	0.212	0.017
	$\pi/2$	0.218	0.228	0.010	0.105	0.116	0.011
	$3\pi/4$	0.188	0.214	0.026	0.059	0.066	0.007
3/8	$\pi/4$	0.349	0.359	0.010	0.253	0.270	0.016
	$\pi/2$	0.291	0.305	0.014	0.128	0.137	0.009
	$3\pi/4$	0.245	0.256	0.011	0.070	0.077	0.007
1/2	$\pi/4$	0.397	0.412	0.015	0.274	0.290	0.016
	$\pi/2$	0.309	0.338	0.028	0.137	0.144	0.007
	$3\pi/4$	0.251	0.288	0.037	0.074	0.081	0.007

Table A.14: $\min P_e$ and $\max P_e$ for a set of $M = 4$ points on a sphere with maximum convex hull.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.115	0.125	0.010	0.082	0.125	0.043
	$\pi/2$	0.092	0.125	0.033	0.027	0.105	0.078
	$3\pi/4$	0.071	0.125	0.054	0.008	0.068	0.060
1/4	$\pi/4$	0.225	0.250	0.025	0.155	0.239	0.084
	$\pi/2$	0.177	0.250	0.073	0.047	0.167	0.119
	$3\pi/4$	0.139	0.250	0.111	0.015	0.098	0.083
3/8	$\pi/4$	0.326	0.375	0.049	0.203	0.318	0.115
	$\pi/2$	0.255	0.375	0.120	0.063	0.198	0.135
	$3\pi/4$	0.205	0.332	0.126	0.022	0.110	0.088
1/2	$\pi/4$	0.390	0.422	0.032	0.223	0.343	0.120
	$\pi/2$	0.296	0.356	0.060	0.076	0.213	0.137
	$3\pi/4$	0.233	0.311	0.078	0.029	0.122	0.092

Table A.15: $\min P_e$ and $\max P_e$ for a set of $M = 5$ points on a sphere with maximum convex hull.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.117	0.125	0.008	0.090	0.125	0.035
	$\pi/2$	0.098	0.125	0.027	0.038	0.107	0.068
	$3\pi/4$	0.082	0.125	0.043	0.015	0.071	0.056
1/4	$\pi/4$	0.230	0.250	0.020	0.168	0.239	0.071
	$\pi/2$	0.192	0.250	0.058	0.068	0.171	0.104
	$3\pi/4$	0.161	0.250	0.089	0.028	0.103	0.075
3/8	$\pi/4$	0.336	0.375	0.039	0.223	0.320	0.097
	$\pi/2$	0.269	0.339	0.071	0.087	0.204	0.116
	$3\pi/4$	0.212	0.301	0.088	0.037	0.118	0.081
1/2	$\pi/4$	0.392	0.434	0.042	0.246	0.350	0.104
	$\pi/2$	0.300	0.377	0.077	0.097	0.224	0.126
	$3\pi/4$	0.239	0.340	0.101	0.043	0.134	0.091

Table A.16: $\min P_e$ and $\max P_e$ for a set of $M = 6$ points on a sphere with maximum convex hull.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.118	0.125	0.007	0.095	0.123	0.028
	$\pi/2$	0.103	0.125	0.022	0.046	0.087	0.041
	$3\pi/4$	0.089	0.125	0.036	0.023	0.052	0.029
1/4	$\pi/4$	0.233	0.250	0.017	0.177	0.224	0.048
	$\pi/2$	0.202	0.250	0.048	0.081	0.133	0.052
	$3\pi/4$	0.176	0.229	0.053	0.042	0.075	0.033
3/8	$\pi/4$	0.342	0.371	0.029	0.234	0.287	0.053
	$\pi/2$	0.2387	0.310	0.023	0.103	0.156	0.052
	$3\pi/4$	0.232	0.273	0.041	0.055	0.086	0.031
1/2	$\pi/4$	0.390	0.436	0.047	0.255	0.317	0.061
	$\pi/2$	0.296	0.382	0.086	0.111	0.174	0.063
	$3\pi/4$	0.233	0.346	0.113	0.063	0.092	0.029

Table A.17: $\min P_e$ and $\max P_e$ for a set of $M = 7$ points on a sphere with maximum convex hull.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.119	0.125	0.006	0.098	0.123	0.025
	$\pi/2$	0.106	0.125	0.019	0.051	0.092	0.041
	$3\pi/4$	0.094	0.125	0.031	0.026	0.055	0.029
1/4	$\pi/4$	0.236	0.250	0.014	0.181	0.228	0.047
	$\pi/2$	0.208	0.250	0.042	0.088	0.143	0.056
	$3\pi/4$	0.175	0.219	0.045	0.045	0.080	0.035
3/8	$\pi/4$	0.343	0.364	0.021	0.239	0.297	0.058
	$\pi/2$	0.278	0.316	0.038	0.112	0.169	0.057
	$3\pi/4$	0.230	0.285	0.055	0.056	0.094	0.038
1/2	$\pi/4$	0.396	0.445	0.049	0.262	0.332	0.070
	$\pi/2$	0.308	0.399	0.091	0.121	0.199	0.079
	$3\pi/4$	0.249	0.368	0.119	0.061	0.119	0.058

Table A.18: $\min P_e$ and $\max P_e$ for a set of $M = 8$ points on a sphere with maximum convex hull.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.120	0.125	0.005	0.100	0.123	0.023
	$\pi/2$	0.108	0.125	0.017	0.053	0.087	0.034
	$3\pi/4$	0.098	0.125	0.027	0.029	0.053	0.024
1/4	$\pi/4$	0.238	0.250	0.012	0.184	0.224	0.041
	$\pi/2$	0.213	0.245	0.032	0.088	0.133	0.045
	$3\pi/4$	0.181	0.218	0.037	0.045	0.075	0.030
3/8	$\pi/4$	0.346	0.363	0.017	0.238	0.288	0.050
	$\pi/2$	0.275	0.318	0.043	0.110	0.157	0.047
	$3\pi/4$	0.222	0.289	0.067	0.055	0.087	0.032
1/2	$\pi/4$	0.396	0.426	0.030	0.257	0.310	0.054
	$\pi/2$	0.308	0.364	0.056	0.118	0.168	0.050
	$3\pi/4$	0.249	0.322	0.073	0.059	0.090	0.031

Table A.19: $\min P_e$ and $\max P_e$ for a set of $M = 9$ points on a sphere with maximum convex hull.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.120	0.125	0.005	0.103	0.122	0.019
	$\pi/2$	0.110	0.125	0.015	0.058	0.084	0.026
	$3\pi/4$	0.101	0.125	0.024	0.033	0.053	0.020
1/4	$\pi/4$	0.239	0.250	0.011	0.188	0.220	0.032
	$\pi/2$	0.216	0.238	0.022	0.096	0.128	0.032
	$3\pi/4$	0.186	0.213	0.027	0.051	0.076	0.025
3/8	$\pi/4$	0.348	0.360	0.012	0.245	0.281	0.036
	$\pi/2$	0.280	0.323	0.043	0.117	0.151	0.033
	$3\pi/4$	0.231	0.287	0.056	0.062	0.089	0.026
1/2	$\pi/4$	0.395	0.413	0.018	0.266	0.301	0.036
	$\pi/2$	0.306	0.340	0.033	0.126	0.157	0.032
	$3\pi/4$	0.247	0.290	0.043	0.066	0.092	0.026

Table A.20: $\min P_e$ and $\max P_e$ for a set of $M = 12$ points on a sphere with maximum convex hull.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.122	0.125	0.003	0.107	0.120	0.012
	$\pi/2$	0.114	0.125	0.011	0.065	0.076	0.011
	$3\pi/4$	0.107	0.120	0.014	0.038	0.046	0.008
1/4	$\pi/4$	0.242	0.250	0.008	0.196	0.213	0.017
	$\pi/2$	0.218	0.228	0.010	0.105	0.116	0.011
	$3\pi/4$	0.188	0.214	0.026	0.059	0.066	0.008
3/8	$\pi/4$	0.349	0.359	0.010	0.253	0.270	0.017
	$\pi/2$	0.291	0.305	0.014	0.128	0.138	0.009
	$3\pi/4$	0.245	0.256	0.011	0.070	0.077	0.007
1/2	$\pi/4$	0.397	0.412	0.015	0.274	0.290	0.016
	$\pi/2$	0.309	0.338	0.028	0.137	0.144	0.008
	$3\pi/4$	0.251	0.288	0.037	0.074	0.081	0.007

Table A.21: $\min P_e$ and $\max P_e$ for a set of $M = 4$ points on a sphere with maximum nearest neighbor distance.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.115	0.125	0.010	0.083	0.125	0.042
	$\pi/2$	0.092	0.125	0.033	0.027	0.105	0.078
	$3\pi/4$	0.071	0.125	0.054	0.008	0.067	0.060
1/4	$\pi/4$	0.225	0.250	0.025	0.155	0.239	0.084
	$\pi/2$	0.177	0.250	0.073	0.047	0.166	0.119
	$3\pi/4$	0.139	0.250	0.111	0.015	0.098	0.083
3/8	$\pi/4$	0.326	0.375	0.049	0.202	0.318	0.115
	$\pi/2$	0.255	0.375	0.120	0.063	0.198	0.135
	$3\pi/4$	0.205	0.332	0.127	0.023	0.109	0.087
1/2	$\pi/4$	0.390	0.422	0.032	0.224	0.344	0.121
	$\pi/2$	0.296	0.356	0.060	0.076	0.214	0.138
	$3\pi/4$	0.233	0.311	0.078	0.029	0.123	0.094

Table A.22: $\min P_e$ and $\max P_e$ for a set of $M = 5$ points on a sphere with maximum nearest neighbor distance.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.117	0.125	0.008	0.090	0.125	0.035
	$\pi/2$	0.098	0.125	0.027	0.038	0.106	0.068
	$3\pi/4$	0.082	0.125	0.043	0.015	0.071	0.056
1/4	$\pi/4$	0.230	0.250	0.020	0.168	0.239	0.071
	$\pi/2$	0.192	0.250	0.058	0.068	0.171	0.103
	$3\pi/4$	0.161	0.250	0.089	0.028	0.103	0.075
3/8	$\pi/4$	0.336	0.375	0.039	0.224	0.319	0.095
	$\pi/2$	0.269	0.339	0.071	0.089	0.204	0.115
	$3\pi/4$	0.212	0.301	0.089	0.037	0.118	0.080
1/2	$\pi/4$	0.392	0.433	0.042	0.240	0.348	0.098
	$\pi/2$	0.300	0.377	0.077	0.101	0.221	0.120
	$3\pi/4$	0.239	0.339	0.101	0.044	0.132	0.088

Table A.23: $\min P_e$ and $\max P_e$ for a set of $M = 6$ points on a sphere with maximum nearest neighbor distance.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.118	0.125	0.007	0.095	0.123	0.028
	$\pi/2$	0.103	0.125	0.022	0.046	0.087	0.041
	$3\pi/4$	0.089	0.125	0.036	0.023	0.052	0.029
1/4	$\pi/4$	0.233	0.250	0.017	0.177	0.224	0.048
	$\pi/2$	0.202	0.250	0.048	0.081	0.135	0.053
	$3\pi/4$	0.176	0.229	0.053	0.042	0.075	0.033
3/8	$\pi/4$	0.342	0.371	0.029	0.233	0.287	0.053
	$\pi/2$	0.287	0.309	0.023	0.103	0.156	0.053
	$3\pi/4$	0.232	0.272	0.040	0.055	0.086	0.031
1/2	$\pi/4$	0.390	0.436	0.047	0.255	0.317	0.061
	$\pi/2$	0.296	0.382	0.086	0.111	0.174	0.063
	$3\pi/4$	0.233	0.346	0.113	0.063	0.094	0.031

Table A.24: $\min P_e$ and $\max P_e$ for a set of $M = 8$ points on a sphere with maximum nearest neighbor distance.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.120	0.125	0.005	0.100	0.124	0.024
	$\pi/2$	0.108	0.125	0.017	0.053	0.096	0.043
	$3\pi/4$	0.098	0.125	0.027	0.029	0.068	0.039
1/4	$\pi/4$	0.238	0.250	0.012	0.183	0.230	0.047
	$\pi/2$	0.214	0.250	0.036	0.086	0.145	0.059
	$3\pi/4$	0.185	0.249	0.064	0.044	0.097	0.052
3/8	$\pi/4$	0.349	0.375	0.026	0.236	0.292	0.056
	$\pi/2$	0.274	0.319	0.045	0.106	0.167	0.061
	$3\pi/4$	0.221	0.290	0.069	0.053	0.113	0.059
1/2	$\pi/4$	0.393	0.417	0.025	0.255	0.311	0.056
	$\pi/2$	0.302	0.347	0.046	0.113	0.173	0.060
	$3\pi/4$	0.241	0.301	0.060	0.057	0.116	0.059

Table A.25: $\min P_e$ and $\max P_e$ for a set of $M = 9$ points on a sphere with maximum nearest neighbor distance.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.120	0.125	0.005	0.102	0.123	0.020
	$\pi/2$	0.110	0.125	0.015	0.057	0.085	0.028
	$3\pi/4$	0.101	0.125	0.024	0.031	0.054	0.023
1/4	$\pi/4$	0.239	0.250	0.011	0.186	0.221	0.035
	$\pi/2$	0.218	0.245	0.028	0.090	0.127	0.037
	$3\pi/4$	0.182	0.223	0.041	0.048	0.076	0.028
3/8	$\pi/4$	0.347	0.365	0.018	0.241	0.279	0.038
	$\pi/2$	0.277	0.322	0.045	0.108	0.147	0.040
	$3\pi/4$	0.228	0.279	0.051	0.058	0.088	0.029
1/2	$\pi/4$	0.396	0.414	0.018	0.263	0.296	0.033
	$\pi/2$	0.308	0.341	0.034	0.116	0.152	0.036
	$3\pi/4$	0.249	0.293	0.044	0.063	0.090	0.027

Table A.26: $\min P_e$ and $\max P_e$ for a set of $M = 12$ points on a sphere with maximum nearest neighbor distance.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.122	0.125	0.003	0.107	0.120	0.012
	$\pi/2$	0.114	0.125	0.011	0.065	0.076	0.011
	$3\pi/4$	0.107	0.120	0.014	0.038	0.046	0.008
1/4	$\pi/4$	0.242	0.250	0.008	0.196	0.213	0.017
	$\pi/2$	0.218	0.228	0.010	0.105	0.116	0.011
	$3\pi/4$	0.188	0.214	0.026	0.059	0.066	0.008
3/8	$\pi/4$	0.349	0.359	0.010	0.253	0.270	0.017
	$\pi/2$	0.291	0.305	0.014	0.128	0.138	0.009
	$3\pi/4$	0.245	0.256	0.011	0.070	0.077	0.007
1/2	$\pi/4$	0.397	0.412	0.015	0.274	0.290	0.016
	$\pi/2$	0.309	0.338	0.028	0.137	0.145	0.008
	$3\pi/4$	0.251	0.288	0.037	0.074	0.081	0.007

Table A.27: $\min P_e$ and $\max P_e$ for a set of $M = 4$ points on a sphere with minimum covering radius.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.115	0.125	0.010	0.083	0.125	0.042
	$\pi/2$	0.092	0.125	0.033	0.027	0.105	0.078
	$3\pi/4$	0.071	0.125	0.054	0.008	0.067	0.059
1/4	$\pi/4$	0.225	0.250	0.025	0.155	0.239	0.084
	$\pi/2$	0.177	0.250	0.073	0.048	0.167	0.119
	$3\pi/4$	0.139	0.250	0.111	0.015	0.098	0.082
3/8	$\pi/4$	0.326	0.375	0.049	0.203	0.318	0.115
	$\pi/2$	0.255	0.375	0.120	0.064	0.198	0.135
	$3\pi/4$	0.205	0.332	0.127	0.023	0.110	0.087
1/2	$\pi/4$	0.390	0.422	0.032	0.224	0.342	0.118
	$\pi/2$	0.296	0.355	0.059	0.077	0.211	0.135
	$3\pi/4$	0.233	0.311	0.078	0.029	0.120	0.091

Table A.28: $\min P_e$ and $\max P_e$ for a set of $M = 5$ points on a sphere with minimum covering radius.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.117	0.125	0.008	0.090	0.125	0.035
	$\pi/2$	0.098	0.125	0.027	0.038	0.107	0.068
	$3\pi/4$	0.082	0.125	0.043	0.015	0.071	0.056
1/4	$\pi/4$	0.230	0.250	0.020	0.167	0.239	0.071
	$\pi/2$	0.192	0.250	0.058	0.067	0.171	0.104
	$3\pi/4$	0.161	0.250	0.089	0.028	0.103	0.076
3/8	$\pi/4$	0.336	0.375	0.039	0.221	0.320	0.098
	$\pi/2$	0.269	0.340	0.071	0.086	0.204	0.118
	$3\pi/4$	0.212	0.301	0.089	0.036	0.118	0.082
1/2	$\pi/4$	0.392	0.433	0.042	0.240	0.350	0.109
	$\pi/2$	0.300	0.377	0.077	0.092	0.223	0.131
	$3\pi/4$	0.239	0.340	0.101	0.043	0.134	0.091

Table A.29: $\min P_e$ and $\max P_e$ for a set of $M = 6$ points on a sphere with minimum covering radius.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.118	0.125	0.007	0.095	0.123	0.028
	$\pi/2$	0.103	0.125	0.022	0.046	0.086	0.040
	$3\pi/4$	0.089	0.125	0.036	0.023	0.052	0.029
1/4	$\pi/4$	0.233	0.250	0.017	0.177	0.224	0.047
	$\pi/2$	0.202	0.250	0.048	0.081	0.133	0.051
	$3\pi/4$	0.176	0.229	0.053	0.042	0.075	0.033
3/8	$\pi/4$	0.342	0.371	0.029	0.234	0.286	0.053
	$\pi/2$	0.287	0.310	0.023	0.103	0.155	0.052
	$3\pi/4$	0.232	0.273	0.041	0.055	0.086	0.031
1/2	$\pi/4$	0.390	0.436	0.047	0.255	0.317	0.061
	$\pi/2$	0.296	0.382	0.086	0.111	0.174	0.063
	$3\pi/4$	0.233	0.346	0.113	0.063	0.092	0.029

Table A.30: $\min P_e$ and $\max P_e$ for a set of $M = 7$ points on a sphere with minimum covering radius.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.119	0.125	0.006	0.098	0.124	0.025
	$\pi/2$	0.106	0.125	0.019	0.052	0.092	0.040
	$3\pi/4$	0.094	0.125	0.031	0.026	0.055	0.029
1/4	$\pi/4$	0.236	0.250	0.014	0.183	0.229	0.045
	$\pi/2$	0.208	0.250	0.042	0.089	0.144	0.055
	$3\pi/4$	0.175	0.219	0.045	0.045	0.080	0.034
3/8	$\pi/4$	0.343	0.364	0.021	0.240	0.298	0.057
	$\pi/2$	0.278	0.316	0.038	0.110	0.169	0.059
	$3\pi/4$	0.230	0.285	0.055	0.057	0.094	0.038
1/2	$\pi/4$	0.396	0.444	0.048	0.262	0.330	0.068
	$\pi/2$	0.308	0.396	0.088	0.120	0.196	0.076
	$3\pi/4$	0.249	0.365	0.116	0.062	0.116	0.055

Table A.31: $\min P_e$ and $\max P_e$ for a set of $M = 8$ points on a sphere with minimum covering radius.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.120	0.125	0.005	0.100	0.122	0.022
	$\pi/2$	0.108	0.125	0.017	0.054	0.084	0.030
	$3\pi/4$	0.098	0.125	0.027	0.030	0.052	0.022
1/4	$\pi/4$	0.238	0.250	0.012	0.183	0.222	0.038
	$\pi/2$	0.205	0.242	0.037	0.089	0.128	0.039
	$3\pi/4$	0.173	0.216	0.044	0.050	0.075	0.025
3/8	$\pi/4$	0.342	0.362	0.021	0.238	0.284	0.046
	$\pi/2$	0.281	0.318	0.037	0.107	0.151	0.044
	$3\pi/4$	0.232	0.286	0.054	0.061	0.089	0.028
1/2	$\pi/4$	0.392	0.423	0.032	0.259	0.307	0.049
	$\pi/2$	0.300	0.358	0.058	0.116	0.160	0.044
	$3\pi/4$	0.238	0.315	0.076	0.064	0.091	0.027

Table A.32: $\min P_e$ and $\max P_e$ for a set of $M = 9$ points on a sphere with minimum covering radius.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.120	0.125	0.005	0.103	0.122	0.019
	$\pi/2$	0.110	0.125	0.015	0.059	0.083	0.024
	$3\pi/4$	0.101	0.125	0.024	0.033	0.050	0.017
1/4	$\pi/4$	0.239	0.250	0.011	0.188	0.219	0.031
	$\pi/2$	0.216	0.237	0.021	0.097	0.126	0.030
	$3\pi/4$	0.185	0.215	0.030	0.052	0.072	0.020
3/8	$\pi/4$	0.349	0.361	0.012	0.245	0.280	0.035
	$\pi/2$	0.279	0.322	0.043	0.119	0.149	0.030
	$3\pi/4$	0.230	0.285	0.054	0.063	0.084	0.021
1/2	$\pi/4$	0.396	0.413	0.017	0.265	0.300	0.034
	$\pi/2$	0.309	0.340	0.031	0.128	0.155	0.028
	$3\pi/4$	0.250	0.290	0.042	0.067	0.088	0.021

Table A.33: $\min P_e$ and $\max P_e$ for a set of $M = 10$ points on a sphere with minimum covering radius.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.121	0.125	0.004	0.105	0.121	0.017
	$\pi/2$	0.112	0.125	0.013	0.061	0.080	0.018
	$3\pi/4$	0.103	0.125	0.022	0.035	0.048	0.013
1/4	$\pi/4$	0.240	0.250	0.010	0.192	0.217	0.025
	$\pi/2$	0.217	0.232	0.015	0.101	0.121	0.020
	$3\pi/4$	0.189	0.213	0.024	0.055	0.069	0.014
3/8	$\pi/4$	0.350	0.359	0.009	0.249	0.276	0.027
	$\pi/2$	0.285	0.315	0.030	0.124	0.143	0.019
	$3\pi/4$	0.232	0.278	0.046	0.067	0.080	0.013
1/2	$\pi/4$	0.398	0.416	0.017	0.269	0.296	0.028
	$\pi/2$	0.312	0.344	0.032	0.132	0.150	0.018
	$3\pi/4$	0.255	0.296	0.042	0.072	0.085	0.013

Table A.34: $\min P_e$ and $\max P_e$ for a set of $M = 12$ points on a sphere with minimum covering radius.

		$L = 1$			$L = 5$		
q_1	α	**Min** P_e	**Max** P_e	**Difference**	**Min** P_e	**Max** P_e	**Difference**
1/8	$\pi/4$	0.122	0.125	0.003	0.107	0.120	0.012
	$\pi/2$	0.114	0.125	0.011	0.065	0.076	0.011
	$3\pi/4$	0.107	0.120	0.014	0.038	0.046	0.008
1/4	$\pi/4$	0.242	0.250	0.008	0.196	0.213	0.017
	$\pi/2$	0.218	0.228	0.010	0.105	0.116	0.011
	$3\pi/4$	0.188	0.214	0.026	0.059	0.066	0.008
3/8	$\pi/4$	0.349	0.359	0.010	0.253	0.270	0.017
	$\pi/2$	0.291	0.305	0.014	0.128	0.138	0.009
	$3\pi/4$	0.245	0.256	0.011	0.070	0.077	0.007
1/2	$\pi/4$	0.397	0.412	0.015	0.274	0.290	0.016
	$\pi/2$	0.309	0.338	0.028	0.137	0.145	0.008
	$3\pi/4$	0.251	0.288	0.037	0.074	0.081	0.007

Bibliography

[1] R. B. A. Adamson and Aephraim M. Steinberg. Improving Quantum State Estimation with Mutually Unbiased Bases. *Physical Review Letters*, 105(3):030406, 2010.

[2] D. M. Appleby. Symmetric Informationally Complete Measurements of Arbitrary Rank. *Optics and Spectroscopy*, 103(3):416–428, 2007.

[3] Joonwoo Bae and Leong-Chuan Kwek. Quantum State Discrimination and Its Applications. *Journal of Physics A: Mathematical and Theoretical*, 48(8):083001, 2015.

[4] Sudipto Banerjee and Anindya Roy. *Linear Algebra and Matrix Analysis for Statistics*. CRC Press, 2014.

[5] J. Robert Beck and Edward K. Shultz. The Use of Relative Operating Characteristic (ROC) Curves in Test Performance Evaluation. *Archives of Pathology & Laboratory Medicine*, 110(1):13–20, 1986.

[6] John J. Benedetto and Andrew Kebo. The Role of Frame Force in Quantum Detection. *Journal of Fourier Analysis and Applications*, 14(3):443–474, 2008.

[7] Sterling K. Berberian. *Notes on Spectral Theory*. D. Van Nostrand Company, Inc., 1966.

[8] Bernhard Bodmann and John Haas. A Short History of Frames and Quantum Designs, 2017.

[9] András Bodor and Mátyás Koniorczyk. Receiver Operation Characteristics of Quantum State Discrimination. *Journal of Russian Laser Research*, 38(2):150–163, 2017.

[10] Sara Botelho-Andrade, Peter G. Casazza, Desai Cheng, John Haas, and Tin T. Tran. The Quantum Detection Problem: A Survey. In *Quantum Theory And Symmetries*, pages 337–352. Springer, 2017.

[11] Sara Botelho-Andrade, Peter G. Casazza, Desai Cheng, and Tin T. Tran. The Solution to the Frame Quantum Detection Problem. *Journal of Fourier Analysis and Applications*, pages 1–56, 2017.

[12] Petros T. Boufounos and Alan V. Oppenheim. Quantization Noise Shaping on Arbitrary Frame Expansions. *EURASIP Journal on Advances in Signal Processing*, 2006(1):053807, 2006.

[13] Peter G. Casazza. The Art of Frame Theory. *Taiwanese Journal of Mathematics*, 4(2):129–201, 2000.

[14] Peter G. Casazza, Matthew Fickus, Dustin G. Mixon, Jesse Peterson, and Ihar Smalyanau. Every Hilbert Space Frame has a Naimark Complement. *Journal of Mathematical Analysis and Applications*, 406(1):111–119, 2013.

[15] Peter G. Casazza and Jelena Kovačević. Equal-Norm Tight Frames with Erasures. *Advances in Computational Mathematics*, 18(2-4):387–430, 2003.

[16] Peter G. Casazza and Gitta Kutyniok. *Finite Frames: Theory and Applications*. Springer, 2012.

[17] Peter G Casazza, Gitta Kutyniok, and Friedrich Philipp. Introduction to Finite Frame Fheory. In *Finite frames*, pages 1–53. Springer, 2013.

[18] Carlton M. Caves, Christopher A. Fuchs, and Rüdiger Schack. Unknown Quantum States: The Quantum de Finetti Representation. *Journal of Mathematical Physics*, 43(9):4537–4559, 2002.

[19] Anthony Chefles. Quantum state discrimination. *Contemporary Physics*, 41(6):401–424, 2000.

[20] Zoran Cvetković and Martin Vetterli. Overcomplete expansions and robustness. In *Wavelet Analysis and Its Applications*, volume 7, pages 301–338. Elsevier, 1998.

[21] Thomas Decker, Dominik Janzing, and Thomas Beth. Quantum Circuits for Single-Qubit Measurements Corresponding to Platonic Solids. *International Journal of Quantum Information*, 2(03):353–377, 2004.

[22] Qi Ding, Catherine A. Medlock, and Alan V. Oppenheim. Povm design for quantum state discrimination. (in press).

[23] G. M. d'Ariano, P. Perinotti, and M. F. Sacchi. Informationally Complete Measurements and Group Representation. *Journal of Optics B: Quantum and Semiclassical Optics*, 6(6):S487, 2004.

[24] Giacomo Mauro D'Ariano and Paolo Perinotti. Optimal Data Processing for Quantum Measurements. *Physical Review Letters*, 98(2):020403, 2007.

[25] Yonina C. Eldar and G. David Forney. On Quantum Detection and the Square-Root Measurement. *IEEE Transactions on Information Theory*, 47(3):858–872, 2001.

[26] Yonina C. Eldar and G. David Forney. Optimal Tight Frames and Quantum Measurement. *IEEE Transactions on Information Theory*, 48(3):599–610, 2002.

[27] Tom Fawcett. An Introduction to ROC Analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.

[28] Jerome Finkelstein. Pure-State Informationally Complete and "Really" Complete Measurements. *Physical Review A*, 70(5):052107, 2004.

[29] Steven T. Flammia, Andrew Silberfarb, and Carlton M. Caves. Minimal Informationally Complete Measurements for Pure States. *Foundations of Physics*, 35(12):1985–2006, 2005.

[30] Christopher A. Fuchs, Michael C. Hoang, and Blake C. Stacey. The SIC Question: History and State of Play. *Axioms*, 6(3):21, 2017.

[31] Vivek K. Goyal, Martin Vetterli, and Nguyen T. Thao. Quantized Overcomplete Expansions in $\mathbb{R}^N$: Analysis, Synthesis, and Algorithms. *IEEE Transactions on Information Theory*, 44(1):16–31, 1998.

[32] Robert M. Gray and David L. Neuhoff. Quantization. *IEEE transactions on information theory*, 44(6):2325–2383, 1998.

[33] David Gross, Yi-Kai Liu, Steven T Flammia, Stephen Becker, and Jens Eisert. Quantum state tomography via compressed sensing. *Physical review letters*, 105(15):150401, 2010.

[34] David J. Hand. Measuring Classifier Performance: A Coherent Alternative to the Area Under the ROC Curve. *Machine Learning*, 77(1):103–123, 2009.

[35] Doug P Hardin, TJ Michaels, and Edward B Saff. A Comparison of Popular Point Configurations on $\mathbb{S}^2$. *arXiv preprint arXiv:1607.04590*, 2016.

[36] Jarvis Haupt, Rui Castro, Robert Nowak, Gerald Fudge, and Alex Yeh. Compressive sampling for signal classification. In *2006 Fortieth Asilomar Conference on Signals, Systems and Computers*, pages 1430–1434. IEEE, 2006.

[37] Jarvis Haupt and Robert Nowak. Compressive sampling for signal detection. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, volume 3, pages III–1509. IEEE, 2007.

[38] Carl W. Helstrom. Detection Theory and Quantum Mechanics. *Information and Control*, 10(3):254–291, 1967.

[39] Carl W. Helstrom. *Elements of Signal Detection and Estimation*. Prentice-Hall, Inc., 1994.

[40] Carl W. Helstrom, Jane W. S. Liu, and James P. Gordon. Quantum-Mechanical Communication Theory. *Proceedings of the IEEE*, 58(10):1578–1598, 1970.

[41] Jelena Kovacevic and Amina Chebira. Life Beyond Bases: The Advent of Frames (Part I). *IEEE Signal Processing Magazine*, 24(4):86–104, 2007.

[42] Jelena Kovacevic and Amina Chebira. Life Beyond Bases: The Advent of Frames (Part I). *IEEE Signal Processing Magazine*, 24(4):86–104, 2007.

[43] Paul C Leopardi. *Distributing Points on the Sphere: Partitions, Separation, Quadrature and Energy*. PhD thesis, University of New South Wales, Sydney, Australia, 2007.

[44] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost. Quantum principal component analysis. *Nature Physics*, 10(9):631–633, 2014.

[45] Jorge M. Lobo, Alberto Jiménez-Valverde, and Raimundo Real. Auc: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17(2):145–151, 2008.

[46] Catherine Medlock, Alan Oppenheim, and Petros Boufounos. Informationally overcomplete povms for quantum state estimation and binary detection, 2020.

[47] Catherine A. Medlock and Alan V. Oppenheim. Optimal ROC Curves from Score Variable Threshold Tests. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 5327–5330. IEEE, 2019.

[48] Catherine A. Medlock and Alan V. Oppenheim. Operating characteristics for classical and quantum binary hypothesis testing. *Foundations and Trends® in Signal Processing*, 15, 2021. in press.

[49] Michael A. Nielsen and Isaac Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2016.

[50] Nancy A. Obuchowski. Receiver Operating Characteristic Curves and Their Use in Radiology. *Radiology*, 229(1):3–8, 2003.

[51] Alan V. Oppenheim and George C. Verghese. *Signals, Systems and Inference*. Pearson, 2015.

[52] Michał Oszmaniec, Leonardo Guerini, Peter Wittek, and Antonio Acín. Simulating positive-operator-valued measures with projective measurements. *Physical review letters*, 119(19):190501, 2017.

[53] Michał Oszmaniec, Filip B Maciejewski, and Zbigniew Puchała. Simulating all quantum measurements using only projective measurements and postselection. *Physical Review A*, 100(1):012351, 2019.

[54] Athanasios Papoulis. Generalized sampling expansion. *IEEE transactions on circuits and systems*, 24(11):652–654, 1977.

[55] Matteo G. A. Paris. The modern tools of quantum mechanics. *The European Physical Journal Special Topics*, 203(1):61–86, 2012.

[56] E. Parzen. On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.

[57] Asher Peres. Neumark's theorem and quantum inseparability. *Foundations of Physics*, 20(12):1441–1453, 1990.

[58] Asher Peres. *Quantum theory: concepts and methods*, volume 57. Springer Science & Business Media, 2006.

[59] Nicola Dalla Pozza and Matteo GA Paris. An effective iterative method to build the naimark extension of rank-n povms. *International Journal of Quantum Information*, 15(04):1750029, 2017.

[60] Provost, F. J., Fawcett, T., *et al.* Analysis and Visualization of Classifier Performance: Comparison under Imprecise Class and Cost Distributions. In *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining*, volume 97, pages 43–48, 1997.

[61] Jaroslav Řeháček, Yong Siah Teo, and Zdeněk Hradil. Determining Which Quantum Measurement Performs Better for State Estimation. *Physical Review A*, 92(1):012108, 2015.

[62] Joseph M. Renes, Robin Blume-Kohout, Andrew J. Scott, and Carlton M. Caves. Symmetric Informationally Complete Quantum Measurements. *Journal of Mathematical Physics*, 45(6):2171–2180, 2004.

[63] M. Rosenblatt. Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics*, 27(3):832–837, 1956.

[64] Mary Beth Ruskai. Some Connections between Frames, Mutually Unbiased Bases, and POVM's in Quantum Information Theory. *Acta Applicandae Mathematicae*, 108(3):709, 2009.

[65] Edward B Saff and Amo BJ Kuijlaars. Distributing Many Points on a Sphere. *The mathematical intelligencer*, 19(1):5–11, 1997.

[66] T. N. Sainath and C. Parada. Convolutional Neural Networks for Small-Footprint Keyword Spotting. In *Proceedings of the Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[67] Andrew J. Scott. Tight Informationally Complete Quantum Measurements. *Journal of Physics A: Mathematical and General*, 39(43):13507, 2006.

[68] Tanmay Singal, Filip B Maciejewski, and Michał Oszmaniec. Implementation of quantum measurements using classical resources and only a single ancillary qubit. *arXiv preprint arXiv:2104.05612*, 2021.

[69] N. J. A. Sloane. Spherical Codes: Nice Arrangements of Points on a Sphere in Various Dimensions. Online. Available: http://neilsloane.com/packings/.

[70] Wojciech Słomczyński and Anna Szymusiak. Highly Symmetric POVMs and Their Informational Power. *Quantum Information Processing*, 15(1):565–606, 2016.

[71] Kent A. Spackman. Signal Detection Theory: Valuable Tools for Evaluating Inductive Learning. In *Proceedings of the Sixth International Workshop on Machine Learning*, pages 160–163. Morgan Kaufmann Publishers Inc., 1989.

[72] Gaetana Spedalieri. *Quantum Hypothesis Testing: Theory and Applications to Quantum Sensing and Data Readout*. PhD thesis, University of York, 2016.

[73] R. M. Stein. The Relationship Between Default Prediction and Lending Profits: Integrating ROC Analysis and Loan Pricing. *Journal of Banking & Finance*, 29(5):1213–1236, 2005.

[74] Gilbert Strang. *Introduction to Linear Algebra"*. Wellesley-Cambridge Press, 2016.

[75] John A Swets, Robyn M Dawes, and John Monahan. Better decisions through science. *Scientific American*, 283(4):82–87, 2000.

[76] J. Prabhu Tej, Syed Raunaq Ahmed, A. R. Devi, and A. K. Rajagopal. Quantum Hypothesis Testing and State Discrimination. *arXiv preprint arXiv:1803.04944*, 2018.

[77] E. Torgersen. *Comparison of Statistical Experiments*, volume 36. Cambridge University Press, 1991.

[78] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory*. Wiley, New York, 1968.

[79] Graeme Weir. *Optimal Discrimination of Quantum States*. PhD thesis, University of Glasgow, 2018.

[80] Horace Yuen, Robert Kennedy, and Melvin Lax. Optimum Testing of Multiple Hypotheses in Quantum Detection Theory. *IEEE Transactions on Information Theory*, 21(2):125–134, 1975.

[81] Huangjun Zhu. *Quantum State Estimation and Symmetric Informationally Complete POMs*. PhD thesis, PhD thesis, National University of Singapore, 2012. 1, 5, 2012.

[82] Huangjun Zhu. Quantum State Estimation with Informationally Overcomplete Measurements. *Physical Review A*, 90(1):012115, 2014.

[83] Huangjun Zhu. Super-Symmetric Informationally Complete Measurements. *Annals of Physics*, 362:311–326, 2015.

[84] Mark H. Zweig and Gregory Campbell. Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine. *Clinical Chemistry*, 39(4):561–577, 1993.

Epilogue

The starting point that led me to this moment of completing my thesis took place at the beginning of the fall semester in 2015. I had enjoyed two signal processing classes in previous semesters and was eager to take their sequel, 6.341 – with *the* Professor Alan Oppenheim!! I was so excited that I remember being scared to even open the textbook, because in my mind that would mark the start of the class and the start of the possibility that I might screw up in a subject I loved so much. At some point during the semester an opportunity was offered to the students to participate in a project involving signal processing education and online courses. I was eager to be involved in any project involving signal processing, but was unsure if I should hold out and apply for UROP positions instead. After much deliberation, I decided to write an email to Professor Oppenheim asking for his advice. It was multiple paragraphs, written and rewritten over several iterations. He responded, "Catherine, I'd be glad to chat about all this if you'd like. -Al Oppenheim" I was floored.

It was pure luck that Al (he told me to call him Al!) had an opening for a UROP student regarding a project in collaboration with Professor Randall Davis. This led to a summer internship at Digital Cognition Technologies, Inc. in Waltham in the summer of 2016. In terms of my thesis, one of the most important things to come out of that internship was a growing familiarity with and curiousity about ROC curves. Al pointed out numerous times over the summer and going into the fall that it was interesting and peculiar how multiple ROCs are so often compared using the areas underneath them, i.e., their AUCs. Over coffee and bagels at the Au Bon Pain in Kendall Square, we talked about what the AUC really meant and did some research into the debate behind it. Eventually this question became both my introduction to

Al's style of research (solutions in search of problems) and my Masters of Engineering thesis.

I also started attending the DSPG meetings that took place every Thursday from 5-6:30 PM in the Jackson Room. They were originally the group meetings of DSPG, but as I would come to learn they had morphed into free-wheeling discussions between Al, a selection of graduate students from the department, and three to five of Al's former students who lived in the area. For over a year I was much too scared to attempt to contribute anything or even ask questions. Eventually I let my guard down and was able to enjoy the discussions to their fullest extent. The spirit of the discussions hinged on complete trust and respect between all participants. "Ridiculous" ideas and free associations were not just accepted but openly encouraged. Sometimes they wouldn't lead anywhere, but more often than not they would result in important connections or "threads" to tug on in a students' thesis. The people I got to know there – Dr. Petros Boufounos, Dr. Sefa Demirtas, Dr. Dan Dudgeon, Dr. James Ward – became role models for me and not just because of their impressive technical knowledge. Rather, I saw firsthand how genuine and kindhearted they were as people, how easily they all laughed at themselves when they made mistakes, and how eager they were to mentor younger students like me.

When it came time to start thinking about a topic for my doctoral thesis, I knew two things – that I loved linear algebra and I loved physics. The former had been further fueled by Petros Boufounos' class on Advanced Topics in Signal Processing. I thoroughly enjoyed the physics classes I took during my undergraduate years and wanted a reason to go back and study some of the topics more carefully. Unfortunately, I didn't know which ones, so Al advised me to look back at several previous DSPG theses that involved physics. These included Professor Yonina Eldar's thesis titled *Quantum Signal Processing* and Professor Andrew Singer's thesis titled *Signal Processing and Communication with Solitons*. Eventually we circled back to talking about hypothesis testing and how the quantum binary state discrimination problem was typically formulated. A key paper was *Receiver Operation Characteristics of Quantum State Discrimination* by Bodor and Koniorczyk, who we eventually reached

out to and had very pleasant and collegial email exchanges with. Classically we were used to thinking about the problem as starting from the conditional distributions of the score variable and ending with the design of the decision region. Quantum mechanically much of the discussion was about the design of the measurement operators. Since it was usually assumed that there were only two possible measurement outcomes, the decision region was essentially trivial with each outcome corresponding to a different final decision. It took quite a while for us to be able to square these two paradigms with each other. We eventually realized that each design problem was actually related to two separate stages of a general binary hypothesis testing problem. We decided to call these the pre-decision operator and the binary decision rule. Essentially our viewpoint became that the design of a quantum measurement was tantamount to designing the shape of the conditional distributions, and once those were fixed classical decision theory could take over to specify an optimal decision region.

At around the same time, we were writing a paper for ICASSP 2019 that summarized the main results of my M.Eng. thesis. The topic was ROCs and classical binary hypothesis testing. Dr. Jim Ward was among those who graciously read our drafts and gave us feedback. And it was in a meeting with him that the question of whether or not a concave SVT ROC was guaranteed to be identical to the LRT ROC of the same score variable. The moment I realized that the answer was yes, which became a main result of this thesis, came in the middle of a lecture for the Quantum Computation course that Al and I were taking together as listeners. It was the first time that I realized that an important technical results might reveal itself at a time when my mind was relaxed and thinking about something completely unrelated.

The Quantum Computation course was also significant because it gave me a more solid foundation in topics surrounding quantum engineering and because of the final project I completed with the guidance of Al and Professor Ike Chuang, who at this point was officially on my committee along with Dr. James Ward. The project topic was the quantum Fourier transform or QFT. Al, Ike, and I discussed how the quantum circuit diagram typically associated with the QFT corresponded directly to the fast

Fourier transform in signal processing, or FFT. The FFT and associated butterfly flow graph is one way of computing the discrete Fourier transform of a signal, but there are many others. Our objective was to explore other quantum circuits that corresponded to other signal flow graphs. We ultimately focused on the Singleton algorithm, and the resulting study was an excellent opportunity for me to practice manipulating basic quantum circuit elements and comparing them to concepts I was already familiar with in signal processing.

The following year in the spring of 2019, Al and I went to ICASSP in Brighton in the UK. It was my first conference and an absolutely wonderful experience. Thomas and I flew to England together and spent a few days in London before the conference started. We then met up with Al in Brighton. Al and I attended a workshop together and afterwards discussed our reactions to the style of the presentation and whether, if we were hypothetically giving a similar presentation, we would want the slides to be standalone with a lot of information or more sparse and intended to be viewed only with the talk. Over the course of the conference we also talked about the different keynote speakers and how they each grabbed the audience and kept our attention using different techniques. Some of the happiest parts of the conference were when I met several DSPG alumni whose names I had heard many times – Professor John Buck and Professor Andrew Singer among them. They assured me that I was always welcome to reach out to them for advice, which I have. I gave my first poster presentation and John Buck, who I had only met the night before, came by for the sole purpose of slipping me a bottle of water.

The next important milestone for me happened later that year, when I attended the 2019 Allerton conference in Monticello, Illinois. Our paper was about quantum binary hypothesis testing and specifically about two different types of operating characteristics which we referred to as QDOCs and QMOCs. The Allerton conference was completely different than ICASSP in both scope and style. Whereas ICASSP had thousands of attendees, tens of booths dedicated to various organizations, and numerous lecture halls, Allerton was held in a remote location with closer to two hundred attendees who all spent the day in a single medium sized building with about

five attached seminar rooms. I met and made friends with fellow graduate students, felt secretly proud but also self-conscious when I recognized some of my professors, and gave my first oral presentation at a conference.

In the time period after the Allerton conference, we came across an important paper that would end up being the source of a large part of our viewpoint about quantum binary state discrimination and informationally complete and overcomplete POVMs. This was Scott's paper titled *Tight Informationally Complete POVMs*. It took several weeks to "decode" the elegant mathematical analysis laid out by Scott. It was well worth it. Doing this gave me an understanding of frames for operator spaces and the importance of decomposing the operator space $\mathcal{V}$ into the span of the identity and its orthogonal complement. It also helped me understand how IC and IOC POVMs are used in quantum state estimation including how the dual frame of an operator-valued frame might be difficult to compute. Finally, it introduced me to generalizations of the Bloch sphere in higher dimensions.

Al and I spent many hours at the whiteboard gaining a common understanding of the postulates of quantum mechanics, the word "state", frames, operator spaces, frames of operator spaces, the word "measurement", projections, POVMs, etc. Every single bit of intellectual content in this thesis was the result of these conversations, in addition to valuable discussions with many other collaborators. These discussions are something I'm really going to miss. We became very close as a result of spending so many hours together, challenging each other without any ego involved, and celebrating our most intriguing insights. Sometimes I was the teacher, sometimes Al was the teacher, and sometimes we were both equally confused.

Once we had more or less settled on a viewpoint for informationally complete and overcomplete POVMs, we set about coming up with ideas about how to demonstrate that overcompleteness could potentially benefit the performance of quantum binary state discrimination system. This lead us to an exploration of POVMs constructed using Platonic solids. We produced simulations showing that a scenario could be constructed where increasing M could lead to superior ROCs. At this point Al pointed out that Platonic solids with more vertices (higher M) were closer approximations

of a sphere than those with fewer vertices. He then asked what might happen if we looked at POVMs constructed from geodesic domes. Would that allow us to observe the trend for a larger range of M or lead to other new insights? This led to a series of very interesting discussions between the two of us along with Andy Ding, a new student and friend who officially joined DSPG in the spring of 2021, although we had known him for much longer. It was eventually Andy who did some initial research into the optimal way of evenly distributing M points on a sphere. Among other insights, he pointed out that to us that the Riesz s-energy was a commonly used criterion.

Another topic we worked on for a short time was the application of first-order noise shaping techniques as described by Boufounos and Oppenheim in *Quantization Noise Shaping on Arbitrary Frame Expansions* to quantum binary state discrimination. This paper helped me understand first-order sigma-delta quantization, which I had originally been exposed to in 6.341, in a completely new light that was rooted in linear algebra. We realized that the methodologies explained in that paper could be applied to the problem of expressing a density operator as accurately as possible as a linear combination of Hermitian operators, with the constraint that the coefficients had to be of the form n/L for some integer $0 \leq n \leq L$. More specifically, the probabilities that represented the true coefficients had to be replaced by a series of quantized counterparts. The problem that Al pointed out is that the sigma-delta technique requires knowledge of the true coefficients (the probabilities) in order to compute the error introduced by quantization. And in quantum binary state discrimination and quantum state, these are of course not available. If I had more time or could write another thesis, I would love to work on this area more.

Up until a few months ago Al and I had never talked about POVMs whose elements all had equal trace and were all rank one, and we certainly would have been confused at the idea of a sphere other than the Bloch sphere. This was even the case while much of the work that Andy was involved in with optimal arrangements of points on a sphere was being done. Al had told me several times over the course of my studies that key insights often came at the end of a student's program, frequently after they were well into the writing of the thesis. So it wasn't surprising when just

a few months ago, during a discussion where we were trying to find a clear way to describe how POVMs were built from points on the Bloch sphere, that Al pointed out that in fact, we weren't talking about the Bloch sphere at all. This obviously ended up becoming an important concept in the thesis. Funnily enough, it took us many tries to come up with a name to call the "other" spheres that we were talking about.

One of the last major pieces of my graduate career that was fundamental to the development of my thesis is the monograph that Al and I wrote together for Foundations and Trends in Signal Processing. This opportunity was graciously presented to us in the early months of 2019 by Professor Yonina Eldar. At the time, we were only just starting to delve into issues of quantum binary state discrimination and the intent was for the article to mainly be a tutorial on operating characteristics for classical binary hypothesis testing. We also planned on including some non-tutorial results such as the optimality criterion for SVT ROCs presented in Chapter 2 of this thesis. As it turned out, we would work on the Foundations and Trends monograph for over two years and it would become an extremely useful way for me to metabolize and deeply understand many concepts related to this thesis. It also provided motivation at crucial points in time where I might otherwise have lost momentum. Al pointed out that it is more typical for a student to write the thesis first before moving on to the corresponding major publication or report containing the main results. By chance, it happened for this thesis in the opposite order. Retrospectively I'm extremely glad that this was the case because it made writing the thesis less daunting. This was a result of the fact that Al and I had had many debates about how to frame or present most if not all of the main topics.

I always assumed that finishing my thesis would feel like tying a big shiny bow on top of a present – joyful, gratifying, and complete. The first two of these turned out to be true. I feel joyful because writing the thesis and especially this epilogue reminded me of all the reasons I loved this subject material to begin with. And it's gratifying to remember how complicated the journey was to get to the final picture. But writing the thesis also reminded me of all of the exciting little pet projects that I never had time to finish, and all of the ideas on the shelf that have come up at one point or another

over the years and that we thought we might look into. As frustrating as it is, I'm happy and somewhat relieved that the research this thesis represents is not a "dead end", at least not with regards to my own personal interests. I see now that I could happily continue this work for several more years. The picture that comes to mind is as follows: If the scope of this thesis were a vector space, then each individual sub-topic would represent a basis vector. The overlapping perspectives, viewpoints, and interpretations given to these topics would be linear combinations of those vectors. And all together the whole picture would not be complete, but overcomplete. These layers of redundancy add depth and enrichment, primarily for my own education if nothing else. With this in mind the six words that summarize my thesis experience are: A finished thesis is always overcomplete.