Foundations and Trends® in Signal Processing

Operating Characteristics for Classical and Quantum Binary Hypothesis Testing

Suggested Citation: Catherine A. Medlock and Alan V. Oppenheim (2021), "Operating Characteristics for Classical and Quantum Binary Hypothesis Testing", Foundations and Trends® in Signal Processing: Vol. 15, No. 1, pp 1–120. DOI: 10.1561/2000000106.

Catherine A. Medlock
Massachusetts Institute of Technology
USA
cmedlock@mit.edu

Alan V. Oppenheim
Massachusetts Institute of Technology
USA
avo@mit.edu

now
the essence of knowledge
Boston — Delft

Contents

Operating Characteristics for Classical and Quantum Binary Hypothesis Testing

Catherine A. Medlock[1] and Alan V. Oppenheim[2]

[1] *Massachusetts Institute of Technology; cmedlock@mit.edu*
[2] *Massachusetts Institute of Technology; avo@mit.edu*

ABSTRACT

This monograph addresses operating characteristics for binary hypothesis testing in both classical and quantum settings and overcomplete quantum measurements for quantum binary state discrimination. We specifically explore decision and measurement operating characteristics defined as the tradeoff between probability of detection and probability of false alarm as parameters of the pre-decision operator and the binary decision rule are varied. In the classical case we consider in detail the Neyman-Pearson optimality of the operating characteristics when they are generated using threshold tests on a scalar score variable rather than threshold tests on the likelihood ratio. In the quantum setting, informationally overcomplete POVMs are explored to provide robust quantum binary state discrimination. We focus on equal trace rank one POVMs which can be specified by arrangements of points on a sphere that we refer to as an Etro sphere.

Catherine A. Medlock and Alan V. Oppenheim (2021), "Operating Characteristics for Classical and Quantum Binary Hypothesis Testing", Foundations and Trends® in Signal Processing: Vol. 15, No. 1, pp 1–120. DOI: 10.1561/2000000106.

Nomenclature

ROC	Receiver operating characteristic	3
POVM	Positive operator-valued measure	4
IC	Informationally complete	5
IOC	Informationally overcomplete	5
Etro	Equal trace rank one	5
LRT	Likelihood ratio test	7
SVT	Score variable threshold test	7
CDOC	Classical decision operating characteristic	7
CMOC	Classical measurement operating characteristic	7
MPE	Minimum probability of error	11
MAP	Maximum a posteriori	11
AUC	Area under an ROC	16
CDF	Cumulative distribution function	18

QMS	Quantum mechanical system	28
QDOC	Quantum decision operating characteristic	29
QMOC	Quantum measurement operating characteristic	29
ENTF	Equal norm tight frame	63
Etro	Equal trace rank one	65

Preamble

Our intention while preparing this monograph has been for it to be readable and interesting to an audience with a wide range of backgrounds. It is written to have a strong tutorial review flavor with some perspectives that hopefully many readers will find to be interesting and somewhat novel. We anticipate that many parts of the monograph will be familiar to readers with a strong background in classical signal processing and other parts to readers with a strong background in quantum mechanics. And it is our hope that both audiences will find the perspectives on the shared issues and overlap between the two fields to be interesting. Some readers may find it helpful in acquiring a broad sense of the scope of the monograph to start by reading the summary remarks and further thoughts given in Section 7. It should be noted however, that the discussion there uses terminology and notation introduced in earlier sections. In writing a monograph intended for an audience with diverse backgrounds part of the challenge is that there are many results referred to in the presentation that will be well-known to readers with backgrounds in one of the disciplines but less so in the other. And with some of these results, we anticipate that some of the readers will want to see or be reminded of a somewhat detailed explanation while others will be very familiar with it. To accommodate these differences we identify these as exercises for the reader to be worked out or not as they choose. The details for verifying those results are contained in the appendix

denoted as Appendix A. A second appendix denoted as Appendix B contains the details of a variety of possibly less familiar results that would be included in a traditional appendix mainly for the purpose of not interrupting the flow of the main body.

1

Introduction

Binary decisions guide our everyday lives in situations both critical and trivial. The choices made by politicians and physicians may have consequential implications 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, the prolongation of life, the ability to keep dry in a downpour, may come at the cost of soldiers' lives, the quality of life of an individual patient, or the wasted effort of toting an umbrella on a rain-free 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, discussed further in Section 2.

In this monograph we focus on a particular set of well-studied metrics for framing the problem of binary hypothesis testing, keeping in mind that there are many alternatives, generalizations, and extensions of the viewpoints and results expressed here. We specifically consider the scenario in which one of two possible hypotheses, denoted as H_0 or H_1, is true. The objective is to make a decision as to which is true using a sample value of a random variable often referred to as the score variable, which is comprised of one or more numerical values associated with the outcome of some measurement or observation. The score variable may be a scalar or a vector and may have been constructed as a composition of multiple measurements and observations. Traditionally H_0 is referred to as the null hypothesis and H_1 as the positive hypothesis, implying that H_1 is the hypothesis of significance (the target is present, the patient has the disease, etc.). In this monograph we use that convention. For convenience we refer to the entire system used to distinguish between the null and positive hypotheses as the discrimination system. The components of the discrimination system are defined in Section 2. Historically a quantity considered to be of significance in binary hypothesis testing is the probability of error, denoted as P_e and defined as the probability of identifying H_0 to be true given that 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 probabilities associated with whether H_0 or H_1 is true apart from any measurement or decision. Various of these probabilities are connected mathematically through the rules of probability. For example, the probability of error can be expressed as a combination of the probability of detection, the probability of false alarm, and the underlying a priori probabilities.

Since in many scenarios the a priori 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 (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 [1]. The use of ROCs has become increasingly prevalent in a very broad set of application areas including biostatistics and machine learning [2]–[8]. In contrast to the problem of radar signal detection for which there are often good mathematical models for the signals and disturbances, in other contexts the score variable is typically a finely-tuned combination of many measurements and is therefore often less amenable to mathematical analysis and modeling.

More generally, the term operating characteristic is used in this monograph to refer to any characterization, such as a curve, table, or graph, of the tradeoff between P_f and P_d as one or more parameters of the discrimination system is varied. When displaying operating characteristics we will choose to utilize a two-dimensional graph of P_f versus P_d. Consequently the parameter or parameters being varied are not immediately visible or explicit. This is especially important in Sections 2.5.4 and 2.5.5 when we consider multiple operating characteristics that were generated using variations of distinct parameters but have identical graphs of P_f versus P_d. We take the viewpoint that an operating characteristic itself is essentially a trajectory in a higher-dimensional space with coordinates corresponding to all of the parameters being varied in addition to P_f and P_d. A graph of P_f versus P_d is the projection of this trajectory onto the P_f-P_d plane. Distinct trajectories including those with different numbers of variable parameters may correspond to the same P_f-P_d projection. For the majority of our discussion we will be concerned only with the characteristics of the P_f-P_d projection of a given operating characteristic. Thus, for the sake brevity we will only explicitly distinguish between an operating characteristic and its P_f-P_d projection when absolutely necessary, as in Sections 2.5.4 and 2.5.5.

Sections 2 and 3 of this monograph address operating characteristics associated with binary hypothesis testing in the classical setting and the setting of quantum mechanics, respectively. By "classical" we mean in

particular that the measurement or observation processes that lead to a realized value of the score variable are not constrained by the postulates of quantum mechanics. The principles of classical binary hypothesis testing are very well-understood and as outlined above, ROCs are widely used in many classical settings. The principles of quantum binary state discrimination are also well-formulated. As we discuss in Section 3, a typical formulation of the quantum binary state discrimination problem consists of a quantum mechanical system that has been prepared in one of two quantum states by two distinct laboratory procedures or physical environments, each corresponding to one of the two hypotheses H_0 or H_1. The objective is to decide which procedure was used based on the outcome of a measurement on the system. An elegant solution to the problem of determining the measurement strategy that achieves minimum probability of error was derived by Helstrom [9].

Just as in the classical setting, the above formulation of the quantum binary hypothesis testing problem naturally involves a tradeoff between P_f and P_d and therefore it also involves the notion of an operating characteristic. But operating characteristics of any kind are significantly less prevalent in the quantum binary hypothesis testing 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 Section 3 are the stipulations made by the postulates of quantum mechanics about the state of a quantum system and about the concept of quantum measurement. Of particular importance is the relationship between a specific quantum measurement and a set of Hermitian operators that form a positive operator-valued measure (POVM).

The theme of Sections 4 through 6 is how quantum measurements that employ redundant, or overcomplete, representations of the state of the system being measured can be used, at least in some cases, to increase the robustness of binary discrimination strategies. We start in Section 4 by describing our viewpoint on some of the basic concepts of frame theory, with the main objective of introducing the mathematical machinery and notation necessary to apply the concepts to quantum

measurement. We then describe how these concepts can be applied to an operator space consisting of all Hermitian operators on another Hilbert space. The relevant operator space $\mathcal{V}$ in quantum mechanics contains all density operators and POVM elements. This leads to a discussion in Section 5 regarding informationally complete (IC) quantum measurements, which are measurements that map every quantum state to a unique probability distribution over the possible measurement outcomes [10]–[22]. IC quantum measurements that are strictly overcomplete are sometimes referred to as informationally overcomplete (IOC) quantum measurements [19]. While the benefits of using IOC measurements have been investigated in the context of quantum state estimation [19], [20], less attention has been given to their utility in quantum binary state discrimination. We review a fundamental result stating that every IC or IOC POVM is a frame for $\mathcal{V}$. IOC POVMs with a larger number M of elements correspond to frame representations of $\mathcal{V}$ that are more overcomplete.

A crucial concept in our discussion of the operator space $\mathcal{V}$ is a specific direct-sum decomposition of $\mathcal{V}$ into two orthogonal subspaces $\mathcal{U}$ and $\mathcal{U}^{\perp}$. All density operators have a constant component in $\mathcal{U}^{\perp}$ and can be distinguished from each other by their components in $\mathcal{U}$. For the density operator of a qubitthe component in $\mathcal{U}$ corresponds to its Bloch vector. We define a counterpart to the Bloch ball and corresponding Bloch sphere in relation to the 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 $\sqrt{2}/M$. This is exactly analogous to how a pure state qubit density operator can be specified by a single point on the Bloch sphere. POVMs constructed using Platonic solids are Etro POVMs in our terminology and are used often in the literature. We provide evidence through simulation that when POVMs constructed from Platonic solids are used for qubit binary state discrimination, there is a tradeoff in probability of error between the number L of identically-prepared quantum mechanical systems and the number M of POVM elements. POVMs constructed from Platonic solids have been of particular interest in the quantum state estimation community because they are all either IC or IOC, and because they

all provide straightforward state reconstruction formulas. Since we are interested in state discrimination rather than estimation, we do not require the state to be reconstructed. Consequently in Section 6 we also performed an exploratory investigation into IC and IOC POVMs constructed using other arrangements of points on an Etro sphere. In particular the problem we consider is that of distinguishing between two pure state qubit density operators. It is assumed that the angle between their Bloch vectors is known but that the overall alignment of the Bloch vectors relative to the Bloch sphere is not. Equivalently, it is assumed that the two Bloch vectors are known and the relative rotational orientation of the Bloch and Etro spheres is unknown. We compare the performance of a variety of POVMs using their minimum and maximum probabilities of error over all possible orientations, as well as the difference between the two. Intuitively it is expected that higher values of M and distributions of points on an Etro sphere that are maximally spread in some sense would lead to POVMs that are less sensitive to changes in the relative orientation of the Bloch and Etro spheres. Indeed, this is what we observed for values of M between 4 and 12 and for distributions of points that were maximally spread with respect to numerous established criteria.

2

Operating Characteristics for Classical Binary Hypothesis Testing

After first defining a simple framework that encompasses general binary hypothesis testing problems in Section 2.1, we review known results regarding optimal binary decision strategies with respect to the minimum probability of error, minimum risk or Bayes' cost, and Neyman-Pearson criteria in Sections 2.2 to 2.4. 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 [23]. In Section 2.5 we compare ROCs generated using LRTs and SVTs and state a condition under which the P_f-P_d projection of an SVT ROC is guaranteed to be Neyman-Pearson optimal and therefore identical to the P_f-P_d projection of the LRT ROC of the same underlying score variable. We also describe a procedure that can be used to recover the optimal ROC from a non-optimal SVT ROC. ROCs can also be classified as classical decision operating characteristics (CDOCs) in our terminology. Finally in Section 2.6 we describe a different type of classical operating characteristic that we refer to as a classical measurement operating characteristic (CMOC). The quantum analogue of a CMOC is defined in Section 3.6.

2.1 Framework for Binary Hypothesis Testing

The framework that we consider in this monograph for binary hypothesis testing is shown in Figure 2.1. We start by defining a random symbol H that is equal to H_0 if the null hypothesis is true and H_1 if the positive hypothesis is true. The input to the system in Figure 2.1 can have a variety of forms as detailed in Examples 2.1 to 2.3, but in all cases the state or value of the input is dependent on the true hypothesis. The objective is to make a binary decision about whether the null or positive hypothesis is true in an optimal way with respect to some optimality criterion. The final decision is represented by a second random symbol $\hat{H}$ that is set to H_1 if we decide that the positive hypothesis is true and H_0 otherwise. An error is made when $\hat{H}$ and H are different. The most general binary hypothesis testing problem might belong to one of a number of more specific problem types. For example, in classical signal processing the objective of a typical binary detection problem is to determine whether an incoming waveform consists only of noise or of noise added to a pre-determined signal. The objective of a typical binary discrimination or classification problem is to determine which out of a pre-determined alphabet of signals an incoming waveform represents. We emphasize that the discussion and results presented in Section 2 pertain to general binary hypothesis testing problems and not to one specific subcategory.

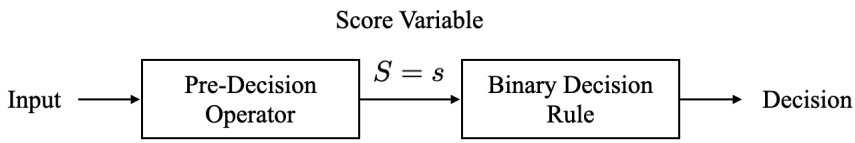


Figure 2.1: Framework for binary hypothesis testing.

The a priori probabilities – or priors, for short – that each hypothesis is true will be denoted as

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

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 2.5, the absence of the use of priors in the formulation of ROCs 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 which itself is sometimes referred to as the score variable. We will denote the score variable 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 distributions, also referred to as likelihood functions, of the score variable given that $H = H_0$ or given that $H = H_1$ will be denoted as $f_i(s)$ for $i \in \{0, 1\}$. The $\{f_i(\cdot)\}$ may be probability density functions (PDFs) or probability mass functions (PMFs) depending on whether S is continuous or discrete, respectively. When the sample value s satisfies $f_0(s) > 0$ and $f_1(s) > 0$, we are unable to identify the true hypothesis with certainty. The decision about the true hypothesis can therefore be thought of as a decision about whether s was drawn from $f_0(\cdot)$ or $f_1(\cdot)$. This problem has been studied extensively in the field of classical decision theory. In that context each decision-making strategy is typically described using a decision region $\mathcal{D}$ that is a subset of the sample space of S. If the sample value s lies in $\mathcal{D}$ then we declare $\hat{H} = H_1$. Otherwise we declare $\hat{H} = H_0$.[1] Of course, there are many different possible decision strategies or equivalently many different ways of choosing $\mathcal{D}$. Two of the most common strategies are LRTs and SVTs. Both will be described in more detail in subsequent sections, but first we describe two examples of classical binary hypothesis testing systems using the terminology of Figure 2.1.

Example 2.1. In a typical radar signal detection problem the input to the pre-decision operator is a waveform, possibly reflected by a target, received by the radar system following the emission of an electromagnetic pulse by the radar transmitter. The pre-decision operator might be a

[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 monograph.

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 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 priors. This choice of decision region corresponds to what we refer to as an SVT.

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. This specific example is elaborated on further in Section 2.5.3.

It will be useful in future sections to write the probabilities of false alarm and detection 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 declare $\hat{H} = H_1$ given that $H = H_0$. The probability of detection is the conditional probability that we declare $\hat{H} = H_1$ given that $H = H_1$. P_f and P_d can equivalently be thought of as the conditional probabilities that s lies in $\mathcal{D}$ given that $H = H_0$ or $H = H_1$, respectively. We have

$$P_f = P(\hat{H} = H_1 | H = H_0) = \int_{s \in \mathcal{D}} ds\, f_0(s) \tag{2.2a}$$

$$P_d = P(\hat{H} = H_1 | H = H_1) = \int_{s \in \mathcal{D}} ds\, f_1(s). \tag{2.2b}$$

Note that in Equations (2.2) and throughout Section 2, we have arbitrarily assumed that S is continuous and have thus used an integral instead of a sum to calculate probability values. Nevertheless, the results of this section can be appropriately modified for the discrete case. In Section 3 we will find it more natural to assume that S is discrete, but again the results can be appropriately modified to apply to the continuous case.

2.2 Minimum Probability of Error Decision Rules

One of the most common optimality criteria used in binary decision making is the minimum probability of error (MPE) criterion. Developing the optimal decision making strategy with respect to this or any other optimality criterion amounts to finding the corresponding optimal decision region $\mathcal{D}$. 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,

$$P_e = \int ds\, f_S(s)\, P_{e|s}. \tag{2.3}$$

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, implying that in order 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 $\hat{H} = H_1$, implying that an error is made when $H = H_0$ and s lies in $\mathcal{D}$. The reverse is true for when s does not lie in $\mathcal{D}$. $P_{e|s}$ can be written as

$$P_{e|s} = \begin{cases} P(H = H_0|S = s) \text{ if } s \in \mathcal{D} \\ P(H = H_1|S = s) \text{ if } s \notin \mathcal{D}. \end{cases} \tag{2.4}$$

The conditional probabilities $P(H = H_i|S = s)$ are typically referred to as the a posteriori probabilities that $H = H_i$ given that $S = s$. To minimize $P_{e|s}$ we should choose the hypothesis that has the maximum a posteriori (MAP) probability conditioned on the observation $S = s$. Thus the optimal decision rule with respect to the MPE criterion is

$$\hat{H} = \begin{cases} H_1 \text{ if } P(H = H_1|S = s) \geq P(H = H_0|S = s) \\ H_0 \text{ if } P(H = H_1|S = s) < P(H = H_0|S = s). \end{cases} \tag{2.5}$$

Note that the values of 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

$$P(H = H_i|S = s) = \frac{P(S = s|H = H_i)P(H = H_i)}{f_S(s)} = \frac{f_i(s)\, q_i}{f_S(s)}. \quad (2.6)$$

Cancelling the factors of $f_S(s)$ on both sides of the inequalities in Equation (2.5) and rearranging leads to the following equivalent decision strategy,

$$\hat{H} = \begin{cases} H_1 \text{ if } \dfrac{f_1(s)}{f_0(s)} \geq \dfrac{q_0}{q_1} \\ H_0 \text{ if } \dfrac{f_1(s)}{f_0(s)} < \dfrac{q_0}{q_1}. \end{cases} \quad (2.7)$$

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

$$\mathcal{D}_{\text{MPE}} = \{s : f_1(s)/f_0(s) \geq q_0/q_1\}. \quad (2.8)$$

As we summarize below, the optimal decision regions for the minimum risk and Neyman-Pearson criteria have a very similar form.

2.3 Minimum Risk Decision Rules

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 risk or Bayes' cost and denoted as R, and the corresponding optimal decision rule is the one that minimizes R. We denote by c_{ij} the cost of declaring $\hat{H} = H_i$ when in truth $H = H_j$. The probability of error corresponds to the special case where $c_{01} = c_{10} = 1$ and $c_{00} = c_{11} = 0$. Similar to the integral in Equation (2.3), the risk can be expressed as

$$R = \int ds\, f_S(s)\, R_s, \quad (2.9)$$

where R_s is the expected conditional cost incurred given that $S = s$. Again since $f_S(s)$ is non-negative for all values of s, to minimize R it is sufficient to minimize R_s for every s. As in Equation (2.4) we consider separately the cases where s lies in $\mathcal{D}$ and where it does not, leading to

$$R_s = \begin{cases} c_{10}\, P(H = H_0|S = s) + c_{11}\, P(H = H_1|S = s) \text{ if } s \in \mathcal{D} \\ c_{00}\, P(H = H_0|S = s) + c_{01}\, P(H = H_1|S = s) \text{ if } s \notin \mathcal{D}. \end{cases} \tag{2.10}$$

The top line above corresponds to the expected cost incurred when we declare $\hat{H} = H_1$, while the bottom line is the same quantity when we declare $\hat{H} = H_0$. To minimize R_s we choose the decision with the smaller expected cost. A parallel analysis to the one in Section 2.2 using Bayes' rule applied to the a posteriori probabilities leads to the optimal decision rule with respect to the minimum risk criterion,

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

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

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

2.4 Neyman-Pearson Optimal Decision Rules

While the MPE and minimum risk 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 risk 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 P_f or $(1 - P_d)$ (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. It can be shown analytically that 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 [24], [25]. In other words, the optimal Neyman-Pearson decision region $\mathcal{D}_{\text{NP}}$ is

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

Recall that the threshold value η_0 affects the value of P_f through the integral given in Equation (2.2a). An informal argument [25] that provides intuition as to why the above decision region is optimal with respect to the Neyman-Pearson criterion is included in Appendix A.1.

It is significant that the optimality criteria of MPE, minimum risk, and maximum P_d for a specified upper bound on P_f all lead to the family of likelihood ratio tests parameterized by an appropriate threshold value η. By definition, each member of the family has the form

$$\mathcal{D}_{\text{LRT}}(\eta) = \{s : f_1(s)/f_0(s) \geq \eta\} \quad (2.14)$$

for some real number $\eta \geq 0$. Explicitly, we have

$$\mathcal{D}_{\text{MPE}} = \mathcal{D}_{\text{LRT}}\left(\frac{q_0}{q_1}\right) \quad (2.15a)$$

$$\mathcal{D}_{\text{MR}} = \mathcal{D}_{\text{LRT}}\left(\frac{q_0(c_{10} - c_{00})}{q_1(c_{01} - c_{11})}\right) \quad (2.15b)$$

$$\mathcal{D}_{\text{NP}} = \mathcal{D}_{\text{LRT}}(\eta_0), \quad (2.15c)$$

where in Equation (2.15c) the threshold η_0 is chosen so that P_f is equal to its upper bound.

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 [24], [25] in which the filter is designed to compute (along with the sampler) 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 added to to a pre-determined signal $y[n]$ of length T,

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

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 T samples, $f_i(x[1], \ldots, x[T])$ for $i \in \{0, 1\}$, are also Gaussian and their ratio can be expressed as

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

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

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

The sum on the left-hand side of the inequality can be obtained as the output of 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 according to Equation (2.15) to be optimal with respect to minimum probability of error, minimum risk, or the Neyman-Pearson criterion. It is also straightforward to generalize this scenario to the case where T tends to infinity.

2.5 Receiver Operating Characteristics

When considering a given error criterion under a specific set of conditions – specific priors, for example – the primary goal is to find the single optimal decision rule with respect to that error criterion and those conditions. But it is often very useful to consider entire families of decision rules that are optimal under potentially different error criteria and for possibly different sets of conditions. ROCs are a useful tool

that allows us to accomplish exactly this. Referring back to Figure 2.1, ROCs are generated by fixing the pre-decision operator and recording the values of P_f and P_d that result from different decision regions of the binary decision rule. 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 operating characteristic introduced in Section 3 for the quantum case, we emphasize that they could equally well be referred to as classical decision operating characteristics or CDOCs.

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 ROCs. 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 (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 more details we refer the reader to [26], [27].

2.5.1 Preliminaries

Two simplifying assumptions used throughout the remainder of Section 2 are as follows. We will always assume that $f_0(\cdot)$ and $f_1(\cdot)$ are continuous, 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. The results presented in Sections 2.5.3 through 2.5.5 can be extended to more general score variables. However, the analysis is more complicated and does not lead to additional insight, so we do not address this more general case.

2.5.2 LRT ROCs and SVT ROCs

The LRT ROC associated with a given score variable may be obtained by recording the values of P_f and P_d corresponding to all possible LRT

thresholds. Each possible operating point on an LRT ROC is optimal with respect to the MPE criterion for some combination of priors, the minimum risk criterion for some combination of priors 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 priors, all possible relative costs, and all possible upper bounds on P_f.

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 [23]. Each member of the SVT family of decision regions has the form

$$\mathcal{D}_{\mathrm{SVT}}(\gamma) = \{s : s \geq \gamma\} \tag{2.19}$$

for some real number γ. Again, the SVT ROC associated with a given score variable may be obtained by recording all possible combinations of γ, 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 [28], [29]. 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 errors in the estimate of the ratio itself. It is in part for this reason that other decision strategies

[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.

besides LRTs, including SVTs as a particularly common choice, are used in many practical binary hypothesis testing situations.

It will be useful to introduce notation for the parametric formulas of the P_f-P_d projections of the LRT or SVT ROC of a given score variable. For an LRT ROC we define the functions $g_f(\cdot)$ and $g_d(\cdot)$ as

$$P_f^{\mathrm{LRT}} = g_f(\eta) = \int_{\mathcal{D}_{\mathrm{LRT}}(\eta)} ds\, f_0(s) \tag{2.20a}$$

$$P_d^{\mathrm{LRT}} = g_d(\eta) = \int_{\mathcal{D}_{\mathrm{LRT}}(\eta)} ds\, f_1(s). \tag{2.20b}$$

When $\eta = +\infty$, $\mathcal{D}_{\mathrm{LRT}}(\eta)$ is empty and $g_f(\eta) = g_d(\eta) = 0$. When $\eta = 0$, $\mathcal{D}_{\mathrm{LRT}}(\eta)$ contains the entire real line and $g_f(\eta) = g_d(\eta) = 1$. The functions $g_f(\cdot)$ and $g_d(\cdot)$ are always non-increasing in η. This follows from the fact that for two threshold values $\eta_0 \leq \eta_1$, the decision region $\mathcal{D}_{\mathrm{LRT}}(\eta_1)$ 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.

Similarly for the P_f-P_d projection of an SVT ROC we define the functions $h_f(\cdot)$ and $h_d(\cdot)$ as

$$P_f^{\mathrm{SVT}} = h_f(\gamma) = \int_{\mathcal{D}_{\mathrm{SVT}}(\gamma)} ds\, f_0(s) \tag{2.21a}$$

$$P_d^{\mathrm{SVT}} = h_d(\gamma) = \int_{\mathcal{D}_{\mathrm{SVT}}(\gamma)} ds\, f_1(s). \tag{2.21b}$$

Equation (2.21) can be simplified by defining $F_i(\cdot)$ to be the cumulative distribution function (CDF) of $f_i(\cdot)$,

$$F_i(u) = \int_{-\infty}^{u} ds\, f_i(s), \quad i \in \{0, 1\}, \tag{2.22}$$

for any real number u. Equation (2.21) can then be rewritten as

$$P_f^{\mathrm{SVT}} = h_f(\gamma) = 1 - F_0(\gamma) \tag{2.23a}$$

$$P_d^{\mathrm{SVT}} = h_d(\gamma) = 1 - F_1(\gamma). \tag{2.23b}$$

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 $\gamma_0 \leq \gamma_1$, $\mathcal{D}_{\text{SVT}}(\gamma_1)$ is always contained within $\mathcal{D}_{\text{SVT}}(\gamma_0)$. Under the current assumptions $h_f(\cdot)$ and $h_d(\cdot)$ are strictly decreasing and therefore invertible.

2.5.3 Properties of LRT and SVT ROCs

We briefly review well-known properties of the P_f-P_d projections 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 their P_f-P_d projections are identical and when, if they are not identical, one can be obtained from the other. The P_f-P_d projections of SVT ROCs and LRT ROCs are always monotonic. The P_f-P_d projections of LRT ROCs have the following additional properties.

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

$$\left.\frac{dP_d^{\text{LRT}}}{dP_f^{\text{LRT}}}\right|_{P_f^{\text{LRT}}=g_f(\eta_0)} = \frac{g_d'(\eta_0)}{g_f'(\eta_0)} = \eta_0, \tag{2.24}$$

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

- P_f-P_d projections of LRT ROCs are concave.

A derivation of Equation (2.24) can be found in many classical decision theory textbooks (see, for example, [24]) and relies mainly on a change of variables in the integrals in Equations (2.20) 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 monograph, so we omit them. The second property follows directly from Equation (2.24) 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. Under the current assumptions, LRT ROCs are necessarily strictly concave.

For a given score variable, we might be interested in whether or not the P_f-P_d projection of its SVT ROC is identical to the P_f-P_d projection of its LRT ROC. When this is the case, we can perform optimal MPE, minimum risk, 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 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, in general the likelihood ratio is not an invertible function of the score variable.

Since the P_f-P_d projections of 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 the P_f-P_d projection of an SVT ROC need be concave. An interesting question is whether or not, if the P_f-P_d projection of the SVT ROC of a given score variable *is* concave, it must be identical to the P_f-P_d projection of the LRT ROC of that score variable. The answer turns out to be surprisingly simple, relying only on a calculation of the slope of the P_f-P_d projection of an SVT ROC as a function of the SVT threshold, and is addressed in Section 2.5.4.

Example 2.5. Consider two conditional distributions $f_0(\cdot)$ and $f_1(\cdot)$ that are 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. The likelihood ratio function $\ell(\cdot) = f_1(\cdot)/f_0(\cdot)$ is strictly monotonic and therefore invertible. Therefore, the LRT and SVT ROCs are identical and we have $\mathcal{D}_{\text{LRT}}(\eta) = \mathcal{D}_{\text{SVT}}(\ell^{-1}(\eta))$ and $\mathcal{D}_{\text{SVT}}(\gamma) = \mathcal{D}_{\text{LRT}}(\ell(\gamma))$ for all η and all γ. The LRT ROC is shown

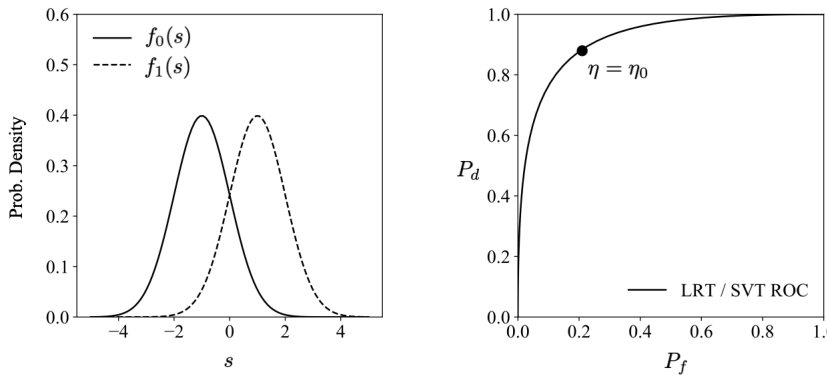


Figure 2.2: (a) Gaussian conditional distributions with variance $\sigma^2 = 1$ and mean $\mu_0 = -1$ or $\mu_1 = 1$. (b) LRT and SVT ROCs, which are identical for these conditional distributions.

in Figure 2.2b. Each point corresponds to a specific LRT threshold. The parametric formulas for the LRT ROC as a function of the LRT threshold are

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

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

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

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

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

Example 2.6. Consider two conditional distributions $f_0(\cdot)$ and $f_1(\cdot)$ that are Gaussian with zero mean but different variances denoted by σ_0^2 and σ_1^2, respectively. An example with $\sigma_0^2 = 0.25$ and $\sigma_1^2 = 2.25$ is shown in Figure 2.3a. 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

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

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

where $u \geq 0$ is the unique non-negative value that satisfies $\ell(u) = \eta$ and $\Phi(\cdot)$ is again 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 curve as a function of the SVT threshold are

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

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

2.5.4 Optimality of a Concave SVT ROC

A principal result presented in [23] is that if the P_f-P_d projection of an ROC that was generated using SVTs on a given score variable is concave, then it is guaranteed to be the P_f-P_d projection of the LRT ROC for that score variable. In other words, concavity is a *sufficient* condition for the Neyman-Pearson optimality of the P_f-P_d projection of the SVT ROC of a given score variable. To show that this is true, recall from Equation (2.23) that the SVT ROC of a given score variable can be written parametrically as

$$P_f^{\text{SVT}} = h_f(\gamma) = 1 - F_0(\gamma) \tag{2.29a}$$

$$P_d^{\text{SVT}} = h_d(\gamma) = 1 - F_1(\gamma). \tag{2.29b}$$

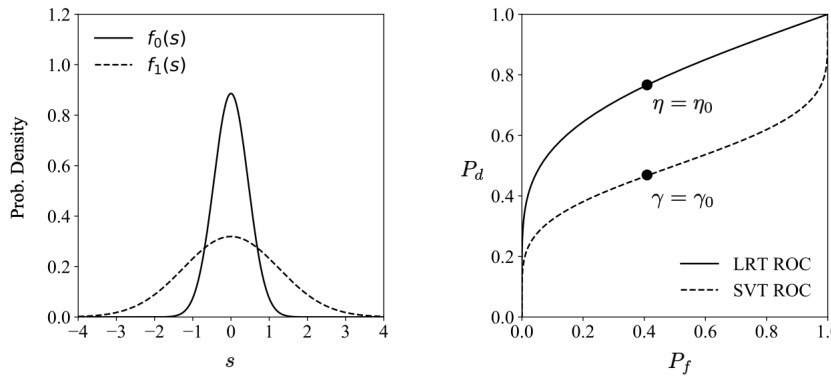


Figure 2.3: (a) Gaussian conditional distributions with mean $\mu = 0$ and variance $\sigma_0^2 = 0.45$ or $\sigma_1^2 = 1.25$. (b) LRT and SVT ROCs.

Let $h'_f(\cdot)$ and $h'_d(\cdot)$ denote the derivatives of $h_f(\cdot)$ and $h_d(\cdot)$, respectively. A key observation is that

$$h'_f(\gamma) = f_0(\gamma) \tag{2.30a}$$

$$h'_d(\gamma) = f_1(\gamma). \tag{2.30b}$$

And thus at the point on the curve corresponding to the SVT threshold γ_0, the derivative of the curve is

$$\left.\frac{dP_d^{\text{SVT}}}{dP_f^{\text{SVT}}}\right|_{P_f^{\text{SVT}}=h_f(\gamma_0)} = \frac{h'_d(\gamma_0)}{h'_f(\gamma_0)} = \frac{f_1(\gamma_0)}{f_0(\gamma_0)}. \tag{2.31}$$

If the P_f-P_d projection of 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. According to Equation (2.31), the slope of the curve as a function of γ is simply equal the likelihood ratio function. In summary, if the P_f-P_d projection of 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 P_f-P_d

projections of the SVT and LRT ROCs 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 P_f-P_d projection of 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 [30], 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 P_f-P_d projection of 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, it is essentially just an arbitrary continuous map from the interval $[0, 1]$ to itself). On the other hand, the result stemming from Equation (2.31) says that if the P_f-P_d projection of the given operating characteristic (i) was generated using SVTs on a *specific* pair of distributions associated with a given score variable and (ii) is strictly concave, then the ROC is optimal *for those distributions*.

The fact that the P_f-P_d projection of 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 the P_f-P_d projection of its SVT ROC is not concave. In this case the P_f-P_d projection of the SVT ROC is not Neyman-Pearson optimal. However, as we show next, 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.

2.5.5 Generation of the Optimal ROC from a Non-Concave SVT ROC

In this section we define 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 2.5.4. Consider first the scenario in

which 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 entire operating characteristic is known as opposed to just its P_f-P_d projection. A straightforward 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, as stated in Equation (2.30). 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 Figure 2.4a were designed specifically to generate distinctly different P_f-P_d projections of the 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 can be found in Appendix B.1.

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.
2. Add the segments together end-to-end to compute the location of the desired point on the LRT ROC. 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 as

$$P_f^{\text{LRT}} = g_f(\eta_0) = \sum_j \Delta P_f^{(j)} \tag{2.32a}$$

$$P_d^{\text{LRT}} = g_d(\eta_0) = \sum_j \Delta P_d^{(j)}. \tag{2.32b}$$

This procedure is fundamentally different than the use of randomization to replace a convex region on the P_f-P_d projection of an ROC by the straight line connecting its endpoints [31], [32]. In that case,

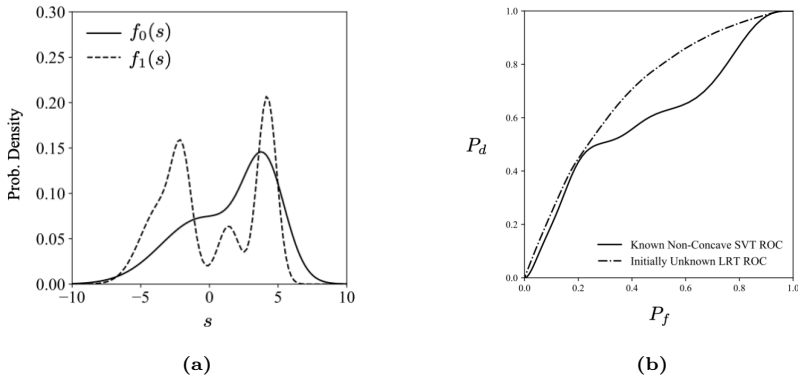


Figure 2.4: (a) Sample conditional PDFs $f_0(\cdot)$ and $f_1(\cdot)$ along with (b) the corresponding SVT and LRT ROCs. Assuming that the SVT ROC is known, the objective is to construct the LRT ROC.

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 operating characteristic is not Neyman-Pearson optimal. One way of seeing this is to observe that the P_f-P_d projection of the LRT ROC of a continuous score variable, even in the absence of the assumptions made in this monograph, can never have any linear regions – it must either be continuous and strictly concave or discontinuous 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 outlined 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 as a function of the SVT threshold 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.

2.6 Classical Measurement Operating Characteristics

It is not uncommon 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 recording the values of P_f and P_d obtained for each individual pre-decision operator. We refer to operating characteristics generated in this way as classical measurement operating characteristics or CMOCs. 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. Operating characteristics analogous to CMOCs generated in the quantum setting are discussed in Section 3.6.

3

Operating Characteristics for Quantum Binary State Discrimination

The objective of Section 3 is to describe a particular binary hypothesis testing problem in the quantum setting using the terminology and notation developed in Section 2. Of particular importance are the concepts of quantum decision and measurement operating characteristics which are analogous to ROCs and CMOCs, respectively. The topics of Section 3 were also described in [33]. Throughout the section, the input to the discrimination system in Figure 2.1, reproduced in Figure 3.1 for convenience, is assumed to be a quantum mechanical system. To avoid ambiguity between the discrimination system and the quantum mechanical system, from this point forward we will abbreviate the latter as the QMS. The null and positive hypotheses correspond to the QMS having been prepared by one of two possible laboratory procedures or physical environments, each of which corresponds to a distinct quantum state. Importantly, the way in which we can obtain information about the QMS through measurement is constrained through the postulates of quantum mechanics. The terminology that we use surrounding quantum measurement is discussed briefly in Section 3.1. The postulates of quantum mechanics that are relevant to this monograph are stated in Section 3.2. In Section 3.3 we use the postulates to describe in detail

one popular formulation of the quantum binary hypothesis testing problem. While in this monograph we only consider this formulation of the problem, we emphasize again that there are many alternate formulations, generalizations, and extensions. In Section 3.4 we review Helstrom's well-known result regarding minimum probability of error decision strategies for quantum binary hypothesis testing. Helstrom's result is the counterpart to the classical MPE decision rules reviewed in Section 2.2. In Sections 3.5 and 3.6 we describe two types of operating characteristics in analogy with classical ROCs (also referred to in this monograph as CDOCs) and CMOCs. We refer to them as either quantum decision operating characteristics (QDOCs) or quantum measurement operating characteristics (QMOCs) depending on the parameter that is varied to generate different values of P_f and P_d.

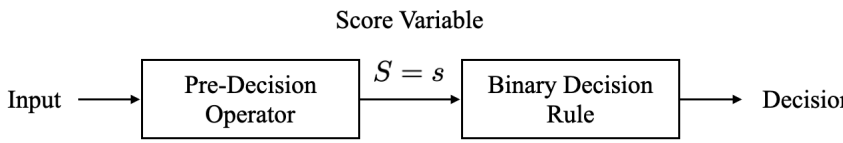


Figure 3.1: Framework for binary hypothesis testing.

3.1 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 [34],

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, which are summarized in Section 3.2. 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 monograph we borrow from the terminology in [35] 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.

3.2 The Postulates of Quantum Mechanics

While there are in total four postulates of quantum mechanics, in Section 3.2 we focus on the two postulates that relate specifically to this monograph. Both are loosely paraphrased from Chapter 2 of [35]. 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, $\langle v|$). 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 $|v\rangle = \sum_n c_n |u_n\rangle$. 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 $||v_1||\,||v_2||\cos\theta = \langle v_1|v_2\rangle$.

Quantum State Postulate. The state of an isolated physical system can be represented by a density operator ρ that is a linear operator 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

$$\rho = \sum_{j=1}^{D} a_j |\psi_j\rangle \langle\psi_j|, \tag{3.1}$$

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 are linear operators 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 monograph 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 $\mathcal{H}$,

$$\sum_{k=1}^{M} A_k^\dagger A_k = I, \tag{3.2}$$

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

$$p(k) = \sum_{j=1}^{D} a_j \langle\psi_j| A_k^\dagger A_k |\psi_j\rangle \tag{3.3}$$

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

$$p(k) = \mathrm{Tr}\left(A_k^\dagger A_k \rho\right). \tag{3.4}$$

A derivation of Equation (3.4) 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 dependent on the pre-measurement state ρ and the measurement operator A_k and is not relevant to this monograph.

A density operator in the form of Equation (3.1) represents a quantum system that has been prepared in the state $\rho_j = |\psi_j\rangle \langle\psi_j|$ with probability a_j [35]. 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 (3.1) implies that ρ is always a positive semidefinite Hermitian operator that has trace 1, since

$$\mathrm{Tr}(\rho) = \sum_{j=1}^{D} a_j \ \mathrm{Tr}\left(|\psi_j\rangle \langle\psi_j|\right) = \sum_{j=1}^{D} a_j = 1, \tag{3.5}$$

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

$$\rho = \sum_{j=1}^{d} \lambda_j \left|x_j\right\rangle \left\langle x_j\right|, \tag{3.6}$$

where the $\{\lambda_j\}$ are real and the $\{|x_j\rangle\}$ are orthogonal. It is straightforward to show that Equation (3.6) can in fact be considered as a special case of Equation (3.1) with $D = d$. Specifically when the $\{|x_j\rangle\}$ are normalized to have unit length, the $\{\lambda_j\}$ must be valid probabilities due to 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. As the discussion presented in this monograph does not depend on the state vectors and probabilities used to represent a given density operator, it will be convenient to specify density operators using the state vectors and probabilities corresponding to their eigenvectors and eigenvalues, respectively.

Regarding the quantum measurement postulate, Equation (3.4) can be derived from Equation (3.3) via

$$p(k) = \sum_{j=1}^{D} a_j \ \mathrm{Tr}\left(\left\langle\psi_j\right| A_k^\dagger A_k \left|\psi_j\right\rangle\right) \tag{3.7a}$$

$$= \mathrm{Tr}\left(\sum_{j=1}^{D} a_j \ \left\langle\psi_j\right| A_k^\dagger A_k \left|\psi_j\right\rangle\right) \tag{3.7b}$$

$$= \mathrm{Tr}\left(A_k^\dagger \, A_k \, \rho\right). \tag{3.7c}$$

In Equation (3.7) we have used both the fact that the trace of a scalar is itself and the cyclic property of the trace, $\mathrm{Tr}(AB) = \mathrm{Tr}(B\,A)$ for any two suitable linear operators A and B. Aside from notational convenience, we will see in Section 4 that the trace operator is also useful because it can be interpreted as an inner product function.

Throughout this monograph 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

$$E_k = E_k^\dagger \qquad \text{(Hermiticity)} \tag{3.8a}$$

$$\langle x| E_k |x\rangle \geq 0, \text{ for all } |x\rangle \in \mathcal{H} \qquad \text{(positive semidefiniteness)} \tag{3.8b}$$

$$\sum_{k=1}^{M} E_k = I \qquad \text{(completeness).} \tag{3.8c}$$

In functional analysis, a collection of operators satisfying these three properties is referred to as a positive operator-valued measure or POVM [36]. 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 (3.3) and (3.4) in terms of the operator E_k and the eigenvectors and eigenvalues of a given density operator ρ,

$$p(k) = \sum_{j=1}^{d} \lambda_j \langle x_j| E_k |x_j\rangle = \text{Tr}(E_k \rho). \tag{3.9}$$

Equation (3.9) is a crucial relation that is the basis for much of the discussion in Section 5. The main focus of Section 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 [10]–[22]. Explicitly, given two density operators ρ_1 and ρ_2 as well as an IC POVM $\{E_k\}$, let the corresponding probabilities be denoted by $\{p_1(k) = \text{Tr}(E_k \rho_1)\}$ and $\{p_2(k) = \text{Tr}(E_k \rho_2)\}$. We have $\{p_1(k) = p_2(k)\}$ if and only if $\rho_1 = \rho_2$. An important result regarding IC POVMs that is reviewed in Section 5 connects each IC POVM to an overcomplete representation of a 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 the POVM elements of a standard measurement are simply $\{E_k = A_k\}$. 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 3.1. 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 (3.8). When the measurement is made, the kth measurement outcome occurs with probability

$$p(k) = \mathrm{Tr}\,(E_k\,\rho) = |\,\langle v_k|\psi\rangle\,|^2. \tag{3.10}$$

Equation (3.10) states that the kth measurement outcome occurs with a probability equal to the squared magnitude of the component of $|\psi\rangle$ in the direction of $|v_k\rangle$. If $|\psi\rangle$ is orthogonal to $|v_k\rangle$ then the kth measurement outcome has zero probability of occurring.

3.3 Quantum Binary State Discrimination

Consider the scenario where the input to the discrimination system in Figure 3.1 is a QMS whose state can be represented by one of two known density operators depending on the true hypothesis, $\rho = \rho_i$ if $H = H_i$, for $i \in \{0, 1\}$. The two hypotheses may correspond, for example, to the QMS being subjected to two distinct laboratory procedures or to the QMS interacting with its environment in two distinct ways. As in Section 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

$$\rho_0 = \sum_{j=1}^{d} a_j\,|x_j\rangle \langle x_j|\,, \tag{3.11a}$$

$$\rho_1 = \sum_{j=1}^{d} b_j\,|y_j\rangle \langle y_j|\,, \tag{3.11b}$$

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. The pre-decision operator is assumed to consist of a quantum measurement $\{A_k, 1 \leq k \leq M\}$. The POVM elements will continue to be denoted by $\{E_k, 1 \leq k \leq M\}$. The score variable is equal to one of the index values $1 \leq k \leq M$ and the decision region $\mathcal{D}$ of the binary decision rule is some subset of $\{1, 2, \ldots, M\}$. For a given decision region $\mathcal{D}$, the conditional distributions of the score variable are

$$f_0(k) = \sum_{j=1}^{d} a_j \langle x_j | E_k | x_j \rangle = \mathrm{Tr}\left(E_k \rho_0\right), \quad 1 \leq k \leq d, \tag{3.12a}$$

$$f_1(k) = \sum_{j=1}^{d} b_j \langle y_j | E_k | y_j \rangle = \mathrm{Tr}\left(E_k \rho_1\right), \quad 1 \leq k \leq d. \tag{3.12b}$$

Then in analogy with Equation (2.2), the probabilities of false alarm and detection are

$$P_f = \sum_{k \in \mathcal{D}} f_0(k) = \sum_{k \in \mathcal{D}} \mathrm{Tr}\left(E_k \rho_0\right), \tag{3.13a}$$

$$P_d = \sum_{k \in \mathcal{D}} f_1(k) = \sum_{k \in \mathcal{D}} \mathrm{Tr}\left(E_k \rho_1\right). \tag{3.13b}$$

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 declaring 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 alternative formulations of quantum binary state discrimination in which this is not the case. In Section 3.5 and 5 we provide examples of operating characteristics generated using pre-decision operators with $M > 2$ outcomes.

3.4 Minimum Probability of Error Decision Rules

In analogy with the classical MPE decision rules described in Section 2.2, we summarize Helstrom's well-known result [9] regarding discrimination between two fixed density operators with minimum probability of error. We refer the reader to [9] for the complete derivation and a generalization of the result to the minimum risk error criterion. For the remainder of Section 3.4, 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 H_0 and if the measurement outcome is $s = 2$ then the final decision is H_1. The probability of error can be expressed as

$$P_e = q_0 \, P_f + q_1 \, (1 - P_d) = q_1 - q_1 \, \mathrm{Tr}\left[E_1 \left(\rho_1 - \frac{q_0}{q_1} \, \rho_0\right)\right]. \tag{3.14}$$

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

$$E_1 = \sum_{j:\lambda_j \geq 0} |z_j\rangle \, \langle z_j| \tag{3.15a}$$

$$E_2 = \sum_{j:\lambda_j < 0} |z_j\rangle \, \langle z_j| \, . \tag{3.15b}$$

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 H_1, otherwise the final decision is H_0. Equivalently, $\mathcal{D} = \{k : \lambda_k \geq 0\}$. Both of these POVMs have the property that the elements form complete sets of orthogonal projectors on $\mathcal{H}$, so they both correspond to standard quantum measurements.

3.5 Decision Operating Characteristics for Quantum Systems

An analogous performance curve 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 Example 3.2 we describe how the result presented in Section 2.5.4 can be applied to QDOCs.

Example 3.2. 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 3.2a. 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 (3.12) the conditional distributions of the score variable are

$$f_0(k) = \sum_{j=1}^{8} a_j \langle e_j|e_k\rangle \langle e_k|e_j\rangle = a_k, \quad 1 \leq k \leq 8, \tag{3.16a}$$

$$f_1(k) = \sum_{j=1}^{8} b_j \langle e_j|e_k\rangle \langle e_k|e_j\rangle = b_k, \quad 1 \leq k \leq 8, \tag{3.16b}$$

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 3.2b. Unlike an LRT decision region, an SVT decision region and

therefore an SVT QDOC inherently 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 3.2a 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 3.2b. 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 [24]. The constructive procedure described in Section 2.5.5 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 along with any POVM $\{E_k\}$.

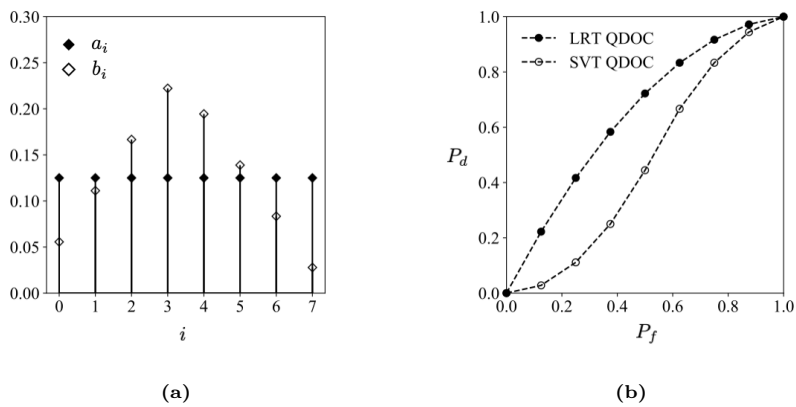


Figure 3.2: (a) Conditional distributions of the score variable as given in Equation (3.11). (b) QDOCs generated using LRT or SVT decision regions.

3.6 Measurement Operating Characteristics for Quantum Systems

An analogous operating characteristic to the CMOCs discussed in Section 2.6 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 [37] are QMOCs in our terminology. Examples 3.3 and 3.4 below below were motivated by the analysis and examples given in [37].

In Example 3.3 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 (which have $M = d = 2$). As noted in [37], the optimal operating points for all possible prior probabilities q_0 and q_1 lie on an ellipse. It is also pointed out in [37] 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. We demonstrate this in Example 3.4. We additionally demonstrate in Example 3.4, as is also shown in [37], 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 3.3. For this example we set $d = 2$, so $\mathcal{H}$ is isomorphic to $\mathbb{C}^2$. As in Equations (3.11) we denote the eigenvectors and eigenvalues of ρ_0 by $\{|x_j\rangle\}$ and $\{a_j\}$, respectively. We arbitrarily set $a_1 = 1/15$ and $a_2 = 14/15$ and

$$|x_1\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad |x_2\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{3.17}$$

where the implied basis is assumed orthonormal but otherwise arbitrary. The eigenvectors $\{|y_j\rangle\}$ and eigenvalues $\{b_j\}$ of ρ_1 are left as parameters to be varied. Without loss of generality the $\{|y_j\rangle\}$ can be assumed to have the form

$$|y_1\rangle = \begin{bmatrix} \cos(\alpha/2) \\ \sin(\alpha/2) \end{bmatrix}, \quad |y_2\rangle = \begin{bmatrix} -\sin(\alpha/2) \\ \cos(\alpha/2) \end{bmatrix} \tag{3.18}$$

for some angle α. The value of α represents (to within a constant factor) the angle of separation between the eigenvectors of ρ_0 and ρ_1. The pre-decision operator is assumed to be a standard measurement with associated POVM $\{E_1, E_2\}$. The decision region of the binary decision rule is $\mathcal{D} = \{2\}$. By definition E_1 and E_2 are orthogonal projectors onto a pair of one-dimensional subspaces spanned by orthogonal vectors. These vectors will be denoted as

$$|v_1\rangle = \begin{bmatrix} -\sin(\theta/2) \\ \cos(\theta/2) \end{bmatrix}, \quad |v_2\rangle = \begin{bmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{bmatrix} \tag{3.19}$$

for some angle θ. We have $E_1 = |v_1\rangle \langle v_1|$ and $E_2 = |v_2\rangle \langle v_2|$. A QMOC can be generated by fixing the values of α, b_1, and b_2 and varying the angle θ. It is straightforward to show that

$$P_f = \mathrm{Tr}(E_2 \rho_0) = a_1 \cos^2\left(\frac{\theta}{2}\right) + a_2 \sin^2\left(\frac{\theta}{2}\right) \tag{3.20a}$$

$$P_d = \mathrm{Tr}(E_2 \rho_1) = b_1 \cos^2\left(\frac{\theta - \alpha}{2}\right) + b_2 \sin^2\left(\frac{\theta - \alpha}{2}\right). \tag{3.20b}$$

In Appendix B.2 we show that Equations (3.20) correspond to the parametric formula for an ellipse. This was stated but not explicitly proven in [37]. Explicit formulas for the parameters of the ellipse in terms of the $\{a_j\}$, the $\{b_j\}$, and α are also given in Appendix B.2.

Figure 3.3 shows a collection of QMOCs each generated by fixing the values of α, b_1, and b_2 and varying the angle θ. In Figure 3.3a, 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 α. As α approaches 0 and the eigenvectors of ρ_0 and ρ_1 become more and more similar, the eccentricity of the ellipse increases. In Figure 3.3b, α is arbitrarily fixed to $\alpha = \pi/5$ 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 $P_d = 1/2$. It is straightforward 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 3.4. 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

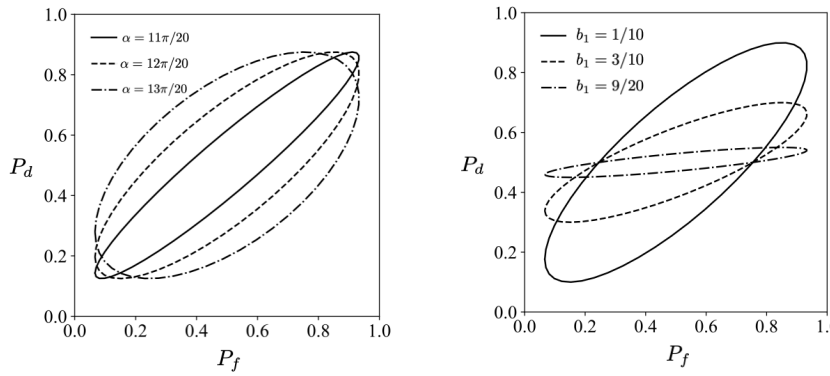


Figure 3.3: QMOCs generated with $d = 2$ for a fixed ρ_0 by varying the parameters of ρ_1 and the standard measurement that constitutes the pre-decison operator.

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 $\{|x_j\rangle\}$ and $\{|y_j\rangle\}$ of ρ_0 and ρ_1 are each set to an arbitrary orthonormal basis for $\mathcal{H}$. The $\{a_j\}$ and $\{b_j\}$ 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 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 (3.15). The MPE operating point then has coordinates

$$P_f = \mathrm{Tr}\left(E_2\,\rho_0\right) = \sum_{j=1}^{8} a_j \left\langle x_j\right| E_2 \left|x_j\right\rangle, \tag{3.21a}$$

$$P_d = \mathrm{Tr}\left(E_2\,\rho_1\right) = \sum_{j=1}^{8} b_j \left\langle y_j\right| E_2 \left|y_j\right\rangle. \tag{3.21b}$$

The result is the collection of upper operating points shown in Figure

3.4. 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 [37], 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 3.4 correspond to changes in the dimension of $\mathcal{U}_2$ (equivalently, the number of non-negative eigenvalues of $(\rho_1 - (q_1/q_0)\rho_0)$ or the rank of E_2). 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 3.3, 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. This is shown by the lower operating points in Figure 3.4. Each cluster corresponds to a different pair of dimensions for the orthogonal subspaces [37]. 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.

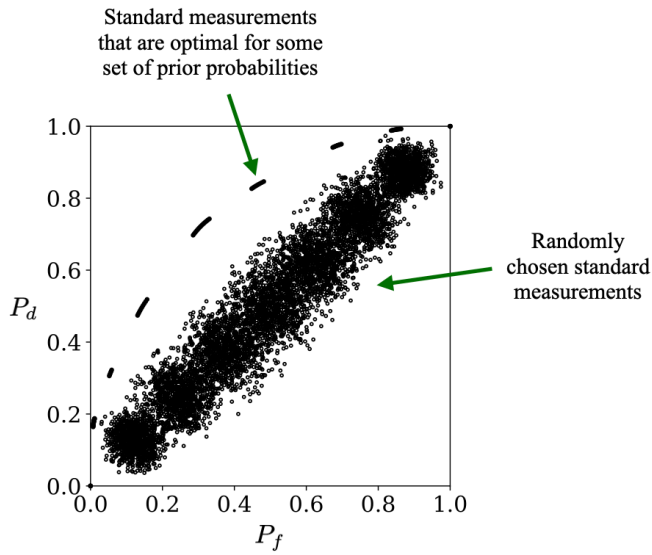


Figure 3.4: Operating points obtained by 2-outcome standard measurements performed on arbitrarily chosen density operators ρ_0 and ρ_1 with $d = 8$. *Upper segments of operating points:* Minimum probability of error operating points for a range of prior probabilities, $0 \leq q_1 \leq 1$. *Lower clusters 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

A Perspective on Frame Representations

The focus of Section 3 was the problem of discriminating between two quantum states using one measurement on a single system. The mathematical framework introduced in Section 4 sets the stage for a less restrictive version of the same problem in which we may perform measurements on many identical systems. This problem is addressed in Sections 5 and 6. One advantage offered by this increased generality is the opportunity to exploit IOC POVMs, which were defined in Section 3.2. IOC POVMs lead to overcomplete representations for density operators. The purpose of Section 4 is to review the key mathematical tools and perspectives that we will want to make use of as well as to establish our notation.

As is well known, a given vector in a finite- or infinite-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 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. 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 [38]–[42].

For the remainder of this monograph, we exploit the mathematics and elegance of frame representations in the Hilbert space characterization of quantum states, and in the Hilbert space representation of operators in the design of IOC POVMs. Specifically in Sections 4.1 to 4.4 we briefly review some core concepts of frame theory that are relevant to this monograph. For the most part, our main motivation is to introduce the mathematical framework necessary for the discussion of IOC POVMs and what we refer to as Etro POVMs in Sections 5 and 6. The many topics that are part of basic frame theory but that are not essential for our discussion of IOC POVMs are not included.

4.1 Preliminaries

Throughout Sections 4 to 6 we consider vectors that lie in a complex Hilbert space $\mathcal{V}$ that is a subspace of a larger space $\mathcal{W}$. For convenience we assume that $\mathcal{V}$ and $\mathcal{W}$ are finite-dimensional with $\mathcal{V}$ having dimension N and $\mathcal{W}$ having dimension M. Clearly $M \geq N$ since $\mathcal{V}$ is a subspace of $\mathcal{W}$. For a given linear operator T on $\mathcal{W}$, we will denote the range of T by $R(T)$ and the nullspace of T by $N(T)$. When a vector $|v\rangle$ is used as the input to a linear operator T on $\mathcal{W}$, the output will denoted interchangeably by $T|v\rangle$ or $|Tv\rangle$. As in Section 3, the superscript $\perp$ will denote an orthogonal complement and the superscript $\dagger$ will denote the Hermitian adjoint of a given linear operator. The orthogonal projection operator onto a subspace $\mathcal{R}$ of $\mathcal{W}$ or $\mathcal{V}$ will be denoted as $\mathcal{P}_{\mathcal{R}}$. An orthogonal projection operator is always Hermitian ($\mathcal{P}_{\mathcal{R}}^{\dagger} = \mathcal{P}_{\mathcal{R}}$) and idempotent ($\mathcal{P}_{\mathcal{R}}^{2} = \mathcal{P}_{\mathcal{R}}$).

Consider any set of vectors $\{|f_k\rangle, 1 \leq k \leq M\}$[1] that lie in and span $\mathcal{V}$ and 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 vectors $\{|f_k\rangle\}$ in $\mathcal{V}$ that satisfy

$$C\,||v||^2 \leq \sum_{k=1}^{M} |\,\langle f_k|v\rangle\,|^2 \leq D\,||v||^2 \tag{4.1}$$

for some $0 < C \leq D < \infty$ and for all $|v\rangle \in \mathcal{V}$ [39]. When C and D are set to form the tightest possible bounds, they are typically referred to as upper and lower frame bounds, respectively. The requirement that $C > 0$ ensures that the frame vectors span $\mathcal{V}$. Unlike in finite dimensions, in infinite dimensions Equation (4.1) is not necessarily satisfied by any set of vectors that lie in and span $\mathcal{V}$. Equation (4.1) can additionally be extended to include continuous frames and frames with a countably infinite number of elements. We will use the notation $\{|w_k\rangle, 1 \leq k \leq M\}$ to denote to an orthonormal basis for $\mathcal{W}$. The $\{|w_k\rangle\}$ are introduced specifically for the purpose of defining the analysis and synthesis maps of a frame in Section 4.2. The notation $\{|f_k\rangle, 1 \leq k \leq M\}$ will always be used to denote a frame for $\mathcal{V}$. As long as the $\{|w_k\rangle\}$ are a basis for $\mathcal{W}$, no generality is lost by assuming that they are orthonormal. If they were not, an inner product could always be constructed under which they were, along with an invertible function relating the original inner product to the new one (see Appendix A.2). 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 $\mathcal{V}$. However, it is of course possible to choose the $\{|w_k\rangle\}$ in such a way that a subset of them is an orthonormal basis for $\mathcal{V}$.

[1] The usage of the letter k to index different frame vectors $\{|f_k\rangle\}$ coincides with our choice of indexing for POVM elements $\{E_k\}$. This is intentional since in Section 5 we will eventually associate each POVM element E_k with a frame vector of an appropriately-defined vector space $\mathcal{V}$.

4.2 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 [38]. 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 $1 \leq k \leq M$.

It will be convenient for the purposes of this monograph to also define 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 will 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

$$|v\rangle \in \mathcal{V} \longrightarrow |w\rangle = A_0\,|v\rangle = \sum_{k=1}^{M} \langle f_k|v\rangle\,|w_k\rangle = \sum_{k=1}^{M} a_k\,|w_k\rangle \in \mathcal{W}. \quad (4.2)$$

It is further defined to satisfy $A_0\,|w\rangle = 0$ for all $|w\rangle \in \mathcal{V}^{\perp}$. For a given frame $\{|f_k\rangle\}$, the analysis map A_0 can always be expressed as

$$A_0 = \sum_{k=1}^{M} |w_k\rangle\,\langle f_k|\,. \quad (4.3)$$

Because the $\{|f_k\rangle\}$ span $\mathcal{V}$, A_0 has full rank and its range $R(A_0)$ is an N-dimensional subspace of $\mathcal{W}$. Its nullspace is $N(A_0) = \mathcal{V}^{\perp}$ has dimension $(M - N)$. The action of A_0 is summarized schematically in Figure 4.1. 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 $\mathcal{W} = R(A_0) \oplus R(A_0)^{\perp}$. Note that since the $\{|w_k\rangle\}$ are orthonormal we have $||A_0\,|v\rangle\,||^2 = \sum_k |a_k|^2$, so Equation 4.1 can be rewritten as

$$C\,||v||^2 \leq ||A_0\,|v\rangle\,||^2 \leq D\,||v||^2 \quad (4.4)$$

for some $0 < C \leq D < \infty$ and for all $|v\rangle \in \mathcal{V}$.

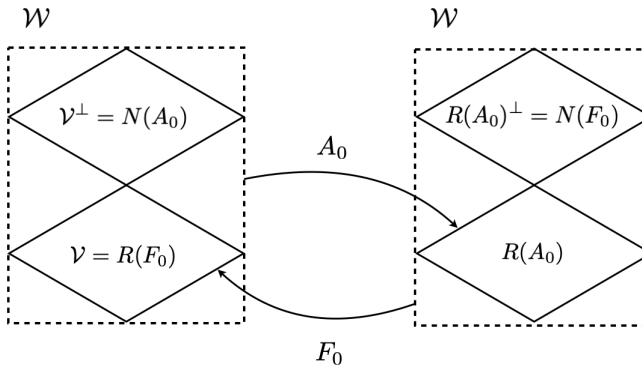


Figure 4.1: The analysis map A_0 takes vectors in $\mathcal{V}$ to a (possibly) different subspace of $\mathcal{W}$ with the same dimension as $\mathcal{V}$. It takes vectors in $\mathcal{V}^{\perp}$ to the zero vector. The synthesis map F_0 takes vectors in $R(A_0)$ to the subspace $\mathcal{V}$. It takes vectors in $R(A_0)^{\perp}$ to the zero vector.

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. Specifically, F_0 is defined by the relation

$$|w\rangle = \sum_{k=1}^{M} c_k |w_k\rangle \in \mathcal{W} \longrightarrow |v\rangle = F_0 |w\rangle = \sum_{k=1}^{M} c_k |f_k\rangle \in \mathcal{V}. \tag{4.5}$$

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

$$F_0 = \sum_{k=1}^{M} |f_k\rangle \langle w_k| . \tag{4.6}$$

The action of F_0 is also shown in Figure 4.1. 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 $\mathcal{W}$. We have $N(A_0) = \mathcal{V}^{\perp} = R(F_0)^{\perp}$ and it can also be shown that $N(F_0) = R(A_0)^{\perp}$. An analogous pair of relations holds for any two linear transformations related by the concept of an adjoint [43], [44] (see Appendix A.4). We have $F_0^{\dagger} = A_0$ and $A_0^{\dagger} = F_0$.[2] For some

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

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 have arrived at the same synthesis map (see Appendix A.3).

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 $\{|w_k\rangle\}$.

4.3 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 $\{|f_k\rangle\}$, and also when considering how $|v\rangle$ can be reconstructed from the collection $\{a_k = \langle f_k|v\rangle\}$ of its frame coefficients. Given some $|v\rangle \in \mathcal{V}$ and a fixed frame $\{|f_k\rangle\}$ for $\mathcal{V}$, consider first the problem of obtaining a set of coefficients $\{\tilde{a}_k\}$ such that

$$|v\rangle = \sum_{k=1}^{M} \tilde{a}_k \, |f_k\rangle \, . \tag{4.7}$$

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

$$|v\rangle = \sum_{k=1}^{M} \langle \tilde{f}_k|v\rangle \, |f_k\rangle \;\; \text{for all} \;\; |v\rangle \in \mathcal{V}. \tag{4.8}$$

A dual frame is always guaranteed to exist [39], and as we will show below if $\{|\tilde{f}_k\rangle\}$ is dual to $\{|f_k\rangle\}$ then the reverse is also true. Comparing Equations (4.7) and (4.8), it is clear that Equation (4.7) 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 (4.8), $\{|\tilde{f}_k\rangle\}$ 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 (4.8) has the equivalent forms

$$|v\rangle = \sum_{k=1}^{M} |f_k\rangle \langle \tilde{f}_k|v\rangle = F_0 \, \tilde{A}_0 \, |v\rangle \;\text{ for all }\; |v\rangle \in \mathcal{V} \tag{4.9a}$$

$$F_0 \, \tilde{A}_0 = \mathcal{P}_{\mathcal{V}}. \tag{4.9b}$$

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 $|\tilde{a}\rangle$ corresponding to a particular dual frame and a particular $|v\rangle \in \mathcal{V}$,

$$|\tilde{a}\rangle = \tilde{A}_0 \, |v\rangle = \sum_{k=1}^{M} \tilde{a}_k \, |w_k\rangle \in \mathcal{W}, \tag{4.10}$$

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

$$|\tilde{f}_k\rangle = (F_0 \, A_0)^{-1} \, |f_k\rangle \,, \quad 1 \leq k \leq M. \tag{4.11}$$

A derivation of this fact is included in Appendix A.5. The dual frame defined by Equation (4.11) is referred to as the canonical dual frame of $\{|f_k\rangle\}$ [39]. According to Equation (4.11) 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 can be shown using these expressions for F_{can} and A_{can}, in addition to the relation we have $(T \, R)^{\dagger} = R^{\dagger} \, T^{\dagger}$, that if $\{|\tilde{f}_k\rangle\}$ is the canonical dual of $\{|f_k\rangle\}$, then the reverse is also true.

It is worth mentioning another notable property of the canonical dual frame relating to its use as a solution to the problem of reconstructing an unknown vector from imprecise versions of its frame coefficients. For some $|v\rangle \in \mathcal{V}$, let $\{|f_k\rangle\}$ be a fixed analysis frame for $\mathcal{V}$ and assume that the coefficients $\{a_k = \langle f_k|v\rangle\}$ are known. Consider the problem of

finding a synthesis frame $\{|\tilde{f}_k\rangle\}$ such that

$$|v\rangle = \sum_{k=1}^{M} \langle f_k|v\rangle \, |\tilde{f}_k\rangle \;\text{ for all }\; |v\rangle \in \mathcal{V}. \tag{4.12}$$

Comparing Equation (4.12) with Equation (4.8), Equation (4.12) states that $\{|f_k\rangle\}$ is a dual frame of $\{|\tilde{f}_k\rangle\}$. Clearly, we have $F_{\text{can}}\, A_0\, |v\rangle = |v\rangle$ for all $|v\rangle \in \mathcal{V}$ and thus the canonical dual frame (unsurprisingly) satisfies Equation (4.12). Equation (4.12) has the equivalent forms

$$|v\rangle = \tilde{F}_0\, A_0\; |v\rangle \;\text{ for all }\; |v\rangle \in \mathcal{V} \tag{4.13a}$$

$$\tilde{F}_0\, A_0 = \mathcal{P}_{\mathcal{V}}. \tag{4.13b}$$

Equations (4.13b) and (4.9b) are in fact also equivalent and a dual frame could be defined according to either one. Their equivalence is straightforward to derive by taking the adjoint of both sides of either equation to show that it implies the other. For example, assume that Equation (4.13b) is true. The adjoint of the left-hand side is $(\tilde{F}_0\, A_0)^\dagger = (F_0\, \tilde{A}_0)$ and the adjoint of the right-hand side is $\mathcal{P}_{\mathcal{V}}^\dagger = \mathcal{P}_{\mathcal{V}}$ (see Appendix A.4). Thus $(F_0\, \tilde{A}_0) = \mathcal{P}_{\mathcal{V}}$.

If the $\{a_k\}$ are instead only known to within some error, the question arises of which dual frame is the optimal synthesis frame with respect to various cost criteria. In Section 4.6 we review how the canonical dual frame is the optimal choice when the error values are additive and uncorrelated. The fact that the nullspace of F_{can} and the range of A_0 are related via $N(F_{\text{can}}) = R(A_0)^\perp$ is a significant part of the derivation.

4.4 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 a special case of tight frames where the constant is equal to one. Naimark's Theorem ensures the existence of a special orthonormal basis $\{|w_k\rangle\}$ for $\mathcal{W}$ defined in relation to a Parseval frame $\{|f_k\rangle\}$ for $\mathcal{V}$.

4.4.1 Parseval Frames

A tight frame is one that satisfies Parseval's identity [39] up to a constant factor,

$$\sum_{k=1}^{M} |\langle f_k|v\rangle|^2 = C\,||v||^2 \tag{4.14}$$

for some $C > 0$ and for all $|v\rangle \in \mathcal{V}$. In reference to Equation (4.1), Equation (4.14) is equivalent to the statement the the frame bounds of $\{|f_k\rangle\}$ are equal, $C = D > 0$. When $C = D = 1$ the frame is referred to as a Parseval frame. Orthonormal bases are a special case of Parseval frames with $M = N$. In much the same way that the energy of the sequence of discrete Fourier transform coefficients of a finite-length sequence is equal to the energy of the original sequence, Equation (4.14) states that when $\{|f_k\rangle\}$ is a tight frame the sum of the squares of the coefficients $\{a_k = \langle f_k|v\rangle\}$ is proportional to the squared norm of the original vector.

Parseval frames are always self-dual. To show that this is true, note that the sum in Equation (4.14) can be alternately expressed as

$$\sum_{k=1}^{M} |\langle f_k|v\rangle|^2 = \sum_{k=1}^{M} \langle v|f_k\rangle\langle f_k|v\rangle = \langle v|\left(\sum_{k=1}^{M} |f_k\rangle\langle f_k|v\rangle\right). \tag{4.15}$$

For Equation (4.14) to be satisfied, the above expression must be equal to $||v||^2 = \langle v|v\rangle$ for all $|v\rangle \in \mathcal{V}$, implying that

$$\sum_{k=1}^{M} |f_k\rangle\langle f_k|v\rangle = |v\rangle \;\text{ for all }\; |v\rangle \in \mathcal{V}. \tag{4.16}$$

And Equation (4.16) states by definition that $\{|f_k\rangle\}$ is a dual frame of itself. In fact, it is the canonical dual frame of itself [39]. 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. In particular there is no concern for issues of computational complexity or instability that might result from the inversion of the operator $(F_0\,A_0)$ as in Equation (4.11) [39]. This is significant in the context of quantum state estimation where $\mathcal{V}$ represents a vector space whose elements are operators and correspondingly $(F_0\,A_0)$ is a so-called "superoperator" [11].

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.

4.4.2 Naimark's Theorem

Naimark's Theorem is well-known in both the frame theory and quantum physics communities. The version stated below will perhaps be most familiar to readers with a background in frame theory (see [45], or Theorem 1.9 in [39]). The version more typically used in the quantum physics community is stated in terms of POVMs and often arises in the context of the physical realizability of non-standard measurements [46], [47]. In the statement of Naimark's Theorem below we continue to assume that $\mathcal{V}$ and $\mathcal{W}$ are finite-dimensional with dimensions N and M, respectively. However, 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

$$\mathcal{P}_{\mathcal{V}} |w_k\rangle = |f_k\rangle, \quad 1 \leq k \leq M. \tag{4.17}$$

A derivation of one direction of the theorem is given in Appendix A.6. A derivation of the other direction can be found in, for example, [45]. We will refer to Equation (4.17) 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 *as long as* 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.

4.4.3 Synthesis and Analysis Maps of a Parseval Frame

We will show that 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 since Parseval frames are self-dual we have $F_0 A_0 = \mathcal{P}_\mathcal{V}$ according to Equation (4.9b). A_0 can be expressed as

$$A_0 = \sum_{k=1}^{M} |w_k\rangle \langle f_k| = \mathcal{P}_\mathcal{V} \left(\sum_{k=1}^{M} |f_k\rangle \langle f_k| \right) = \mathcal{P}_\mathcal{V} F_0 A_0 = \mathcal{P}_\mathcal{V}, \tag{4.18}$$

where we have used the idempotency of $\mathcal{P}_\mathcal{V}$, i.e., $\mathcal{P}_\mathcal{V}^2 = \mathcal{P}_\mathcal{V}$. This establishes that $A_0 = \mathcal{P}_\mathcal{V}$. Since orthogonal projection operators are Hermitian (see Appendix A.4), taking the adjoint of both sides leads to the conclusion that $A_0^\dagger = F_0 = \mathcal{P}_\mathcal{V}$.

An alternative derivation of the fact that $A_0 = F_0 = \mathcal{P}_\mathcal{V}$ relies on the observation that any vector $|v\rangle \in \mathcal{V}$ can be represented in terms of its basis coefficients $\{b_k = \langle w_k|v\rangle\}$ or in terms of its frame coefficients $\{a_k = \langle f_k|v\rangle\}$. When the $\{|w_k\rangle\}$ satisfy Naimark's identity, the two sets of coefficients are identical,

$$b_k = \langle w_k|v\rangle = \langle f_k|v\rangle = a_k, \quad 1 \le k \le M. \tag{4.19}$$

The reason Equation (4.19) is true is because a given basis vector $|w_k\rangle$ can always be written as the sum of its orthogonal projection onto $\mathcal{V}$, which is equal to $|f_k\rangle$, and its orthogonal projection onto $\mathcal{V}^\perp$. The component in $\mathcal{V}^\perp$ has no impact on the value of the inner product of $|w_k\rangle$ with $|v\rangle$.

We next note that since the $\{|w_k\rangle\}$ are orthonormal, we have

$$|v\rangle = \sum_{k=1}^{M} \langle w_k|v\rangle |w_k\rangle = \sum_{k=1}^{M} a_k |w_k\rangle \text{ for all } |v\rangle \in \mathcal{V}. \tag{4.20}$$

A_0 maps every $|v\rangle \in \mathcal{V}$ to $A_0|v\rangle = \sum_k b_k |w_k\rangle$, and since $\{b_k = a_k\}$ this implies that $A_0|v\rangle = |v\rangle$ for all $|v\rangle \in \mathcal{V}$. By definition A_0 also maps all elements of $\mathcal{V}^\perp$ to the zero vector. Taken together these two properties imply that $A_0 = \mathcal{P}_\mathcal{V}$. Similarly, F_0 maps every $|w\rangle \in \mathcal{W}$ to $F_0|w\rangle = \sum_k a_k |f_k\rangle$. Applying Naimark's identity and noting that Equation (4.20) holds for all $|w\rangle \in \mathcal{W}$ leads to $F_0|w\rangle = \mathcal{P}_\mathcal{V}(\sum_k a_k |w_k\rangle) = \mathcal{P}_\mathcal{V}|w\rangle$, and thus $F_0 = \mathcal{P}_\mathcal{V}$.

4.5 Frame Representations of Operator Spaces

The goal of Section 4.5 is to extend the discussion of frame representations to vector spaces $\mathcal{V}$ whose elements are Hermitian operators acting on a given Hilbert space $\mathcal{H}$. Such a vector space is sometimes referred to as an operator space. Unlike in Section 3, in Section 4.5 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 Sections 5 and 6.

For the remainder of the monograph, $\mathcal{V}$ 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 [11] and [22], 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 (4.25) below. The same notation carries over to elements of $\mathcal{W}$. Linear operators on $\mathcal{W}$ ("superoperators" [11]) 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 $\mathcal{V}$.

4.5.1 Defining $\mathcal{V}$ and $\mathcal{W}$

Assume that $\mathcal{H}$ is a vector-valued 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

$$\mathcal{U}^\perp = \text{span}\{I\}, \tag{4.21a}$$

$$\mathcal{U} = \text{span}\{V \in \mathcal{V} : \langle\langle I|V\rangle\rangle = \text{Tr}(V) = 0\}, \tag{4.21b}$$

where I denotes the identity operator on $\mathcal{H}$. $\mathcal{U}$ is the span of all trace 0 operators in $\mathcal{V}$. It has dimension $(N-1) = (d^2-1)$ and is always isomorphic to $\mathbb{R}^{d^2-1}$ [11]. $\mathcal{U}^\perp$ 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 $\mathcal{P}_{\mathcal{U}^\perp}(V) = \langle\langle I|V\rangle\rangle\,|I\rangle\rangle/d = \mathrm{Tr}(V)\,|I\rangle\rangle/d$, where the factor of $1/d$ accounts for the fact that $|I\rangle\rangle/\sqrt{d}$ has unit norm with respect to the inner product defined below. Therefore, V can always be written as

$$|V\rangle\rangle = \mathcal{P}_{\mathcal{U}^\perp}\,|V\rangle\rangle + \mathcal{P}_{\mathcal{U}}\,|V\rangle\rangle = \frac{\mathrm{Tr}(V)}{\sqrt{d}}\,\frac{|I\rangle\rangle}{\sqrt{d}} + \mathcal{P}_{\mathcal{U}}\,|V\rangle\rangle. \tag{4.22}$$

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 thus 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

$$V = \sum_{i=1}^{d} a_i\,|x_i\rangle\,\langle x_i|\,, \tag{4.23}$$

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

$$U = \sum_{i=1}^{d'-d} b_i\,|y_i\rangle\,\langle y_i|\,, \tag{4.24}$$

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 (4.23) or (4.24) – 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

$$\langle\langle W_1|W_2\rangle\rangle = \sum_{i=1}^{d'} \gamma_i \langle e_i| W_2 |e_i\rangle \text{ where } W_1 = \sum_{i=1}^{d'} \gamma_i |e_i\rangle \langle e_i| . \tag{4.25}$$

In Equation (4.25), 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 (4.23) and (4.24), 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 (4.25) 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 [10], [35].

4.5.2 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

$$C\,||V||^2 \leq \sum_{k=1}^{M} |\langle\langle F_k|V\rangle\rangle|^2 \leq D\,||V||^2 \tag{4.26}$$

for some $0 < C \leq D < \infty$ and for all $|V\rangle\rangle \in \mathcal{V}$ [11]. 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 (4.26) may also be extended to include to the notion of a generalized operator frame with respect to a given measure (see Appendix A.9). In the terminology of [11], a set of operators satisfying Equation (4.26) is referred to as a generalized operator frame with respect to the counting measure.

4.5.3 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 $\mathcal{V}$. While none of the concepts presented in Section 4.5.3 are specific to the context of quantum mechanics, they are relevant to the simulations presented in Sections 5 and 6 involving qubit density operators. For generalizations to values of $d > 2$, we refer the reader to, for example, [11], [15].

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 [35], is a commonly used basis for $\mathcal{V}$. It is an orthonormal basis with respect to the inner product defined by Equation (4.25). We have $\mathcal{U}^\perp = \text{span}\{I\}$ and $\mathcal{U} = \text{span}\{\sigma_1, \sigma_2, \sigma_3\}$. 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 $\{I/\sqrt{2}, \sigma_1/\sqrt{2}, \sigma_2/\sqrt{2}, \sigma_3/\sqrt{2}\}$,

$$V = c_0 \frac{|I\rangle\rangle}{\sqrt{2}} + c_1 \frac{|\sigma_1\rangle\rangle}{\sqrt{2}} + c_2 \frac{|\sigma_2\rangle\rangle}{\sqrt{2}} + c_3 \frac{|\sigma_3\rangle\rangle}{\sqrt{2}}. \tag{4.27}$$

In Equation (4.27) the basis expansion coefficients are $c_0 = \langle\langle I|V\rangle\rangle/\sqrt{2} = \text{Tr}(V)/\sqrt{2}$ and $c_i = \langle\langle \sigma_i|V\rangle\rangle/\sqrt{2} = \text{Tr}(\sigma_i V)/\sqrt{2}$ for $1 \leq i \leq 3$. A crucial

relation that forms the foundation of much of the discussion in Sections 5 and 6 is that if V is positive semidefinite then we always have

$$\sqrt{c_1^2 + c_2^2 + c_3^2} \leq \mathrm{Tr}(V)/\sqrt{2}, \tag{4.28}$$

with equality if and only if V has rank one. Equation (4.28) 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 (4.28) 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 4.1. 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

$$|x\rangle = c_0 \, |b_0\rangle + c_1 \, |b_1\rangle + c_2 \, |b_1\rangle , \tag{4.29}$$

where $c_i = \langle b_i | x \rangle$ for $0 \leq i \leq 2$. As shown in Figure 4.2, 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.

4.6 Robustness of Frame Representations

Given an unknown vector $|v\rangle \in \mathcal{V}$ and an analysis frame $\{|f_k\rangle\}$ for $\mathcal{V}$, $|v\rangle$ can always be written as

$$|v\rangle = \sum_{k=1}^{M} a_k \, |\tilde{f}_k\rangle , \tag{4.30}$$

where $a_k = \langle f_k | v \rangle$ for $1 \leq k \leq M$ and the synthesis frame $\{|\tilde{f}_k\rangle\}$ is any dual frame of $\{|f_k\rangle\}$. An important problem in classical signal

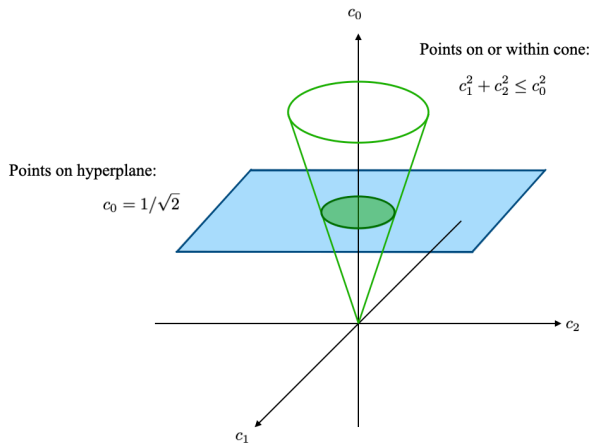


Figure 4.2: Illustration of the constraints described in Example 4.1.

processing is that of reconstructing $|v\rangle$ given only imprecise versions of the $\{a_k\}$ after they have been affected by some source of error. In the remainder of Section 4.6 we describe one version of this problem that incorporates a specific model for the error source in more detail. As the discussion is not directly relevant to binary hypothesis testing, some readers may wish to proceed directly to Section 5.

Throughout Section 4.6 we use notation corresponding to vector-valued vector spaces, but we emphasize that all of the analysis applies equally well to operator spaces. The key conclusion is that when $\{|f_k\rangle\}$ is what is referred to as an equal norm tight frame and the error values are additive and uncorrelated, the quality of reconstruction is reduced when the variance of the error values is reduced and (separately) when the number M of frame vectors is increased. In Section 5 we briefly address how an analogous problem can be formulated in the context of quantum state estimation.

4.6.1 Optimality of the Canonical Dual

We assume that the observed coefficients are $\{a_k + e_k\}$, where the individual error values $\{e_k\}$ have zero mean, variance σ^2, and collectively

are pairwise uncorrelated. That is,

$$\mathbb{E}[e_k] = 0, \quad 1 \leq k \leq M, \tag{4.31a}$$

$$\mathbb{E}[e_j e_k] = \begin{cases} \Delta^2 \text{ if } j = k \\ 0 \text{ if } j \neq k \end{cases}, \quad 1 \leq j, k \leq M. \tag{4.31b}$$

Equations (4.31) have been shown to be a useful model mathematically in certain scenarios, despite not always being literally true in practice (see, for example, Section 4 of [25]). The observed coefficients can be assembled into a vector that is the sum of the true coefficient vector $A_0 |v\rangle = \sum_k a_k |w_k\rangle \in \mathcal{W}$ with the error vector, defined by $|w_e\rangle = \sum_k e_k |w_k\rangle \in \mathcal{W}$. For a given synthesis frame $\{|\tilde{f}_k\rangle\}$ with synthesis map $\tilde{F}_0$, the reconstructed vector $|\hat{v}\rangle$ is obtained by applying $\tilde{F}_0$ to the observed coefficient vector,

$$|\hat{v}\rangle = \tilde{F}_0 \left(A_0 |v\rangle + |w_e\rangle \right) = |v\rangle + |v_e\rangle . \tag{4.32}$$

In Equation (4.32) we have defined the final error vector $|v_e\rangle = \tilde{F}_0 |w_e\rangle = \sum_k e_k |\tilde{f}_k\rangle$. The objective is to find the synthesis frame that minimizes the expected value of the squared norm of $|v_e\rangle$, i.e., we want to minimize $\mathcal{E}$ where

$$\mathcal{E} = \mathbb{E}\left[||v_e||^2 \right] = \mathbb{E}\left[||\tilde{F}_0 |w_e\rangle ||^2 \right] . \tag{4.33}$$

It is well-known that as long as the error values are uncorrelated, the optimal synthesis frame that minimizes $\mathcal{E}$ is the canonical dual of the analysis frame (see [48] and Appendix A.5). This is true even if each of the $\{e_k\}$ have possibly different variances denoted by $\{\Delta_k^2\}$. The underlying concept is that the nullspace of the synthesis operator of the canonical dual frame contains the largest portion of the error vector $|w_e\rangle$ as compared to other dual frames.

4.6.2 Application to Equal-Norm Tight Frames

In this monograph we will be particularly interested in the case where $\{|f_k\rangle\}$ is a tight frame for $\mathcal{V}$ with frame bound C, with the additional property that all of the frame vectors have the same norm, denoted by

B. Such a frame is typically referred to as an equal norm tight frame (ENTF) [39], [49]. Mathematically, we have

$$\sum_{k=1}^{M} |\langle f_k|v\rangle|^2 = C\,||v||^2 \text{ for all } |v\rangle \in \mathcal{V}, \tag{4.34a}$$

$$||f_k|| = B, \quad 1 \le k \le M. \tag{4.34b}$$

ENTFs are of interest in the quantum physics community in the form of tight IC POVMs as used for quantum state estimation. This is discussed further in Section 5. They are also utilized, for example, in the context of oversampling in classical signal processing (see [39], [41], [42] and Appendix A.7).

The canonical dual of an ENTF $\{|f_k\rangle\}$ is $\{|\tilde{f}_k\rangle = |f_k\rangle/C\}$. Its synthesis map, which we will denote as F_{can}, is equal to $F_{\text{can}} = F_0/C$. It can be shown [49] that for an ENTF the following relationship holds,

$$C\,N = M\,B^2. \tag{4.35}$$

To find the minimum value of $\mathcal{E}$ when $\{|f_k\rangle\}$ is an ENTF, we first evaluate the squared norm of $F_{\text{can}}\,|w_e\rangle$ for a specific error vector $|w_e\rangle$. Consider writing the error vector as $|w_e\rangle = |w_1\rangle + |w_2\rangle$ where $|w_1\rangle \in R(A_0)$ and $|w_2\rangle \in R(A_0)^{\perp}$. Since $N(F_{\text{can}}) = R(A_0)^{\perp}$ we have

$$F_{\text{can}}\,|w_e\rangle = F_{\text{can}}\,|w_1\rangle + F_{\text{can}}\,|w_2\rangle = F_{\text{can}}\,|w_1\rangle\,. \tag{4.36}$$

For any vector $|w\rangle \in \mathcal{W}$ that lies in $R(A_0)$, it can be shown that $||F_{\text{can}}\,|w\rangle\,||^2 = ||w||^2/C$ (see Appendix A.5). Moreover, using the fact that the individual error values $\{e_k\}$ satisfy Equations (4.31) it can be verified that the expected value of $||e_1||^2$ is $\mathbb{E}[||e_1||^2] = N\Delta^2$ (see Appendix A.8). The minimum value $\mathcal{E}^*$ of $\mathcal{E}$ is thus

$$\mathcal{E}^* = \mathbb{E}\left[||F_{\text{can}}\,|e\rangle\,||^2\right] = \frac{||e_1||^2}{C} = \frac{N\,\Delta^2}{C} = \frac{N^2\,\Delta^2}{M\,B^2}. \tag{4.37}$$

Equation (4.37) states that fixed N and B, $\mathcal{E}^*$ can be reduced by reducing Δ or by increasing M. When the variances of the $\{e_k\}$ are not assumed to be identical for all values of k, we have $\mathbb{E}[||\,|e_1\rangle\,||^2] = (N/M)\sum_k \Delta_k^2$, so Equation (4.37) becomes

$$\mathcal{E}^* = \frac{N}{MC}\sum_{k=1}^{M}\Delta_k^2 = \frac{N^2}{M^2B^2}\sum_{k=1}^{M}\Delta_k^2. \tag{4.38}$$

As expected, Equation (4.38) reduces to Equation (4.37) when $\Delta_k^2 = \Delta^2$ for all $1 \leq k \leq M$.

5

An Operator Frame View of Quantum Measurement

The main objective of Section 5 is to utilize the mathematical methodology developed in Section 4 to interpret the process of quantum measurement as it is defined by the postulates of quantum mechanics. Given an arbitrary finite-dimensional QMS with state space $\mathcal{H}$, density operators and POVM elements are linear operators on $\mathcal{H}$ can always be interpreted as elements of a common operator space $\mathcal{V}$. As we describe in Section 5.1, when a quantum measurement with POVM $\{E_k\}$ is performed on a QMS with denisty operator ρ, the probabilities $\{p(k) = \mathrm{Tr}(E_k\rho)\}$ of the possible measurement outcomes can then be expressed using an inner product defined on $\mathcal{V}$. Crucially, when the POVM is IC the $\{p(k)\}$ can be interpreted as the frame coefficients of ρ with respect to the $\{E_k\}$, since every IC POVM satisfies the definition of a frame for $\mathcal{V}$. The connection between IC POVMs and frames for $\mathcal{V}$ is a fundamental and established result and is reviewed in Section 5.1.1. POVMs corresponding to qubit measurements, which we refer to as qubit POVMs for brevity, are of particular interest throughout Sections 5 and 6. Specifically, the focus is on a class of qubit POVMs defined in Section 5.1.2 that we refer to as equal trace rank one (Etro) POVMs. Every M-element Etro POVM can be fully specified by M points on a sphere of radius $\sqrt{2}/M$ that we refer

to as an Etro sphere. The representation of a qubit POVM through M points on the Etro sphere is exactly analogous to the representation of a pure state qubit density operator through a point on the Bloch sphere. In Section 5.2 we use qubit Etro POVMs corresponding to one of the five Platonic solids to discriminate between two qubit density operators. In Section 6 we generalize this to POVMs specified by other sets of points on an Etro sphere.

5.1 Operator Spaces in Quantum Mechanics

Throughout Section 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 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 (4.25), 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 $\{E_k\}$.

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 5.1.2. According to Equation (4.22), 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 5.1.2 we apply these concepts as well as those described in Section 4.5.3 specifically to POVMs of a qubit, leading to the definition of an Etro sphere.

5.1.1 Informationally Complete and Overcomplete POVMs

The definition of an 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 3.2 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 $\mathcal{V}$ [10], [11], [13]. This statement can be generalized to include the case where $\mathcal{V}$ is infinite-dimensional and to include generalized operator-valued frames [11], but for simplicity we do not consider those scenarios in this monograph. It will be convenient moving forward to consider the following statements separately,

(i) $\{U_k\}$ is a POVM,

(ii) $\{U_k\}$ span $\mathcal{V}$,

(iii) $\{U_k\}$ maps every density operator $|\rho\rangle\rangle \in \mathcal{V}$ to a unique sequence of coefficients $\{\langle\langle U_k|\rho\rangle\rangle\}$.

The fundamental result states that if conditions (i) and (ii) are satisfied, then (iii) must be satisfied. And separately that if (i) and (iii) are satisfied, then (ii) must be satisfied. It does *not* state that (ii) implies

(iii) or vice versa. 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 [11], [15], [17], [19], [20], [50].

We first show that (i) and (ii) imply (iii). Assume that a set of operators $\{U_k\}$ in $\mathcal{V}$ satisfies (i) and (ii). 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 $\rho_1, \rho_2 \in \mathcal{V}$. We have

$$\mathbf{A}_0 |\rho_1\rangle\rangle = \sum_{k=1}^{M} |W_k\rangle\rangle\langle\langle U_k|\rho_1\rangle\rangle = \sum_{k=1}^{M} p_1(k) |W_k\rangle\rangle, \tag{5.4a}$$

$$\mathbf{A}_0 |\rho_2\rangle\rangle = \sum_{k=1}^{M} |W_k\rangle\rangle\langle\langle U_k|\rho_2\rangle\rangle = \sum_{k\in\mathcal{K}} p_2(k) |W_k\rangle\rangle, \tag{5.4b}$$

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 $\mathbf{A}_0 |\rho_1 - \rho_2\rangle\rangle = 0$ implying that $|\rho_1 - \rho_2\rangle\rangle = 0$, i.e., $\rho_1 = \rho_2$.

The statement that (i) and (iii) imply (ii) 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 (iii) is satisfied, 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 (iii). 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 $\mathcal{U}$.

IC POVMs are commonly studied in the context of quantum state estimation [11], [12], [19], [20], [50]–[52], 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 main motivation for the simulations presented in Section 5.2.

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". 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 [11] and are some of the most extensively studied.

5.1.2 The Bloch Sphere and The Etro Spheres

We now apply the discussion given in Section 4.5.3 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 4.5.3, 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

$$\rho = \frac{1}{\sqrt{2}}\frac{|I\rangle\rangle}{\sqrt{2}} + r_1\frac{|\sigma_1\rangle\rangle}{\sqrt{2}} + r_2\frac{|\sigma_2\rangle\rangle}{\sqrt{2}} + r_3\frac{|\sigma_3\rangle\rangle}{\sqrt{2}}, \tag{5.5a}$$

$$E_k = \frac{\mathrm{Tr}(E_k)}{\sqrt{2}}\frac{|I\rangle\rangle}{\sqrt{2}} + c_{k1}\frac{|\sigma_1\rangle\rangle}{\sqrt{2}} + c_{k2}\frac{|\sigma_2\rangle\rangle}{\sqrt{2}} + c_{k3}\frac{|\sigma_3\rangle\rangle}{\sqrt{2}}, \quad 1 \le k \le M, \tag{5.5b}$$

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 one and thus represents a pure state. The ball and sphere correspond within a constant factor to the very commonly used Bloch ball, which is typically assumed to have unit radius, and the corresponding Bloch sphere. The column vector $\mathbf{r}$ is proportional to the Bloch vector of ρ [35]. 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. 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 (5.5) into Equation (5.1) and utilizing the orthonormality of the $\{\sigma_i\}$ yields

$$p(k) = \frac{\mathrm{Tr}(E_k)}{2} + \mathbf{c_k}\cdot\mathbf{r} = \frac{1}{M} + \mathbf{c_k}\cdot\mathbf{r}, \tag{5.6}$$

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

Of particular interest in this monograph are qubit POVMs that we will refer to as equal trace rank one (Etro) POVMs. Unsurprisingly, an Etro POVM is one for which $\mathrm{Tr}(E_k) = 2/M$ for $1 \le k \le M$ and for which each of the $\{E_k\}$ has rank one, implying from Equation (4.28) that $\sqrt{c_{k1}^2 + c_{k2}^2 + c_{k3}^2} = \mathrm{Tr}(E_k)/\sqrt{2} = \sqrt{2}/M$ for $1 \le k \le M$. 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 the sphere corresponding to the

surface of the ball, which 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 M. An M-element Etro POVM can be fully specified by M vectors $\{\mathbf{c}_k\}$ extending from the origin to 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 5.1.1 that a qubit Etro POVM is IC if and only if the $\{\mathbf{c_k}\}$ span $\mathbb{R}^3$ [15]. The definition of an Etro POVM could easily be generalized to higher dimensions. However, in this monograph we use the term Etro POVM to refer specifically to those corresponding to qubit measurements.

Example 5.1. POVMs constructed using Platonic solids are used often in the literature [19], [20], [53], [54]. In our terminology, 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 for the state space of the qubit [55]. An Etro POVM satisfies the definition of a a tight IC POVM when its Etro vectors $\{\mathbf{c_k}\}$ form a tight frame for $\mathbb{R}^3$ [11]. All POVMs constructed from Platonic solids are tight IC POVMs.

Example 5.2. Consider a qubit POVM $\{E_1, E_2\}$ whose elements form a complete set of orthogonal projectors onto $\mathcal{H}$. As stated in Section 3.2, 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 such example.

5.2 Qubit State Discrimination using Platonic Solids

The topic of Section 5.2 is the following variation on the binary hypothesis testing problem considered in Section 3. Consider L identical QMSs whose states are all described by the density operator $\rho = \rho_i$ if $H = H_i$, for $i \in \{0, 1\}$. In Section 3 we assumed that $L = 1$, i.e., we assumed that there was a single QMS prepared in a state corresponding to the density operator ρ_0 or ρ_1. 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. Throughout Section 5.2, 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 $f_i(\mathbf{S})$ for $i \in \{0, 1\}$ are multinomial distributions (see Appendix A.10). A final decision of $\hat{H} = H_0$ or $\hat{H} = H_1$ is made based on the outcome of an LRT with threshold $\eta = q_0/q_1$ performed on the score variable.

In Example 5.3 below, we utilize POVMs constructed using Platonic solids, which were defined in Example 5.1. These POVMs are of significant interest in the context of qubit state estimation [11], [12], [19]–[21], [50]–[52] but have been utilized less often for binary hypothesis testing. We present evidence through simulation that discrimination performance improves with increasing L, the number of identically-prepared QMSs and (separately) with increasing M, the number of POVM elements. Note that as stated in Example 5.1, all POVMs constructed using Platonic solids are Etro POVMs. It is straightforward to show that they are all also IOC POVMs, with the exception of the POVM with $M = 4$ vertices constructed from a tetrahedron, which is IC but not IOC.

Example 5.3. 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 5.1.2, 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 of $L = 5$, 10, 20 are shown in Figure 5.1. The plots reflect an improvement in discrimination performance when either M or L is increased. 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.

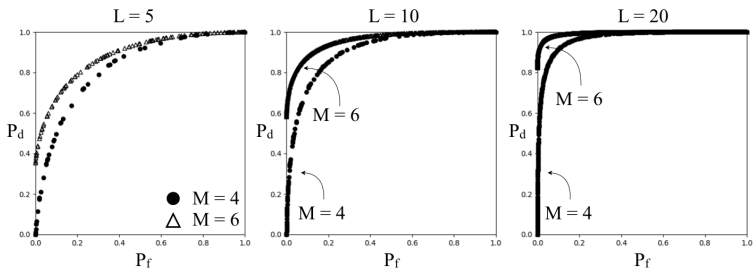


Figure 5.1: LRT QDOCs for ensemble sizes $L = 5, 10, 15$ and tight IC POVMs constructed from Platonic solids with $M = 4$ (tetrahedron) and $M = 6$ (octahedron) vertices.

5.3 Robustness of IOC POVMs for Quantum State Estimation

In Section 5.3 we focus on the following problem that was designed specifically to demonstrate that the robustness properties of ENTFs described in Section 4.6 can be applied in the context of quantum mechanics. Once again, readers interested only in binary hypothesis testing and not estimation may wish to proceed to Section 6. 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 4.6, the vector $|v\rangle$ lying in $\mathcal{V}$ is replaced by the shifted operator $(\rho - I/d)$ lying in the subspace $\mathcal{U}$ of $\mathcal{V}$. The analysis frame $\{|f_k\rangle\}$ is replaced by the operators $\{Q_k = E_k/\sqrt{\mathrm{Tr}(E_k)} - \sqrt{\mathrm{Tr}(E_k)}I/d\}$, which by assumption 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, the $\{Q_k\}$ are an ENTF for $\mathcal{U}$ satisfying

$$\sum_{k=1}^{M} |\langle\langle Q_k|V\rangle\rangle|^2 = C\,||V||^2 \text{ for all } V \in \mathcal{U}, \tag{5.7a}$$

$$||Q_k|| = B, \quad 1 \leq k \leq M. \tag{5.7b}$$

The set of operators $\{Q_k/C\}$ is the canonical dual frame of $\{Q_k\}$. The operator $(\rho - I/d)$ is an element of $\mathcal{U}$, so it can always be expressed as

$$\rho - \frac{I}{d} = \sum_{k=1}^{M} \langle\langle Q_k|\rho - I/d\rangle\rangle\, \tilde{Q}_k = \sum_{k=1}^{M} a_k\, \tilde{Q}_k \tag{5.8}$$

where $\{a_k = \langle\langle Q_k|\rho - I/d\rangle\rangle\}$ and $\{\tilde{Q}_k\}$ is any dual frame of $\{Q_k\}$. In terms of the probabilities $\{p(k) = \langle\langle E_k|\rho\rangle\rangle\}$, it can be shown through direct substitution that the $\{a_k\}$ can be expressed as

$$a_k = \langle\langle Q_k|\rho - I/d\rangle\rangle = \frac{p(k)}{\sqrt{\mathrm{Tr}(E_k)}} - \frac{\sqrt{\mathrm{Tr}(E_k)}}{d}. \tag{5.9}$$

We assume that the observed, imprecise values $\{\hat{a}_k = a_k + e_k\}$ of the frame coefficients are obtained as follows. Given L identically prepared quantum systems all in the state ρ, the true probabilities $\{p(k)\}$ are estimated by performing a quantum measurement with associated POVM $\{E_k\}$ on each system. The set of all measurement outcomes are used to compute estimates of the true probabilities in the form of the relative frequencies. If n_k is the number of times the measurement outcome associated with E_k occurred, the relative frequencies are $\{\hat{p}(k) = n_k/L\}$. The $\{\hat{a}_k\}$ are obtained by replacing $p(k)$ with $\hat{p}(k)$ in Equation (5.9), yielding

$$\hat{a}_k = \frac{\hat{p}(k)}{\sqrt{\mathrm{Tr}(E_k)}} - \frac{\sqrt{\mathrm{Tr}(E_k)}}{d} = a_k + e_k. \tag{5.10}$$

where $e_k = (\hat{p}_k - p(k))/\sqrt{\mathrm{Tr}(E_k)}$ for $1 \leq k \leq M$. An estimate $(\hat{\rho} - I/d)$ of $(\rho - I/d)$ is constructed by replacing the $\{a_k\}$ with the $\{\hat{a}_k\}$ in Equation (5.8),

$$\hat{\rho} - \frac{I}{d} = \sum_{k=1}^{M} \hat{a}_k\, \tilde{Q}_k = \rho - \frac{I}{d} + \rho_e, \tag{5.11}$$

where we have defined $\rho_e = \hat{\rho} - \rho = \sum_k e_k \tilde{Q}_k$. The objective is to find the synthesis frame $\{\tilde{Q}_k\}$ that minimizes the expected squared norm of $||\rho_e||$, i.e., to minimize $\mathcal{E}$ where

$$\mathcal{E} = \mathbb{E}\left[||\hat{\rho} - I/d||^2\right]. \tag{5.12}$$

Unlike in Section 4.6.1, setting $\{\tilde{Q}_k\}$ equal to the canonical dual of $\{Q_k\}$ is not necessarily optimal in terms of minimizing $\mathcal{E}$ because the error values $\{e_k\}$ are not pairwise uncorrelated. In fact, it can be shown (see Appendix A.10) that

$$\mathbb{E}[e_k] = 0, \quad 1 \le k \le M \tag{5.13a}$$

$$\mathbb{E}[e_j e_k] = \begin{cases} \frac{p(k)\,(1-p(k))}{L\,\mathrm{Tr}(E_k)} = \Delta_k^2 \text{ if } j = k \\ \frac{-p(j)\,p(k)}{L\sqrt{\mathrm{Tr}(E_j)\,\mathrm{Tr}(E_k)}} \text{ if } j \neq k \end{cases}, \quad 1 \le j,\, k \le M. \tag{5.13b}$$

The optimal synthesis frame $\{\tilde{Q}_k\}$ could be found by first whitening the $\{e_k\}$ and then computing the canonical dual of the effective analysis frame. However, as demonstrated next in Example 5.4, the conclusion reached in Section 4.6 under the assumption that the $\{e_k\}$ are uncorrelated is still supported by our simulations. This suggests that the correlations present in the simulations are small enough that they can be disregarded for the purpose of high-level predictions and modeling. According to Equation (4.38), the value $\mathcal{E}_{\text{can}}$ of $\mathcal{E}$ obtained by setting $\{\tilde{Q}_k = Q_k/C\}$ to be the canonical dual frame of $\{Q_k\}$ is

$$\mathcal{E}_{\text{can}} = \frac{N^2}{M^2 B^2} \sum_{k=1}^{M} \Delta_k^2 = \frac{N^2}{L\,M^2\,B^2} \sum_{k=1}^{M} \frac{p(k)\,(1-p(k))}{\mathrm{Tr}(E_k)}. \tag{5.14}$$

Example 5.4. In this example we demonstrate the utility of Equation (5.14) for estimating the state of a qubit using tight IC POVMs $\{E_k\}$ corresponding to Platonic solids with $M = 4, 6, 8, 12$ vertices. The value of B was chosen so that all of the $\{E_k\}$ are positive semidefinite. As a consequence it can be verified that B^2 must be proportional to M. Equation (5.14) suggests that with all else fixed, we would expect $\mathcal{E}_{\text{can}}$ to scale as $1/(LM^2)$. Indeed, this is what we observe.

For the purposes of illustration, we chose the density operator ρ to be $\rho = |\psi\rangle\langle\psi|$ with $|\psi\rangle = \cos(\theta/2)\,|0\rangle + \sin(\theta/2)\,|1\rangle$ and $\theta = 2\pi/3$.

The collection sizes used were $L = 5, 10, 50$. For a given value of L and a given POVM $\{E_k\}$ with M elements, 500 independent trials of the following procedure was performed. First L independent samples were drawn from the true probability distribution $\{p(k)\}$, in order to simulate an experiment in which L quantum measurements, each with POVM $\{E_k\}$, were performed on L identically prepared QMSs in the state ρ. This resulted in a collection of relative frequencies $\{\hat{p}(k)\}$ from which an estimated value of $(\hat{\rho} - I/d)$ and thus an error operator $\rho_e = \hat{\rho} - \rho$ were constructed. Finally, the value of $||\rho_e||^2$ was computed at the end of each trial. The compilation of these values after all trials were complete were used to compute estimates of the expected value $\mathcal{E}_{\text{can}} = \mathbb{E}[||\rho_e||^2]$ and the variance $\text{var}(||\rho_e||^2)$.

The mean values $\mathcal{E}_{\text{can}} = \mathbb{E}[||\rho_e||^2]$ and standard deviations $\text{var}(||\rho_e||^2)^{-1/2}$ over all trials and for all combinations of M and L are presented in Table 5.1 and shown in Figure 5.2. The results clearly indicate that for fixed values of M, increasing the value of L reduces both the mean value and standard deviation of $||\rho_e||^2$. And, for fixed values of L, increasing the value of M also reduces the mean value and standard deviation of $||\rho_e||^2$. However, the tradeoff is not entirely symmetric. For fixed values of M, doubling the value of L roughly halves both the mean value and standard deviation of $||\rho_e||^2$. In other words, both quantities are roughly inversely proportional to L. For the mean value, this is as expected from Equation (5.14). But for fixed values of L, doubling the value of M causes the mean value and standard deviation of $||\rho_e||^2$ to become reduced nearly by a factor of 4, suggesting that the two quantities are roughly inversely proportional to M^2. Again, this is as expected for the mean value according to Equation (5.14) when $\text{Tr}(E_k) = 2/M$ and B^2 is proportional to M.

Table 5.1: Mean values and standard deviations of $||\rho_{se}||^2$, rounded to the nearest tenth, over all trials and for different combinations of M and L.

Number of POVM Elements (M)	Collection Size (L)		
	5	10	50
4	23.8 ± 17.0	11.2 ± 8.2	2.3 ± 1.9
6	9.5 ± 6.7	5.4 ± 4.2	1.0 ± 0.8
8	5.3 ± 3.9	2.6 ± 2.0	0.5 ± 0.4
12	2.4 ± 1.8	1.2 ± 0.9	0.2 ± 0.2

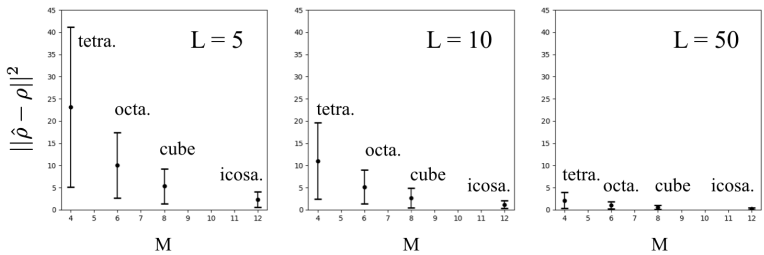


Figure 5.2: Mean (circular marker) and standard deviation (upper and lower error bars) of $||\hat{\rho} - \rho||^2$ over 500 independent trials for ensemble sizes $L = 5, 10, 50$ and tight IC POVMs constructed from Platonic solids with $M = 4$ (tetrahedron), $M = 6$ (octahedron), $M = 8$ (cube), and $M = 12$ (icosahedron) vertices.

6

Qubit State Discrimination on the Etro Spheres

As mentioned briefly in Section 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, which 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 Section 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 [19], [20], [53], [54], [56]. In our terminology these are Etro POVMs constructed from sets of points on the Etro sphere of radius $\sqrt{2}/M$ corresponding to the vertices of one of the Platonic solids. A main motivation of Section 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. The findings can also be found in [57]. As in Section 5.2 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 Section 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 [58]–[60]. We first consider distributions of points that correspond to the vertices of a Platonic solid, in addition 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

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

In Equation (6.1), $\|\mathbf{c}_j - \mathbf{c}_k\|$ denotes the Euclidean 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. [61] 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

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

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

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 $||\epsilon|| \le 10^{-8}$.

6.2 Results and Simulations

We simulated a range of the parameter values M, $0 \leq \alpha \leq \pi$, and $0 \leq q_1 \leq 1$ all with $L = 5$. 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. 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.[1] For fixed q_1, the decrease in sensitivity with M is more pronounced for smaller values of α, 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 $\max P_e$.

Our exploratory investigation suggests that various arrangements of points that 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, 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.

[1] The Platonic solid with $M = 20$ vertices, the dodecahedron, was not simulated due to computational constraints.

7

Summary, Reflections, and Further Thoughts

Our key objectives in writing this monograph have been to develop and present a framework for binary hypothesis testing as it applies to both the classical and quantum mechanical environments. As we formulate it in Section 2, this framework consists of two stages, the first of which we refer to as the pre-decision operator which generates a scalar- or vector-valued score variable or more generally and the second the binary decision rule applied to the score vector. And throughout this monograph we have specifically focused on issues related to the design of the pre-decision operator, the choice of decision rule and performance of the overall process as one or more parameters are varied. In this monograph, performance is specifically characterized by the tradeoff between the probability of making the correct decision given that the positive hypothesis is true, commonly referred to as the probability of detection P_d, and the probability of making the incorrect decision when the null hypothesis is true, commonly referred to as the probability of false alarm P_f. The characterization of an "optimal" binary decision-making system is a well-studied area of statistics with a long history and is often formulated in terms of a variety of criteria, e.g., minimizing probability of error, minimizing risk or Bayes' cost, or maximizing

P_d for a specified upper bound on P_f. Conveniently, mathematically all of these criteria lead to an optimal binary decision rule in the form of an LRT. This corresponds to evaluating the likelihood ratio against an appropriate threshold. The likelihood ratio will of course be a function of one or more of the parameters characterizing the underlying problem which then leads to the operating characteristics. More generally, we used the term operating characteristic to refer to any characterization of the tradeoff between P_f and P_d as one or more parameters of the discrimination system is varied. Essentially an operating characteristic viewed in this light can be thought of as a trajectory in a high-dimensional space with coordinates corresponding to P_f, P_d, and the parameters being varied. For a given trajectory, a two-dimensional graph of P_f vs. P_d is the projection of the trajectory onto the P_f-P_d plane. While such P_f-P_d projections are often used as visual representations of operating characteristics, it is important to note that they do not explicitly provide information about the parameter or parameters being varied. When there is only one parameter to vary, the P_f-P_d projection is traditionally referred to as an ROC. This terminology stems from its widespread use in signal detection scenarios and has since been adopted and adapted to broader settings.

In many application contexts, it is difficult or impossible to determine the likelihood ratio because doing so requires an assessment of underlying conditional probabilities, prior probabilities, etc. Partially for this reason, the decision-making system often applies a threshold test directly to a scalar score variable generated by combining together a variety of measurements. We refer to this as an SVT. In Section 2 of this monograph, much of this background along with our perspective on it for classical binary hypothesis testing is summarized. Of particular importance is the development in Section 2.5 that addresses the relationship between operating characteristics generated using LRTs and SVTs. In many situations, it is not unreasonable to expect or assume that the P_f-P_d projection of an SVT operating characteristic is equivalent to that of an LRT operating characteristic or that one can be obtained from the other through a simple and reversible change of variables. In Section 2.5.4 we show that in the case of a P_f-P_d projection of an SVT ROC, if that curve is concave then it is guaranteed to be equivalent to the

P_f-P_d projection of the LRT ROC for the same underlying conditional distributions. Also of significance is the discussion in Section 2.5.5 in which we outline a procedure for generating the P_f-P_d projection of the LRT ROC from that of the SVT ROC when the SVT does not correspond to the LRT through a simple change of variables.

In classical binary hypothesis testing, the operating characteristics are associated with the decision rule with the viewpoint that the pre-decision operator has previously been specified. However there is also the possibility of specifying the decision rule and exploring the operating characteristics that result from varying the pre-decision operator. To distinguish these two classes of operating characteristics in the classical case, we introduce in Section 2.6 the terminology of CDOCs and CMOCs, the first referring to operating characteristics when the pre-decision operator has been specified and some parameter or parameters of the binary decision rule is varied, and the second for which the binary decision rule is fixed and some aspect of the pre-decision operator is varied. In Section 3, when we address quantum binary state discrimination, the corresponding terminology is QDOCs and QMOCs. In many classical settings the pre-decision operator is based on measurements that can be made or processing that can be applied prior to making the decision. For example, in signal detection environments such as with the use of radar and sonar, the pre-decision operator is a filter which is designed based on knowledge of the signal to be detected. In medical applications, the pre-decision operator might consistent of combining multiple measurements in some appropriate way. In these various contexts the traditional focus is on the operating characteristics of the decision rule. CMOCs for which parameters of the pre-decision operator are varied are less typical and perhaps worthy of serious exploration in the classical case.

In Section 3 the same basic structure of a pre-decision operator followed by a binary decision rule is applied to exploring operating characteristics for quantum binary state discrimination. But 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 Section 3.2 we summarize the key postulates of quantum mechanics that govern the formulation and development of the quantum binary state discrimination problem that is a main focus of this monograph. One specific way of viewing the underlying problem as we formulate it is to imagine two possible physical environments that we would like to distinguish between. Any QMS prepared by or associated with each would have associated with it one of two known density operators. The decision 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 quantum measurement with an associated POVM. We assume that L independent QMSs prepared by only one of the environments are available for measurement and that we are able to determine the index of the resulting post-measurement state. 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. 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 corresponding to the chosen quantum measurement, has been specified and the operating characteristics correspond to parameters of the binary decision rule being varied. 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 relative frequency 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 operator $(\rho_1 - (q_0/q_1)\rho_0)$ can equivalently be used. (Recall that d is the dimension of the state space $\mathcal{H}$ of the QMS.) As introduced in Section 3.6, QMOCs correspond to keeping fixed the decision region for

the binary decision rule and varying the pre-decision operator through the POVM. Our presentation of QMOCs in Section 3.6 is primarily intended to introduce the concept and to illustrate it with a simple example.

Sections 5 and 6 are 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 Hilbert space $\mathcal{H}$. 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 set 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 Section 5 and 6 in designing IOC POVMs. In preparation for those discussions, in Section 4 we develop the representation of density operators and POVM elements as vectors in a vector space whose elements are Hermitian operators on the state space of a QMS. Consequently before utilizing frame representation of vector spaces, we summarize in Section 4 our perspective and the notation and key properties that we exploit in Sections 5 and 6. This includes the basics of frame theory. In effect, frames correspond to sets of linearly dependent vectors that span the space and thus every vector in the space can be constructed as a linear combination of the frame vectors. A basis 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 $\mathcal{W}$. In Section 4.2 we review the notion of analysis and synthesis operators and introduce the notion of analysis and synthesis maps. The analysis operator applied to a vector in $\mathcal{V}$ generates a set of coefficients representing the vector and the synthesis operator constructs a vector in $\mathcal{V}$ through a linear combination of the frame vectors. In other words,

the analysis operator generates a vector in $\mathbb{R}^M$ representing a vector in the N-dimensional space $\mathcal{V}$. The synthesis operator generates a vector in $\mathcal{V}$ from an appropriate subspace in $\mathbb{R}^M$. The analysis map generates a vector in $\mathcal{W}$ using the frame coefficients in a basis expansion of $\mathcal{W}$. And the synthesis map generates a vector in $\mathcal{V}$ by applying the coefficients of a basis expansion in $\mathcal{W}$ to the frame vectors in $\mathcal{V}$. This is a particular viewpoint that we've taken in this monograph in the context of Sections 5 and 6. Sections 4.3 and 4.4 then summarize the well-known notions in frame theory related to the use of dual frames, Parseval frames, and Naimark's Theorem. Section 4.5 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 Section 4.5 forms the basis for the discussion in Sections 5 and 6.

Section 5 applies the methodology in Section 4 to density operators, POVM elements and in particular utilizing frame theory to characterize informationally overcomplete (IOC) POVMs. Section 5.1.2 specifically addresses 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 Section 5.1.2, 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) in operator space. We refer to these as the Etro ball and Etro sphere which, for an M-element POVM, has 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 Section 6 is specifically based on selecting the M points on the Etro sphere from which the POVM is constructed. The concepts of IC and IOC POVMs are conveniently characterized in operator space with the methodology of frame theory as discussed in Section 5.1.1. 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 be positive semidefinite and they must collectively 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. In this case, the points on the Etro sphere are the M vertices of the corresponding Platonic solid. Example 5.3 in Section 5.2 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. 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.

The inspiration for Section 6 comes from the role of Platonic solids in quantum measurement and their potential use as illustrated in Example 5.3 for quantum state discrimination. As stated in Section 5.1.2 the vertices of each Platonic solid can be inscribed in an Etro sphere and define valid POVMs. In Section 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 Section 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, Section 6 offers an initial approach to the design of IOC POVMs. As we indicate throughout this monograph, and in this summary, there continue to be many opportunities for a continuing exploration of operating characteristics for binary discrimination and hypothesis testing in both the classical and quantum mechanics domains.

Acknowledgements

The preparation of this manuscript began several years ago and evolved in ways that we didn't imagine at the beginning. Throughout this evolution we had the good fortune of having rich discussions with a number of colleagues. We are particularly eager to thank Isaac Chuang and Qi Ding for their collaborative efforts on various aspects of the work and James Ward for his insightful comments and detailed suggestions for improvement. We also had very helpful and in depth technical discussions with Petros Boufounos. We received valuable technical perspective from George Verghese, Yonina Eldar, and Meir Feder. Finally, we would like to thank Yonina Eldar for inviting us write this monograph and Mike Casey for patiently encouraging us and accepting moving deadlines as we continued to develop the content.

Appendices

A

Optional Appendices

Since this monograph was intended for an audience with a diverse set of backgrounds, the purpose of the derivations contained in the following Appendices A.1 to A.10 is to provide some level of detail surrounding concepts and results that are likely familiar to some readers but perhaps not to others. Many of the derivations can also be found in some form in many classical signal processing, linear algebra, or quantum mechanics textbooks and review monographs. The title of each section contains a reference to the section in the main body of the monograph where the concept was first mentioned.

A.1 Optimal Neyman-Pearson Decision Regions

The following reasoning was adapted from [25]. 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 $[s', s'+ds']$. The resulting changes in P_f and P_d are

$$\Delta P_f = f_0(s')\,ds' - f_0(s)\,ds \tag{A.1a}$$

$$\Delta P_d = f_1(s')\,ds' - f_1(s)\,ds. \tag{A.1b}$$

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 (A.1) and combining them together leads to the requirement that

$$\frac{f_1(s')\,ds'}{f_0(s')\,ds'} \geq \frac{f_1(s)\,ds}{f_0(s)\,ds}. \tag{A.2}$$

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 (A.2) 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.

A.2 Orthonormality of the $\{|w_k\rangle\}$ (Section 4.1)

We wish to show that no generality is lost by assuming that the basis vectors $\{|w_k\rangle, 1 \leq k \leq M\}$ for $\mathcal{W}$ are orthonormal with respect to the $\langle\cdot|\cdot\rangle$ inner product. Assume that they are not and let $\{|e_k\rangle\}$ be a basis for $\mathcal{W}$ that is orthonormal with respect to the $\langle\cdot|\cdot\rangle$ inner product. Thus

$$\langle e_j|e_k\rangle = \delta_{jk}, \quad 1 \leq j,\, k \leq M, \tag{A.3}$$

where δ_{jk} takes the value 1 if $j = k$ and 0 otherwise. Now we define a function $f(\cdot,\cdot)$ that takes two vectors in $\mathcal{W}$ as input and outputs a

complex number. It is defined to satisfy the following properties,

$$f(a\,u_j + b\,u_k, u_\ell) = a^*\,f(u_j, u_\ell) + b^*\,f(u_k, u_\ell) \tag{A.4a}$$

$$f(u_j, a\,u_k + b\,u_\ell) = a\,f(u_j, u_k) + b\,f(u_j, u_\ell) \tag{A.4b}$$

$$f(u_j, u_k) = f(u_k, u_j)^* \tag{A.4c}$$

$$f(w_j, w_k) = \delta_{jk}, \quad 1 \le j,\, k \le M. \tag{A.4d}$$

In Equations (A.4), $|u_j\rangle$, $|u_k\rangle$, and $|u_\ell\rangle$ are arbitrary vectors in $\mathcal{W}$, a and b are arbitrary complex numbers, and the superscript $*$ indicates complex conjugation. Equations (A.4) imply that the function $f(\cdot,\cdot)$ is also positive definite. That is, if $|u\rangle = \sum_k c_k\,|w_k\rangle$ is an arbitrary nonzero vector in $\mathcal{W}$, then $f(u,u) > 0$ since

$$|u\rangle = \sum_{k=1}^{M} c_k\,|w_k\rangle \neq 0 \in \mathcal{W} \text{ (arbitrary nonzero vector in } \mathcal{W}) \tag{A.5a}$$

$$f(u,u) = \left(\sum_{j=1}^{M} c_j\,|w_j\rangle\,, \sum_{k=1}^{M} c_k\,|w_k\rangle\right) \tag{A.5b}$$

$$= \sum_{j,k=1}^{M} c_j^* c_k f(w_j, w_k) \text{ by conjugate bilinearity} \tag{A.5c}$$

$$= \sum_{j=1}^{M} |c_j|^2 \text{ since } f(w_j, w_k) = \delta_{jk} \tag{A.5d}$$

$$> 0 \text{ since } |u\rangle \neq 0. \tag{A.5e}$$

Thus $f(\cdot,\cdot)$ is a valid inner product function on $\mathcal{W}$ and $\{|w_k\rangle\}$ is orthonormal with respect to this inner product. We will now show that there is an invertible linear operator L on $\mathcal{W}$ such that

$$f(L\,|u_1\rangle\,, L\,|u_2\rangle) = \langle u_1|u_2\rangle \tag{A.6}$$

for all $|u_1\rangle\,, |u_2\rangle \in \mathcal{W}$. This implies that all calculations can be made with the $f(\cdot,\cdot)$ inner product, and then the inverse of L can be used to "translate" the answers back to the $\langle\cdot|\cdot\rangle$ inner product. Note that

because the $\{|w_k\rangle\}$ form a basis for $\mathcal{W}$, to satisfy Equation (A.6) it is sufficient to have a linear operator L such that

$$f(L\,|w_j\rangle\,, L\,|w_\ell\rangle) = \langle w_j|w_\ell\rangle\,, \quad 1 \le j, \ell \le M. \tag{A.7}$$

Equation (A.6) follows from Equation (A.7) by the properties of the inner product functions $f(\cdot,\cdot)$ and $\langle\cdot|\cdot\rangle$. To find an appropriate linear operator L, let $|w_j\rangle$ and $|w_\ell\rangle$ be any two of the $\{|w_k\rangle\}$ (possibly with $j = \ell$) and write them in terms of the $\{|e_k\rangle\}$,

$$|w_j\rangle = \sum_{k=1}^{M} a_k\,|e_k\rangle\,, \quad |w_\ell\rangle = \sum_{k=1}^{M} c_k\,|e_k\rangle\,. \tag{A.8}$$

Substituting into the left- and right-hand sides of Equation (A.6) and simplifying, we find

$$\begin{aligned} f(L\,|w_j\rangle\,, L\,|w_\ell\rangle) &= f\left(\sum_{k=1}^{M} a_k L\,|e_k\rangle\,, \sum_{m=1}^{M} d_m L\,|e_m\rangle\right) \\ &= \sum_{k,m=1}^{M} a_k^*\, d_m\, f(L\,|e_k\rangle\,, L\,|e_m\rangle) \end{aligned} \tag{A.9a}$$

$$\langle w_j|w_\ell\rangle = \left\langle \sum_{k=1}^{M} a_k\,|e_k\rangle \,\middle|\, \sum_{m=1}^{M} d_m\,|e_m\rangle \right\rangle = \sum_{k=1}^{M} a_k^* d_k \tag{A.9b}$$

For the two to be equal, it is sufficient to have $(L\,|e_k\rangle\,, L\,|e_m\rangle) = \delta_{km}$ for all $1 \le k, m \le M$. One operator that satisfies this condition is the linear operator L that is also defined to satisfy

$$L\,|e_k\rangle = |w_k\rangle\,, \quad 1 \le k \le M. \tag{A.10}$$

This operator is clearly invertible, and using it to further simplify the left-hand side of Equation (A.7) leads to

$$(L\,|e_k\rangle\,, L\,|e_m\rangle) = \sum_{k,m=1}^{M} a_k^*\, d_m\, f(w_k, w_m) = \sum_{k=1}^{M} a_k^* d_k, \tag{A.11}$$

so Equation (A.7) is satisfied.

A.3 Expressions for a Synthesis Map (Section 4.2)

To see why F_0 can sometimes be written in the form $F_0 = \sum_k |f_k\rangle \langle g_k|$ where the $\{|g_k\rangle\}$ are different from the $\{|w_k\rangle\}$, note that instead of defining F_0 using Equation (4.5) we could equivalently define it according to the relation

$$F\,|w_k\rangle = |f_k\rangle\,, \quad 1 \le k \le M. \tag{A.12}$$

Equation (4.5) then follows by linearity. For Equation (A.12) to be true the $\{|g_j\rangle\}$ must satisfy the relation

$$\sum_{j=1}^{M} |f_j\rangle \langle g_j|w_k\rangle = |f_k\rangle\,, \quad 1 \le k \le M. \tag{A.13}$$

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 $\{b_k\}$ be a set of coefficients such that

$$\sum_{k=1}^{M} b_k\,|f_k\rangle = 0. \tag{A.14}$$

Then to satisfy Equation (A.13) it is sufficient to have

$$\langle g_j|w_k\rangle = \delta_{jk}\,b_k, \quad 1 \le j,k \le M, \tag{A.15}$$

where $\delta_{jk} = 1$ if $j = k$ and 0 otherwise. This is again a system of linear equations that may or may not have more than one solution depending on the frame.

A.4 The Adjoint of a Linear Transformation (Sections 4.2 and 4.3)

Assume that $\{|f_k\rangle\}$ is a given frame for $\mathcal{V}$ and that A_0 and F_0 are its analysis and synthesis maps, respectively. The adjoint of F_0, denoted by $F_0^\dagger$, is defined as the linear operator on $\mathcal{W}$ that satisfies

$$\langle u_1|F_0 u_2\rangle = \langle F_0^\dagger u_1|u_2\rangle \;\text{ for all } |u_1\rangle\,, |u_2\rangle \in \mathcal{W}. \tag{A.16}$$

We wish to show that $F_0^\dagger = A_0$. Substituting Equation (4.6) into the left-hand side of Equation (A.16) and using the linearity of the inner product, we find

$$\langle u_1 | F_0 u_2 \rangle = \sum_{k=1}^{M} \langle u_1 | f_k \rangle \langle w_k | u_2 \rangle . \tag{A.17}$$

On the other hand, since $\langle x|y\rangle = \langle y|x\rangle^*$ for any two vectors $|x\rangle, |y\rangle \in \mathcal{W}$, we have $\langle F_0^\dagger u_1 | u_2 \rangle = \langle u_2 | F_0^\dagger u_1 \rangle^*$. Substituting back into Equation (A.16) leads to

$$\langle u_2 | F_0^\dagger u_1 \rangle = \left(\sum_{k=1}^{M} \langle u_1 | f_k \rangle \langle w_k | u_2 \rangle \right)^* \tag{A.18a}$$

$$= \sum_{k=1}^{M} \langle u_2 | w_k \rangle \langle f_k | u_1 \rangle \tag{A.18b}$$

$$= \langle u_2 | \left(\sum_{k=1}^{M} |w_k\rangle \langle f_k| \right) |u_1\rangle , \tag{A.18c}$$

and since this must be true for all $|u_1\rangle, |u_2\rangle \in \mathcal{W}$, we must have

$$F_0^\dagger = \sum_{k=1}^{M} |w_k\rangle \langle f_k| = A_0. \tag{A.19}$$

Next let $\mathcal{R}$ be an arbitrary subspace of $\mathcal{W}$ and consider the orthogonal projection operator $\mathcal{P}_\mathcal{R}$ from $\mathcal{W}$ onto $\mathcal{R}$. We wish to show that $\mathcal{P}_\mathcal{R}^\dagger = \mathcal{P}_\mathcal{R}$, i.e.,

$$\langle u_1 | \mathcal{P}_\mathcal{R} u_2 \rangle = \langle \mathcal{P}_\mathcal{R}^\dagger u_1 | u_2 \rangle \text{ for all } |u_1\rangle, |u_2\rangle \in \mathcal{W}. \tag{A.20}$$

Equation (A.20) follows directly from decomposing $|u_1\rangle$ and $|u_2\rangle$ into their components in $\mathcal{R}$ and $\mathcal{R}^\perp$.

The notion of the adjoint of a linear transformation applies much more broadly beyond linear transformations acting on finite-dimensional Hilbert spaces (see, for example, [36]). We consider one extension below to linear transformations whose input and output vector spaces may be different, although we still assume that both spaces are finite-dimensional for simplicity. We will continue to use the superscript †

to denote the adjoint. Consider two finite-dimensional Hilbert spaces $\mathcal{W}_1$, $\mathcal{W}_2$ and two linear transformations $T : \mathcal{W}_1 \mapsto \mathcal{W}_2$, $R : \mathcal{W}_2 \mapsto \mathcal{W}_1$ satisfying $R = T^\dagger$. We assume for simplicity that $\mathcal{W}_1$ and $\mathcal{W}_2$ have the same inner product denoted by $\langle \cdot | \cdot \rangle$, although this is solely for notational clarity. By definition we have

$$\langle u_2 | T u_1 \rangle = \langle R u_2 | u_1 \rangle \text{ for all } |u_1\rangle \in \mathcal{W}_1, |u_2\rangle \in \mathcal{W}_2. \tag{A.21}$$

As stated in Section 4.2, taking the complex conjugate of both sides of Equation (A.21) implies that $T = R^\dagger$. It is well-known that T can always be written as a sum of rank-one operators of the form $|u_2\rangle \langle u_1|$ where $|u_1\rangle \in \mathcal{W}_1$ and $|u_2\rangle \in \mathcal{W}_2$. As an example, one possibility for expressing A in this form would be to use its singular value decomposition. It is straightforward to show using a derivation exactly analogous to the one given above that $T^\dagger$ is the same sum of rank-one operators but with each term of the form $|u_2\rangle \langle u_1|$ replaced by $|u_1\rangle \langle u_2|$.

We will now show that $N(R) = R(T)^\perp$. It will follow by symmetry that $N(T) = R(R)^\perp$. Given an arbitrary vector $|u_2\rangle \in N(R)$, $|u_2\rangle$ must also be an element of $R(T)^\perp$. To see why this is true, let $|y\rangle$ be an arbitrary vector in $R(T)$. By definition there is some $|u_1\rangle \in \mathcal{W}_1$ such that $|y\rangle = T|u_1\rangle$. Then $|u_2\rangle$ is orthogonal to $|y\rangle$,

$$\langle y | u_2 \rangle = \langle T u_1 | u_2 \rangle = \langle u_1 | R u_2 \rangle = 0. \tag{A.22}$$

Since $|y\rangle$ was arbitrary, this implies that $N(R)$ is contained in $R(T)^\perp$. On the other hand, let $|u_2\rangle \in \mathcal{W}_2$ be an arbitrary element of $R(T)^\perp$. By definition it must satisfy $\langle y | u_2 \rangle = 0$ for all $|y\rangle \in R(T)$, i.e., $\langle T u_1 | u_2 \rangle = 0$ for all $|u_1\rangle \in \mathcal{W}_1$. Then $R|u_2\rangle$ must equal 0,

$$\langle T u_1 | u_2 \rangle = 0 \text{ for all } |u_1\rangle \in \mathcal{W}_1 \tag{A.23a}$$

$$\langle u_1 | R u_2 \rangle = 0 \text{ for all } |u_1\rangle \in \mathcal{W}_1 \tag{A.23b}$$

$$R|u_2\rangle = 0. \tag{A.23c}$$

This implies that $R(T)^\perp$ is contained in $N(R)$, and because the reverse is also true it must be that the two are identical.

A.5 The Canonical Dual Frame (Sections 4.3 and 4.6)

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

$$\underset{\tilde{A}_0 : \mathcal{V} \to \mathcal{W}}{\text{minimize}} \quad ||\tilde{A}_0 |v\rangle||^2 \tag{A.24a}$$

$$\text{subject to} \quad F_0 \tilde{A}_0 |v\rangle = |v\rangle \tag{A.24b}$$

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 $R(A_0)^\perp = N(F_0)$,

$$\tilde{A}_0 |v\rangle = |w_1\rangle + |w_2\rangle \tag{A.25}$$

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 (A.25) 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 (A.24b) 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 (A.24b) implies that $\dim R(\tilde{A}_0) \geq N$. Since $\dim R(A_0) = N$ according to Section 4.2, 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 (A.24b), we find that $F_0 \tilde{A}_0 |v\rangle = F_0 A_0 |x\rangle$. It is straightforward to show that the operator $(F_0 A_0)$, often referred to as the frame operator of $\{|f_k\rangle\}$, is always invertible. Thus, $|x\rangle = (F_0 A_0)^{-1} |v\rangle$ and so $\tilde{A}_0 |v\rangle = A_0 |x\rangle = A_0 (F_0 A_0)^{-1} |v\rangle$. Again using the fact that $|v\rangle$ was assumed to be arbitrary, this implies that $\tilde{A}_0 = A_0 (F_0 A_0)^{-1}$, which is exactly equal to the analysis map of the canonical dual frame.

Next we wish to show that of all dual frames, the canonical dual frame minimizes the expected reconstruction error $\mathcal{E}$ as defined in

Equation (4.33). Note that the derivation given below does not assume that the analysis frame is an ENTF. The problem can be formulated as

$$\underset{\tilde{F}_0 : \mathcal{W} \to \mathcal{V}}{\text{minimize}} \quad \mathbb{E}\left[||\tilde{F}_0 \, |w_e\rangle \, ||^2\right] \tag{A.26a}$$

$$\text{subject to} \quad \tilde{F}_0 \, A_0 \, |v\rangle = |v\rangle \text{ for all } |v\rangle \in \mathcal{V} \tag{A.26b}$$

where the minimization is performed over all linear operators $\tilde{F}_0$ from $\mathcal{W}$ to $\mathcal{V}$. Equation (A.26b), which in effect specifies that $\tilde{F}_0$ must be the synthesis operator of a frame that is dual to the analysis frame, amounts to the requirement that $\tilde{F}_0$ is a left-inverse of A_0. A left-inverse is guaranteed to exist because as stated in Section 4.2, A_0 has rank N.

Let $\tilde{F}_0$ be an arbitrary left-inverse of A_0 and assume that $\{|w_k\rangle \, , \, 1 \leq k \leq M\}$ is an orthonormal basis for $\mathcal{W}$. Further assume that the $\{|w_k\rangle\}$ can be partitioned into an orthonormal $\{|w_k\rangle \, , \, 1 \leq k \leq N\}$ for $R(A_0)$ and an orthonormal $\{|w_k\rangle \, , \, N+1 \leq k \leq M\}$ for $R(A_0)^{\perp}$. To fully specify the operator $\tilde{F}_0$, it is both necessary and sufficient to specify its action on each of the $\{|w_k\rangle\}$. Its action on $R(A_0)$ must be chosen to satisfy Equation (A.26b) while its action on $R(A_0)^{\perp}$ can be chosen to minimize $\mathbb{E}\left[||\tilde{F}_0 \, |w_e\rangle \, ||^2\right]$.

We first consider its action on $R(A_0)$. For each $\{|w_k\rangle \, , \, 1 \leq k \leq N\}$, there is a unique vector $|v_k\rangle \in \mathcal{V}$ satisfying $A_0 \, |v_k\rangle = |w_k\rangle$. Equation (A.26b) implies that $\tilde{F}_0 \, |w_k\rangle = |v_k\rangle$ for all $1 \leq k \leq N$. The action of $\tilde{F}_0$ on $R(A_0)^{\perp}$ can now be chosen to minimize $||\tilde{F}_0 \, |w_e\rangle \, ||^2$. Note that any error vector $|w_k\rangle \in \mathcal{W}$ can be written uniquely as the sum of a component $|w_1\rangle \in R(A_0)$ and a component $|w_2\rangle \in R(A_0)^{\perp}$,

$$|w_e\rangle = |w_1\rangle + |w_2\rangle = \sum_{k=1}^{N} c_k \, |w_k\rangle + \sum_{k=N+1}^{M} c_k \, |w_k\rangle \, , \tag{A.27}$$

where $\{c_k\}$ are the coefficients of $|w_e\rangle$ in the $\{|w_k\rangle\}$ basis. 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. The expected value of $||\tilde{F}_0 \, |w_e\rangle \, ||^2$ is

$$\mathbb{E}\left[||\tilde{F}_0 \, |w_e\rangle \, ||^2\right] = \mathbb{E}\left[||\tilde{F}_0 \, |w_1\rangle + \tilde{F}_0 \, |w_2\rangle \, ||^2\right] . \tag{A.28a}$$

As we will show below, the expected value is minimized when $\tilde{F}_0 |w_2\rangle$ is set to zero for all possible values of $\vec{w_2}$. The vector $\tilde{F}_0 |w_e\rangle$ is equal to

$$\tilde{F}_0 |w_e\rangle = \sum_{k=1}^{N} c_k \tilde{F}_0 |w_k\rangle + \sum_{k=N+1}^{M} c_k \tilde{F}_0 |w_k\rangle \tag{A.29a}$$

$$= \sum_{k=1}^{N} c_k |v_k\rangle + \sum_{k=N+1}^{M} c_k \tilde{F}_0 |w_k\rangle . \tag{A.29b}$$

Its squared norm is equal to $\langle \tilde{F}_0 |w_e\rangle \, |\tilde{F}_0 |w_e\rangle\rangle$, and since the $\{c_k\}$ are pairwise uncorrelated all cross terms are equal to zero. Thus,

$$\mathbb{E}\left[||\tilde{F}_0 |w_e\rangle ||^2 \right] = \mathbb{E}\left[\sum_{k=1}^{N} c_k^2 ||v_k||^2 + \sum_{k=N+1}^{M} c_k^2 ||\tilde{F}_0 |w_k\rangle ||^2 \right] \tag{A.30a}$$

$$= \sum_{k=1}^{N} \mathbb{E}[c_k^2] \, ||v_k||^2 + \sum_{k=N+1}^{M} \mathbb{E}[c_k^2] \, ||\tilde{F}_0 |w_k\rangle ||^2 \tag{A.30b}$$

$$= \sigma^2 \sum_{k=1}^{N} ||v_k||^2 + \sigma^2 \sum_{k=N+1}^{M} ||\tilde{F}_0 |w_e\rangle ||^2. \tag{A.30c}$$

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 $\tilde{F}_0 |w_k\rangle = 0$ for all $N+1 \leq k \leq M$. Thus, the optimal left-inverse $\tilde{F}_0$ inverts A_0 over its range and acts as the zero operator on $R(A_0)^{\perp}$. The unique left-inverse with these properties is the Moore-Penrose pseudoinverse of A_0 (see, for example, Section 1 of [39]). Explicitly, the pseudoinverse is equal to

$$A_0^* = (A_0^\dagger A_0)^{-1} A_0^\dagger = (F_0 A_0)^{-1} F_0, \tag{A.31}$$

and this corresponds exactly to the synthesis operator of the canonical dual frame [39].

A.6 Naimark's Theorem (Section 4.4.2)

Let $\{|f_k\rangle\}$ be an arbitrary frame for $\mathcal{V}$ and assume that there exists and orthonormal basis $\{|w_k\rangle\}$ for $\mathcal{W}$ that satisfies $\mathcal{P}_{\mathcal{V}} |w_k\rangle = |f_k\rangle$ for

$1 \leq k \leq M$. As explained in Section 4.4.3, an arbitrary vector $|v\rangle \in \mathcal{V}$ can always be written as

$$|v\rangle = \sum_{k=1}^{M} \langle w_k|v\rangle \, |w_k\rangle = \sum_{k=1}^{M} b_k \, |w_k\rangle \,, \tag{A.32}$$

where we have defined $b_k = \langle w_k|v\rangle$. Since the $\{|w_k\rangle\}$ are an orthonormal basis for $\mathcal{W}$, the squared norm of $|v\rangle$ is equal to the sum of the squared magnitudes of the $\{b_k\}$, $||v||^2 = \sum_k |b_k|^2$. On the other hand and as also explained in Section 4.4.3, since the $\{|w_k\rangle\}$ satisfy the property $\mathcal{P}_{\mathcal{V}} \, |w_k\rangle = |f_k\rangle$ for $1 \leq k \leq M$, we also have

$$b_k = \langle w_k|v\rangle = \langle f_k|v\rangle \,, \quad 1 \leq k \leq M. \tag{A.33}$$

Thus,

$$\sum_{k=1}^{M} |\, \langle f_k|v\rangle \,|^2 = \sum_{k=1}^{M} |b_k|^2 = ||v||^2 \text{ for all } |v\rangle \in \mathcal{V}, \tag{A.34}$$

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

A.7 An Oversampling Frame in Classical Signal Processing (Section 4.6.2)

It is common in many classical signal processing scenarios to sample a bandlimited continuous-time (CT) signal at an integer multiple of its Nyquist rate. This tactic is sometimes referred to as oversampling [25]. In Appendix A.7 we verify explicitly that a particular set of shifted sinc functions forms an ENTF for a space of bandlimited CT signals. Assume that $\mathcal{V}$ is the set of all finite-energy CT signals bandlimited to $(-\Omega_N, \Omega_N)$ and let $T = \pi/(r\,\Omega_N)$ for an arbitrary positive integer r. We wish to show that $\{f_k(t)\}$, where

$$f_k(t) = \frac{\sin(\Omega_N(t - kT))}{\pi(t - kT)}, \quad k \text{ an integer}, \tag{A.35}$$

is an ENTF for $\mathcal{V}$ with frame bound $C = 1/T$ and $||f_k|| = (rT)^{-1/2}$. By definition, this means that

$$\sum_{k=-\infty}^{\infty} |\langle f_k(t), v(t)\rangle|^2 = \frac{||v(t)||^2}{T} \quad \text{for all } v(t) \in \mathcal{V} \tag{A.36a}$$

$$||f_k(t)|| = \frac{1}{\sqrt{r\,T}} \quad \text{for all } k. \tag{A.36b}$$

Note that we are assuming the following standard inner product on $\mathcal{V}$,

$$\langle v_1(t), v_2(t)\rangle = \int_{-\infty}^{\infty} dt\, v_1(t)\, v_2(t) \quad \text{for all } v_1(t),\, v_2(t) \in \mathcal{V}. \tag{A.37}$$

To verify that $\{f_k(t)\}$ satisfies Equation (A.36a), consider $f_k(t)$ for a specific value of k and an arbitrary element $v(t) \in \mathcal{V}$ with CT Fourier transform (CTFT) $V(j\Omega)$. Since $v(t)$ is an element of $\mathcal{V}$, $V(j\Omega)$ is only nonzero for $|\Omega| \leq \Omega_N$. We will first show that the inner product of $f_k(t)$ with $v(t)$ is equal to $v(t)$ sampled at time $t = kT$, i.e., $\langle f_k(t), v(t)\rangle = v(kT)$. Then we will use Parseval's theorem to show that

$$\sum_{k=-\infty}^{\infty} |\langle f_k(t), v(t)\rangle|^2 = \sum_{k=-\infty}^{\infty} |v(kT)|^2 = \frac{||v(t)||^2}{T}. \tag{A.38}$$

Let $y(t)$ be the convolution of $f_k(t)$ with $v(t)$,

$$y(t) = \int_{-\infty}^{\infty} d\tau\, v(\tau)\, f_k(t-\tau). \tag{A.39}$$

It is well-known that the CTFT of $y(t)$ is $Y(j\Omega) = F_k(j\Omega)\, V(j\Omega)$ where $F_k(j\Omega)$ is the CTFT of $f_k(t)$, defined by

$$F_k(j\Omega) = \begin{cases} e^{-j\,\Omega\,kT} \text{ if } |\Omega| \leq \Omega_N, \\ 0 \text{ else.} \end{cases} \tag{A.40}$$

Since $f_k(t) = f_k(-t)$, the inner product of $f_k(t)$ with $v(t)$ is equal to $y(t)$ evaluated at $t = 0$,

$$\langle f_k(t), v(t)\rangle = \int_{-\infty}^{\infty} dt\, f_k(t)\, v(t) = \int_{-\infty}^{\infty} dt\, f_k(-t)\, v(t) = y(0). \tag{A.41}$$

Using the definition of the inverse CTFT, we may express $y(0)$ as

$$y(0) = \left[\frac{1}{2\pi}\int_{-\infty}^{\infty} d\Omega\, Y(j\Omega)\, e^{j\Omega t}\right]_{t=0} = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\Omega\, Y(j\Omega) \tag{A.42a}$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} d\Omega\, F_k(j\Omega)\, V(j\Omega). \tag{A.42b}$$

And now substituting Equation (A.40) into Equation (A.42) yields

$$\langle f_k(t), v(t)\rangle = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\Omega\, V(j\Omega)\, e^{-j\,\Omega\, kT} = v(kT), \tag{A.43}$$

where we have again used the definition of the inverse CTFT. Next we use Parseval's theorem to show that $\sum_k |v(kT)|^2 = ||v(t)||^2/T$. The discrete time Fourier transform of the sequence $\{v(kT)\}$, denoted by $\hat{V}(e^{j\omega})$, is 2π-periodic and is related to $V(j\Omega)$ via

$$\hat{V}(e^{j\omega}) = \frac{1}{T} V\left(j\frac{\omega}{T}\right), \quad -\pi < \omega \leq \pi. \tag{A.44}$$

Parseval's theorem for discrete time sequences states that

$$\sum_{k=-\infty}^{\infty} |v(kT)|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} d\omega \left|\hat{V}(e^{j\omega})\right|^2. \tag{A.45}$$

Substituting Equation (A.44) into Equation (A.45) and changing the variable of integration to $\Omega = \omega/T$, we find

$$\sum_{k=-\infty}^{\infty} |v(kT)|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} d\omega\, \frac{1}{T^2}\left|V\left(j\frac{\omega}{T}\right)\right|^2 = \frac{1}{2\pi}\int_{-\pi/T}^{\pi/T} d\Omega\, \frac{1}{T}\,|V(j\Omega)|^2 \tag{A.46a}$$

$$= \frac{1}{2\pi\, T}\int_{-\infty}^{\infty} d\Omega\, |V(j\Omega)|^2 \tag{A.46b}$$

Note that in going from Equation (A.46a) to (A.46b), we have used the fact that $\pi/T = r\,\Omega_N$ and the fact that by assumption, $V(j\Omega)$ is only nonzero for $|\Omega| \leq \Omega_N$. Finally, Parseval's theorem for CT signals states that

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} d\Omega\, |V(j\Omega)|^2 = \int_{-\infty}^{\infty} dt\, |v(t)|^2 = ||v(t)||^2 \text{ for all } v(t) \in \mathcal{V}. \tag{A.47}$$

In summary, we have

$$\sum_{k=-\infty}^{\infty} |\langle f_k(t), v(t)\rangle|^2 = \sum_{k=-\infty}^{\infty} |v(kT)|^2 = \frac{||v(t)||^2}{T}. \tag{A.48}$$

Since Equation (A.48) is true for any $v(t) \in \mathcal{V}$, $\{f_k(t)\}$ is a tight frame for $\mathcal{V}$ with frame bound $C = 1/T$.

To verify Equation (A.36b), we use Parseval's theorem for CT signals to show that $||f_k(t)||^2$ is identical for all values of k,

$$||f_k(t)||^2 = \int_{-\infty}^{\infty} dt\, |f_k(t)|^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\Omega\, |F_k(j\Omega)|^2 \tag{A.49a}$$

$$= \frac{1}{2\pi} \int_{-\Omega_N}^{\Omega_N} d\Omega\, |e^{-j\Omega kT}|^2 = \frac{\Omega_N}{\pi}. \tag{A.49b}$$

Since $\Omega_N = \pi/(r\,T)$, we have

$$||f_k(t)|| = \frac{1}{\sqrt{rT}} \text{ for all } k. \tag{A.50}$$

A.8 Change of Basis in $\mathcal{W}$ (Section 4.6.2)

Consider $|e\rangle = \sum_k \Delta_k\, |w_k\rangle$ where the $\{\Delta_k\}$ each have zero mean and variance σ^2 and are pairwise uncorrelated, as specified in Equation (4.31). Let $\{|u_k\rangle\}$ be any orthonormal basis of $\mathcal{W}$. We wish to show that the components of $|e\rangle$ with respect to $\{|u_k\rangle\}$, which we denote by $\{\Delta'_k\}$, have these same properties. We start by expanding each of the $\{|w_k\rangle\}$ as a linear combination of the $\{|u_k\rangle\}$,

$$|w_k\rangle = \sum_{\ell=1}^{M} c_{k\ell}\, |u_k\rangle\,, \quad 1 \le k \le M. \tag{A.51}$$

Since the $\{|w_k\rangle\}$ are orthonormal, the $\{c_{k\ell}\}$ satisfy

$$\sum_{\ell=1}^{M} c_{j\ell}\, c_{k\ell} = \delta_{jk}, \quad 1 \le j, k \le M, \tag{A.52}$$

where δ_{jk} takes the value 1 if $j = k$ and 0 otherwise. Equation (A.52) implies that the $\{c_{k\ell}\}$ also satisfy

$$\sum_{k=1}^{M} c_{k\ell}\, c_{km} = \delta_{\ell m}, \quad 1 \le \ell, m \le M. \tag{A.53}$$

To see why this is true, consider the $M \times M$ matrix D whose kth row and ℓth column contains the element $c_{k\ell}$ for $1 \leq k, \ell \leq M$. Equation (A.52) states that the columns of D are orthonormal with respect to the standard inner product (often referred to as the dot product) on $\mathbb{C}^M$, the canonical M-dimensional complex coordinate space. A square matrix whose columns are orthonormal also has the property that its rows are orthonormal, which is exactly the meaning of Equation (A.53).

Substituting Equation (A.51) into the expression for $|e\rangle$ and rearranging, we find

$$|e\rangle = \sum_{k=1}^{M} \Delta_k \left(\sum_{\ell=1}^{M} c_{k\ell} \, |u_k\rangle \right) = \sum_{\ell=1}^{M} \left(\sum_{k=1}^{M} \Delta_k \, c_{k\ell} \right) |u_k\rangle \, . \tag{A.54}$$

The components $\{\Delta'_k\}$ can thus be expressed as $\Delta'_k = \sum_\ell \Delta_k \, c_{k\ell}$ for $1 \leq k \leq M$. We may now derive the desired result using the linearity of expectation and the properties of the $\{\Delta_k\}$,

$$\mathbb{E}[\Delta'_k] = \sum_{k=1}^{M} \mathbb{E}[\Delta_k] \, c_{k\ell} = 0 \tag{A.55a}$$

$$\mathbb{E}[\Delta'_j \, \Delta'_k] = \sum_{j,k=1}^{M} \mathbb{E}[\Delta_j \Delta_k] \, c_{j\ell} \, c_{k\ell} \tag{A.55b}$$

$$= \begin{cases} \sigma^2 \sum_{k=1}^{M} c_{k\ell}^2 \text{ if } j = k \\ 0 \text{ if } j \neq k \end{cases} \tag{A.55c}$$

$$= \begin{cases} \sigma^2 \text{ if } j = k \\ 0 \text{ if } j \neq k \end{cases} \quad .. \tag{A.55d}$$

Note that in Equation (A.55d) we have used Equation (A.53). Thus, we have shown that the $\{\Delta'_k\}$ have zero mean and variance σ^2 and are pairwise uncorrelcated.

The fact that the $\{\Delta'_k\}$ are uncorrelated can also be interpreted in terms of the matrix D. This viewpoint is the one most commonly used when considering a general decorrelation transformation of a vector-valued random variable. Consider the vector-valued random variable $\vec{\Delta}$ whose kth component is the random variable Δ_k for $1 \leq k \leq M$.

Similarly, let $\vec{\Delta}'$ be the vector-valued random variable with components $\{\Delta'_k\}$. Since $\Delta'_k = \sum_\ell \Delta_k \, c_{k\ell}$ for $1 \le k \le M$, we have $\vec{\Delta}' = D\,\vec{\Delta}$. Since the components of $\vec{\Delta}'$ clearly have zero mean, the covariance matrix of $\vec{\Delta}'$ can be expressed as

$$\mathbb{E}[\vec{\Delta}'(\vec{\Delta}')^T] = D\,\mathbb{E}[\vec{\Delta}\,\vec{\Delta}^T]\,D^T. \tag{A.56}$$

The $\{\Delta_k\}$ satisfy $\mathbb{E}[\Delta_j \Delta_k] = \sigma^2 \delta_{jk}$ by assumption, implying that $\mathbb{E}[\vec{\Delta}\vec{\Delta}^T] = \sigma^2 I_M$ where I_M is the $M \times M$ identity matrix. Thus, Equation (A.56) can be simplified to

$$\mathbb{E}[\vec{\Delta}'(\vec{\Delta}')^T] = \sigma^2\, D\, I_M\, D^T = \sigma^2 DD^T = \sigma^2 I_M, \tag{A.57}$$

where in the last line we have used that fact that $DD^T = I_M$ because the rows of D are orthonormal. The component in the jth row and kth column of the matrix $\mathbb{E}[\vec{\Delta}'(\vec{\Delta}')^T]$ is $\mathbb{E}[\Delta'_j \Delta'_k]$, and Equation (A.56) states that it is equal to σ^2 if $j = k$ and 0 otherwise, as expected.

A.9 Generalized Operator Frames (Section 4.5.2)

The definition of a special class of IC POVMs referred to as tight IC POVMs relies on the notion of a generalized operator frame with respect to (w.r.t.) a given measure, as introduced in [11]. Given a measure $\alpha(\cdot)$ that maps each $1 \le k \le M$ to a non-negative number $\alpha(k) \ge 0$, a set of operators $\{|F_k\rangle\rangle\}$ in $\mathcal{V}$ is a generalized operator frame for $\mathcal{V}$ w.r.t. $\alpha(\cdot)$ if

$$C\,||V||^2 \le \sum_{k=1}^{M} \alpha(k)\,|\langle\langle F_k|V\rangle\rangle|^2 \le D\,||V||^2 \text{ for all } |V\rangle\rangle \in \mathcal{V}, \tag{A.58}$$

for some $0 < C \le D < \infty$.

Example A.1. Equation (4.26) is a special case of Equation (A.58) in which $\alpha(\cdot)$ is the counting measure, defined by

$$\alpha(k) = 1, \quad 1 \le k \le M. \tag{A.59}$$

Thus, a set of operators satisfying Equation (4.26) is a frame for $\mathcal{V}$ w.r.t. the counting measure.

Example A.2. The trace measure [11] is defined by

$$\alpha(k) = \mathrm{Tr}(F_k), \quad 1 \leq k \leq M. \tag{A.60}$$

Note that Equation (A.60) only represents a valid measure when $\mathrm{Tr}(F_k) \geq 0$ for all values of k. One instance in which this is true is when the $\{|F_k\rangle\rangle\}$ are the elements of a POVM. Substituting Equation (A.60) into Equation (A.58) leads to

$$C\,||V||^2 \leq \sum_{k=1}^{M} \mathrm{Tr}(F_k)\,|\langle\langle F_k|V\rangle\rangle|^2 \leq D\,||V||^2 \text{ for all } |V\rangle\rangle \in \mathcal{V} \tag{A.61}$$

for some $0 < C \leq D < \infty$. A set of operators $\{|F_k\rangle\rangle\}$ in $\mathcal{V}$ satisfying Equation (A.61) is a frame for $\mathcal{V}$ w.r.t. the trace measure. A tight frame for $\mathcal{V}$ w.r.t. the trace measure is one for which the upper and lower bounds in Equation (A.61) can both be set to the same value. Note that in finite dimensions, if $\{|F_k\rangle\rangle\}$ is a frame for $\mathcal{V}$ w.r.t. the trace measure, then $\{\sqrt{|\,\mathrm{Tr}(F_k)|}\;|F_k\rangle\rangle\}$ is a frame for $\mathcal{V}$ w.r.t. the counting measure.

A.10 Distribution of Relative Frequencies (Section 5.3)

While the following derivation is motivated by the quantum state estimation problem considered in Section 5.3, the concepts and conclusions rely only on the laws of probability and not on the postulates of quantum mechanics. Therefore we state the results without any reference to density operators or quantum measurement. Let X be a discrete random variable that takes values in the set $\{1, \ldots, M\}$ with probability mass function (PMF) $\{p(1), \ldots, p(M)\}$, i.e.,

$$X = k \text{ with probability } p(k),\ 1 \leq k \leq M. \tag{A.62}$$

Assume that $\{x_i, 1 \leq i \leq L\}$ is a set of L independent realizations of X and consider the set of relative frequencies $\{\hat{p}(k) = \ell_k/L\}$, where ℓ_k is the number of realizations $\{x_i\}$ that are equal to k. Defining $d_k = \hat{p}(k) - p(k)$ for $1 \leq k \leq M$, the goal is to evaluate the expected values $\mathbb{E}[d_k]$ and $\mathbb{E}[d_j d_k]$ for all $1 \leq j, k \leq M$.

We first address the case where $j = k$. Let k be a fixed integer between 1 and M. To compute $\mathbb{E}[d_k]$, note that the value of ℓ_k is

binomially distributed with parameters $p(k)$ and L [ref]. Its expected value is $\mathbb{E}[\ell_k] = L\,p(k)$ and its variance is $\mathrm{var}(\ell_k) = L\,p(k)\,(1-p(k))$. Using linearity of expectation we find that, unsurprisingly, the expected value of d_k is equal to zero,

$$\mathbb{E}[d_k] = \mathbb{E}\left[p(k) - \frac{\ell_k}{L}\right] = p(k) - \frac{L\,p(k)}{L} = 0. \tag{A.63}$$

The variance of d_k is

$$\mathrm{var}(d_k) = \mathrm{var}\left(p(k) - \frac{\ell_k}{L}\right) = \frac{\mathrm{var}(\ell_k)}{L^2} = \frac{p(k)\,(1-p(k))}{L}. \tag{A.64}$$

Furthermore, since $\mathbb{E}[d_k] = 0$ we have $\mathbb{E}[d_k^2] = \mathrm{var}(d_k)$.

Now let j and k be fixed integers between 1 and M with $j \neq k$. To compute $\mathbb{E}[d_j\,d_k]$, note that the joint distribution of $\{\ell_1, \ldots, \ell_M\}$ is given by a multinomial distribution with parameters L and $\{p_1, \ldots, p_M\}$ [ref]. It can be shown using the properties of the multinomial distribution that

$$\mathbb{E}[\ell_j \ell_k] = L\,p(j)\,p(k)\,(L-1). \tag{A.65}$$

Using linearity of expectation and the fact that $\mathbb{E}[\ell_j] = L\,p(j)$ and $\mathbb{E}[\ell_k] = L\,p(k)$, we find that the value of $\mathbb{E}[d_j d_k]$ is

$$\mathbb{E}[d_j d_k] = \mathbb{E}\left[\left(\frac{\ell_j}{L} - p(j)\right)\left(\frac{\ell_k}{L} - p(k)\right)\right] \tag{A.66a}$$

$$= \frac{\mathbb{E}[\ell_j \ell_k]}{L^2} - p(j)\,p(k) = -\frac{p(j)\,p(k)}{L}. \tag{A.66b}$$

B

Traditional Appendices

B.1 Generation of P_f-P_d Projection of LRT ROC from Suboptimal SVT ROC

An explanation of the procedure is shown in Figure B.1. The graphs in Figure B.1a show P_f^{SVT}, P_d^{SVT}, and the derivative $dP_d^{\mathrm{SVT}}/dP_f^{\mathrm{SVT}}$ as functions of the score variable. According to Equation (2.31) 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 $\mathcal{D}_{\mathrm{LRT}}(\eta_0)$. Each individual region j covers an interval $[a_j, b_j]$ with $a_j < b_j$ and corresponds to the segment in the P_f-P_d projection of the SVT ROC 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 B.1b,

can be expressed as

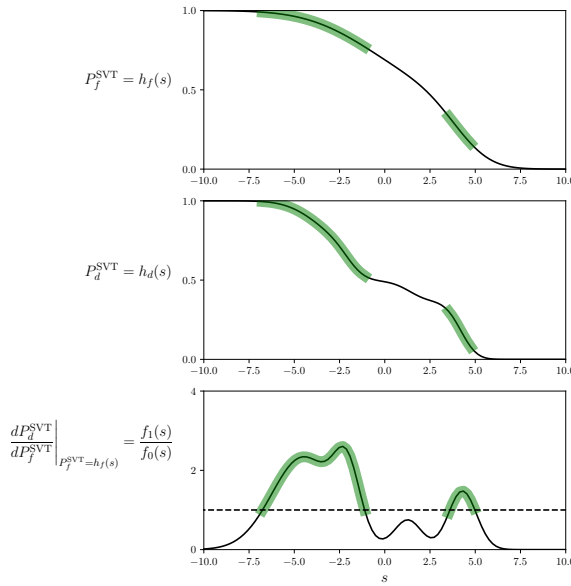
$$\int_{a_j}^{b_j} ds\, f_0(s) = (1 - F_0(a_j)) - (1 - F_0(b_j)) = h_f(a_j) - h_f(b_j) \quad \text{(B.1a)}$$

$$\int_{a_j}^{b_j} ds\, f_1(s) = (1 - F_1(a_j)) - (1 - F_1(b_j)) = h_d(a_j) - h_d(b_j) \quad \text{(B.1b)}$$

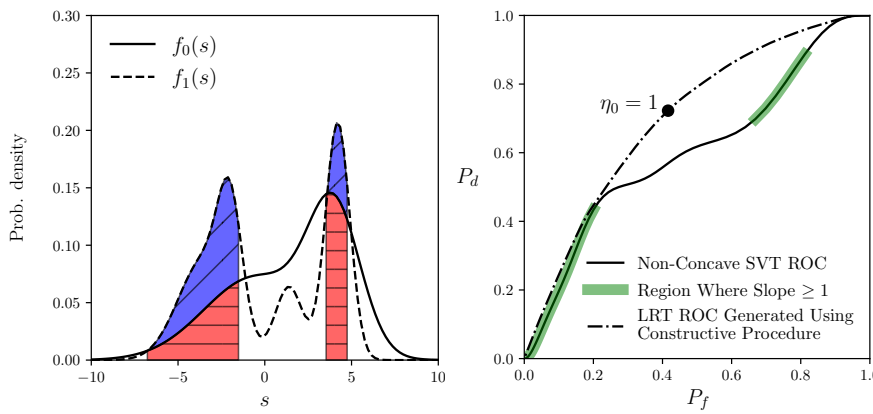
which are simply the changes in P_f^{SVT} and P_d^{SVT} between the endpoints of the segment. Summing these changes over all regions corresponds to summing the integrals of $f_0(\cdot)$ and $f_1(\cdot)$ over each disjoint portion of $\mathcal{D}_{\text{LRT}}(\eta)$. The resulting P_f-P_d projection of the LRT ROC made by varying η_0 over its entire range is illustrated in Figure B.1c.

B.2 QMOCs Generated using Standard Measurements are Ellipses

We show that any QMOC generated according to the method described in Example 3.3 of Section 3.6, 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 (3.19), 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 H_1 if the measurement outcome associated with E_2 occurs and 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 $d = 2$.



(a)



(b) **(c)**

Figure B.1: (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) P_f-P_d projections of non-concave SVT ROC and LRT ROC generated using the procedure given in the text.

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

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

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

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,

$$P_f - \frac{1}{2} = \frac{a_0 - a_1}{2}\cos\theta \tag{B.3a}$$

$$P_f - \frac{1}{2} = \frac{b_0 - b_1}{2}\cos(\theta-\alpha). \tag{B.3b}$$

For ease of notation we now make the substitutions

$$x = P_f - \frac{1}{2}, \quad y = P_d - \frac{1}{2}, \quad a = \frac{a_0 - a_1}{2}, \quad b = \frac{b_0 - b_1}{2}, \tag{B.4}$$

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

$$x = f_x(\theta) = a\cos\theta \tag{B.5a}$$

$$y = f_y(\theta) = b\cos(\theta-\alpha). \tag{B.5b}$$

(Note that the x and y above should not be confused with the $\{|x_i\rangle\}$ and $\{|y_i\rangle\}$ in Equations (3.11).) 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

$$x = g_x(t) = q\cos\beta\cos t - r\sin\beta\sin t \tag{B.6a}$$

$$y = g_y(t) = q\sin\beta\cos t + r\cos\beta\sin t \tag{B.6b}$$

for some angle of rotation β 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, β in terms of the

known values of a, b, α by using Equations (B.5) and (B.6) 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 $\theta_y = \alpha$ and has coordinates $(f_x(\alpha), f_y(\alpha)) = (a\cos\alpha, b)$. Similarly, the point on the ellipse described by Equations (B.6) with maximum x-value occurs at $t_x = \tan^{-1}(-(r/q)\tan\beta)$ and has coordinates

$$g_x(t_x) = \sqrt{q^2\cos^2\beta + r^2\sin^2\beta} \tag{B.7a}$$

$$g_y(t_x) = \frac{q^2 - r^2}{\sqrt{q^2/\sin^2\beta + r^2/\cos^2\beta}}. \tag{B.7b}$$

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

$$g_x(t_y) = \frac{q^2 - r^2}{\sqrt{q^2/\cos^2\beta + r^2/\sin^2\beta}} \tag{B.8a}$$

$$g_y(t_y) = \sqrt{q^2\sin^2\beta + r^2\cos^2\beta}. \tag{B.8b}$$

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

$$\beta = \frac{1}{2}\tan^{-1}\left(\frac{2ab\cos\alpha}{a^2 - b^2}\right) \tag{B.9a}$$

$$q = \left[\frac{1}{2}\left(a^2 + b^2 + \frac{a^2 - b^2}{\cos(2\beta)}\right)\right]^{1/2} \tag{B.9b}$$

$$r = \left[\frac{1}{2}\left(a^2 + b^2 - \frac{a^2 - b^2}{\cos(2\beta)}\right)\right]^{1/2}. \tag{B.9c}$$

It can be verified through straightforward algebra that when β, q, and r are given by Equations (B.9), the coordinates $x = f_x(\theta)$, $y = f_y(\theta)$ in Equation (B.5) satisfy the equation that defines an ellipse:

$Ax^2 + Bxy + Cy^2 + D = 0$ with $B^2 - 4AC < 0$, where

$$A = q^2 \sin^2 \beta + r^2 \cos^2 \beta \tag{B.10a}$$

$$B = 2(q^2 - r^2) \sin \beta \cos \beta \tag{B.10b}$$

$$C = q^2 \cos^2 \beta + r^2 \sin^2 \beta \tag{B.10c}$$

$$D = -q^2 r^2. \tag{B.10d}$$

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 β 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 $(1/2, 1/2)$.

References

[1] J. A. Swets, R. M. Dawes, and J. Monahan, "Better decisions through science," *Scientific American*, vol. 283, no. 4, 2000, pp. 82–87.

[2] T. Fawcett, "An Introduction to ROC Analysis," *Pattern Recognition Letters*, vol. 27, no. 8, 2006, pp. 861–874.

[3] N. A. Obuchowski, "Receiver Operating Characteristic Curves and Their Use in Radiology," *Radiology*, vol. 229, no. 1, 2003, pp. 3–8.

[4] M. H. Zweig and G. Campbell, "Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine," *Clinical Chemistry*, vol. 39, no. 4, 1993, pp. 561–577.

[5] K. A. Spackman, "Signal Detection Theory: Valuable Tools for Evaluating Inductive Learning," in *Proceedings of the Sixth International Workshop on Machine Learning*, Morgan Kaufmann Publishers Inc., pp. 160–163, 1989.

[6] J. R. Beck and E. K. Shultz, "The Use of Relative Operating Characteristic (ROC) Curves in Test Performance Evaluation," *Archives of Pathology & Laboratory Medicine*, vol. 110, no. 1, 1986, pp. 13–20.

[7] 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.

[8] R. M. Stein, "The Relationship Between Default Prediction and Lending Profits: Integrating ROC Analysis and Loan Pricing," *Journal of Banking & Finance*, vol. 29, no. 5, 2005, pp. 1213–1236.

[9] C. W. Helstrom, "Detection Theory and Quantum Mechanics," *Information and Control*, vol. 10, no. 3, 1967, pp. 254–291.

[10] G. M. D'Ariano, P. Perinotti, and M. F. Sacchi, "Informationally Complete Measurements and Group Representation," *Journal of Optics B: Quantum and Semiclassical Optics*, vol. 6, no. 6, 2004, S487–S491.

[11] A. J. Scott, "Tight Informationally Complete Quantum Measurements," *Journal of Physics A: Mathematical and General*, vol. 39, no. 43, 2006, pp. 13 507–13 530.

[12] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves, "Symmetric Informationally Complete Quantum Measurements," *Journal of Mathematical Physics*, vol. 45, no. 6, 2004, pp. 2171–2180.

[13] B. Bodmann and J. Haas, *A Short History of Frames and Quantum Designs*, 2017. arXiv: 1709.01958 [quant-ph].

[14] M. B. Ruskai, "Some Connections between Frames, Mutually Unbiased Bases, and POVM's in Quantum Information Theory," *Acta Applicandae Mathematicae*, vol. 108, no. 3, 2009, pp. 709–719.

[15] S. T. Flammia, A. Silberfarb, and C. M. Caves, "Minimal Informationally Complete Measurements for Pure States," *Foundations of Physics*, vol. 35, no. 12, 2005, pp. 1985–2006.

[16] J. Finkelstein, "Pure-State Informationally Complete and 'Really' Complete Measurements," *Physical Review A*, vol. 70, no. 5, 2004, 052107.

[17] D. M. Appleby, "Symmetric Informationally Complete Measurements of Arbitrary Rank," *Optics and Spectroscopy*, vol. 103, no. 3, 2007, pp. 416–428.

[18] C. A. Fuchs, M. C. Hoang, and B. C. Stacey, "The SIC Question: History and State of Play," *Axioms*, vol. 6, no. 3, 2017, 21.

[19] H. Zhu, "Quantum State Estimation and Symmetric Informationally Complete POMs," Ph.D. dissertation, PhD thesis, National University of Singapore, 2012. 1, 5, 2012.

[20] H. Zhu, "Quantum State Estimation with Informationally Overcomplete Measurements," *Physical Review A*, vol. 90, no. 1, 2014, 012115.

[21] H. Zhu, "Super-Symmetric Informationally Complete Measurements," *Annals of Physics*, vol. 362, 2015, pp. 311–326.

[22] G. M. D'Ariano and P. Perinotti, "Optimal Data Processing for Quantum Measurements," *Physical Review Letters*, vol. 98, no. 2, 2007, 020403.

[23] C. A. Medlock and A. V. Oppenheim, "Optimal ROC Curves from Score Variable Threshold Tests," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, pp. 5327–5330, 2019.

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

[25] A. V. Oppenheim and G. C. Verghese, *Signals, Systems and Inference*. Pearson, 2015.

[26] J. M. Lobo, A. Jiménez-Valverde, and R. Real, "Auc: A misleading measure of the performance of predictive distribution models," *Global Ecology and Biogeography*, vol. 17, no. 2, 2008, pp. 145–151.

[27] D. J. Hand, "Measuring Classifier Performance: A Coherent Alternative to the Area Under the ROC Curve," *Machine Learning*, vol. 77, no. 1, 2009, pp. 103–123.

[28] M. Rosenblatt, "Remarks on Some Nonparametric Estimates of a Density Function," *The Annals of Mathematical Statistics*, vol. 27, no. 3, 1956, pp. 832–837.

[29] E. Parzen, "On Estimation of a Probability Density Function and Mode," *The Annals of Mathematical Statistics*, vol. 33, no. 3, 1962, pp. 1065–1076.

[30] E. Torgersen, *Comparison of Statistical Experiments*, vol. 36. Cambridge University Press, 1991.

[31] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory*. Wiley, New York, 1968.

[32] 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*, vol. 97, pp. 43–48, 1997.

[33] C. Medlock, A. Oppenheim, I. Chuang, and Q. Ding, "Operating characteristics for binary hypothesis testing in quantum systems," in *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, IEEE, pp. 1136–1145, 2019.

[34] A. Peres, *Quantum theory: concepts and methods*, vol. 57. Springer Science & Business Media, 2006.

[35] M. A. Nielsen and I. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2016.

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

[37] A. Bodor and M. Koniorczyk, "Receiver Operation Characteristics of Quantum State Discrimination," *Journal of Russian Laser Research*, vol. 38, no. 2, 2017, pp. 150–163.

[38] P. G. Casazza, "The Art of Frame Theory," *Taiwanese Journal of Mathematics*, vol. 4, no. 2, 2000, pp. 129–201.

[39] P. G. Casazza and G. Kutyniok, *Finite Frames: Theory and Applications*. Springer, 2012.

[40] P. G. Casazza, G. Kutyniok, and F. Philipp, "Introduction to finite frame theory," in *Finite frames*, Springer, 2013, pp. 1–53.

[41] J. Kovacevic and A. Chebira, "Life Beyond Bases: The Advent of Frames (Part I)," *IEEE Signal Processing Magazine*, vol. 24, no. 4, 2007, pp. 86–104.

[42] J. Kovacevic and A. Chebira, "Life Beyond Bases: The Advent of Frames (Part II)," *IEEE Signal Processing Magazine*, vol. 24, no. 4, 2007, pp. 86–104.

[43] S. Banerjee and A. Roy, *Linear Algebra and Matrix Analysis for Statistics*. CRC Press, 2014.

[44] G. Strang, *Introduction to Linear Algebra"*. Wellesley-Cambridge Press, 2016.

[45] P. G. Casazza, M. Fickus, D. G. Mixon, J. Peterson, and I. Smalyanau, "Every Hilbert Space Frame has a Naimark Complement," *Journal of Mathematical Analysis and Applications*, vol. 406, no. 1, 2013, pp. 111–119.

[46] M. G. Paris, "The modern tools of quantum mechanics," *The European Physical Journal Special Topics*, vol. 203, no. 1, 2012, pp. 61–86.

[47] A. Peres, "Neumark's theorem and quantum inseparability," *Foundations of Physics*, vol. 20, no. 12, 1990, pp. 1441–1453.

[48] V. K. Goyal, M. Vetterli, and N. T. Thao, "Quantized Overcomplete Expansions in $\mathbb{R}^N$: Analysis, Synthesis, and Algorithms," *IEEE Transactions on Information Theory*, vol. 44, no. 1, 1998, pp. 16–31.

[49] P. G. Casazza and J. Kovačević, "Equal-Norm Tight Frames with Erasures," *Advances in Computational Mathematics*, vol. 18, no. 2-4, 2003, pp. 387–430.

[50] C. M. Caves, C. A. Fuchs, and R. Schack, "Unknown Quantum States: The Quantum de Finetti Representation," *Journal of Mathematical Physics*, vol. 43, no. 9, 2002, pp. 4537–4559.

[51] J. Řeháček, Y. S. Teo, and Z. Hradil, "Determining Which Quantum Measurement Performs Better for State Estimation," *Physical Review A*, vol. 92, no. 1, 2015, 012108.

[52] R. B. A. Adamson and A. M. Steinberg, "Improving Quantum State Estimation with Mutually Unbiased Bases," *Physical Review Letters*, vol. 105, no. 3, 2010, 030406.

[53] T. Decker, D. Janzing, and T. Beth, "Quantum Circuits for Single-Qubit Measurements Corresponding to Platonic Solids," *International Journal of Quantum Information*, vol. 2, no. 03, 2004, pp. 353–377.

[54] W. Słomczyński and A. Szymusiak, "Highly Symmetric POVMs and Their Informational Power," *Quantum Information Processing*, vol. 15, no. 1, 2016, pp. 565–606.

[55] S. Brandsen, M. Dall'Arno, and A. Szymusiak, "Communication capacity of mixed quantum t-designs," *Physical Review A*, vol. 94, no. 2, 2016, pp. 022335-1–022335-8.

[56] C. Medlock, A. Oppenheim, and P. Boufounos, *Informationally overcomplete povms for quantum state estimation and binary detection*, 2020. arXiv: 2012.05355 [quant-ph].

[57] Q. Ding, C. A. Medlock, and A. V. Oppenheim, "POVM design for quantum state discrimination," *2021 Asilomar Conference on Signals, Systems, and Computing*, forthcoming.

[58] D. P. Hardin, T. Michaels, and E. B. Saff, "A Comparison of Popular Point Configurations on $\mathbb{S}^2$," *arXiv:1607.04590*, 2016.

[59] E. B. Saff and A. B. Kuijlaars, "Distributing Many Points on a Sphere," *The mathematical intelligencer*, vol. 19, no. 1, 1997, pp. 5–11.

[60] P. C. Leopardi, "Distributing Points on the Sphere: Partitions, Separation, Quadrature and Energy," Ph.D. dissertation, University of New South Wales, Sydney, Australia, 2007.

[61] N. J. A. Sloane, *Spherical Codes: Nice Arrangements of Points on a Sphere in Various Dimensions*. [Online]. Available: http://neilsloane.com/packings/.